PARTICIPATORY APPROACHES TO A NEW ETHICAL AND LEGAL FRAMEWORK FOR ICT

Lignes directrices sur la protection des données Questions éthiques et juridiques dans la recherche et l'innovation en matière de TIC.

INTELLIGENCE ARTIFICIELLE (IA)

Ce projet a reçu un financement du programme de recherche et d'innovation Horizon 2020 de l'Union européenne sous la convention de subvention n° 788039. Ce document ne reflète que le point de vue de l'auteur et l'Agence n'est pas responsable de l'usage qui pourrait être fait des informations qu'il contient.

IA : exigences pour les développeurs et les innovateurs

Iñigo de Miguel Beriain[386] *(UPV/EHU), Felix Schaber*[387] *(OEAW), Gianclaudio Malgieri et Andrés Chomczyk Penedo*[388] *(VUB)*

Remerciements : Les auteurs remercient José Antonio Castillo Parrilla, Eduard Fosch Villaronga et Lorena Perez Campillo pour leur aimable soutien dans la rédaction de ce document. Il va sans dire que toutes les erreurs sont de notre entière responsabilité.

Cette partie des lignes directrices a été revue et validée par Marko Sijan, conseiller principal spécialiste (DPA RH).

Introduction partie A

La première partie de ce chapitre s'articule autour des sept exigences éthiques incluses dans les recommandations publiées par le groupe d'experts de haut niveau sur l'IA ([389]) dans son document "Ethics guidelines for trustworthy AI".[390] Ces recommandations ont été récemment analysées dans le cadre du projet SHERPA,[391] , qui comprend une analyse approfondie des questions éthiques liées au développement d'outils adéquats pour relever ces défis. À la lumière de cet effort louable, il serait redondant d'inclure ici une analyse approfondie du même sujet (IA) du même point de vue (éthique). Au lieu de cela, ce que nous avons essayé de faire dans ce document est de fournir une analyse complémentaire. Ces lignes directrices visent à trouver le chevauchement entre les

[386] Auteur de l'ensemble du document à l'exception des sections 2 et 7 de cette partie.

[387] Auteur de la section 2 de cette partie.

[388] Les auteurs de la section 7 de cette partie.

[389] Le groupe d'experts de haut niveau sur l'IA a été créé par la Commission européenne en 2018. Il a pour objectif général de soutenir la mise en œuvre de la stratégie européenne sur l'intelligence artificielle. Cela inclut l'élaboration de recommandations sur l'élaboration de politiques liées à l'avenir et sur les questions éthiques, juridiques et sociétales liées à l'IA, y compris les défis socio-économiques. Disponible à l'adresse : https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence (consulté le 15 mai 2020).

[390] Groupe d'experts de haut niveau sur l'IA (2019) Lignes directrices en matière d'éthique pour une IA digne de confiance, p.15 et suivantes. Bruxelles, Commission européenne, Bruxelles. Disponible à l'adresse : https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (consulté le 15 mai 2020).

[391] SHERPA (2019) Lignes directrices pour l'utilisation éthique de l'IA et des systèmes de big data. SHERPA, www.project-sherpa.eu/wp-content/uploads/2019/12/use-final.pdf (consulté le 5 mai 2020).

recommandations éthiques formulées par le groupe d'experts de haut niveau sur l'IA et le cadre juridique créé par le règlement général sur la protection des données (RGPD) sur les questions de protection des données.

Avant de commencer l'analyse, il est toutefois nécessaire d'inclure quelques notes préliminaires. Tout d'abord, ce rapport se concentre principalement sur les développeurs d'IA : les organisations désireuses de développer des outils d'IA. Ces organisations deviennent des responsables du traitement dès qu'elles commencent à traiter des données à caractère personnel. Dans le même ordre d'idées, les termes "outil", "solution", "modèle" et "développement" doivent être considérés comme synonymes dans le contexte de cette analyse.

Deuxièmement, cette partie des lignes directrices ne peut être comprise que dans le contexte de l'ensemble de l'outil (les lignes directrices). Plusieurs concepts ne sont pas abordés dans ce document, car ils sont traités dans d'autres sections des lignes directrices ; nous y avons fait référence lorsque cela était nécessaire (les références sont surlignées en jaune). À l'avenir, toutes les sections seront disponibles sur un site web, ce qui rendra les lignes directrices beaucoup plus conviviales.

Les différentes sections de cette partie du document ne contiennent que ce que nous considérons comme strictement nécessaire pour comprendre les arguments fondamentaux des questions éthiques et juridiques en jeu. Nous avons inclus des listes de contrôle qui devraient aider les responsables du traitement à déterminer s'ils abordent les questions avec précision, ainsi qu'une section de lectures complémentaires que les lecteurs pourront consulter si nécessaire. Des notes de bas de page fournissent des références supplémentaires aux déclarations les plus importantes.

Enfin, ce document a été structuré de manière à être facile à comprendre. Comme mentionné précédemment, il est basé sur les sept exigences décrites par le groupe d'experts de haut niveau sur l'IA. Nous commençons notre analyse par une brève description des principales questions éthiques en jeu, puis nous résumons les principaux problèmes et défis éthiques qui y sont liés. Ceux-ci servent de base commune sur laquelle notre analyse juridique est construite, fournissant le contexte de l'analyse juridique effectuée.

1 Capacité d'action humaine et surveillance

"Les systèmes d'IA devraient favoriser l'autonomie humaine et la prise de décision, comme le prescrit le principe du respect de l'autonomie humaine. Cela exige que les systèmes d'IA agissent à la fois comme des facilitateurs d'une société démocratique, florissante et équitable en soutenant la capacité d'action de l'utilisateur et en favorisant les droits fondamentaux et en permettant le contrôle par l'humain."

- Groupe d'experts de haut niveau sur l'IA[392]

[392] Groupe d'experts de haut niveau sur l'IA (2019) Lignes directrices en matière d'éthique pour une IA digne de confiance, p.15 et suivantes. Bruxelles, Commission européenne, Bruxelles. Disponible à l'adresse : https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (consulté le 15 mai 2020).

1.1 Principes éthiques

Cette première exigence pour le développement de l'IA repose sur trois grands principes différents : [393]

- **Les droits fondamentaux.** Les systèmes d'IA peuvent permettre ou entraver les droits fondamentaux. Dans ces circonstances, une évaluation de l'impact sur les droits fondamentaux devrait être entreprise avant de développer une solution d'IA.
- **La capacité d'action humaine.** Les utilisateurs de systèmes d'IA devraient être en mesure de prendre des décisions éclairées et autonomes à ce sujet. Les systèmes d'IA doivent aider les individus à faire des choix plus judicieux, plus éclairés et conformes à leurs objectifs. Le principe général de l'autonomie de l'utilisateur doit être au cœur de la fonctionnalité du système d'IA. Par exemple, les personnes concernées doivent savoir que leurs données pourraient être utilisées pour le profilage, si cela devait arriver. En outre, leur droit de ne pas faire l'objet d'une décision fondée uniquement sur un traitement automatisé, lorsque celui-ci produit des effets juridiques ou des effets d'une importance similaire, doit être respecté. Toutefois, il faut garder à l'esprit que cela se réfère, en général, à des fins commerciales. Il n'en va donc pas de même pour les organismes chargés de l'application de la loi qui traitent des données à caractère personnel sur une base légale et peuvent utiliser l'IA pour lutter efficacement contre différents crimes et remplir les obligations stipulées par la loi.
- **Supervision humaine.** Elle permet de s'assurer qu'un système d'IA ne porte pas atteinte à l'autonomie humaine ou ne provoque pas d'autres effets indésirables. Cette surveillance peut être assurée par divers mécanismes de gouvernance. Ceci étant dit, moins un humain peut exercer de surveillance sur un système d'IA, plus les tests doivent être approfondis et la gouvernance rigoureuse.

1.2 Dispositions du RGPD

L'exigence d'une capacité d'action humaine et d'une supervision lors du développement d'outils d'IA est clairement liée au droit d'obtenir une intervention humaine, au droit de ne pas faire l'objet d'une décision fondée uniquement sur un traitement automatisé, au droit à l'information sur la prise de décision automatique et à la logique impliquée, qui sont tous inclus dans le RGPD. Ces droits sont remis en question par l'utilisation d'outils d'IA. L'IA implique souvent une forme de traitement automatisé et, dans certains cas, les décisions sont directement prises par le modèle d'IA. En effet, il arrive que les outils d'IA apprennent et prennent des décisions sans supervision humaine et que la logique impliquée dans leurs performances soit difficile à comprendre. [394]

À cet égard, **le profilage est particulièrement problématique dans le développement de l'IA** (voir encadré 1), car le processus de profilage "est souvent invisible pour la

---

[393] Ibid. p.15.

[394] Burrell, J. (2016) " How the machine 'thinks' : understanding opacity in machine learning algorithms ", *Big Data & Society* 3(1) : 1-12.

personne concernée. Il fonctionne en créant des données dérivées ou déduites sur les individus - de "nouvelles" données personnelles qui n'ont pas été fournies directement par les personnes concernées elles-mêmes. Les gens ont des niveaux de compréhension très différents de ce sujet et peuvent trouver difficile de comprendre les techniques sophistiquées impliquées dans les processus de profilage et de prise de décision automatisée"[395] (voir section 'e

Comprendre la transparence et l'opacité').

Bien entendu, le **RGPD n'empêche pas toute forme de profilage et/ou de prise de décision automatisée** : il offre seulement aux individus un droit qualifié d'être informés à ce sujet, et un droit de ne pas faire l'objet d'une décision basée sur une prise de décision purement automatisée, y compris le profilage. Leur droit à l'information (voir "Droit à l'information" dans la section "Droits des personnes concernées" de la partie II des présentes lignes directrices) doit être satisfait par l'application du principe de licéité, de loyauté et de transparence (voir "Principe de licéité, de loyauté et de transparence" dans la section "Principes" de la partie II des présentes lignes directrices). Cela signifie que, au **minimum,** les responsables du traitement doivent informer la personne concernée qu'"ils s'engagent dans ce type d'activité, fournir des informations significatives sur la logique impliquée et sur l'importance et les conséquences envisagées du profilage pour la personne concernée"[396] (voir articles 13 et 14 du RGPD).

Les **informations sur la logique d'un système** et les explications des décisions doivent donner aux individus le contexte nécessaire pour prendre des décisions sur le traitement de leurs données personnelles. Dans certains cas, des explications insuffisantes peuvent inciter les personnes à recourir inutilement à d'autres droits. Les demandes d'intervention, l'expression de points de vue ou les oppositions au traitement sont plus susceptibles de se produire si les personnes ne pensent pas avoir une compréhension suffisante de la manière dont la décision a été prise. Les personnes concernées doivent être en mesure d'**exercer leurs droits de manière simple et conviviale.** Par exemple, "si le résultat d'une décision exclusivement automatisée est communiqué par le biais d'un site web, la page devrait contenir un lien ou des informations claires permettant à la personne de contacter un membre du personnel qui peut intervenir, sans délai ni complication excessifs".[397] Il est toutefois difficile de définir concrètement l'étendue des informations à fournir. En effet, cette question fait actuellement l'objet d'un débat académique animé. [398]

---

[395] Groupe de travail Article 29 (2017) Lignes directrices sur la prise de décision individuelle automatisée et le profilage aux fins du règlement 2016/679, WP 251, p.9. Commission européenne, Bruxelles.

[396] Ibid, pp. 13-14.

[397] ICO (2020) AI auditing framework : draft guidance for consultation, p.94. Bureau du commissaire à l'information, Wilmslow. Disponible à l'adresse : https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf (consulté le 15 mai 2020).

[398] Wachter, S., Mittelstadt, B. et Floridi, L. (2017) 'Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation', *International Data Privacy Law*. Disponible à l'adresse suivante https://ssrn.com/abstract=2903469 ou http://dx.doi.org/10.2139/ssrn.2903469 (consulté le 15 mai 2020) ; Selbst, A.D. et Powles, J. (2017) 'Meaningful information and the right to explanation', *International Data Privacy Law* 7(4) : 233-242, https://doi.org/10.1093/idpl/ipx022 (consulté le 15 mai 2020).

Encadré 0. La question du classement

Les fournisseurs de services ou de biens qui participent à l'économie dite "collaborative" (ou "économie de plateforme") doivent comprendre le fonctionnement du classement dans le contexte de leur utilisation de services d'intermédiation en ligne ou de moteurs de recherche en ligne spécifiques. Il peut s'agir, par exemple, d'un hôtel - petit ou grand - qui propose son hébergement par l'intermédiaire de Booking.com ou TripAdvisor. Pour permettre aux entreprises de participer en tant que prestataires sur la plateforme, il n'est pas nécessaire que les plateformes divulguent le fonctionnement détaillé de leurs mécanismes de classement, y compris les algorithmes utilisés. Il suffit de fournir une description générale des principaux paramètres de classement (y compris la possibilité d'influencer le classement contre toute rémunération directe ou indirecte versée par le prestataire), pour autant que cette description soit facilement et publiquement disponible, et rédigée dans un langage clair et intelligible.[399]

En outre, conformément à l'article 22, paragraphe 1, les personnes concernées ont le droit de ne pas faire l'objet d'une décision fondée uniquement sur un traitement automatisé, y compris le profilage, qui produit des effets juridiques les concernant ou des effets similaires significatifs. Ainsi, les responsables du traitement doivent toujours s'assurer que les outils d'IA qu'ils utilisent ou développent **ne favorisent en aucun cas une prise de décision automatique inévitable.** En effet, selon le groupe de travail Article 29, "[s]i le responsable du traitement envisage un "modèle" dans lequel il prend uniquement des décisions automatisées ayant un impact élevé sur les personnes sur la base de profils établis à leur sujet et qu'il ne peut pas s'appuyer sur le consentement de la personne, sur un contrat avec elle ou sur une loi l'autorisant, le responsable du traitement ne doit pas poursuivre son action. Le responsable du traitement peut toujours envisager un "modèle" de prise de décision fondé sur le profilage, en augmentant de manière significative le niveau d'intervention humaine de sorte que le modèle ne soit plus un processus décisionnel entièrement automatisé, bien que le traitement puisse toujours présenter des risques pour les droits et libertés fondamentaux des personnes."[400]

Encadré 1. Comprendre le profilage

Les recherches de Kosinski et al. (2013)[401] ont montré qu'en 2011, les enregistrements numériques accessibles du comportement (comme les pages "aimées" sur Facebook)

[399] Règlement UE 1159/2019 du 20 juin 2019 visant à promouvoir l'équité et la transparence pour les utilisateurs professionnels de services d'intermédiation en ligne, article 5 et considérant 27. Disponible à l'adresse suivante : https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R1150&from=EN

[400] Groupe de travail Article 29 (2018) Lignes directrices sur la prise de décision individuelle automatisée et le profilage aux fins du règlement 2016/679. Commission européenne, Bruxelles, p. 30. Disponible à l'adresse : https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053.

[401] Kosinski M., Stillwell, D. et Graepel, T. (2013) "Digital records of behavior expose personal traits", *Proceedings of the National Academy of Sciences* 110(15) : 5802- 5805, DOI:10.1073/pnas.1218772110.

pouvaient être utilisés pour prédire avec précision une série d'attributs personnels très sensibles. Il s'agissait notamment de l'orientation sexuelle, de l'origine ethnique, des opinions religieuses et politiques, des traits de personnalité, de l'intelligence, du bonheur, de la consommation de substances addictives, de la séparation des parents, de l'âge et du sexe. L'analyse était basée sur un ensemble de données de plus de 58 000 volontaires qui ont fourni leurs "J'aime" sur Facebook, des profils démographiques détaillés et les résultats de plusieurs tests psychométriques.

Le modèle a correctement discriminé entre les hommes homosexuels et hétérosexuels dans 88% des cas ; entre les Afro-américains et les Américains de type caucasien dans 95% des cas ; et entre les électeurs démocrates et républicains dans 85% des cas. Pour le trait de personnalité "Ouverture", la précision de la prédiction était proche de la précision test-retest d'un test de personnalité standard. Les auteurs ont également fourni des exemples d'association entre des attributs et des "J'aime" et ont discuté des implications pour la personnalisation en ligne et la vie privée.

Ce cas constitue un excellent exemple du fonctionnement du profilage : les informations relatives aux personnes concernées ont servi à les classer et à faire des prédictions à leur sujet.

En outre, un responsable du traitement doit toujours se rappeler que, conformément à l'article 9, paragraphe 2, point a), du RGPD, les décisions automatisées qui impliquent le traitement de catégories particulières de données à caractère personnel ne sont autorisées que si la personne concernée a donné son consentement explicite au traitement de ces données à caractère personnel pour une ou plusieurs finalités déterminées, ou s'il existe une base juridique pour le traitement mentionné. Cette exception s'applique non seulement lorsque les données observées entrent dans cette catégorie, mais aussi si le rapprochement de différents types de données à caractère personnel peut révéler des informations sensibles sur des personnes, ou si des données déduites entrent dans cette catégorie. En effet, **une étude capable de déduire des catégories spéciales de données est soumise aux mêmes obligations légales, en vertu du RGPD, qu'une étude dans laquelle des données personnelles sensibles sont traitées dès le départ.** Dans tous ces cas, nous devons prendre en compte la réglementation applicable au traitement des catégories spéciales de données personnelles et l'application nécessaire de garanties appropriées, capables de protéger les droits, intérêts et libertés des personnes concernées. La proportionnalité entre l'objectif de la recherche et l'utilisation des catégories particulières de données doit être garantie. En outre, les responsables du traitement doivent être conscients que leurs États membres peuvent maintenir ou introduire des conditions supplémentaires, y compris des limitations, en ce qui concerne le traitement des données génétiques, des données biométriques ou des données relatives à la santé (article 9, paragraphe 4, du RGPD).

Si le profilage déduit des données personnelles qui n'ont pas été fournies par la personne concernée, les responsables du traitement doivent s'assurer que le traitement n'est pas incompatible avec la finalité initiale (voir "Protection des données et recherche scientifique" dans la partie II, section "Concepts principaux") ; qu'ils ont identifié une base juridique pour le traitement des données de catégorie spéciale ; et qu'ils informent

la personne concernée du traitement[402] (voir "Limitation de la finalité" dans la partie II, section "Principes").

La réalisation d'une "**analyse d'impact sur la protection des données" (AIPD) (**voir "AIPD" dans la section "Principaux outils et actions" de la partie II) est **obligatoire s'il existe un risque réel de profilage non autorisé ou de prise de décision automatisée.** L'article 35(3)(a) du RGPD stipule la nécessité pour le responsable du traitement de réaliser une AIPD dans le cas d'une évaluation systématique et étendue des aspects personnels relatifs aux personnes physiques. Cela doit être fait pour les outils basés sur un traitement automatisé, y compris le profilage, et pour ceux sur lesquels sont fondées des décisions produisant des effets juridiques concernant la personne physique, ou affectant de manière significative la personne physique.

Conformément à l'article 37, paragraphe 1, point b)5, du RGPD, une exigence supplémentaire en matière de responsabilité est la **désignation d'un délégué à la protection des données (DPD)**, lorsque le profilage ou la prise de décision automatisée est une activité essentielle du responsable du traitement et nécessite un suivi régulier et systématique des personnes concernées à grande échelle. Les responsables du traitement sont également tenus de conserver un **registre de toutes les décisions prises par un système d'IA dans le cadre** de leurs obligations en matière de responsabilité et de documentation (voir la section Responsabilité du chapitre Principes). Ce registre doit indiquer si une personne a demandé une intervention humaine, a exprimé son point de vue, a contesté la décision et si celle-ci a été modifiée en conséquence. [403]

Voici quelques actions supplémentaires qui pourraient être extrêmement utiles pour éviter la prise de décision automatisée. [404]

- Prenez en compte les exigences du système nécessaires pour permettre un examen humain significatif dès la phase de conception.
- En particulier, il faut tenir compte des exigences d'interprétabilité et de la conception d'une interface utilisateur efficace pour soutenir les examens et les interventions humaines.
- Concevoir et dispenser une formation et un soutien appropriés aux investigateurs humains.
- Donner au personnel l'autorité, les incitations et le soutien appropriés pour répondre aux préoccupations des individus ou les transmettre à un échelon supérieur et, si nécessaire, passer outre la décision du système d'IA.

En tout état de cause, les responsables du traitement doivent savoir que les États membres introduisent certaines **références concrètes à cette question dans leurs**

---

[402] Groupe de travail Article 29 (2017) Lignes directrices sur la prise de décision individuelle automatisée et le profilage aux fins du règlement 2016/679, WP 251, p.15. Commission européenne, Bruxelles.

[403] ICO (2020) AI auditing framework : draft guidance for consultation, p.94-95. Bureau du commissaire à l'information, Wilmslow. Disponible à l'adresse : https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf (consulté le 15 mai 2020).

[404] Ibid. p.95.

réglementations nationales, en utilisant différents outils pour assurer une conformité adéquate.[405]

Liste de contrôle : profilage et prise de décision automatisée[406]

 Les responsables du traitement disposent d'une base juridique pour effectuer le profilage et/ou la prise de décision automatisée, et le documentent dans leur politique de protection des données.

 Les responsables du traitement envoient aux personnes un lien vers leur déclaration de confidentialité lorsqu'ils ont obtenu leurs données personnelles de manière indirecte.

 Les responsables du traitement expliquent comment les personnes peuvent accéder aux détails des informations qu'elles ont utilisées pour créer leur profil.

 Les responsables du traitement indiquent aux personnes qui leur fournissent leurs données personnelles et comment elles peuvent s'opposer au profilage.

 Les responsables du traitement disposent de procédures permettant aux clients d'accéder aux données personnelles saisies dans leurs profils, afin qu'ils puissent les examiner et les modifier pour tout problème d'exactitude.

 Les responsables du traitement ont mis en place des contrôles supplémentaires pour leurs systèmes de profilage/de prise de décision automatisée afin de protéger tout groupe vulnérable (y compris les enfants).

 Les responsables du traitement ne collectent que le minimum de données nécessaires et ont une politique de conservation claire pour les profils qu'ils créent.

En tant que modèle de bonnes pratiques

 Les responsables du traitement effectuent une AIPD pour examiner et traiter les risques lorsqu'ils commencent toute nouvelle prise de décision automatisée ou tout nouveau profilage.

 Les responsables du traitement informent leurs clients du profilage et de la prise de décision automatisée qu'ils effectuent, des informations qu'ils utilisent pour créer les profils et de la provenance de ces informations.

 Les responsables du traitement utilisent des données anonymisées dans le cadre de leurs activités de profilage.

 Les responsables garantissent le droit à la lisibilité des décisions algorithmiques.

---

[405] Malgieri, G. (2018) La prise de décision automatisée dans les lois des États membres de l'UE : le droit à l'explication et autres "garanties appropriées" pour les décisions algorithmiques dans les législations nationales de l'UE. Disponible à l'adresse : https://ssrn.com/abstract=3233611 ou http://dx.doi.org/10.2139/ssrn.3233611 (consulté le 2 mai 2020).

[406] ICO (aucune date) Droits liés à la prise de décision automatisée, y compris le profilage. Bureau du commissaire à l'information, Wilmslow. Disponible à l'adresse : https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/rights-related-to-automated-decision-making-including-profiling/ (consulté le 15 mai 2020).

 Les décideurs disposent d'un mécanisme capable de notifier et d'expliquer les raisons lorsqu'une contestation de la décision algorithmique n'est pas acceptée en raison de l'absence d'intervention humaine.

 Les décideurs disposent d'un modèle d'évaluation des droits de l'Homme dans la prise de décision automatisée.

 Une supervision humaine qualifiée est mise en place dès la phase de conception, notamment sur les exigences d'interprétation et la conception efficace de l'interface, et les investigateurs sont formés.

 Des vérifications sont effectuées en ce qui concerne les déviations possibles des résultats des déductions dans les systèmes adaptatifs ou évolutifs.

 La certification du système d'IA est, ou a été, effectuée.

Informations complémentaires

Groupe de travail Article 29 (2018) Lignes directrices sur la prise de décision individuelle automatisée et le profilage aux fins du règlement 2016/679. Commission européenne, Bruxelles. Disponible à l'adresse suivante : https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053

ICO (2020) AI auditing framework : draft guidance for consultation, p.94-95. Bureau du commissaire à l'information, Wilmslow. Disponible à l'adresse : https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf

ICO (2019) Analyses d'impact sur la protection des données et IA. Bureau du commissaire à l'information, Wilmslow. Disponible à l'adresse : https://ico.org.uk/about-the-ico/news-and-events/ai-blog-data-protection-impact-assessments-and-ai/

Kosinski M., Stillwell, D. et Graepel, T. (2013) "Digital records of behavior expose personal traits", *Proceedings of the National Academy of Sciences* 110(15) : 5802- 5805, DOI:10.1073/pnas.1218772110.

Malgieri, G. (2018) La prise de décision automatisée dans les lois des États membres de l'UE : le droit à l'explication et autres "garanties appropriées" pour les décisions algorithmiques dans les législations nationales de l'UE. Disponible à l'adresse : https://ssrn.com/abstract=3233611 ou http://dx.doi.org/10.2139/ssrn.3233611

Autorité norvégienne de protection des données (2018) Intelligence artificielle et vie privée. Autorité norvégienne de protection des données, Oslo. Disponible à l'adresse : https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf

Selbst, A.D. et Powles, J. (2017) " Meaningful information and the right to explanation ", International Data Privacy Law 7(4) : 233-242, https://doi.org/10.1093/idpl/ipx022.

Wachter, S., Mittelstadt, B. et Floridi, L. (2017) 'Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation', International Data Privacy Law. Disponible à l'adresse suivante https://ssrn.com/abstract=2903469 ou http://dx.doi.org/10.2139/ssrn.2903469

2 Robustesse technique et sécurité

"La robustesse technique, qui est étroitement liée au principe de prévention des dommages, est un élément crucial de la réalisation d'une IA digne de confiance. La robustesse technique exige que les systèmes d'IA soient développés selon une approche préventive des risques et de manière à ce qu'ils se comportent de manière fiable comme prévu, tout en minimisant les dommages involontaires et inattendus, et en empêchant les dommages inacceptables. Cela devrait également s'appliquer aux changements potentiels de leur environnement de fonctionnement ou à la présence d'autres agents (humains et artificiels) susceptibles d'interagir avec le système de manière contradictoire. En outre, l'intégrité physique et mentale des humains doit être assurée."

- *Groupe d'experts de haut niveau sur l'IA*[407]

2.1 Principes éthiques et dispositions du RGPD

Le groupe d'experts de haut niveau sur l'IA divise l'exigence de robustesse technique et de sécurité en quatre sous-composantes : (1) résilience aux attaques et sécurité ; (2) plan de repli et sécurité générale ; (3) précision ; et (4) fiabilité et reproductibilité.

Pour faciliter les références, cette section reprend cette structure, tout en reliant ces sous-composantes aux exigences et recommandations légales (RGPD). Ce point est important, car si les exigences du RGPD ne s'appliquent généralement qu'au traitement des données personnelles, de nombreux systèmes d'IA pratiques sont conçus pour produire un résultat personnalisé (par exemple, les systèmes de recommandation) et doivent donc traiter des données personnelles à un moment donné.

2.1.1 Résilience aux attaques et sécurité

La résilience aux attaques devrait être un objectif de tous les systèmes TIC, y compris les systèmes d'IA. Lors du traitement des données à caractère personnel, l'article 32 du RGPD exige explicitement la mise en œuvre de mesures techniques et organisationnelles appropriées pour assurer la sécurité des données (voir "Mesures en faveur de la confidentialité" dans la sous-section "Intégrité et confidentialité" des "Principes" de la partie II.

Les mesures de sécurité requises dépendent de l'impact probable d'un dysfonctionnement du système d'IA. Ces mesures doivent également inclure les dispositions prises pour assurer la résilience des systèmes de traitement.[408] Pour certains types de systèmes d'IA, le processus de prise de décision peut être particulièrement vulnérable aux attaques. Par exemple, un acteur malveillant peut créer une entrée trompeuse pour exploiter les différences fondamentales de perception entre les humains et les systèmes d'IA, comme le montre l'exemple de l'encadré 2.

---

[407] Groupe d'experts de haut niveau sur l'IA (2019) Lignes directrices en matière d'éthique pour une IA digne de confiance, p.16 et suivantes. Commission européenne, Bruxelles. Disponible sur : https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (consulté le 28 mai 2020).

[408] Article 32(1)(b) du RGPD.

Encadré 2. Exemple du besoin de sécurité dans les systèmes d'IA

Un véhicule autonome doit reconnaître automatiquement les panneaux de signalisation des rues, en utilisant des caméras embarquées et en ajustant sa vitesse en conséquence. Bien que les algorithmes d'IA basés sur les réseaux neuronaux profonds puissent exceller dans cette tâche, il faut veiller à protéger le système contre les attaques ciblées d'un adversaire malveillant. Par exemple, de petites modifications ciblées des panneaux de signalisation pourraient amener le système d'IA à confondre un panneau d'arrêt avec un panneau de limitation de vitesse, ce qui entraînerait des situations potentiellement dangereuses. Dans le même temps, la modification peut apparaître comme un simple graffiti pour l'observateur humain courant. Il est donc de la plus haute importance de protéger un système d'IA utilisé à cette fin contre de telles attaques, augmentant ainsi sa résilience.[409]

Les modèles d'IA entraînés peuvent également constituer une source de données précieuse. Dans certaines circonstances, il peut être possible d'obtenir des informations sur les données d'entrée originales en utilisant uniquement le modèle formé.[410] Une telle "fuite d'informations" pourrait être exploitée par des acteurs internes et externes. Il est donc important que les responsables du traitement **prennent des mesures pour limiter l'accès au modèle et aux données de formation sous-jacentes, et ce pour toutes les catégories d'acteurs** (voir "Mesures en faveur de la confidentialité" dans la sous-section "Intégrité et confidentialité" des "Principes" de la partie II des présentes lignes directrices).

Une fois formé, le système d'IA résultant peut être utilisé à des fins très différentes de celles prévues à l'origine. Par exemple, un système d'IA à reconnaissance faciale peut être utilisé pour reconnaître et regrouper des photos contenant une personne spécifique dans un album photo en ligne. Le même système d'IA pourrait également être utilisé pour rechercher sur internet des photos d'une personne spécifique, en révélant potentiellement des détails personnels sensibles (c'est-à-dire en utilisant l'emplacement de la photo ou le contexte de capture). Ce type d'utilisation polyvalente est souvent possible avec les systèmes d'IA, et il incombe au concepteur du système de *prévoir un éventuel traitement illégal des données à caractère personnel et de mettre en œuvre des mesures de sécurité qui l'empêcheraient ou le minimiseraient.* Il peut s'agir de mesures telles que la restriction des sources de données utilisables ou l'interdiction de certains modes d'utilisation par le biais de conditions de licence. Le cadre juridique de la protection des données peut compléter ces restrictions, mais ne les remplace en aucun cas.

---

[409] Eykholt, K. et al. (2018) 'Robust physical-world attacks on deep learning models', 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition 10.4.2018, arXiv:1707.08945.

[410] Fredrikson, M. et al. (2015) 'Model inversion attacks that exploit confidence information and basic countermeasures', CCS '15 : Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, October 2015. Université Cornell, Ithaca. Disponible à l'adresse : https://rist.tech.cornell.edu/papers/mi-ccs.pdf (consulté le 20 mai 2020).

2.1.2 Plan de repli et sécurité générale

> "L'erreur est humaine mais pour vraiment tout gâcher, il faut un ordinateur."
>
> - *Paul Ehrlich / Bill Vaughan*[411]

Comme tous les systèmes TIC, les systèmes d'IA peuvent échouer et fournir des résultats ou des prédictions incorrects. Cependant, dans le cas des systèmes d'IA, il peut être particulièrement difficile d'expliquer de manière tangible et humaine pourquoi une conclusion particulière (fausse) a été atteinte. Un exemple de comportement indésirable serait un système d'IA qui prend des décisions qui affectent de manière significative une personne (par exemple, le refus automatique d'une demande de crédit). Le RGPD exige que les responsables du traitement mettent en œuvre des plans de repli appropriés protégeant les personnes concernées de telles situations, y compris le droit de contester une décision de l'IA et d'obtenir une intervention humaine prenant en compte le point de vue des personnes concernées.[412] Ces mesures de protection doivent être prises en compte lors de la conception des systèmes. Même dans les cas où le RGPD n'exige pas explicitement un tel plan de repli, il est souhaitable que les responsables du traitement envisagent d'en mettre un en place.

Les responsables du traitement doivent également être conscients des questions de sécurité. Les nouvelles technologies entraînent souvent de nouveaux risques. Il est important d'être conscient que la protection des données personnelles dépend des mesures de sécurité informatique et donc que les risques liés aux données personnelles sont ceux liés à l'informatique. Par conséquent, les mesures techniques et organisationnelles appropriées mises en œuvre dans les TI assureront la protection des données, comme le stipule le RGPD, et ces mesures devraient être régulièrement testées et mises à jour pour prévenir ou minimiser les risques de sécurité. (voir la sous-section "Principale différence par rapport aux autres risques du RGPD et aux risques liés à la sécurité informatique" dans la section "Intégrité et confidentialité" du chapitre "Principes").

Pour évaluer ces risques et en déduire des garanties appropriées, le RGPD exige qu'une AIPD soit réalisée avant le traitement lorsqu'il existe un risque élevé pour les droits et libertés d'une personne physique[413] (voir " AIPD" dans la partie II section "Principaux outils et actions" des présentes lignes directrices). L'utilisation de nouvelles technologies telles que l'IA augmente la probabilité que le traitement entre dans la catégorie des risques élevés. Certaines agences nationales de protection des données ont émis des directives exigeant un AIPD lors de l'utilisation de certains algorithmes d'IA.[414] En cas de doute, il est recommandé aux responsables du traitement de réaliser une AIPD.[415]

---

[411] La paternité de la citation semble être contestée, voir par exemple https://quoteinvestigator.com/2010/12/07/foul-computer/#note-1699-18 (consulté le 2 juin 2020).

[412] Article 33(3) du RGPD.

[413] Article 35, paragraphe 1, du RGPD

[414] Voir, par exemple, la situation juridique en Autriche § 2(2)(4) DSFA-V.

[415] Groupe de travail Article 29 (2017) WP248, Lignes directrices sur l'analyse d'impact sur la protection des données (AIPD) et la détermination du fait que le traitement est "susceptible d'entraîner un risque élevé" aux fins du règlement 2016/679, p.8. Commission européenne, Bruxelles.

2.1.3 Précision

Une grande précision du système est généralement l'un des objectifs de conception des systèmes d'IA. De nombreux systèmes d'IA ont besoin de données de formation précises et fiables pour obtenir les meilleurs résultats. Lors du traitement des données personnelles, les tenir à jour et corriger les entrées erronées est également une obligation légale.[416] La personne concernée peut également exiger la rectification de données personnelles inexactes.[417] Les systèmes d'IA devraient donc être conçus en tenant compte de la nécessité d'un recyclage, au cours duquel des données peuvent non seulement être ajoutées, mais aussi supprimées (voir "Droit de rectification" dans la section "Droits de la personne concernée" de la partie II des présentes lignes directrices et "Droit de rectification" dans la section "Droits de la personne concernée" de la troisième partie des présentes lignes directrices). Loyauté, diversité et non-discrimination " de la présente partie III sur l'IA, ainsi que le "principe de licéité, de loyauté et de transparence" de la partie II, section "Principes").

En outre, la production d'un système d'IA ne doit pas seulement être un résultat, mais aussi une mesure de la confiance du système dans l'exactitude du résultat. Une telle mesure n'est pas seulement un indicateur technique de la précision du système, mais aussi une indication précieuse de la nécessité éventuelle d'une intervention humaine (voir la section "Principe de précision" dans les "Principes" de la partie II des présentes lignes directrices).

2.1.4 Fiabilité et reproductibilité

De nombreux systèmes d'IA sont conçus avec un cas d'utilisation spécifique en tête. Cependant, comme on l'a dit, il peut évoluer avec le temps et s'éloigner lentement des intentions initiales des concepteurs. Il est donc important de documenter clairement les hypothèses initiales et les conditions dans lesquelles le système d'IA était destiné à être utilisé. Par exemple, le système d'IA s'attend-il à un environnement spécifique, ou l'ensemble d'apprentissage contient-il des biais connus ? Si un système d'IA est accessible au public, la documentation sur la fiabilité du système devrait l'être également.

En plus de la fiabilité, la reproductibilité des résultats d'un système d'IA est importante. Non seulement la reproductibilité est une propriété technique souhaitable d'un système d'IA (par exemple, pour rechercher la raison de résultats erronés), mais c'est aussi une condition préalable importante pour la confiance. Si un résultat ne peut être reproduit, son explicabilité - et donc la confiance dans le système d'IA - peut en souffrir.

Liste de contrôle : robustesse technique et sécurité[418]

Résilience aux attaques et sécurité

[416] Article 5, paragraphe 1, point d), du RGPD.

[417] Article 16 du RGPD.

[418] Cette liste de contrôle a été adaptée de celle élaborée par le Groupe d'experts de haut niveau sur l'intelligence artificielle (2019) Lignes directrices en matière d'éthique pour une IA digne de confiance. Commission européenne, Bruxelles. Disponible à l'adresse : https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (consulté le 20 mai 2020).

 Le responsable du traitement a évalué les formes potentielles d'attaques auxquelles le système d'IA pourrait être vulnérable.

 Le responsable du traitement a examiné différents types et natures de vulnérabilités, comme la pollution des données, les infrastructures physiques et les cyberattaques.

 Le responsable du traitement a mis en place des mesures ou des systèmes pour garantir l'intégrité et la résilience du système d'IA contre les attaques potentielles.

 Le responsable du traitement a vérifié comment le système se comporte dans des situations et des environnements inattendus.

 Le responsable du traitement a examiné dans quelle mesure le système pouvait être à double usage. Dans l'affirmative, le responsable du traitement a pris des mesures préventives appropriées à cet égard (par exemple, ne pas publier la recherche ou déployer le système).

Plan de repli et sécurité générale

 Le responsable du traitement s'est assuré que le système dispose d'un plan de repli suffisant s'il est confronté à des attaques adverses ou à d'autres situations inattendues (par exemple, procédures de commutation technique ou demande d'un opérateur humain avant de poursuivre).

 Le responsable du traitement a examiné le niveau de risque soulevé par le système d'IA dans ce cas d'utilisation spécifique.

 Le responsable du traitement a mis en place tout processus pour mesurer et évaluer les risques et la sécurité.

 Le responsable du traitement a fourni les informations nécessaires en cas de risque pour l'intégrité physique humaine.

 Le responsable du traitement a envisagé une police d'assurance pour faire face aux dommages potentiels du système d'IA.

 Le responsable du traitement a identifié les risques potentiels pour la sécurité des (autres) utilisations prévisibles de la technologie, y compris les utilisations accidentelles ou malveillantes. Existe-t-il un plan pour atténuer ou gérer ces risques ?

 Le responsable du traitement a évalué s'il existe une chance probable que le système d'IA puisse causer des dommages ou des préjudices aux utilisateurs ou à des tiers. Le responsable du traitement a évalué la probabilité, les dommages potentiels, le public impacté et la gravité.

 Le responsable du traitement a examiné les règles de responsabilité et de protection des consommateurs, et en tient compte.

 Le responsable du traitement a considéré l'impact potentiel ou le risque de sécurité pour l'environnement ou les animaux.

 L'analyse des risques du responsable du traitement comprenait la question de savoir si des problèmes de sécurité ou de réseau (par exemple, des risques de cybersécurité) pouvaient poser des risques de sécurité ou des dommages dus à un comportement non intentionnel du système d'IA.

 Le responsable du traitement a estimé l'impact probable d'une défaillance du système d'IA lorsqu'il fournit des résultats erronés, devient indisponible ou fournit des résultats socialement inacceptables (par exemple, la discrimination).

 Le responsable du traitement a défini des seuils et mis en place des procédures de gouvernance pour déclencher des plans alternatifs/de repli.

 Le responsable du traitement a défini et testé des plans de repli.

Précision

 Le responsable du traitement a évalué le niveau et la définition de la précision qui seraient nécessaires dans le contexte du système d'IA et du cas d'utilisation.

 Le responsable du traitement a évalué comment la précision est mesurée et assurée.

 Le responsable du traitement a mis en place des mesures pour s'assurer que les données utilisées sont complètes et à jour.

 Le responsable du traitement a mis en place des mesures pour évaluer s'il est nécessaire de disposer de données supplémentaires, par exemple pour améliorer l'exactitude ou éliminer les biais.

 Le responsable du traitement a vérifié quel dommage serait causé si le système d'IA fait des prédictions inexactes.

 Le responsable du traitement a mis en place des moyens pour mesurer si le système fait une quantité inacceptable de prédictions inexactes.

 Le responsable du traitement a mis en place une série d'étapes pour augmenter la précision du système.

Fiabilité et reproductibilité

 Le responsable du traitement a mis en place une stratégie pour surveiller et tester si le système d'IA atteint ses objectifs, ses buts et les applications prévues.

 Le responsable du traitement a testé si des contextes spécifiques ou des conditions particulières doivent être pris en compte pour assurer la reproductibilité.

 Le responsable du traitement a mis en place des méthodes de vérification pour mesurer et garantir différents aspects de la fiabilité et de la reproductibilité du système.

 Le responsable du traitement a mis en place des processus pour décrire quand un système d'IA échoue dans certains contextes.

 Le responsable du traitement a clairement documenté et opérationnalisé ces processus pour le test et la vérification de la fiabilité des systèmes d'IA.

 Le responsable du traitement a établi des mécanismes de communication pour assurer les utilisateurs (finaux) de la fiabilité du système.

Informations complémentaires

Groupe de travail Article 29 (2017) WP248, Lignes directrices sur l'analyse d'impact sur la protection des données (AIPD) et la détermination du fait que le traitement est "susceptible d'entraîner un risque élevé" aux fins du règlement 2016/679. Commission européenne, Bruxelles. Disponible à l'adresse suivante : https://ec.europa.eu/newsroom/document.cfm?doc_id=47711

Eykholt, K. et al. (2018) " Robust physical-world attacks on deep learning models ", 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition 10.4.2018, arXiv:1707.08945.

Fredrikson, M. et al. (2015) 'Model inversion attacks that exploit confidence information and basic countermeasures', CCS '15 : Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, October 2015. Université Cornell, Ithaca. Disponible à l'adresse : https://rist.tech.cornell.edu/papers/mi-ccs.pdf

3 Vie privée et gouvernance des données

"La vie privée, un droit fondamental particulièrement touché par les systèmes d'IA, est étroitement liée au principe de prévention des dommages. La prévention des atteintes à la vie privée nécessite également une gouvernance adéquate des données qui couvre la qualité et l'intégrité des données utilisées, leur pertinence au regard du domaine dans lequel les systèmes d'IA seront déployés, leurs protocoles d'accès et la capacité de traiter les données de manière à protéger la vie privée."

- *Groupe d'experts de haut niveau sur l'IA*[419]

3.1 Principes éthiques

Cette exigence englobe trois grands principes différents, à savoir : [420]

- **Protection de la vie privée et des données.** Les systèmes d'IA doivent garantir la protection de la vie privée et des données tout au long du cycle de vie du système. Cela inclut les informations fournies initialement par l'utilisateur, ainsi que les informations générées sur l'utilisateur au cours de son interaction avec le système (par exemple, les résultats que le système d'IA a générés pour des utilisateurs spécifiques ou la façon dont les utilisateurs ont répondu à des recommandations particulières). Pour que les personnes puissent avoir confiance dans le processus de collecte des données, il faut s'assurer que les données recueillies à leur sujet ne seront pas utilisées pour les discriminer de manière illégale ou injuste.
- **Qualité et intégrité des données.** La qualité des ensembles de données utilisés est primordiale pour les performances des systèmes d'IA. Lorsque les données

[419] Groupe d'experts de haut niveau sur l'IA (2019) Lignes directrices en matière d'éthique pour une IA digne de confiance, p.17. Commission européenne, Bruxelles. Disponible à l'adresse : https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (consulté le 28 mai 2020).

[420] Ibid. , p.15 et ff.

sont recueillies, elles peuvent contenir des biais, des inexactitudes, des erreurs et des fautes d'origine sociale. Il faut y remédier avant de s'entraîner avec un ensemble de données donné. En outre, l'intégrité des données doit être garantie. L'introduction de données malveillantes dans un système d'IA peut modifier son comportement, en particulier avec les systèmes d'auto-apprentissage. Les processus et les ensembles de données utilisés doivent être testés et documentés à chaque étape, comme la planification, la formation, les tests et le déploiement. Cela devrait également s'appliquer aux systèmes d'IA qui n'ont pas été développés en interne mais acquis ailleurs.

- **Accès aux données.** Dans toute organisation qui traite des données à caractère personnel, des documents/politiques internes stipulant qui peut accéder aux données à caractère personnel (et dans quelles conditions), y compris les mesures organisationnelles et techniques de contrôle d'accès, doivent être en place. Seul le personnel dûment qualifié ayant la compétence et le besoin d'accéder aux données personnelles des individus doit être autorisé à le faire. En outre, tout le personnel auquel l'accès est accordé doit signer une déclaration de confidentialité.

3.2 Dispositions du RGPD

Le RGPD fait référence au traitement des données personnelles des personnes concernées. Certaines dispositions sont particulièrement pertinentes pour la vie privée et la gouvernance des données. La qualité et l'intégrité des données et l'accès aux données étant analysés dans la section précédente, nous nous concentrons ici sur quatre concepts extrêmement pertinents pour garantir une gouvernance adéquate des données. Il s'agit de : (1) la limitation de la finalité ; (2) la licéité ; (3) la minimisation des données ; et (4) la loyauté, un principe général qui exige de protéger les droits des personnes concernées.

Il est inutile de parler de protection des données si le traitement n'est pas licite, et une finalité spécifique et explicite est une condition préalable à un traitement licite. Toutefois, même si le traitement est autorisé (c'est-à-dire licite et légitime), la protection des données reste impossible à mettre en œuvre si les finalités du traitement ne sont pas claires. En outre, le traitement n'est pas licite s'il n'est pas lié aux finalités pour lesquelles les données ont été collectées. Par conséquent, le principe de limitation de la finalité est directement lié à la gouvernance des données.

Parallèlement, la minimisation des données est essentielle à la protection de la vie privée. La meilleure façon de s'assurer que "les données collectées sur [les personnes concernées] ne seront pas utilisées pour les discriminer de manière illégale ou injuste"[421] est de minimiser la quantité et l'étendue des données personnelles collectées. Enfin, une mise en œuvre adéquate des droits des personnes concernées, tels qu'ils sont

[421] Groupe d'experts de haut niveau sur l'IA (2019) Lignes directrices en matière d'éthique pour une IA digne de confiance, p.17. Commission européenne, Bruxelles. Disponible à l'adresse : https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (consulté le 28 mai 2020).

[421] Ibid. , p.15 et ff.

intégrés dans le RGPD, est essentielle pour les responsabiliser et renforcer le cadre de gouvernance des données.

3.2.1 Limitation de la finalité

Le principe de limitation de la finalité limite l'utilisation des données à caractère personnel à la/les finalités initiales, ou aux finalités compatibles avec celles-ci. Cependant, le développement de l'IA exige que les données soient réutilisées assez souvent. En outre, il peut arriver que l'outil d'IA réutilise les données automatiquement (cela se produit certainement dans le cas de l'apprentissage profond). Ces situations créent des tensions entre les techniques d'apprentissage de l'IA et le principe de limitation de la finalité (voir "Principe de limitation de la finalité" dans la partie II, section "Principes" des présentes lignes directrices).

Afin d'éviter tout traitement de données illicite, les responsables du traitement utilisant des systèmes d'IA **doivent déterminer la finalité du traitement "dès le début de sa formation ou de son déploiement, et procéder à une réévaluation de cette détermination si le traitement du système donne des résultats inattendus**, puisqu'il exige que les données à caractère personnel ne soient collectées que pour des "finalités spécifiques, explicites et légitimes" et ne soient pas utilisées d'une manière incompatible avec la finalité initiale"[422] (voir la section "Protection des données dès la conception et par défaut" dans "Concepts principaux", dans la partie II des présentes lignes directrices).

La **réutilisation des données** dans le cadre du développement d'un outil d'IA soulève des questions très complexes en termes de limitation de la finalité. Les systèmes d'IA traitent les données à caractère personnel à différents stades et à des fins diverses. En conséquence, un responsable du traitement peut ne pas distinguer chaque opération de traitement distincte et traiter des données à des fins autres que celles pour lesquelles elles ont été initialement collectées. Les responsables du traitement doivent être particulièrement préoccupés par ces situations, car elles peuvent entraîner un non-respect du principe de licéité en matière de protection des données[423] (voir la sous-section "Utilisation à des fins incompatibles" dans la section "Principe de limitation de la finalité" des "Principes" de la partie II des présentes lignes directrices).

Les responsables du traitement doivent considérer que l'identification de la base légale appropriée **est liée aux principes de loyauté et de limitation de la finalité** (voir "Principe de licéité, de loyauté et de transparence" dans la partie II, section "Principes" des présentes lignes directrices).[424] Ils doivent choisir la base légale qui reflète le mieux

---

[422] CIPL (2020) Intelligence artificielle et protection des données : comment le RGPD réglemente l'IA. Centre for Information Policy Leadership, Washington DC / Bruxelles / Londres, p.6. Mis en évidence par l'auteur. Disponible à l'adresse : www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl-hunton_andrews_kurth_legal_note_-_how_gdpr_regulates_ai__12_march_2020_.pdf (consulté le 17 mai 2020).

[423] ICO (2020) Guidance on the AI auditing framework : draft guidance for consultation. Information Commissioner's Office, Wilmslow. Disponible à l'adresse : https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf (consulté le 15 mai 2020).

[424] EDPB (2018) Lignes directrices 2/2019 sur le 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, adoptées le 9 avril 2019, p.6. Conseil européen de la protection des données,

la véritable nature de leur relation avec la personne et la finalité du traitement. Cette décision est essentielle, car il est impossible de modifier la base juridique du traitement s'il n'y a pas de raisons substantielles qui le justifient, en raison du principe de limitation de la finalité. Si les développeurs d'IA prévoient d'utiliser un ensemble de données à différentes étapes (par exemple, formation, validation ou déploiement), ils doivent considérer ces étapes comme ayant des objectifs distincts et séparés.[425] En outre, **ils doivent tenir compte du type de relation qu'ils entretiennent avec la personne concernée.** Par exemple, le consentement peut constituer une base légale appropriée pour le traitement s'il existe un contact permanent avec les personnes concernées et si les responsables du traitement sont en mesure d'obtenir des consentements successifs pour différentes utilisations ou d'obtenir le consentement de la personne concernée pour plusieurs traitements avant le début du traitement. Toutefois, dans le cas de l'IA, il est souvent difficile de maintenir ce type de relation, car l'IA est souvent construite en agrégeant et en fusionnant de grands ensembles de données.

Enfin, les responsables du traitement doivent être conscients que pour le traitement de données à caractère personnel à des fins scientifiques, de recherche historique ou de statistiques, la législation ou les règles de l'Union ou des États membres peuvent prévoir des dérogations aux droits des personnes concernées stipulés aux articles 15, 16, 18 et 21. Par conséquent, le traitement de ces données à des fins autres que celles pour lesquelles elles ont été initialement collectées devrait être licite pour autant que des mesures techniques et organisationnelles appropriées soient en place, en particulier la minimisation des données. (Voir la section "Protection des données et recherche scientifique" dans les "Concepts principaux" de la partie Ii des présentes lignes directrices).

Liste de contrôle : limitation de la finalité [426]

- Les responsables du traitement ont clairement identifié la ou les finalités de leur traitement.
- Les responsables du traitement ont documenté ces finalités.
- Les responsables du traitement incluent le détail de leurs finalités dans les informations relatives à la vie privée des personnes.
- Les responsables du traitement revoient régulièrement leurs traitements et, le cas échéant, mettent à jour leur documentation et les informations relatives à la vie privée des personnes.

Bruxelles. Disponible à l'adresse : https://edpb.europa.eu/sites/edpb/files/consultation/edpb_draft_guidelines-art_6-1-b-final_public_consultation_version_en.pdf (consulté le 15 mai 2020).

[425] ICO (2020) Orientations sur le cadre d'audit de l'IA : projet pour consultation. 2020. Bureau du commissaire à l'information, Wilmslow. Disponible à l'adresse : https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf (consulté le 15 mai 2020).

[426] ICO (aucune date) Principe (b) : limitation de la finalité. Bureau du commissaire à l'information, Wilmslow. Disponible à l'adresse : https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/purpose-limitation/ (consulté le 17 mai 2020).

- Si les responsables du traitement prévoient d'utiliser les données à caractère personnel pour une nouvelle finalité autre qu'une obligation légale ou une fonction prévue par la loi, ils vérifient que celle-ci est compatible avec leur finalité initiale ou obtiennent un consentement spécifique pour cette nouvelle finalité.

Informations complémentaires

Groupe de travail Article 29 sur la protection des données (2013) Avis 03/2013 sur la limitation de la finalité. Commission européenne, Bruxelles. Disponible sur : https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

CIPL (2020) Intelligence artificielle et protection des données : comment le RGPD réglemente l'IA. Centre for Information Policy Leadership, Washington DC / Bruxelles / Londres. Disponible à l'adresse : www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl-hunton_andrews_kurth_legal_note_-_how_gdpr_regulates_ai__12_march_2020_.pdf

EDPB (2018) Lignes directrices 2/2019 sur le 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, adoptées le 9 avril 2019, p.6. Conseil européen de la protection des données, Bruxelles. Disponible à l'adresse suivante : https://edpb.europa.eu/sites/edpb/files/consultation/edpb_draft_guidelines-art_6-1-b-final_public_consultation_version_en.pdf

ICO (2020) Guidance on the AI auditing framework : draft guidance for consultation. Information Commissioner's Office, Wilmslow. Disponible à l'adresse : https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf

ICO (aucune date) Principe (b) : limitation de la finalité. Bureau du commissaire à l'information, Wilmslow. Disponible sur : https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/purpose-limitation/

3.2.2 Licéité

La licéité est un principe essentiel en matière de protection des données. Il implique que les responsables du traitement doivent s'assurer qu'ils disposent d'une **base légale pour traiter les données à caractère personnel**. **Si tel n'est pas le cas,** le traitement **ne doit pas être effectué**.[427] En général, et y compris pour les données de catégories spéciales, les bases légales du traitement sont décrites à l'article 6 et à l'article 9 du RGPD. Dans le cas de l'IA, les bases légales habituellement invoquées pour justifier le traitement sont : le consentement ; l'intérêt légitime ; la nécessité contractuelle ; et l'obligation légale ou

[427] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial : Una introducción, p.20. Agencia Espanola Proteccion Datos, Madrid. Disponible sur : www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (consulté le 15 mai 2020).

l'intérêt vital. Le traitement pour l'intérêt public peut également être un motif légal, mais nous ne nous y attarderons pas ici puisque nous abordons largement ce sujet dans la section "Protection des données et recherche scientifique" de la section "Concepts principaux" de la partie II des présentes lignes directrices. Par conséquent, nous nous concentrerons sur les quatre motifs juridiques énumérés.

a) Consentement

Le traitement des données est souvent fondé sur le consentement fourni par les personnes concernées. Cependant, le consentement ne s'accorde pas bien avec la nature essentielle de la plupart des développements de l'IA, pour une raison simple : le consentement est, par nature, lié à un objectif concret et bien défini.[428] Dans le cas de l'IA, l'utilisation de big data et les actions d'agrégation, de partage ou de réaffectation qui sont souvent effectuées créent un scénario qui ne correspond pas aux principes sous-jacents du concept de consentement et au principe de limitation de la finalité (voir "Principe de limitation de la finalité" dans la partie II, section "Principes" des présentes lignes directrices).

