Institutional Repository - Research Portal
Dépôt Institutionnel - Portail de la Recherche
researchportal.unamur.be
University of Namur

RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

Big Data dans l'IA et principe de minimisation

Everarts de Velp, Sophie

Published in:
Time to reshape the digital society

Publication date:
2021

Document Version
le PDF de l'éditeur

Link to publication

Citation for pulished version (HARVARD):
Everarts de Velp, S 2021, Big Data dans l'IA et principe de minimisation: défis et risques . dans *Time to reshape the digital society: 40th anniversary of the CRIDS.* Collection du CRIDS, numéro 52, Larcier , Bruxelles, pp. 289-298.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 02. May. 2026

strada lex

Chapitre 6
Big Data dans l'IA et principe de minimisation : Défis et risques[1]

Sophie Everarts de Velp[2]

Par définition, le Big Data[3] désigne une quantité de données devenue si volumineuse qu'il est devenu presque impossible pour un outil informatique classique, et encore moins pour un cerveau humain, d'en analyser le contenu. Un système d'intelligence artificielle (« IA ») peut être défini comme un « logiciel [...] capable, pour un ensemble donné d'objectifs définis par l'homme, de générer des résultats tels que du contenu, des prédictions, des recommandations ou des décisions influençant les environnements avec lesquels ils interagissent »[4]. Le Big Data et l'IA peuvent être considérés comme complémentaires. D'une part, l'IA s'améliore au fur et à mesure que l'on lui fournit des données. D'autre part, le Big Data est tout simplement sans valeur sans logiciel pour l'analyser. Le Big Data continuera à se développer à mesure que l'IA deviendra une option viable pour une plus grande automatisation et l'IA deviendra un domaine plus large à mesure que davantage de données seront disponibles pour l'apprentissage et l'analyse.

Cependant, l'utilisation de l'IA pourrait affecter un certain nombre de droits fondamentaux dont ceux inclus dans la Charte des droits fondamentaux de l'Union européenne. De ce fait, la nouvelle proposition de

[1] This work has been done with the financial support from the European Union's Horizon 2020 research and innovation program under Grant Agreements n° 830892 (SPARTA). La publication ne reflète que l'opinion de ses auteurs et la Commission européenne ne peut être tenue responsable de l'usage qui en serait fait.

[2] Chercheuse au CRIDS- UNamur.

[3] Pour plus d'informations sur le Big Data : B. Frenay, « Démystifier le machine learning », *R.D.T.I.*, 2018, 70, pp. 5 et s.

[4] Art. 3 de la Proposition de Règlement du Parlement européen et du Conseil du 21 avril 2021 établissant des règles harmonisées en matière d'intelligence artificielle (loi sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union.

Université de Namur - Bibliothèque de la Faculté de droit - On Campus / quentin.houbion@unamur.be
Chapitre 6 - Big Data dans l'IA et principe de minimisation : Défis et risques
www.stradalex.com - 25/11/2021

strada lex

Règlement européen relative à l'intelligence artificielle établit des règles harmonisées en matière d'intelligence artificielle, garantissant un niveau élevé de protection de ces droits fondamentaux, notamment le droit à la vie privée et à la protection des données à caractère personnel. Elle traite également les différentes sources de risque au moyen d'une approche fondée sur le risque clairement identifiée[5].

Rappelons que l'article 5 du Règlement général sur la protection des données (RGPD) dispose que les données à caractère personnel collectées et traitées doivent être adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées[6]. Cette exigence découle de l'article 8 de la CEDH qui dispose que toute personne a droit au respect de sa vie privée et familiale[7]. Nous pouvons en conclure que la limitation de collecte de données à caractère personnel à ce qui est strictement nécessaire à la réalisation de la finalité est une obligation fondamentale de respect de la vie privée des personnes concernées.

Ainsi, en raison de la nécessité de collecter et d'utiliser d'énormes quantités de données, dont certaines sont considérées comme des données à caractère personnel, l'intelligence artificielle pourrait être par nature opposée au principe de minimisation des données à caractère personnel promu par le RGPD.