Le consentement peut être une base juridique utile pour le traitement des données en vue du développement de l'IA, en particulier si les responsables du traitement ont une **relation directe avec le sujet qui fournit les données à utiliser pour la formation, la validation et le déploiement du modèle**.[429] Par exemple, si l'outil d'IA vise à fournir des diagnostics de pneumonie, et que les médecins obtiennent des données de patients dans leur établissement de santé, le consentement pourrait bien servir de base juridique au traitement. Cependant, si le traitement implique l'utilisation d'un outil d'IA complexe qui peut avoir d'autres utilisations des données (par exemple, le profilage et la prise de décision automatisée peuvent se produire par inadvertance, les données sont susceptibles d'être déduites pendant le traitement, ces données déduites peuvent être utilisées à diverses fins, etc.), il est difficile de voir comment un seul consentement pourrait justifier tous ces traitements. À cette fin, les responsables du traitement doivent appliquer les lignes directrices sur le consentement fournies par le groupe de travail Article 29[430] .

Dans le cadre de la recherche scientifique (voir la section "Protection des données et recherche scientifique" dans les "Concepts principaux" de la partie II des présentes lignes directrices), le RGPD prévoit une dérogation spécifique aux attributs du consentement, permettant aux responsables du traitement de faire usage du **consentement général** comme base juridique du traitement. Le consentement général

---

[428] Comité international de bioéthique (2017) Rapport du CIB sur le big data et la santé, p.20. UNESCO. Disponible à l'adresse : http://unesdoc.unesco.org/images/0024/002487/248724e.pdf (consulté le 13 mars 2020).

[429] ICO (aucune date) Comment appliquer les intérêts légitimes dans la pratique ? Information Commissioner's Office, Wilmslow. Disponible à l'adresse : https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/how-do-we-apply-legitimate-interests-in-practice/ (consulté le 15 mai 2020). En outre, l'évaluation de la nature de cette relation doit inclure une enquête sur le rapport de force entre la personne concernée et le responsable du traitement des données.

[430] Groupe de travail Article 29 (2018) Lignes directrices sur le consentement en vertu du règlement 2016/679. Commission européenne, Bruxelles, p.29. Disponible à l'adresse : https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051 (consulté le 5 mai 2020).

doit être compris en lien avec le considérant 33 du RGPD, qui stipule qu'il "n'est souvent pas possible d'identifier pleinement la finalité du traitement des données personnelles à des fins de recherche scientifique au moment de la collecte des données. Par conséquent, les personnes concernées devraient être autorisées à donner leur consentement à certains domaines de la recherche scientifique lorsque cela est conforme aux normes éthiques reconnues en matière de recherche scientifique."

Toutefois, le consentement général n'est pas une sorte de consentement générique ou équivoque par rapport au consentement ouvert. Il s'agit d'un **outil exceptionnel qui ne peut être acceptable que si plusieurs conditions sont réunies.** Si le consentement général est utilisé pour des catégories spéciales de données, les responsables du traitement doivent s'assurer que leur réglementation nationale l'autorise. Ils doivent également être conscients des garanties qui doivent être mises en œuvre. La proportionnalité entre l'objectif de la recherche et l'utilisation de catégories particulières de données doit être garantie. En outre, les responsables du traitement doivent s'assurer que la réglementation de leurs États membres ne protège pas les données génétiques, biométriques et de santé en introduisant des conditions ou des limitations supplémentaires, puisqu'ils sont autorisés à le faire par le RGPD.

En outre, chaque fois qu'un consentement général est utilisé pour atteindre la finalité de la recherche, certaines mesures essentielles doivent être envisagées pour compenser la définition abstraite des finalités de la recherche. Le respect des normes éthiques reconnues en matière de recherche scientifique, conformément au considérant 33 du RGPD, semble particulièrement pertinent à cette fin.

Encadré 3 : Consentement général et garanties supplémentaires

L'autorité allemande de protection des données a récemment dressé une liste de garanties supplémentaires à mettre en œuvre en cas de consentement général.[431] Il s'agit de :

1. Garanties pour assurer la transparence :

- Utilisation de règlements d'utilisation ou de plans de recherche qui illustrent les méthodes de travail prévues et les questions qui doivent faire l'objet du projet de recherche.
- Évaluation et documentation de la question de savoir pourquoi, dans ce projet de recherche particulier, une spécification plus détaillée des finalités de la recherche n'est pas possible.
- Mettre en place des présences sur le web pour informer les participants aux études sur les études en cours et futures.

2. Garde-fous pour instaurer la confiance :

[431] DSK, Beschluss der 97. Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder zu Auslegung des Begriffs "bestimmte Bereiche wissenschaftlicher Forschung" im Erwägungsgrund 33 der DS-GVO 3. avril 2019, à l'adresse : www.datenschutzkonferenz-online.de/media/dskb/20190405_auslegung_bestimmte_bereiche_wiss_forschung.pdf (consulté le 20 mai 2020). La traduction anglaise provient d'un beau résumé des mesures qui peut être consulté ici : www.technologylawdispatch.com/2019/04/privacy-data-protection/german-dpas-publish-resolution-on-concept-of-broad-consent-and-the-interpretation-of-certain-areas-of-scientific-research/.

- Vote positif d'un comité d'éthique avant l'utilisation des données à des fins de recherche ultérieure.
- Évaluation de la possibilité de travailler avec un consentement dynamique ou de la possibilité pour une personne concernée de s'opposer avant que les données ne soient utilisées pour de nouvelles questions de recherche.

3. Garanties de sécurité :

- Pas de transfert de données vers des pays tiers dont le niveau de protection des données est inférieur.
- Mesures supplémentaires concernant la minimisation, le cryptage, l'anonymisation ou la pseudonymisation des données
- Mise en œuvre de politiques spécifiques pour limiter l'accès aux données personnelles.

En tout état de cause, les participants à la recherche doivent avoir la **possibilité de retirer leur consentement,** d'accepter ou de refuser certaines recherches ou parties de recherches, et d'être assurés que leurs droits sont protégés par le respect des normes éthiques de la recherche scientifique.[432] Parfois, cela peut nuire à la solution d'IA ou obliger les responsables du traitement à effectuer des actions complexes. Par conséquent, les responsables du traitement doivent se demander si d'autres bases juridiques ne pourraient pas mieux les aider à développer l'outil tout en respectant la loi.

En résumé, les responsables du traitement doivent être **prudents lorsqu'ils utilisent le consentement comme base juridique pour justifier le traitement des données**, car le consentement n'invalide pas leurs responsabilités concernant la loyauté, la nécessité et la proportionnalité du traitement.[433] En outre, dans le cas de l'IA utilisant le Big Data, il est souvent difficile de justifier que le consentement remplit toutes les exigences nécessaires : librement donné, spécifique, informé et sans ambiguïté, et un acte affirmatif clair de la part de la personne concernée. En général, plus les développeurs d'IA veulent faire de choses avec les données, plus il est difficile de garantir que le consentement est véritablement spécifique et éclairé. Tout cela doit être pris en compte lors du choix du consentement comme base juridique du traitement des données.

Encadré 4. Le consentement comme base juridique : l'affaire OkCupid

En 2016, un groupe de chercheurs danois a publié un ensemble de données concernant environ 70 000 utilisateurs. Ces données avaient été obtenues sur le site de rencontres en ligne OkCupid[434] et comprenaient des catégories de données telles que les noms d'utilisateur, l'âge, le sexe, la localisation, le type de relation (ou de sexe) qui intéressait les personnes

[432] Kuyumdzhieva, A. (2018) 'Ethical challenges in the digital era : focus on medical research', pp.45-62 in : Koporc, Z. (ed.) *Ethics and integrity in health and life sciences research*. Emerald, Bingley.

[433] Groupe de travail Article 29 (2018) Lignes directrices sur le consentement en vertu du règlement 2016/679. WP259. Commission européenne, Bruxelles, p.3. Disponible à l'adresse : https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051 (consulté le 15 mai 2020).

[434] www.okcupid.com (consulté le 5 mai 2020).

concernées, leurs traits de personnalité, etc.

Les chercheurs ont estimé que le simple fait que ces données étaient accessibles au public (sur les profils de rencontre des utilisateurs) constituait une base juridique pour un traitement ultérieur. Il s'agit d'un excellent exemple des terribles conséquences de l'argument selon lequel ″les données sont déjà publiques". Les personnes concernées ont vu leurs données personnelles traitées, et des informations très sensibles exposées au public, sans leur consentement.

Malheureusement, cette association entre données publiques et données ouvertes est encore trop étendue. Les chercheurs devraient être conscients que le consentement fourni pour un traitement concret ne sert pas de base juridique pour d'autres traitements, et que "accessible au public" n'est pas synonyme de "données ouvertes", c'est-à-dire de données et de contenus qui peuvent être librement utilisés, modifiés et partagés par quiconque, à n'importe quelle fin, comme le définit l'Open Data Institute.[435]

Liste de contrôle : consentement

 Les responsables du traitement ont vérifié que le consentement est la base juridique la plus appropriée pour le traitement.

 Les responsables du traitement demandent le consentement des intéressés de manière libre, spécifique, éclairée et non équivoque.

 Le consentement général est utilisé uniquement lorsqu'il est difficile ou improbable de prévoir comment ces données seront traitées à l'avenir.

 Le consentement général utilisé pour le traitement de catégories spéciales de données est compatible avec les réglementations nationales.

 Lorsque le consentement général est utilisé, les personnes concernées ont la possibilité de retirer leur consentement et de choisir de participer ou non à certaines recherches et parties de celles-ci.

 Les responsables du traitement ont une relation directe avec le sujet qui fournit les données à utiliser pour la formation, la validation et le déploiement du modèle IA.

 Il n'y a pas de déséquilibre de pouvoir entre les responsables du traitement et les personnes concernées.

 Les responsables du traitement demandent aux personnes de s'inscrire positivement.

 Les responsables du traitement n'utilisent pas de cases pré-cochées ou tout autre type de consentement par défaut.

 Les responsables du traitement utilisent un langage clair et simple, facile à comprendre.

 Les responsables du traitement précisent pourquoi ils veulent les données et ce qu'ils vont en faire.

 Les responsables du traitement donnent des options distinctes ("granulaires") pour consentir séparément à différentes finalités et types de traitement.

[435] http://opendefinition.org/ (consulté le 5 mai 2020).

 Les responsables du traitement indiquent aux personnes qu'elles peuvent retirer leur consentement et comment le faire.

 Les responsables du traitement veillent à ce que les personnes puissent refuser de donner leur consentement sans subir de préjudice.

 Les responsables du traitement évitent de faire du consentement une condition préalable à un service.

b) Intérêt légitime

L'utilisation de l'intérêt légitime comme base juridique du traitement pour le développement de l'IA est applicable, à condition que le résultat du test de mise en balance le justifie (voir "Intérêt légitime et test de mise en balance" dans la partie II, section "Principales actions et outils" des présentes lignes directrices). Cela peut impliquer de définir l'objectif du traitement de l'IA dès le départ, et de veiller à ce que l'objectif initial du traitement soit réévalué si le système d'IA fournit un résultat inattendu, de manière à pouvoir identifier les intérêts légitimes poursuivis ou à pouvoir recueillir un consentement valable auprès des personnes.[436] Le test de mise en balance **doit être documenté de manière adéquate dans les registres de traitement.** Toutefois, dans certains cas, l'intérêt légitime peut ne pas être utile aux fins du traitement de l'IA. Par exemple, si les responsables du traitement prévoient de rassembler une quantité considérable de données à caractère personnel "au cas où", ils ne devraient pas considérer l'intérêt légitime comme un motif légal pour le traitement des données, car la mise en balance entre la nécessité du traitement et les impacts possibles du traitement sur les personnes ne le justifierait guère.[437]

Liste de contrôle : l'intérêt légitime comme base juridique

 Les responsables de traitement ont vérifié que l'intérêt légitime est la base la plus appropriée.

 Les responsables du traitement comprennent leur responsabilité de protéger les intérêts des personnes.

 Les responsables du traitement tiennent un registre des décisions prises et de leur motivation, afin de s'assurer qu'ils peuvent justifier leur décision.

 Les responsables du traitement ont identifié les intérêts légitimes pertinents.

 Les responsables du traitement ont vérifié que le traitement est nécessaire et qu'il n'existe pas de moyen moins intrusif pour parvenir au même résultat.

 Les responsables du traitement ont effectué un test d'équilibre et sont convaincus que les

---

[436] CIPL (2020) Intelligence artificielle et protection des données. Comment le RGPD réglemente l'IA. Centre for Information Policy Leadership, Washington, DC/Bruxelles/Londres, p.5. Disponible sur : www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl-hunton_andrews_kurth_legal_note_-_how_gdpr_regulates_ai__12_march_2020_.pdf (consulté le 15 mai 2020).

[437] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, p.22. Disponible sur : www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (consulté le 15 mai 2020).

intérêts de la personne ne l'emportent pas sur ces intérêts légitimes.

- Les responsables du traitement n'utilisent les données des personnes que de la manière à laquelle ils peuvent raisonnablement s'attendre, sauf si les responsables du traitement ont une très bonne raison.
- Les responsables du traitement n'utilisent pas les données des personnes d'une manière qu'elles trouveraient intrusive, ou qui pourrait leur causer un préjudice, à moins que les responsables du traitement aient une très bonne raison.
- Si les responsables du traitement traitent les données d'enfants, ils prennent des précautions supplémentaires pour s'assurer qu'ils protègent les intérêts des enfants.
- Les responsables du traitement ont envisagé des mesures de sauvegarde pour réduire l'impact, dans la mesure du possible.
- Les responsables du traitement ont examiné s'ils pouvaient proposer un opt-out.
- Les responsables du traitement ont examiné s'ils devaient également réaliser une AIPD.

c) Nécessité contractuelle

L'exécution d'un contrat auquel la personne concernée est partie, ou l'accomplissement de démarches à la demande de la personne concernée avant la conclusion d'un contrat, peut servir de base juridique au traitement, si l'utilisation de l'IA est objectivement nécessaire à l'une de ces finalités. Cela pourrait être le cas pour les développeurs qui engagent des sujets pour utiliser leurs données personnelles dans la phase de formation du système. Il pourrait également s'agir du responsable du traitement, qui fournit à des tiers intéressés un service comprenant la solution d'IA, et qui utilise les données de ces sujets dans le cadre du contrat de service.[438] Toutefois, cette base juridique ne devrait pas être utilisée pour des finalités différentes (telles que l'amélioration du système ou similaire) selon le principe de limitation de la finalité (voir "Principe de limitation de la finalité" dans la partie II section "Principes" des présentes lignes directrices), puisque les données utilisées pour exécuter le contrat ne sont pas nécessaires pour ces finalités alternatives.[439] Ainsi, les responsables du traitement peuvent traiter les données qui sont intrinsèquement nécessaires à l'exécution d'un contrat sous l'égide de cette base juridique si elles sont objectivement nécessaires à l'exécution du contrat, mais pas à d'autres fins.[440] En résumé, il semble difficile de voir comment l'exécution d'un contrat pourrait servir de base juridique pour la recherche et l'innovation en matière d'IA.

---

[438] Ibid. , p.20.

[439] Groupe de travail Article 29 sur la protection des données (2014) Avis 06/2014 sur la notion d'intérêts légitimes du responsable du traitement des données au titre de l'article 7 de la directive 95/46/CE. Commission européenne, Bruxelles, p.16-17. Disponible à l'adresse : https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf (consulté le 16 mai 2020).

[440] EDPB (2019) Lignes directrices 2/2019 sur le 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. Conseil européen de la protection des données, Bruxelles, p.14. Disponible à l'adresse : https://edpb.europa.eu/sites/edpb/files/consultation/edpb_draft_guidelines-art_6-1-b-final_public_consultation_version_en.pdf (consulté le 15 mai 2020).

d) Obligation légale ou intérêt vital

Selon l'article 6, paragraphe 1, point d), du RGPD, les données peuvent être traitées si elles sont "nécessaires à la sauvegarde des intérêts vitaux de la personne concernée ou d'une autre personne physique". De même, le traitement est licite s'il est "nécessaire au respect d'une obligation légale à laquelle le responsable du traitement est soumis" (article 6, paragraphe 1, point c). Si nous parlons de catégories spéciales de données, il existe alors d'autres motifs légaux de traitement, comme l'exprime l'article 9, paragraphe 2. Il est encore une fois **difficile d'imaginer un seul cas où l'une de ces bases pourrait constituer un fondement juridique pour la formation d'un système d'IA** à l'heure actuelle, même si des révisions des réglementations existantes aux niveaux national et européen pourraient changer la donne à l'avenir. En tout état de cause, pour la formation de systèmes d'IA susceptibles de sauver des vies, il serait préférable de s'appuyer sur d'autres bases juridiques, telles que le consentement ou l'intérêt public.[441]

Encadré 5. Exemples d'intérêt vital comme base juridique du traitement des données par un outil d'IA

Imaginons que, lors de la pandémie de COVID-19, une organisation développe un outil d'IA capable de diagnostiquer la maladie en utilisant la radiologie. Dans ce cas, les données relatives aux patients pourraient être traitées sur la base de l'intérêt vital, comme le prévoit l'article 9, paragraphe 2, point c), du RGPD. Toutefois, d'autres bases juridiques, tels que l'intérêt public substantiel (article 9, paragraphe 2, point g) ou i), pourraient être plus appropriés.

Informations complémentaires

AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, p.20 Disponible sur : www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf

Groupe de travail Article 29 (2014) Avis 6/2014 sur la notion d'intérêts légitimes du responsable du traitement au titre de l'article 7 de la directive 95/46. Commission européenne, Bruxelles. Disponible à l'adresse suivante : www.dataprotection.ro/servlet/ViewDocument?id=1086

CIPL (2020) Intelligence artificielle et protection des données. Comment le RGPD réglemente l'IA. Centre for Information Policy Leadership, Washington, DC / Bruxelles / Londres. Disponible à l'adresse : www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl-hunton_andrews_kurth_legal_note_-_how_gdpr_regulates_ai__12_march_2020_.pdf

EDPB (2019) Lignes directrices 2/2019 sur le 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. Conseil européen de la protection des données, Bruxelles. Disponible à l'adresse suivante : https://edpb.europa.eu/sites/edpb/files/consultation/edpb_draft_guidelines-art_6-1-b-

[441] Groupe de travail Article 29 (2014) Avis 06/2014 sur la notion d'intérêts légitimes du responsable du traitement des données au titre de l'article 7 de la directive 95/46/CE. Commission européenne, Bruxelles, p.20. Disponible à l'adresse : https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf (consulté le 15 mai 2020).

final_public_consultation_version_en.pdf

EDPB (2020) Lignes directrices 05/2020 sur le consentement au titre du règlement 2016/679 Version 1.1 adoptée le 4 mai 2020. Disponible à l'adresse : https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202005_consent_en.pdf

CEPD (2017) Necessity toolkit. Contrôleur européen de la protection des données, Bruxelles. Disponible sur : https://edps.europa.eu/data-protection/our-work/publications/papers/necessity-toolkit_en

Vous trouverez dans les documents suivants d'autres lectures sur l'intérêt légitime, avec des cas pratiques et plusieurs références aux arrêts de la Cour de justice de l'Union européenne.

Future of Privacy Forum (pas de date) Traitement des données personnelles sur la base d'intérêts légitimes en vertu du RGPD. Réseau européen de formation judiciaire, Bruxelles. Disponible à l'adresse : www.ejtn.eu/PageFiles/17861/Deciphering_Legitimate_Interests_Under_the_GDPR%20(1).pdf

ICO (aucune date) Comment appliquer les intérêts légitimes dans la pratique ? Information Commissioner's Office, Wilmslow. Disponible à l'adresse : https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/how-do-we-apply-legitimate-interests-in-practice/

ICO (aucune date) Base légale du traitement. Information Commissioner's Office, Wilmslow. Disponible à l'adresse : https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/

Kuyumdzhieva, A. (2018) 'Ethical challenges in the digital era : focus on medical research', pp. 45-62 in : Koporc, Z. (ed.) Ethics and integrity in health and life sciences research. Emerald, Bingley.

Autorité norvégienne de protection des données (2018) Intelligence artificielle et vie privée. Autorité norvégienne de protection des données, Oslo. Disponible à l'adresse : https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf

3.2.3 Minimisation des données

Le principe de minimisation des données stipule que les données à caractère personnel doivent être "adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées".[442] Dans le contexte de l'IA, cela signifie, en premier lieu, que les **responsables du traitement doivent éviter d'utiliser des données à caractère personnel si cela n'est pas nécessaire**, c'est-à-dire si l'objectif visé par le responsable du traitement peut être obtenu sans traiter de données à caractère personnel (voir la section "Licéité, loyauté et transparence" dans la partie "Principes" de la partie II des présentes lignes directrices). En effet, il arrive que des données personnelles puissent être remplacées par des données non personnelles sans que cela n'affecte les finalités de la recherche. Dans ces circonstances, l'utilisation de données anonymes est obligatoire, conformément à l'article 89.1 du RGPD.

Si l'anonymisation n'est pas possible, les responsables du traitement doivent au moins essayer de travailler avec des données pseudonymisées. En fin de compte, chaque

[442] Article 5(1)(c) du RGPD.

responsable du traitement doit définir quelles données à caractère personnel sont réellement nécessaires (et lesquelles ne le sont pas) aux fins du traitement, y compris les périodes de conservation des données pertinentes. En effet, les responsables du traitement doivent garder à l'esprit que la nécessité du traitement doit être prouvée dans le cas de la plupart des bases juridiques - y compris toutes les bases énoncées à l'article 6 du RGPD, à l'exception du consentement, et la plupart des bases incluses dans l'article 9, paragraphe 2, concernant les catégories particulières de données. En d'autres termes, pour la majorité des bases juridiques du traitement des données personnelles, les principes de minimisation des données et de licéité exigent que les responsables du traitement s'assurent que le développement de l'IA ne peut se faire sans utiliser de données personnelles.

La notion de nécessité est toutefois complexe et a une signification indépendante dans le droit de l'Union européenne.[443] En général, elle exige que le traitement soit un moyen ciblé et proportionné d'atteindre une finalité spécifique. Il ne suffit pas de faire valoir que le traitement est nécessaire parce que les responsables du traitement ont choisi d'exercer leur activité d'une manière particulière. Ils doivent être en mesure de démontrer que le traitement est **nécessaire à la réalisation de l'objectif poursuivi** et qu'il est **moins intrusif que d'autres options** pour atteindre le même objectif ; et non pas qu'il s'agit d'une partie nécessaire des méthodes qu'ils ont choisies.[444] S'il existe des alternatives réalistes et moins intrusives, le traitement des données personnelles n'est pas considéré comme nécessaire. [445]

Par conséquent, le principe de minimisation des données exige que les développeurs d'IA optent pour les outils dont le développement implique une utilisation minimale de données personnelles par rapport aux alternatives disponibles. Une fois cet objectif atteint, des processus spécifiques doivent être mis en place pour exclure la collecte et/ou le transfert de données personnelles inutiles, réduire les champs de données et prévoir des mécanismes de suppression automatisée.[446] La minimisation des données peut être particulièrement complexe dans le cas de l'apprentissage profond, où la discrimination par caractéristiques peut être impossible. Par conséquent, si des solutions alternatives peuvent donner les mêmes résultats, il est préférable d'éviter l'apprentissage profond.

En outre, le CIPL note que "les données personnelles considérées comme "nécessaires" varient selon le système d'IA et l'objectif pour lequel il est utilisé, mais la gouvernance du RGPD dans ce domaine devrait empêcher le parfait d'être l'ennemi du bien pour les

---

[443] Voir CJUE, affaire C524/06-, Heinz Huber c. Bundesrepublik Deutschland, 18 décembre 2008, para. 52.

[444] CEPD (2017) Necessity toolkit : assessing the necessity of measures that limit the fundamental right to the protection of personal data, p.5. Contrôleur européen de la protection des données, Bruxelles. Disponible à l'adresse : https://edps.europa.eu/data-protection/our-work/publications/papers/necessity-toolkit_en (consulté le 15 mai 2020) ; ICO (aucune date) Lawful basis for processing. Bureau du commissaire à l'information, Wilmslow. Disponible à l'adresse : https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/ (consulté le 15 mai 2020).

[445] Voir CJUE, Affaires jointes C92/09 -et C93/09-, Volker und Markus Schecke GbR et Hartmut Eifert c. Land Hessen, 9. novembre 2010.

[446] ENISA (2015) Le respect de la vie privée dès la conception dans le domaine du big data. Agence de l'Union européenne pour la cybersécurité, Athènes / Héraklion, p.23. Disponible à l'adresse : www.enisa.europa.eu/publications/big-data-protection (consulté le 28 mai 2020).

concepteurs d'IA - le fait que les données personnelles doivent être limitées ne signifie pas que le système d'IA lui-même devient inutile, d'autant plus que tous les systèmes d'IA n'ont pas besoin de fournir une sortie précise."[447] Afin de déterminer précisément l'éventail et la quantité de données personnelles nécessaires, le **fait d'avoir un expert capable de sélectionner les caractéristiques pertinentes devient extrêmement important.** Cela devrait réduire considérablement le risque pour la vie privée des personnes concernées - sans perdre en qualité.

Il existe un outil efficace pour réguler la quantité de données recueillies et ne l'augmenter que si cela semble nécessaire : la **courbe d'apprentissage**.[448] Le responsable du traitement doit commencer par recueillir et utiliser une quantité limitée de données d'apprentissage, puis surveiller la précision du modèle à mesure qu'il est alimenté en nouvelles données. Cela aidera également le responsable des données à éviter la "malédiction de la dimensionnalité", c'est-à-dire "une mauvaise performance des algorithmes et leur grande complexité associées à un cadre de données ayant un grand nombre de dimensions/caractéristiques, ce qui rend souvent la fonction cible assez complexe et peut conduire à un surajustement du modèle tant que l'ensemble de données se trouve souvent dans la courbe de dimensionnalité inférieure". [449]

Parmi les mesures supplémentaires liées au principe de minimisation, on peut citer :

- limiter l'extension des catégories de données (par exemple, les noms, les adresses physiques et les adresses, les champs concernant leur santé, leur situation professionnelle, leurs croyances, leur idéologie, etc.)
- limiter le degré de détail ou de précision des informations, la granularité de la collecte dans le temps et la fréquence, et l'ancienneté des informations utilisées
- limiter l'extension du nombre de parties intéressées de ceux qui traitent les données
- limiter l'accès aux différentes catégories de données au personnel du responsable du traitement/gestionnaire ou même à l'utilisateur final (si les modèles d'IA contiennent des données de tiers) à toutes les étapes du traitement.[450]

Bien entendu, l'adoption de ces mesures pourrait nécessiter un effort considérable en termes d'unification et d'homogénéisation des données, etc., mais elle contribuera à la

---

[447] CIPL (2020) Intelligence artificielle et protection des données : comment le RGPD réglemente l'IA. Centre for Information Policy Leadership, Washington DC / Bruxelles / Londres, p.13. Disponible à l'adresse : www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl-hunton_andrews_kurth_legal_note_-_how_gdpr_regulates_ai__12_march_2020_.pdf (consulté le 15 mai 2020).

[448] Voir : www.ritchieng.com/machinelearning-learning-curve/ (consulté le 28 mai 2020).

[449] Oliinyk, H. (2018) Pourquoi et comment se débarrasser correctement de la malédiction de la dimensionnalité (avec visualisation d'un ensemble de données sur le cancer du sein). Vers la science des données, 20 mars. Disponible à l'adresse : https://towardsdatascience.com/why-and-how-to-get-rid-of-the-curse-of-dimensionality-right-with-breast-cancer-dataset-7d528fb5f6c0 (consulté le 15 mai 2020).

[450] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, p.39-40. Disponible sur : www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (consulté le 15 mai 2020).

mise en œuvre du principe de minimisation des données de manière beaucoup plus efficace. [451]

Enfin, il est utile de rappeler que les responsables du traitement doivent **éviter de conserver de longues bases de données historiques,** par exemple au-delà de la période requise à des fins commerciales normales, ou pour remplir des obligations légales, ou simplement parce que leur outil analytique est capable de produire un grand nombre de données et que sa capacité de stockage le permet. Au lieu de cela, les entreprises utilisant le big data doivent appliquer des calendriers de conservation appropriés (voir la section "Limitation du stockage" dans les "Principes", partie II des présentes lignes directrices).

Encadré 6. Un exemple des avantages de la minimisation des données dans l'IA

Un outil d'IA développé par l'administration fiscale norvégienne pour filtrer les erreurs dans les déclarations d'impôts a testé cinq cents variables lors de la phase de formation. Cependant, seules trente d'entre elles ont été incluses dans le modèle d'IA final, car elles se sont avérées les plus pertinentes pour la tâche à accomplir. Il est probable que les développeurs de l'outil auraient pu éviter de collecter autant de données personnelles s'ils avaient effectué une meilleure sélection des variables pertinentes au début du processus de développement.

Source : Autorité norvégienne de protection des données (2018) Intelligence artificielle et vie privée. Autorité norvégienne de protection des données, Oslo. Disponible à l'adresse : https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf

Liste de contrôle : minimisation des données

- Les responsables du traitement ont veillé à n'utiliser les données personnelles qu'en cas de besoin.
- Les responsables du traitement ont réfléchi à la proportionnalité entre la quantité de données et la précision de l'outil d'IA.
- Les responsables du traitement examinent périodiquement les données qu'ils détiennent et suppriment tout ce dont ils n'ont pas besoin.
- Les responsables du traitement au stade de la formation du système d'IA débuguent toutes les informations qui ne sont pas strictement nécessaires à cette formation.
- Les responsables du traitement vérifient si des données à caractère personnel sont traitées au stade de la distribution du système IA et les suppriment, sauf s'il existe un besoin justifié et une légitimité à les conserver à d'autres fins compatibles.

Informations complémentaires

[451] Autorité norvégienne de protection des données (2018) Intelligence artificielle et vie privée. Autorité norvégienne de protection des données, Oslo. Disponible à l'adresse : https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf (consulté le 15 mai 2020).

ENISA (2015) Le respect de la vie privée dès la conception dans le domaine du big data. Agence de l'Union européenne pour la cybersécurité, Athènes / Héraklion, p.23. Disponible sur : www.enisa.europa.eu/publications/big-data-protection

ICO (pas de date) Principe (c) : minimisation des données. Information Commissioner's Office, Wilmslow. Disponible sur : https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/data-minimisation/

Autorité norvégienne de protection des données (2018) Intelligence artificielle et vie privée. Autorité norvégienne de protection des données, Oslo. Disponible à l'adresse : https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf

Pure Storage (2015) Le grand échec du big data : les difficultés rencontrées par les entreprises pour accéder aux informations dont elles ont besoin. Pure Storage, Mountain View, CA. Disponible à l'adresse : http://info.purestorage.com/rs/225-USM-292/images/Big%20Data%27s%20Big%20Failure_UK%281%29.pdf

3.2.4 La loyauté dans le respect des droits des personnes concernées

La loyauté est un concept essentiel en matière de protection des données (voir la sous-section "Loyauté" dans la section "Licéité, loyauté et transparence" des "Principes", ainsi que "Droits des personnes concernées", tous deux dans la partie II des présentes lignes directrices), un concept qui peut difficilement être atteint sans une prise de conscience que le développement d'outils d'IA peut porter atteinte aux intérêts, aux droits et aux libertés des personnes concernées. C'est pourquoi il est logique de veiller à ce que des garanties adéquates soient mises en œuvre non seulement pour éviter les conséquences injustes, mais aussi pour fournir aux personnes concernées des droits exécutoires qui garantissent une protection adéquate contre le traitement déloyal.

Dans cette section, nous examinons comment les principaux droits reconnus par le RGPD s'appliquent au cadre de développement de l'IA. À cette fin, nous nous concentrerons sur certains droits qui sont particulièrement pertinents dans ce domaine : (a) le droit à l'information ; (b) le droit d'accès ; (c) le droit à la portabilité des données ; (d) le droit de rectification ; et (e) le droit à l'effacement ; et (f) le droit d'opposition.

Toutefois, avant d'envisager cela, les chercheurs doivent vérifier si leur recherche **peut être considérée comme une recherche scientifique au sens de l'article 89 du RGPD**. C'est extrêmement important : si c'est le cas, le droit de l'UE ou des États membres peut prévoir des dérogations aux droits visés aux articles 15, 16, 18 et 21 (adresse, rectification, restriction et objet - et, indirectement, portabilité). Celles-ci sont soumises aux conditions et garanties visées à l'article 89, paragraphe 1, dans la mesure où ces droits sont susceptibles de rendre impossible ou de nuire gravement à la réalisation des finalités spécifiques, et où ces dérogations sont nécessaires à la réalisation de ces finalités (voir la section "Protection des données et recherche scientifique" dans les "Concepts principaux", partie II des présentes lignes directrices).

a) Droit à l'information

Selon l'article 13 du RGPD, avant de traiter des données à caractère personnel, le responsable du traitement doit fournir aux personnes concernées des informations complètes sur le traitement et leurs droits dans un format compréhensible. Si "le responsable du traitement a l'intention de traiter ultérieurement les données à caractère personnel pour une finalité autre que celle pour laquelle les données à caractère personnel ont été collectées, le responsable du traitement fournit à la personne concernée, avant ce traitement ultérieur, des informations sur cette autre finalité et toute autre information pertinente" (article 13, paragraphe 3).

Toutefois, les responsables du traitement sont dispensés de fournir des informations aux personnes concernées si : la fourniture de ces informations s'avère impossible ; elle impliquerait un effort disproportionné, notamment pour le traitement à des fins d'archivage dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques ; ou l'obligation de fournir des informations est susceptible de rendre impossible ou de nuire gravement à la réalisation des finalités de ce traitement. Dans ces circonstances, les responsables du traitement doivent prendre des mesures appropriées pour protéger les droits, les libertés et les intérêts légitimes de la personne concernée, y compris en rendant les informations accessibles au public. Cette dérogation est toutefois subordonnée à l'adoption des garanties imposées par l'article 89 (voir la section "Protection des données et recherche scientifique" de la partie II "Concepts principaux" des présentes lignes directrices).

b) Droit d'accès

Le droit d'accès des personnes concernées à leurs données doit être **garanti à toutes les étapes du cycle de vie d'un outil d'IA.** Les responsables du traitement sont encouragés à mettre en œuvre des mesures techniques adéquates pour s'assurer que cet accès est facilement accessible par la personne concernée. En effet, l'article 15 du RGPD donne à la personne concernée le droit d'obtenir des détails sur toute donnée personnelle utilisée pour le profilage, y compris les catégories de données utilisées pour construire un profil. En outre, en vertu de l'article 15, paragraphe 3, le responsable du traitement a l'obligation de mettre à disposition les données utilisées en entrée pour créer le profil, ainsi que l'accès aux informations sur le profil et les détails des segments dans lesquels la personne concernée a été placée. De même, le considérant 63 du RGPD stipule que "[l]orsque cela est possible, le responsable du traitement devrait être en mesure de fournir un accès à distance à un système sécurisé qui permettrait aux personnes concernées d'accéder directement à leurs données personnelles". **Cela inclut les données observées, dérivées et déduites**. [452]

Encadré 7. La question des données déduites

L'une des questions les plus urgentes auxquelles nous sommes confrontés dans le domaine de l'IA est le statut concret des données déduites. Il s'agit de données

[452] ICO (2014) Big data et protection des données. Bureau du commissaire à l'information, Wilmslow, p.99-10. Disponible à l'adresse : https://rm.coe.int/big-data-and-data-protection-ico-information-commissioner-s-office/1680591220 (consulté le 28 mai 2020).

qui ne sont pas fournies par les personnes concernées, mais qui leur sont "attribuées" à partir de données disponibles, provenant soit des mêmes personnes, soit d'autres personnes. Parfois, ces données déduites fournissent des informations sur une personne identifiable. Que ces informations soient exactes ou non, ces données doivent être considérées comme des données à caractère personnel et le RGPD s'applique donc à elles. En conséquence, les droits des personnes concernées doivent être strictement respectés, y compris le droit d'accès à ces données.[453] Toutefois, comme indiqué ailleurs dans les présentes lignes directrices, les données déduites ne sont pas incluses dans le droit à la portabilité (voir "Droit à la portabilité" dans la partie II section "Droits des personnes concernées" des présentes lignes directrices).

L'un des principaux problèmes inhérents au traitement de l'IA et du big data est que le droit d'accès peut parfois entrer en conflit avec l'intérêt d'une entreprise à conserver ses secrets commerciaux. En effet, le considérant 63 du RGPD prévoit une certaine protection pour les responsables du traitement qui ne souhaitent pas dévoiler leurs secrets commerciaux ou leur propriété intellectuelle, ce qui peut être particulièrement pertinent en ce qui concerne le profilage.[454] Toutefois, les développeurs d'IA **ne peuvent pas invoquer la protection de leurs secrets commerciaux comme excuse pour refuser l'accès ou refuser de fournir des informations** aux personnes concernées**.** Les organisations doivent plutôt trouver des solutions pragmatiques.[455]

Le droit d'accès peut être plus ou moins applicable, selon le stade du cycle de vie auquel se trouve le développement de l'IA. Par exemple, il peut être difficile de donner accès aux données d'entraînement à une personne concernée, car elles ne contiennent généralement que des informations pertinentes pour les prédictions (par exemple, des transactions antérieures, des données démographiques, une localisation), mais pas de coordonnées ou d'identifiants uniques du client. De plus, elles sont souvent prétraitées pour les rendre plus accessibles aux algorithmes d'apprentissage automatique. Cependant, cela ne signifie pas du tout que ces données peuvent être considérées comme anonymes. Ainsi, elles continuent d'être des données à caractère personnel. Par exemple, dans le cas d'un modèle de prédiction d'achat, la formation peut inclure un modèle d'achat propre à un client. Dans cet exemple, si un client devait fournir une liste de ses achats récents dans le cadre de sa demande, l'organisation pourrait être en mesure d'identifier la partie des données d'apprentissage qui se rapporte à cet individu.

Dans ces circonstances, les **responsables du traitement doivent répondre aux demandes d'accès des personnes concernées à leurs données personnelles, en**

---

[453] Voir : Custers, B. (2018) 'Profiling as inferred data. Amplifier effects and positive feedback loops', pp.112-115 in Bayamlıoğlu, E. et al. (eds) *Being profiled : cogitas ergo sum. 10 ans de profilage du citoyen européen*. Amsterdam University Press, Amsterdam. DOI 10.5117/9789463722124/CH19. Disponible sur : https://ssrn.com/abstract=3466857 ou http://dx.doi.org/10.2139/ssrn.3466857 (consulté le 28 mai 2020).

[454] A29WP (2016) Lignes directrices sur la prise de décision individuelle automatisée et le profilage aux fins du règlement 2016/679. Commission européenne, Bruxelles, p.17. Disponible à l'adresse : https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053 (consulté le 28 mai 2020).

[455] Autorité norvégienne de protection des données (2018) Intelligence artificielle et vie privée. Autorité norvégienne de protection des données, Oslo, p.19. Disponible à l'adresse : https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf (consulté le 28 mai 2020).

supposant qu'ils aient pris des mesures raisonnables pour vérifier l'identité de la personne concernée, et qu'aucune autre exception ne s'applique. Et, comme l'indique l'ICO, "les demandes d'accès, de rectification ou d'effacement de données de formation ne doivent pas être considérées comme manifestement infondées ou excessives simplement parce qu'elles peuvent être plus difficiles à satisfaire ou que la motivation de la demande peut être peu claire par rapport aux autres demandes d'accès qu'une organisation reçoit habituellement".[456] Toutefois, les **organisations ne sont pas tenues de collecter ou de conserver des données à caractère personnel supplémentaires pour permettre l'identification des personnes concernées par les données de formation dans le seul but de se conformer au règlement.** Si les responsables du traitement ne peuvent pas identifier une personne concernée dans les données de formation, et que la personne concernée ne peut pas fournir d'informations supplémentaires qui permettraient son identification, ils ne sont pas obligés de satisfaire une demande qu'il n'est pas possible de satisfaire.[457]

Liste de contrôle : droit d'accès[458]

Préparation des demandes d'accès aux données

- Les responsables du traitement savent comment reconnaître une demande d'accès à un sujet et ils comprennent quand le droit d'accès s'applique.
- Les responsables du traitement comprennent que le droit d'accès doit être appliqué à chaque étape du cycle de vie de la solution d'IA, si celle-ci utilise des données à caractère personnel.
- Les responsables du traitement ont une politique sur la façon d'enregistrer les demandes qu'ils reçoivent verbalement.
- Les responsables du traitement comprennent quand ils peuvent refuser une demande et sont conscients des informations qu'ils doivent fournir aux personnes lorsqu'ils le font.
- Les responsables du traitement comprennent la nature des informations supplémentaires qu'ils doivent fournir en réponse à une demande d'accès à un sujet.

Respecter les demandes d'accès aux données

- Les responsables du traitement ont mis en place des processus pour s'assurer qu'ils répondent à une demande d'accès d'un sujet sans retard excessif et dans un délai d'un mois après réception.
- Les responsables du traitement sont conscients des circonstances dans lesquelles ils

[456] ICO (2019) Permettre les droits d'accès, d'effacement et de rectification dans les systèmes d'IA. Bureau du commissaire à l'information, Wilmslow. Disponible à l'adresse : https://ico.org.uk/about-the-ico/news-and-events/ai-blog-enabling-access-erasure-and-rectification-rights-in-ai-systems/ (consulté le 28 mai 2020).

[457] Ibid.

[458] ICO (pas de date) Droit d'accès. Information Commissioner's Office, Wilmslow. Disponible à l'adresse : https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-of-access/ (consulté le 28 mai 2020).

peuvent prolonger le délai de réponse à une demande.

- Les responsables du traitement comprennent qu'il est particulièrement important d'utiliser un langage clair et simple s'ils divulguent des informations à un enfant.
- Les responsables du traitement comprennent ce qu'ils doivent prendre en compte si une demande comprend des informations sur d'autres personnes.
- Les responsables du traitement comprennent comment appliquer le droit d'accès lors de stages de formation.

Informations complémentaires

Groupe de travail Article 29 (2014) Avis 06/2014 sur la notion d'intérêts légitimes du responsable du traitement des données au titre de l'article 7 de la directive 95/46/CE. Commission européenne, Bruxelles. Disponible à l'adresse suivante : https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf

ICO (2013) Big data, intelligence artificielle, apprentissage automatique et protection des données. Bureau du commissaire à l'information, Wilmslow. Disponible à l'adresse suivante : https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

ICO (pas de date) Droit d'accès. Information Commissioner's Office, Wilmslow. Disponible sur : https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-of-access/

Autorité norvégienne de protection des données. (2018) Intelligence artificielle et vie privée. Autorité norvégienne de protection des données, Oslo. Disponible à l'adresse : https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf

c) Droit à la portabilité des données

L'article 20 du RGPD a créé un nouveau droit : le droit à la portabilité des données.[459] Ce droit permet aux personnes concernées de contrôler l'utilisation de leurs données en les redirigeant là où elles sont le plus utiles (voir la section "Droit à la portabilité des données" dans les "Droits des personnes concernées" de la partie II des présentes lignes directrices). Cependant, le droit à la portabilité des données pourrait être difficile à mettre en œuvre dans le domaine de l'IA, pour plusieurs raisons. Il faut garder à l'esprit le coût et la faisabilité de la fourniture d'ensembles de données extrêmement vastes et complexes accumulés sur de nombreuses années. Il pourrait donc être difficile pour une entreprise de satisfaire à ses exigences en matière de droit à la portabilité des données.

Il existe différents types de données personnelles qu'un système d'apprentissage automatique peut traiter. Selon le groupe de travail Article 29 sur la protection des données, certaines catégories de données sont liées au droit à la portabilité des données, à savoir : les données à caractère personnel concernant la personne concernée et les

[459] Groupe de travail Article 29 (2015) Lignes directrices sur le droit à la portabilité des données. Commission européenne, Bruxelles. Disponible à l'adresse : http://ec.europa.eu/newsroom/document.cfm?doc_id=45685 (consulté le 28 mai 2020).

données qu'elle a fournies à un responsable du traitement. En général, le terme "fournies par la personne concernée" doit être interprété de **manière large**. Ainsi, il inclut les données recueillies en observant le comportement des personnes concernées (par exemple, les données brutes traitées par les compteurs intelligents, les journaux d'activité ou l'historique des sites web). Toutefois, les "données déduites" et les "données dérivées" doivent être exclues si elles comprennent des données à caractère personnel créées par un prestataire de services (par exemple, les résultats algorithmiques). Contrairement aux données observées ou recueillies, les **données déduites sont créées par le service lui-même, sur la base des données observées, et non fournies par la personne concernée.**[460] Par conséquent, le droit à la portabilité des données **ne s'applique pas aux données déduites par un processus d'apprentissage automatique.**

Liste de contrôle : portabilité des données[461]

Se préparer aux demandes de portabilité des données

- Les responsables du traitement savent reconnaître une demande de portabilité des données et comprennent quand ce droit s'applique.
- Les responsables de traitement prennent en compte l'exigence de portabilité des données dès les premières étapes de la conception et du design du traitement de l'IA.
- Les responsables du traitement ont une politique sur la façon d'enregistrer les demandes qu'ils reçoivent verbalement.
- Les responsables du traitement comprennent quand ils peuvent refuser une demande et sont conscients des informations qu'ils doivent fournir aux personnes s'ils procèdent à un tel refus.

Respecter les demandes de portabilité des données

- Les responsables du traitement peuvent transmettre les données à caractère personnel dans des formats structurés, couramment utilisés et lisibles par machine.
- Les responsables du traitement informent au préalable les utilisateurs lorsqu'il n'est pas techniquement possible d'exercer le droit à la portabilité au moyen d'un protocole.
- Les responsables du traitement utilisent des méthodes sécurisées pour transmettre les données personnelles.
- Les responsables du traitement disposent de processus permettant de garantir qu'ils répondent à une demande de portabilité des données sans retard excessif et dans un délai d'un mois à compter de sa réception.

[460] Groupe de travail Article 29 (2015) Lignes directrices sur le droit à la portabilité des données. Commission européenne, Bruxelles, p.8. Disponible à l'adresse : http://ec.europa.eu/newsroom/document.cfm?doc_id=45685 (consulté le 28 mai 2020).

[461] ICO (pas de date) Droit à la portabilité des données. Information Commissioner's Office, Wilmslow. Disponible à l'adresse : https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-data-portability/ (consulté le 28 mai 2020).

 Les responsables du traitement connaissent les circonstances dans lesquelles ils peuvent prolonger le délai de réponse à une demande.

Informations complémentaires

Groupe de travail Article 29 (2016) Lignes directrices sur le droit à la portabilité des données. Commission européenne, Bruxelles. Disponible à l'adresse : https://ec.europa.eu/information_society/newsroom/image/document/2016-51/wp242_en_40852.pdf

EBF (2017) Commentaires de la Fédération bancaire européenne sur les lignes directrices du groupe de travail 29 sur le droit à la portabilité des données. Fédération bancaire européenne, Bruxelles, p.4, disponible à l'adresse : www.ebf.eu/wp-content/uploads/2017/04/EBF_025448E-EBF-Comments-to-the-WP-29-Guidelines_Right-of-data-portabi.._.pdf.

ICO (pas de date) Droit à la portabilité des données. Information Commissioner's Office, Wilmslow. Disponible à l'adresse : https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-data-portability/

Wallace, N. et Castro, D. (2018) L'impact du nouveau règlement de l'UE sur la protection des données sur l'IA. Center for Data Innovation, Washington, DC / Bruxelles / Londres. Disponible à l'adresse : www2.datainnovation.org/2018-impact-gdpr-ai.pdf.

d) Droit de rectification

Le droit de corriger des données inexactes est particulièrement important dans le cas de l'IA, car les algorithmes d'apprentissage automatique déduisent souvent des données. Ces données peuvent affecter la personne concernée, surtout si elles sont produites à des étapes avancées du cycle de vie de l'IA. Les données inexactes déduites pendant la phase de formation ne sont pas aussi préoccupantes que dans les phases finales. Étant donné que l'objectif des données de formation est d'entraîner des modèles basés sur des modèles généraux dans de grands ensembles de données, les inexactitudes individuelles sont moins susceptibles d'avoir un effet direct sur une personne concernée.[462] Par exemple, si les données personnelles utilisées pour fournir des informations aux clients ne sont pas correctes, comme un numéro de téléphone erroné dans un ensemble de données, la personne concernée pourrait subir un préjudice plus grave que si un numéro de téléphone déduit est utilisé pour entraîner un modèle. Toutefois, cela ne signifie certainement pas que le droit de rectification ne s'applique pas à ce stade.

Certains types concrets d'algorithmes, tels que les machines à vecteur de support (SVM), utilisent certains exemples clés des données d'entraînement afin d'aider à distinguer les nouveaux exemples pendant le déploiement. Si la personne concernée demande la rectification ou l'effacement de l'une de ces données, il ne sera pas possible

[462] Binns, R. (2019) Permettre les droits d'accès, d'effacement et de rectification dans les systèmes d'IA. Bureau du commissaire à l'information, Wilmslow. Disponible à l'adresse : https://ico.org.uk/about-the-ico/news-and-events/ai-blog-enabling-access-erasure-and-rectification-rights-in-ai-systems/ (consulté le 15 mai 2020).

de réaliser ce qui précède sans devoir réentraîner le modèle avec les données rectifiées ou sans effacer complètement le modèle.[463] Cela ne rend toutefois pas le droit de rectification inapplicable.

Il est particulièrement important de garder à l'esprit que si le responsable du traitement constate que, contrairement à l'avis de la personne concernée, les données ne sont pas inexactes au regard des finalités du traitement, il **n'est pas tenu de les rectifier.**[464] Toutefois, la charge de la preuve repose sur les épaules des responsables du traitement. Ils doivent fournir une bonne raison pour refuser la rectification, et il est difficile de conclure que le dommage que cela pourrait causer au système d'IA pourrait servir de raison convaincante. Le CEPD a critiqué les systèmes qui ne prévoient pas la possibilité de faire rectifier un ensemble de données personnelles individuelles sans créer un préjudice considérable à l'ensemble du système.[465] En tout état de cause, si le responsable du traitement choisit de refuser la demande de la personne concernée, il doit répondre à cette dernière en lui fournissant une raison justifiée de ne pas rectifier les données et, si elle le souhaite, la personne concernée peut alors saisir l'autorité de contrôle. [466]

Liste de contrôle : droit de rectification[467]

Préparation des demandes de rectification

 Les responsables du traitement savent comment reconnaître une demande de rectification et comprennent quand ce droit s'applique.

 Les responsables du traitement ont une politique sur la façon d'enregistrer les demandes qu'ils reçoivent verbalement.

 Les responsables du traitement comprennent quand ils peuvent refuser une demande, et sont conscients des informations qu'ils doivent fournir aux personnes lorsqu'elles leur sont demandées.

Respecter les demandes de rectification

 Les responsables du traitement sont prêts à aborder le droit de rectification des données

[463] Ibid.

[464] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, p.27. Disponible sur : www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (consulté le 28 mai 2020).

[465] CEPD (2014) Lignes directrices sur les droits des personnes à l'égard du traitement des données à caractère personnel. Contrôleur européen de la protection des données, Bruxelles, p.18. Disponible sur : https://edps.europa.eu/sites/edp/files/publication/14-02-25_gl_ds_rights_en.pdf (consulté le 10 mai 2020).

[466] Office of the Data Protection Ombudsman (pas de date) Right to Rectification. Office of the Data Protection Ombudsman, Helsinki. Disponible à l'adresse : https://tietosuoja.fi/en/right-to-rectification (consulté le 28 mai 2020).

[467] ICO (aucune date) Droit de rectification. Information Commissioner's Office, Wilmslow. Disponible à l'adresse : https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-rectification/ (consulté le 28 mai 2020).

des personnes concernées, notamment celles générées par les déductions et les profils réalisés par la solution d'IA.

 Les responsables du traitement ont mis en place des processus pour s'assurer qu'ils répondent à une demande de rectification sans retard excessif et dans un délai d'un mois après réception.

 Les responsables du traitement connaissent les circonstances dans lesquelles ils peuvent prolonger le délai de réponse à une demande.

 Les responsables du traitement disposent de systèmes appropriés pour rectifier ou compléter les informations, ou fournir une déclaration complémentaire.

 Les responsables du traitement ont mis en place des procédures pour informer tout destinataire en cas de rectification des données qu'ils ont partagées avec eux.

Informations complémentaires

Binns, R. (2019) Permettre les droits d'accès, d'effacement et de rectification dans les systèmes d'IA. Bureau du commissaire à l'information, Wilmslow. Disponible à l'adresse : https://ico.org.uk/about-the-ico/news-and-events/ai-blog-enabling-access-erasure-and-rectification-rights-in-ai-systems/

ICO (aucune date) Droit de rectification. Information Commissioner's Office, Wilmslow. Disponible à l'adresse : https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-rectification/

e) Droit à l'effacement

Les personnes concernées ont le droit permanent de demander au responsable du traitement l'effacement de leurs données personnelles. Toutefois, cela peut s'avérer extrêmement compliqué dans certains cas.[468] En effet, il faut garder à l'esprit qu'il peut parfois être impossible d'atteindre les objectifs juridiques du droit à l'effacement - également connu sous le nom de droit à l'oubli - dans les environnements IA, car l'opacité du traitement peut cacher certaines données à caractère personnel au sous-traitant (voir "e

Comprendre la transparence et l'opacité" dans la présente partie III sur l'IA).

Cependant, le principal problème du droit à l'effacement est qu'**il pourrait ruiner tout un système d'IA formé sur la base des données qu'un sujet demande à effacer.** Pour faire simple, les algorithmes ont besoin de conserver les données qu'ils ont utilisées pour leur formation. Si ces données sont effacées, les algorithmes risquent d'être moins précis, voire de tomber en panne. Les responsables du traitement doivent donc garder à l'esprit qu'il pourrait être impossible de modifier une base de données sérieusement affectée par l'effacement des données.

Les responsables du traitement pourraient considérer cela comme inacceptable, mais le fait est que le RGPD ne prévoit aucune exception au droit à l'effacement sur la base des

[468] Fosch-Villaronga, E., Kieseberg, P. et Li, T. (2018) " Humans forget, machines remember : artificial intelligence and the right to be forgotten ", *Computer Law & Security Review* 34(2) : 304-313.

dommages causés à une base de données contenant des données personnelles. Certains auteurs, comme Humerick, ont suggéré que "plutôt que d'exiger un effacement complet des données à caractère personnel, les responsables du traitement et les sous-traitants devraient pouvoir conserver les informations jusqu'au moment de l'effacement. De cette façon, l'apprentissage automatique de l'IA resterait au point où il a progressé, plutôt que de créer une amnésie forcée." Selon lui, cela pourrait bien servir à protéger les intérêts des personnes concernées sans faire régresser les données de l'IA. Cependant, il n'est pas facile d'être sûr que cette solution est conforme aux exigences du RGPD.

Toute recommandation devrait toujours se concentrer sur les premières étapes du cycle de vie du produit. Techniquement, il est difficile de trouver des solutions sûres aux dilemmes posés par le droit à l'effacement une fois qu'une base de données a été créée. Par conséquent, les responsables du traitement devraient toujours essayer d'arriver à une conclusion simple : **la meilleure façon d'éviter des dommages catastrophiques est de se préparer à une éventuelle perte de données dès le début**.

Enfin, le responsable du traitement doit toujours garder à l'esprit les restrictions au droit à l'effacement introduites par l'article 17, paragraphe 3, du RGPD. En outre, les autorités nationales pourraient poser des restrictions supplémentaires qui doivent être prises en compte.

Liste de contrôle : droit à l'effacement[469]

Préparation des demandes d'effacement

 Les responsables du traitement savent comment reconnaître une demande d'effacement et ils comprennent quand ce droit s'applique.

 Les responsables du traitement ont une politique sur la façon d'enregistrer les demandes qu'ils reçoivent verbalement.

 Les responsables du traitement comprennent quand ils peuvent refuser une demande et sont conscients des informations qu'ils doivent fournir aux personnes lorsqu'ils le font.

Traitement des demandes d'effacement

 Les responsables du traitement ont mis en place des processus pour s'assurer qu'ils répondent à une demande d'effacement sans retard excessif et dans un délai d'un mois à compter de sa réception.

 Les responsables du traitement connaissent les circonstances dans lesquelles ils peuvent prolonger le délai de réponse à une demande.

 Les responsables du traitement comprennent qu'un accent particulier est mis sur le droit à l'effacement si la demande concerne des données collectées auprès d'enfants.

 Les responsables du traitement ont des procédures pour informer tout destinataire s'ils effacent les données qu'ils ont partagées avec eux.

[469] ICO (pas de date) Droit à l'effacement. Information Commissioner's Office, Wilmslow. Disponible à l'adresse : https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-erasure/ (consulté le 28 mai 2020).

 Les responsables du traitement disposent de méthodes appropriées pour effacer les informations.

Informations complémentaires

Une interview de Tiffany Li sur le droit à l'effacement et l'IA peut être consultée ici : www.youtube.com/watch?v=Sozg6yJJkHk.

Binns, R. (2019) Permettre les droits d'accès, d'effacement et de rectification dans les systèmes d'IA. Blog de l'ICO, 15 octobre. Bureau du commissaire à l'information, Wilmslow. Disponible à l'adresse : https://ico.org.uk/about-the-ico/news-and-events/ai-blog-enabling-access-erasure-and-rectification-rights-in-ai-systems/

Fosch-Villaronga, E., Kieseberg, P. et Li, T. (2018) " Humans forget, machines remember : artificial intelligence and the right to be forgotten ", *Computer Law & Security Review* 34(2) : 304-313.

Humerick, M. (2018) Taking AI personally : how the E.U. must learn to balance the interests of personal data privacy & artificial intelligence, 34 Santa Clara High Tech. L.J.393. Disponible à l'adresse : https://digitalcommons.law.scu.edu/chtlj/vol34/iss4/3

ICO (pas de date) Droit à l'effacement. Information Commissioner's Office, Wilmslow. Disponible sur : https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-erasure/

Wallace, N. et Castro, D. (2018) L'impact du nouveau règlement de l'UE sur la protection des données sur l'IA. Center for Data Innovation, Washington, DC / Bruxelles / Londres. Disponible à l'adresse : www2.datainnovation.org/2018-impact-gdpr-ai.pdf.

e) Droit d'opposition

Les personnes concernées ont le droit de s'opposer au traitement de leurs données personnelles lorsque le responsable du traitement les traite sur la base d'un intérêt légitime, ou pour une tâche d'intérêt public. Cela ne s'applique pas aux cas où le motif juridique du traitement était le consentement éclairé, car dans ces cas, les personnes concernées peuvent simplement retirer leur consentement et le responsable du traitement ne peut plus traiter leurs données. Une fois que les personnes concernées ont fait leur demande, les responsables du traitement doivent cesser de traiter les données, à moins qu'ils ne puissent prouver qu'ils ont des raisons impérieuses et justifiables de continuer à le faire, et que ces raisons l'emportent sur les intérêts, les droits et les libertés des personnes concernées. [470]

Lorsque les responsables du traitement reçoivent une opposition au traitement des données à caractère personnel, et à condition qu'aucun motif de refus ne s'applique, **ils doivent cesser immédiatement de traiter les données**. Cela peut signifier qu'ils doivent effacer les données personnelles stockées car la définition large du traitement en vertu du RGPD inclut le stockage des données.

[470] Autorité norvégienne de protection des données (2018) Intelligence artificielle et vie privée. Autorité norvégienne de protection des données, Oslo, p.29. Disponible à l'adresse : https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf (consulté le 28 mai 2020).

Liste de contrôle : droit d'opposition

Préparation aux oppositions au traitement

 Les responsables du traitement savent reconnaître une opposition et ils comprennent quand le droit s'applique.

 Les responsables du traitement ont une politique concernant la manière d'enregistrer les oppositions qu'ils reçoivent verbalement.

 Les responsables du traitement comprennent quand ils peuvent refuser une opposition et sont conscients des informations qu'ils doivent fournir aux personnes lorsqu'ils le font.

 Les responsables du traitement disposent d'une information claire dans leur avis de confidentialité sur le droit d'opposition des personnes, qui est présenté séparément des autres informations sur leurs droits.

 Les responsables du traitement comprennent quand ils doivent informer les personnes de leur droit d'opposition, en plus de l'inclure dans leur avis de confidentialité.

Respecter les demandes d'opposition au traitement

 Les responsables du traitement ont mis en place des processus pour s'assurer qu'ils répondent à une opposition sans retard excessif et dans un délai d'un mois après réception.

 Les responsables du traitement sont conscients des circonstances dans lesquelles ils peuvent prolonger le délai de réponse à une opposition.

 Les responsables du traitement ont mis en place des méthodes appropriées pour effacer, supprimer ou cesser de traiter les données à caractère personnel.

Informations complémentaires

CEPD (2020) Un avis préliminaire sur la protection des données et la recherche scientifique. Contrôleur européen de la protection des données, Bruxelles. Disponible à l'adresse : https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf

ICO (aucune date) Le droit de s'opposer à l'utilisation de vos données. Information Commissioner's Office, Wilmslow. Disponible à l'adresse : https://ico.org.uk/your-data-matters/the-right-to-object-to-the-use-of-your-data/

Autorité norvégienne de protection des données (2018) Intelligence artificielle et vie privée. Autorité norvégienne de protection des données, Oslo. Disponible à l'adresse : https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf

4 Transparence

"Cette exigence est étroitement liée au principe d'explicabilité et englobe la transparence des éléments pertinents pour un système d'IA : les données, le système et les modèles commerciaux."

- *Groupe d'experts de haut niveau sur l'IA*[471]

4.1 Questions éthiques et juridiques

Cette exigence s'appuie sur trois grands principes différents : la traçabilité, l'explicabilité et la communication. [472]

Traçabilité

Les ensembles de données et les processus qui aboutissent à la décision du système d'IA, y compris ceux de la collecte et de l'étiquetage des données ainsi que les algorithmes utilisés, doivent être documentés selon la meilleure norme possible afin de permettre la traçabilité et d'accroître la transparence. Cela s'applique également aux décisions prises par le système d'IA. Cela permet d'identifier les raisons pour lesquelles une décision de l'IA était erronée, ce qui, à son tour, pourrait aider à prévenir de futures erreurs. La traçabilité facilite l'auditabilité ainsi que l'explicabilité.

Explicabilité

Il s'agit de la capacité à expliquer à la fois les processus techniques d'un système d'IA et les décisions humaines correspondantes (par exemple, les domaines d'application d'un système). L'explicabilité technique exige que les décisions prises par un système d'IA puissent être comprises et retracées par des êtres humains. En outre, il peut être nécessaire de faire des compromis entre l'amélioration de l'explicabilité d'un système (qui peut réduire sa précision) et l'augmentation de sa précision (au prix de l'explicabilité). Chaque fois qu'un système d'IA a un impact significatif sur la vie des gens, il devrait être possible d'exiger une explication appropriée du processus décisionnel du système d'IA. Cette explication devrait être fournie en temps utile et adaptée à l'expertise de la partie prenante concernée (par exemple, un profane, un régulateur ou un chercheur). En outre, des explications sur la mesure dans laquelle un système d'IA influence et façonne le processus de prise de décision de l'organisation, les choix de conception du système et la justification de son déploiement devraient être disponibles (assurant ainsi la transparence du modèle d'entreprise).

Communication

Les systèmes d'IA ne doivent pas se faire passer pour des humains auprès des utilisateurs ; les humains ont le droit d'être informés qu'ils interagissent avec un système d'IA. Cela implique que les systèmes d'IA doivent être identifiables en tant que tels. En outre, la possibilité de renoncer à cette interaction au profit d'une interaction humaine

---

[471] Groupe d'experts de haut niveau sur l'IA (2019) Lignes directrices en matière d'éthique pour une IA digne de confiance, p.18. Bruxelles, Commission européenne, Bruxelles. Disponible à l'adresse : https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (consulté le 15 mai 2020).

[472] Ibid.

doit être prévue lorsque cela est nécessaire pour garantir le respect des droits fondamentaux. En outre, les capacités et les limites du système d'IA doivent être communiquées aux praticiens de l'IA ou aux utilisateurs finaux d'une manière adaptée au cas d'utilisation concerné. Cela pourrait englober la communication du niveau de précision du système d'IA, ainsi que de ses limites.

4.2 Dispositions du RGPD : Transparence

4.2.1 Comprendre la transparence et l'opacité

En général, la transparence signifie que les personnes concernées reçoivent des informations claires sur le traitement des données (voir "Transparence" dans la sous-section "Licéité, loyauté et transparence" des "Principes" de la partie II des présentes lignes directrices). Elles doivent être informées de la manière dont leurs informations (y compris les données observées et déduites les concernant) sont utilisées et à quelles fins, que ces informations soient collectées auprès des personnes concernées elles-mêmes ou par des tiers.[473] Les personnes concernées doivent toujours savoir comment et pourquoi une décision prise à leur sujet par l'IA a été prise, ou si leurs données personnelles ont été utilisées pour former et tester un système d'IA. Les responsables du traitement doivent garder à l'esprit que dans de tels cas, la transparence est encore plus importante que lorsqu'ils n'ont pas de relation directe avec les personnes concernées.

D'une manière générale, la transparence doit être garantie en utilisant un certain nombre d'outils complémentaires. La désignation d'un DPD, qui sert alors de point de contact unique pour les questions des personnes concernées, est une excellente option. La préparation de registres adéquats du traitement à l'intention des autorités de contrôle ou la réalisation d'analyses d'impact sur la protection des données sont également des mesures hautement recommandées pour promouvoir la transparence. Enfin, la réalisation d'analyses visant à évaluer l'efficacité et l'accessibilité des informations fournies aux personnes concernées contribue à garantir la mise en œuvre efficace de ce principe.

Le principal défi de l'IA est qu'elle englobe un éventail de techniques très différentes les unes des autres. Certaines sont très simples, il est donc facile pour le responsable des données de fournir toutes les informations nécessaires. D'autres, comme l'apprentissage profond, ont de sérieux problèmes en termes de transparence. C'est ce qu'on appelle souvent le problème de la "boîte noire", qui introduit le problème de l'opacité dans le cadre de l'IA, une circonstance qui rend la transparence difficile à atteindre. En effet, l'opacité est l'une des principales menaces contre l'IA équitable, car elle va directement à l'encontre du besoin de transparence. Il existe au moins trois types d'opacité qui sont inhérents à l'IA dans une mesure plus ou moins grande : (1) le secret intentionnel des entreprises ou des États ; (2) l'analphabétisme technique ; et (3) l'opacité épistémique.

4.2.1.1 Opacité en tant que secret intentionnel d'entreprise ou d'État

Ce type d'opacité peut être légitime en vertu de la protection de la réglementation sur le secret industriel. Elle peut également répondre à des intérêts légitimes, tels que la

[473] Articles 13 et 14 du RGPD.

préservation des avantages concurrentiels, la préservation de la sécurité du système, ou la prévention de l'utilisation du système par des utilisateurs malveillants. Toutefois, elle doit être compatible avec l'incorporation de systèmes de certification indépendants capables d'accréditer que le mécanisme répond aux exigences du RGPD. Dans la plupart des cas, fournir à la personne concernée les informations dont elle a besoin pour protéger ses intérêts, sans divulguer en même temps des secrets commerciaux, ne devrait pas poser de problème.[474] Cela est dû au simple fait que les sujets n'ont pas besoin de comprendre en détail comment le système fonctionne, mais seulement comment il pourrait porter atteinte à leurs intérêts, droits et libertés.

4.2.1.2 L'opacité comme analphabétisme technique

Ce type d'opacité découle des compétences spécifiques requises pour concevoir et programmer les algorithmes, ainsi que de la capacité à lire et à écrire le code. Nous pourrions dire que les codes utilisés dans l'IA sont un mystère pour la grande majorité de la population, qui ne dispose pas de ces connaissances spécifiques, mais cela ne doit pas être un obstacle au respect de l'obligation d'information stipulée par le RGPD. La capacité à comprendre le langage informatique ne doit pas être un obstacle pour fournir une explication compréhensible de la finalité d'un système d'IA, non seulement aux parties prenantes qui font l'objet d'un profilage ou d'une décision automatisée, mais à toutes les autres personnes.

4.2.1.3 Opacité épistémique

Cette opacité découle des caractéristiques des algorithmes d'apprentissage automatique et de l'échelle requise pour les appliquer utilement. Elle est liée au fait que certains modèles algorithmiques ne sont pas interprétables par l'Homme. En clair, le transit entre les entrées que le modèle reçoit et les sorties qu'il émet est impénétrable en termes de compréhension humaine. Au niveau réglementaire, il n'y a pas d'interdiction d'utiliser ce type de modèle, bien qu'il soit conseillé de suivre le principe de précaution lors de son utilisation, car le manque d'interprétabilité pourrait aggraver les difficultés d'identification des biais du modèle, qui pourraient à leur tour conduire à des résultats discriminatoires, ou à des corrélations fausses ou fallacieuses. Bien entendu, tous les modèles d'apprentissage automatique ne sont pas opaques dans ce sens.

4.2.1.4 La préférence pour des outils transparents

En général, les responsables du traitement doivent toujours prévoir le développement d'algorithmes plus compréhensibles que ceux qui le sont moins. Les compromis entre l'**explicabilité, la transparence et les meilleures performances du système** doivent être équilibrés en fonction du contexte d'utilisation. Par exemple, dans le domaine des soins de santé, la précision et les performances du système peuvent être plus importantes que son explicabilité ; dans le domaine du maintien de l'ordre, l'explicabilité est beaucoup plus cruciale pour justifier le comportement et les résultats des forces de l'ordre. Dans d'autres domaines, comme le recrutement, la précision et

[474] Autorité norvégienne de protection des données (2018) Intelligence artificielle et vie privée. Autorité norvégienne de protection des données, Oslo. Disponible à l'adresse : https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf (consulté le 20 mai 2020).

l'explicabilité sont toutes deux appréciées de la même manière.[475] Si un service peut être offert à la fois par un algorithme facile à comprendre et un algorithme opaque - c'est-à-dire lorsqu'il n'y a pas de compromis entre l'explicabilité et la performance - le responsable du traitement devrait opter pour celui qui est le plus interprétable.

Si les responsables du traitement n'ont d'autre choix que d'utiliser un modèle opaque, ils devraient au moins essayer de trouver des solutions techniques au manque d'interprétabilité. Bien entendu, il est extrêmement difficile de mesurer précisément dans quelle mesure une extension de l'explicabilité est réalisée. Pour plus d'informations, voir la section "Droit de ne pas faire l'objet d'une prise de décision automatisée" dans la partie II, section "Droits de la personne concernée" des présentes lignes directrices. Si les responsables du traitement ont du mal à trouver des explications, ils doivent demander un avis extérieur. La possibilité de recourir à des audits indépendants peut à nouveau être une option raisonnable.

Informations complémentaires

CEPD (2015) Avis 7/2015. Relever les défis du big data : Un appel à la transparence, au contrôle par l'utilisateur, à la protection des données dès la conception et à la responsabilité. Contrôleur européen de la protection des données, Bruxelles. Disponible à l'adresse : https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf

Groupe d'experts de haut niveau sur l'intelligence artificielle (2019) Lignes directrices en matière d'éthique pour une IA digne de confiance. Commission européenne, Bruxelles. Disponible à l'adresse : https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

ICO (2020) Expliquer les décisions prises avec l'IA. Information Commissioner's Office, Wilmslow. Disponible à l'adresse : https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-artificial-intelligence-1-0.pdf

Autorité norvégienne de protection des données (2018) Intelligence artificielle et vie privée. Autorité norvégienne de protection des données, Oslo. Disponible à l'adresse : https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf

SHERPA (2019) Lignes directrices pour l'utilisation éthique de l'IA et des systèmes de big data. Projet Sherpa. Disponible à l'adresse : https://www.project-sherpa.eu/wp-content/uploads/2019/12/use-final.pdfwww.project-sherpa.eu/wp-content/uploads/2019/12/use-final.pdf

5 Loyauté, diversité et non-discrimination

[475] SHERPA (2019) Lignes directrices pour l'utilisation éthique de l'IA et des systèmes de big data. Projet SHERPA, p.26. Disponible à l'adresse : www.project-sherpa.eu/wp-content/uploads/2019/12/use-final.pdf (consulté le 15 mai 2020).

"Pour parvenir à une IA digne de confiance, nous devons permettre l'inclusion et la diversité tout au long du cycle de vie du système d'IA. Outre la prise en compte et la participation de toutes les parties prenantes concernées tout au long du processus, il s'agit également de garantir l'égalité d'accès par des processus de conception inclusifs ainsi que l'égalité de traitement. Cette exigence est étroitement liée au principe de loyauté."

- *Groupe d'experts de haut niveau sur l'IA*[476]

5.1 Principes éthiques

Éviter les biais injustes

Les ensembles de données utilisés par les systèmes d'IA, tant pour la formation que pour le fonctionnement, peuvent souffrir de l'inclusion de biais historiques involontaires, d'incomplétudes et de modèles de mauvaise gouvernance. Le maintien de ces biais pourrait entraîner des biais et des discriminations involontaires (in)directs à l'encontre de certains groupes ou personnes, ce qui pourrait exacerber les biais et la marginalisation. Le préjudice peut également résulter de l'exploitation intentionnelle des biais (des consommateurs) ou de l'exercice d'une concurrence déloyale, telle que l'homogénéisation des prix par le biais de la collusion ou d'un marché non transparent.

Les biais identifiables et discriminatoires doivent être supprimés dans la mesure du possible lors de la phase de collecte. La manière dont les systèmes d'IA sont développés (par exemple, la programmation des algorithmes) peut également souffrir de biais injustes. Cela peut être contré en mettant en place des processus de surveillance pour analyser et traiter l'objectif, les contraintes, les exigences et les décisions du système d'une manière claire et transparente. En outre, l'embauche de développeurs issus de divers milieux, cultures et disciplines peut garantir une diversité d'opinions - et doit donc être encouragée.

Accessibilité et conception universelle

Les systèmes doivent être centrés sur l'utilisateur et conçus de manière à permettre à tous d'utiliser les produits ou services d'IA, quels que soient leur âge, leur sexe, leurs capacités ou leurs caractéristiques. L'accessibilité à cette technologie pour les mineurs ou les personnes handicapées, qui sont présents dans tous les groupes sociétaux, revêt une importance particulière. Les systèmes d'IA ne devraient pas avoir une approche unique et devraient tenir compte des principes de conception universelle, en s'adressant au plus grand nombre d'utilisateurs possible, conformément aux normes d'accessibilité pertinentes. Cela permettra un accès équitable et une participation active de tous aux activités humaines existantes et émergentes assistées par ordinateur, ainsi qu'en ce qui concerne les technologies d'assistance.

---

[476] Groupe d'experts de haut niveau sur l'intelligence artificielle (2019) Lignes directrices éthiques pour une IA digne de confiance, p. 15 et suivantes. Commission européenne, Bruxelles. Disponible à l'adresse : https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (consulté le 15 mai 2020).

Participation des parties prenantes

Pour développer des systèmes d'IA dignes de confiance, il est conseillé de consulter les parties prenantes qui peuvent être directement ou indirectement affectées par le système tout au long de son cycle de vie. Il est bénéfique de solliciter un retour d'information régulier, même après le déploiement, et de mettre en place des mécanismes à plus long terme pour la participation des parties prenantes, par exemple en assurant l'information, la consultation et la participation des travailleurs tout au long du processus de mise en œuvre des systèmes d'IA dans les organisations.

5.2 Dispositions du RGPD

Selon le considérant 71 du RGPD, "le responsable du traitement devrait utiliser des procédures mathématiques ou statistiques appropriées pour le profilage, mettre en œuvre des mesures techniques et organisationnelles appropriées pour garantir, en particulier, que les facteurs qui entraînent des inexactitudes dans les données à caractère personnel sont corrigés et que le risque d'erreurs est réduit au minimum", sécuriser les données à caractère personnel d'une manière qui tienne compte des risques potentiels qu'elles comportent pour les intérêts et les droits de la personne concernée et qui prévienne, entre autres, les effets discriminatoires à l'égard des personnes physiques sur la base de l'origine raciale ou ethnique, des opinions politiques, de la religion ou des convictions, de l'appartenance syndicale, du statut génétique ou de santé ou de l'orientation sexuelle, ou qui aboutissent à des mesures ayant un tel effet."

Ce paragraphe est strictement lié à l'**article 21 de la Charte des droits fondamentaux de l'UE**, qui stipule que "[t]oute discrimination fondée notamment sur le sexe, la race, la couleur, les origines ethniques ou sociales, les caractéristiques génétiques, la langue, la religion ou les convictions, les opinions politiques ou toute autre opinion, l'appartenance à une minorité nationale, la fortune, la naissance, un handicap, l'âge ou l'orientation sexuelle est interdite". Parallèlement, l'EDPB fournit une définition de la loyauté dans ses lignes directrices sur la protection des données dès la conception et par défaut, qui stipulent que "[l]a loyauté est un principe général qui exige que les données à caractère personnel ne soient pas traitées d'une manière qui soit préjudiciable, discriminatoire, inattendue ou trompeuse pour la personne concernée". [477]

La discrimination est donc une violation dramatique du principe de loyauté. Or, dans le domaine de l'IA, les biais constituent une menace redoutable contre ce principe, car ils peuvent conduire à la stigmatisation ou à la discrimination potentielle d'individus isolés ou de communautés entières. [478]

5.2.1.1 Biais : les causes

Les biais peuvent être causés par un certain nombre de problèmes différents, et lorsque des données sont recueillies, elles peuvent contenir des **biais, des inexactitudes, des**

[477] EDPB (2019) Lignes directrices 4/2019 sur l'article 25 Protection des données par conception et par défaut (version pour consultation publique). Conseil européen de la protection des données, Bruxelles. Disponible à l'adresse : https://edpb.europa.eu/our-work-tools/public-consultations-art-704/2019/guidelines-42019-article-25-data-protection-design_es (consulté le 20 mai 2020).

[478] Mittelstadt, B. et L. Floridi, L. (2016) " The ethics of big data : current and foreseeable issues in biomedical context ", *Science and Engineering Ethics* 22(2) : 303-341.

erreurs et des fautes construits par la société. Les raisons qui expliquent ces biais sont multiples. Parfois, il peut arriver que les ensembles de données soient biaisés en raison d'**actions malveillantes.** L'introduction de données malveillantes dans un système d'IA peut modifier son comportement, en particulier dans le cas des systèmes d'auto-apprentissage.[479] Par exemple, dans le cas du chatbot Tay, développé par Microsoft, un grand nombre d'internautes ont commencé à poster des commentaires racistes et sexistes qui ont servi à alimenter l'algorithme. En conséquence, Tay a commencé à envoyer des tweets racistes et sexistes après seulement quelques heures de fonctionnement. Dans d'autres cas, les **données sont tout simplement de mauvaise qualité**, ce qui crée un biais. Par exemple, les données issues de la plateforme de médias sociaux présentent de sérieux risques pour les chercheurs, en raison des caractéristiques de l'environnement en ligne, qui ne garantit pas l'exactitude et la représentativité des données.

Le **déséquilibre des données de formation** (voir encadré 8) est une autre cause de biais, qui survient lorsque la proportion des différentes catégories dans les données de formation n'est pas équilibrée. Par exemple, dans le contexte des essais cliniques, il peut y avoir beaucoup plus de données provenant d'hommes que de femmes. Dans ce cas, les femmes risquent d'être discriminées par le modèle d'IA résultant. Par conséquent, les questions liées à la composition des bases de données utilisées pour la formation soulèvent des problèmes éthiques et juridiques cruciaux, et pas seulement des questions liées à l'efficacité ou de nature technique.

Encadré 8. Biais causés par une formation déséquilibrée des données

L'affaire Beauty.AI

Lancé en 2016, l'outil Beauty.AI a été créé pour sélectionner "la première reine ou roi de beauté jugé par des robots", en utilisant des algorithmes de reconnaissance de l'âge et du visage. Sept mille personnes ont envoyé leurs photos par le biais d'une application, mais la plupart des 44 gagnants étaient blancs ; seule une poignée était asiatique, et un seul avait la peau foncée. Et ce, malgré le fait que, même si la majorité des participants étaient blancs, de nombreuses personnes de couleur ont envoyé des photos, y compris des groupes importants d'Afrique et d'Inde. Ce résultat a immédiatement été considéré comme raciste, en raison d'une mauvaise sélection de l'ensemble de données d'entraînement. Le principal problème était que les données utilisées par le projet pour établir les normes de beauté étaient principalement composées de personnes blanches. Bien que les développeurs n'aient pas conçu l'algorithme pour que la peau claire soit considérée comme un signe de beauté, les données d'entrée ont effectivement conduit les juges robots à parvenir à cette conclusion.[480]

---

[479] Groupe d'experts de haut niveau sur l'IA (2019) Lignes directrices en matière d'éthique pour une IA digne de confiance. Commission européenne, Bruxelles, p.17. Disponible à l'adresse : https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (consulté le 20 mai 2020).

[480] LEVIN, S. (2016) 'UN CONCOURS DE BEAUTE A ETE JUGE PAR L'IA ET LES ROBOTS N'ONT PAS AIME LA PEAU FONCEE', *THE GUARDIAN*, 8 SEPTEMBRE. DISPONIBLE A L'ADRESSE : WWW.THEGUARDIAN.COM/TECHNOLOGY/2016/SEP/08/ARTIFICIAL-INTELLIGENCE-BEAUTY-CONTEST-DOESNT-LIKE-BLACK-PEOPLE (CONSULTE LE 26 MAI 2020).

L'outil de recrutement d'Amazon

En décembre 2018, Amazon a mis au rebut son outil de recrutement d'IA lorsque l'entreprise a découvert que le système d'IA présentait des biais contre les femmes. Amazon construisait des programmes informatiques depuis 2014 pour examiner les CV des candidats à un emploi, dans le but de mécaniser la recherche des meilleurs talents. L'outil utilisait l'IA pour noter les candidats à l'emploi d'une à cinq étoiles. En 2015, cependant, Amazon a découvert que l'outil ne notait pas les candidats aux postes de développeurs de logiciels et à d'autres postes techniques de manière non sexiste. En effet, les modèles informatiques d'Amazon ont été formés pour évaluer les candidats en observant les modèles de CV soumis à l'entreprise sur une période de 10 ans. La plupart provenaient d'hommes, ce qui reflète la domination masculine dans le secteur de la technologie.[481]

Troisièmement, les données de formation peuvent refléter une **discrimination passée produite par des tendances sociétales** (voir encadré 9). Si les responsables du traitement utilisent des données historiques, ils doivent être conscients des différences probables entre les contextes sociaux par rapport à l'époque actuelle. Sinon, les biais seront inévitables. Parfois, les biais proviennent des différents contextes sociaux de la communauté qui a fourni les données et de la communauté qui est censée utiliser l'algorithme. Si le responsable du traitement n'y prête pas une attention particulière, des biais seront probablement présents dans l'outil.

Encadré 10. Biais produits par les tendances sociétales

Dans le passé, les demandes de prêt des femmes étaient rejetées plus fréquemment que celles des hommes, en raison de biais. Dans ce cas, tout modèle d'IA formé sur des données historiques est susceptible de reproduire le même schéma de discrimination. Ces problèmes peuvent survenir même si les données de formation ne contiennent aucune caractéristique protégée, comme le sexe ou la race. Diverses caractéristiques des données d'apprentissage sont souvent étroitement corrélées aux caractéristiques protégées (par exemple, la profession, la race, etc.). Ces "variables de substitution" permettent au modèle de reproduire des schémas de discrimination associés à ces caractéristiques, même si ses concepteurs n'en avaient pas l'intention.

Ces problèmes peuvent se produire dans tout modèle statistique. Cependant, ils sont plus susceptibles de se produire dans les systèmes d'IA parce qu'ils peuvent inclure un plus grand nombre de caractéristiques, et peuvent identifier des combinaisons complexes de caractéristiques qui sont des substituts de caractéristiques protégées. De nombreuses méthodes modernes d'apprentissage automatique sont plus puissantes que les approches statistiques traditionnelles parce qu'elles sont plus aptes à découvrir des modèles non linéaires dans des données de grande dimension. Toutefois, celles-ci comprennent également des modèles qui reflètent la discrimination.[482]

[481] Dastin, J. (2018) " Amazon scraps secret AI recruiting tool that showed bias against women ", *Reuters*, 10 octobre. À l'adresse : www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

[482] ICO (2020) AI auditing framework : draft guidance for consultation, p.54. Bureau du commissaire à l'information, Wilmslow. Disponible à l'adresse : https://ico.org.uk/media/about-the-

Enfin, il est possible que les biais soient causés **par un outil d'IA mal conçu** (voir encadré 11). Il peut arriver que le concepteur introduise des corrélations par procuration qui ne fonctionnent pas bien avec la réalité. Si c'est le cas, le modèle fera des prédictions inexactes, car ses bases conceptuelles ne sont pas solides.

Encadré 11. Biais causé par un outil d'IA mal conçu : les algorithmes

Le système de santé américain utilise des algorithmes commerciaux pour guider les décisions en matière de santé. Obermeyer et al.[483] ont trouvé des preuves de biais racial dans un algorithme largement utilisé, ce qui signifie que, parmi les patients noirs et blancs auxquels l'algorithme a attribué le même niveau de risque, les patients noirs étaient plus malades que les blancs. Les auteurs ont estimé que ce biais racial réduisait de plus de la moitié le nombre de patients noirs identifiés pour des soins supplémentaires. Le biais s'est produit parce que l'algorithme a utilisé les coûts de santé comme un indicateur des besoins de santé. Moins d'argent a été dépensé pour les patients noirs ayant le même niveau de besoin que les patients blancs, et l'algorithme a donc faussement conclu que les patients noirs étaient en meilleure santé que les patients blancs tout aussi malades. En réalité, ces dépenses moindres étaient dues à un certain nombre de facteurs à caractère racial, tels qu'un accès différent aux traitements, des niveaux de confiance dans le système, des déséquilibres causés par les prestataires de soins, etc.

5.2.1.2 Résoudre les biais

Il existe différentes stratégies qui peuvent aider à éviter les biais ou à les corriger. Lors de la création des bases de données qui serviront à construire un modèle d'IA, les responsables du traitement **doivent s'efforcer d'éviter les données déséquilibrées ou erronées.** Les biais identifiables et discriminatoires doivent être supprimés dans la mesure du possible lors de la phase de création des bases de données. [484]

Si l'origine du biais est liée à l'ensemble de données de formation, le responsable du traitement doit rechercher une **sélection adéquate de données à utiliser dans la phase de formation,** afin d'éviter que les résultats du modèle ultérieur soient incorrects ou discriminatoires.[485] Un modèle d'IA doit "être formé à l'**aide de données pertinentes et correctes et il doit apprendre quelles sont les données à privilégier.** Le modèle ne doit pas mettre l'accent sur des informations relatives à l'origine raciale ou ethnique, aux opinions politiques, à la religion ou aux convictions, à l'appartenance syndicale, au

ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf (consulté le 26 mai 2020).

[483] Obermeyer, Z. et al. (2019) 'Dissecting racial bias in an algorithm used to manage the health of populations', *Science,* 25 octobre, 447-453.

[484] Considérant 71 du RGPD.

[485] Pour une définition de la discrimination directe et indirecte, voir, par exemple, l'article 2 de la directive 2000/78/CE du Conseil du 27 novembre 2000, qui porte création d'un cadre général en faveur de l'égalité de traitement en matière d'emploi et de travail. Voir également l'article 21 de la Charte des droits fondamentaux de l'UE.

statut génétique, à l'état de santé ou à l'orientation sexuelle si cela devait conduire à un traitement discriminatoire arbitraire (c'est nous qui soulignons). [486]

En outre, les personnes handicapées devraient être incluses dans le processus d'approvisionnement en données pour construire des modèles, et dans les tests, afin de créer un système plus inclusif et plus robuste. Si ce processus est réalisé de manière adéquate, le biais disparaîtra probablement. Par exemple, dans l'étude de cas sur le biais racial dans les algorithmes de santé (voir encadré 11), il a été possible de reformuler l'algorithme (dans ce cas, afin qu'il n'utilise plus les coûts comme indicateur des besoins) et d'éliminer le biais racial dans la prédiction des personnes nécessitant des soins supplémentaires. En effet, en changeant l'indicateur de santé, en passant des coûts prévus au nombre d'affections chroniques, le pourcentage de patients noirs bénéficiant de meilleurs soins est passé de 17 % à 46 %. Il s'agit d'un excellent exemple d'amélioration de la loyauté par la reformulation d'un algorithme.

Toutefois, les responsables du traitement doivent toujours garder à l'esprit que ce qui rend la lutte contre les biais particulièrement complexe, c'est que la sélection d'un ensemble de données implique de prendre des décisions et de faire des choix - ce qui peut, parfois, être fait presque **inconsciemment**. En revanche, le codage d'un algorithme traditionnel et déterministe est toujours une opération délibérée. En effet, les humains sont toujours l'intelligence qui se cache derrière un développement - même lorsqu'il est intégré à des algorithmes que nous pensons neutres. Quiconque construit un ensemble de données le construit, dans une certaine mesure, à son image, pour refléter sa propre vision du monde, ses valeurs ou, à tout le moins, les valeurs qui sont plus ou moins inhérentes aux données recueillies dans le passé. [487]

Dans cette optique, il est important que les équipes chargées de sélectionner les données à intégrer dans un ensemble de données soient composées de **personnes qui reflètent la diversité que le développement de l'IA est censé présenter.** À l'heure actuelle, il s'agit d'un défi majeur. En termes de genre, par exemple, les femmes ne représentent que 15 % du personnel de recherche en IA chez Facebook et 10 % chez Google, et il n'existe aucune donnée publique sur les travailleurs transgenres ou les autres minorités de genre. En termes de race, l'écart est encore plus marqué : seuls 2,5 % des effectifs de Google sont noirs, tandis que Facebook et Microsoft en comptent chacun 4 %.[488] Les responsables du traitement devraient tout mettre en œuvre pour que leurs **équipes reflètent mieux la diversité et mettre en place des données précises qui en témoignent.**

En résumé, les processus de développement des algorithmes **devraient toujours inclure un contrôle minutieux des biais possibles.** Les examens internes et externes

---

[486] Autorité norvégienne de protection des données (2018) Intelligence artificielle et vie privée. Autorité norvégienne de protection des données, Oslo, p.16. Disponible à l'adresse : https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf (consulté le 15 mai 2020).

[487] CNIL (2017) Comment l'humain peut-il garder la main ? Les questions éthiques soulevées par les algorithmes et l'intelligence artificielle. Commission nationale de l'informatique et des libertés, Paris, p.34. Disponible sur : www.cnil. fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf (consulté le 15 mai 2020).

[488] West, S.M., Whittaker, M. et Crawford, K. (2019) Discriminating systems : gender, race and power in AI. AI Now Institute, New York, p.3. Disponible à l'adresse : https://ainowinstitute.org/discriminatingsystems.html (consulté le 15 mai 2020).

devraient accorder une attention particulière à cette question. Les ensembles de données construits à des fins de validation doivent être soigneusement sélectionnés pour garantir une incorporation adéquate de données relatives à des sujets issus de différents secteurs de la société, en termes d'âge, de race, de sexe, de handicap, etc. Heureusement, il existe de nombreux outils techniques consacrés à l'éradication des biais dans les modèles d'IA.[489] La norme IEEE P7003TM pour la prise en compte des biais algorithmiques est particulièrement intéressante en ce moment. [490]

Cependant, aucun d'entre eux n'offre une solution magique, ou "solution miracle", applicable à tous les types d'algorithmes. Dans la plupart des cas, la bonne solution dépendra des multiples variables impliquées dans l'algorithme. Les responsables du traitement doivent s'efforcer d'éradiquer les biais autant que possible et être honnêtes quant aux résultats finaux de leurs efforts. Si des biais sont découverts, la solution d'IA doit être entraînée à nouveau. **S'il n'est pas possible d'éliminer les biais injustes du modèle, son déploiement ne doit pas avoir lieu.**

Liste de contrôle : partialité

 Le responsable du traitement a établi une stratégie ou un ensemble de procédures pour éviter de créer ou de renforcer un biais injuste dans le système d'IA, tant en ce qui concerne l'utilisation des données d'entrée que pour la conception des algorithmes.

 Le responsable du traitement évalue et reconnaît les éventuelles limitations découlant de la composition des ensembles de données utilisés.

 Le responsable du traitement a considéré la diversité et la représentativité des données utilisées.

 Le responsable du traitement a fait des tests pour des populations spécifiques ou des cas d'utilisation problématiques.

 Les responsables du traitement ont utilisé les outils techniques disponibles pour améliorer leur compréhension des données, du modèle et des performances.

 Le responsable du traitement a mis en place des processus pour tester et surveiller les biais potentiels pendant les phases de développement, de déploiement et d'utilisation du système d'IA.

 Le responsable du traitement a mis en place un mécanisme qui permet à d'autres personnes de signaler les problèmes liés aux biais, à la discrimination ou aux mauvaises performances du système d'IA.

 Le responsable du traitement a établi des étapes et des moyens de communication clairs sur la manière et la personne à qui ces questions peuvent être soulevées.

 Le responsable du traitement a pris en compte les autres personnes, potentiellement affectées indirectement par le système d'IA, en plus des utilisateurs (finaux).

[489] ICO (2020) AI auditing framework : draft guidance for consultation. Information Commissioner's Office, Wilmslow, p.55-56. Disponible sur : https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf (consulté le 15 mai 2020).

[490] Voir : https://ethicsinaction.ieee.org/ (consulté le 17 mai 2020).

 Le responsable du traitement a évalué s'il y a une variabilité possible de la décision qui peut se produire dans les mêmes conditions.

 En cas de variabilité, le responsable du traitement a mis en place un mécanisme de mesure ou d'évaluation de l'impact potentiel de cette variabilité sur les droits fondamentaux.

 Le responsable du traitement a mis en place une analyse quantitative ou des métriques pour mesurer et tester la définition appliquée de la loyauté.

 Le responsable du traitement a mis en place des mécanismes pour garantir la loyauté concernant les systèmes d'IA, et a envisagé d'autres mécanismes potentiels.

Informations complémentaires

CNIL (2017) Comment l'humain peut-il garder la main ? Les questions éthiques soulevées par les algorithmes et l'intelligence artificielle. Commission nationale de l'informatique et des libertés, Paris. Disponible sur : www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf

EDPB (2019) Lignes directrices 4/2019 sur l'article 25 Protection des données par conception et par défaut (version pour consultation publique). Conseil européen de la protection des données, Bruxelles. Disponible à l'adresse : https://edpb.europa.eu/our-work-tools/public-consultations-art-704/2019/guidelines-42019-article-25-data-protection-design_es

ICO (2020) AI auditing framework : draft guidance for consultation. Information Commissioner's Office, Wilmslow. Disponible à l'adresse : https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf

Mittelstadt, B. et Floridi, L. (2016) " The ethics of big data : current and foreseeable issues in biomedical context ", *Science and Engineering Ethics* 22(2) : 303-341.

Autorité norvégienne de protection des données (2018) Intelligence artificielle et vie privée. Autorité norvégienne de protection des données, Oslo. Disponible à l'adresse : https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf

West, S.M., Whittaker, M. et Crawford, K. (2019) Discriminating systems : gender, race and power in AI. AI Now Institute, New York, p.3. Disponible à l'adresse : https://ainowinstitute.org/discriminatingsystems.html.

6 Bien-être sociétal et environnemental

"Conformément aux principes de loyauté et de prévention des dommages, la société au sens large, les autres êtres sensibles et l'environnement devraient également être considérés comme des parties prenantes tout au long du cycle de vie du système d'IA. Il convient d'encourager la durabilité et la responsabilité écologique des systèmes d'IA et de favoriser la recherche de solutions d'IA dans des domaines d'intérêt mondial, tels que

les objectifs de développement durable. Idéalement, les systèmes d'IA devraient être utilisés au profit de tous les êtres humains, y compris les générations futures."

- *Groupe d'experts de haut niveau sur l'IA*[491]

6.1 Principes éthiques

Une IA durable et respectueuse de l'environnement

Les systèmes d'IA promettent d'aider à résoudre certaines de nos préoccupations sociétales les plus pressantes, mais cela doit être réalisé de la manière la plus respectueuse possible de l'environnement. Les processus de développement, de déploiement et d'utilisation du système, ainsi que l'ensemble de sa chaîne d'approvisionnement, doivent être évalués à cet égard. Cela comprend des mesures telles qu'un examen critique de l'utilisation des ressources et de la consommation d'énergie pendant la formation, et l'adoption de solutions moins dommageables pour l'environnement lorsqu'elles sont disponibles. Les mesures visant à garantir le respect de l'environnement dans l'ensemble de la chaîne d'approvisionnement des systèmes d'IA doivent également être encouragées.

Impact social

L'exposition omniprésente aux systèmes d'IA sociale dans tous les domaines de notre vie - qu'il s'agisse d'éducation, de travail, de soins ou de divertissement - peut modifier notre conception de l'action sociale ou avoir un impact sur nos relations sociales et notre attachement. Si les systèmes d'IA peuvent être utilisés pour améliorer les compétences sociales, ils peuvent également contribuer à leur détérioration. Cela pourrait également affecter le bien-être physique et mental des personnes. Les effets de ces systèmes doivent donc être soigneusement surveillés et examinés.

Société et démocratie

Au-delà de l'évaluation de l'impact du développement, du déploiement et de l'utilisation d'un système d'IA sur les individus, cet impact devrait également être évalué d'un point de vue sociétal, en tenant compte de son effet sur les institutions, la démocratie et la société en général. L'utilisation de systèmes d'IA devrait toujours faire l'objet d'une réflexion approfondie, en particulier dans les situations liées aux processus démocratiques, y compris non seulement la prise de décision politique mais aussi les contextes électoraux.

---

[491] Groupe d'experts de haut niveau sur l'intelligence artificielle (2019) Lignes directrices en matière d'éthique pour une IA digne de confiance, p.19. Commission européenne, Bruxelles. Disponible à l'adresse : https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (consulté le 15 mai 2020).

6.2 Dispositions du RGPD : légitimité

Le RGPD ne comprend pas de dispositions spécifiques liées au bien-être sociétal et environnemental. Toutefois, l'article 5, paragraphe 1, point b), stipule que "les données à caractère personnel sont collectées pour des finalités spécifiques, explicites et légitimes". Par cette clause, le RGPD introduit le concept de légitimité dans le contexte de la protection des données.

Cependant, la légitimité est un concept flou qui n'est pas du tout défini par le RGPD (voir "Principe de licéité, de loyauté et de transparence" dans la section "Principes" de la partie II des présentes lignes directrices). Le groupe de travail Article 29 indique que cela "signifie que les finalités doivent être "conformes à la loi" au sens le plus large. Cela inclut toutes les formes de droit écrit et de common law, la législation primaire et secondaire, les décrets municipaux, les précédents judiciaires, les principes constitutionnels, les droits fondamentaux, les autres principes juridiques, ainsi que la jurisprudence, telle que cette "loi" serait interprétée et prise en compte par les tribunaux compétents". Par conséquent, elle doit être comprise comme un concept très large qui englobe des considérations de bien-être social et environnemental.

Dans le "Livre blanc sur l'intelligence artificielle : une approche européenne de l'excellence et de la confiance", les auteurs notent que "compte tenu de l'importance croissante de l'IA, l'impact environnemental des systèmes d'IA doit être dûment pris en compte tout au long de leur cycle de vie et dans l'ensemble de la chaîne d'approvisionnement, par exemple en ce qui concerne l'utilisation des ressources pour la formation des algorithmes et le stockage des données". [492]

D'autres recommandations concrètes pour le développement de l'IA, orientées vers le bien-être sociétal et environnemental, figurent dans le "Rapport de la Commission au Parlement européen, au Conseil et au Comité économique et social européen : rapport sur les implications de l'intelligence artificielle, de l'internet des objets et de la robotique en matière de sécurité et de responsabilité".[493] Ce type de recommandations éthiques devrait être soigneusement examiné par les développeurs d'IA avant de traiter des données personnelles, car elles sont clairement liées à leur légitimité.

7 Responsabilité

"L'exigence de responsabilité complète les exigences ci-dessus, et est étroitement liée au principe de loyauté. Elle nécessite que des mécanismes soient mis en place pour garantir la responsabilité et l'obligation de rendre compte des systèmes d'IA et de leurs résultats, avant et après leur développement, leur déploiement et leur utilisation."

---

[492] Commission européenne (2020) Livre blanc sur l'intelligence artificielle : une approche européenne de l'excellence et de la confiance. Commission européenne, Bruxelles, p.3. Disponible sur : https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf (consulté le 26 mai 2020).

[493] Commission européenne (2020) Rapport de la Commission au Parlement européen, au Conseil et au Comité économique et social européen : rapport sur les implications en matière de sécurité et de responsabilité de l'intelligence artificielle, de l'internet des objets et de la robotique. Commission européenne, Bruxelles. Disponible à l'adresse : https://ec.europa.eu/info/sites/info/files/report-safety-liability-artificial-intelligence-feb2020_en_1.pdf (consulté le 26 mai 2020).

- *Groupe d'experts de haut niveau sur l'IA*[494]

7.1 Principes éthiques

Auditabilité

L'auditabilité signifie permettre l'évaluation des algorithmes, des données et des processus de conception. L'évaluation par des auditeurs internes et externes, et la disponibilité de ces rapports d'évaluation, peuvent contribuer à la fiabilité de la technologie. Dans les applications affectant les droits fondamentaux, y compris les applications critiques pour la sécurité, les systèmes d'IA devraient être ouverts à un audit indépendant. Cela n'implique toutefois pas nécessairement que les informations sur les modèles commerciaux et la propriété intellectuelle liés au système d'IA doivent toujours être ouvertement disponibles.

Minimisation et notification des impacts négatifs

Il est essentiel de garantir à la fois la capacité de rendre compte des actions ou des décisions qui contribuent à un certain résultat du système, et de réagir aux conséquences d'un tel résultat. Il est particulièrement crucial d'identifier, d'évaluer, de documenter et de minimiser les impacts négatifs potentiels des systèmes d'IA pour les personnes (in)directement concernées. Les dénonciateurs, les ONG, les syndicats ou d'autres entités doivent bénéficier d'une protection adéquate lorsqu'ils font part de préoccupations légitimes concernant un système d'IA. L'utilisation d'évaluations d'impact (par exemple, le red teaming ou des formes d'évaluation d'impact algorithmique), à la fois avant et pendant le développement, le déploiement et l'utilisation des systèmes d'IA, peut contribuer à minimiser les impacts négatifs. Ces évaluations doivent être proportionnelles au risque que posent les systèmes d'IA.

Compromis

Lors de la mise en œuvre des exigences susmentionnées, des tensions peuvent apparaître entre elles, ce qui peut conduire à des compromis inévitables. Ces compromis doivent être traités de manière rationnelle et méthodologique dans le cadre de l'état de l'art. Cela signifie que les intérêts et les valeurs concernés par le système d'IA doivent être identifiés et que, en cas de conflit, les compromis doivent être explicitement reconnus et évalués en termes de risque pour les principes éthiques, y compris les droits fondamentaux. Dans les situations où aucun compromis acceptable sur le plan éthique ne peut être identifié, le développement, le déploiement et l'utilisation du système d'IA ne devraient pas se poursuivre sous cette forme. Toute décision concernant le compromis à faire doit être expliquée et correctement documentée. Le décideur doit être responsable de la manière dont le compromis approprié est fait, et doit continuellement examiner la pertinence de la décision qui en résulte afin de s'assurer que les changements nécessaires peuvent être apportés au système, le cas échéant.

Réparation

---

[494] Groupe d'experts de haut niveau sur l'intelligence artificielle (2019) Lignes directrices en matière d'éthique pour une IA digne de confiance, p.19. Commission européenne, Bruxelles. Disponible à l'adresse : https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (consulté le 15 mai 2020).

Lorsqu'un impact négatif injuste se produit, des mécanismes accessibles doivent être mis en place pour garantir une réparation adéquate. Savoir qu'il est possible d'obtenir réparation lorsque les choses tournent mal est essentiel pour garantir la confiance. Une attention particulière doit être accordée aux personnes ou groupes vulnérables.

7.2 Dispositions du RGPD

7.2.1 Responsabilité

Conformément à l'article 5, paragraphe 2, du RGPD, le responsable du traitement est responsable du respect de tous les principes du RGPD mentionnés à l'article 5, paragraphe 1, et doit être en mesure de le démontrer. Cela inclut le principe de responsabilité (voir "Principe de responsabilité" dans la partie II, section "Principes" des présentes lignes directrices).

Le principe de responsabilité du RGPD est fondé sur le risque : plus le risque du traitement des données pour les droits et libertés fondamentaux des personnes concernées est élevé, plus les mesures nécessaires pour atténuer ces risques sont importantes.[495] Le principe de responsabilité repose sur plusieurs obligations de conformité pour les responsables du traitement des données, notamment : des obligations de transparence (articles 12-14) ; la garantie de l'exercice des droits en matière de protection des données (articles 15-22) ; la tenue de registres des opérations de traitement des données (article 30) ; la notification des éventuelles violations de données à une autorité de contrôle nationale (article 33) et aux personnes concernées (article 34) ; et, en cas de risque plus élevé, le recrutement d'un DPD et la réalisation d'une AIPD (article 35).

Étant donné que le traitement des données à caractère personnel dans les systèmes d'IA peut souvent être considéré comme à haut risque,[496] le développeur de l'IA devra souvent avoir un DPD et effectuer une AIPD. Les deux sections suivantes traitent de ces deux obligations spécifiques de responsabilité.

7.2.2 Évaluation des risques et AIPD

Une AIPD est un processus dans lequel le responsable du traitement des données, avant de lancer une procédure de traitement des données présentant un risque élevé pour les libertés et droits fondamentaux des personnes concernées, évalue l'impact des opérations de traitement envisagées sur la protection des données à caractère personnel (article 35, paragraphe 1).

Déterminer si le traitement des données présente un risque élevé n'est cependant pas une tâche facile. L'article 35, paragraphe 3, énumère trois cas : (1) une évaluation systématique et extensive d'aspects personnels concernant des personnes physiques, qui

---

[495] Voir les articles 24, 25 et 32 du RGPD, qui exigent que les responsables du traitement prennent en compte les "risques de probabilité et de gravité variables pour les droits et libertés des personnes physiques" lorsqu'ils adoptent des mesures spécifiques de protection des données.

[496] Voir, en particulier, l'article 35, paragraphe 3, point a), selon lequel le traitement des données est considéré comme présentant un risque élevé dans les cas, entre autres, "d'une évaluation systématique et extensive d'aspects personnels concernant des personnes physiques, fondée sur un traitement automatisé, y compris le profilage, et sur laquelle sont fondées des décisions produisant des effets juridiques à l'égard de la personne physique ou l'affectant de manière significative de façon similaire".

est fondée sur un traitement automatisé, y compris le profilage, et sur laquelle sont fondées des décisions produisant des effets juridiques concernant la personne physique ou l'affectant de manière significative de façon similaire ; (2) un traitement à grande échelle de catégories particulières de données visées à l'article 9, paragraphe 1, ou de données à caractère personnel relatives aux condamnations pénales et aux infractions visées à l'article 10 ; et (3) une surveillance systématique à grande échelle d'un domaine accessible au public.

En ce qui concerne les technologies innovantes, le groupe de travail "Article 29" a précisé certains exemples, tels que "l'utilisation combinée de la reconnaissance des empreintes digitales et du visage pour améliorer le contrôle d'accès physique" et "certaines applications de l'"Internet des objets"". Ces opérations de traitement des données sont considérées comme à haut risque "parce que l'utilisation de ces technologies peut impliquer de nouvelles formes de collecte et d'utilisation des données, éventuellement avec un risque élevé pour les droits et libertés des individus. En effet, les conséquences personnelles et sociales du déploiement d'une nouvelle technologie peuvent être inconnues."[497]

Si le traitement présente un risque élevé, il convient alors de réaliser une AIPD conformément à l'article 35, paragraphe 7, du RGPD. Le considérant 90 du RGPD précise en outre que l'évaluation du risque doit être effectuée à l'aide de deux paramètres : la **probabilité** et la **gravité** du risque élevé, compte tenu de la nature, de la portée, du contexte et des finalités du traitement, ainsi que des sources de risque. Plusieurs autorités nationales de contrôle ont publié des orientations sur la manière d'évaluer ces risques, comme l'Agencia Española de Protección de Datos Personales, l'Information Commissioner's Office, la Commission irlandaise de protection des données, la Commission nationale de l'informatique et des libertés, entre autres (voir "AIPD" dans la partie II, section "Principaux outils et actions" de la partie II des présentes lignes directrices).

Dans certaines situations, si le résultat de la AIPD est que l'activité de traitement envisagée présente un risque élevé de porter atteinte aux libertés et droits fondamentaux des personnes concernées, le responsable du traitement doit demander l'avis de l'autorité de contrôle nationale, comme le prescrit l'article 36 du RGPD. Certains États membres ont publié des listes qui contiennent des exemples d'activités de traitement des données qui déclencheraient cette consultation obligatoire ; parmi ces exemples, nous pouvons identifier des situations qui correspondent aux techniques d'IA et, dans certains cas, aller jusqu'à inclure expressément l'IA. Les autorités de contrôle peuvent exiger l'adoption de certaines mesures pour atténuer le risque, si possible, ou interdire l'utilisation de l'IA si cela n'est pas possible.

Liste de contrôle : une AIPD est-elle nécessaire ?

Le responsable du traitement a déterminé les juridictions où les activités de traitement des données auront lieu.

Le responsable du traitement a vérifié si ces juridictions ont adopté des listes indiquant les

[497] Groupe de travail Article 29 (2017) Lignes directrices relatives à l'analyse d'impact sur la protection des données, WP248, p. 10. Commission européenne, Bruxelles. Disponible à l'adresse : https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236 (consulté le 20 mai 2020).

traitements qui nécessitent une analyse de l'impact sur la protection des données et a vérifié si les activités de traitement des données prévues qui impliquent l'IA sont couvertes par ces dispositions.

Si les responsables du traitement ne sont pas sûrs de la nécessité d'effectuer une analyse de l'impact sur la protection des données, ils consultent le DPD ou, à défaut, le service juridique du responsable du traitement.

Le cas échéant, le responsable du traitement a procédé à une AIPD.

Si nécessaire, le responsable du traitement a déposé une consultation préalable auprès de l'autorité de contrôle compétente.

Si des modifications étaient suggérées, le responsable du traitement suivait l'avis de l'autorité de contrôle.

7.2.3 Diligence raisonnable du sous-traitant

Le principe de responsabilité (voir "Principe de responsabilité" dans la partie II, section "Principes" des présentes lignes directrices) est également présent lorsqu'un responsable du traitement choisit de faire appel aux services d'un sous-traitant. À cet égard, l'article 28, paragraphe 1, du RGPD[498] exige que les responsables du traitement effectuent certaines actions de diligence raisonnable, et ce avant de donner aux sous-traitants l'accès aux données à caractère personnel pour l'exécution d'activités de traitement des données. Comme pour les autres dispositions du RGPD, il n'est pas précisé quelles actions spécifiques un responsable du traitement doit mener lors de l'évaluation des sous-traitants. Le seul critère fourni par le RGPD est que les **responsables du traitement doivent juger les sous-traitants sur la base de leur capacité à démontrer qu'ils peuvent effectuer des activités de traitement en conformité avec le RGPD.**

Par conséquent, un chercheur qui développe une IA et qui doit faire appel à un tiers pour certaines activités de traitement devrait se poser deux questions : (1) quel type de comportement est attendu pour démontrer le respect de cette obligation ; et (2), si une certaine forme d'action positive est attendue, comment les responsables du traitement doivent-ils procéder pour effectuer cette diligence raisonnable ?

Pour la première question, le RGPD indique que si les responsables du traitement entendent rester en conformité avec le RGPD, ils ne peuvent retenir qu'un sous-traitant capable de démontrer sa conformité avec le RGPD. Par conséquent, les responsables du traitement doivent demander des informations pour l'évaluer. En d'autres termes, le RGPD attend des responsables du traitement qu'ils interrogent activement leur sous-traitant potentiel à ce sujet ; il ne suffit pas de s'appuyer sur une clause de déclaration et de garantie dans l'accord de traitement des données (voir "Principe d'intégrité et de confidentialité" dans la partie II, section "Principes" des présentes lignes directrices).