La présente contribution analyse les notions de Big Data, du principe de minimisation, de la protection des données dès la conception et par défaut, le tout dans un contexte d'intelligence artificielle. Nous développerons également une solution envisagée pour remédier à cette contrainte de minimisation de données là où la collecte massive est de vigueur, à savoir l'anonymisation. Nous verrons que cette dernière n'est pas la solution idéale tant par les pièges auxquels nous pourrions être confrontés qu'aux risques à considérer, tel que celui de la ré-identification des données à caractère personnel.

Big Data. Cette notion est souvent reliée à trois caractéristiques, dites « les 3 V » : le volume massif de données, la variété des données et la vitesse de collecte et de traitement des données[8]. Certains auteurs ajoutent

[5] Pt 3 du mémorandum explicatif de la Proposition de Règlement du Parlement européen et du Conseil du 21 avril 2021 établissant des règles harmonisées en matière d'intelligence artificielle (loi sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union.

[6] Art. 5 du Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (RGPD), *J.O.U.E.*, L 119, 4 mai 2016.

[7] Art. 8 de la Convention européenne des Droits de l'homme.

[8] A. Beelen, « Fiche de guidance n° 35 – Le big data et le RGPD », in *Guide pratique du RGPD*, 1re éd., Bruxelles, Bruylant, 2018, pp. 251-263.

Université de Namur - Bibliothèque de la Faculté de droit - On Campus / quentin.houbion@unamur.be
Chapitre 6 - Big Data dans l'IA et principe de minimisation : Défis et risques
www.stradalex.com - 25/11/2021

strada lex

un quatrième critère qui est celui de la valeur des données collectées, traitées et interprétées[9]. Il est devenu indéniable qu'au plus les données ont de la valeur, au plus le résultat sera intéressant pour le responsable du traitement[10].

En particulier, le premier « V » pose la question de la quantité massive de données à caractère personnel collectées et traitées. En effet, cette collecte massive de données (ex : via des objets connectés, via les réseaux sociaux, les sites de vente, l'historique des recherches, des données publiques, …) permet au responsable du traitement de connaître le plus précisément possible les préférences de ses clients. Il s'agit « d'une opportunité pour les entreprises de développer leurs ventes, de fidéliser, d'améliorer la qualité de leurs produits/services et donc à terme de se transformer »[11]. Ces collectes massives de données sont principalement utilisées pour mieux comprendre le comportement des utilisateurs d'une technologie et pour la développer en conséquence. D'après Axel Beelen : « Ce ne sont pas les mégadonnées qui sont en elles-mêmes inquiétantes, mais bien les utilisations des résultats qui en découlent »[12]. En effet, une mauvaise gestion des résultats engrangés par le Big Data pourrait mener à des manipulations (par exemple : Cambridge Analytica[13], influence sur le processus démocratique[14] etc.), des fausses prédictions (par exemple : une mauvaise idée d'un profil utilisateur) ou à des biais (discrimination, etc.).

Le développement exponentiel de l'intelligence artificielle et du Big Data pose de nombreuses questions relatives à la vie privée. Il est désormais clair que nous serons de plus en plus entourés d'outils intelligents (smart cities, voitures autonomes, robots domestiques, objets connectés, etc.). Des milliards de données sont et seront collectées. Un cadre législatif

[9] T. Zarsky, « Incompatible: The GDPR in the Age of Big Data », 47 4(2) *Seton Hall Law Review*, 2017.

[10] Selon l'article 4.7 du RGPD, le responsable du traitement est « la personne physique ou morale, l'autorité publique, le service ou un autre organisme qui, seul ou conjointement avec d'autres, détermine les finalités et les moyens du traitement ; lorsque les finalités et les moyens de ce traitement sont déterminés par le droit de l'Union ou le droit d'un État membre, le responsable du traitement peut être désigné ou les critères spécifiques applicables à sa désignation peuvent être prévus par le droit de l'Union ou par le droit d'un État membre ».

[11] A. Beelen, « Fiche de guidance n° 35 – Le big data et le RGPD », *op. cit.*

[12] *Ibid.*

[13] Les utilisateurs de Facebook qui ont répondu à un test de personnalité pensaient le faire dans le cadre d'une étude, alors qu'en réalité le but de la récolte des données était commercial et de prospection politique.