Quant à la manière dont les responsables du traitement doivent effectuer cette diligence raisonnable, là encore le RGPD ne fournit pas de points concrets à analyser. Néanmoins, certaines autorités de contrôle nationales ont proposé des sujets à prendre en compte, comme le fait de savoir si le sous-traitant suit les normes du secteur, de

[498] "Article 28 Sous-traitant 1. "Lorsque le traitement doit être effectué pour le compte d'un responsable du traitement, celui-ci ne fait appel qu'à des sous-traitants présentant des garanties suffisantes pour mettre en œuvre les mesures techniques et organisationnelles appropriées de telle sorte que le traitement réponde aux exigences du présent règlement et assure la protection des droits de la personne concernée."

demander la fourniture d'informations tant juridiques que techniques sur la manière dont le sous-traitant traite les données à caractère personnel, s'il adhère à un code de conduite ou s'il a suivi un programme de certification. [499]

Outre ces considérations générales, et en fonction de la manière dont le traitement demandé à ce tiers s'intègre dans le cadre de l'IA développée, d'autres questions doivent être posées. À cet égard, **toute question que les responsables du traitement se poseraient lors du développement de l'IA devrait être posée au sous-traitant.** Nous nous en remettons aux questions posées dans la liste de contrôle pour plus d'indications.

Liste de contrôle : diligence raisonnable du sous-traitant

Les responsables du traitement ont demandé des informations concernant le lieu où les activités de traitement des données auront lieu, et : (1) procéder à l'examen de la jurisprudence suggéré ci-dessous ; et (2) évaluer si les juridictions, dans le cas de pays non membres de l'UE, sont considérées comme adéquates par la Commission européenne.

Les responsables du traitement ont examiné la jurisprudence des autorités de contrôle nationales où le sous-traitant opère afin de vérifier les sanctions potentielles.

Les responsables du traitement ont exigé la preuve de l'adhésion à un code de conduite ou à une certification.

Les responsables du traitement ont exigé la preuve d'une certification ISO pertinente.

Les responsables du traitement ont demandé une copie des registres des activités de traitement.

Les responsables du traitement se sont enquis du processus de développement de l'IA, en particulier du type de données utilisées pour l'entraînement de l'IA et des données dont l'IA a besoin pour fonctionner et fournir un résultat utile.

7.2.4 DPD

Les DPD jouent un rôle crucial lors de la conception et de la mise en œuvre des activités de traitement des données dans le respect du RGPD. Ils constituent une autre garantie que le RGPD rend obligatoire à certaines occasions et, en général, il est recommandé de nommer une telle personnalité. Le groupe de travail Article 29 considère qu'il s'agit "d'une pierre angulaire de la responsabilité et que la nomination d'un DPD peut faciliter la conformité". [500]

---

[499] ICO (aucune date) Guide du Règlement général sur la protection des données (RGPD), Quelles sont les responsabilités et les obligations des contrôleurs lorsqu'ils font appel à un sous-traitant ? Information Commissioner's Office, Wilmslow. Disponible à l'adresse : https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/contracts-and-liabilities-between-controllers-and-processors-multi/responsibilities-and-liabilities-for-controllers-using-a-processor/ (consulté le 20 mai 2020).

[500] Groupe de travail Article 29 (2017) Lignes directrices sur les délégués à la protection des données ("DPD"), p.4. Commission européenne, Bruxelles.

L'article 37, paragraphe 1, du RGPD [501] indique dans quels cas les responsables du traitement et les sous-traitants doivent désigner un DPD. Dans le cas du développement de l'IA, et comme expliqué précédemment**, la désignation d'un DPD est (presque) certainement nécessaire, car de nombreux systèmes d'IA traitent des données à caractère personnel, ce qui les rendrait applicables aux conditions décrites à l'article 37, paragraphe 1, points a) et b), dans la plupart des situations.** Cette opinion est partagée, par exemple, par l'autorité de contrôle espagnole.[502] Cependant, ni le groupe de travail Article 29 ni l'EDPB n'ont spécifiquement déclaré qu'un DPD est obligatoire si un responsable du traitement ou un sous-traitant s'engage dans des activités de traitement de données qui impliquent l'IA. Néanmoins, le groupe de travail Article 29 a souligné que les **activités de profilage peuvent être considérées comme des activités qui déclenchent la nomination obligatoire d'un DPD**[503] si, comme indiqué ci-dessus, ces activités de profilage impliquent l'IA.

Il serait utile que la réglementation de chaque État membre relative à la nécessité de désigner un DPD élargisse la liste des activités qui exigent la désignation d'un DPD ou, au moins, fournisse des exemples clairs qui pourraient aider à interpréter quelles activités de traitement des données effectuées par les responsables du traitement et les sous-traitants exigent une telle désignation.

Si un DPD doit être désigné, pour l'une des raisons mentionnées ci-dessus, il est nécessaire qu'il participe à l'AIPD (requise par l'article 39, paragraphe 1, point c)) ainsi qu'à toute autre question liée à la protection des données au sein de l'entité (comme le prescrit l'article 39, paragraphe 1, point a)). Cela peut inclure l'examen d'un sous-traitant potentiel, comme décrit dans le point précédent. Par conséquent, les chercheurs impliqués dans le développement de l'IA devraient consulter le DPD concernant les questions de protection des données qui pourraient se poser au cours du développement de l'IA. Par exemple, le rôle du DPD, en ce qui concerne les systèmes d'IA, est également pertinent pour collaborer à la rédaction d'une notification appropriée, comme l'exigent les articles 13 et 14 correspondants, afin de communiquer correctement aux personnes concernées le fonctionnement de l'IA et les conséquences qu'elle pourrait avoir sur elles.

Liste de contrôle : DPD

---

[501] Article 37. Désignation du délégué à la protection des données. 1. Le responsable du traitement et le sous-traitant désignent un délégué à la protection des données dans tous les cas où : (a) le traitement est effectué par une autorité ou un organisme public, à l'exception des juridictions agissant dans l'exercice de leurs fonctions juridictionnelles ; b) les activités principales du responsable du traitement ou du sous-traitant consistent en des traitements qui, en raison de leur nature, de leur portée et/ou de leurs finalités, nécessitent un suivi régulier et systématique des personnes concernées à grande échelle ; ou c) les activités principales du responsable du traitement ou du sous-traitant consistent en des traitements à grande échelle de catégories particulières de données conformément à l'article 9 et de données à caractère personnel relatives aux condamnations pénales et aux infractions visées à l'article 10.

[502] Agencia Española de Protección de Datos Personales (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción, p.35. Agencia Española de Protección de Datos Personales, Madrid. Disponible sur : www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (consulté le 20 mai 2020).

[503] Groupe de travail Article 29 (2017) Lignes directrices sur les délégués à la protection des données ("DPD"), p.4. Commission européenne, Bruxelles.

Les responsables du traitement ont vérifié si l'institution a déjà nommé un DPD.

Si ce n'est pas le cas, ils ont vérifié auprès du service juridique si les activités de traitement des données envisagées nécessitent la désignation d'un DPD, soit en examinant les interprétations européennes faisant autorité, les réglementations locales, les interprétations locales faisant autorité, la jurisprudence - tant locale qu'européenne - et, enfin, les interprétations universitaires.

Les responsables du traitement ont exigé la nomination de DPD si nécessaire, et leur implication dans le processus de développement de l'IA si nécessaire.

En règle générale, le DPD doit être informé de chaque démarche entreprise afin de pouvoir intervenir s'il le juge utile.

Informations complémentaires

Agencia Española de Protección de Datos Personales (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción, p.35. Agencia Española de Protección de Datos Personales, Madrid. Disponible sur : www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf

Groupe de travail Article 29 (2010) Avis 3/2010 sur le principe de responsabilité. Commission européenne, Bruxelles. Disponible à l'adresse : https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2010/wp173_en.pdf

Groupe de travail Article 29 (2017) Lignes directrices sur l'analyse d'impact sur la protection des données (AIPD), pp. 9-10. Commission européenne, Bruxelles. Disponible à l'adresse : https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236

L'IA : un processus étape par étape

Iñigo de Miguel Beriain (UPV/EHU)

Remerciements : L'auteur tient à remercier Andres Chomsky, Oliver Feeney, Gianclaudio Malgieri, Aurélie Pols et Marko Sijan pour leurs conseils, leur contribution et leurs commentaires sur les versions préliminaires. Il va sans dire que toutes les erreurs sont de mon entière responsabilité.

Cette partie des lignes directrices a été revue et validée par Marko Sijan, conseiller principal spécialiste (DPA RH).

Introduction partie B

Cette deuxième partie des lignes directrices est construite sur la base d'un modèle étape par étape, le modèle CRISP-DM,[504] , qui est largement utilisé pour expliquer les étapes du développement des outils d'analyse de données et d'IA à forte intensité de données. En effet, il s'agit de l'outil sélectionné par le projet SHERPA pour élaborer ses lignes directrices pour le développement éthique de l'IA et des systèmes de Big Data.[505] Ces six étapes sont les suivantes : compréhension de l'activité ; compréhension des données ; préparation des données ; modélisation ; évaluation ; et déploiement. Il ne s'agit pas d'une classification fixe, car il arrive très souvent que les développeurs mélangent certaines de ces étapes. Par exemple, un algorithme entraîné peut être amélioré après l'étape de validation par un nouvel entraînement.

Néanmoins, il faut souligner que certaines des exigences éthiques et légales concernant le développement de l'IA doivent être évaluées tout au long du cycle de vie d'un développement d'IA de manière continue. Les responsables du traitement doivent surveiller la légitimité éthique du traitement, et ses effets inattendus. Ils doivent également évaluer l'impact collatéral possible d'un tel traitement dans un environnement social, au-delà des limites initialement conçues de la finalité, de la durée dans le temps et de l'extension.[506] Et cela doit être fait tout au long du cycle de vie d'un outil d'IA, conformément à l'article 25 du RGPD. Comme l'a déclaré le groupe de travail Article 29,

"Les responsables du traitement devraient procéder à des évaluations fréquentes des ensembles de données qu'ils traitent afin de vérifier l'absence de tout parti pris et de mettre au point des moyens de remédier à tout élément préjudiciable, y compris tout recours excessif aux corrélations. Les systèmes qui vérifient les algorithmes et les examens réguliers de l'exactitude et de la pertinence de la prise de décision automatisée, y compris le profilage, sont d'autres mesures utiles. Les responsables du traitement doivent mettre en place des procédures et des mesures appropriées pour prévenir les erreurs, les inexactitudes ou la discrimination sur la base de données de catégorie spéciale. Ces mesures doivent être utilisées de manière cyclique, non seulement au stade de la conception, mais aussi de manière continue, au fur et à mesure

---

[504] Shearer, C. (2000) 'The CRISP-DM model : the new blueprint for data mining', *Journal of Data Warehousing* 5(4) : 13-23. Disponible à l'adresse : https://mineracaodedados.files.wordpress.com/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf (consulté le 15 mai 2020).

[505] Projet SHERPA (2019) Lignes directrices pour le développement éthique des systèmes d'IA et de big data : une approche éthique par la conception. Projet SHERPA. Disponible à l'adresse : www.project-sherpa.eu/wp-content/uploads/2019/12/development-final.pdf (consulté le 15 mai 2020).

[506] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, p.7. Disponible sur : www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (consulté le 15 mai 2020).

que le profilage est appliqué aux personnes. Le résultat de ces tests devrait être pris en compte dans la conception du système."[507]

Une autre idée qui mérite réflexion est que l'IA est un label commun qui englobe une variété de technologies différentes. Il convient de faire une distinction fondamentale entre l'apprentissage automatique supervisé (des données d'entrée étiquetées par des humains sont transmises à un algorithme, qui définit ensuite les règles sur la base d'exemples qui sont des cas validés) et l'apprentissage non supervisé (des données d'entrée non étiquetées sont transmises à un algorithme, qui effectue sa propre classification et est libre de produire ses propres résultats lorsqu'on lui présente un modèle ou une variable). L'apprentissage supervisé exige que les superviseurs enseignent à la machine les résultats qu'elle doit produire, c'est-à-dire qu'ils doivent la "former". En principe, l'apprentissage supervisé est plus facile à comprendre et à contrôler.[508] De plus, étant donné que les ensembles de données utilisés dans les processus de formation sont sélectionnés par les formateurs, nous pourrions traiter certains des défis les plus inquiétants posés par ces technologies de manière tout à fait raisonnable. En revanche, l'IA non supervisée, et plus particulièrement les techniques telles que l'apprentissage profond, nécessite un suivi et un contrôle plus sophistiqués, car l'obscurité, les biais ou le profilage sont beaucoup plus difficiles à détecter, du moins à certaines étapes du cycle de vie du développement de l'IA.

Dans cette partie des lignes directrices, nous essayons de fournir un soutien à l'IA supervisée et non supervisée. Nous sommes conscients qu'il est presque impossible de fournir des conseils sur toutes les situations possibles. Cependant, nous espérons être en mesure de mettre en évidence les éléments fondamentaux et d'inclure des sources d'information supplémentaires utiles. Enfin, nous comprenons parfaitement que certains experts puissent considérer que certaines des recommandations que nous formulons pourraient être déplacées d'une étape à l'autre. En outre, certaines d'entre elles pourraient s'appliquer à plusieurs étapes différentes. Par conséquent, nous leur recommandons vivement d'adapter ces lignes directrices à leur convenance et selon leurs connaissances.

La structure du document est facile à suivre. Tout d'abord, nous introduisons une citation au chapitre de Colin Shearer,[509] suivie d'une description des tâches à accomplir à chaque étape concrète du processus, selon le même auteur. Ensuite, nous présentons quelques recommandations qui devraient être mises en œuvre à ce stade. Les références à d'autres chapitres des lignes directrices sont mises en évidence, tandis que les références à d'autres parties de ce chapitre font l'objet de renvois. Enfin, les annexes

---

[507] Groupe de travail Article 29 (2017) Lignes directrices sur la prise de décision individuelle automatisée et le profilage aux fins du règlement 2016/679. Adoptées le 3 octobre 2017, telles que révisées en dernier lieu et adoptées le 6 février 2018. Commission européenne, Bruxelles, p.28. Disponible à l'adresse : https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053 (consulté le 15 mai 2020).

[508] CNIL (2017) Comment l'humain peut-il garder la main ? Les questions éthiques soulevées par les algorithmes et l'intelligence artificielle. Commission nationale de l'informatique et des libertés, Paris, p.17. Disponible sur : www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf (consulté le 15 mai 2020).

[509] Shearer, C. (2000) 'The CRISP-DM model : the new blueprint for data mining', *Journal of Data Warehousing* 5(4) : 13-23. Disponible à l'adresse : https://mineracaodedados.files.wordpress.com/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf (consulté le 15 mai 2020).

comprennent des références à certains outils qui pourraient servir les objectifs de cette partie des lignes directrices. L'annexe I présente les recommandations pour l'audit des outils d'IA élaborées par l'Agence espagnole de protection des données. L'annexe II est plus spécifique, puisqu'elle fait référence à l'utilisation de l'IA dans le secteur des soins de santé. Cependant, elle constitue un excellent guide pour ceux qui souhaitent développer un outil d'IA dans ce secteur. À l'avenir, nous essaierons d'intégrer davantage d'annexes, dès qu'un mécanisme efficace pour le faire sera produit.

1 Compréhension de l'entreprise

"La phase initiale de compréhension de l'entreprise se concentre sur la compréhension des objectifs du projet d'un point de vue commercial, en convertissant cette connaissance en une définition du problème d'exploration de données, puis en développant un plan préliminaire conçu pour atteindre les objectifs. Afin de comprendre quelles données doivent être analysées plus tard, et comment, il est vital pour les praticiens de l'exploration de données de comprendre pleinement l'entreprise pour laquelle ils trouvent une solution. La phase de compréhension de l'entreprise comprend plusieurs étapes clés, notamment la détermination des objectifs de l'entreprise, l'évaluation de la situation, la détermination des objectifs de l'exploration de données et la production du plan de projet." [510]

1.1 Description

Cet objectif général implique quatre tâches principales :

1. Déterminez les objectifs de l'entreprise :
    a. Découvrez l'objectif principal de l'entreprise ainsi que les questions connexes auxquelles l'entreprise souhaite répondre.
    b. Déterminez la mesure du succès.
2. Évaluez la situation :
    a. Identifiez les ressources disponibles pour le projet, tant matérielles que personnelles.
    b. Identifiez les données disponibles pour atteindre l'objectif principal de l'entreprise.
    c. Dressez la liste des hypothèses formulées dans le cadre du projet.
    d. Dressez la liste des risques liés au projet, dressez la liste des solutions potentielles à ces risques, créez un glossaire de termes commerciaux et de termes relatifs à l'extraction de données, et réalisez une analyse coûts-avantages du projet.

[510] Shearer, C. (2000) 'The CRISP-DM model : the new blueprint for data mining', *Journal of Data Warehousing* 5(4) : 13-23, p. 14. disponible sur : https://mineracaodedados.files.wordpress.com/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf (consulté le 15 mai 2020).

3. Déterminer les objectifs de l'exploration des données :
    a. Décidez du niveau de précision prédictive attendu pour considérer le projet comme réussi.
4. Produire un plan de projet :
    a. Décrivez le plan prévu pour atteindre les objectifs de l'exploration de données, y compris la description des étapes spécifiques et un calendrier proposé, une évaluation des risques potentiels, et une évaluation initiale des outils et des techniques nécessaires pour soutenir le projet.

1.2 Principales mesures à prendre

1.2.1 Déterminer l'objectif de votre entreprise

Les développeurs d'IA devraient savoir dès le départ ce qu'ils attendent de l'outil. Plus ils sont imprécis sur ces attentes, plus il devient difficile de déterminer les finalités précises du traitement (voir "Conditions préalables à la licéité - finalités spécifiques et explicites" dans la sous-section "Licéité, loyauté et transparence" des "Principes" de la partie II des présentes lignes directrices). Si l'on garde à l'esprit que les responsables du traitement doivent rendre les finalités du traitement explicites, c'est-à-dire "révélées, expliquées ou exprimées d'une manière intelligible",[1] des attentes précises sont fortement recommandées. Il faut toutefois distinguer les différentes étapes du cycle de vie du développement d'une IA. Au stade de la formation, l'utilisation de grandes quantités de données peut être essentielle pour estimer l'utilité concrète de l'outil. Par conséquent, le traitement de grands ensembles de données peut être acceptable même si l'objectif spécifique (le développement de l'outil d'IA) n'est pas aussi précis. Bien entendu, cela ne serait pas aussi facilement acceptable si nous nous trouvions à la dernière étape du processus, c'est-à-dire le déploiement et l'utilisation de l'outil. Si, à ce moment-là, le responsable du traitement devait utiliser une grande quantité de données, une justification beaucoup plus détaillée serait nécessaire.

Dans tous les cas, il est nécessaire de souligner que certaines idées clés doivent être gardées à l'esprit dès le début. Par exemple, pour décider du niveau de précision prédictive attendu, afin de considérer le projet comme un succès, il est essentiel d'évaluer la quantité de données qui seront nécessaires pour développer l'outil d'IA ou la nature de ces données. Le niveau de prévisibilité ou de précision de l'algorithme, les critères de validation pour le tester, la quantité maximale ou la qualité minimale des données qui seront nécessaires pour l'utiliser dans le monde réel, etc. sont des caractéristiques fondamentales d'un développement d'IA. Ces décisions clés doivent être prises en compte dès la première étape du cycle de vie de la solution. Cela sera extrêmement utile pour mettre en œuvre une politique de protection des données dès la conception (voir la section "Protection des données dès la conception et par défaut" dans la partie II, section "Concepts principaux" des présentes lignes directrices).

Ainsi, le développeur d'IA doit fixer des seuils ou des fourchettes acceptables de faux positifs/négatifs, en fonction du cas d'utilisation, puis effectuer un bilan d'utilité. Le développeur d'IA doit être conscient que la détermination du niveau de précision attendu est clairement liée à la quantité de données nécessaires. Ce n'est pas la même

chose de développer, par exemple, un produit pour la santé ou pour la recommandation de séries télévisées. En outre, même dans le secteur de la santé, ce n'est pas la même chose de développer un outil capable d'effectuer un premier triage (c'est-à-dire de recommander l'intervention d'un médecin de premier recours ou d'un spécialiste) ou une solution visant à soutenir les radiologues dans leur diagnostic. En fonction de la finalité du mécanisme, des exigences de précision plus ou moins élevées seront adoptées.

S'il est possible d'atteindre un niveau de précision acceptable en utilisant beaucoup moins de données à caractère personnel que ne l'exige un niveau de précision plus élevé, il convient d'y réfléchir sérieusement. En outre, les développeurs d'IA doivent garder à l'esprit que toute augmentation marginale en termes de précision de la prédiction appelle parfois une augmentation significative de la quantité de données personnelles nécessaires.[2] Par conséquent, s'ils envisagent une modification fondamentale du niveau de précision de la prédiction requise, ils doivent examiner attentivement si cela s'accorde bien avec le principe de minimisation des données (voir "Principe de minimisation des données" dans la partie II, section "Principes" des présentes lignes directrices) .

1.2.2 Opter pour la solution technique

En général, les développeurs d'IA devraient toujours prévoir le développement d'algorithmes plus compréhensibles plutôt que d'algorithmes moins compréhensibles (voir la section **Dispositions du RGPD : Transparenc**). Les compromis entre l'explicabilité/la transparence et les meilleures performances du système doivent être équilibrés de manière appropriée en fonction du contexte d'utilisation. Par exemple, dans le domaine des soins de santé, la précision et les performances du système peuvent être plus importantes que son explicabilité, alors que, dans le domaine du maintien de l'ordre, l'explicabilité est beaucoup plus cruciale pour justifier les comportements et les résultats de l'application de la loi. Dans d'autres domaines, comme le recrutement, la précision et l'explicabilité sont toutes deux appréciées de la même manière.[511] Si un service peut être offert à la fois par un algorithme facile à comprendre et par un algorithme opaque, c'est-à-dire lorsqu'il n'y a pas de compromis entre l'explicabilité et la performance, le responsable du traitement doit opter pour celui qui est le plus interprétable (voir la section "Licéité, loyauté et transparence" dans les "Principes" de la partie II des présentes lignes directrices).

Encadré 13 : Interprétation de l'interprétabilité

Même si l'interprétabilité semble être recommandée, il faut reconnaître que ce n'est pas un concept clair. La littérature académique montre différentes motivations pour l'interprétabilité et, plus important encore, offre une myriade de notions sur les attributs qui rendent les modèles interprétables. Ce que recouvre le terme "interprétation" n'est toujours pas clair. À première vue, il semble raisonnable de supposer que les algorithmes simples et linéaires sont plus faciles à comprendre. Cependant, "pour certains types d'interprétation post-hoc, les réseaux neuronaux profonds présentent un avantage clair. Ils apprennent des représentations riches qui peuvent être visualisées,

[511] Projet SHERPA (2019) Lignes directrices pour le développement éthique des systèmes d'IA et de big data : une approche éthique par la conception. SHERPA, p. 26. Disponible à l'adresse : www.project-sherpa.eu/wp-content/uploads/2019/12/development-final.pdf (consulté le 15 mai 2020).

verbalisées ou utilisées pour le regroupement. Si l'on considère les critères d'interprétabilité, les modèles linéaires semblent avoir de meilleurs résultats pour l'étude du monde naturel, mais nous ne connaissons pas de raison théorique expliquant pourquoi il en est ainsi. Il est concevable que des interprétations post-hoc puissent s'avérer utiles dans des scénarios similaires." Il est donc difficile de formuler des recommandations spécifiques sur le type de modèles à privilégier en fonction de leur "interprétabilité".[512]

1.2.3 Mise en œuvre d'un programme de formation

Cette action est l'un des conseils les plus importants à prendre en compte dès le premier moment du développement d'une entreprise d'IA. Les concepteurs d'algorithmes (développeurs, programmeurs, codeurs, data scientists, ingénieurs), qui occupent le premier maillon de la chaîne algorithmique, sont susceptibles d'ignorer les implications éthiques et juridiques de leurs actions. En outre, l'un des principaux problèmes que soulève l'IA est qu'elle utilise généralement des données personnelles incluses dans de vastes ensembles de données. Cela brouille en quelque sorte la relation entre les données et la personne concernée, entraînant des violations de la réglementation qui se produisent rarement lorsque le responsable des données et le sujet ont une relation directe.[513] Cela pourrait avoir des conséquences en termes de respect adéquat des normes de protection des données. Il est primordial que ces travailleurs clés aient la plus grande conscience possible des implications éthiques et sociales de leur travail, et du fait même que celles-ci peuvent s'étendre à des choix de société,[514] même si l'alibi de l'"ingénierie dévoyée" peut difficilement fonctionner après l'affaire Google Street View.[515]

Afin d'éviter que la mauvaise représentation des questions éthiques et juridiques ne provoque des conséquences indésirables, deux grandes lignes d'action peuvent être adoptées. Tout d'abord, les développeurs peuvent essayer de faire en sorte que les concepteurs d'algorithmes soient en mesure de comprendre les implications de leurs actions, tant pour les individus que pour la société, et qu'ils soient conscients de leurs responsabilités en apprenant à faire preuve d'une attention et d'une vigilance constantes.[516] En ce sens, une formation optimale de tous les sujets impliqués dans le projet (développeurs, programmeurs, codeurs, data scientists, ingénieurs, chercheurs) avant même qu'il ne commence pourrait être l'un des outils les plus efficaces pour économiser du temps et des ressources en termes de conformité avec la réglementation sur la protection des données. Ainsi, la mise en œuvre de programmes de formation de

---

[512] Lipton, Z. C. (2017) 'The mythos of model interpretability', 2016 ICML workshop on human interpretability in machine learning (WHI 2016), New York, NY. Disponible à l'adresse : https://arxiv.org/pdf/1606.03490.pdf (consulté le 15 mai 2020).

[513] Kuyumdzhieva, A. (2018) 'Ethical challenges in the digital era : focus on medical research', pp. 45-62 in : Koporc, Z. (ed.) *Ethics and integrity in health and life sciences research*. Emerald, Bingley.

[514] CNIL (2017) Comment l'humain peut-il garder la main ? Les questions éthiques soulevées par les algorithmes et l'intelligence artificielle. Commission nationale de l'informatique et des libertés, Paris, p.55. Disponible sur : www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf (consulté le 15 mai 2020).

[515] Voir, par exemple : https://www.slashgear.com/googles-rogue-engineer-street-view-excuse-blown-apart-30225200/

[516] Ibid. p. 55.

base qui comprennent au moins les principes fondamentaux de la Charte des droits fondamentaux, les principes exposés à l'article 5 du RGPD, la nécessité d'une base légale pour le traitement (y compris les contrats entre les parties), etc.

Cependant, il peut être difficile de former des personnes qui n'ont jamais été en contact avec les questions de protection des données. Une autre solution consiste à impliquer un expert de la protection des données et des questions éthiques et juridiques dans l'équipe de développement, de manière à créer une équipe interdisciplinaire. Pour ce faire, on peut engager un expert à cette fin (un travailleur interne ou un consultant externe) pour concevoir la stratégie et les décisions ultérieures sur les données personnelles requises par le développement des outils, avec la participation étroite du délégué à la protection des données.

Il est également fortement recommandé d'adopter des mesures adéquates pour garantir la confidentialité, l'intégrité et la disponibilité des données (voir "Mesures en faveur de la confidentialité" dans la sous-section "Intégrité et confidentialité" des "Principes" de la partie II des présentes lignes directrices).

1.2.4 **Conception d'outils de traitement légitime des données** Selon l'article 5, paragraphe 1, point a), du RGPD, les données à caractère personnel sont "collectées pour des finalités spécifiques, explicites et légitimes et ne sont pas traitées ultérieurement de manière incompatible avec ces finalités". Le concept de légitimité n'est pas bien défini dans le RGPD, mais le groupe de travail Article 29 a déclaré que la légitimité implique que les données doivent être traitées "conformément à la loi", et que la "loi" doit être comprise comme un concept large qui inclut "toutes les formes de droit écrit et de common law, la législation primaire et secondaire, les décrets municipaux, les précédents judiciaires, les principes constitutionnels, les droits fondamentaux, les autres principes juridiques, ainsi que la jurisprudence, telle que cette "loi" serait interprétée et prise en compte par le tribunal compétent". [517]

Il s'agit donc d'un concept plus large que la licéité. Il implique le respect des principales valeurs de la réglementation applicable et des grands principes éthiques en jeu. Par exemple, certains développements concrets de l'IA nécessiteront l'intervention d'un comité d'éthique. Dans d'autres cas, des lignes directrices ou tout autre type de réglementation non contraignante peuvent être applicables. Les développeurs d'IA doivent s'assurer de la conformité à cette exigence en élaborant un plan pour cette étape préliminaire du cycle de vie de l'outil (voir "Légitimité et licéité" dans la sous-section "Licéité, loyauté et transparence" des "Principes" de la partie II des présentes lignes directrices).

1.2.6 **Adopter une approche de réflexion fondée sur le risque**

Les responsables du traitement doivent minimiser les risques pour les droits, intérêts et libertés des personnes concernées. À cette fin, ils doivent travailler selon une approche fondée sur le risque (voir "Principe d'intégrité et de confidentialité" dans la partie II, section "Principes").

---

[517] Groupe de travail Article 29 (2013) Avis 03/2013 sur la limitation de la finalité Adopté le 2 avril 2013, WP203. Commission européenne, Bruxelles, p.20. Disponible sur : https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf (consulté le 15 mai 2020).

L'approche fondée sur le risque de la législation sur la protection des données exige que les responsables du traitement se conforment à leurs obligations et mettent en œuvre des mesures appropriées dans le cadre de leur situation particulière - la nature, la portée, le contexte et les finalités du traitement qu'ils ont l'intention d'effectuer, et les risques que cela représente pour les droits et libertés des personnes. Leurs considérations de conformité impliquent donc d'évaluer les risques pour les droits et libertés des personnes et de juger de ce qui est approprié dans ces circonstances. Dans tous les cas, les responsables du traitement doivent s'assurer qu'ils respectent les exigences en matière de protection des données (voir "Principe de responsabilité" dans la partie II, section "Principes").

Une réflexion fondée sur le risque en ce qui concerne la confidentialité des données, ou une approche fondée sur le risque en ce qui concerne les questions relatives aux dommages qui peuvent être causés aux personnes, doit être intégrée dès les premières étapes du processus. Il pourrait être trop tard si elle n'est envisagée que plus tard. Pour gérer les risques pour les personnes qui découlent du traitement des données personnelles dans les outils d'IA, il est important que les responsables du traitement développent une compréhension et une articulation matures des droits fondamentaux, des risques et de la manière d'équilibrer ces intérêts et d'autres. En définitive, il est nécessaire que les responsables du traitement évaluent les risques que l'utilisation de l'IA fait peser sur les droits des personnes, qu'ils déterminent la manière dont ils doivent y faire face et qu'ils établissent l'impact que cela a sur leur utilisation de l'IA.[518] À cette fin, deux facteurs clés doivent être pris en considération : [519]

- Risques découlant du traitement lui-même, tels que l'apparition de biais associés au profilage ou aux systèmes de prise de décision automatisés (cf. **Dispositions du**).
- Les risques découlant du traitement par rapport au contexte social et les effets secondaires indirectement liés à l'objet du traitement qui peuvent survenir.

[518] ICO (2020) AI auditing framework - draft guidance for consultation. Information Commissioner's Office, Wilmslow, p.13-14. Disponible à l'adresse : https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf (consulté le 15 mai 2020).

[519] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, p.30. Disponible sur : www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (consulté le 15 mai 2020).

Encadré 14 : L'importance des fournisseurs de logiciels en termes de sécurité

En août 2015, une société de logiciels médicaux de l'Indiana a signalé au gouvernement fédéral que ses réseaux avaient été piratés et que les informations privées de 3,9 millions de personnes avaient été exposées. Cela comprenait des données personnelles telles que des noms, des adresses, des dates de naissance, des numéros de sécurité sociale et des dossiers médicaux. Selon IBM X-Force Research, il s'agit de l'une des plus importantes violations de données de santé de ces dernières années. Selon l'entreprise, l'attaque a été détectée dix-neuf jours après que les auteurs ont obtenu un accès non autorisé à son réseau. Les clients n'ont été avertis que près d'un mois après le début de l'attaque.[520]

En outre, les responsables du traitement doivent veiller à ce que des mesures techniques et organisationnelles appropriées soient mises en œuvre pour éliminer, ou au moins atténuer, le risque de sécurité, en réduisant la probabilité que les menaces identifiées se concrétisent, ou en réduisant leur impact. La description générale des mesures de sécurité techniques et organisationnelles doit faire partie des registres du traitement, si possible (article 30, paragraphe 1, point g), pour les responsables du traitement, et article 30, paragraphe 2, point d), pour les sous-traitants) et toutes les mesures mises en œuvre font partie de la AIPD, en tant que mesures correctives pour limiter le risque. Enfin, une fois les mesures sélectionnées mises en œuvre, le risque résiduel restant doit être évalué et maintenu sous contrôle. L'analyse des risques et l'AIPD sont les outils qui s'appliquent.

Une AIPD est très souvent obligatoire dans le cas du développement de l'IA (voir "Dans quels cas dois-je réaliser un AIPD" dans la sous-section "Analyse de l'impact sur la protection des données" des "Principaux outils et actions", partie II). Cela dépend si les risques associés au traitement sont élevés ou non, conformément à l'article 35, paragraphe 3, du RGPD. Cependant, elle est fortement recommandée car elle soutient la responsabilisation. En cas de doute, la consultation de l'autorité de contrôle compétente avant le traitement est fortement recommandée. Enfin, n'oubliez pas que lors de l'utilisation du big data et de l'IA, il est difficile de prévoir quels seront les risques futurs, de sorte que faire une évaluation des implications éthiques ne sera pas suffisant pour traiter tous les risques possibles. Il est donc important d'envisager une réévaluation des risques et il est également fortement recommandé d'intégrer une méthode plus dynamique d'évaluation des risques liés à la recherche. N'hésitez pas à effectuer des AIPD supplémentaires à d'autres étapes du processus si nécessaire.

1.2.7 Préparer la documentation du traitement

Quiconque traite des données à caractère personnel (y compris les responsables du traitement[521] et les sous-traitants[522]) doit documenter ses activités, principalement à

[520] IBM X-Force® Research (2017) Tendances en matière de sécurité dans le secteur de la santé : Le vol de données et les ransomwares tourmentent les organisations de santé. IBM Security, Somers, NY, p.7. Disponible à l'adresse : www.ibm.com/downloads/cas/PLWZ76MM (consulté le 17 mai 2020).

[521] Voir article 30, paragraphe 1, du RGPD.

[522] Voir article 30, paragraphe 2, du RGPD.

l'intention des autorités de contrôle qualifiées/pertinentes.[523] Cela doit se faire au moyen de registres du traitement qui sont conservés de manière centralisée par l'organisation pour l'ensemble de ses activités de traitement, et de documents supplémentaires qui se rapportent à une activité individuelle de traitement des données (voir "Documentation du traitement" dans la section "Principaux outils et actions" de la partie II des présentes lignes directrices). Cette étape préliminaire est le moment idéal pour mettre en place une méthode systématique de collecte de la documentation nécessaire, puisque c'est à ce moment-là que l'organisation conçoit et planifie l'activité de traitement[524] .

En effet, les responsables du traitement devraient créer une politique de protection des données qui permette la traçabilité des informations (s'il existe des codes de conduite approuvés, ceux-ci devraient être mis en œuvre ; voir la sous-section "Économie d'échelle pour la conformité et sa démonstration" dans la section "Responsabilité" des "Principes" de la partie II des présentes lignes directrices). Cette politique devrait également préciser les responsabilités attribuées aux sous-traitants, et inclure dans l'accord de traitement les tâches qui lui seront déléguées en ce qui concerne l'exécution des droits des personnes concernées. Les développeurs d'IA doivent toujours se rappeler que l'article 32, paragraphe 4, du RGPD précise qu'un élément important de la sécurité consiste à s'assurer que les employés n'agissent que sur instruction et selon les instructions du responsable du traitement (voir "Intégrité et confidentialité" dans la partie II, section "Principes").

Les responsables du traitement doivent toujours garder à l'esprit que le développement d'outils d'IA implique souvent l'utilisation de différents ensembles de données. La traçabilité du traitement, les informations sur l'éventuelle réutilisation des données et l'utilisation de données appartenant à différents ensembles de données dans différentes ou dans les mêmes étapes du cycle de vie doivent être assurées par les registres.

1.2.8 Documentation du traitement

Comme indiqué dans les exigences et les tests d'acceptation pour l'achat et/ou le développement des logiciels, du matériel et de l'infrastructure employés (voir la sous-section de la section "Documentation du traitement"), l'évaluation des risques et les décisions prises "doivent être documentées afin de respecter l'exigence de protection des données dès la conception" (de l'article 25 du RGPD).

Enfin, les responsables du traitement doivent toujours être conscients que, conformément à l'article 32, paragraphe 1, point d), du RGPD, la protection des données est un processus. Par conséquent, **ils doivent tester, apprécier et évaluer régulièrement l'efficacité des mesures techniques et organisationnelles.** C'est à ce moment-là qu'il convient de créer des procédures permettant aux responsables du traitement d'identifier les changements qui déclencheraient le réexamen de l'évaluation des risques avant traitement. Dans la mesure du possible, les responsables du traitement doivent essayer d'imposer un modèle dynamique de suivi des mesures en jeu (voir "Intégrité et confidentialité" dans la partie II des présentes lignes directrices, section "Principes").

---

[523] Voir les articles 58(1)(a), 30(4) et 5(2) du RGPD.

[524] L'article 25, paragraphe 1, du RGPD appelle cela "le moment de la détermination des moyens de traitement".

Encadré 15 : L'extrême difficulté de la responsabilisation dans le développement de l'IA

Même si la responsabilisation est un objectif nécessaire et que l'attribution de responsabilités à un sous-traitant spécifique est absolument nécessaire, les responsables du traitement doivent toujours être conscients que le fonctionnement de l'IA peut rendre extrêmement difficile la surveillance d'un système. Comme l'a déclaré la CNIL, "la question de savoir où la responsabilité et la prise de décision peuvent être mises en place doit être abordée d'une manière légèrement différente lorsqu'il s'agit de systèmes d'apprentissage automatique". Ainsi, les responsables de traitement devraient plutôt penser en termes de chaîne de responsabilité, du concepteur du système jusqu'à son utilisateur, en passant par la personne qui va alimenter ce système en données d'apprentissage. Ce dernier fonctionnera différemment en fonction de ces données d'entrée.

À ce sujet, on peut citer le chatbot Tay de Microsoft. Il a été fermé vingt-quatre heures à peine après son lancement lorsque, s'inspirant des messages des utilisateurs des médias sociaux, il a commencé à diffuser ses propres commentaires racistes et sexistes. Inutile de dire que déterminer la part exacte de responsabilité entre ces différents maillons de la chaîne pourrait être une tâche laborieuse.[525]

1.2.9 Vérification du cadre réglementaire

Le RGPD comprend un cadre réglementaire spécifique concernant le traitement à des fins de recherche scientifique (voir " Protection des données et recherche scientifique " dans la partie II, section "Concepts principaux").[526] Si le développement de l'IA peut être considéré comme de la recherche scientifique, le "droit de l'Union ou des États membres peut prévoir des dérogations aux droits visés aux articles 15, 16, 18 et 21, sous réserve des conditions et garanties visées au paragraphe 1 du présent article, dans la mesure où ces droits sont susceptibles de rendre impossible ou de nuire gravement à la réalisation des finalités spécifiques, et où ces dérogations sont nécessaires à la réalisation de ces finalités" (article 89, paragraphe 2). En outre, selon l'article 5, point b), "un traitement ultérieur des données collectées, conformément à l'article 89, paragraphe 1, ne serait pas considéré comme incompatible avec les finalités initiales ("limitation de la finalité")". Il existe d'autres exceptions particulières au cadre général applicable au traitement à des fins de recherche (comme la limitation du stockage) qui doivent également être prises en compte. Néanmoins, les développeurs d'IA doivent être conscients du cadre réglementaire concret qui s'applique à leurs recherches. Celui-ci peut comporter des changements importants en fonction de leur réglementation nationale. La consultation de leurs DPD est fortement recommandée à cet effet.

---

[525] CNIL (2017) Comment l'humain peut-il garder la main ? Les questions éthiques soulevées par les algorithmes et l'intelligence artificielle. Commission nationale de l'informatique et des libertés, Paris, p.29. Disponible sur : www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf (consulté le 15 mai 2020).

[526] Ce cadre spécifique comprend également des objectifs de recherche historique ou des objectifs statistiques. Toutefois, les recherches sur les TIC ne sont généralement pas liées à ces objectifs. Par conséquent, nous ne les analyserons pas.

1.2.10 Définition des politiques de stockage des données

Selon l'article 5, paragraphe 1, point e), du RGPD, les données à caractère personnel doivent être "conservées sous une forme permettant l'identification des personnes concernées pendant une durée n'excédant pas celle nécessaire à la réalisation des finalités pour lesquelles elles sont traitées". Cette exigence est double. D'une part, elle concerne l'identification : les données doivent être conservées sous une forme permettant l'identification des personnes concernées pendant une durée n'excédant pas celle nécessaire. Par conséquent, les développeurs d'IA devraient mettre en œuvre des politiques visant à éviter l'identification dès qu'elle n'est pas nécessaire au traitement. Cela implique l'adoption de mesures adéquates pour garantir qu'à tout moment, seul **le degré minimal d'identification nécessaire à la réalisation des finalités doit être utilisé** (voir "Limitation du stockage" dans la partie II, section "Principes").

D'autre part, la conservation des données implique que les données ne peuvent être stockées que pour une **période limitée** : le temps strictement nécessaire aux finalités pour lesquelles les données sont traitées. Toutefois, le RGPD autorise la "conservation pour des périodes plus longues" si la seule finalité est la recherche scientifique (ou l'archivage dans l'intérêt public, la recherche historique ou les fins statistiques) (voir "Protection des données et recherche scientifique" dans la partie II, section "Concepts principaux").

Dans le cas du développement de l'IA, cette exception soulève le risque que les développeurs décident de conserver les données plus longtemps que ce qui est strictement nécessaire, afin de s'assurer qu'elles sont disponibles pour des raisons autres que les finalités initiales pour lesquels elles ont été collectées. Les responsables du traitement doivent savoir que, même si le RGPD autorise la conservation des données pendant des périodes plus longues, **ils doivent avoir une bonne raison d'opter pour une telle prolongation (**voir "Aspect temporel" dans la sous-section "Limitation du stockage" des "Principes" de la partie II). Pour autant qu'un risque réel découle du non-respect du principe de limitation de la finalité, **le test de compatibilité devrait faire partie de toute réutilisation potentielle des données**.

L'intention du législateur semble avoir été de dissuader le stockage illimité, même dans le cadre de ce régime spécial, et de veiller à ce que la recherche scientifique ne serve pas de prétexte à un stockage prolongé à d'autres fins, privées. En cas de doute, le responsable du traitement doit examiner si une nouvelle base juridique est appropriée. Ainsi, les durées de conservation doivent être proportionnées aux finalités du traitement : "Pour définir les périodes de stockage (délais), il convient de prendre en compte des critères tels que la durée et la finalité de la recherche. Il convient de noter que les dispositions nationales peuvent également prévoir des règles concernant la période de stockage." [527]

[527] CEPD (2020) Lignes directrices 03/2020 sur le traitement des données relatives à la santé à des fins de recherche scientifique dans le cadre de l'épidémie de COVID-19 Adoptées le 21 avril 2020. Contrôleur européen de la protection des données, Bruxelles, p.10. Disponible sur https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf (consulté le 23 avril 2020).

Ainsi, si les responsables du traitement n'ont pas besoin des données, et qu'il n'y a pas de raisons légales obligatoires qui les obligent à conserver les données, il vaut mieux les anonymiser ou les supprimer. Les chercheurs devraient consulter leur DPD s'ils souhaitent conserver des données pendant une longue période et s'informer de la réglementation nationale applicable. Ce pourrait également être un excellent moment pour **envisager des délais d'effacement des différentes catégories de données et documenter ces décisions** (voir la section "Responsabilité" des "Principes" de la partie II des présentes lignes directrices.

1.2.11 Nomination d'un délégué à la protection des données

La nomination d'un DPD est l'une des meilleures mesures que peut prendre le responsable du traitement pour mettre en œuvre correctement les mesures qui garantissent le respect des droits des personnes concernées. La désignation d'un DPD n'est pas une conséquence nécessaire de l'utilisation d'outils d'IA. Il est toutefois indéniable que la désignation d'un DPD n'est obligatoire que si les conditions établies par l'article 37, paragraphe 1, points b) ou c), s'appliquent. Par conséquent, il n'est pas nécessaire que tous les développeurs d'IA désignent un DPD. Cependant, il **est toujours recommandé de le faire, au moins en termes de transparence** (voir la section "Transparence" des "Principes" de la partie II)**.**

En tout état de cause, les responsables des données devraient développer ce point en décrivant le rôle du DPD par rapport à la gestion globale du projet, en veillant à ce que le rôle du DPD ne soit pas marginal, mais qu'il soit cimenté dans les processus décisionnels de l'organisation/du projet. Il devrait préciser ce que pourrait être ce rôle en termes de supervision, de prise de décision et autres.

2 Compréhension des données

"La phase de compréhension des données commence par une collecte initiale des données. L'analyste procède ensuite à une familiarisation accrue avec les données, à l'identification des problèmes de qualité des données, à la découverte des premiers aperçus des données, ou à la détection de sous-ensembles intéressants pour former des hypothèses sur des informations cachées. La phase de compréhension des données comporte quatre étapes, à savoir la collecte des données initiales, la description des données, l'exploration des données et la vérification de la qualité des données".[528]

[528] Shearer, C. (2000) 'The CRISP-DM model : the new blueprint for data mining', *Journal of Data Warehousing* 5(4) : 13-23, p. 15. Disponible à l'adresse : https://mineracaodedados.files.wordpress.com/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf (consulté le 15 mai 2020).

2.1 Description

À ce stade, la collecte initiale des données a lieu et une première étude des données est réalisée. Elle comporte quatre tâches séquentielles :

- Collecter les données initiales
- Décrire les données
- Analyser les données
- Vérifier la qualité des données.

Toutes ces tâches visent à identifier les données disponibles. À ce stade, les développeurs doivent être conscients des données avec lesquelles ils auront à travailler et commencer à prendre des décisions sur la manière dont les grands principes liés à la protection des données seront mis en œuvre.

2.2 Principales mesures à prendre

À ce stade, un très grand nombre de questions fondamentales liées à la protection des données personnelles doivent être abordées. En fonction des décisions prises, des principes tels que la minimisation des données, la protection de la vie privée dès la conception ou par défaut, la licéité, la loyauté et la transparence, etc. seront réglés de manière adéquate.

2.2.1 Type de données collectées

Selon le RGPD, le responsable du traitement "met en œuvre les mesures techniques et organisationnelles appropriées pour garantir que, par défaut, seules les données à caractère personnel qui sont nécessaires à chaque finalité spécifique du traitement sont traitées". Cette obligation s'applique à la quantité de données à caractère personnel collectées, à l'étendue de leur traitement, à la durée de leur conservation et à leur accessibilité. En particulier, ces mesures doivent garantir que, par défaut, les données à caractère personnel ne sont pas rendues accessibles, sans l'intervention de la personne concernée, à un nombre indéfini de personnes physiques"[529] (voir la section "Protection des données dès la conception et par défaut", dans les "Concepts principaux" de la partie II). Il convient de garder cela à l'esprit, notamment au cours de cette étape, car c'est souvent à ce moment-là que sont prises les décisions relatives au type de données qui seront utilisées.

Les responsables du traitement doivent considérer qu'il est toujours préférable d'éviter d'utiliser les données personnelles si cela est possible. En effet, selon le principe de minimisation des données, l'utilisation des données personnelles doit être adéquate, pertinente et limitée à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées. Par conséquent**, si la même finalité peut être atteinte sans utiliser de données personnelles, le traitement doit être évité**.

Dans un deuxième niveau de précaution, **si les développeurs doivent utiliser des données personnelles, ils doivent essayer d'éviter d'utiliser des données de**

[529] Article 24 du RGPD.

catégorie spéciale. C'est parfois faisable, parfois non. Cela dépend souvent du domaine d'application du modèle. Ce n'est pas la même chose de travailler sur un modèle qui sera utilisé pour l'analyse de l'influence de l'épigénétique sur la santé humaine, un modèle utilisé pour surveiller une épidémie ou un modèle qui servira à cibler les publicités avec précision. Si ces données de catégorie spéciale sont finalement utilisées, les responsables du traitement doivent tenir compte de la réglementation applicable à leur traitement et de l'application nécessaire de garanties appropriées, capables de protéger les droits, les intérêts et les libertés des personnes concernées. La proportionnalité entre l'objectif de la recherche et l'utilisation des catégories particulières de données doit être garantie. En outre, les responsables du traitement doivent s'assurer que la réglementation de leur État membre ne protège pas les données génétiques, biométriques et de santé en introduisant des conditions ou des limitations supplémentaires, puisqu'ils sont habilités à le faire par le RGPD.

Si des données à caractère personnel sont nécessaires, le développeur d'IA devrait au moins essayer de réduire autant que possible la quantité de données considérées (voir la section "Minimisation des données" dans les "Principes" de la partie II). Il ne doit jamais oublier qu'il ne peut traiter des données que si le traitement est adéquat et pertinent. Par conséquent, ils doivent éviter d'utiliser une quantité excessive de données à caractère personnel. Trop souvent, cela est plus facile à faire qu'il n'y paraît. Comme l'indique l'Agence norvégienne de protection des données, "[i]l convient de noter que la qualité des données d'entraînement, ainsi que les caractéristiques utilisées, peuvent dans de nombreux cas être beaucoup plus importantes que la quantité. Lors de la formation d'un modèle, il est important que la sélection des données de formation soit représentative de la tâche à résoudre ultérieurement. D'énormes volumes de données sont de peu d'utilité s'ils ne couvrent qu'une fraction de ce sur quoi le modèle travaillera par la suite."[530] Il est donc particulièrement important de **ne pas collecter de données inutiles**. Un étiquetage correct pourrait être un bon antidote contre la collecte inutile. Notez que **si les données sont déjà stockées, la sélection implique la suppression des éléments de données inutiles.**

Le développeur doit toujours essayer d'éviter la "malédiction de la dimensionnalité", c'est-à-dire "une performance médiocre des algorithmes et leur complexité élevée associées à un cadre de données ayant un grand nombre de dimensions/caractéristiques, ce qui rend souvent la fonction cible assez complexe et peut conduire à un surajustement du modèle tant que l'ensemble de données se trouve souvent dans le collecteur de dimension inférieure".[531] À cette fin**, il peut être extrêmement important de disposer d'un expert capable de sélectionner les caractéristiques pertinentes.** Cela permettrait de réduire considérablement la quantité de données personnelles utilisées sans perdre en qualité. Cela ne devrait pas être difficile si le spécialiste des données connaît bien l'ensemble de données et la signification de ses caractéristiques numériques. Dans ces conditions, il serait facile de déterminer si certaines des variables

---

[530] Autorité norvégienne de protection des données (2018) Intelligence artificielle et vie privée. Autorité norvégienne de protection des données, Oslo. Disponible à l'adresse : https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf (consulté le 15 mai 2020).

[531] Oliinyk, H. (2018) Pourquoi et comment se débarrasser correctement de la malédiction de la dimensionnalité (avec visualisation d'un ensemble de données sur le cancer du sein). Vers la science des données, 20 mars. Disponible à l'adresse : https://towardsdatascience.com/why-and-how-to-get-rid-of-the-curse-of-dimensionality-right-with-breast-cancer-dataset-7d528fb5f6c0 (consulté le 15 mai 2020).

sont nécessaires ou non. Toutefois, une telle approche n'est possible que si l'ensemble de données est facile à interpréter et si les dépendances entre les variables sont bien connues. Par conséquent, le développeur aura besoin d'une plus petite quantité de données si elles ont été correctement classées. Les données intelligentes pourraient être beaucoup plus utiles que les données volumineuses. Bien sûr, cela pourrait impliquer un effort énorme en termes d'unification, d'homogénéisation, etc., mais cela aidera à mettre en œuvre le principe de minimisation des données (voir "Principe de minimisation des données" dans la partie II, section "Principes" des présentes lignes directrices) d'une manière beaucoup plus efficace.

En outre, les responsables du traitement devraient essayer de **limiter la résolution des données** à ce qui est minimalement nécessaire aux fins poursuivies par le traitement. Ils doivent également **déterminer un niveau optimal d'agrégation des données** avant de commencer le traitement (voir la section "Adéquat, pertinent et limité" de la section "Minimisation des données" des "Principes" de la partie II).

La minimisation des données peut être complexe dans le cas de l'apprentissage profond, où la discrimination par caractéristiques peut être impossible. Il existe un moyen efficace de réguler la quantité de données recueillies et de ne l'augmenter que si cela semble nécessaire : la courbe d'apprentissage[532] . Le développeur doit commencer par collecter et utiliser une quantité limitée de données d'apprentissage, puis surveiller la précision du modèle lorsqu'il est alimenté par de nouvelles données.

Encadré 16 : Une pratique de minimisation des données qui n'a pas été mise en œuvre de manière adéquate

Un outil développé par l'administration fiscale norvégienne pour filtrer les erreurs dans les déclarations d'impôts a testé 500 variables lors de la phase d'entraînement. Toutefois, seules 30 d'entre elles ont été incluses dans le modèle d'IA final, car elles se sont avérées les plus pertinentes pour la tâche à accomplir. Cela signifie qu'ils auraient probablement pu éviter de collecter autant de données personnelles s'ils avaient effectué une meilleure sélection des variables pertinentes dès le début.[533]

2.2.2 Sélection de base juridique appropriée pour le traitement

Les responsables du traitement doivent décider de la base juridique qu'ils utiliseront pour le traitement avant de le commencer, documenter leur décision dans l'avis de confidentialité (ainsi que les finalités) et inclure les raisons pour lesquelles ils ont fait ces choix (voir la section "Responsabilité" dans les "Principes" de la partie II). En principe, ils doivent choisir la **base juridique qui reflète le mieux la véritable nature de leur relation avec la personne et la finalité du traitement.** Cette décision est essentielle, car il n'est pas possible de changer la base juridique du traitement s'il n'y a pas de raisons solides qui le justifient (voir la section "Limitation de la finalité" des "Principes" dans la partie II).

---

[532] Ng, R. (pas de date) Learning curve. Disponible à l'adresse : www.ritchieng.com/machinelearning-learning-curve/ (consulté le 15 mai 2020).

[533] Autorité norvégienne de protection des données (2018) Intelligence artificielle et vie privée. Autorité norvégienne de protection des données, Oslo. Disponible à l'adresse : https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf (consulté le 15 mai 2020).

En principe, le consentement est l'un des fondements juridiques les plus courants du traitement (voir la section "Consentement" des "Concepts principaux" de la partie II). Toutefois, il comporte certains risques. En effet, le consentement est toujours lié à des finalités spécifiques. Par conséquent, "l'élargissement" des finalités du traitement au-delà du consentement explicite des personnes concernées est considéré comme un traitement illicite. Afin de déterminer si un traitement ultérieur est compatible ou non avec le traitement initial, les responsables du traitement doivent utiliser les critères inclus dans l'article 6, paragraphe 4, du RGPD (voir la sous-section "Quand les finalités sont-elles compatibles ?" dans la section "Limitation des finalités"). Comme mentionné, le traitement à des fins de recherche scientifique ou historique ou à des fins statistiques ne doit pas être considéré comme incompatible avec les finalités initiales (voir la section "Protection des données et recherche scientifique" dans les "Concepts principaux" de la partie II).