[14] « L'intelligence artificielle (et les algorithmes) et l'impact sur les médias sociaux dans le processus démocratique, rapport fait au nom du Comité d'avis des questions scientifiques et technologiques », *Doc. parl.,* Chambre, 26 avril 2021, n° 1947/001.

Université de Namur - Bibliothèque de la Faculté de droit - On Campus / quentin.houbion@unamur.be
Chapitre 6 - Big Data dans l'IA et principe de minimisation : Défis et risques
www.stradalex.com - 25/11/2021

strada lex

important ainsi que des procédés techniques devraient être développés pour contrôler au mieux ces développements technologiques afin de garantir aux individus une réelle sécurité juridique et la sauvegarde de leurs droits fondamentaux[15]. Bien entendu, il existe déjà des règles en ce sens, mais il convient de les faire évoluer pour encadrer au mieux ces technologies.

Principes de minimisation. Comme nous l'avons précisé, la limitation de la collecte de données à caractère personnel (pseudonymisées[16] ou non) aux données strictement nécessaires à la réalisation de la finalité est une obligation fondamentale pour le respect de la vie privée des personnes[17]. Le responsable du traitement doit maintenir la proportionnalité au regard des données collectées et des traitements mis en œuvre. Ainsi, l'exigence de minimisation concerne tant la quantité de données à caractère personnel collectées que leur durée de conservation et leur accessibilité[18].

Par conséquent, les données ne devraient être traitées que si la finalité du traitement envisagé ne peut être raisonnablement atteinte par d'autres moyens que par l'utilisation de données à caractère personnel. Il convient au responsable du traitement d'effectuer une analyse d'impact afin de déterminer si les données qu'il compte utiliser sont réellement pertinentes, adéquates et utiles pour la conception et le développement de sa technologie[19]. Souvent, la finalité poursuivie pourra être atteinte autrement par des mesures plus respectueuses des droits des personnes concernées.

Enfin, une solution intéressante quant à la durée de conservation à minimiser serait un système d'effacement des données inutilisées par

[15] C. DE TERWANGNE, K. ROSIER et B. LOSDYCK, « Lignes de force du nouveau Règlement relatif à la protection des données à caractère personnel », *Journal de droit européen*, 2016, pp. 32-33.

[16] Selon l'article 4.5 du RGPD, la pseudonymisation désigne : « le traitement de données à caractère personnel de telle sorte que les données à caractère personnel ne puissent plus être attribuées à une personne concernée spécifique sans l'utilisation d'informations supplémentaires, à condition que ces informations supplémentaires soient conservées séparément et fassent l'objet de mesures techniques et organisationnelles visant à garantir que les données à caractère personnel ne sont pas attribuées à une personne physique identifiée ou identifiable ».

[17] Art. 5, c), et cons. n° 39 RGPD.

[18] Art. 25 RGPD.

[19] Art. 35 RGPD.

Université de Namur - Bibliothèque de la Faculté de droit - On Campus / quentin.houbion@unamur.be
Chapitre 6 - Big Data dans l'IA et principe de minimisation : Défis et risques
www.stradalex.com - 25/11/2021

strada lex

l'intelligence artificielle pendant un certain temps[20]. Une autre piste serait celle de l'anonymisation de ces données[21].

Protection des données dès la conception et par défaut. L'article 25.1 du RGPD impose au responsable du traitement de mettre en œuvre, tant au moment de la détermination des moyens du traitement qu'au moment du traitement même, des mesures techniques et organisationnelles appropriées. Cette disposition donne des exemples concrets tels que la pseudonymisation et la minimisation des données. Le RGPD fait état des conditions à prendre en compte en prenant ces mesures : « l'état des connaissances, des coûts de mise en œuvre et de la nature, de la portée, du contexte et des finalités du traitement ainsi que des risques, dont le degré de probabilité et de gravité varie, que présente le traitement pour les droits et libertés des personnes physiques »[22]. La minimisation est dès lors considérée comme une mesure technique et organisationnelle par le législateur européen visant au respect de ce principe de protection des données dès la conception[23]. Un second exemple de mesure technique de minimisation serait que l'intelligence artificielle floute automatiquement les visages ou les plaques de voiture dans la rue dès lors que la finalité poursuivie par le responsable du traitement peut être atteinte sans ces informations[24].

Par ailleurs, l'article 25, alinéa 2, du RGPD rappelle ce principe de minimisation en indiquant que le responsable du traitement doit mettre en place des mesures techniques et organisationnelles pour que, par défaut, seules les données à caractère personnel qui sont nécessaires au regard de chaque finalité spécifique du traitement soient traitées.

Enfin, le RGPD impose des mesures de sécurité appropriées et adopte donc une approche basée sur les risques. Ainsi, en fonction de la nature et de la quantité de données à caractère personnel et du traitement effectué, le responsable du traitement doit déterminer les risques, la probabilité que ces risques se produisent et la gravité des risques pour les personnes dans une analyse d'impact sur la protection des données. Il s'agit non seulement des risques liés à la vie privée et à la protection des données

---

[20] A. Delforge et L. Gérard, « Chapitre 2. – Le GDPR, source de solutions ou de blocages ? Une question de point de vue », in A. de Streel et H. Jacquemin (dir.), *L'intelligence artificielle et le droit*, Bruxelles, Larcier, 2017, pp. 156-188.

[21] Voy. *infra*.

[22] Art. 25.1 RGPD.

[23] Voy. également ENISA, « Recommendations on shaping technology according to GDPR provisions », 28 janvier 2019, https://www.enisa.europa.eu/publications/recommendations-on-shaping-technology-according-to-gdpr-provisions-part-2.

[24] A. Delforge et L. Gérard, « Chapitre 2. – Le GDPR, source de solutions ou de blocages ? Une question de point de vue », *op. cit.*

Université de Namur - Bibliothèque de la Faculté de droit - On Campus / quentin.houbion@unamur.be
Chapitre 6 - Big Data dans l'IA et principe de minimisation : Défis et risques
www.stradalex.com - 25/11/2021

strada lex

personnelles, mais aussi des risques liés à la liberté d'expression, à la liberté de pensée, à la liberté de circulation, à la discrimination, etc. (Par exemple, vol, discrimination, perte financière, suppression non autorisée de la pseudonymisation, vol d'identité, etc.)[25].

Personne responsable. Le RGPD indique que le responsable du traitement doit mettre en place des mesures techniques et organisationnelles afin de respecter notamment les principes de minimisation et de protection dès la conception et par défaut. Dans la pratique, les concepteurs et fabricants de logiciels IA ne traitent pas forcément de données à caractère personnel et ne sont dès lors pas systématiquement considérés comme responsables du traitement. Il s'agira bien souvent de l'utilisateur de cette intelligence artificielle qui traitera ces données, ce qui en fera le responsable du traitement devant respecter les principes énoncés dans le RGPD. Malgré cette responsabilité qui incombe à l'utilisateur, le concepteur de l'outil d'intelligence artificielle devrait prendre en compte les aspects de protection des données à caractère personnel, bien qu'il n'y soit pas légalement lié[26]. En effet, le considérant n° 78 du RGPD mentionne que « lors de l'élaboration, de la conception, de la sélection et de l'utilisation d'applications, de services et de produits qui reposent sur le traitement de données à caractère personnel ou traitent des données à caractère personnel pour remplir leurs fonctions, il convient d'inciter les fabricants de produits, les prestataires de services et les producteurs d'applications à prendre en compte le droit à la protection des données lors de l'élaboration et de la conception de tels produits, services et applications et, compte dûment tenu de l'état des connaissances, à s'assurer que les responsables du traitement et les sous-traitants sont en mesure de s'acquitter des obligations qui leur incombent en matière de protection des données »[27]. Dans la pratique, les véritables choix liés au traitement des données sont généralement posés par les concepteurs de logiciels, d'applications, de robots et pas forcément par ceux qui les utilisent pour traiter des données à caractère personnel[28]. En effet, les mesures dites techniques, imposées

[25] F. Dumortier, « La sécurité des traitements de données, les analyses d'impact et les violations de données », in C. De Terwangne et K. Rosier (coord.), *Le règlement général sur la protection des données (RGPD/GDPR). Analyse approfondie*, Bruxelles, Larcier, 2018, p. 188 ; cons. n° 75 RGPD.