Les motifs alternatifs les plus courants pour le traitement des données dans l'IA sont les intérêts légitimes, l'exécution du contrat et l'obligation légale ou l'intérêt vital. Tous comportent des caractéristiques spécifiques qui doivent être soigneusement analysées.

2.2.3 Vérification de l'utilisation légitime des jeux de données

Les ensembles de données peuvent être obtenus de différentes manières. Tout d'abord, le développeur peut choisir d'acquérir ou d'obtenir l'accès à une base de données qui a déjà été construite par quelqu'un d'autre. Si tel est le cas, le responsable du traitement doit être particulièrement prudent, car l'acquisition de l'accès à une base de données soulève de nombreuses questions juridiques (voir la section "Achat de l'accès à une base de données" dans la partie "Principaux outils et actions" de la partie II). [534]

Deuxièmement, l'alternative la plus courante consiste à créer une base de données. De toute évidence, dans ce cas, les responsables du traitement doivent s'assurer qu'ils se conforment à toutes les exigences légales imposées par le RGPD pour créer une base de données (voir la section "Création d'une base de données" dans les "Principaux outils et actions" de la partie II des présentes lignes directrices).

Troisièmement, les développeurs choisissent parfois une autre voie. Ils **mélangent des données sous licence provenant de tiers entre elles ou avec l'ensemble de données des responsables du traitement, de manière à créer un énorme ensemble de données de formation et un autre à des fins de validation.** Cela peut poser certains problèmes, comme par exemple la possibilité que la combinaison de ces données personnelles fournisse des informations supplémentaires sur les personnes concernées. Par exemple, elle pourrait permettre au responsable du traitement d'identifier les personnes concernées, ce qui n'était pas possible auparavant. Cela pourrait impliquer la désanonymisation de données anonymes et la création de nouvelles informations personnelles qui ne figuraient pas dans l'ensemble de données d'origine, ce qui poserait des problèmes éthiques et juridiques considérables. Par conséquent, la

[534] Yeong Z. K. (2019) Legal issues in AI deployment. Law Gazette, février. Disponible à l'adresse : https://lawgazette.com.sg/feature/legal-issues-in-ai-deployment/ (consulté le 15 mai 2020).

réidentification doit être testée par des méthodes telles que les techniques de k-anonymat, de l-diversité ou de t-proximité.[535]

Un autre problème courant est que la base initiale du traitement des données recueillies dans chaque ensemble de données est différente. Si un responsable du traitement fusionne les ensembles de données et qu'ensuite l'une des bases juridiques n'est plus applicable, il se retrouve dans une situation terrible. Par exemple, si l'une des bases de données a été construite sur la base du consentement et que certaines des personnes concernées retirent leur consentement, le responsable du traitement devra les supprimer de l'ensemble de données fusionné. Cela pourrait être très difficile à réaliser dans la pratique.

En outre, les nouvelles informations ainsi produites peuvent également être fondées sur des probabilités ou des conjectures, et donc être fausses, ou contenir des biais dans la représentation des personnes (voir la section "Dispositions du **Dispositions** du ").[536] Par conséquent, les responsables du traitement doivent essayer d'éviter de telles conséquences en s'assurant que la fusion des ensembles de données ne va pas à l'encontre des droits et des intérêts des personnes concernées.

Enfin, si les responsables du traitement utilisent plusieurs ensembles de données qui poursuivent des finalités différentes, ils doivent mettre en œuvre des mesures adéquates pour séparer les différentes activités de traitement. Sinon, ils pourraient facilement utiliser des données collectées pour une seule finalité à des fins différentes. Cela pourrait poser des problèmes liés au principe de limitation de la finalité.

3 Préparation des données

"La phase de préparation des données couvre toutes les activités visant à construire l'ensemble de données final ou les données qui seront introduites dans le ou les outils de modélisation à partir des données brutes initiales. Les tâches comprennent la sélection des tables, des enregistrements et des attributs, ainsi que la transformation et le nettoyage des données pour les outils de modélisation. Les cinq étapes de la préparation des données sont la sélection des données, le nettoyage des données, la construction des données, l'intégration des données et le formatage des données."[537]

---

[535] Rajendran, K., Jayabalan, M. et Rana, M. E. (2017) "A study on k-anonymity, l-diversity, and t-closeness techniques focusing medical data", International Journal of Computer Science and Network Security 17(12) : 172-177.

[536] Projet SHERPA (2019) Lignes directrices pour le développement éthique des systèmes d'IA et de big data : une approche éthique par la conception. SHERPA, p. 38. Disponible à l'adresse : www.project-sherpa.eu/wp-content/uploads/2019/12/development-final.pdf (consulté le 15 mai 2020).

[537] Shearer, C. (2000) 'The CRISP-DM model : the new blueprint for data mining', *Journal of Data Warehousing* 5(4) : 13-23, p. 16. Disponible à l'adresse : https://mineracaodedados.files.wordpress.com/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf (consulté le 15 mai 2020).

3.1 Description

Cette étape comprend toutes les activités nécessaires pour construire l'ensemble de données final qui est introduit dans le modèle, à partir des données brutes initiales. Elle comprend les cinq tâches suivantes, qui ne sont pas nécessairement exécutées de manière séquentielle.

1. Sélectionner les données. Décidez des données à utiliser pour l'analyse, en fonction de leur pertinence par rapport aux objectifs de l'exploration de données, de leur qualité et des contraintes techniques telles que les limites du volume ou des types de données.
2. Nettoyer les données. Amenez la qualité des données à un niveau requis, par exemple en sélectionnant des sous-ensembles de données propres, en insérant des valeurs par défaut et en estimant les données manquantes par modélisation.
3. Construire des données. La construction de nouvelles données par la production d'attributs dérivés, de nouveaux enregistrements ou de valeurs transformées pour des attributs existants.
4. Intégrer des données. Combinez les données de plusieurs tables ou enregistrements pour créer de nouveaux enregistrements ou valeurs.
5. Formater les données. Apportez des modifications syntaxiques aux données qui pourraient être requises par l'outil de modélisation.

3.2 Principales mesures à prendre

3.2.1 Garantir la précision des données personnelles

Selon le RGPD, les données doivent être exactes (voir la section "Précision" dans les "Principes" de la partie II). Cela signifie que les données sont correctes et mises à jour, mais fait également référence à l'exactitude des analyses effectuées. L'EDPB a souligné l'importance de l'exactitude du profilage ou du processus décisionnel (non exclusivement) automatisé à toutes les étapes (de la collecte des données à l'application du profil à l'individu).[538]

Les responsables du traitement sont chargés de garantir la précision des données à caractère personnel. Par conséquent, une fois qu'ils ont terminé la collecte des données personnelles, ils doivent mettre en place des outils adéquats pour garantir la précision de ces données. Cela implique essentiellement la mise en œuvre de mesures techniques et

[538] *Lignes directrices* sur la *prise de décision* individuelle *automatisée et le profilage* aux fins du règlement 2016/679 (wp251rev.01). 22/08/2018, *p.* 13 ; Ducato, Rossana, Private Ordering of Online Platforms in Smart Urban Mobility The Case of Uber's Rating System, CRIDES Working Paper Series no. 3/20202 February 2020 Updated on 26 July 2020, p. 20-21, at: https://poseidon01.ssrn.com/delivery.php?ID=247104118003073117118086021112071111102048023015008020118084071112086000027097102088036101006014057116105116119119026079007006118044033055000114023106007076115096073024007094081002078064098028091093003078095099082108113086098120001079015123027083125024&EXT=pdf&INDEX=TRUE

organisationnelles qui garantiront l'application de ce principe (voir la sous-section "Mesures techniques et organisationnelles connexes" dans la section "Précision" des "Principes" de la partie II). Si les données à caractère personnel proviennent des personnes concernées, le responsable du traitement peut supposer qu'elles sont exactes (sauf si le responsable considère que la personne concernée pourrait avoir une raison de fournir des données inexactes). Si les données à caractère personnel n'ont pas été collectées auprès de la personne concernée, les responsables du traitement sont tenus "de vérifier la précision des données obtenues, au moins en ce qui concerne leur adéquation aux finalités déclarées du traitement et les conséquences négatives que les inexactitudes peuvent avoir pour les personnes concernées." (voir la sous-section "Comment l'inexactitude des données est-elle découverte ?" dans la section "Précision" des "Principes" de la partie II). En tout état de cause, la précision exige une mise en œuvre adéquate des mesures consacrées à faciliter le droit de rectification des personnes concernées (voir "Droit de rectification" dans la partie II, section "Droits des personnes concernées").

3.2.2 Se concentrer sur les questions de profilage

Dans le cas d'une base de données qui servira à former ou à valider un outil d'IA, il existe une obligation particulièrement pertinente d'informer les personnes concernées que **leurs données pourraient entraîner une prise de décision automatisée ou un profilage à leur égard, à moins que les responsables du traitement ne puissent garantir que l'outil ne produira en aucun cas ces conséquences.**

Même si la prise de décision automatique peut difficilement se produire dans le contexte de la recherche, les développeurs doivent prêter attention à cette question.

Le profilage, en revanche, pourrait poser certains problèmes au développement de l'IA.

Cela est dû à une raison simple : le processus de profilage est "souvent invisible pour la personne concernée. Il fonctionne en créant des données dérivées ou déduites sur les individus - de "nouvelles" données personnelles qui n'ont pas été fournies directement par les personnes concernées elles-mêmes. Les individus ont des niveaux de compréhension différents et peuvent trouver difficile de comprendre les techniques complexes impliquées dans les processus de profilage et de prise de décision automatisée."[539] Ainsi, "si le responsable du traitement envisage un "modèle" dans lequel il prend uniquement des décisions automatisées ayant un impact élevé sur les personnes sur la base de profils établis à leur sujet et qu'il ne peut pas s'appuyer sur le consentement de la personne, sur un contrat avec elle ou sur une loi l'autorisant, le responsable du traitement ne doit pas poursuivre."[540] Le risque pour les droits, intérêts et libertés de la personne est un facteur très important qui doit toujours être pris en compte. Ce n'est pas le même type de profilage que de prendre une décision sur les goûts d'une personne en matière de séries télévisées, par rapport au profilage consacré à l'approbation de sa police d'assurance maladie. Ainsi, si le traitement présente des risques pour les libertés et droits fondamentaux des personnes, les responsables du

[539] Groupe de travail Article 29 (2017) Lignes directrices sur la prise de décision individuelle automatisée et le profilage aux fins du règlement 2016/679. Adopté le 3 octobre 2017, tel que révisé en dernier lieu et adopté le 6 février 2018. Commission européenne, Bruxelles, p.9. Disponible à l'adresse : https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053 (consulté le 15 mai 2020).

[540] Ibid. p.30.

traitement doivent s'assurer qu'ils peuvent faire face à ces risques et répondre aux exigences.

En cas de traitement et/ou de prise de décision automatisée, les personnes concernées **doivent être informées de manière adéquate sur ce traitement et sur le fonctionnement de l'algorithme.** En d'autres termes, leur droit à l'information doit être satisfait en application du principe de licéité, de loyauté et de transparence. Cela signifie que, au minimum, les responsables du traitement doivent informer la personne concernée qu'"ils se livrent à ce type d'activité, fournir des informations significatives sur la logique impliquée et sur l'importance et les conséquences envisagées du profilage pour la personne concernée". [541]

Les informations sur la logique d'un système et les explications des décisions doivent donner aux individus le contexte nécessaire pour décider s'ils souhaitent demander une intervention humaine, et pour quels motifs. Dans certains cas, des explications insuffisantes peuvent inciter les personnes à recourir inutilement à d'autres droits, ou à retirer leur consentement. Les demandes d'intervention, l'expression de points de vue ou les contestations sont plus susceptibles de se produire si les individus estiment ne pas avoir une compréhension suffisante de la manière dont la décision a été prise. [542]

Enfin, un responsable du traitement doit toujours se rappeler que, conformément à l'article 22, paragraphe 3, les décisions automatisées qui concernent des catégories particulières de données à caractère personnel ne sont autorisées que si la personne concernée a donné son consentement ou si elles sont fondées sur une base juridique (voir la section '**Capacité d'action humaine et surveillance** de cette partie des lignes directrices). Cette exception s'applique non seulement lorsque les données observées entrent dans cette catégorie, mais **aussi si le rapprochement de différents types de données personnelles peut révéler des informations sensibles sur des personnes ou si des données déduites entrent dans cette catégorie.** Dans tous ces cas, il faut parler de traitement de catégories particulières de données à caractère personnel. En effet, une étude capable de déduire des catégories spéciales de données est soumise aux mêmes obligations légales en vertu du RGPD que si des données personnelles sensibles avaient été traitées dès le départ. Si le profilage déduit des données personnelles qui n'ont pas été fournies par la personne concernée, les responsables du traitement doivent s'assurer que le traitement n'est pas incompatible avec la finalité initiale, qu'ils ont identifié une base légale pour le traitement des données de catégorie spéciale et qu'ils informent la personne concernée du traitement.[543]

Encadré 17 : Exemple d'inférence de données de catégories spéciales

---

[541] Ibid. p.13-14.

[542] ICO (2020) AI auditing framework - draft guidance for consultation. Information Commissioner's Office, Wilmslow, p.94. Disponible sur : https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf (consulté le 15 mai 2020).

[543] Groupe de travail Article 29 (2018) Lignes directrices sur la prise de décision individuelle automatisée et le profilage aux fins du règlement 2016/679. Adoptées le 3 octobre 2017, telles que révisées en dernier lieu et adoptées le 6 février 2018. Commission européenne, Bruxelles, p.15. Disponible à l'adresse : https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053 (consulté le 15 mai 2020).

La recherche a montré qu'en 2011, des enregistrements numériques facilement accessibles du comportement, les "J'aime" sur Facebook, pouvaient être utilisés pour prédire automatiquement et avec précision une série d'attributs personnels très sensibles, notamment l'orientation sexuelle, l'origine ethnique, les opinions religieuses et politiques, les traits de personnalité, l'intelligence, le bonheur, la consommation de substances addictives, la séparation des parents, l'âge et le sexe. L'analyse s'est basée sur un ensemble de données de plus de 58 000 volontaires qui ont fourni leurs "J'aime" sur Facebook, des profils démographiques détaillés et les résultats de plusieurs tests psychométriques. Le modèle a correctement distingué les hommes homosexuels des hommes hétérosexuels dans 88 % des cas, les Afro-Américains des Américains de type caucasien dans 95 % des cas, et les démocrates des républicains dans 85 % des cas. Pour le trait de personnalité "Ouverture", la précision de la prédiction était proche de la précision test-retest d'un test de personnalité standard. Les auteurs ont fourni des exemples d'associations entre les attributs et les goûts et discutent des implications pour la personnalisation en ligne et la vie privée.[544]

La réalisation d'une AIPD est obligatoire s'il existe un risque réel de profilage non autorisé ou de prise de décision automatisée. L'article 35(3) (a) du RGPD stipule la nécessité pour le responsable du traitement d'effectuer une AIPD dans le cas d'une évaluation systématique et extensive d'aspects personnels concernant des personnes physiques qui est basée sur un traitement automatisé, y compris le profilage, et sur laquelle sont fondées des décisions produisant des effets juridiques concernant la personne physique ou l'affectant de manière significative de façon similaire. Les responsables du traitement doivent savoir qu'à l'heure actuelle, chaque pays a soumis à l'EDPB sa liste des cas dans lesquels un AIPD est requis. Si le responsable du traitement se trouve dans l'EEE, cette liste doit également être vérifiée localement[545] **(**voir "AIPD" dans la partie II section "Principales actions et outils" de ces lignes directrices).

Selon l'article 37, paragraphe 1, point b), et paragraphe 5 du RGPD, les responsables du traitement désignent un délégué à la protection des données lorsque "les activités principales du responsable du traitement ou du sous-traitant consistent en des opérations de traitement qui, en raison de leur nature, de leur portée et/ou de leurs finalités, nécessitent un suivi régulier et systématique des personnes concernées à grande échelle." Les responsables du traitement sont également tenus de conserver un registre de toutes les décisions prises par un outil d'IA dans le cadre de leurs obligations de responsabilité et de documentation. Cela devrait également indiquer si une personne a demandé une intervention humaine, a exprimé un point de vue, a contesté la décision, et si une décision a été modifiée à la suite[546] (voir la section "Principe de responsabilité" dans les "Principes" de la partie II).

---

[544] Kosinski, M., Stillwell, D. et Graepel, T. (2013) 'Digital records of behavior expose personal traits', *Proceedings of the National Academy of Sciences* 110 (15) : 5802-5805, DOI : 10.1073/pnas.1218772110.

[545] EDPB (2019) Analyse d'impact sur la protection des données. Conseil européen de la protection des données, Bruxelles. Disponible à l'adresse : https://edpb.europa.eu/our-work-tools/our-documents/topic/data-protection-impact-assessment-DPIA_es (consulté le 3 juin 2020).

[546] ICO (2020) Guidance on the AI auditing framework - draft guidance for consultation. Information Commissioner's Office, Wilmslow, p.94-95. Disponible à l'adresse : https://ico.org.uk/media/about-the-

Voici quelques actions supplémentaires qui pourraient être extrêmement utiles pour éviter la prise de décision automatisée : [547]

- Prenez en compte les exigences du système nécessaires pour soutenir une révision humaine significative **dès la phase de conception.** En particulier, les exigences d'interprétabilité et la conception d'une interface utilisateur efficace pour soutenir les examens et les interventions humaines.
- Concevez et dispensez une formation et un soutien appropriés aux examinateurs humains.
- Donnez au personnel l'autorité, les incitations et le soutien appropriés pour répondre aux préoccupations des personnes ou les transmettre à un échelon supérieur et, si nécessaire, passer outre la décision de l'outil d'IA.

3.2.3 Sélection de données non biaisées

Les biais sont l'un des principaux problèmes liés au développement de l'IA, un problème qui va à l'encontre du principe de loyauté (voir "Principe de licéité, de loyauté et de transparence" dans la partie II, section "Principes" des présentes lignes directrices). Les biais peuvent être causés par un grand nombre de problèmes différents. Lorsque des données sont recueillies, elles peuvent contenir des biais, des inexactitudes, des erreurs et des fautes construits par la société. Parfois, il peut arriver que les ensembles de données soient biaisés en raison d'actions malveillantes. L'introduction de données malveillantes dans un outil d'IA peut modifier son comportement, en particulier avec les systèmes d'auto-apprentissage.[548] Par exemple, dans le cas du chatbot Tay, développé par Microsoft, un grand nombre d'internautes ont commencé à poster des commentaires racistes et sexistes qui ont servi à alimenter l'algorithme. Le résultat final est que Tay a commencé à envoyer des tweets racistes et sexistes après seulement quelques heures de fonctionnement. En d'autres occasions, le principal problème est que l'ensemble de données ne représente pas bien la population considérée et l'objectif visé. Par conséquent, il contient des biais cachés qui seront transposés à l'outil entraîné qui reflétera ces biais, et cela pourrait conduire à des résultats du modèle incorrects ou discriminatoires. [549]

Par conséquent, les questions liées à la composition des bases de données utilisées pour la formation soulèvent des problèmes éthiques et juridiques cruciaux, et pas seulement des questions d'efficacité ou de nature technique. Elles doivent donc être abordées avant la formation de l'algorithme. Les modèles d'IA doivent "être entraînés à l'aide de

---

ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf (consulté le 15 mai 2020).

[547] Ibid, p. 95.

[548] Groupe d'experts de haut niveau sur l'IA (2019) Lignes directrices en matière d'éthique pour une IA digne de confiance. Commission européenne, Bruxelles, p.17. Disponible à l'adresse : https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (consulté le 15 mai 2020).

[549] Pour une définition de la discrimination directe et indirecte, voir, par exemple, l'article 2 de la directive 2000/78/CE du Conseil du 27 novembre 2000 portant création d'un cadre général en faveur de l'égalité de traitement en matière d'emploi et de travail. Voir également l'article 21 de la Charte des droits fondamentaux de l'UE.

données pertinentes et correctes et ils doivent apprendre quelles sont les données à privilégier. Le modèle ne doit pas mettre en avant les informations relatives à l'origine raciale ou ethnique, aux opinions politiques, à la religion ou aux convictions, à l'appartenance syndicale, au statut génétique, à l'état de santé ou à l'orientation sexuelle si cela conduit à un traitement discriminatoire arbitraire."[550] Les biais identifiables et discriminatoires doivent être supprimés lors de la phase de constitution des ensembles de données, dans la mesure du possible.

Encadré 18 : Comprendre les biais : le cas du gorille

En 2015, un ingénieur logiciel, Jacky Alciné, a dénoncé les algorithmes de reconnaissance d'images utilisés dans Google Photos qui classaient certaines personnes noires comme des "gorilles." Google a immédiatement reconnu le problème et a promis de le corriger.

Ce problème a été provoqué par une grave erreur lors de la phase d'entraînement. L'algorithme a été entraîné à reconnaître des personnes à l'aide d'un ensemble de données principalement composé de photographies de personnes caucasiennes. En conséquence, l'algorithme a considéré qu'une personne noire était beaucoup plus similaire à l'objet "gorille" qu'il avait été entraîné à reconnaître, qu'à l'objet "humain". Cet exemple montre parfaitement l'importance de la sélection des données à des fins d'entraînement.

Ainsi, pour intégrer les exigences éthiques dans cette phase, le développeur d'IA devrait évaluer les conséquences éthiques de la sélection des données par rapport à la diversité et apporter des modifications, si nécessaire. En effet, le responsable du traitement "devrait utiliser des procédures mathématiques ou statistiques appropriées pour le profilage, mettre en œuvre des mesures techniques et organisationnelles appropriées pour garantir, en particulier, que **les facteurs qui entraînent des inexactitudes dans les données à caractère personnel soient corrigés et que le risque d'erreurs soit réduit au minimum",** sécuriser les données à caractère personnel d'une manière qui tienne compte des risques potentiels qu'elles comportent pour les intérêts et les droits de la personne concernée et qui prévienne, entre autres, les effets discriminatoires à l'égard des personnes physiques sur la base de l'origine raciale ou ethnique, des opinions politiques, de la religion ou des convictions, de l'appartenance syndicale, du statut génétique ou de santé ou de l'orientation sexuelle, ou qui aboutissent à des mesures ayant un tel effet."[551]

Les responsables du traitement doivent toujours garder à l'esprit que ce qui rend cette question si spécifique, c'est que la sélection d'un ensemble de données pour l'entraînement implique de prendre des décisions et de faire des choix, parfois de manière presque inconsciente (alors que le codage d'un algorithme traditionnel et déterministe est toujours une opération délibérée). Celui qui entraîne un algorithme y intègre en quelque sorte sa propre façon de voir le monde, ses valeurs ou, à tout le moins, les valeurs qui sont plus ou moins directement inhérentes aux données recueillies

[550] Autorité norvégienne de protection des données (2018) Intelligence artificielle et vie privée. Autorité norvégienne de protection des données, Oslo. Disponible à l'adresse : https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf (consulté le 15 mai 2020).

[551] Considérant 71 du RGPD.

dans le passé.[552] Cela signifie que les **équipes chargées de sélectionner les données à intégrer dans les jeux de données devraient être composées de personnes qui garantissent la diversité dont le développement de l'IA est censé faire preuve.** Dans tous les cas, une expertise juridique sur la réglementation anti-discrimination pourrait être pertinente sur ce point.

4 Modélisation (formation)

"Dans cette phase, diverses techniques de modélisation sont sélectionnées et appliquées et leurs paramètres sont calibrés à des valeurs optimales. Généralement, plusieurs techniques existent pour le même type de problème d'exploration de données. Certaines techniques ont des exigences spécifiques sur la forme des données. Par conséquent, il peut être nécessaire de revenir à la phase de préparation des données. Les étapes de modélisation comprennent la sélection de la technique de modélisation, la génération du plan de test, la création de modèles et l'évaluation des modèles."[553]

4.1 Description

Cette phase comporte plusieurs tâches essentielles. Globalement, le développeur doit effectuer les tâches suivantes :

- **Sélectionner la technique de modélisation qui sera utilisée.** Selon le type de technique, des conséquences telles que l'inférence des données, l'obscurité ou les biais sont plus ou moins susceptibles de se produire.
- **Prendre une décision sur l'outil de formation à utiliser.** Cela permet au développeur de mesurer la capacité du modèle à prédire l'histoire avant de l'utiliser pour prédire l'avenir. La formation implique toujours l'exécution de tests empiriques avec des données personnelles. Parfois, les développeurs testent le modèle avec des données différentes de celles utilisées pour le générer. Par conséquent, à ce stade, on peut parler de différents types d'ensembles de données. Il est parfois difficile d'identifier les personnes auxquelles se rapportent les données de formation. Cela crée des problèmes pour le respect des droits des personnes, qui doivent être traités de manière appropriée.

---

[552] CNIL (2017) Comment l'humain peut-il garder la main ? Les questions éthiques soulevées par les algorithmes et l'intelligence artificielle. Commission nationale de l'informatique et des libertés, Paris, p.34. Disponible sur : www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf (consulté le 15 mai 2020).

[553] Shearer, C. (2000) 'The CRISP-DM model : the new blueprint for data mining', *Journal of Data Warehousing* 5(4) : 13-23, p. 17. Disponible à l'adresse : https://mineracaodedados.files.wordpress.com/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf (consulté le 15 mai 2020).

4.2 Principales mesures à prendre

4.2.1 Mise en œuvre du principe de minimisation des données

Selon le principe de limitation de la finalité (voir "Principe de limitation de la finalité" dans la partie II, section "Principes" des présentes lignes directrices), les responsables du traitement utilisant des outils d'IA déterminent la finalité de l'utilisation de l'outil d'IA dès le début de sa formation ou de son déploiement, et réévaluent cette détermination si le traitement du système donne des résultats inattendus, car il exige que les données à caractère personnel ne soient collectées que pour des "finalités déterminées, explicites et légitimes" et ne soient pas utilisées d'une manière incompatible avec la finalité initiale.

Selon le principe de minimisation des données, les responsables du traitement doivent procéder à la réduction de la quantité de données et/ou de l'éventail d'informations sur la personne concernée qu'ils fournissent dès que possible. Par conséquent, les données à caractère personnel utilisées pendant la phase de formation doivent être expurgées de toutes les informations qui ne sont pas strictement nécessaires à la formation du modèle (voir la sous-section "Aspect temporel" dans la section "Minimisation des données" des "Principes" de la partie II). Il existe de multiples stratégies pour assurer la minimisation des données au stade de la formation. Les techniques évoluent en permanence. Toutefois, certaines des plus courantes sont présentées ci-dessous ;[554] voir aussi la section "Intégrité et confidentialité" dans les "Principes" de la partie II) :

- Analyse des conditions que les données doivent remplir pour être considérées comme de haute qualité et dotées d'une grande capacité de prédiction pour l'application spécifique.
- Analyse critique de l'étendue de la typologie des données utilisées à chaque étape de l'outil d'IA.
- Suppression des données non structurées et des informations inutiles recueillies lors du prétraitement de l'information.
- Identification et suppression des catégories de données qui n'ont pas d'influence significative sur l'apprentissage ou sur le résultat de l'inférence.
- Suppression des conclusions non pertinentes associées aux informations personnelles pendant le processus de formation, par exemple, dans le cas d'une formation non supervisée.
- Utilisation de techniques de vérification qui nécessitent moins de données, comme la validation croisée.
- Analyse et configuration des hyperparamètres algorithmiques pouvant influencer la quantité ou l'étendue des données traitées afin de les minimiser.
- Utilisation de modèles d'apprentissage fédérés plutôt que centralisés.
- Application de stratégies de confidentialité différentielle.

---

[554] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, p.40. Disponible sur : www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (consulté le 15 mai 2020).

- Entraînement avec des données cryptées en utilisant des techniques homomorphiques.
- Agrégation de données.
- Anonymisation et pseudonymisation, non seulement dans la communication des données, mais aussi dans les données de formation, les éventuelles données personnelles contenues dans le modèle et dans le traitement de l'inférence.

4.2.2 Détecter et effacer les biais

Même si les mécanismes de lutte contre les biais sont convenablement adoptés lors des étapes précédentes (voir la section précédente sur la formation), il faut encore s'assurer que les résultats de la phase de formation minimisent les biais. Cela peut être difficile, car certains types de biais et de discrimination sont souvent particulièrement difficiles à détecter. Les membres de l'équipe qui conservent les données d'entrée n'en sont parfois pas conscients, et les utilisateurs qui sont leurs sujets n'en sont pas nécessairement conscients non plus. Ainsi, les systèmes de contrôle mis en place par le développeur d'IA lors de la phase de validation sont des facteurs extrêmement importants pour éviter les biais.

Il existe de nombreux outils techniques qui pourraient servir à détecter les biais, comme l'évaluation de l'impact algorithmique.[555] Le développeur d'IA doit envisager leur mise en œuvre effective.[556] Cependant, comme le montre la littérature[557] , il peut arriver qu'un algorithme ne puisse pas être totalement purgé de tous les différents types de biais. Le développeur doit cependant essayer d'être au moins conscient de leur existence et des implications que cela peut avoir (voir les sections "Licéité, loyauté et transparence" et "Précision" dans la partie II des présentes lignes directrices, "Principes").

4.2.3 Exercice des droits des personnes concernées

De toute évidence, les responsables du traitement doivent faciliter l'exercice de tous les droits des personnes concernées tout au long du cycle de vie. Toutefois, à ce stade spécifique, les droits d'accès, de rectification et d'effacement sont particulièrement sensibles et comportent certaines caractéristiques dont les responsables du traitement doivent être conscients.

a) Droit d'accès

En général, les données d'apprentissage peuvent difficilement être reliées à une personne concernée, car elles ne contiennent généralement que des informations

---

[555] Reisman, D., Crawford, K. et Whittaker, M. (2018) Algorithmic impact assessments : a practical framework for public agency accountability. AI Now Institute, New York, NY. Disponible à l'adresse : https://ainowinstitute.org/aiareport2018.pdf (consulté le 15 mai 2020).

[556] ICO (2020) AI auditing framework - draft guidance for consultation. Information Commissioner's Office, Wilmslow. Disponible sur : https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf (consulté le 15 mai 2020).

[557] Chouldechova, A. (2017) " Fair prediction with disparate impact : a study of bias in recidivism prediction instruments ", *Big Data* 5(2) : 153-163, http://doi.org/10.1089/big.2016.0047.

pertinentes pour les prédictions, telles que les transactions passées, les données démographiques ou la localisation, mais pas les coordonnées ou les identifiants uniques des clients. De plus, elles sont souvent prétraitées afin d'être plus facilement exploitables par les algorithmes. Toutefois, cela ne signifie pas que ces données peuvent être considérées comme entièrement pseudonymisées ou anonymisées. Elles restent donc des données à caractère personnel. Par exemple, dans le cas d'un modèle de prédiction d'achat, la formation peut inclure un modèle d'achat propre à un client. Dans cet exemple, si un client devait fournir une liste de ses achats récents dans le cadre de sa demande, l'organisation pourrait être en mesure d'identifier la partie des données d'apprentissage qui concerne cet individu.

Dans ces circonstances, les développeurs d'IA devraient répondre aux demandes des personnes concernées pour accéder à leurs données personnelles, en supposant qu'ils aient pris des mesures raisonnables pour vérifier l'identité de la personne concernée, et qu'aucune autre exception ne s'applique. Et, comme l'indique l'ICO, "les demandes d'accès, de rectification ou d'effacement des données de formation ne devraient pas être considérées comme manifestement infondées ou excessives simplement parce qu'elles peuvent être plus difficiles à satisfaire ou que la motivation de la demande peut être peu claire par rapport aux autres demandes d'accès qu'une organisation reçoit habituellement." [558]

Cependant, il est clair que les organisations ne sont pas tenues de collecter ou de conserver des données personnelles supplémentaires pour permettre l'identification des personnes concernées dans les données de formation dans le seul but de se conformer au règlement. Si les développeurs d'IA ne peuvent pas identifier une personne concernée dans les données de formation et que la personne concernée ne peut pas fournir d'informations supplémentaires qui permettraient son identification, les développeurs d'IA ne sont pas obligés de répondre à une demande qu'il n'est pas possible de satisfaire.

b) Droit de rectification

Dans le cas du droit de rectification, le responsable du traitement doit garantir le droit de rectification des données, notamment celles générées par les inférences et les profils établis par le développement de l'IA.

Même si l'objectif des données de formation est de former des modèles basés sur des modèles généraux dans de grands ensembles de données et que les inexactitudes individuelles sont donc moins susceptibles d'avoir un effet direct sur une personne concernée, le droit de rectification ne peut être limité. Au maximum, le responsable du traitement pourrait demander un délai plus long (deux mois supplémentaires) pour procéder à la rectification si la procédure technique est particulièrement complexe (article 11, paragraphe 3).

Encadré 19 : Rectification

Par exemple, il peut être plus important de rectifier une adresse de livraison

[558] ICO (2019) Permettre les droits d'accès, d'effacement et de rectification dans les outils d'IA. Bureau du commissaire à l'information, Wilmslow. Disponible à l'adresse : https://ico.org.uk/about-the-ico/news-and-events/ai-blog-enabling-access-erasure-and-rectification-rights-in-ai-systems/ (consulté le 15 mai 2020).

client enregistrée de manière incorrecte que de rectifier la même adresse incorrecte dans les données de formation. En effet, la première pourrait entraîner l'échec de la livraison, tandis que la seconde n'affecterait guère la précision globale du modèle.[559]

c) Droit à l'effacement

Les personnes concernées ont le droit de demander la suppression de leurs données personnelles. Toutefois, ce droit peut être limité si certaines circonstances concrètes s'appliquent. Selon l'ICO, "les organisations peuvent également recevoir des demandes d'effacement de données de formation. Les organisations doivent répondre aux demandes d'effacement lorsque les personnes concernées fournissent des motifs appropriés, sauf si une exemption légale pertinente s'applique. Par exemple, si les données de formation ne sont plus nécessaires parce que le modèle ML a déjà été formé, l'organisation doit répondre à la demande. Toutefois, dans certains cas, lorsque le développement du système est en cours, il peut encore être nécessaire de conserver les données de formation aux fins du réentraînement, du perfectionnement et de l'évaluation d'un outil d'IA. Dans ce cas, l'organisation doit adopter une approche au cas par cas pour déterminer si elle peut satisfaire les demandes. Se conformer à une demande de suppression des données d'entraînement n'entraînerait pas l'effacement des modèles ML basés sur ces données, sauf si les modèles eux-mêmes contiennent ces données ou peuvent être utilisés pour les déduire." [560]

5 Évaluation (validation)

"Avant de procéder au déploiement final du modèle construit par l'analyste de données, il est important de procéder à une évaluation plus approfondie du modèle et de revoir la construction du modèle pour s'assurer qu'il atteint correctement les objectifs de l'entreprise. Il est essentiel de déterminer si certaines questions importantes n'ont pas été suffisamment prises en compte. À la fin de cette phase, le chef de projet doit alors décider exactement comment utiliser les résultats de l'exploration de données. Les étapes clés ici sont l'évaluation des résultats, la révision du processus et la détermination des prochaines étapes."[561]

---

[559] ICO (2019) Permettre les droits d'accès, d'effacement et de rectification dans les systèmes d'IA. Bureau du commissaire à l'information, Wilmslow. Disponible à l'adresse : https://ico.org.uk/about-the-ico/news-and-events/ai-blog-enabling-access-erasure-and-rectification-rights-in-ai-systems/ (consulté le 15 mai 2020).

[560] ICO (2019) Permettre les droits d'accès, d'effacement et de rectification dans les systèmes d'IA. Bureau du commissaire à l'information, Wilmslow. Disponible à l'adresse : https://ico.org.uk/about-the-ico/news-and-events/ai-blog-enabling-access-erasure-and-rectification-rights-in-ai-systems/ (consulté le 15 mai 2020).

[561] Shearer, C. (2000) 'The CRISP-DM model : the new blueprint for data mining', *Journal of Data Warehousing* 5(4) : 13-23, p. 17. Disponible à l'adresse : https://mineracaodedados.files.wordpress.com/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf (consulté le 15 mai 2020).

5.1 Description

Cette phase comporte plusieurs tâches qui soulèvent d'importantes questions relatives à la protection des données. Globalement, le développeur doit :

- évaluer les résultats du modèle, par exemple, s'il est exact ou non ; à cette fin, le développeur d'IA peut le tester dans le monde réel
- revoir le processus : le développeur pourrait revoir la mission d'exploration de données pour déterminer s'il y a un facteur ou une tâche importante qui a été en quelque sorte négligée. Cela inclut les questions d'assurance qualité.

5.2 Principales actions à mener

5.2.1 Processus de validation dynamique

La validation du traitement comprenant un composant d'IA doit être effectuée dans des conditions qui reflètent l'environnement réel dans lequel le traitement est destiné à être déployé. En outre, le processus de validation nécessite un examen périodique si les conditions changent ou si l'on soupçonne que la solution elle-même peut être altérée. Les développeurs d'IA doivent s'assurer que la validation reflète fidèlement les conditions dans lesquelles l'algorithme a été validé.

Pour atteindre cet objectif, la validation doit inclure tous les composants d'un outil d'IA, y compris les données, les modèles pré-entraînés, les environnements et le comportement du système dans son ensemble, et être effectuée dès que possible. Dans l'ensemble, il faut s'assurer que les résultats ou les actions sont cohérents avec les résultats des processus précédents, en les comparant aux politiques préalablement définies pour s'assurer qu'elles ne sont pas violées.[562] La validation nécessite parfois la collecte de nouvelles données à caractère personnel. Dans d'autres cas, les responsables du traitement utilisent les données à des fins autres que celles prévues à l'origine. Dans tous ces cas, les responsables du traitement doivent s'assurer du respect du RGPD (voir la section "Limitation de la finalité" dans les "Principes" et "Protection des données et recherche scientifique" dans la section "Concepts principaux", toutes deux dans la partie II).

5.2.2 Suppression d'un jeu de données inutile

Très souvent, les processus de validation et de formation sont en quelque sorte liés. Si la validation recommande des améliorations du modèle, la formation doit être effectuée à nouveau.

En principe, une fois le développement de l'IA achevé, l'étape de formation de l'outil d'IA est terminée. À ce moment-là, vous devriez mettre en œuvre la suppression de

[562] Groupe d'experts de haut niveau sur l'IA (2019) Lignes directrices en matière d'éthique pour une IA digne de confiance. Commission européenne, Bruxelles, p.22. Disponible à l'adresse : https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (consulté le 15 mai 2020).

l'ensemble des données utilisées à cette fin, à moins qu'il n'existe un besoin légitime de les conserver pour affiner ou évaluer le système, ou pour d'autres fins compatibles avec celles pour lesquelles elles ont été collectées conformément aux conditions de l'article 6, paragraphe 4, du RGPD (voir la section "Définir des politiques de stockage des données adéquates"). Cependant, les développeurs d'IA doivent toujours considérer que la suppression des données personnelles peut aller à l'encontre de la nécessité de mettre à jour la précision des outils basés sur l'auto-apprentissage en temps réel des algorithmes : si une erreur est trouvée, ils devront probablement rappeler les données précédemment utilisées dans la phase de formation.

Dans le cas où les personnes concernées demandent son effacement, le responsable du traitement doit adopter une approche au cas par cas en tenant compte des limitations à ce droit prévues par le règlement (voir article 17, paragraphe 3). [563]

5.2.3 Réalisation d'un audit externe du traitement des données

Dans les cas où les risques de traitement des données personnelles au sein de l'outil d'IA sont élevés, **un audit du système par un tiers indépendant doit être envisagé**. Différents types d'audits peuvent être utilisés. Ils peuvent être internes ou externes, ne porter que sur le produit final ou être réalisés sur des prototypes moins évolués. Ils peuvent être considérés comme une forme de contrôle ou un outil de transparence. L'annexe I, à la fin de ce document, contient certaines recommandations de l'agence espagnole de protection des données qui pourraient servir de modèle.

En termes d'exactitude juridique, les outils d'IA doivent être audités afin de vérifier si le traitement des données personnelles au sein de leur système remplit les obligations stipulées dans le RGPD, compte tenu du large éventail de questions qui se posent. Le groupe d'experts de haut niveau sur l'IA a déclaré que "les processus de test devraient être conçus et réalisés par un groupe de personnes aussi diversifié que possible. Des mesures multiples devraient être développées pour couvrir les catégories qui sont testées pour différentes perspectives. On peut envisager des tests contradictoires effectués par des "équipes rouges" fiables et diverses qui tentent délibérément de "casser" le système pour trouver des vulnérabilités, ainsi que des "primes aux bogues" qui incitent les personnes extérieures à détecter et à signaler de manière responsable les erreurs et les faiblesses du système."[564] Cependant, il existe de bonnes raisons d'être sceptique quant à la capacité d'un auditeur à vérifier le fonctionnement d'un système d'apprentissage automatique.

C'est pourquoi il est judicieux de se concentrer sur les éléments inclus par l'AEPD dans sa liste de contrôle recommandée : il serait plus simple de se concentrer sur les mesures mises en œuvre pour éviter les biais, l'obscurité, le profilage caché, etc., se concentrer

---

[563] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, p.26. Disponible sur : www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (consulté le 15 mai 2020).

[564] Groupe d'experts de haut niveau sur l'IA (2019) Lignes directrices en matière d'éthique pour une IA digne de confiance. Commission européenne, Bruxelles, p.22. Disponible à l'adresse : https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (consulté le 15 mai 2020).

sur la mise en œuvre de principes tels que la protection des données dès la conception et par défaut (voir "Protection des données dès la conception et par défaut" dans la partie II, section "Concepts principaux" des présentes lignes directrices) ou la minimisation des données (voir "Principe de minimisation des données" dans la partie II, section "Principes" des présentes lignes directrices et l'utilisation adéquate d'outils tels que le AIPD ou l'intervention d'un DPD compétent, plutôt que d'essayer de comprendre en profondeur le fonctionnement d'un algorithme complexe (le problème de la "boîte noire" est évidemment très important à cet égard). La mise en œuvre de politiques de protection des données adéquates dès les premières étapes du cycle de vie de l'outil est le meilleur moyen d'éviter les problèmes de protection des données.

Encadré 20 : La difficulté d'auditer un système d'apprentissage automatique : La plateforme Watson d'IBM

La politique d'IBM souligne que Watson est formé par "apprentissage supervisé". En d'autres termes, le système est guidé, étape par étape, dans son apprentissage. Cela devrait signifier que son processus peut être surveillé, contrairement à l'apprentissage non supervisé, dans lequel la machine est totalement autonome pour déterminer ses critères de fonctionnement. IBM affirme également pouvoir vérifier ce que les systèmes ont fait, avant toute décision de retenir un certain type d'apprentissage. Mais les spécialistes de la recherche sur ce sujet qui se sont exprimés lors des différents débats organisés (notamment par le comité de recherche en éthique d'Allistene, le CERNA) ont insisté à maintes reprises sur le fait que ces affirmations sont erronées. Selon les recherches actuelles, les "résultats" produits par les algorithmes d'apprentissage automatique les plus récents ne sont pas explicables, l'IA explicable étant un concept sur lequel les recherches se poursuivent. Ils soulignent également qu'il est très difficile de vérifier un système d'apprentissage automatique dans la pratique.[565]

6 Déploiement

"Le déploiement est le processus qui consiste à rendre un système informatique opérationnel dans son environnement, y compris l'installation, la configuration, l'exécution, les tests et les modifications nécessaires. Le déploiement n'est généralement pas effectué par les développeurs d'un système mais par l'équipe informatique du client. Néanmoins, même si c'est le cas, les développeurs auront la responsabilité de fournir au client des informations suffisantes pour un déploiement réussi du modèle. Cela comprendra normalement un plan de déploiement (générique), avec les étapes nécessaires pour un déploiement réussi et la manière de les réaliser, et un plan de surveillance et de maintenance (générique) pour la maintenance du système, et pour la

[565] CNIL (2017) Comment l'humain peut-il garder la main ? Les questions éthiques soulevées par les algorithmes et l'intelligence artificielle. Commission nationale de l'informatique et des libertés, Paris, p.28. Disponible sur : www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf (consulté le 15 mai 2020).

surveillance du déploiement et de l'utilisation correcte des résultats de l'exploration de données."[566]

6.1 Principales mesures à prendre

6.1.1 Remarques générales

Au moment de la distribution de l'outil IA, s'il incorpore des données à caractère personnel, il sera nécessaire de procéder comme suit (voir également la section "Acheter ou promouvoir l'accès à une base de données" dans les "Principaux outils et actions" de la partie II) :

- Les supprimer ou, au contraire, justifier l'impossibilité de le faire, en tout ou partie, en raison de la dégradation que cela impliquerait pour le modèle (voir la section "Limitation du stockage" dans les "Principes" de la partie II).
- Déterminer la base juridique de la communication de données à caractère personnel à des tiers, en particulier si des catégories spéciales de données sont concernées (voir la sous-section "Licéité" de la section "Licéité, loyauté et transparence" de la partie II des présentes lignes directrices).
- Informer les personnes concernées du traitement ci-dessus.
- Démontrer que les politiques de protection des données dès la conception et par défaut ont été mises en œuvre (notamment la minimisation des données).
- En fonction des risques qu'elle pourrait présenter pour les parties prenantes et compte tenu du volume ou des catégories de données à caractère personnel à utiliser, envisager la réalisation d'une analyse d'impact sur la protection des données (AIPD).[567] (Voir "AIPD" dans la partie II, section "Principales actions et outils" des présentes lignes directrices).

En principe, une fois que le modèle est mis en service, les données d'apprentissage sont retirées de l'algorithme et le modèle ne traite que les données à caractère personnel auxquelles il est appliqué. Le responsable du traitement pourrait conserver les données de la personne concernée pour la personnalisation du service offert par l'outil d'IA. Toutefois, une fois ce service terminé, ces données doivent être supprimées, sauf si des raisons convaincantes font qu'il est recommandé de les conserver. Bien entendu, cela ne signifie pas que les données doivent être conservées éternellement.

Le développeur de l'IA doit s'assurer que l'algorithme n'inclut pas de données personnelles de manière cachée (ou prendre les mesures nécessaires si cela est inévitable). Dans tous les cas, le développeur doit effectuer une évaluation formelle évaluant quelles données personnelles des personnes concernées pourraient être

---

[566] Projet SHERPA (2020) Lignes directrices pour le développement éthique des systèmes d'IA et de big data : une approche éthique par la conception. SHERPA, p. 13. Disponible à l'adresse : www.project-sherpa.eu/wp-content/uploads/2019/12/development-final.pdf (consulté le 15 mai 2020).

[567] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, p.26. Disponible sur : www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (consulté le 15 mai 2020).

identifiables.[568] Cela peut parfois être compliqué. Par exemple, certains outils d'IA, tels que les machines à support vectoriel (VSM), pourraient contenir des exemples de données d'entraînement par conception dans la logique du modèle. Dans d'autres cas, des modèles peuvent être trouvés dans le modèle qui identifient un individu unique.[569] Dans tous ces cas, des parties non autorisées peuvent être en mesure de récupérer des éléments des données d'entraînement ou de déduire qui s'y trouvait, en analysant la façon dont le modèle se comporte.

Dans ces conditions, il pourrait être difficile de s'assurer que les personnes concernées sont en mesure d'exercer et de respecter leurs droits d'accès, de rectification et d'effacement (voir les sections "Droit d'accès, de rectification et d'effacement" dans la section "Droits de la personne concernée" de la partie II des présentes lignes directrices). En effet, "à moins que la personne concernée ne présente des preuves que ses données personnelles pourraient être déduites du modèle, le responsable du traitement peut ne pas être en mesure de déterminer si des données personnelles peuvent être déduites et donc si la demande est fondée."[570] Toutefois, les responsables du traitement devraient prendre des mesures régulières pour évaluer de manière proactive la probabilité de la possibilité d'inférer des données à caractère personnel à partir de modèles à la lumière de l'état de la technologie, de sorte que le risque de divulgation accidentelle soit minimisé. Si ces actions révèlent une possibilité substantielle de divulgation des données, les mesures nécessaires pour l'éviter doivent être mises en œuvre (voir la section "Intégrité et confidentialité" dans les "Principes", partie II des présentes lignes directrices).

6.1.2 Mise à jour des informations

Si l'algorithme est mis en œuvre par un tiers, les développeurs de l'IA devraient communiquer les résultats du système de validation et de suivi employé et proposer leur collaboration pour continuer à suivre la validation des résultats. Il serait également souhaitable d'établir ce type de coordination avec les tiers auprès desquels ils acquièrent des bases de données ou tout autre composant pertinent dans le cycle de vie du système. Si cela implique le traitement de données par un tiers, le responsable du traitement doit s'assurer que l'accès est fourni dans le cadre d'une base légale.

Il est nécessaire d'offrir à l'utilisateur final des informations en temps réel sur les valeurs de précision et/ou de qualité des informations inférées à chaque étape (voir la section "Principe de précision" dans les "Principes" de la partie II). Lorsque les informations inférées n'atteignent pas les seuils de qualité minimum, il doit être explicitement indiqué

---

[568] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, p.41. Disponible sur : www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (consulté le 15 mai 2020).

[569] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, p.13. Disponible sur : www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (consulté le 15 mai 2020).

[570] ICO (2019) Permettre les droits d'accès, d'effacement et de rectification dans les outils d'IA. Bureau du commissaire à l'information, Wilmslow. Disponible à l'adresse : https://ico.org.uk/about-the-ico/news-and-events/ai-blog-enabling-access-erasure-and-rectification-rights-in-ai-systems/ (consulté le 15 mai 2020).

que ces informations n'ont aucune valeur.[571] Cette exigence implique souvent que les développeurs doivent fournir des informations détaillées sur les étapes de formation et de validation. Les informations sur les jeux de données utilisés à ces fins sont particulièrement importantes. Dans le cas contraire, l'utilisation de la solution risque d'apporter des résultats décevants aux utilisateurs finaux, qui se retrouvent à spéculer sur la cause.

[571] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, p.35. Disponible sur : www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (consulté le 15 mai 2020).

Annexe I : Outils d'audit IA[572]

Selon l'Agence espagnole de protection des données, les audits doivent couvrir une large liste d'éléments, à savoir :

- L'existence ou non de données personnelles, de profilage ou de décisions automatiques sur les personnes concernées sans intervention humaine.
- L'efficacité des méthodes d'anonymisation et de pseudonymisation.
- L'existence et la légitimité du traitement de catégories spéciales de données, notamment les informations déduites.
- La base juridique du traitement et l'identification des responsabilités.
- En particulier, lorsque la base juridique est l'intérêt légitime, l'évaluation de l'équilibre entre les différents intérêts et les impacts sur les droits et libertés des personnes concernées à la lumière des garanties adoptées.
- L'information et l'efficacité des mécanismes de transparence mis en œuvre.
- L'application du principe de responsabilité proactive et de gestion des risques pour les droits et libertés des personnes concernées et, en particulier, l'obligation ou la nécessité d'effectuer des AIPD et, le cas échéant, leurs résultats.
- L'application de mesures de protection des données dès la conception et par défaut, telles que :
  - l'analyse de la nécessité de la quantité et de l'extension du traitement des données personnelles aux différentes étapes du développement de l'IA ;
  - l'analyse de la précision, de la fiabilité, de la qualité et des biais des données utilisées ou saisies pour le développement ou le fonctionnement de la composante IA, ainsi que les méthodes de nettoyage des données utilisées ;
  - le suivi et la mise en œuvre de processus de test et de validation concernant la précision, l'exactitude, la convergence, la cohérence, la prévisibilité et toute autre mesure de la qualité des algorithmes utilisés, du profilage et des déductions effectuées. En outre, il faut vérifier que ces paramètres répondent aux exigences du traitement.
- L'adéquation des mesures de sécurité pour éviter les risques pour la vie privée.
- La formation et l'éducation du personnel du responsable du traitement liées au développement ou à la mise en œuvre de la composante IAI, le cas échéant, dans ce dernier cas avec une attention particulière à l'interprétation correcte des déductions.

[572] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, p. 45-47. Disponible sur : www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (consulté le 3 juin 2020).

Annexe II. Recherche sur l'apprentissage automatique et l'intelligence artificielle au profit des patients : 20 questions critiques sur la transparence, la reproductibilité, l'éthique et l'efficacité.[573]

Inception

1. Quelle est la question de santé relative au bénéfice du patient ?

2. Quelles sont les preuves que le développement de l'algorithme s'est appuyé sur les bonnes pratiques en matière de recherche clinique et de conception d'études épidémiologiques ?

Étude

1. Quand et comment les patients doivent-ils être impliqués dans la collecte, l'analyse, le déploiement et l'utilisation des données ?

2. Les données sont-elles appropriées pour répondre à la question clinique - c'est-à-dire, capturent-elles l'hétérogénéité pertinente du monde réel, et sont-elles suffisamment détaillées et de qualité ?

3. La méthodologie de validation reflète-t-elle les contraintes du monde réel et les procédures opérationnelles associées à la collecte et au stockage des données ?

4. Quelles sont les ressources informatiques et logicielles requises pour cette tâche, et les ressources disponibles sont-elles suffisantes pour résoudre ce problème ?

Méthodes statistiques

1. Les mesures de performance rapportées sont-elles pertinentes pour le contexte clinique dans lequel le modèle sera utilisé ?

2. L'algorithme ML/AI est-il comparé à la meilleure technologie actuelle et à d'autres bases de référence appropriées ?

3. Le gain rapporté en performance statistique avec l'algorithme ML/AI est-il justifié dans le contexte de tout compromis ?

Reproductibilité

1. Sur quelle base les données sont-elles accessibles aux autres chercheurs ?

2. Le code, le logiciel et toutes les autres parties pertinentes du pipeline de modélisation des prédictions sont-ils disponibles pour les autres afin de faciliter la reproductibilité ?

3. Y a-t-il une transparence organisationnelle concernant le flux des données et des résultats ?

Évaluation de l'impact

1. Les résultats sont-ils généralisables à d'autres contextes que celui où le système a été développé (c'est-à-dire la reproductibilité des résultats/la validité externe) ?

2. Le modèle crée-t-il ou exacerbe-t-il des inégalités dans les soins de santé en fonction de

[573] Vollmer, S. et al. (2020) " Machine learning and artificial intelligence research for patient benefit : 20 critical questions on transparency, replicability, ethics, and effectiveness ", *BMJ* 2020;368:l6927, http://dx.doi.org/10.1136/bmj.l6927.

l'âge, du sexe, de l'origine ethnique ou d'autres caractéristiques protégées ?

3. Quelles preuves existe-t-il que les cliniciens et les patients trouvent le modèle et ses résultats (raisonnablement) interprétables ?

4. Comment les preuves de l'efficacité du modèle en situation réelle dans le cadre clinique proposé seront-elles générées, et comment les conséquences involontaires seront-elles évitées ?

Mise en œuvre

1. Comment le modèle est-il régulièrement réévalué et mis à jour en fonction de l'évolution de la qualité des données et des pratiques cliniques (c'est-à-dire le suivi post-déploiement) ?

2. La construction, la mise en œuvre et la maintenance du modèle ML/AI sont-elles rentables ?

3. Comment les bénéfices financiers potentiels seront-ils distribués si le modèle ML/AI est commercialisé ?

4. Comment les exigences réglementaires en matière d'accréditation/approbation ont-elles été traitées ?

Annexe III : Listes de contrôle

Liste de contrôle : compréhension de l'entreprise

 Les responsables du traitement ont évalué la quantité de données qui seront nécessaires pour développer la solution d'IA ou la nature de ces données et se sont assurés qu'elles fonctionnent bien avec le principe de minimisation.

 Les responsables du traitement ont fixé des seuils acceptables de faux positifs/négatifs ou des fourchettes, selon le cas d'utilisation, puis ont effectué un bilan d'utilité.

 Les responsables du traitement ont équilibré de manière adéquate le niveau de précision nécessaire et l'éventail de données à caractère personnel requis pour l'atteindre.

 Les responsables du traitement ont prévu le développement d'algorithmes plus compréhensibles que ceux qui le sont moins, chaque fois que cela est possible.

 Les responsables du traitement ont assuré une formation optimale pour tous les sujets impliqués dans le projet ou une évaluation interne ou externe adéquate sur les questions éthiques et juridiques.