[26] A. Delforge et L. Gerard, « Notre vie privée est-elle réellement mise en danger par les robots ? Étude des risques et analyse des solutions apportées par le GDPR », in H. Jacquemin et A. De Streel (coord.), *L'intelligence artificielle et le droit*, Bruxelles, Larcier, 2017, p. 169.

[27] Cons. n° 78 RGPD ; voy. également : Conseil de l'Europe, *Lignes directrices sur l'intelligence artificielle et la protection des données*, disponible sur https://rm.coe.int/lignes-directrices-sur-l-intelligence-artificielle-et-la-protection-de/168091ff40.

[28] A. Delforge et L. Gérard, « Chapitre 2. – Le GDPR, source de solutions ou de blocages ? Une question de point de vue », *op. cit.*

Université de Namur - Bibliothèque de la Faculté de droit - On Campus / quentin.houbion@unamur.be
Chapitre 6 - Big Data dans l'IA et principe de minimisation : Défis et risques
www.stradalex.com - 25/11/2021

strada lex

par le RGPD, sont plus facilement mises en place par un spécialiste (le concepteur de l'IA) plutôt que par l'utilisateur. Prenons l'exemple d'une de ces mesures techniques permettant une coexistence du Big Data et du principe de minimisation des données en matière d'intelligence artificielle : l'anonymisation.

Anonymisation vs pseudonymisation. Une des possibilités envisagées pour minimiser l'impact du traitement de données à caractère personnel lors d'une collecte massive en vue de développer un logiciel d'IA est l'anonymisation de ces données. L'intérêt de la distinction entre anonymisation et pseudonymisation se trouve dans le fait que les données anonymisées ne tombent pas dans le champ d'application du RGPD car elles ne sont plus considérées comme des données à caractère personnel. Au contraire, les données pseudonymisées sont toujours considérées comme des données à caractère personnel, ce qui implique l'application du RGPD avec quelques particularités.

La directive (UE) 2019/1024 sur les données ouvertes et la réutilisation des informations du secteur public[29] propose une définition formelle de l'anonymisation. Selon l'article 2.7 de la directive, on entend par anonymisation : « le processus consistant à changer des documents en documents anonymes qui ne se rapportent pas à une personne physique identifiée ou identifiable, ou le processus consistant à rendre des données à caractère personnel anonymes de telle sorte que la personne concernée ne soit pas ou plus identifiable ». Différentes techniques existent pour anonymiser les données, par exemple « par randomisation »[30] ou « par généralisation »[31]. La première consiste à modifier certaines informations dans un jeu de données de telle sorte qu'elles soient moins précises, par exemple en modifiant une date de naissance. La seconde consiste à modifier l'échelle de certaines informations des jeux de données ou leur ordre de grandeur afin de s'assurer qu'ils soient communs à un ensemble de personnes. Cela pourrait être le cas de dates de naissance qu'on limiterait à l'année de naissance de l'individu.

Il est vrai que la notion d'anonymisation est souvent confondue avec la notion de pseudonymisation. Selon l'article 4.5 du RGPD, la pseudonymisation désigne : « le traitement de données à caractère personnel de telle

[29] Directive (UE) 2019/1024 du Parlement européen et du Conseil du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur public, *J.O.U.E*, 26 juin 2019.

[30] Cette technique permet de protéger le jeu de données du risque d'inférence (voy. : Groupe 29, Avis 05/2014 du 10 avril 2014 sur les techniques d'anonymisation, WP216).

[31] Cette technique permet d'éviter l'individualisation d'un jeu de données. Elle limite également les possibles corrélations du jeu de données avec d'autres (voy. : Groupe 29, Avis 05/2014 du 10 avril 2014 sur les techniques d'anonymisation, WP216).