 Les responsables du traitement ont conçu avec soin les outils qui permettront de légitimer le traitement des données. À cette fin, ils ont vérifié si l'intervention d'un comité d'éthique est nécessaire ou si un type de réglementation douce est applicable.

 Les responsables du traitement ont adopté une approche fondée sur les risques (y compris les mesures de sécurité techniques et organisationnelles) qui minimise les risques pour les droits, intérêts et libertés des personnes concernées.

 Les responsables du traitement ont mis en place des outils et des politiques visant à

évaluer et à apprécier régulièrement l'efficacité des mesures techniques et organisationnelles.

 Les responsables du traitement ont examiné si le cadre réglementaire concernant la recherche scientifique s'applique.

 Les politiques de stockage conservent les données à caractère personnel sous une forme permettant l'identification des personnes concernées pendant une durée n'excédant pas celle nécessaire aux finalités pour lesquelles les données à caractère personnel sont traitées.

 Les responsables du traitement ont envisagé la désignation d'un DPD.

Liste de contrôle : compréhension des données

 Les responsables du traitement ont mis en œuvre des mesures techniques et organisationnelles appropriées pour garantir que, par défaut, seules les données à caractère personnel qui sont nécessaires pour chaque finalité spécifique du traitement sont traitées.

 Les responsables du traitement ont mis en place des politiques qui minimisent la quantité de données personnelles collectées, l'étendue de leur traitement, la période de leur stockage et leur accessibilité. Ces mesures garantissent que, par défaut, les données à caractère personnel ne sont pas rendues accessibles sans l'intervention de l'individu à un nombre indéfini de personnes physiques.

 Les responsables du traitement ne collectent pas de données inutiles. Si des données sont déjà stockées, ils ont pris des mesures visant à supprimer les éléments de données inutiles.

 Les responsables du traitement ont limité la résolution des données à ce qui est minimalement nécessaire aux fins poursuivies par le traitement.

 Les responsables du traitement ont choisi la base juridique qui reflète le mieux la véritable nature de leur relation avec la personne et la finalité du traitement.

 Les responsables du traitement ont soigneusement analysé si le traitement implique la désanonymisation de données anonymisées et la création de nouvelles informations personnelles qui n'étaient pas contenues dans l'ensemble de données d'origine et prennent des mesures adéquates pour faire face à ces défis.

 Les responsables du traitement se sont assurés que la fusion des ensembles de données ne crée pas de problèmes éthiques ou juridiques concernant les droits et libertés des personnes concernées.

Liste de contrôle : Préparation des données

 Les responsables du traitement se sont assurés que les données sont précises, c'est-à-dire qu'elles sont correctes et à jour.

 Si un profilage ou une prise de décision automatisée est prévu :

 Les responsables du traitement ont envoyé aux personnes un lien vers leur déclaration de confidentialité lorsqu'ils ont obtenu leurs données personnelles de manière indirecte.

 Les responsables du traitement ont expliqué comment les personnes peuvent accéder aux détails des informations qu'elles ont utilisées pour créer leur profil.

 Les responsables du traitement ont communiqué aux personnes concernées qui leur fournissent leurs données à caractère personnel et la manière dont elles peuvent s'opposer au profilage.

 Les responsables du traitement ont mis en place des procédures permettant aux clients d'accéder aux données personnelles saisies dans leurs profils, afin qu'ils puissent les examiner et les modifier en cas de problème de précision.

 Les responsables du traitement ont mis en place des contrôles supplémentaires pour leurs systèmes de profilage/décision automatisée afin de protéger tout groupe vulnérable (y compris les enfants).

 Les responsables du traitement se sont assurés de ne collecter que le minimum de données nécessaires et d'avoir une politique de conservation claire pour les profils qu'ils créent.

 Les responsables du traitement ont réalisé une AIPD pour examiner et traiter les risques lorsqu'ils commencent toute nouvelle prise de décision ou tout nouveau profilage automatisé.

 Les responsables du traitement ont associé le DPD correspondant à ces activités.

 Les responsables du traitement ont pris en compte les exigences du système nécessaires pour soutenir une révision humaine significative **dès la phase de conception.** En particulier, les exigences d'interprétabilité et la conception efficace de l'interface utilisateur pour soutenir les examens et les interventions humaines.

 Les responsables du traitement ont conçu et dispensé une formation et un soutien appropriés aux réviseurs humains.

 Les responsables du traitement ont donné au personnel impliqué dans le traitement l'autorité, les incitations et le soutien appropriés pour traiter ou faire remonter les préoccupations des personnes et, si nécessaire, passer outre la décision du système d'IA.

 Les responsables du traitement ont veillé à ce que les équipes chargées de sélectionner les données à intégrer dans les jeux de données soient composées de personnes assurant la diversité dont le développement de l'IA est censé faire preuve.

 Les responsables du traitement ont veillé à ce que les facteurs qui entraînent des inexactitudes dans les données à caractère personnel soient corrigés et que le risque d'erreurs soit réduit au minimum.

 Les responsables de traitement ont mis en place des outils visant à prévenir les effets discriminatoires à l'égard des personnes physiques sur la base de l'origine raciale ou

ethnique, des opinions politiques, de la religion ou des convictions, de l'appartenance syndicale, du statut génétique ou de santé ou de l'orientation sexuelle, ou qui aboutissent à des mesures ayant un tel effet.

Liste de contrôle : Modélisation (formation)

 Les responsables du traitement ont déterminé l'objectif de l'utilisation du système d'IA dès le début de sa formation ou de son déploiement, et ont procédé à une réévaluation de cette détermination si le traitement du système donnait des résultats inattendus.

 Les responsables du traitement ont purgé les données utilisées lors de la phase d'entraînement de toutes les informations non strictement nécessaires à l'entraînement du modèle.

 Les responsables du traitement ont envisagé de mettre en place des outils techniques qui pourraient bien servir à détecter les biais, comme l'Algorithmic Impact Assessment.

 Les responsables du traitement ont envisagé de réaliser une AIPD à ce stade.

 Les responsables de traitement se sont assurés qu'ils sont en mesure de répondre aux demandes des personnes concernées pour que les exceptions au droit d'accès s'appliquent.

 Les responsables du traitement peuvent garantir le droit de rectification des données, notamment celles générées par les inférences et les profils établis par le développement de l'IA.

 Les responsables du traitement sont en mesure de répondre aux demandes d'effacement, sauf si une exemption pertinente s'applique et à condition que la personne concernée ait des motifs appropriés.

Liste de contrôle : évaluation (validation)

 Les responsables du traitement se sont assurés que la validation reflète fidèlement les conditions dans lesquelles l'algorithme a été validé.

 Les responsables du traitement ont informé les personnes concernées des traitements supplémentaires à ce stade.

 Les responsables du traitement ont veillé à la suppression de l'ensemble des données utilisées à des fins de validation, sauf s'il existe un besoin légitime de les conserver pour affiner ou évaluer le système, ou pour d'autres fins compatibles avec celles pour lesquelles elles ont été collectées.

 Les responsables du traitement ont envisagé de réaliser une AIPD à ce stade.

 Si les personnes concernées demandent la suppression de leurs données, le responsable du traitement a adopté une approche au cas par cas en tenant compte des éventuelles limitations à ce droit prévues par le règlement.

 Les responsables du traitement ont envisagé un audit du système par un tiers indépendant.

Liste de contrôle : déploiement

 Les responsables du traitement ont supprimé toutes les données personnelles inutiles ou, au contraire, ont justifié l'impossibilité de le faire.

 Les responsables du traitement ont informé les personnes concernées des traitements supplémentaires à ce stade.

 Les responsables du traitement ont déterminé la base juridique adéquate pour effectuer la communication de données à caractère personnel à des tiers, en particulier si des catégories

spéciales de données sont concernées.

 Les responsables du traitement ont envisagé de réaliser une AIPD.

 Les responsables du traitement se sont assurés que l'algorithme n'inclut pas de données personnelles de manière cachée (ou ont pris les mesures nécessaires si cela est inévitable).

 Les développeurs de l'IA ont mis en place des outils visant à communiquer les résultats du système de validation et de suivi employé et ont proposé leur collaboration pour continuer.

☐ Les développeurs de l'IA s'engagent à offrir aux utilisateurs finaux des informations en temps réel sur les valeurs de précision et/ou de qualité des informations déduites à chaque étape.

Premier scénario : construction d'un outil d'IA dédié au diagnostic de la maladie COVID-19

Iñigo de Miguel Beriain (UPV/EHU)

Cette partie des lignes directrices a été revue et validée par Marko Sijan, conseiller principal spécialiste (DPA RH).

Description

La réponse à la pandémie a créé des situations dans lesquelles de nombreux patients avaient besoin de soins de santé mais ceux-ci étaient difficiles à fournir en raison de la forte incidence de la maladie parmi le personnel de santé. Dans cette situation, un radiologue, par exemple, ne pouvait pas faire face au grand nombre de radiographies à analyser en raison de l'absence de ses collègues en congé de maladie. L'utilisation de l'IA à de telles fins pourrait être d'une grande aide pour l'avenir, mais de nombreuses questions éthiques et juridiques doivent être prises en compte. Dans ce scénario, nous analyserons les différentes étapes que doit franchir une équipe de chercheurs désireuse de former un algorithme capable d'aider au diagnostic des maladies pulmonaires.

Remarques préliminaires

La recherche sur les données de santé présente des défis éthiques particulièrement importants. Si nous parlons également d'un cas où les patients souffrent d'une maladie telle que la COVID, le dilemme est particulièrement pressant. Dans le contexte des soins de santé, il est facile de mélanger le consentement éclairé associé à la pratique clinique et le consentement à la recherche biomédicale. C'est toujours un sujet de préoccupation. Les deux choses sont extrêmement différentes. La planification d'une activité telle que le développement d'un outil d'IA pour le diagnostic doit en tenir compte. Cela est particulièrement vrai pour les patients qui se trouvent dans des situations plus vulnérables que d'habitude. Il ne faut jamais oublier que les objectifs de la recherche biomédicale ne peuvent empiéter sur les intérêts et le bien-être des personnes.

Il existe plusieurs outils essentiels que les chercheurs doivent toujours garder à l'esprit lorsqu'ils élaborent un plan pour le développement d'un outil d'IA. La liste de contrôle des questions éthiques incluse dans le guide du programme Horizon 2020 "Comment réaliser votre auto-évaluation éthique", page 6[574] est fortement recommandée. Parmi les documents essentiels à consulter figurent :

- Groupe d'experts de haut niveau sur l'IA : "Lignes directrices éthiques pour une IA digne de confiance".[575]
- Commission européenne, Livre blanc sur l'intelligence artificielle - Une approche européenne de l'excellence et de la confiance[576]
- Formation et ressources pour l'évaluation de l'éthique de la recherche (TRREE)[577] est un outil en ligne qui permet d'accéder gratuitement à des informations :
    - **e-Learning :** un programme de formation à distance et une certification sur l'évaluation de l'éthique de la recherche,
    - **e-Ressources :** un site web participatif contenant des ressources internationales, régionales et nationales en matière de réglementation et de politique.
- D'autres outils de formation en ligne sont disponibles sur la page web EUREC[578] .

[574] ttps://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-self-assess_fr.pdf

[575] https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[576] https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

[577] https://elearning.trree.org/mod/page/view.php?id=70

[578] http://www.eurecnet.org/materials/index.html

Analyse étape par étape

1 Compréhension de l'entreprise

1.1 Description

"La phase initiale de compréhension de l'entreprise se concentre sur la compréhension des objectifs du projet d'un point de vue commercial, en convertissant cette connaissance en une définition du problème d'exploration de données, puis en développant un plan préliminaire conçu pour atteindre les objectifs. Afin de comprendre quelles données doivent être analysées plus tard, et comment, il est vital pour les praticiens de l'exploration de données de comprendre pleinement l'entreprise pour laquelle ils trouvent une solution. La phase de compréhension de l'entreprise comprend plusieurs étapes clés, notamment la détermination des objectifs de l'entreprise, l'évaluation de la situation, la détermination des objectifs de l'exploration de données et la production du plan de projet." [579]

Cet objectif général implique quatre tâches principales :

1. Déterminer les objectifs de l'entreprise. Cela signifie :
    a. Découvrir l'objectif principal de l'entreprise ainsi que les questions connexes auxquelles l'entreprise souhaite répondre.
    b. Déterminer la mesure du succès.
2. Évaluer la situation
    a. Identifier les ressources disponibles pour le projet, tant matérielles que personnelles.
    b. Identifier les données disponibles pour atteindre l'objectif principal de l'entreprise.
    c. Dresser la liste des hypothèses formulées dans le cadre du projet.
    d. Dresser la liste des risques du projet, énumérer les solutions potentielles à ces risques, créer un glossaire de termes commerciaux et d'exploration de données, et construire une analyse coûts-avantages pour le projet.
3. Déterminer les objectifs de l'extraction de données : décider du niveau de précision prédictive attendu pour considérer le projet comme réussi.
4. Produire un plan de projet : décrire le plan prévu pour atteindre les objectifs de l'exploration de données, y compris la description des étapes spécifiques et un calendrier proposé, une évaluation des risques potentiels, et une évaluation initiale des outils et des techniques nécessaires pour soutenir le projet.

[579] Shearer, Colin, Le modèle CRISP-DM : The New Blueprint for Data Mining, p. 14.

1.2 Principales mesures à prendre

1.2.1 Définir les objectifs de l'entreprise

La première chose à clarifier lorsque vous voulez créer un outil d'IA est ce que vous voulez réaliser. Dans le cas d'un outil qui identifie une pathologie à partir d'une radiographie, il peut s'agir, par exemple, de

1) Il est destiné à servir de support au travail du radiologue.
2) Il peut être utilisé pour soutenir le travail d'un médecin de soins primaires, c'est-à-dire pour déterminer s'il faut adresser le patient à un spécialiste.
3) Il peut également être conçu pour remplacer le médecin et poser seul un diagnostic de COVID, par exemple.
4) Il peut être utilisé pour effectuer un premier triage (c'est-à-dire recommander l'intervention d'un médecin de soins primaires ou d'un spécialiste).

Chacun de ces scénarios présente des caractéristiques très différentes. Certains d'entre eux exigent un niveau de précision plus élevé que d'autres. Ainsi, par exemple, si vous avez l'intention de remplacer le professionnel de santé, il est nécessaire que l'IA atteigne un niveau de précision impressionnant.

Les implications éthiques et juridiques de ces différentes finalités sont, en même temps, très différentes. Si le mécanisme doit être utilisé à des fins de prise de décision automatisée, comme dans les cas 3) ou 4), le traitement des données sera soumis à un régime juridique considérablement plus strict. En fait, dans de nombreux pays, cette utilisation peut être directement illégale.

Toutes ces considérations doivent être gardées à l'esprit dès le départ. Le processus de développement ne doit pas être lancé si vous, en tant que responsable du traitement, ne clarifiez pas les résultats à atteindre, car cette question est essentielle pour déterminer si le traitement des données prévu est conforme ou non au RGPD. Décider du niveau de précision prédictive attendu pour considérer le projet comme réussi est essentiel pour évaluer la quantité de données qui sera nécessaire pour développer l'outil d'IA ou la nature de ces données. Le niveau de prévisibilité ou de précision de l'algorithme, les critères de validation pour le tester, la quantité maximale ou la qualité minimale des données personnelles qui seront nécessaires pour l'utiliser dans le monde réel, etc. sont des caractéristiques fondamentales d'un développement d'IA.

Ces éléments clés du développement doivent être pris en compte dès la première étape du cycle de vie de la solution. Cela sera extrêmement utile pour mettre en œuvre une politique de protection des données dès la conception (voir "Protection des données dès la conception et par défaut" dans la partie II, section "Concepts principaux" des présentes lignes directrices). S'il est possible d'atteindre un niveau de précision acceptable en utilisant une quantité de données à caractère personnel nettement inférieure à ce qu'exige un niveau plus élevé, il convient d'y réfléchir sérieusement. Plus ces évaluations sont imprécises, plus il devient difficile de déterminer les finalités précises poursuivies par le traitement (voir la sous-section "Conditions préalables à la licéité - finalités spécifiques et explicites" dans la section "Licéité, loyauté et transparence" de la partie II "Principes"). Si l'on garde à l'esprit que les responsables du

traitement doivent rendre les finalités du traitement explicites, c'est-à-dire "révélées, expliquées ou exprimées d'une manière intelligible", il est fortement recommandé d'avoir des attentes précises.

1.2.2 Opter pour les solutions techniques

En général, il faut toujours prévoir le développement d'algorithmes plus compréhensibles que d'algorithmes moins compréhensibles. Les compromis entre l'explicabilité/la transparence et les meilleures performances du système doivent être équilibrés de manière appropriée en fonction du contexte d'utilisation. Même si, dans le domaine des soins de santé, la précision et les performances du système peuvent être plus importantes que sa facilité d'explication, vous devez toujours garder à l'esprit que l'explication d'une recommandation peut être utile pour former les médecins, fournir des informations adéquates aux patients qui doivent faire un choix entre différents traitements possibles ou justifier une décision de triage, par exemple. Ainsi, si un service tout à fait similaire peut être offert soit par un algorithme facile à comprendre, soit par un algorithme opaque, c'est-à-dire lorsqu'il n'y a pas de compromis entre l'explicabilité et la performance, vous devez opter pour celui qui est le plus interprétable (voir la section "Licéité, loyauté et transparence" dans "Principes" de la partie II).

1.2.3 Mettre en place un programme de formation sur les questions éthiques et juridiques

Cette action est l'un des conseils les plus importants à prendre en compte dès le premier moment du développement commercial de l'IA. Les concepteurs d'algorithmes (développeurs, programmeurs, codeurs, data scientists, ingénieurs), qui occupent le premier maillon de la chaîne algorithmique, sont susceptibles de ne pas avoir conscience des implications éthiques et juridiques de leurs actions. Si tout le personnel intervenant est en contact étroit avec les personnes concernées, les considérations éthiques sont plus faciles à mettre en œuvre. Cependant, ce ne sera probablement pas votre cas. En effet, l'un des principaux problèmes que rencontre un outil d'IA consacré au traitement des questions de santé est qu'il utilise généralement des données personnelles incluses dans de grands ensembles de données. Cela brouille en quelque sorte la relation entre les données et la personne concernée, ce qui entraîne des violations de la réglementation qui se produisent rarement lorsque le responsable du traitement et le sujet ont une relation directe.

Cela pourrait avoir des conséquences terribles en termes de respect adéquat des normes de protection des données, notamment parce que des données de catégories spéciales sont en jeu. Il est primordial que ces travailleurs clés aient la plus grande conscience possible des implications éthiques et sociales de leur travail, et du fait même que celles-ci peuvent aller jusqu'à des choix de société, qu'ils ne devraient pas, de droit, pouvoir juger seuls. La mentalité de silo doit être soigneusement combattue.

Afin d'éviter que la mauvaise représentation des questions éthiques et juridiques ne provoque des conséquences indésirables, deux grandes lignes d'action peuvent être adoptées. Tout d'abord, les développeurs peuvent essayer de faire en sorte que les concepteurs d'algorithmes soient en mesure de comprendre les implications de leurs actions, tant pour les individus que pour la société, et qu'ils soient conscients de leurs responsabilités en apprenant à faire preuve d'une attention et d'une vigilance

constantes.[580] Dans ce sens, une formation optimale de tous les sujets impliqués dans le projet (développeurs, programmeurs, codeurs, data scientists, ingénieurs, chercheurs) avant même qu'il ne commence pourrait être l'un des outils les plus efficaces pour économiser du temps et des ressources en termes de conformité avec la réglementation sur la protection des données. Ainsi, la mise en œuvre de programmes de formation de base qui incluent au moins les principes fondamentaux de la Charte des droits fondamentaux, les principes exposés à l'article 5 du RGPD, la nécessité d'une base légale pour le traitement (y compris les contrats entre les parties), etc.

Cependant, il peut être difficile de former des personnes qui n'ont jamais été en contact avec les questions de protection des données. Une autre solution consiste à impliquer un expert de la protection des données et des questions éthiques et juridiques dans l'équipe de développement, de manière à créer une équipe interdisciplinaire. Pour ce faire, on peut engager un expert à cette fin (un travailleur interne ou un consultant externe) pour concevoir la stratégie et les décisions ultérieures sur les données personnelles requises par le développement des outils, avec la participation étroite du délégué à la protection des données.

Il est également fortement recommandé d'adopter des mesures adéquates pour garantir la confidentialité (voir les sous-sections "Mesures en faveur de la confidentialité" de la section "Intégrité et confidentialité" dans les "Principes" de la partie II des présentes lignes directrices).

1.2.4 Conception d'outils de traitement des données légitimes

Selon l'article 5, paragraphe 1, point a), du RGPD, les données à caractère personnel sont "collectées pour des finalités spécifiques, explicites et légitimes et ne sont pas traitées ultérieurement de manière incompatible avec ces finalités". Le concept de légitimité n'est pas bien défini dans le RGPD, mais le groupe de travail Article 29 a déclaré que la légitimité implique que les données doivent être traitées "conformément à la loi", et que la "loi" doit être comprise comme un concept large qui inclut "toutes les formes de droit écrit et de common law, la législation primaire et secondaire, les décrets municipaux, les précédents judiciaires, les principes constitutionnels, les droits fondamentaux, les autres principes juridiques, ainsi que la jurisprudence, telle que cette "loi" serait interprétée et prise en compte par le tribunal compétent".

Il s'agit donc d'un concept plus large que la licéité. Il implique le respect des principales valeurs de la réglementation applicable et des grands principes éthiques en jeu. Par exemple, certains développements concrets de l'IA nécessiteront l'intervention d'un comité d'éthique. Dans d'autres cas, des lignes directrices ou tout autre type de réglementation non contraignante peuvent être applicables. Vous devez vous assurer de la conformité à cette exigence en élaborant un plan pour cette étape préliminaire du cycle de vie de l'outil (voir "Légitimité et licéité" dans "Licéité, loyauté et transparence" des "Principes" de la partie II). À cette fin, vous devez être particulièrement attentif aux exigences posées par la réglementation applicable au niveau national. Dans de nombreux États membres, le développement d'un algorithme lié aux soins de santé impliquera certainement l'intervention de comités d'éthique, très probablement à un stade préliminaire. Assurez-vous que votre plan de recherche répond bien à ces exigences.

---

[580] Ibid. p. 55.

1.2.5 Adopter une approche de réflexion fondée sur le risque

Étant donné que la création de votre algorithme impliquera certainement l'utilisation d'une quantité énorme de catégories spéciales de données personnelles, principalement des données relatives à la santé, vous devez vous assurer que vous mettez en œuvre des mesures appropriées pour minimiser les risques pour les droits et libertés des personnes concernées (voir "Intégrité et confidentialité" des "Principes" dans la partie II). À cette fin, vous devez évaluer les risques pour les droits et libertés des personnes participant au processus de recherche et de développement et juger ce qui est approprié pour les protéger. Dans tous les cas, vous devez vous assurer qu'ils sont conformes aux exigences en matière de protection des données.

Une réflexion fondée sur le risque en ce qui concerne la confidentialité des données, ou une approche fondée sur le risque des questions relatives aux préjudices qui peuvent être causés aux personnes/aux personnes concernées, doit être incluse dès les premières étapes du processus. Elle pourrait avoir des conséquences juridiques pour le responsable du traitement des données par rapport aux obligations stipulées dans le RGPD si elle n'est prise en compte que plus tard. Ainsi, vous devez identifier les menaces implicites qui pèsent sur le traitement des données prévu et évaluer le niveau de risque intrinsèque qu'il comporte. Si vous prévoyez d'utiliser un logiciel à des fins de traitement, vous devez vous assurer que des mesures adéquates à l'appui de la confidentialité sont mises en œuvre. Si votre IA doit utiliser un logiciel tiers ou un logiciel standard, il est essentiel d'exclure les fonctions de traitement des données personnelles qui n'ont pas de base juridique ou qui ne sont pas compatibles avec les finalités visées.

Dans la mesure du possible, essayez d'éviter d'utiliser des services de stockage de données ou de logiciels qui sont situés dans un pays tiers. Si cela est inévitable, vous devez vous assurer que vos contrats de traitement des données avec ces tiers offrent une protection adéquate conforme au RGPD ou, si ce n'est pas le cas, vous assurer que les participants à la recherche sont pleinement conscients des risques de confidentialité/sécurité pour leurs données. *Vous devez également être conscient et informé des mesures de sécurité appropriées mises en œuvre par les* fournisseurs de services de *stockage de données et de* logiciels, et que les omissions en matière de sécurité peuvent entraîner une violation du traitement sécurisé.

En outre, vous devez vous assurer que des mesures techniques et organisationnelles appropriées sont mises en œuvre pour éliminer, ou au moins atténuer le risque, en réduisant la probabilité que les menaces identifiées se concrétisent ou en réduisant leur impact. Les mesures de sécurité doivent faire partie de vos documents de traitement (voir la section "Documentation du traitement" dans la section "Principaux outils et actions" de la partie II des présentes lignes directrices) et toutes les mesures mises en œuvre feront partie de l'AIPD **(**voir "AIPD" dans la section "Principaux outils et actions" de la partie II des présentes lignes directrices).

Une fois les mesures sélectionnées mises en œuvre, le risque résiduel restant doit être évalué et gardé sous contrôle. L'analyse des risques et l'AIPD sont les outils qui s'appliquent. Dans votre cas concret, vous devez réaliser une AIPD, car la création de l'outil d'IA impliquera le traitement à grande échelle de catégories spéciales de données.

Enfin, il ne faut pas oublier que lorsqu'on utilise le big data et l'IA, il est difficile de prévoir quels seront les risques futurs, de sorte que l'évaluation des implications éthiques ne suffira pas à traiter tous les risques possibles. Il est donc important

d'envisager une réévaluation des risques et il est également fortement recommandé d'intégrer une méthode plus dynamique d'évaluation des risques liés à la recherche. N'hésitez pas à effectuer des AIPD supplémentaires à d'autres étapes du processus si nécessaire.

1.2.6 Préparer la documentation du traitement

Quiconque traite des données à caractère personnel (qu'il s'agisse de responsables du traitement ou de sous-traitants) doit documenter ses activités, principalement à l'intention des autorités de contrôle compétentes. Vous devez le faire par le biais de registres du traitement qui sont conservés de manière centralisée par votre organisation pour l'ensemble de ses activités de traitement, et de documents supplémentaires qui se rapportent à une activité individuelle de traitement des données (voir la section "Documentation du traitement" dans "Principaux outils et actions" de la partie II des présentes lignes directrices). Cette phase préliminaire est le moment idéal pour mettre en place une méthode systématique de collecte de la documentation nécessaire, puisque c'est à ce moment-là que vous pourrez concevoir et planifier l'activité de traitement.

En effet, vous devez créer une politique de protection des données (voir la sous-section "Économie d'échelle pour la conformité et sa démonstration" de la section "Responsabilité" des "Principes" de la partie II) qui permet la traçabilité des informations (s'il existe des codes de conduite approuvés, ceux-ci doivent être mis en œuvre, voir la sous-section "Économie d'échelle pour la conformité et sa démonstration" de la section "Responsabilité" des "Principes" de la partie II). Cette politique doit également préciser les responsabilités attribuées aux sous-traitants, si vous souhaitez les associer à votre projet, et inclure les tâches de l'accord de traitement qui lui seront déléguées en ce qui concerne l'exécution des droits des personnes concernées. Vous devez toujours vous rappeler que l'art. 32(4) du RGPD précise qu'un élément important de la sécurité consiste à s'assurer que les employés n'agissent que sur instruction et selon vos instructions (voir la section "Intégrité et confidentialité" dans "Principes", partie II des présentes lignes directrices).

Le développement de votre outil d'IA peut impliquer l'utilisation de différents ensembles de données. La traçabilité du traitement, les informations sur la réutilisation éventuelle des données et l'utilisation de données appartenant à des ensembles de données différents dans des étapes différentes ou identiques du cycle de vie doivent être garanties par les registres.

Comme indiqué dans la section "Exigences et tests d'acceptation pour l'achat et/ou le développement des logiciels, du matériel et de l'infrastructure utilisés" (sous-section de la section "Documentation du traitement"), l'évaluation des risques et les décisions prises "doivent être documentées afin de respecter l'exigence de protection des données dès la conception (voir "Protection des données dès la conception et par défaut" dans la partie II, section "Concepts principaux" des présentes lignes directrices). En pratique, cela peut prendre la forme de :

- **Exigences** de protection des données spécifiques pour l'achat (par exemple, un appel d'offres) ou le développement de logiciels, de matériel et d'infrastructures,
- **Tests d'acceptation** qui vérifient que les logiciels, les systèmes et l'infrastructure choisis sont adaptés à l'usage prévu et offrent une protection et des garanties adéquates.

Cette documentation peut faire partie intégrante de l'AIPD."

Enfin, vous devez toujours être conscient que, conformément à l'art. 32(1)(d) du RGPD, la protection des données est un processus. Par conséquent, **vous devez tester, évaluer et apprécier régulièrement l'efficacité des mesures techniques et organisationnelles.** C'est le moment idéal pour élaborer une stratégie visant à relever ces défis.

1.2.7 Utilisation du cadre réglementaire

Le RGPD comprend un cadre réglementaire spécifique concernant le traitement à des fins de recherche scientifique (voir la section "Protection des données et recherche scientifique" dans les "Concepts principaux" de la partie II).[581] Le développement de votre IA constitue une recherche scientifique, indépendamment du fait qu'elle soit créée dans un but lucratif ou non. Par conséquent, "le droit de l'Union ou des États membres peut prévoir des dérogations aux droits visés aux articles 15, 16, 18 et 21, sous réserve des conditions et garanties visées au paragraphe 1 du présent article, dans la mesure où ces droits sont susceptibles de rendre impossible ou de nuire gravement à la réalisation des finalités spécifiques, et où ces dérogations sont nécessaires à la réalisation de ces finalités" (article 89, paragraphe 2). En outre, selon l'article 5, point b), "le traitement ultérieur des données recueillies, conformément à l'article 89, paragraphe 1, ne serait pas considéré comme incompatible avec les finalités initiales ("limitation de la finalité"). D'autres exceptions particulières au cadre général applicable au traitement à des fins de recherche (comme la limitation du stockage) devraient également être envisagées".

Vous pouvez certainement bénéficier de ce cadre favorable. Néanmoins, vous devez être conscient du cadre réglementaire concret qui s'applique à cette recherche (principalement, les garanties à mettre en œuvre). Il peut inclure des changements importants en fonction des réglementations nationales respectives. La consultation de votre DPD est fortement recommandée à cet effet.

1.2.8 Définir des politiques adéquates de stockage des données

Conformément à l'article 5, paragraphe 1, point e), du RGPD, les données à caractère personnel doivent être "conservées sous une forme permettant l'identification des personnes concernées pendant une durée n'excédant pas celle nécessaire à la réalisation des finalités pour lesquelles elles sont traitées" (voir la section "Limitation du stockage" des "Principes" de la partie II). Cette exigence est double. D'une part, elle concerne l'identification : les données doivent être conservées sous une forme permettant l'identification des personnes concernées pendant une durée n'excédant pas celle nécessaire. Par conséquent, vous devez mettre en œuvre des politiques visant à éviter l'identification dès qu'elle n'est pas nécessaire au traitement. Cela implique l'adoption de mesures adéquates pour garantir qu'à tout moment, seul **le degré minimal d'identification nécessaire à la réalisation des finalités doit être utilisé** (voir la sous-section "Aspect temporel" dans la section "Limitation du stockage" des "Principes" de la partie II).

[581] Ce cadre spécifique comprend également des objectifs de recherche historique ou des objectifs statistiques. Toutefois, la recherche sur les TIC n'est généralement pas liée à ces objectifs. Par conséquent, nous ne les analyserons pas ici.

D'autre part, la conservation des données implique que les données ne peuvent être stockées que pendant une **période limitée** : le temps strictement nécessaire aux fins pour lesquelles les données sont traitées. Toutefois, le RGPD permet un "stockage pour des périodes plus longues" si la seule finalité est la recherche scientifique (comme dans votre cas concret).

Ainsi, cette exception soulève le risque que vous décidiez de conserver les données plus longtemps que strictement nécessaire afin de garantir qu'elles soient disponibles pour des raisons autres que les finalités initiales pour lesquelles elles ont été collectées. Ne le faites pas, s'il n'y a pas de bonnes raisons qui le recommandent (par exemple, si des radiographies proviennent d'un dossier médical, vous devez les conserver dans le dossier clinique du patient). Vous devez être conscient que même si le RGPD peut autoriser le stockage pour des périodes plus longues, **vous devez avoir une bonne raison d'opter pour une telle période prolongée.** Ainsi, si vous n'avez pas besoin des données, et qu'aucune raison légale obligatoire ne vous oblige à les conserver, il est préférable de les anonymiser ou de les supprimer. Ce pourrait également être le moment idéal pour **envisager des délais d'effacement des différentes catégories de données et documenter ces décisions** (voir "Principe de responsabilité" dans la partie "Principes" de la partie II).

1.2.9 Nomination d'un délégué à la protection des données

Conformément à l'article 37 du RGPD, vous devez désigner un DPD puisque vous allez traiter un grand nombre de catégories spéciales de données conformément à l'article 9. Dans tous les cas, le personnel clé du responsable du traitement doit définir le rôle du DPD par rapport à la gestion globale du projet, en veillant à ce que le rôle du DPD ne soit pas marginal, mais qu'il soit intégré dans les processus décisionnels de l'organisation/du projet. Ils devraient également préciser ce que pourrait être ce rôle en termes de supervision, de prise de décision et autres.

1.2.10 Assurer la conformité avec le cadre juridique des dispositifs médicaux

Même si ces lignes directrices sont principalement orientées vers les questions de protection des données, nous ne pouvons éviter de mentionner que vous devez être bien conscient dès cette étape préliminaire que vous devez assurer une conformité adéquate avec le cadre juridique lié aux dispositifs médicaux. Nous faisons principalement référence au règlement (UE) 2017/745 - Règlement sur les dispositifs médicaux (MDR) et au règlement (UE) 2017/746 - Règlement sur les dispositifs médicaux de diagnostic in vitro (IVDR). Il existera très probablement des réglementations nationales applicables à ces questions. Veuillez prendre des mesures visant à vous mettre en conformité. Vous trouverez des lignes directrices utiles à cet effet ici : https://ec.europa.eu/docsroom/documents/40323

En ce qui concerne la réglementation sur les données relatives à la santé au niveau des États membres, cette ressource pourrait être particulièrement pertinente :

https://ec.europa.eu/health/sites/health/files/ehealth/docs/ms_rules_health-data_en.pdf

2 Compréhension des données

2.1 Description

"La phase de compréhension des données commence par une collecte initiale des données. L'analyste procède ensuite à une familiarisation accrue avec les données, à l'identification des problèmes de qualité des données, à la recherche d'idées générales sur les données, ou à la détection de sous-ensembles intéressants pour former des hypothèses sur des informations cachées. La phase de compréhension des données comporte quatre étapes, à savoir la collecte des données initiales, la description des données, l'exploration des données et la vérification de la qualité des données".[582]

À ce stade, la collecte des données initiales se fait au palais, et une première étude des données est réalisée. Elle comporte quatre tâches séquentielles :

- Collecter les données initiales
- Décrire les données
- Analyser les données
- Vérifier la qualité des données.

Toutes ces tâches ont pour but d'identifier les données disponibles. À ce stade, vous devez être conscient des données avec lesquelles vous aurez à travailler et commencer à prendre des décisions sur la manière dont les grands principes liés à la protection des données seront mis en œuvre.

2.2 Principales mesures à prendre

À ce stade, un très grand nombre de questions fondamentales liées à la protection des données personnelles doivent être abordées. En fonction des décisions prises, des principes tels que la minimisation des données, la protection de la vie privée dès la conception ou par défaut, la licéité, la loyauté et la transparence, etc. seront réglés de manière adéquate.

2.2.1 Type de données collectées

Selon le RGPD, vous "mettez en œuvre les mesures techniques et organisationnelles appropriées pour garantir que, par défaut, seules les données à caractère personnel qui sont nécessaires à chaque finalité spécifique du traitement sont traitées". Cette obligation s'applique à la quantité de données à caractère personnel collectées, à l'étendue de leur traitement, à la durée de leur conservation et à leur accessibilité. En particulier, ces mesures garantissent que, par défaut, les données à caractère personnel ne sont pas rendues accessibles sans l'intervention de la personne concernée à un nombre indéfini de personnes physiques."[583] (Voir "Protection des données dès la conception et par défaut" dans la partie II, section "Concepts principaux") Il faut en tenir compte tout particulièrement à ce stade, car c'est souvent à ce moment-là que sont prises les décisions concernant le type de données qui seront utilisées. En général, la manière la plus simple de construire votre IA en termes de protection des données

---

[582] Colin Shearer, Le modèle CRISP-DM : Le nouveau plan directeur pour l'extraction de données, p. 15

[583] Article 24.

consisterait à utiliser exclusivement des images radiologiques. Néanmoins, il pourrait également être intéressant d'introduire des données relatives à des pathologies antérieures, à l'âge ou au sexe, par exemple. En outre, on pourrait envisager d'utiliser des données telles que les habitudes alimentaires, le code postal, les habitudes sportives, etc. Il se peut que l'ajout d'un grand nombre de nouvelles caractéristiques au modèle augmente sa précision de manière significative. Cependant, il est également possible que cela ne se produise pas**. Vous devez évaluer si l'introduction de données supplémentaires, en dehors des images radiographiques, par exemple, apporte au diagnostic un niveau de précision accru suffisant pour justifier leur utilisation.** Cela peut être difficile à évaluer à l'avance, mais au moins la phase de formation devrait clarifier cette question. Si l'augmentation de la précision ne justifie pas une utilisation disproportionnée des données à caractère personnel, elle devrait être évitée.

Assurez-vous donc que vous avez réellement besoin d'énormes quantités de données. Les données intelligentes peuvent être beaucoup plus utiles que les données volumineuses. Bien sûr, l'utilisation de données intelligentes et bien préparées peut impliquer un effort considérable en termes d'unification, d'homogénéisation, etc., mais elle aidera à mettre en œuvre le principe de minimisation des données (voir "Principe de minimisation des données" dans la partie II, section "Principes" des présentes lignes directrices) de manière beaucoup plus efficace. À cette fin, **il peut être extrêmement important de faire appel à un expert capable de sélectionner les caractéristiques pertinentes.**

En outre, vous devez essayer de **limiter la résolution des données** à ce qui est minimalement nécessaire aux fins poursuivies par le traitement. Vous devez également **déterminer un niveau optimal d'agrégation des données** avant de commencer le traitement (voir la section "Adéquate, pertinente et limitée" de la section "Minimisation des données" dans la Partie II, section "Principes").

La minimisation des données peut être complexe dans le cas de l'apprentissage profond, où la discrimination par caractéristiques peut être impossible. Il existe un moyen efficace de réguler la quantité de données recueillies et de ne l'augmenter que si cela semble nécessaire : la courbe d'apprentissage. Le développeur doit commencer par collecter et utiliser une quantité limitée de données d'apprentissage, puis surveiller la précision du modèle lorsqu'il est alimenté par de nouvelles données.

2.2.2 Vérification de l'utilisation légitime des jeux de données

Les ensembles de données peuvent être obtenus de différentes manières. Tout d'abord, le développeur peut choisir d'accéder à une base de données qui a déjà été construite par quelqu'un d'autre. Si tel est le cas, vous devez être particulièrement prudent, car l'acquisition de l'accès à une base de données soulève de nombreuses questions juridiques (voir "Comment accéder à une base de données" dans la Partie II, section "Principaux outils et actions"). [584]

Ensuite, l'alternative la plus courante consiste à créer une base de données. Bien évidemment, dans ce cas, vous devez vous assurer que vous respectez toutes les exigences légales imposées par le RGPD pour créer une base de données (voir "Créer une base de données" dans la partie II, section "Principaux outils et actions").

---

[584] Yeong Zee Kin, Legal Issues in AI Deployment, à l'adresse : https://lawgazette.com.sg/feature/legal-issues-in-ai-deployment/ consulté le 15 mai 2020.

Troisièmement, vous pouvez choisir une autre voie. Vous pouvez **mélanger différents ensembles de données de manière à créer un énorme ensemble de données de formation et un autre à des fins de validation.** Cela peut poser certains problèmes, comme par exemple la possibilité que la combinaison de ces données personnelles fournisse des informations supplémentaires sur les personnes concernées. Par exemple, elle pourrait vous permettre d'identifier les personnes concernées, ce qui n'était pas possible auparavant. Cela pourrait impliquer de désanonymiser des données anonymes et de créer de nouvelles informations personnelles qui n'étaient pas contenues dans l'ensemble de données d'origine, une circonstance qui soulèverait des questions éthiques et juridiques dramatiques. Par exemple, "si les personnes concernées ont donné leur consentement éclairé au traitement des informations personnelles contenues dans les ensembles de données d'origine à des fins particulières, elles n'ont pas nécessairement donné leur autorisation par extension à la fusion des ensembles de données et à l'exploration des données qui révèle de nouvelles informations. Les nouvelles informations produites de cette manière peuvent également être basées sur des probabilités ou des conjectures, et donc être fausses, ou contenir des biais dans la représentation des personnes."[585] Par conséquent, vous devez essayer d'éviter de telles conséquences en vous assurant que la fusion des ensembles de données ne va pas à l'encontre des droits et des intérêts des personnes concernées.

Enfin, si vous utilisez plusieurs ensembles de données qui poursuivent des finalités différentes, vous devez mettre en œuvre des mesures adéquates pour séparer les différentes activités de traitement. Sinon, vous pourriez facilement utiliser des données collectées pour une seule finalité dans le cadre de différentes activités. Cela pourrait poser des problèmes liés au principe de limitation de la finalité (voir "Principe de limitation de la finalité" dans la partie II, section "Principes" des présentes lignes directrices).

2.2.3 Sélection de la base juridique appropriée

Les responsables du traitement doivent décider de la base juridique utilisée pour le traitement avant de le commencer, documenter leur décision dans l'avis de confidentialité (ainsi que les finalités) et inclure les raisons pour lesquelles ces choix ont été fait (voir "Responsabilité" dans la partie II, section "Principes").

Vous devez choisir la **base juridique qui reflète le mieux la véritable nature de votre relation avec la personne et la finalité du traitement.** Cette décision est essentielle, car il n'est pas possible de changer la base juridique du traitement s'il n'y a pas de raisons solides qui le justifient (voir "Limitation de la finalité" dans la partie II, section "Principes").

Dans le cas d'un outil d'IA impliquant des données de patients, les développeurs sont généralement tentés d'utiliser le consentement comme fondement juridique du traitement. Cela pourrait avoir un sens si vous réutilisez des données qui ont déjà été collectées à une autre fin et que le consentement était la base qui permettait l'utilisation primaire des données. En effet, le RGPD autorise la réutilisation des données à des fins

---

[585] SHERPA, Lignes directrices pour le développement éthique des systèmes d'IA et de Big Data : Une approche d'éthique par la conception, 2020, p 38. À l'adresse : https://www.project-sherpa.eu/wp-content/uploads/2019/12/development-final.pdf Consulté le 15 mai 2020

scientifiques et l'article 5.1 (b) stipule que le traitement ultérieur à des fins de recherche scientifique ne doit pas être considéré comme incompatible avec les finalités initiales ("limitation de la finalité"). Ainsi, en principe, vous pourriez réutiliser ces données sur la base du consentement initial. Cependant, vous devez garder à l'esprit que, selon l'article 9.4 du RGPD, "les États membres peuvent maintenir ou introduire des conditions supplémentaires, y compris des limitations, en ce qui concerne le traitement des données génétiques, des données biométriques ou des données relatives à la santé." Ainsi, il se pourrait bien que votre réglementation nationale pertinente introduise des exceptions ou des conditions spécifiques à la réutilisation des données personnelles. En tout état de cause, vous devez toujours vous rappeler que vos devoirs d'information demeurent. Vous devez fournir à la personne concernée, avant tout traitement ultérieur de ses données, des informations sur cette autre finalité et toute autre information pertinente visée au paragraphe 2 de l'article 13 du RGPD.

La discussion sur la réutilisation des données

En ce moment, la réutilisation des données à des fins de recherche fait l'objet d'un débat animé. Selon l'article 5.1 (b) du RGPD, le traitement ultérieur à des fins scientifiques ne doit pas être considéré comme incompatible avec les finalités initiales. Ainsi, à moins que votre réglementation nationale ne stipule le contraire, vous pouvez réutiliser les données disponibles à des fins de recherche, puisque celles-ci sont compatibles avec la finalité initiale pour laquelle elles ont été collectées.

Toutefois, le CEPD a fait valoir que, "afin de garantir le respect des droits de la personne concernée, le test de compatibilité prévu à l'article 6, paragraphe 4, devrait toujours être examiné avant la réutilisation des données aux fins de la recherche scientifique, en particulier lorsque les données ont été initialement collectées pour des finalités très différentes ou en dehors du domaine de la recherche scientifique. En effet, selon une analyse du point de vue de la recherche médicale, l'application de ce test devrait être simple"[586] . Selon cette interprétation, vous ne devez réutiliser les données à caractère personnel que si les circonstances de l'article 6.4 s'appliquent.

Cette interprétation est en quelque sorte en contradiction avec l'interprétation de cette question par l'EDPB, qui a déclaré que l'article 5, paragraphe 1, point b), du RGPD prévoit que lorsque des données sont traitées ultérieurement à des fins scientifiques, "celles-ci ne sont a priori pas considérées comme incompatibles avec la finalité initiale, à condition que cela se fasse conformément aux dispositions de l'article 89, qui prévoit des garanties adéquates et des dérogations spécifiques dans ces cas. Lorsque c'est le cas, le responsable du traitement pourrait être en mesure, sous certaines conditions, de poursuivre le traitement des données sans qu'une nouvelle base juridique soit nécessaire. Ces conditions, en raison de leur nature horizontale et complexe, nécessiteront une attention et des orientations spécifiques de la part du CEPD à l'avenir. Pour l'heure, la présomption de compatibilité, sous réserve des conditions énoncées à l'article 89, ne

[586] CEPD, un avis préliminaire sur la protection des données et la recherche scientifique, 6 janvier 2020, p. 23.

devrait pas être exclue, en toutes circonstances, pour l'utilisation secondaire de données d'essais cliniques en dehors du protocole d'essai clinique à d'autres fins scientifiques"[587].

La situation reste donc floue à l'heure actuelle, même si nous considérons que l'interprétation de l'EDPB est plus logique et qu'elle prévaudra probablement à l'avenir.

Si vous pouvez collecter de nouvelles données pour votre recherche, nous vous recommandons d'éviter le consentement comme base légale, surtout si les données sont collectées dans une situation où les patients ont besoin de soins de santé urgents, comme dans le cas, par exemple, où ils souffrent de symptômes associés au COVID. Dans le contexte des essais cliniques, l'EDPB[588] a déclaré qu'"il faut garder à l'esprit que même si les conditions d'un consentement éclairé en vertu du CTR sont réunies, une situation claire de déséquilibre des pouvoirs entre le participant et le promoteur/chercheur impliquera que le consentement n'est pas "librement donné" au sens du RGPD. À titre d'exemple, l'EDPB considère que ce sera le cas lorsqu'un participant n'est pas en bonne santé, lorsque les participants appartiennent à un groupe économiquement ou socialement défavorisé ou dans toute situation de dépendance institutionnelle ou hiérarchique. Par conséquent, et comme expliqué dans les lignes directrices sur le consentement du groupe de travail 29, le consentement ne sera pas la base juridique appropriée dans la plupart des cas, et d'autres bases juridiques que le consentement doivent être utilisées (voir ci-dessous les autres bases juridiques). Par conséquent, l'EDPB considère que les responsables du traitement doivent procéder à une évaluation particulièrement approfondie des circonstances de l'essai clinique avant de s'appuyer sur le consentement des personnes comme base juridique pour le traitement des données à caractère personnel aux fins des activités de recherche de cet essai."

De notre point de vue, cet avis pourrait être étendu à d'autres scénarios où le rapport de force est biaisé. Cependant, il se peut que le comité d'éthique correspondant ne partage pas notre critère. Veuillez être conscient de ces circonstances et essayez d'éviter les inconvénients éventuels en consultant le comité et/ou votre DPD et les autorités de contrôle si nécessaire.

3 Préparation des données

[587] EDPB, Avis 3/2019 concernant les questions et réponses sur l'interaction entre le règlement sur les essais cliniques (CTR) et le règlement général sur la protection des données (RGPD) (art. 70.1.b)). Adopté le 23 janvier 2019, p. 8.

[588] AVIS 3/2019 CONCERNANT LES QUESTIONS ET REPONSES SUR L'INTERACTION ENTRE LE REGLEMENT SUR LES ESSAIS CLINIQUES (CTR) ET LE REGLEMENT GENERAL SUR LA PROTECTION DES DONNEES (RGPD), A L'ADRESSE : **https://edpb.europa.eu/our-work-tools/our-documents/dictamen-art-70/opinion-32019-concerning-questions-and-answers_en.**

3.1 Description

"La phase de préparation des données couvre toutes les activités visant à construire l'ensemble de données final ou les données qui seront introduites dans le ou les outils de modélisation à partir des données brutes initiales. Les tâches comprennent la sélection des tables, des enregistrements et des attributs, ainsi que la transformation et le nettoyage des données pour les outils de modélisation. Les cinq étapes de la préparation des données sont la sélection des données, le nettoyage des données, la construction des données, l'intégration des données et le formatage des données."[589]

Cette étape comprend toutes les activités nécessaires pour construire l'ensemble de données final qui est introduit dans le modèle, à partir des données brutes initiales. Elle comprend les cinq tâches suivantes, qui ne sont pas nécessairement exécutées de manière séquentielle :

1. Sélectionner les données. Décidez des données à utiliser pour l'analyse, en fonction de leur pertinence par rapport aux objectifs de l'exploration de données, de leur qualité et des contraintes techniques telles que les limites du volume ou des types de données.

2. Nettoyer les données. Amenez la qualité des données à un niveau requis, par exemple en sélectionnant des sous-ensembles de données propres, en insérant des valeurs par défaut et en estimant les données manquantes par modélisation.

3. Construire des données. La construction de nouvelles données par la production d'attributs dérivés, de nouveaux enregistrements ou de valeurs transformées pour des attributs existants.

4. Intégrer des données. Combiner les données de plusieurs tables ou enregistrements pour créer de nouveaux enregistrements ou valeurs.

5. Formater les données. Apporter des modifications syntaxiques aux données qui pourraient être requises par l'outil de modélisation.

3.2 Principales mesures à prendre

3.2.1 Introduire les garanties prévues à l'article 89 du RGPD.

Puisque vous utilisez des données à des fins scientifiques, vous devez les préparer selon les garanties prévues par le RGPD dans son article 89. Si les finalités de votre recherche peuvent être atteintes par un traitement ultérieur qui ne permet pas ou plus l'identification des personnes concernées, c'est-à-dire par la pseudonymisation, ces finalités doivent être atteintes de cette manière. Si cela n'est pas possible, vous devez introduire des garanties assurant que les mesures techniques et organisationnelles qui permettent une mise en œuvre adéquate du principe de minimisation des données. Veuillez prendre en considération les règles concrètes établies par votre réglementation nationale concernant les garanties. Consultez votre DPD.

[589] Colin Shearer, Le modèle CRISP-DM : The New Blueprint for Data Mining, p. 16.

3.2.2 Garantir la précision du traitement des données à caractère personnel

Selon le RGPD, les données doivent être exactes (voir "Précision" dans la partie II, section "Principes").

Cela signifie que les données doivent être correctes et à jour, mais aussi que les analyses effectuées doivent être exactes. L'EDPB a souligné l'importance de l'exactitude du profilage ou du processus décisionnel (non exclusivement) automatisé à tous les stades (de la collecte des données à l'application du profil à l'individu).[590]

Les responsables du traitement sont chargés de garantir la précision des données. Par conséquent, une fois que vous avez terminé la collecte des données, vous devez mettre en place des outils adéquats pour garantir la précision des données. Cela implique généralement que vous deviez prendre des décisions fondamentales sur les mesures techniques et organisationnelles qui rendront ce principe applicable (voir la sous-section "Mesures techniques et organisationnelles connexes" dans la section "Précision" du chapitre "Principes"). Puisque la plupart des données proviennent des patients et que la plupart d'entre elles sont quantitatives, vous pouvez supposer qu'elles sont exactes. En tout état de cause, la précision exige une mise en œuvre adéquate des mesures destinées à faciliter le droit de rectification des personnes concernées (voir "Droit de rectification" dans la partie II, section "Droits des personnes concernées".

3.2.3 Se concentrer sur les questions de profilage

Dans le cas d'une base de données qui servira à former ou à valider un outil d'IA, il existe une obligation particulièrement pertinente d'informer les personnes concernées que **leurs données pourraient entraîner une prise de décision automatisée ou un profilage à leur égard, à moins que vous puissiez garantir que l'outil ne produira en aucun cas ces conséquences.** Même si la prise de décision automatique ne peut guère se produire dans le contexte de la recherche, les développeurs doivent garder un œil ouvert sur cette question. Le profilage, quant à lui, pourrait poser certains problèmes au développement de l'IA.

Selon l'article 22, paragraphe 3, les décisions automatisées qui portent sur des catégories particulières de données à caractère personnel, telles que les données relatives à la santé que vous utilisez, ne sont autorisées que si la personne concernée a donné son consentement ou si elles sont fondées sur une base juridique. Cette exception s'applique non seulement lorsque les données observées entrent dans cette catégorie, mais **aussi si le rapprochement de différents types de données à caractère personnel peut révéler des informations sensibles sur des personnes ou si des données déduites entrent dans cette catégorie.**

Voici quelques actions supplémentaires qui pourraient être extrêmement utiles pour éviter le profilage s'il n'est pas nécessaire :

---

[590] *Lignes directrices* sur la *prise de décision* individuelle *automatisée et le profilage* aux fins du règlement 2016/679 (wp251rev.01). 22/08/2018, *p.* 13 ; Ducato, Rossana, Private Ordering of Online Platforms in Smart Urban Mobility The Case of Uber's Rating System, CRIDES Working Paper Series no. 3/20202 February 2020 Updated on 26 July 2020, p. 20-21, at: https://poseidon01.ssrn.com/delivery.php?ID=247104118003073117118086021112071111102048023015008020118084071112086000027097102088036101006014057116105116119119026079007006118044033055000114023106007076115096073024007094081002078064098028091093003078095099082108113086098120001079015123027083125024&EXT=pdf&INDEX=TRUE

- Tenir compte des exigences du système nécessaires pour soutenir un examen humain significatif **dès la phase de conception.** En particulier, les exigences d'interprétabilité et la conception d'une interface utilisateur efficace pour soutenir les examens et les interventions humaines ;
- Concevoir et offrir une formation et un soutien appropriés aux examinateurs humains ; et
- Donner au personnel l'autorité, les incitations et le soutien appropriés pour répondre aux préoccupations des personnes ou les transmettre à un échelon supérieur et, si nécessaire, passer outre la décision de l'outil d'IA.

Si vous procédez à un profilage ou à des décisions automatisées, vous devez informer les personnes concernées de votre décision et fournir toutes les informations nécessaires conformément au RGPD et à la réglementation nationale, le cas échéant.

3.2.4 Sélection de données non biaisées

La partialité est l'un des principaux problèmes liés au développement de l'IA, un problème qui va à l'encontre du principe de loyauté (voir "Principe de licéité, de loyauté et de transparence" dans la partie II, section "Principes" des présentes lignes directrices). Les biais peuvent être causés par de nombreux facteurs différents. Lorsque des données sont recueillies, elles peuvent contenir des biais, des inexactitudes, des erreurs et des fautes construits par la société. Parfois, il peut arriver que les ensembles de données soient biaisés en raison d'actions malveillantes. L'introduction de données malveillantes dans un outil d'IA peut modifier son comportement, en particulier avec les systèmes d'auto-apprentissage.[591] Par conséquent, les questions liées à la composition des bases de données utilisées pour la formation soulèvent des problèmes éthiques et juridiques cruciaux, et pas seulement des questions d'efficacité ou de nature technique.

Vous devez résoudre ces problèmes avant de former l'algorithme. Dans la mesure du possible, les biais identifiables et discriminatoires doivent être supprimés lors de la phase de constitution des ensembles de données. Dans le cas de COVID, des distinctions pourraient être faites entre les patients en fonction de leur âge, de leur genre ou de leur groupe ethnique, par exemple. Vous devez vous assurer que l'algorithme tient compte de ce facteur lors de la sélection des données. Cela signifie que **les équipes chargées de sélectionner les données à intégrer dans les jeux de données doivent être composées de personnes qui garantissent la diversité dont le développement de l'IA est censé faire preuve.** Enfin, gardez toujours à l'esprit que, si vos données sont principalement liées à un groupe concret, par exemple la population caucasienne de plus de quarante ans, vous devez déclarer que l'algorithme a été formé sur cette base et, par conséquent, il pourrait ne pas fonctionner aussi bien dans d'autres groupes de population.

---

[591] Groupe d'experts de haut niveau sur l'IA, Lignes directrices en matière d'éthique pour une IA digne de confiance, 2019, p. 17. À l'adresse : https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai Consulté le 15 mai 2020

4 Modélisation (formation)

4.1 Description

"Dans cette phase, diverses techniques de modélisation sont sélectionnées et appliquées et leurs paramètres sont calibrés à des valeurs optimales. Généralement, plusieurs techniques existent pour le même type de problème d'exploration de données. Certaines techniques ont des exigences spécifiques sur la forme des données. Par conséquent, il peut être nécessaire de revenir à la phase de préparation des données. Les étapes de modélisation comprennent la sélection de la technique de modélisation, la génération du plan de test, la création de modèles et l'évaluation des modèles."[592]

Cette phase comporte plusieurs tâches essentielles. Dans l'ensemble, vous devez :

- Sélectionner la technique de modélisation qui sera utilisée. Selon le type de technique, des conséquences telles que l'inférence des données, l'obscurité ou les biais sont plus ou moins susceptibles de se produire.
- Prendre une décision sur l'outil de formation à utiliser. Cela permet au développeur de mesurer la capacité du modèle à prédire l'histoire avant de l'utiliser pour prédire l'avenir. La formation implique toujours l'exécution de tests empiriques avec des données. Parfois, les développeurs testent le modèle avec des données différentes de celles utilisées pour le générer. Par conséquent, à ce stade, on peut parler de différents types d'ensembles de données. Il est parfois difficile d'identifier les personnes auxquelles se rapportent les données de formation. Cela crée des problèmes pour le respect des droits des personnes, qui doivent être traités de manière appropriée.

4.2 Principales mesures à prendre

4.2.1 Mise en œuvre du principe de minimisation des données

Selon le principe de minimisation des données, vous devez procéder à la réduction de la quantité de données et/ou de l'éventail d'informations sur la personne concernée qu'ils fournissent dès que possible. Par conséquent, vous devez purger les données utilisées pendant la phase de formation de toutes les informations qui ne sont pas strictement nécessaires à la formation du modèle. (voir la sous-section "Aspect temporel dans la minimisation des données" dans la partie II, section "Principes"). Il existe plusieurs stratégies pour assurer la minimisation des données lors de la phase de formation. Bien sûr, vous devriez commencer par effacer toutes les données personnelles liées aux rayons X que vous utilisez, mais ce ne serait qu'une première étape pour suivre le principe de minimisation. Des mesures plus strictes devraient être soigneusement mises en œuvre à cette fin. Les techniques évoluent en permanence. Toutefois, certaines des

[592] Colin Shearer, Le modèle CRISP-DM : The New Blueprint for Data Mining, p. 17.

plus courantes sont les suivantes[593] (voir également "Intégrité et confidentialité" dans la partie II, section "Principes") :

-Analyse des conditions que les données doivent remplir pour être considérées comme de haute qualité et avec une grande capacité de prédiction pour l'application spécifique.

- Analyse critique de l'étendue de la typologie des données utilisées à chaque étape de l'outil d'IA.

- Suppression des données non structurées et des informations inutiles recueillies lors du prétraitement de l'information.

-Identification et suppression des catégories de données qui n'ont pas d'influence significative sur l'apprentissage ou sur le résultat de l'inférence.

- Suppression des conclusions non pertinentes associées aux informations personnelles pendant le processus de formation, par exemple, dans le cas d'une formation non supervisée.

- Utilisation de techniques de vérification qui nécessitent moins de données, comme la validation croisée.

- Analyse et configuration des hyperparamètres algorithmiques pouvant influencer la quantité ou l'étendue des données traitées afin de les minimiser.

- Utilisation de modèles d'apprentissage fédérés plutôt que centralisés.

- Application de stratégies de confidentialité différentielle.

- Entraînement avec des données cryptées en utilisant des techniques homomorphiques.

- Agrégation de données.

- Anonymisation et pseudonymisation, non seulement dans la communication des données, mais aussi dans les données de formation, les éventuelles données personnelles contenues dans le modèle et dans le traitement de l'inférence.

4.2.2 Détecter et effacer les biais

Même si les mécanismes de lutte contre les biais sont convenablement adoptés lors des étapes précédentes (voir la section précédente sur la formation), il faut encore s'assurer que les résultats de la phase de formation minimisent les biais. Cela peut être difficile, car certains types de biais et de discrimination sont souvent particulièrement difficiles à détecter. Les membres de l'équipe qui traite les données d'entrée n'en sont parfois pas conscients, et les utilisateurs qui sont leurs sujets n'en sont pas nécessairement conscients non plus. Ainsi, les systèmes de contrôle mis en place par le développeur d'IA lors de la phase de validation sont des facteurs extrêmement importants pour éviter les biais.

Il existe de nombreux outils techniques qui peuvent servir à détecter les biais, comme l'évaluation de l'impact algorithmique.[594] Vous devez envisager leur mise en œuvre

[593] AEPD, Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción, 2020, p.40. À l'adresse : https://www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf Consulté le 15 mai 2020.

effective.[595] Cependant, comme le montre la littérature[596] , il peut arriver qu'un algorithme ne puisse pas être totalement purgé de tous les différents types de biais. Vous devez cependant essayer d'être au moins conscient de leur existence et des implications que cela peut entraîner (voir "Léicité, loyauté et transparence" et "Précision" dans la partie II, section "Principes").

4.2.3 Exercice des droits des personnes concernées

Parfois, les développeurs complètent les données disponibles par inférence. Par exemple, si vous ne disposez pas des données concrètes correspondant à la pression artérielle d'un patient, vous pouvez utiliser un autre algorithme pour la déduire du reste des données. Toutefois, cela ne signifie pas que ces données peuvent être considérées comme entièrement pseudonymisées ou anonymisées. C'est particulièrement vrai dans le cas des données génomiques, car leur anonymisation est presque impossible. Elles restent donc des données personnelles. En outre, les données déduites doivent également être considérées comme des données à caractère personnel. Par conséquent, les personnes concernées ont certains droits fondamentaux sur ces données que vous devez respecter.

En effet, vous devez faciliter tous les droits des personnes concernées tout au long du cycle de vie. Dans cette étape spécifique, les droits d'accès, de rectification et d'effacement sont particulièrement sensibles et comportent certaines caractéristiques que les responsables de traitement doivent connaître. Toutefois, dans le cas de recherches à des fins scientifiques telles que celle que vous développez, le RGPD prévoit certaines garanties et dérogations relatives au traitement (art. 89). Vous devez être au courant de la réglementation concrète de votre État membre. Selon le RGPD, le droit de l'Union ou des États membres peut prévoir des dérogations aux principaux droits inclus dans les articles 15 et suivants, dans la mesure où ces droits sont susceptibles de rendre impossible ou de nuire gravement à la réalisation des finalités spécifiques, et où ces dérogations sont nécessaires à la réalisation de ces finalités.

-Droit d'accès (voir "Droit d'accès" dans la partie II, section "Droits de la personne concernée")

En principe, vous devez répondre aux demandes d'accès des personnes concernées à leurs données personnelles, à condition d'avoir pris des mesures raisonnables pour vérifier l'identité de la personne concernée, et qu'aucune autre exception ne s'applique. Toutefois, vous n'êtes pas tenu de collecter ou de conserver des données à caractère personnel supplémentaires pour permettre l'identification des personnes concernées dans les données de formation dans le seul but de vous conformer au règlement. Si vous ne pouvez pas identifier une personne concernée dans les données de formation et que la

---

[594] Reisman, D., Crawford, K., Whittaker, M., Algorithmic impact assessments : Un cadre pratique pour la responsabilité des agences publiques, 2018, à l'adresse : https://ainowinstitute.org/aiareport2018.pdf Consulté le 15 mai 2020.

[595] https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf consulté le 15 mai 2020

[596] Chouldechova. Alexandra, Fair Prediction with Disparate Impact : Une étude des biais dans les instruments de prédiction de la récidive, Big Data. Volume : 5 Numéro 2 : 1er juin 2017. 153-163.http://doi.org/10.1089/big.2016.0047

personne concernée ne peut pas fournir d'informations supplémentaires qui permettraient son identification, elle n'est pas obligée de satisfaire une demande qu'il n'est pas possible de satisfaire.

-*Droit de* ***rectification*** (voir "Droit de rectification" dans la partie II, section "Droits de la personne concernée").

Dans le cas du droit de rectification, vous devez garantir le droit de rectification des données, notamment celles générées par les déductions et les profils établis par un outil d'IA. Même si l'objectif des données d'entraînement est de former des modèles basés sur des modèles généraux dans de grands ensembles de données et que, par conséquent, les inexactitudes individuelles sont moins susceptibles d'avoir un effet direct sur une personne concernée, le droit de rectification ne peut pas être limité. Au maximum, vous pouvez demander un délai plus long (deux mois supplémentaires) pour procéder à la rectification si la procédure technique est particulièrement complexe (article 11, paragraphe 3).

-Droit à l'effacement (voir "Droit à l'effacement" dans la partie II, section "Droits de la personne concernée").

Les personnes concernées ont le droit de demander la suppression de leurs données personnelles. Toutefois, ce droit peut être limité si certaines circonstances concrètes s'appliquent. Selon l'ICO, "les organisations peuvent également recevoir des demandes d'effacement de données de formation. Les organisations doivent répondre aux demandes d'effacement, sauf si une exemption pertinente s'applique et à condition que la personne concernée ait des motifs appropriés. Par exemple, si les données de formation ne sont plus nécessaires parce que le modèle ML a déjà été formé, l'organisation doit satisfaire la demande. Toutefois, dans certains cas, lorsque le développement du système est en cours, il peut encore être nécessaire de conserver les données de formation aux fins du réentraînement, du perfectionnement et de l'évaluation d'un outil d'IA. Dans ce cas, l'organisation doit adopter une approche au cas par cas pour déterminer si elle peut satisfaire les demandes. Se conformer à une demande d'effacement des données d'entraînement n'entraînerait pas l'effacement des modèles ML basés sur ces données, sauf si les modèles eux-mêmes contiennent ces données ou peuvent être utilisés pour les déduire." [597]

5 Évaluation (validation)

5.1 Description

"Avant de procéder au déploiement final du modèle construit par l'analyste de données, il est important de procéder à une évaluation plus approfondie du modèle et de revoir la

[597] ICO, Enabling access, erasure, and rectification rights in AI tools (Permettre les droits d'accès, d'effacement et de rectification dans les outils d'IA), à l'adresse suivante : https://ico.org.uk/about-the-ico/news-and-events/ai-blog-enabling-access-erasure-and-rectification-rights-in-ai-systems/. Consulté le 15 mai 2020.

construction du modèle pour s'assurer qu'il atteint correctement les objectifs de l'entreprise. Il est essentiel de déterminer si certaines questions importantes n'ont pas été suffisamment prises en compte. À la fin de cette phase, le chef de projet doit alors décider exactement comment utiliser les résultats de l'exploration de données. Les étapes clés ici sont l'évaluation des résultats, la révision du processus et la détermination des prochaines étapes."[598]

Cette phase comporte plusieurs tâches qui soulèvent d'importantes questions relatives à la protection des données. Globalement, vous devez :

- Évaluer les résultats du modèle, par exemple pour savoir s'il est précis ou non. À cette fin, le développeur d'IA peut le tester dans le monde réel.
- Révisez le processus. Vous devez examiner la mission d'exploration de données pour déterminer s'il y a un facteur ou une tâche importante qui a été en quelque sorte négligée. Cela inclut les questions d'assurance qualité.

5.2 Principales mesures à prendre

5.2.1 Processus de validation dynamique

La validation du traitement comprenant un composant d'IA doit être effectuée dans des conditions qui reflètent l'environnement réel dans lequel le traitement est destiné à être déployé. Ainsi, si vous savez à l'avance où l'outil d'IA sera utilisé, vous devez adapter le processus de validation à cet environnement. Par exemple, si l'outil sera déployé en Italie, vous devez le valider avec des données obtenues auprès de la population italienne ou, si ce n'est pas possible, auprès d'une population similaire. Dans le cas contraire, les résultats pourraient être totalement erronés. Dans tous les cas, vous devez informer tout utilisateur potentiel des conditions de la validation.

En outre, le processus de validation nécessite un examen périodique si les conditions changent ou si l'on soupçonne que la solution elle-même peut être altérée. Par exemple, si l'algorithme est alimenté par des données provenant de personnes âgées, vous devez évaluer si cela modifie ou non sa précision dans une population jeune. Vous devez vous assurer que la validation reflète fidèlement les conditions dans lesquelles l'algorithme a été validé.

Pour atteindre cet objectif, la validation doit inclure tous les composants d'un outil d'IA, y compris les données, les modèles pré-entraînés, les environnements et le comportement du système dans son ensemble. La validation doit également être effectuée le plus tôt possible. Dans l'ensemble, il faut s'assurer que les résultats ou les actions sont cohérents avec les résultats des processus précédents, en les comparant aux politiques préalablement définies pour s'assurer qu'elles ne sont pas violées.[599] La validation nécessite parfois la collecte de nouvelles données à caractère personnel. Dans d'autres cas, les responsables du traitement utilisent les données à des fins autres que celles prévues à l'origine. Dans tous ces cas, les responsables du traitement doivent

---

[598] Colin Shearer, Le modèle CRISP-DM : Le nouveau plan directeur pour l'extraction de données, p. 17

[599] Groupe d'experts de haut niveau sur l'IA, Lignes directrices en matière d'éthique pour une IA digne de confiance, 2019, p. 22. À l'adresse : https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai v

s'assurer du respect du RGPD (voir la section "Limitation de la finalité" dans la partie "Principes" et "Protection des données et recherche scientifique" dans la partie "Concepts principaux" de la partie II des présentes lignes directrices).

5.2.2 Suppression des jeux de données inutiles s

Très souvent, les processus de validation et de formation sont en quelque sorte liés. Si la validation recommande des améliorations du modèle, la formation doit être effectuée à nouveau. En principe, une fois le développement de l'IA achevé, l'étape de formation de l'outil d'IA est terminée. À ce moment-là, vous devriez mettre en œuvre la suppression de l'ensemble des données utilisées à cette fin, à moins qu'il n'existe un besoin légitime de les conserver pour affiner ou évaluer le système, ou pour d'autres finalités compatibles avec celles pour lesquelles elles ont été collectées conformément aux conditions de l'article 6, paragraphe 4, du RGPD. Cependant, vous devez toujours considérer que la suppression des données personnelles peut aller à l'encontre de la nécessité de mettre à jour la précision des outils basés sur l'auto-apprentissage en temps réel des algorithmes : si une erreur est trouvée, vous devrez probablement rappeler les données précédemment utilisées dans la phase de formation. Dans le cas où les personnes concernées demandent leur suppression, vous devrez adopter une approche au cas par cas en tenant compte des éventuelles limitations à ce droit prévues par le règlement (voir art. 17, paragraphe 3). [600]

5.2.3 Réalisation d'un audit externe du traitement des données

Étant donné que les risques du système que vous développez sont élevés, **un audit du système par un tiers indépendant doit être envisagé**. Différents audits peuvent être utilisés. Ils peuvent être internes ou externes, ils peuvent couvrir uniquement le produit final, ou être réalisés avec des prototypes moins évolués. Ils peuvent être considérés comme une forme de contrôle ou un outil de transparence.

En termes d'exactitude juridique, les outils d'IA doivent être vérifiés pour voir s'ils traitent les données personnelles conformément aux dispositions du RGPD, en tenant compte d'un large éventail de questions qui pourraient être liées à ce traitement. Le groupe d'experts de haut niveau sur l'IA a déclaré que "les processus de test devraient être conçus et réalisés par un groupe de personnes aussi diversifié que possible. Des mesures multiples devraient être développées pour couvrir les catégories qui sont testées pour différentes perspectives. On peut envisager des tests contradictoires réalisés par des "équipes rouges" fiables et diverses qui tentent délibérément de "casser" le système pour trouver des vulnérabilités, ainsi que des "primes aux bogues" qui incitent les personnes extérieures à détecter et à signaler de manière responsable les erreurs et les faiblesses du système."[601] Cependant, il existe de bonnes raisons d'être sceptique quant à la capacité d'un auditeur à vérifier le fonctionnement d'un système d'apprentissage automatique.

---

[600] AEPD, Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción, 2020, p.26. À l'adresse : https://www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf

[601] Groupe d'experts de haut niveau sur l'IA, Lignes directrices en matière d'éthique pour une IA digne de confiance, 2019, p. 22. À l'adresse : https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai Consulté le 15 mai 2020

C'est pourquoi il est judicieux de se concentrer sur les éléments inclus par l'AEPD dans sa liste de contrôle recommandée : il serait plus simple de se concentrer sur les mesures mises en œuvre pour éviter le biais, l'obscurité, le profilage caché, etc., et sur l'utilisation adéquate d'outils tels que le AIPD, qui peut être effectué plusieurs fois, que d'essayer d'avoir une compréhension approfondie du fonctionnement d'un algorithme complexe. La mise en œuvre de politiques de protection des données adéquates dès les premières étapes du cycle de vie de l'outil est le meilleur moyen d'éviter les problèmes de protection des données.

5.2.4 Assurer la conformité avec le cadre juridique des dispositifs médicaux

Avant de déployer votre dispositif, vous devez vous assurer que vous avez bien suivi la réglementation concernant le développement des dispositifs médicaux. Veuillez vous assurer que c'est le cas. Une évaluation clinique et une évaluation des performances doivent également être développées. Le guide sur l'évaluation clinique (MDR) / l'évaluation des performances (IVDR) des logiciels de dispositifs médicaux (https://ec.europa.eu/docsroom/documents/40323 pourrait être un excellent outil à cet effet).

5.2.5 Informer les travailleurs de la santé qui participent au développement des problèmes possibles

Il est fréquent que les mécanismes d'IA soient validés en comparant leurs performances à celles d'éléments humains, en l'occurrence des professionnels de la santé. Cela peut conduire subrepticement à ce que leur participation induise une évaluation de leur propre capacité professionnelle. Si l'on compare le taux de réussite de certains professionnels avec d'autres, certains d'entre eux peuvent avoir l'impression d'être testés par inadvertance. Il est très important d'essayer d'éviter cet effet. S'il doit se produire, les participants doivent en être avertis et l'accepter.

6 Déploiement

6.1 Description

"Le déploiement est le processus qui consiste à rendre un système informatique opérationnel dans son environnement, y compris l'installation, la configuration, l'exécution, les tests et les modifications nécessaires. Le déploiement n'est généralement pas effectué par les développeurs d'un système mais par l'équipe informatique du client. Néanmoins, même si c'est le cas, les développeurs auront la responsabilité de fournir au client des informations suffisantes pour un déploiement réussi du modèle. Cela comprendra normalement un plan de déploiement (générique), avec les étapes nécessaires pour un déploiement réussi et la manière de les réaliser, et un plan de surveillance et de maintenance (générique) pour la maintenance du système, et pour la

surveillance du déploiement et de l'utilisation correcte des résultats de l'exploration de données."[602]

6.2 Principales mesures à prendre

6.2.1 Remarques générales

Une fois que vous avez créé votre algorithme, vous êtes confronté à un problème important. Il se peut qu'il incorpore des données personnelles, ouvertement ou de manière cachée. Vous devez procéder à une évaluation formelle pour déterminer quelles données personnelles des personnes concernées pourraient être identifiables. Cela peut parfois être compliqué. Par exemple, certains outils d'IA, tels que les machines à support vectoriel (VSM), peuvent contenir des exemples de données d'entraînement dans la logique du modèle. Dans d'autres cas, des modèles peuvent être trouvés dans le modèle qui identifie un individu unique. Dans tous ces cas, des parties non autorisées peuvent être en mesure de récupérer des éléments des données d'entraînement ou de déduire qui y figurait, en analysant le comportement du modèle. Si vous savez ou soupçonnez que l'outil d'IA contient des données personnelles (voir la section "Achat ou promotion de l'accès à une base de données" dans "Principaux outils et actions", partie II des présentes lignes directrices), vous devez :

- Les supprimer ou, au contraire, justifier l'impossibilité de le faire, en tout ou partie, en raison de la dégradation que cela impliquerait pour le modèle (voir la section "Limitation du stockage" dans "Principes" de la partie II).

- Déterminer la base juridique de la communication de données à caractère personnel à des tiers, en particulier si des catégories spéciales de données sont concernées (voir la sous-section "Licéité" dans "Licéité, loyauté et transparence" dans la section "Principes" de la partie II).

- Informer les personnes concernées du traitement ci-dessus.

- Démontrer que les politiques de protection des données dès la conception et par défaut ont été mises en œuvre (notamment la minimisation des données) (voir "Protection des données dès la conception et par défaut" dans la partie II, section "Concepts principaux" des présentes lignes directrices).

- Réaliser une analyse d'impact sur la protection des données (AIPD) **(**voir "AIPD" dans la partie II, section "Principales actions et outils" des présentes lignes directrices).

Enfin, vous devez prendre des mesures régulières pour évaluer de manière proactive la probabilité que des données à caractère personnel soient déduites de modèles à la lumière de l'état de la technologie, afin de minimiser le risque de divulgation accidentelle. Si ces actions révèlent une possibilité substantielle de divulgation des données, les mesures nécessaires pour l'éviter doivent être mises en œuvre (voir la section "Intégrité et confidentialité" dans les "Principes" de la partie II des présentes lignes directrices).

---

[602] SHERPA, Lignes directrices pour le développement éthique des systèmes d'IA et de Big Data : Une approche d'éthique par la conception, 2020, p 13. À l'adresse : https://www.project-sherpa.eu/wp-content/uploads/2019/12/development-final.pdf Consulté le 15 mai 2020

6.2.2 Mise à jour des informations

Si l'algorithme est mis en œuvre par un tiers, vous devez communiquer les résultats du système de validation et de suivi employé lors des phases de développement et proposer votre collaboration pour continuer à suivre la validation des résultats. Il serait également souhaitable d'établir ce type de coordination avec les tiers auprès desquels vous acquérez des bases de données ou tout autre composant pertinent dans le cycle de vie du système. Si cela implique le traitement de données par un tiers, vous devez vous assurer que l'accès est fourni dans le cadre d'une base légale.

Il est nécessaire d'offrir à l'utilisateur final des informations en temps réel sur les valeurs de précision et/ou de qualité des informations déduites à chaque étape (voir la section "Précision" dans "Principes", partie II des présentes lignes directrices). Lorsque les informations déduites n'atteignent pas les seuils de qualité minimum, vous devez souligner que ces informations n'ont aucune valeur. Cette exigence implique souvent que vous devez fournir des informations détaillées sur les étapes de formation et de validation. Les informations sur les ensembles de données utilisés à ces fins sont particulièrement importantes. Dans le cas contraire, l'utilisation de la solution risque d'apporter des résultats décevants aux utilisateurs finaux, qui se retrouvent à spéculer sur la cause.

Deuxième scénario : L'IA pour la prédiction et la prévention des infractions pénales

Johann Čas (ITA/OEAW)

Cette partie des lignes directrices a été revue et validée par Marko Sijan, conseiller principal spécialiste (DPA RH).

Introduction et remarques préliminaires

L'utilisation de TIC avancées joue - en tant que technologie essentielle pour toutes les activités économiques, gouvernementales ou sociétales - un rôle de plus en plus important dans la prévision, la prévention, l'investigation et la poursuite d'activités criminelles ou terroristes. En conséquence, la recherche visant à développer et à améliorer les capacités techniques des services répressifs (LEA) constitue un domaine prioritaire des programmes de financement passés, actuels et futurs de la CE. Les TIC

avancées et émergentes possèdent des pouvoirs de contrôle et d'analyse sans précédent d'ensembles de données vastes et diversifiés, notamment en relation avec les technologies d'IA[603] . La recherche sur ces technologies, ainsi que la mise en œuvre de TIC avancées dans le contexte de la sécurité, soulèvent de sérieuses préoccupations en matière d'éthique et de conformité juridique. Les programmes de recherche sur la sécurité financés par l'UE exigent explicitement le respect total des dispositions de la Charte des droits fondamentaux de l'Union européenne,[604] la prise en compte du respect de la vie privée dès la conception, de la protection des données dès la conception, du respect de la vie privée par défaut et de la protection des données par défaut,[605] et, en plus du tableau d'auto-évaluation éthique[606] également de remplir un tableau d'impact sociétal. Un "tableau d'impact sociétal" est une caractéristique spécifique de cette partie du programme de travail. Ce tableau met l'accent sur les aspects sociétaux de la recherche en sécurité. Il vérifie si la recherche en matière de sécurité proposée répond aux besoins de la société, lui est bénéfique et n'a pas d'impact négatif sur elle. Les candidats doivent remplir le 'tableau d'impact sociétal' dans le cadre du processus de soumission." [607] Des procédures similaires devraient également être mises en œuvre au niveau de la conception des programmes de travail. Des garanties supplémentaires devraient être prévues pour que les programmes ne contiennent pas d'appels qu'il est difficile ou impossible de satisfaire sans soulever de graves problèmes d'éthique ou provoquer des atteintes disproportionnées aux droits de l'Homme. Cela pourrait être réalisé par une participation obligatoire de représentants de la société civile et d'experts

[603] L'IA est un terme (trop) fréquemment utilisé, sans définition unique. Nous nous référons ici à la définition large de l'IA, élaborée par le groupe d'experts de haut niveau sur l'IA :
"Les systèmes d'intelligence artificielle (IA) sont des systèmes logiciels (et éventuellement matériels) conçus par des humains qui, compte tenu d'un objectif complexe, agissent dans la dimension physique ou numérique en percevant leur environnement par l'acquisition de données, en interprétant les données structurées ou non structurées collectées, en raisonnant sur les connaissances, ou en traitant les informations, dérivées de ces données et en décidant de la ou des meilleures actions à entreprendre pour atteindre l'objectif donné. Les outils d'IA peuvent soit utiliser des règles symboliques, soit apprendre un modèle numérique, et ils peuvent également adapter leur comportement en en analysant comment l'environnement est affecté par leurs actions précédentes.
En tant que discipline scientifique, l'IA comprend plusieurs approches et techniques, telles que l'apprentissage automatique (dont l'apprentissage profond et l'apprentissage par renforcement sont des exemples spécifiques), le raisonnement automatique (qui comprend la planification, l'ordonnancement, la représentation et le raisonnement des connaissances, la recherche et l'optimisation) et la robotique (qui comprend le contrôle, la perception, les capteurs et les actionneurs, ainsi que l'intégration de toutes les autres techniques dans les systèmes cyber-physiques)." https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines

[604] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2010:083:0389:0403:en:PDF

[605] Pour plus de détails, voir EDPB. (2019). Lignes directrices 4/2019 relatives à l'article 25 Protection des données dès la conception et par défaut Version 2.0. Adopté le 20 octobre 2020.
<
https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf>

[606]
https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-self-assess_en.pdf

[607] Voir p.5 https://ec.europa.eu/research/participants/data/ref/h2020/wp/2018-2020/main/h2020-wp1820-security_en.pdf

en éthique et en droit au sein des groupes d'experts qui élaborent les programmes de recherche financés par l'UE.

Ces précautions sont essentielles pour mettre la recherche sur la sécurité en conformité avec des principes tels que les droits de l'Homme et la démocratie ; néanmoins, des inquiétudes subsistent quant au fait qu'elles pourraient accroître la légitimité des projets de recherche sur la sécurité sans garantir le respect des règles éthiques et juridiques dans la pratique.[608] L'utilisation de l'IA dans le contexte de la prédiction ou de la prévention des infractions pénales fait peser de graves menaces sur les libertés civiles. Un simple compromis entre sécurité et liberté n'est ni approprié ni suffisant. Cette relation complexe doit être traitée comme une sorte de symbiose hostile,[609] impliquant que les deux sont nécessaires à la survie de l'autre.

Pour tenir compte de ces préoccupations, ce scénario intègre également des informations provenant des appels de recherche H2020 sur la sécurité existants, en particulier de l'appel H2020-SEC-2016-2017 et des projets actuellement en cours ou récemment terminés. MAGNETO[610] (Moteur d'analyse et de corrélation multimédia pour la prévention et l'investigation du crime organisé), CONNEXIONs[611] (Plateforme IdO immersive de génération NEXt interconnectée de services de détection, de prédiction, d'investigation et de prévention du crime et du terrorisme) ou RED-Alert[612] (Système de détection précoce et d'alerte en temps réel pour le contenu terroriste en ligne basé sur le traitement du langage naturel, l'analyse des réseaux sociaux, l'intelligence artificielle et le traitement des événements complexes) sont des exemples de projets pertinents pour cette étude de cas. Ils sont financés par l'appel 2016-2017 Technologies for prevention, investigation, and mitigation in the context of the fight against crime and terrorism.[613] Le projet initial de prendre l'un de ces projets comme base concrète pour ce scénario a été abandonné car la plupart, voire la quasi-totalité des résultats des projets mentionnés sont, conformément à la réglementation H2020,[614] classifiés et non accessibles au public. Si la classification des résultats spécifiques des projets de recherche sur la sécurité peut être nécessaire et compréhensible, elle limite certainement aussi la possibilité d'un examen et d'un débat publics sur ces technologies, qui devraient être obligatoires au vu des violations potentielles des droits de l'Homme et des valeurs européennes.

La complexité de ce cas d'utilisation est encore accrue par le fait que différentes réglementations s'appliquent à la phase de recherche et de développement, d'une part, et à la phase de mise en œuvre et d'utilisation, d'autre part. Les activités de recherche sont soumises au RGPD; les applications futures des résultats de la recherche sont soumises à la *directive d'application de la loi sur la protection des données* (directive 2016/680)

---

[608] Leese, M., Lidén, K. und Nikolova, B., 2019, Putting critique to work : Ethics in EU security research, Security Dialogue 50(1), 59-76 <https://journals.sagepub.com/doi/abs/10.1177/0967010618809554>.
[609] Wittes, B. (2011). Contre un équilibre brut : La sécurité des plateformes et la symbiose hostile entre liberté et sécurité. Projet sur le droit et la sécurité, Harvard Law School et Brookings, <https://www.brookings.edu/wp-content/uploads/2016/06/0921_platform_security_wittes.pdf>
[610] http://www.magneto-h2020.eu/
[611] https://www.connexions-project.eu/
[612] https://redalertproject.eu/
[613] https://cordis.europa.eu/programme/id/H2020_SEC-12-FCT-2016-2017
[614] http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/secur/h2020-hi-guide-classif_en.pdf

,[615] permettant une mise en œuvre et une législation spécifiques dans les différents États membres.

Le développement de l'IA pour des objectifs de sécurité exige une prise en compte et un respect particulièrement attentifs et stricts des exigences éthiques en général, c'est-à-dire des orientations du programme Horizon 2020 déjà mentionnées - Comment réaliser votre auto-évaluation éthique, des documents clés respectifs liés à l'IA, par exemple le Groupe d'experts de haut niveau sur l'IA : "Lignes directrices en matière d'éthique pour une IA digne de confiance"[616] et le Livre blanc de la Commission européenne sur l'intelligence artificielle - Une approche européenne de l'excellence et de la confiance,[617] et d'autres considérations et documents spécifiques à la sécurité, tels que le tableau d'impact sociétal, l'avis n°28 du GEE - Éthique des technologies de sécurité et de surveillance[618] ou les documents pertinents publiés par le CEPD (Contrôleur européen de la protection des données).[619] La proposition de loi sur l'intelligence artificielle traite spécifiquement de l'utilisation des technologies d'IA à des fins répressives et "...établit une méthodologie de risque solide pour définir les outils d'IA à "haut risque" qui présentent des risques importants pour la santé et la sécurité ou les droits fondamentaux des personnes. Ces outils d'IA devront se conformer à un ensemble d'exigences horizontales obligatoires pour une IA digne de confiance et suivre des procédures d'évaluation de la conformité avant que ces systèmes puissent être mis sur le marché de l'Union."[620] L'annexe III énumère un certain nombre d'utilisations de l'IA pour le maintien de l'ordre en tant qu'outils d'IA à haut risque pour lesquels des procédures d'évaluation de la conformité sont obligatoires.

L'analyse étape par étape suivante suit la structure et la terminologie du modèle CRISP-DM[621] , comme indiqué dans la description ci-dessous. Afin d'accroître la comparabilité des approches et des résultats, cette structure est communément appliquée à toutes les études de cas présentées et discutées dans le cadre des MLE (Mutual Learning Encounters) menées par le PANELFIT. L'adoption d'une structure commune implique que les termes individuels ne doivent pas être compris littéralement. La compréhension commerciale peut, par exemple, signifier le développement d'une vision holistique des objectifs du projet et des moyens et étapes pour les atteindre dans le cas où le projet

---

[615] Parlement européen et Conseil, 2016, Directive (UE) 2016/680 du Parlement européen et du Conseil du 27 avril 2016 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel par les autorités compétentes à des fins de prévention et de détection des infractions pénales, d'enquêtes et de poursuites en la matière ou d'exécution de sanctions pénales, et à la libre circulation de ces données, et abrogeant la décision-cadre 2008/977/JAI du Conseil, Journal officiel <http://eur-lex.europa.eu/legal-content/EL/TXT/?uri=OJ:L:2016:119:TOC>.

[616] https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[617] https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

[618] Groupe européen d'éthique des sciences et des nouvelles technologies. (2014). Avis n° 28 : éthique des technologies de sécurité et de surveillance (10.2796/22379). Récupéré de Luxembourg : Bruxelles : https://publications.europa.eu/en/publication-detail/-/publication/6f1b3ce0-2810-4926-b185-54fc3225c969/language-en/format-PDF/source-77404258

[619] https://edps.europa.eu/data-protection/our-work/subjects_en

[620] Commission européenne. (2021). COM(2021) 206 final. Proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées en matière d'intelligence artificielle (loi sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union. , p. 3 <https://ec.europa.eu/newsroom/dae/redirection/document/75788>

[621] Shearer, Colin, Le modèle CRISP-DM : The New Blueprint for Data Mining, p. 14.

prévu n'a pas (principalement) d'intentions commerciales. Cela implique également que certaines des étapes ou des tâches incluses dans le cadre commun ne sont pas applicables ou moins pertinentes pour les différents contextes des études de cas. Par exemple, la première des quatre tâches principales composant l'objectif général, c'est-à-dire la détermination des objectifs commerciaux, se caractérise par une liberté de choix faible ou réduite si les objectifs sont définis et décrits dans un appel à soumettre des propositions de recherche, comme c'est le cas ici. Cette affirmation ne doit cependant pas laisser entendre que les libertés de choix n'existent pas du tout ou ne doivent pas être prises en compte, mais que les options disponibles pour les demandeurs de projets sont limitées par rapport à celles disponibles lors de la décision sur les sujets des appels à la recherche.

Lors de la discussion de la version préliminaire avec des experts externes, nous avons également reçu des recommandations allant au-delà de ce scénario spécifique, par exemple l'élaboration de programmes d'enseignement de l'éthique et leur intégration obligatoire dans les études techniques ou des offres de formation à la protection des données et à l'éthique pour les ingénieurs. Des programmes de formation correspondants devraient également être proposés aux forces de police (déployant l'IA) à titre d'activité de sensibilisation générale.

Analyse étape par étape

7 Compréhension de l'entreprise

7.1 Description

"La phase initiale de compréhension de l'entreprise se concentre sur la compréhension des objectifs du projet d'un point de vue commercial, en convertissant cette connaissance en une définition du problème d'exploration de données, puis en développant un plan préliminaire conçu pour atteindre les objectifs. Afin de comprendre quelles données doivent être analysées plus tard, et comment, il est vital pour les praticiens de l'exploration de données de comprendre pleinement l'entreprise pour laquelle ils trouvent une solution. La phase de compréhension de l'entreprise comprend plusieurs étapes clés, notamment la détermination des objectifs de l'entreprise, l'évaluation de la situation, la détermination des objectifs de l'exploration de données et la production du plan de projet."[622]

Dans le contexte de la R&D sur les technologies de prédiction et de prévention de la criminalité menée dans le cadre de H2020, la description générale et la structure des tâches doivent être adaptées en conséquence. Cela peut impliquer que la terminologie et le contenu concret de la tâche doivent être interprétés et modifiés pour répondre aux objectifs particuliers.

[622] Shearer, Colin, Le modèle CRISP-DM : The New Blueprint for Data Mining, p. 14.

Les objectifs généraux susmentionnés impliquent quatre tâches principales :

1. Déterminer les objectifs du projet. Cela signifie :
    a. Découvrir les objectifs principaux ainsi que les questions connexes auxquelles le projet (solution envisagée) voudrait répondre.
    b. Déterminer la mesure du succès.
2. Évaluer la situation
    a. Identifier les ressources disponibles pour le projet, tant matérielles que personnelles.
    b. Identifier les données disponibles pour atteindre l'objectif principal.
    c. Dresser la liste des hypothèses formulées dans le cadre du projet.
    d. Dresser la liste des risques du projet, énumérer les solutions potentielles à ces risques, créer un glossaire des termes relatifs au projet et au traitement des données, et réaliser une analyse coûts-avantages du projet.
3. Déterminer les objectifs du traitement des données : décider du niveau de précision prédictive attendu pour considérer le projet comme réussi.
4. Produire un plan de projet : Décrire le plan prévu pour atteindre les objectifs de traitement des données, y compris les étapes spécifiques et le calendrier proposé. Fournir une évaluation des risques potentiels et une évaluation initiale des outils et techniques nécessaires pour soutenir le projet.

7.2 Principales mesures à prendre

7.2.1 Définir les objectifs du projet

Pour notre scénario, les objectifs généraux sont définis par l'appel respectif. Les projets mentionnés ci-dessus se rapportent à l'appel SEC-12-FCT-2016-2017 : Technologies pour la prévention, l'investigation et l'atténuation dans le cadre de la lutte contre le crime et le terrorisme.[623] Le défi spécifique est décrit comme suit : "Les organisations criminelles et terroristes sont souvent à la pointe de l'innovation technologique pour planifier, exécuter et dissimuler leurs activités criminelles et les revenus qui en découlent. Les organismes chargés de l'application de la loi (LEA) sont souvent à la traîne lorsqu'ils s'attaquent à des activités criminelles soutenues par des technologies "avancées" ".

Le champ d'application de cet appel comprend :

- De nouvelles connaissances et des technologies ciblées pour lutter contre les formes anciennes et nouvelles de criminalité et les comportements terroristes, soutenues par des technologies avancées ;
- Le test et la démonstration de la nouvelle technologie développée par les LEA impliqués dans les propositions ;
- Les programmes d'études innovants, la formation et les exercices (conjoints) à utiliser pour faciliter l'adoption de ces nouvelles technologies dans toute l'UE, en particulier dans les domaines des sous-thèmes suivants :

[623] https://ec.europa.eu/research/participants/data/ref/h2020/wp/2016_2017/main/h2020-wp1617-security_en.pdf

1. cybercriminalité : monnaies virtuelles/crypto désanonymisation/traçage/altération lorsqu'elles soutiennent des marchés souterrains dans le darknet.

2. détection et neutralisation des drones légers/UAV malveillants survolant des zones restreintes, et impliquant comme bénéficiaires, le cas échéant, les opérateurs d'infrastructures.

3. analyse vidéo dans le contexte de l'enquête juridique

et un quatrième sous-thème ouvert.

Les conditions fixées dans le présent appel laissent une certaine marge de manœuvre, bien que limitée, pour la conception du projet. Les candidats sont libres de choisir le type de technologies ; cependant, les stratégies de solutions non techniques semblent ne pas être éligibles au financement. Même si l'éventail des technologies reste ouvert, l'appel exige clairement des solutions techniques, excluant ainsi les approches visant à résoudre les problèmes de sécurité spécifiques abordés sans l'implication de technologies potentiellement très intrusives. Le terme "technologies avancées" suggère au moins une enquête sur le développement et l'utilisation de technologies d'intelligence artificielle et d'apprentissage automatique. Le choix est également limité en ce qui concerne l'objectif, par exemple les formes de criminalité ou les comportements terroristes visés par le projet. Il est donc essentiel d'impliquer les utilisateurs finaux, c'est-à-dire les LEA (services répressifs), dès la phase de décision sur les objectifs et les moyens de les atteindre.

La sélection de technologies spécifiques ou, dans un contexte plus général, de méthodes spécifiques, influe également sur l'éventail des questions d'éthique ou de conformité juridique que pose le projet. Dans le cas de la recherche sur la sécurité, les technologies spécifiquement sélectionnées, dans notre cas des approches particulières d'IA ou d'apprentissage automatique, peuvent, outre les questions d'éthique habituelles comme le traitement des données à caractère personnel, soulever des problèmes d'éthique liés au double usage, à la concentration exclusive de la recherche sur des applications civiles ou à l'utilisation abusive, ce qui nécessite d'examiner les réglementations particulières correspondantes.

7.2.2 Opter pour des solutions techniques explicables et transparentes

Alors que l'explicabilité et la transparence constituent des exigences génériques pour les outils d'IA, elles forment des exigences obligatoires dans le cas des technologies d'IA appliquées aux humains ou ayant des conséquences pour eux (voir également la section "Licéité, loyauté et transparence" dans les "Principes" de la partie II). Dans le cas de l'IA utilisée pour le profilage ou l'aide à la décision dans un contexte de sécurité, ces principes sont fondamentaux. Les outils d'IA sont susceptibles d'être biaisés ; l'explicabilité et la transparence peuvent aider à détecter et à supprimer les biais des algorithmes créés par ces méthodes. Les technologies soutenant la prévention, la détection et la poursuite des infractions doivent fournir des résultats prouvables et attestables en tant que preuves valables, y compris devant les tribunaux. Des résultats inexacts peuvent avoir de graves conséquences pour les individus, notamment sous la forme de faux positifs ou d'issues fatales dans le cas de faux négatifs. Par conséquent, il peut être nécessaire de mettre en œuvre l'outil d'IA comme un soutien aux décisions prises par les humains, ainsi que des mesures obligatoires accompagnant l'emploi. Ainsi, il faut s'assurer que les responsables ne se contentent pas d'appliquer la

suggestion du système à leur propre décision, mais qu'ils comprennent qu'ils doivent justifier leur décision, qu'ils suivent la suggestion ou qu'ils s'opposent à une suggestion du système. Pour permettre aux humains de comprendre la suggestion d'un outil d'IA, ces systèmes doivent être très transparents quant aux facteurs qui influencent le résultat d'un calcul. En fin de compte, l'homme doit assumer la responsabilité d'une décision. La transparence est également essentielle pour garantir une compréhension suffisante du modèle et des données utilisés ainsi que des résultats produits, notamment en cas de plainte ou de besoin de preuve.

Les développeurs d'outils d'IA utilisés dans ce contexte pourraient faciliter la mise en œuvre en programmant des applications de soutien pour l'ensemble du processus de décision, par exemple en prévoyant un champ obligatoire à remplir lorsqu'une décision est prise sur la base d'une suggestion du système avant que le résultat puisse être traité ultérieurement.

7.2.3 Mise en œuvre d'un programme de formation

Dans notre cas, "la formation et les exercices (conjoints) à utiliser pour faciliter l'adoption de ces nouvelles technologies à l'échelle de l'UE" figurent déjà dans la description de l'appel. Ces exercices de formation ne doivent pas se limiter à l'utilisation des technologies développées, mais commencer au tout début des activités de recherche et, en particulier, comprendre toutes les personnes impliquées dans la conception des technologies d'IA (par exemple, les concepteurs d'algorithmes, les développeurs, les programmeurs, les codeurs, les scientifiques des données, les ingénieurs). Cette action est l'un des conseils essentiels à prendre en compte dès les premiers instants d'un projet de prédiction et de prévention de la criminalité. Les concepteurs d'algorithmes, qui occupent le premier maillon de la chaîne algorithmique, sont susceptibles de ne pas être conscients des implications éthiques et juridiques de leurs actions. L'un des principaux problèmes des outils d'IA consacrés à la lutte contre la criminalité et le terrorisme est qu'ils utilisent souvent des données personnelles incluses dans de grands ensembles de données, comprenant de grandes fractions de citoyens, par exemple les utilisateurs de réseaux sociaux spécifiques. Alors que l'analyse de données de surveillance de masse par des outils d'IA peut être autorisée par des juridictions nationales spécifiques ou des transpositions de la *directive relative à l'application de la loi sur la protection des données* (directive 2016/680), elle reste très problématique pour plusieurs raisons. Premièrement, la conformité juridique peut être une condition nécessaire à la conformité aux principes éthiques, mais ne peut jamais être considérée comme une condition suffisante. Un document d'information fourni par la commission européenne sur "l'éthique et la protection des données"[624] indique clairement que "le fait que certaines données soient accessibles au public ne signifie pas qu'il n'y a pas de limites à leur utilisation" (voir encadré 4, page 13). Deuxièmement, la conformité avec la législation nationale ou européenne n'implique pas nécessairement la conformité juridique avec les droits fondamentaux. La directive sur la conservation des données[625] est un exemple

---

[624] https://ec.europa.eu/info/sites/default/files/5._h2020_ethics_and_data_protection_0.pdf

[625] Directive 2006/24/CE du Parlement européen et du Conseil du 15 mars 2006 sur la conservation de données générées ou traitées dans le cadre de la fourniture de services de communications électroniques accessibles au public ou de réseaux publics de

connexe important, car elle a été annulée par la Cour de justice de l'Union européenne (CJUE) dans un arrêt du 8 avril 2014[626] parce que la Cour a estimé que la directive "comporte une ingérence de grande ampleur et particulièrement grave dans les droits fondamentaux au respect de la vie privée et à la protection des données à caractère personnel, sans que cette ingérence soit limitée au strict nécessaire". Troisièmement, l'opinion publique et l'acceptabilité par les citoyens doivent être respectées. Des consultations citoyennes à grande échelle sur les technologies de surveillance ont révélé que les citoyens acceptent en général les intrusions sérieuses dans leur vie privée si elles sont fondées sur des soupçons concrets et plausibles, mais rejettent les mesures de surveillance de masse non ciblées.[627] L'application de l'extraction de données pour détecter des activités criminelles ou terroristes peut être comparée à la recherche d'une aiguille dans une botte de foin[628] . Cela signifie également que le traitement comprendra des données à caractère personnel de personnes concernées qui ne sont pas actuellement ou n'ont pas été dans le passé impliquées dans des activités criminelles ou terroristes. En fonction du ciblage des données analysées, les données traitées peuvent concerner principalement ou presque exclusivement des personnes innocentes. Ce traitement de données viole la présomption d'innocence, modifie la relation entre les citoyens et l'État et peut avoir de graves conséquences sociétales et individuelles (en cas de faux positifs).

En tant que concepteur d'algorithmes, vous devez donc être capable de comprendre les implications de vos actions, tant pour les individus que pour la société, et être conscient de vos responsabilités en apprenant à faire preuve d'une attention et d'une vigilance constantes. Suivre ces conseils peut vous aider à éviter ou à atténuer de nombreux problèmes éthiques et juridiques. En ce sens, une formation optimale de toutes les personnes impliquées dans le projet, avant même qu'il ne démarre, pourrait être l'un des outils les plus efficaces pour économiser du temps et des ressources en termes de conformité avec la protection des données, l'éthique, le droit européen et national ou l'acceptabilité sociétale. Cela implique également la participation d'experts en éthique et en droit, tant dans les activités de formation que dans l'exécution du projet. Des mesures adéquates pour garantir la confidentialité méritent également une attention particulière (voir "Mesures en faveur de la confidentialité" dans la section "Intégrité et confidentialité" sous "Principes" dans la partie II). La sécurité et la confidentialité des données traitées, d'une part, sont essentielles ; la connaissance générale des types de données exploitées, des personnes concernées ou des algorithmes appliqués, d'autre part, est obligatoire pour garantir le respect des droits de l'Homme et des valeurs européennes. La conformité avec l'État membre le plus restrictif soutient également les objectifs de l'entreprise, en permettant la mise en œuvre et l'utilisation des systèmes développés sans qu'il soit nécessaire de procéder à des ajustements individuels.

---

communications, et modifiant la directive 2002/58/CE, Journal officiel de l'Union européenne <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32006L0024&from=en>.

626 https://www.europarl.europa.eu/legislative-train/theme-area-of-justice-and-fundamental-rights/file-data-retention-directive

627 Strauß, S. (2015). *D 6.10-Sommets citoyens sur la vie privée, la sécurité et la surveillance : Rapport de synthèse*. <http://surprise-project.eu/wp-content/uploads/2015/02/SurPRISE-D6.10-Synthesis-report.pdf>

628 Ce qui signifie également que la recherche d'un plus grand nombre de données ne fait qu'augmenter la botte de foin, pas nécessairement le nombre d'aiguilles.

7.2.4 Utilisation du cadre juridique applicable au traitement des données

Pour les projets de R&D liés à la sécurité, cette étape est particulièrement complexe et difficile. Pour le projet de recherche en tant que tel, la réglementation RGPD s'applique ; pour les mises en œuvre ultérieures, les règles et les dispositions de la *directive d'application de la loi sur la protection des données* (directive 2016/680) doivent être suivies. En outre, il convient de tenir compte des éventuelles divergences entre les législations des États (membres) concernés. Par conséquent, les technologies et systèmes développés doivent au moins prévoir une adaptabilité et une flexibilité pour faire face aux différentes réglementations. Du point de vue des droits de l'Homme et de l'éthique, la conformité avec les règles les plus restrictives devrait être incorporée dans les technologies créées, favorisant ainsi un respect maximal des droits fondamentaux et des valeurs connexes, tout en réduisant ou en éliminant, comme nous l'avons déjà mentionné, le besoin de modifications en cas d'application dans des pays où les réglementations sont divergentes.

Selon l'article 5, paragraphe 1, point a), du RGPD, les données à caractère personnel sont "collectées pour des finalités déterminées, explicites et légitimes et ne sont pas traitées ultérieurement de manière incompatible avec ces finalités". Le concept de légitimité n'est pas bien défini dans le RGPD, mais le groupe de travail Article 29 a déclaré que la légitimité implique que les données doivent être traitées "conformément à la loi", et que la "loi" doit être comprise comme un concept large qui inclut "toutes les formes de droit écrit et de common law, la législation primaire et secondaire, les décrets municipaux, les précédents judiciaires, les principes constitutionnels, les droits fondamentaux, les autres principes juridiques, ainsi que la jurisprudence, telle que cette "loi" serait interprétée et prise en compte par les tribunaux compétents".[629]

Il s'agit donc d'un concept plus large que la licéité. Elle implique le respect des principales valeurs des réglementations applicables et des grands principes éthiques en jeu. Par exemple, certains outils d'IA concrets nécessiteront l'intervention d'un comité d'éthique. Dans d'autres cas, des directives ou tout autre type de réglementation non contraignante peuvent être applicables. Vous devez veiller à respecter cette exigence en élaborant un plan pour cette étape préliminaire du cycle de vie de l'outil (voir "Légitimité et licéité" dans "Licéité, loyauté et transparence" sous "Principes" dans la partie II). À cette fin, vous devez être particulièrement attentif aux exigences posées par la réglementation applicable au niveau national. Le développement d'algorithmes liés à la prédiction et à la prévention de la criminalité nécessite clairement l'implication des comités d'éthique dès le début et, conformément à l'art. 35 du RGPD, une analyse de l'impact sur la protection des données doit être réalisée. Comme nous l'avons déjà mentionné, l'art. 10 du RGPD exige de vérifier si le traitement est autorisé par le droit de l'Union ou des États membres dans le cas du traitement de données à caractère personnel relatives à des condamnations pénales et à des infractions ou à des mesures de sécurité connexes. Assurez-vous que votre plan de recherche répond bien à toutes ces exigences pour les deux phases, la conduite du projet de recherche et les futures mises en œuvre des systèmes développés.

Le guide d'éthique fourni pour les recherches financées par l'UE (voir note de bas de page 606) constitue un cadre complet pour vérifier la conformité à l'éthique, qui doit

[629] Groupe de travail Article 29 (2013) Avis 03/2013 sur la limitation de la finalité Adopté le 2 avril 2013, WP203. Commission européenne, Bruxelles, p.20. Disponible sur : https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

être consulté en plus des règlements éthiques ou des codes de conduite institutionnels, que votre recherche reçoive ou non un financement de la CE. Sachez que l'évaluation de l'éthique n'est pas une activité qui se limite à une liste de contrôle, mais qu'elle comprend toujours une pesée des normes potentiellement conflictuelles. En particulier, l'application des TIC émergentes et la prise en compte du respect de la vie privée dès la conception dans un domaine aussi sensible exigent une réflexion prospective de la part des deux parties, les chercheurs concernés et les évaluateurs éthiques.

Même si votre projet ou votre institution de recherche n'est pas soumis à des réglementations éthiques spécifiques, il est essentiel d'observer et de respecter les réglementations nationales ou européennes pertinentes. Dès que vous mettez les technologies et systèmes développés sur le marché, la conformité est essentielle à la fois pour la mise en œuvre au sein de l'UE et pour l'obtention de licences d'exportation pour l'exploitation commerciale en dehors de l'UE.

7.2.5 Adopter une approche de réflexion fondée sur le risque

La création de votre algorithme impliquera probablement l'utilisation de plusieurs catégories particulières de données à caractère personnel, par exemple les opinions politiques, les convictions religieuses ou philosophiques ou les données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique dans le cas de l'exploration de données sur les réseaux sociaux. Par conséquent, vous devez vous assurer que vous mettez en œuvre des mesures appropriées pour minimiser les risques pour les droits, les intérêts et les libertés des personnes concernées (voir "Principe d'intégrité et de confidentialité" dans la partie II section "Principes" des présentes lignes directrices). À cette fin, vous devez évaluer les risques pour les droits et libertés des personnes participant au processus de recherche et de développement et juger de ce qui est approprié pour les protéger. Dans tous les cas, vous devez veiller au respect des exigences en matière de protection des données.

Dans le contexte des technologies de prédiction, de prévention, de détection ou d'investigation de la criminalité, une approche fondée sur le risque rend obligatoire une analyse d'impact sur la protection des données (AIPD), car au moins l'une des trois conditions spécifiques de l'art. 35(3) du RGPD s'applique nécessairement :

"*3. L'analyse d'impact sur la protection des données visée au paragraphe 1 est notamment requise dans les cas suivants :*

(a) une évaluation systématique et extensive d'aspects personnels concernant des personnes physiques, fondée sur un traitement automatisé, y compris le profilage, et sur laquelle sont fondées des décisions produisant des effets juridiques concernant la personne physique ou l'affectant de manière significative de façon similaire ;

(b) le traitement à grande échelle de catégories particulières de données visées à l'article 9, paragraphe 1, ou de données à caractère personnel relatives aux condamnations pénales et aux infractions visées à l'article 10 ; ou

(c) une surveillance systématique d'une zone accessible au public à grande échelle. "

L'analyse fondée sur les risques doit également inclure les questions d'éthique potentielles liées aux utilisations abusives[630] des technologies développées et au double usage[631] des restrictions à l'exportation qui peuvent s'appliquer aux systèmes développés.