Université de Namur - Bibliothèque de la Faculté de droit - On Campus / quentin.houbion@unamur.be
Chapitre 6 - Big Data dans l'IA et principe de minimisation : Défis et risques
www.stradalex.com - 25/11/2021

strada lex

sorte que les données à caractère personnel ne puissent plus être attribuées à une personne concernée spécifique sans l'utilisation d'informations supplémentaires, à condition que ces informations supplémentaires soient conservées séparément et fassent l'objet de mesures techniques et organisationnelles visant à garantir que les données à caractère personnel ne sont pas attribuées à une personne physique identifiée ou identifiable »[32]. Le groupe de travail Article 29 (remplacé par le Comité européen à la protection des données) soulignait que la pseudonymisation n'était pas une méthode d'anonymisation mais plutôt d'une réduction de la « possibilité de liaison d'un ensemble de données »[33] avec l'identité originale d'une personne concernée. La grande différence entre l'anonymisation et la pseudonymisation est donc que l'anonymisation complète ne permet plus d'identifier la personne concernée avec des moyens raisonnables. *A contrario*, il suffit de disposer d'une information initiale supplémentaire (par exemple, une clé pour déchiffrer un fichier crypté) pour identifier la personne dont les données sont pseudonymisées.

Par conséquent, pour pouvoir déclarer qu'il travaille avec des données anonymes, le responsable du traitement doit être certain qu'il n'est plus possible, avec des moyens raisonnables, d'identifier les individus, même en croisant les données. Le Big Data rend de plus en plus complexe la possibilité de traiter des données « anonymes ».

En conclusion, l'anonymisation constitue une solution intéressante en matière d'intelligence artificielle et de Big Data étant donné que les données collectées et traitées ne sont plus considérées comme des données à caractère personnel tombant sous le champ d'application du Règlement européen et, par conséquent, du principe de minimisation. La prudence est toutefois de mise étant entendu que l'anonymisation comprend certains pièges et risques[34]. En outre, en raison de la quantité massive d'informations qui existent, il semble compliqué de procéder à l'anonymisation de la plupart d'entre elles, vu le travail colossal que cela demanderait.

Pièges à éviter lors de l'anonymisation. Lorsqu'ils envisagent d'utiliser des techniques d'anonymisation, les responsables du traitement des données doivent tenir compte de différentes difficultés[35]. Premièrement, il ne faut pas confondre l'anonymisation et la peudonymisation. N'oublions pas que les personnes concernées peuvent toujours être identifiées dans

[32] Art. 4.5 RGPD.

[33] Groupe 29, Avis 05/2014 du 10 avril 2014 sur les techniques d'anonymisation, WP216.

[34] Voy. *infra.*

[35] Groupe 29, Avis 05/2014 du 10 avril 2014 sur les techniques d'anonymisation, WP216.

Université de Namur - Bibliothèque de la Faculté de droit - On Campus / quentin.houbion@unamur.be
Chapitre 6 - Big Data dans l'IA et principe de minimisation : Défis et risques
www.stradalex.com - 25/11/2021

strada lex

l'hypothèse de données pseudonymisées. Deuxièmement, le piège de penser que des données correctement anonymisées priveraient les personnes concernées de toute garantie. En effet, bien que le RGPD ne s'applique plus en cas d'anonymisation, d'autres textes législatifs pourraient s'appliquer, comme la Directive vie privée et communication électronique[36]. Troisièmement, le piège de ne pas se rendre compte de l'impact de l'utilisation de ces données, mêmes anonymisées, en matière de vie privée. Il convient de ne pas se contenter de « publier et oublier »[37]. Les responsables du traitement des données devraient réévaluer régulièrement les risques, examiner si les contrôles des risques identifiés sont suffisants et les ajuster en conséquence[38].

Risques de réidentification. L'anonymisation des données peut sembler être une solution idéale à une collecte massive des données respectant le principe de minimisation imposé par le RGPD. Néanmoins, elle n'est pas exempt de risques. La réidentification des données anonymisées est une possibilité à ne pas négliger. En effet, le responsable du traitement doit s'assurer que la réidentification de la personne concernée n'est plus possible par des moyens raisonnables[39]. Ainsi, l'analyse de ce risque nécessite une évaluation au cas par cas basée à la fois sur l'état de l'art des technologies d'anonymisation et sur les moyens raisonnables dont disposent les tiers pour parvenir à l'identification d'un individu[40]. Le groupe de travail Article 29 sur la protection des données avait rendu un avis[41] à ce propos, analysant les techniques pouvant fournir des garanties en matière de respect de la vie privée, mais pouvant également être utilisées pour générer des « processus d'anonymisation efficaces »[42].