Il faut également considérer que les risques ne se limitent pas aux impacts des systèmes développés sur la protection des données et la vie privée. Les droits constitutionnels et autres droits de l'Homme tels que la présomption d'innocence, l'égalité d'accès à la justice, la non-discrimination ou la liberté d'expression peuvent également être violés ou altérés. En outre, ces effets ne se limitent pas aux suspects potentiels, mais touchent la société dans son ensemble. Ils sont exacerbés par le manque de transparence et de contrôlabilité humaine de nombreux outils d'IA.

7.2.6 Préparer la documentation du traitement

Quiconque traite des données personnelles (qu'il s'agisse de responsables du traitement ou de sous-traitants) doit documenter ses activités, principalement à l'intention des autorités de contrôle compétentes. Vous devez le faire au moyen de registres du traitement qui sont conservés de manière centralisée par votre organisation pour l'ensemble de ses activités de traitement, et d'une documentation supplémentaire qui se rapporte aux activités individuelles de traitement des données (voir la section Documentation du traitement dans le chapitre Actions et outils). Cette étape préliminaire est le moment idéal pour mettre en place une méthode systématique de collecte de la documentation nécessaire, puisque c'est à ce moment-là que vous pourrez concevoir et planifier l'activité de traitement.

Le développement de votre outil d'IA peut impliquer l'utilisation de différents jeux de données. Les registres doivent assurer la traçabilité du traitement, l'information sur la réutilisation possible des données, et l'utilisation de données appartenant à différents jeux de données dans différentes ou dans les mêmes étapes du cycle de vie.

Pour les systèmes utilisés à des fins répressives, la documentation du traitement doit également comprendre la documentation de l'accès au système une fois celui-ci mis en œuvre, afin de prévenir et de détecter d'éventuelles utilisations abusives, par exemple l'accès non autorisé aux résultats générés.

Comme indiqué dans les Exigences et tests d'acceptation pour l'achat et/ou le développement du logiciel, du matériel et de l'infrastructure employés (sous-section de la section Documentation du traitement), l'évaluation des risques et les décisions prises *"doivent être documentées afin de se conformer à l'exigence de protection des données dès la conception (de l'article 25 du RGPD). En pratique, cela peut prendre la forme de :*

> ***Exigences** de protection des données spécifiques pour l'achat (par exemple, un appel d'offres) ou le développement de logiciels, de matériel et d'infrastructures,*

[630] Voir https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/guide_research-misuse_en.pdf

[631] Voir https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/guide_research-dual-use_en.pdf

Tests d'acceptation *qui vérifient que les logiciels, les systèmes et l'infrastructure choisis sont adaptés à l'usage prévu et offrent une protection et des garanties adéquates.*

Cette documentation doit faire partie intégrante de l'évaluation des risques et des opportunités de développement.

Enfin, vous devez toujours être conscient que, conformément à l'art. 32(1)(d) du RGPD, la protection des données est un processus. Par conséquent, **vous devez tester, évaluer et apprécier l'efficacité des mesures techniques et organisationnelles régulièrement.** Cette étape est le moment idéal pour construire une stratégie visant à relever ces défis.

7.2.7 Vérification du cadre réglementaire

Le RGPD comprend des règles spécifiques concernant le traitement à des fins de recherche scientifique (voir la section "Protection des données et recherche scientifique" du chapitre "Concepts principaux").[632] Votre outil d'IA pourrait être classé dans la catégorie de la recherche scientifique, indépendamment du fait qu'il soit créé dans un but lucratif ou non. *"Le droit de l'Union ou des États membres peut prévoir des dérogations aux droits visés aux articles 15, 16, 18 et 21, sous réserve des conditions et garanties visées au paragraphe 1 du présent article, dans la mesure où ces droits sont susceptibles de rendre impossible ou de nuire gravement à la réalisation des finalités spécifiques, et où ces dérogations sont nécessaires à la réalisation de ces finalités"* (article 89, paragraphe 2, du RGPD). En outre, selon l'article 5, point b), *"le traitement ultérieur des données collectées, conformément à l'article 89, paragraphe 1, ne serait pas considéré comme incompatible avec les finalités initiales ("limitation de la finalité")". Certaines autres exceptions particulières au cadre général applicable au traitement à des fins de recherche (telles que la limitation du stockage) devraient également être envisagées".*

Il est possible que vous puissiez bénéficier de ce cadre favorable, en fonction des pays où la recherche est menée et de la forme juridique des partenaires impliqués, par exemple s'il s'agit d'entités universitaires ou commerciales. Néanmoins, vous devez être conscient des réglementations (nationales) concrètes qui s'appliquent à cette recherche (principalement, les garanties à mettre en œuvre). Elles peuvent inclure des exigences spécifiques, en fonction des lois nationales respectives.

Être prudent implique également que vous devez tenir compte des limites juridiques et éthiques de la recherche envisagée. Ce n'est pas parce que des réglementations (nationales) spécifiques autorisent le traitement des données prévu qu'il est également acceptable ou conforme du point de vue de l'éthique. Par analogie, la conformité à l'éthique ne doit pas être utilisée à tort comme une échappatoire[633] aux réglementations.

---

[632] Ce cadre spécifique comprend également des objectifs de recherche historique ou des objectifs statistiques. Toutefois, la recherche sur les TIC n'est généralement pas liée à ces objectifs. Par conséquent, nous ne les analyserons pas ici.

[633] Wagner, B. (2018). L'éthique comme échappatoire à la réglementation : De l'ethics-washing à l'ethics-shopping ? Dans E. BayamlioĞLu, I. Baraliuc, L. Janssens, & M. Hildebrandt (Eds.), Being Profiled (pp. 84-89) : Amsterdam University Press.

7.2.8 Définition des politiques de stockage des données

Selon l'article 5, paragraphe 1, point e), du RGPD, les données à caractère personnel doivent être *"conservées sous une forme permettant l'identification des personnes concernées pendant une durée n'excédant pas celle nécessaire à la réalisation des finalités pour lesquelles elles sont traitées"*. Cette exigence est double. D'une part, elle concerne l'identification : les données doivent être conservées sous une forme permettant l'identification des personnes concernées pendant une durée n'excédant pas celle nécessaire. Par conséquent, vous devez mettre en œuvre des politiques visant à éviter l'identification dès qu'elle n'est pas nécessaire au traitement. Ces politiques impliquent l'adoption de mesures adéquates pour garantir qu'à tout moment, seul **le degré minimal d'identification nécessaire à la réalisation des finalités doit être utilisé** (voir la sous-section "Aspect temporel" de la section "Principe de limitation du stockage" de la partie II, section "Principes" des présentes lignes directrices).

D'autre part, le stockage des données implique que les données ne peuvent être conservées que pendant une **période limitée** : le temps strictement nécessaire aux fins pour lesquelles les données sont traitées. Toutefois, le RGPD autorise "le stockage pour des périodes plus longues si la seule finalité est la recherche scientifique" (ce qui pourrait être le cas pour la phase de R&D).

L'exception relative à la recherche scientifique augmente le risque que vous décidiez de conserver les données plus longtemps que ce qui est strictement nécessaire. Vous devez être conscient que même si le RGPD peut autoriser le stockage pour des périodes plus longues, **vous devez avoir des raisons justifiables d'opter pour une telle période prolongée.** Pour les systèmes développés, vous devez inclure des précautions organisationnelles et techniques pour pouvoir vous conformer aux différentes réglementations légales nationales concernant les périodes maximales de stockage des données. Ce pourrait également être le moment idéal pour **envisager des délais pour l'effacement (automatique) de différentes catégories de données et pour documenter ces décisions** (voir "Principe de responsabilité" dans la partie II, section "Principes" des présentes lignes directrices).

7.2.9 Nomination d'un délégué à la protection des données

Conformément à l'art. 37(1) du RGPD, vous devez désigner un DPD :

" 1) Le responsable du traitement et le sous-traitant désignent un délégué à la protection des données dans tous les cas où :

(a) le traitement est effectué par une autorité ou un organe public, à l'exception des tribunaux agissant dans l'exercice de leurs fonctions judiciaires ;

(b) les activités principales du responsable du traitement ou du sous-traitant consistent en des opérations de traitement qui, en raison de leur nature, de leur portée et/ou de leurs finalités, requièrent un suivi régulier et systématique des personnes concernées à grande échelle ; ou

(c) les activités principales du responsable du traitement ou du sous-traitant consistent à traiter à grande échelle des catégories particulières de données conformément à l'article 9 et des données à caractère personnel relatives aux condamnations pénales et aux infractions visées à l'article 10."

8 Compréhension des données

8.1 Description

"La phase de compréhension des données commence par une collecte initiale des données. L'analyste procède ensuite à une familiarisation accrue avec les données, à l'identification des problèmes de qualité des données, à la découverte d'aperçus initiaux sur les données, ou à la détection de sous-ensembles intéressants pour former des hypothèses sur des informations cachées. La phase de compréhension des données comporte quatre étapes, à savoir la collecte des données initiales, la description des données, l'exploration des données et la vérification de la qualité des données".[634]

Toutes ces étapes visent à identifier les données disponibles. À ce stade, vous devez être conscient des données avec lesquelles vous devrez travailler et commencer à prendre des décisions sur la manière dont les grands principes liés à la protection des données seront mis en œuvre. Vous devez consulter le document Éthique et protection des données du 14 novembre 2018[635] pour vous conformer aux exigences légales et éthiques. Dans le cas de l'utilisation de données issues de réseaux sociaux, les informations fournies dans l'encadré 4 Utiliser des données "open source", page 13, sont particulièrement pertinentes.

Vous devez également savoir que les bases de données qui contiennent des données personnelles sur les poursuites liées à des condamnations pénales et à des infractions sont sensibles et que vous, en tant que développeur, ne pourrez normalement pas y accéder.

8.2 Principales mesures à prendre

À ce stade, un grand nombre de questions fondamentales liées à la protection des données personnelles doivent être abordées. En fonction des décisions prises, des principes tels que la minimisation des données, le respect de la vie privée dès la conception ou par défaut, la licéité, la loyauté et la transparence, etc. seront réglés de manière adéquate. Une communication entre les experts éthiques et juridiques, d'une part, et les développeurs de projets, d'autre part, doit être établie pour pouvoir réaliser les principes de "vie privée dès la conception" ou "par défaut".

8.2.1 Prise de décision sur les types de données à traiter

Selon le RGPD, le "responsable du traitement *met en œuvre les mesures techniques et organisationnelles appropriées pour garantir que, par défaut, seules les données à caractère personnel qui sont nécessaires à chaque finalité spécifique du traitement sont traitées. Cette obligation s'applique à la quantité de données à caractère personnel collectées, à l'étendue de leur traitement, à la durée de leur conservation et à leur*

[634] Colin Shearer, Le modèle CRISP-DM : Le nouveau plan directeur pour l'extraction de données, p. 15
[635] https://ec.europa.eu/info/sites/info/files/5._h2020_ethics_and_data_protection_0.pdf

accessibilité. En particulier, ces mesures garantissent que, par défaut, les données à caractère personnel ne sont pas rendues accessibles sans l'intervention de la personne concernée à un nombre indéfini de personnes physiques."[636] (Voir Protection des données dès la conception et par défaut dans le chapitre Concepts) Cette exigence doit être spécialement gardée à l'esprit au cours de cette étape, car c'est souvent à ce moment que sont prises les décisions concernant le type de données qui seront utilisées.

Il faut donc s'assurer que vous avez vraiment besoin de grandes quantités de données. Des "données intelligentes" ciblées pourraient être beaucoup plus utiles que des données volumineuses. Bien sûr, l'utilisation de données intelligentes et bien préparées peut impliquer un effort considérable en termes d'unification, d'homogénéisation, etc., mais elle aidera à mettre en œuvre le principe de minimisation des données (voir "Principe de minimisation des données" dans la partie II, section "Principes" des présentes lignes directrices) de manière beaucoup plus efficace. À cette fin, il est **essentiel de disposer d'une expertise pour sélectionner les caractéristiques pertinentes.** Cette étape consiste également à vérifier la nécessité du traitement pour chaque catégorie de données ; cela implique de prouver qu'aucune mesure ou méthode alternative, moins attentatoire du point de vue de la protection des données et des droits de l'Homme, ne pourrait être appliquée pour atteindre le même résultat.

En outre, vous devez essayer de **limiter la résolution des données** à ce qui est minimalement nécessaire aux fins poursuivies par le traitement. Vous devez également **déterminer un niveau optimal d'agrégation des données** avant de commencer le traitement (voir la section "Partie adéquate, pertinente et limitée de la minimisation des données" du chapitre "Principes"). Dans le cas de l'IA appliquée à la prédiction, la prévention ou l'investigation de la criminalité, le niveau possible d'agrégation des données, c'est-à-dire l'anonymisation des données, est sans aucun doute limité, du moins pour les implémentations et utilisations ultérieures des systèmes développés. L'objectif premier étant d'identifier les auteurs (potentiels) de crimes, il doit au moins être possible de (re)personnaliser les données sur les menaces potentielles.

La minimisation des données peut être compliquée dans le cas de l'apprentissage profond, où la différenciation par caractéristiques peut être impossible. Il existe un moyen efficace de réguler la quantité de données recueillies et de ne l'augmenter que si cela semble nécessaire : la courbe d'apprentissage. Vous devez commencer par collecter et utiliser une quantité limitée de données d'apprentissage, puis surveiller la précision du modèle à mesure qu'il est alimenté en nouvelles données.

8.2.2 Vérification de l'utilisation légitime des jeux de données

Les ensembles de données peuvent être obtenus de différentes manières. Tout d'abord, le développeur peut opter pour l'acquisition ou l'accès à une base de données qui a déjà été construite par quelqu'un d'autre. Si c'est le cas, vous devez être particulièrement prudent car l'acquisition de l'accès à une base de données soulève de nombreuses questions juridiques (voir la section "Achat de l'accès à une base de données" du chapitre "Principaux outils et actions"). [637]

[636] Article 25(2).
[637] Yeong Zee Kin, Legal Issues in AI Deployment, à l'adresse : https://lawgazette.com.sg/feature/legal-issues-in-ai-deployment/ consulté le 15 mai 2020.

Ensuite, l'alternative la plus courante consiste à créer une base de données. Bien évidemment, dans ce cas, vous devez vous assurer que vous respectez toutes les exigences légales imposées par le RGPD pour créer une base de données (voir la section "Créer une base de données" dans le chapitre "Principaux outils et actions").

Troisièmement, vous pouvez choisir une autre voie. Vous pouvez mélanger des données sous licence provenant de tiers avec votre propre ensemble de données de manière à créer un énorme ensemble de données de formation et un autre à des fins de validation. Cela peut poser certains problèmes, notamment la possibilité que la combinaison de différents ensembles de données fournisse des informations supplémentaires sur les personnes concernées. Par exemple, cela pourrait vous permettre d'identifier les personnes concernées, ce qui n'était pas possible auparavant, en utilisant un seul des ensembles de données. Cela pourrait impliquer la désanonymisation de données anonymes et la création de nouvelles informations personnelles qui ne figuraient pas dans l'ensemble de données initial. Cette situation entraînerait d'importants problèmes éthiques et juridiques. Par exemple, *"si les personnes concernées ont donné leur consentement éclairé au traitement des informations personnelles contenues dans les ensembles de données d'origine à des fins particulières, elles n'ont pas nécessairement donné leur autorisation par extension à la fusion d'ensembles de données et à l'exploration de données qui révèle de nouvelles informations. Les nouvelles informations produites de cette manière peuvent également être basées sur des probabilités ou des conjectures, et donc être fausses, ou contenir des biais dans la représentation des personnes."*[638] Par conséquent, vous devriez essayer d'éviter de telles conséquences en vous assurant que la fusion des ensembles de données ne va pas à l'encontre des droits et des intérêts des personnes concernées.

Enfin, si vous utilisez plusieurs ensembles de données qui poursuivent des objectifs différents, vous devez mettre en œuvre des mesures adéquates pour séparer les différentes activités de traitement. Sinon, vous pourriez facilement utiliser des données dans un but pour lequel elles n'ont pas été collectées. Cela pourrait poser des problèmes liés au principe de limitation de la finalité (voir "Principe de limitation de la finalité" dans la partie II, section "Principes" des présentes lignes directrices).

Sachez que les mesures susmentionnées ne sont suffisantes que pour la phase d'exécution du projet de recherche. Le consentement éclairé sera généralement d'une utilité très limitée dans le cadre d'une activité répressive. Il en va de même pour la création et l'utilisation de données factices ou synthétiques. L'utilisation de données synthétiques peut toujours poser des problèmes de ré-identification potentielle, ainsi que la question de savoir si l'on peut faire confiance à ces données lors de l'entraînement d'algorithmes d'IA. Toutes ces mesures peuvent effectivement contribuer à atténuer ou à éliminer les problèmes éthiques ou juridiques de la phase de recherche. Il est essentiel de s'assurer que les ensembles de données nécessaires aux mises en œuvre dans le monde réel sont également conformes aux exigences éthiques et juridiques imposées par les réglementations de l'UE et des États membres ; cela vaut également pour l'utilisation d'ensembles de données appartenant à la police ou au gouvernement. Sachez également qu'il peut s'avérer difficile, voire impossible, d'accéder à des ensembles de données réelles de taille suffisante pour la formation pratique de l'outil d'IA.

---

[638] SHERPA, Lignes directrices pour le développement éthique des systèmes d'IA et de Big Data : Une approche d'éthique par la conception, 2020, p 38. À l'adresse : https://www.project-sherpa.eu/wp-content/uploads/2019/12/development-final.pdf Consulté le 15 mai 2020

8.2.3 Sélection de la base juridique appropriée pour le traitement

Vous devez décider de la base juridique que vous utiliserez pour le traitement avant de le commencer, documenter votre décision (ainsi que les finalités) et inclure les raisons pour lesquelles vous avez fait votre choix (voir "Principe de responsabilité" dans la partie II, section "Principes" des présentes lignes directrices).

Vous devez choisir la base juridique qui reflète le mieux la véritable nature de votre traitement des données à caractère personnel. Si des participants humains sont impliqués, il faut également tenir compte de la relation avec les participants et de la finalité du traitement. Cette décision est essentielle, car il n'est pas possible de changer la base juridique du traitement s'il n'y a pas de raisons solides qui le justifient (voir la section Limitation de la finalité dans le chapitre Principes).

Dans le cas d'outils d'IA développés à des fins de prédiction ou de prévention de la criminalité, etc. vous devez à nouveau faire la distinction entre la phase de recherche et les mises en œuvre ultérieures. Pour la phase de recherche, vous pouvez être en mesure d'utiliser le consentement comme fondement juridique du traitement, en fonction de l'implication concrète de participants humains. Il peut s'agir, par exemple, d'outils d'IA utilisant l'identification biométrique ou l'interprétation de données vidéo, qui nécessitent l'intervention de participants humains pour les tests. Le consentement pourrait également constituer un fondement juridique valable si vous réutilisez des données qui ont déjà été recueillies à une autre fin et que le consentement était la base qui permettait l'utilisation primaire des données. Le RGPD autorise la réutilisation des données à des fins scientifiques et l'article 5.1 (b) stipule que le traitement ultérieur à des fins de recherche scientifique ne doit pas être considéré comme incompatible avec les finalités initiales ("limitation de la finalité"). Ainsi, en principe, vous pourriez réutiliser ces données sur la base du consentement initial. Cependant, vous devez garder à l'esprit que, selon l'article 9.4 du RGPD, *"les États membres peuvent maintenir ou introduire des conditions supplémentaires, y compris des limitations, en ce qui concerne le traitement des données génétiques, des données biométriques ou des données relatives à la santé."* Ainsi, il se pourrait bien que votre réglementation nationale pertinente introduise des exceptions ou des conditions spécifiques à la réutilisation des données personnelles. En tout état de cause, vous devez toujours vous rappeler que vos devoirs d'information demeurent. Vous devez fournir à la personne concernée, avant tout traitement ultérieur de ses données, des informations sur cette autre finalité et toute autre information pertinente visée au paragraphe 2 de l'article 13 du RGPD.

Veuillez garder à l'esprit que les dispositions ci-dessus ne s'appliquent qu'à la conduite de la recherche en tant que telle. Les utilisations futures des systèmes développés doivent être conformes à la législation en vigueur dans l'UE et dans les États membres concernant les activités répressives. Soyez également conscient que le développement de technologies qui ne sont pas conformes aux réglementations applicables ou aux principes éthiques ou aux valeurs européennes impliquerait un gaspillage d'efforts et de ressources.

8.2.4 Réutilisation des données

Actuellement, la réutilisation des données à des fins de recherche fait l'objet d'un débat animé. Selon l'article 5.1 (b) du RGPD, le traitement ultérieur à des fins scientifiques ne doit pas être considéré comme incompatible avec les finalités initiales. Ainsi, à moins que votre réglementation nationale ne stipule le contraire, vous pouvez réutiliser les

données disponibles à des fins de recherche, puisque celles-ci sont compatibles avec la finalité initiale pour laquelle elles ont été collectées.

Toutefois, le CEPD fait valoir que, *"afin de garantir le respect des droits de la personne concernée, le test de compatibilité prévu à l'article 6, paragraphe 4, devrait toujours être pris en considération avant la réutilisation des données aux fins de la recherche scientifique, en particulier lorsque les données ont été initialement collectées pour des finalités très différentes ou en dehors du domaine de la recherche scientifique. En effet, selon une analyse du point de vue de la recherche médicale, l'application de ce test devrait être simple".*[639] Selon cette interprétation, vous ne devez réutiliser les données à caractère personnel que si les circonstances de l'article 6.4 s'appliquent. Veuillez vérifier dans ce contexte également l'applicabilité de l'article 10 *"Le traitement des données à caractère personnel relatives aux condamnations pénales et aux infractions ou aux mesures de sûreté connexes fondé sur l'article 6, paragraphe 1, n'est effectué que sous le contrôle de l'autorité publique ou lorsque le traitement est autorisé par le droit de l'Union ou des États membres qui prévoit des garanties appropriées pour les droits et libertés des personnes concernées."*

9 Préparation des données

9.1 Description

"La phase de préparation des données couvre toutes les activités visant à construire l'ensemble de données final ou les données qui seront introduites dans le ou les outils de modélisation à partir des données brutes initiales. Les tâches comprennent la sélection des tables, des enregistrements et des attributs, ainsi que la transformation et le nettoyage des données pour les outils de modélisation. Les cinq étapes de la préparation des données sont la sélection des données, le nettoyage des données, la construction des données, l'intégration des données et le formatage des données."[640]

Cette étape comprend toutes les activités nécessaires pour construire l'ensemble de données final qui est introduit dans le modèle, à partir des données brutes initiales. Elle comprend les cinq tâches suivantes, qui ne sont pas nécessairement exécutées de manière séquentielle :

1. Sélectionner les données : Décider des données à utiliser pour l'analyse, en fonction de leur pertinence par rapport aux objectifs de l'exploration de données, de leur qualité et des contraintes techniques telles que les limites du volume ou des types de données.
2. Nettoyer les données : Amener la qualité des données à un niveau requis, par exemple en sélectionnant des sous-ensembles de données propres, en insérant des valeurs par défaut et en estimant les données manquantes par modélisation.
3. Construire des données : La construction de nouvelles données par la production d'attributs dérivés, de nouveaux enregistrements ou de valeurs transformées pour des attributs existants.

[639] CEPD, un avis préliminaire sur la protection des données et la recherche scientifique, 6 janvier 2020, p. 23.
[640] Colin Shearer, Le modèle CRISP-DM : The New Blueprint for Data Mining, p. 16.

4. Intégrer des données : Combiner les données de plusieurs tables ou enregistrements pour créer de nouveaux enregistrements ou valeurs.
5. Formater les données : Apporter des modifications syntaxiques aux données qui pourraient être requises par l'outil de modélisation.

9.2 Principales mesures à prendre

9.2.1 Introduction des mesures de protection prévues à l'article 89 du RGPD

Puisque vous utilisez des données à des fins scientifiques, vous devez les préparer selon les garanties prévues par le RGPD à l'article 89. Si les finalités de votre recherche peuvent être atteints par un traitement ultérieur qui ne permet pas ou plus l'identification des personnes concernées, c'est-à-dire par la pseudonymisation, ces finalités doivent être atteints de cette manière. Si cela n'est pas possible, vous devez introduire des garanties assurant que les mesures techniques et organisationnelles permettent une mise en œuvre adéquate du principe de minimisation des données. Veuillez prendre en considération les règles concrètes établies par votre réglementation nationale concernant les garanties. Consultez votre DPD.

9.2.2 Garantir la précision du traitement des données à caractère personnel

Selon le RGPD, les données doivent être exactes (voir la section "Précision" du chapitre "Principes"). Cela signifie que les données de traitement sont correctes et à jour. Les responsables du traitement sont chargés de garantir la précision. Par conséquent, une fois que vous avez terminé la collecte des données, vous devez mettre en place des outils adéquats pour garantir la précision des données. Cela implique généralement que vous deviez prendre certaines décisions fondamentales sur les mesures techniques et organisationnelles qui rendront ce principe applicable (voir la sous-section "Mesures techniques et organisationnelles connexes" dans la section "Précision" du chapitre "Principes"). Étant donné que la plupart des données proviennent de sources probablement très différentes, sans exigences de qualité normalisées, et que la plupart d'entre elles seront probablement qualitatives dans le cas de la prédiction de la criminalité, vous ne pouvez pas supposer qu'elles sont exactes en soi. En effet, ces données peuvent être basées sur des évaluations individuelles de différentes personnes, alors que les personnes concernées ne savent peut-être même pas que ce type de données est stocké à leur sujet.

En tout état de cause, la précision exige une mise en œuvre adéquate des mesures destinées à faciliter le droit de rectification des personnes concernées (voir (voir "Droit de rectification" dans la partie II, section "Droits des personnes concernées" des présentes lignes directrices).

Assurez-vous également qu'ils produisent des résultats aussi précis que possible. Les types de faux positifs et de faux négatifs doivent être définis à l'avance lors de la phase de préparation des données. Les faux résultats sont l'une des questions essentielles ayant un impact sur les droits fondamentaux des individus.

9.2.3 Se concentrer sur les questions de profilage

En général, dans le cas d'une base de données qui servira à former ou à valider un outil d'IA, il existe une obligation particulièrement pertinente d'informer les personnes concernées que **leurs données pourraient donner lieu à une prise de décision automatisée ou à un profilage les concernant.** Le profilage est particulièrement problématique dans le développement de l'IA, cela vaut également pour les outils d'IA développés à des fins de LEA.

Selon l'article 22, paragraphe 2, point c), les décisions automatisées qui portent sur des catégories particulières de données à caractère personnel, telles que les *données qui révèlent l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données relatives à la santé ou des données relatives à la vie sexuelle ou à l'orientation sexuelle d'une personne physique* (article 9, paragraphe 1) ne sont autorisées que si la personne concernée a donné son consentement ou si elles sont fondées sur une base juridique. Cette exception s'applique non seulement lorsque les données observées entrent dans cette catégorie, mais **aussi si le rapprochement de différents types de données à caractère personnel peut révéler des informations sensibles sur des personnes ou si des données déduites entrent dans cette catégorie.** Dans le cas de la prédiction et de la prévention de la criminalité, le consentement explicite des personnes concernées ne s'appliquera normalement qu'aux participants humains volontaires pendant la phase de recherche et développement. Le traitement de catégories particulières de données à caractère personnel, par exemple les opinions politiques ou les croyances religieuses, peut faire partie du noyau de données des outils d'IA appliqués dans le domaine de la prévention du terrorisme.

Certaines actions supplémentaires peuvent être extrêmement utiles pour éviter la prise de décision automatisée si elle n'est pas nécessaire :

- Tenir compte des exigences du système nécessaires pour soutenir un examen humain significatif **dès la phase de conception.** En particulier, les exigences d'interprétabilité et la conception d'une interface utilisateur efficace pour soutenir les examens et les interventions humaines ;
- Concevoir et offrir une formation et un soutien appropriés aux examinateurs humains ; et
- Donner au personnel l'autorité, les incitations et le soutien appropriés pour répondre aux préoccupations des personnes ou les transmettre à un échelon supérieur et, si nécessaire, passer outre la décision de l'outil d'IA.[641]

Si vous procédez à un profilage ou à des décisions automatisées, vous devez informer les personnes concernées de votre décision et fournir toutes les informations nécessaires conformément au RGPD et à la réglementation nationale, le cas échéant.

9.2.4 Sélection de données non biaisées

La partialité est l'un des principaux problèmes liés au développement de l'IA, un problème qui va à l'encontre du principe de loyauté (voir "Principe de licéité, de loyauté et de transparence" dans la partie II, section "Principes" des présentes lignes directrices).

[641] https://ico.org.uk/about-the-ico/news-and-events/ai-blog-fully-automated-decision-making-ai-systems-the-right-to-human-intervention-and-other-safeguards/

Les biais peuvent être causés par de nombreux facteurs différents. Lorsque des données sont recueillies, elles peuvent contenir des biais, des inexactitudes, des erreurs et des fautes construits par la société. Parfois, il peut arriver que les ensembles de données soient biaisés en raison d'actions malveillantes. L'introduction de données malveillantes dans un outil d'IA peut modifier son comportement, en particulier avec les systèmes d'auto-apprentissage.[642] Par conséquent, les questions liées à la composition des bases de données utilisées pour la formation soulèvent des problèmes éthiques et juridiques cruciaux, et pas seulement des questions d'efficacité ou de nature technique.

Vous devez résoudre ces problèmes avant de former l'algorithme. Les biais identifiables et discriminatoires doivent être supprimés dans la mesure du possible lors de la phase de constitution des ensembles de données. Comme nous l'avons vu par le passé, l'idée que certains groupes de personnes (Noirs, Arabes ou étrangers en général, musulmans...) sont plus souvent condamnés parce qu'ils enfreignent la loi plus fréquemment dans la plupart des cas n'est pas valable. Ils sont plus souvent fouillés, plus souvent discriminés par la police, plus souvent confrontés à une violence excessive, à l'arbitraire ou à l'hostilité de la police et se retrouvent donc plus souvent dans des situations problématiques. Cette observation serait très probablement valable pour tout autre sous-ensemble de la population s'il était traité de la même manière. Par conséquent, déduire un taux de criminalité plus élevé dans les zones où vivent de nombreux étrangers pourrait devenir une prophétie auto-réalisatrice.

Un autre exemple pourrait être l'hypothèse selon laquelle un outil d'IA produit les bons résultats dès qu'ils correspondent aux résultats obtenus par les humains. Or, les décisions humaines sont souvent biaisées, et l'outil d'IA perpétuerait très probablement ces pratiques discriminatoires au lieu de produire des résultats plus objectifs.

Si l'algorithme est biaisé, il peut également augmenter le nombre de faux positifs ou de faux négatifs. Les faux positifs peuvent avoir des effets négatifs graves sur les personnes concernées, les faux négatifs sur la société et, bien sûr, sur les victimes d'activités criminelles ou terroristes qui auraient pu être évitées.

Vous devez vous assurer que l'algorithme évalue ces facteurs en conséquence lorsque vous sélectionnez les données. Cela signifie que **les équipes chargées de sélectionner les données à intégrer dans les jeux de données doivent être composées de personnes qui garantissent la diversité dont l'outil d'IA est censé faire preuve.** Enfin, gardez toujours à l'esprit que, si vos données concernent principalement un groupe concret, vous devez déclarer que l'algorithme a été formé sur cette base et qu'il pourrait donc ne pas fonctionner aussi bien dans d'autres groupes de population.

10 Modélisation (formation)

10.1 Description

[642] Groupe d'experts de haut niveau sur l'IA, Lignes directrices en matière d'éthique pour une IA digne de confiance, 2019, p. 17. À l'adresse : https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai Consulté le 15 mai 2020

"Dans cette phase, diverses techniques de modélisation sont sélectionnées et appliquées et leurs paramètres sont calibrés à des valeurs optimales. Généralement, plusieurs techniques existent pour le même type de problème d'exploration de données. Certaines techniques ont des exigences spécifiques sur la forme des données. Par conséquent, il peut être nécessaire de revenir à la phase de préparation des données. Les étapes de modélisation comprennent la sélection de la technique de modélisation, la génération du plan de test, la création de modèles et l'évaluation des modèles."[643]

Cette phase comporte plusieurs tâches essentielles. Dans l'ensemble, vous devez

- Sélectionner la technique de modélisation qui sera utilisée. Selon le type de technique, des conséquences telles que l'inférence des données, l'obscurité ou les biais sont plus ou moins susceptibles de se produire.
- Prendre une décision sur l'outil de formation à utiliser. Cela permet au développeur de mesurer la capacité du modèle à prédire l'histoire avant de l'utiliser pour prédire l'avenir. Dans le cas de la prédiction de la criminalité, cela pourrait constituer un problème en soi. Ce n'est pas comme prédire que quelqu'un qui aime les yaourts en achètera à nouveau. Nous parlons d'êtres humains et de leurs chances dans la vie. En supposant qu'une personne récidivera parce qu'elle a fait quelque chose d'illégal dans le passé, on néglige presque le fait que nous considérons les citoyens comme des êtres humains dotés de libre arbitre et de la possibilité de prendre une meilleure décision la prochaine fois. Il est intrinsèquement problématique de supposer que l'avenir sera une extrapolation du passé. En fonction des conséquences individuelles et sociétales, cela peut être moins problématique dans certains cas et injustifiable dans d'autres.

  La formation implique toujours la réalisation de tests empiriques avec des données. Parfois, les développeurs testent le modèle avec des données différentes de celles utilisées pour le générer. Par conséquent, à ce stade, on peut parler de différents types d'ensembles de données.

10.2 Principales mesures à prendre

10.2.1 Mise en œuvre du principe de minimisation des données

Selon le principe de minimisation des données, vous devez procéder à la réduction de la quantité de données et/ou de l'éventail d'informations sur la personne concernée qu'ils fournissent dès que possible. Par conséquent, vous devez purger les données utilisées pendant la phase d'entraînement de toutes les informations qui ne sont pas strictement nécessaires à l'entraînement du modèle. (voir la sous-section "Aspect temporel" dans "Minimisation des données" dans "Principes" dans la partie II. Il existe de multiples stratégies pour assurer la minimisation des données lors de la phase de formation. Les techniques évoluent en permanence. Toutefois, certaines des plus courantes sont[644] (voir "Principe d'intégrité et de confidentialité" dans la section "Principes" de la partie II des présentes lignes directrices) :

---

[643] Colin Shearer, Le modèle CRISP-DM : The New Blueprint for Data Mining, p. 17.

[644] AEPD, Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción, 2020, p.40. À l'adresse : https://www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf Consulté le 15 mai 2020.

- Analyse des conditions que les données doivent remplir pour être considérées comme de haute qualité et dotées d'une grande capacité de prédiction pour l'application spécifique.
- Analyse critique de l'étendue de la typologie des données utilisées à chaque étape de la solution d'IA.
- Suppression des données non structurées et des informations inutiles recueillies lors du prétraitement de l'information.
- Identification et suppression des catégories de données qui n'ont pas d'influence significative sur l'apprentissage ou sur le résultat de l'inférence.
- Suppression des conclusions non pertinentes associées aux informations personnelles pendant le processus de formation, par exemple, dans le cas d'une formation non supervisée.
- Utilisation de techniques de vérification qui nécessitent moins de données, comme la validation croisée.
- Analyse et configuration des hyperparamètres algorithmiques qui pourraient influencer la quantité ou l'étendue des données traitées afin de les minimiser
- Utilisation de modèles d'apprentissage fédérés plutôt que centralisés
- Application de stratégies de confidentialité différentielle.
- Entraînement avec des données cryptées en utilisant des techniques homomorphiques.
- Agrégation de données.
- Anonymisation et pseudonymisation, non seulement dans la communication des données, mais aussi dans les données de formation, les éventuelles données personnelles contenues dans le modèle et dans le traitement de l'inférence.

10.2.2 **Détecter et effacer les biais**

Même si les mécanismes de lutte contre les biais sont convenablement adoptés lors des étapes précédentes (voir la section sur la formation ci-dessus), il faut encore s'assurer que les résultats de la phase de formation minimisent les biais. Cela peut être difficile car certains types de biais et de discrimination sont souvent particulièrement difficiles à détecter. Les membres de l'équipe qui traite les données d'entrée n'en sont parfois pas conscients, et les utilisateurs qui sont leurs sujets n'en sont pas nécessairement conscients non plus. Ainsi, les systèmes de contrôle mis en place par le développeur d'IA lors de la phase de validation sont des facteurs extrêmement importants pour éviter les biais.

Il existe de nombreux outils techniques qui peuvent servir à détecter les biais, comme l'évaluation de l'impact algorithmique.[645] Il faut envisager leur mise en œuvre effective.[646] Cependant, comme le montre la littérature,[647] il peut arriver qu'un

[645] Reisman, D., Crawford, K., Whittaker, M., Algorithmic impact assessments : Un cadre pratique pour la responsabilité des agences publiques, 2018, à l'adresse : https://ainowinstitute.org/aiareport2018.pdf Consulté le 15 mai 2020.

[646] https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf consulté le 15 mai 2020

[647] Chouldechova. Alexandra, Fair Prediction with Disparate Impact : Une étude des biais dans les instruments de prédiction de la récidive, Big Data. Volume : 5 Numéro 2 : 1er juin 2017. 153-163.http://doi.org/10.1089/big.2016.0047

algorithme ne puisse être totalement purgé de tous les différents types de biais. Vous devez cependant essayer d'être au moins conscient de leur existence et des implications que cela peut entraîner (voir "Principe de licéité, de loyauté et de transparence" dans la partie II, section "Principes" des présentes lignes directrices).

10.2.3 Exercice des droits des personnes concernées

Parfois, les développeurs complètent les données disponibles par inférence. Par exemple, si vous ne disposez pas des données factuelles correspondant aux opinions politiques d'un délinquant, vous pouvez utiliser un autre algorithme pour les déduire du reste des données, comme la participation observée à des manifestations. Toutefois, cela ne signifie en aucun cas que ces données peuvent être considérées comme pseudonymisées ou anonymisées. Elles restent donc des données à caractère personnel. De même, les données déduites doivent également être considérées comme des données à caractère personnel. Par conséquent, les personnes concernées ont certains droits fondamentaux sur ces données que vous devez respecter.

En effet, vous devez respecter les droits des personnes concernées tout au long de leur cycle de vie. Dans cette étape spécifique, le droit d'accès, de rectification et d'effacement sont particulièrement sensibles et comportent certaines caractéristiques que les responsables de traitement doivent connaître. Toutefois, dans le cas de recherches à des fins scientifiques telles que celle que vous développez, le RGPD inclut certaines garanties et dérogations relatives au traitement (Art. 89). Vous devez être au courant de la réglementation concrète de votre État membre. Selon le RGPD, le droit de l'Union ou des États membres peut prévoir des dérogations aux principaux droits inclus dans les articles 15 et suivants, dans la mesure où ces droits sont susceptibles de rendre impossible ou de nuire gravement à la réalisation des finalités spécifiques, et où ces dérogations sont nécessaires à la réalisation de ces finalités.

-Droit d'accès (voir "Droit d'accès" dans la partie II, section "Droits des personnes concernées" des présentes lignes directrices).

En principe, vous devez répondre aux demandes d'accès des personnes concernées à leurs données personnelles, à condition d'avoir pris des mesures raisonnables pour vérifier l'identité de la personne concernée, et qu'aucune autre exception ne s'applique. Toutefois, vous n'êtes pas tenu de collecter ou de conserver des données à caractère personnel supplémentaires pour permettre l'identification des personnes concernées dans les données de formation dans le seul but de vous conformer au règlement. Si vous ne pouvez pas identifier une personne concernée dans les données de formation et que la personne concernée ne peut pas fournir d'informations supplémentaires qui permettraient son identification, elle n'est pas obligée de satisfaire une demande qu'il n'est pas possible de satisfaire.

-Droit de ***rectification*** (voir "Droit de rectification" dans la partie II, section "Droits de la personne concernée" des présentes lignes directrices).

Dans le cas du droit de rectification, vous devez garantir le droit de rectification des données, notamment celles générées par les déductions et les profils établis par un outil d'IA. Même si l'objectif des données d'entraînement est de former des modèles basés sur des modèles généraux dans de grands ensembles de données, et donc que les

inexactitudes individuelles sont moins susceptibles d'avoir un effet direct sur une personne concernée, le droit de rectification ne peut pas être limité. Au maximum, vous pourriez demander un délai plus long (deux mois supplémentaires) pour procéder à la rectification si la procédure technique est particulièrement complexe (article 11, paragraphe 3).

-Droit à l'effacement (voir "Droit à l'effacement" dans la partie II, section "Droits de la personne concernée" des présentes lignes directrices).

Les personnes concernées ont le droit de supprimer leurs données personnelles. Toutefois, ce droit peut être limité si certaines circonstances concrètes s'appliquent. Selon l'ICO britannique, "les organisations peuvent également recevoir des demandes d'effacement de données de formation. Les organisations doivent répondre aux demandes d'effacement, sauf si une exemption pertinente s'applique et à condition que la personne concernée ait des motifs appropriés. Par exemple, si les données de formation ne sont plus nécessaires parce que le modèle ML a déjà été formé, l'organisation doit satisfaire la demande. Toutefois, dans certains cas, lorsque le développement du système est en cours, il peut encore être nécessaire de conserver les données de formation aux fins du réentraînement, du perfectionnement et de l'évaluation d'un outil d'IA. Dans ce cas, l'organisation doit adopter une approche au cas par cas pour déterminer si elle peut satisfaire les demandes. Se conformer à une demande de suppression des données d'entraînement n'entraînerait pas l'effacement des modèles ML basés sur ces données, sauf si les modèles eux-mêmes contiennent ces données ou peuvent être utilisés pour les déduire." [648]

11 Évaluation (validation)

11.1 Description

"Avant de procéder au déploiement final du modèle construit par l'analyste de données, il est important de procéder à une évaluation plus approfondie du modèle et de revoir la construction du modèle pour s'assurer qu'il atteint correctement les objectifs de l'entreprise. Il est essentiel de déterminer si certaines questions importantes n'ont pas été suffisamment prises en compte. À la fin de cette phase, le chef de projet doit alors décider exactement comment utiliser les résultats de l'exploration de données. Les étapes clés ici sont l'évaluation des résultats, la révision du processus et la détermination des prochaines étapes."[649]

Cette phase comporte plusieurs tâches qui soulèvent des questions importantes en matière de protection des données. Globalement, vous devez :

[648] ICO, Enabling access, erasure, and rectification rights in AI tools (Permettre les droits d'accès, d'effacement et de rectification dans les outils d'IA), à l'adresse suivante : https://ico.org.uk/about-the-ico/news-and-events/ai-blog-enabling-access-erasure-and-rectification-rights-in-ai-systems/. Consulté le 15 mai 2020.

[649] Colin Shearer, Le modèle CRISP-DM : Le nouveau plan directeur pour l'extraction de données, p. 17

- Évaluer les résultats de votre modèle, par exemple pour savoir s'il est précis ou non. À cette fin, le développeur d'IA peut le tester dans le monde réel. Ce test peut souvent être réalisé en coordination avec un partenaire lié au projet et appartenant au domaine dans lequel le système doit être déployé (par exemple, LEA).
- Examiner le processus. Vous devez examiner le système de traitement des données afin de déterminer s'il existe un facteur ou une tâche critique qui a été négligé d'une manière ou d'une autre. Cela inclut les questions d'assurance qualité. Il s'agit en fait de la phase la plus récente pour impliquer les utilisateurs finaux potentiels dans le processus de développement. Cependant, vous devez impliquer et connaître les besoins de l'utilisateur final à un stade très précoce de votre projet (compréhension de l'activité). À ce stade, les parties prenantes et les utilisateurs finaux peuvent donner un aperçu des forces et des faiblesses du système dans le monde réel.

11.2 Principales mesures à prendre

11.2.1 Processus de validation dynamique

La validation du traitement, y compris d'une composante IA, doit être effectuée dans des conditions qui reflètent l'environnement réel dans lequel le traitement est destiné à être déployé. Ainsi, si vous savez à l'avance où l'outil d'IA sera utilisé, vous devez adapter le processus de validation à cet environnement. La meilleure façon d'y parvenir est d'impliquer les partenaires respectifs du domaine concerné. Si l'outil doit être déployé dans le pays x, vous devez le valider avec des données obtenues auprès de la population concernée ou, si ce n'est pas possible, auprès d'une population similaire. Sinon, les résultats pourraient être totalement erronés. Dans tous les cas, vous devez informer tout utilisateur potentiel des conditions de validation.

En outre, le processus de validation nécessite un examen périodique si les conditions changent ou si l'on soupçonne que la solution elle-même peut être altérée. Par exemple, si l'algorithme est alimenté par les données d'un groupe spécifique de personnes, vous devez évaluer si cela modifie ou non sa précision dans une autre partie de la population. Vous devez vous assurer que la validation reflète fidèlement les conditions dans lesquelles l'algorithme a été validé.

Pour atteindre cet objectif, la validation doit inclure tous les composants d'un outil d'IA, y compris les données, les modèles pré-entraînés, les environnements et le comportement du système dans son ensemble. La validation doit également être effectuée le plus tôt possible. Globalement, il faut s'assurer que les résultats ou les actions sont cohérents avec les résultats des processus précédents, en les comparant aux politiques préalablement définies pour s'assurer qu'elles ne sont pas violées.[650] La validation nécessite parfois la collecte de nouvelles données à caractère personnel. Dans d'autres cas, les responsables du traitement utilisent les données à des fins autres que

[650] Groupe d'experts de haut niveau sur l'IA, Lignes directrices en matière d'éthique pour une IA digne de confiance, 2019, p. 22. À l'adresse : https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai Consulté le 15 mai 2020

celles prévues à l'origine. Dans tous ces cas, les responsables du traitement doivent veiller au respect du RGPD (voir la section "Limitation de la finalité" dans "Principes" et "Protection des données et recherche scientifique" dans "Concepts principaux", partie II des présentes lignes directrices).

11.2.2 Suppression d'un jeu de données inutile

Très souvent, les processus de validation et de formation sont en quelque sorte liés. Si la validation recommande des améliorations du modèle, la formation doit être effectuée à nouveau. Une fois que l'outil d'IA a finalement été réalisé, l'étape de formation de l'outil d'IA est terminée. À ce moment-là, vous devez mettre en œuvre la suppression de l'ensemble des données utilisées à cette fin, à moins qu'il n'existe un besoin légal de les conserver pour affiner ou évaluer le système, ou à d'autres fins compatibles avec celles pour lesquelles elles ont été collectées conformément aux conditions de l'article 6, paragraphe 4, du RGPD (voir la section "Définir des politiques adéquates de stockage des données").

Dans le cas où les personnes concernées demandent son effacement, vous devrez adopter une approche au cas par cas en tenant compte des éventuelles limitations à ce droit prévues par le règlement (voir art. 17, paragraphe 3). [651]

11.2.3 Réalisation d'un audit externe du traitement des données

Étant donné que les risques du système que vous développez sont élevés, **un audit du système par une tierce partie indépendante doit être impliqué**. Différents audits peuvent être utilisés. Ils peuvent être internes ou externes ; ils peuvent couvrir uniquement le produit final ou être réalisés avec des prototypes moins évolués. Ils peuvent être considérés comme une forme de contrôle et un outil de transparence, qui est censé être une caractéristique de qualité également.

En termes d'exactitude juridique, les solutions d'IA doivent être auditées pour voir si elles fonctionnent bien avec le RGPD en considérant un large éventail de questions. Le groupe d'experts de haut niveau sur l'IA a déclaré que *"les processus de test devraient être conçus et réalisés par un groupe de personnes aussi diversifié que possible. Des mesures multiples devraient être développées pour couvrir les catégories qui sont testées pour différentes perspectives. On peut envisager des tests contradictoires effectués par des "équipes rouges" fiables et diverses qui tentent délibérément de "casser" le système pour trouver des vulnérabilités, ainsi que des "primes aux bogues" qui incitent les personnes extérieures à détecter et à signaler de manière responsable les erreurs et les faiblesses du système."*[652] L'audit doit également comprendre le respect du principe d'explicabilité. "*Le degré auquel l'explicabilité est nécessaire dépend fortement du contexte et de la gravité des conséquences si cette sortie est erronée ou autrement inexacte.* "[653] Compte tenu des conséquences très graves pour les personnes soupçonnées ou condamnées pour des activités criminelles, les technologies de ML appliquées doivent permettre l'explicabilité, parmi d'autres mesures requises, afin que

---

[651] AEPD, Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción, 2020, p.26. À l'adresse : https://www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf

[652] Groupe d'experts de haut niveau sur l'IA, Lignes directrices en matière d'éthique pour une IA digne de confiance, 2019, p. 22. À l'adresse : https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai Consulté le 15 mai 2020

[653] Ibidem, p.15

les systèmes développés respectent les droits fondamentaux. L'audit doit également porter sur les mesures mises en œuvre pour éviter les biais, l'obscurité, le profilage caché, etc., et sur l'utilisation correcte d'outils tels que le AIPD, qui peut être réalisée plusieurs fois. La mise en œuvre de politiques de protection des données adéquates dès les premières étapes du cycle de vie de l'outil est le meilleur moyen d'éviter les problèmes de protection des données.

12 Déploiement

12.1 Description

"Le déploiement est le processus qui consiste à rendre un système informatique opérationnel dans son environnement, y compris l'installation, la configuration, l'exécution, les tests et les modifications nécessaires. Le déploiement n'est généralement pas effectué par les développeurs d'un système mais par l'équipe informatique du client. Néanmoins, même si c'est le cas, les développeurs auront la responsabilité de fournir au client des informations suffisantes pour un déploiement réussi du modèle. Cela comprendra normalement un plan de déploiement (générique), avec les étapes nécessaires pour un déploiement réussi et la manière de les réaliser, et un plan de surveillance et de maintenance (générique) pour la maintenance du système, et pour la surveillance du déploiement et de l'utilisation correcte des résultats de l'exploration de données."[654]

12.2 Principales mesures à prendre

12.2.1 Remarques générales

Une fois que vous avez créé votre algorithme, vous êtes confronté à un problème important. Il se peut qu'il incorpore des données personnelles, ouvertement ou de manière cachée. Vous devez procéder à une évaluation formelle pour déterminer quelles données personnelles des personnes concernées pourraient être identifiables. Cela peut parfois être compliqué. Par exemple, certaines solutions d'IA, telles que les machines à support vectoriel (VSM), peuvent contenir des exemples de données d'entraînement dans la logique du modèle. Dans d'autres cas, des modèles peuvent être trouvés dans le modèle qui identifient un individu unique. Dans tous ces cas, des parties non autorisées peuvent être en mesure de récupérer des éléments des données d'apprentissage, ou de déduire qui s'y trouvait, en analysant la façon dont le modèle se comporte. Si vous savez ou soupçonnez que l'outil d'IA contient des données personnelles (voir la section "Achat ou promotion de l'accès à une base de données" dans "Principaux outils et actions", Partie II), vous devez :

[654] SHERPA, Lignes directrices pour le développement éthique des systèmes d'IA et de Big Data : Une approche d'éthique par la conception, 2020, p 13. À l'adresse : https://www.project-sherpa.eu/wp-content/uploads/2019/12/development-final.pdf Consulté le 15 mai 2020

- Les supprimer ou, au contraire, justifier l'impossibilité de le faire, en tout ou partie, en raison de la dégradation que cela impliquerait pour le modèle (voir la section "Limitation du stockage" du chapitre "Principes").
- Déterminer la base juridique de la communication de données à caractère personnel à des tiers, en particulier si des catégories spéciales de données sont concernées (voir la sous-section "Licéité" de la section "Principes de licéité, de loyauté et de transparence" de la partie II, section "Principes" des présentes lignes directrices).
- Informer les personnes concernées du traitement ci-dessus.
- Démontrer que les politiques de protection des données dès la conception et par défaut ont été mises en œuvre (notamment la minimisation des données) (voir "Protection des données dès la conception et par défaut" dans la partie II, section "Concepts principaux" des présentes lignes directrices).
- Réaliser une analyse d'impact sur la protection des données (AIPD) **(**voir "AIPD" dans la partie II, section "Principaux outils et actions" des présentes lignes directrices).

Enfin, vous devez prendre des mesures régulières pour évaluer de manière proactive la probabilité que des données à caractère personnel soient déduites de modèles à la lumière de l'état de la technologie, afin de minimiser le risque de divulgation accidentelle. Si ces actions révèlent une possibilité substantielle de divulgation des données, les mesures nécessaires pour l'éviter doivent être mises en œuvre (voir "Principe d'intégrité et de confidentialité" dans la partie II section "Principes" des présentes lignes directrices).

12.2.2 Mise à jour des informations

Si l'algorithme est mis en œuvre par un tiers, vous devez communiquer les résultats du système de validation et de suivi employé lors des phases de développement et proposer votre collaboration pour continuer à suivre la validation des résultats. Il serait également souhaitable d'établir ce type de coordination avec les tiers auprès desquels vous acquérez des bases de données ou tout autre composant pertinent dans le cycle de vie du système. Si cela implique le traitement de données par un tiers, vous devez vous assurer que l'accès est fourni sur une base légale.

Il est nécessaire d'offrir à l'utilisateur final des informations en temps réel sur les valeurs de précision et/ou de qualité des informations déduites à chaque étape (voir "Principe de précision" dans la partie II, section "Principes" des présentes lignes directrices). Lorsque les informations déduites n'atteignent pas les seuils de qualité minimum, vous devez souligner que ces informations n'ont aucune valeur. Cette exigence implique souvent que vous devez fournir des informations détaillées sur les étapes de formation et de validation. Les informations sur les ensembles de données utilisés à ces fins sont particulièrement importantes. Dans le cas contraire, l'utilisation de la solution risque d'apporter des résultats décevants aux utilisateurs finaux, qui se retrouvent à spéculer sur la cause.

Vous devez également vous assurer que toute mise en œuvre dans le monde réel est également conforme à la *directive relative à l'application de la loi sur la protection des*

données (directive 2016/680)[655] et à leur mise en œuvre spécifique dans les différents États membres. Sachez que cela implique généralement pour les LEA des réglementations moins restrictives concernant l'utilisation des données personnelles. Dans le domaine de la justice pénale, la fourniture de preuves est souvent une activité contraignante. Il s'agit donc d'une tendance naturelle à collecter et traiter autant de données que possible qui pourraient éventuellement s'avérer utiles. Cette tendance est même renforcée par les possibilités techniques croissantes d'analyse automatique d'énormes quantités de données par des outils d'IA. Cependant, la minimisation des données est nécessaire et des contre-mesures efficaces contre la collecte et le traitement extensifs des données doivent donc être intégrées dès la conception des outils d'IA.

Le respect des droits de l'Homme et des principes éthiques exige la réalisation d'autres conditions essentielles :

"En ce qui concerne les technologies de surveillance, la charge de la preuve devrait incomber aux États et/ou aux entreprises, qui doivent faire des démonstrations publiques et transparentes, avant d'introduire des options de surveillance,

- qu'elles sont nécessaires

- qu'elles sont efficaces

- qu'elles respectent la proportionnalité (par exemple, la limitation de la finalité)

- qu'il n'existe pas de meilleures alternatives qui pourraient remplacer ces technologies de surveillance

Ces critères doivent ensuite également être soumis à une évaluation a posteriori, soit au niveau de l'analyse politique normale, soit par le biais des politiques des États membres en la matière. "[656]

[655] Parlement européen et Conseil, 2016, Directive (UE) 2016/680 du Parlement européen et du Conseil du 27 avril 2016 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel par les autorités compétentes à des fins de prévention et de détection des infractions pénales, d'enquêtes et de poursuites en la matière ou d'exécution de sanctions pénales, et à la libre circulation de ces données, et abrogeant la décision-cadre 2008/977/JAI du Conseil, Journal officiel <http://eur-lex.europa.eu/legal-content/EL/TXT/?uri=OJ:L:2016:119:TOC>.

[656] Groupe européen d'éthique des sciences et des nouvelles technologies. (2014). Avis n° 28 : éthique des technologies de sécurité et de surveillance (10.2796/22379). Récupéré de Luxembourg : Bruxelles : https://publications.europa.eu/en/publication-detail/-/publication/6f1b3ce0-2810-4926-b185-54fc3225c969/language-en/format-PDF/source-77404258