En outre, le groupe de travail avait établi des critères permettant d'évaluer le risque de réidentification d'une personne concernée et donc d'évaluer le degré d'anonymisation ou de pseudonymisation d'une donnée, à

[36] Art. 5, § 3, de la Directive 2002/58/CE du Parlement européen et du Conseil du 12 juillet 2002 concernant le traitement des données à caractère personnel et la protection de la vie privée dans le secteur des communications électroniques (Directive vie privée et communications électroniques), *JO*, L 201, 31 juillet 2002, p. 37.

[37] Groupe 29, Avis 05/2014 du 10 avril 2014 sur les techniques d'anonymisation, WP216.

[38] *Idem.*

[39] Cons. n° 26 RGPD.

[40] C. De Terwangne, K. Rosier et B. Losdyck, « Le règlement européen relatif à la protection des données à caractère personnel : quelles nouveautés », *J.D.E*, 2017, p. 309. ; ENISA, « Handbook on Security of Personal Data Processing », décembre 2017, https://www.enisa.europa.eu/publications/handbook-on-security-of-personal-data-processing.

[41] Groupe 29, Avis 05/2014 du 10 avril 2014 sur les techniques d'anonymisation, WP216.

[42] *Idem.*

Université de Namur - Bibliothèque de la Faculté de droit - On Campus / quentin.houbion@unamur.be
Chapitre 6 - Big Data dans l'IA et principe de minimisation : Défis et risques
www.stradalex.com - 25/11/2021

strada lex

savoir la disponibilité d'autres données, la probabilité d'une tentative de réidentification, et la probabilité que cette dernière s'achève avec succès. Selon le groupe Article 29, des tests étaient nécessaires afin de vérifier qu'une réidentification n'était pas réalisable.

Conclusion. En raison de la nécessité de collecter et de traiter d'énormes quantités de données, dont certaines sont considérées comme des données à caractère personnel, l'intelligence artificielle pourrait être par nature opposée au principe de minimisation promu par le RGPD. Pour éviter justement qu'on ne traite ces données sans raison, il existe différents principes de protection dans le RGPD dont, entre autres, le principe de minimisation. Pour respecter ce principe, l'anonymisation des données semble être la solution idéale pour la collecte massive de données liée au développement des technologies d'IA, transformant les données à caractère personnel utilisées en données à caractère non personnel, et dès lors non soumises au RGPD. Cependant, cette technique comporte des pièges à éviter et des risques importants dont, en particulier, le risque de réidentification qui est une composante essentielle du concept de données anonymes. Assurément, avant de pouvoir envisager de travailler avec des données anonymisées (et donc des données pour lesquelles le RGPD ne s'applique pas), le responsable du traitement doit s'assurer que la réidentification de la personne concernée n'est plus possible par des moyens raisonnables. Dans l'exemple de l'IA, il est fondamental pour le responsable du traitement d'identifier les risques de réidentification des personnes concernées avant de pouvoir déclarer que sa technologie utilise des données anonymisées et que, par conséquent, elle n'est plus soumise au RGPD, et notamment au principe de minimisation des données à caractère personnel. Enfin, le développement exponentiel de l'intelligence artificielle et du Big Data, nous pousse à nous poser ces questions en matière de protection des données et de vie privée. Il est désormais clair que notre avenir (c'est même déjà le cas aujourd'hui) sera rempli d'outils collectant des milliards de données. Il nous semble primordial qu'un cadre législatif important ainsi que des procédés techniques évoluent en même temps que ces technologies pour les contrôler au mieux, afin de permettre aux individus d'avoir une réelle sécurité juridique et la sauvegarde de leurs droits fondamentaux.

Université de Namur - Bibliothèque de la Faculté de droit - On Campus / quentin.houbion@unamur.be
Chapitre 6 - Big Data dans l'IA et principe de minimisation : Défis et risques
www.stradalex.com - 25/11/2021