



RGPD

## IA et anonymisation : la commune avait tout faux

Comme chaque mois, Alexandre Fievée tente d'apporter des réponses aux questions que tout le monde se pose en matière de protection des données personnelles, en s'appuyant sur les décisions rendues par les différentes autorités nationales de contrôle au niveau européen et les juridictions européennes. Ce mois-ci, il se penche sur la question de l'anonymisation, par une municipalité, des données personnelles d'individus avant leur analyse par une solution d'Intelligence Artificielle (IA).

Un traitement de données personnelles est soumis aux exigences du règlement européen sur la protection des données (dit « RGPD »). Il en résulte pour le responsable du traitement – celui qui détermine les finalités et les moyens dudit traitement – un certain nombre d'obligations, dont le respect des principes suivants : licéité, loyauté et transparence ; limitation des finalités ; minimisation ; exactitude ; limitation de la conservation et sécurité (article 5 du RGPD). En revanche, si les données sont anonymisées, le RGPD ne s'applique pas.

Que faut-il entendre par « anonymisation » ? La Cnil la définit comme « un traitement qui consiste à utiliser un ensemble de techniques de manière à rendre impossible, en pratique, toute identification de la personne par quelque moyen que ce soit et de manière irréversible »<sup>1</sup>.

L'anonymisation ne doit donc pas être confondue avec la « pseudonymisation », qui est un traitement de données personnelles réalisé de telle manière qu'on ne puisse plus

attribuer les données relatives à une personne physique sans information supplémentaire. En d'autres termes, « la pseudonymisation consiste à remplacer les données directement identifiantes (nom, prénom, etc.) d'un jeu de données par des données indirectement identifiantes (alias, numéro séquentiel, etc.) »<sup>2</sup>. Si la première est irréversible, la seconde ne l'est pas. Par conséquent, le traitement de données pseudonymisées reste soumis au RGPD (contrairement au traitement de données anonymisées, comme indiqué supra).

La difficulté, en matière d'anonymisation, consiste à éliminer toute possibilité de réidentification des personnes dont les données font l'objet du traitement. Afin de s'assurer que les techniques d'anonymisation utilisées ont été effectives, il convient d'appliquer les trois critères suivants : l'individualisation (il est impossible d'isoler une personne dans un jeu de données)<sup>3</sup> ; la corrélation (il est impossible de relier entre eux des ensembles de données distincts concernant une même personne<sup>4</sup> ; l'inférence (il est

impossible de déduire, de façon quasi certaine, de nouvelles informations sur une personne)<sup>5</sup>. L'expérience montre qu'il est très compliqué de garantir une anonymisation parfaite des données. L'affaire ci-dessous en est l'illustration.

### L'affaire<sup>6</sup>

L'histoire se passe en Italie, dans la commune de Trente. Avec le soutien d'une fondation, la municipalité avait lancé deux projets – « Marvel » et « Protector » - reposant sur une solution d'intelligence artificielle. Le premier impliquait la collecte d'informations (images et sons) dans les lieux publics au moyen de microphones et de caméra de vidéosurveillance. Selon la commune, responsable du traitement, les données étaient immédiatement anonymisées après leur collecte pour être ensuite analysées dans l'optique de détecter, au moyen d'une solution d'IA, des événements pertinents pour la protection de la sécurité publique (rassemblements, agressions, etc.). Le second projet comprenait notamment la collecte de messages haineux

postés sur la plateforme Twitter (« X ») et de commentaires postés sur YouTube afin, après analyse par une solution d'IA, de détecter d'éventuelles émotions négatives sur la religion (agressivité, colère, etc.) et donc d'identifier les risques et menaces pour la sécurité des lieux de culte.

L'autorité italienne de protection des données a été saisie de ces deux dossiers, qui, selon la municipalité, étaient conformes au RGPD. En tout état de cause, elle estimait que leur impact sur les droits et libertés fondamentaux des personnes était minime, puisque, même si des données personnelles étaient collectées, celles-ci étaient, avant leur analyse, immédiatement anonymisées. L'autorité de contrôle a eu une vision quelque peu différente de la situation, considérant que les techniques utilisées ne permettaient pas une anonymisation effective des données.

S'agissant des données audio, elle a estimé que « la simple substitution de la voix du locuteur n'est en aucun cas adaptée à l'anonymisation des données à caractère personnel relatives à une conversation, étant donné qu'il est possible de deviner du contenu de la conversation des informations relatives au locuteur ainsi qu'à des tiers et que ces informations peuvent rendre identifiables le locuteur, ses interlocuteurs ou les tiers auxquels il est fait référence dans la conversation ». S'agissant des images, l'autorité de contrôle a considéré comme insuffisante la technique du floutage des visages, les personnes concernées pouvant être identifiées par notamment « d'autres caractéristiques ou éléments contextuels (tels que la corpulence, l'habillement, la position dans la scène filmée, des caractéristiques physiques particulières, etc.) ». Quant aux auteurs des messages et commentaires publiés sur les réseaux X et YouTube, ces derniers étaient simplement pseudonymisés, puisque la municipalité

s'était contentée de remplacer leur nom par un numéro ID.

L'autorité italienne de protection des données a donc considéré que les deux projets impliquaient des traitements de données personnelles et que, par conséquent, ces traitements auraient dû, « tout au long de leur cycle vie », être conformes au RGPD, ce qui n'était pas le cas en l'espèce, faute notamment d'avoir respecté les principes du RGPD. Le montant de l'amende a été fixé à 50 000 euros.

### Quelles recommandations ?

L'IA – « tout outil utilisé par une machine afin de reproduire des comportements liés aux humains, tels que le raisonnement, la planification et la créativité » – est (et sera) de plus en plus présente dans nos vies, avec le développement de nouveaux produits ou services. Elle repose sur des algorithmes extrêmement consommateurs de données et son usage nécessite donc le respect de certaines précautions. De deux choses l'une : soit les données utilisées par ces algorithmes sont personnelles et, dans ce cas, il est vivement recommandé d'appliquer les principes du RGPD et de la loi « Informatique et Libertés » (qui peut prévoir des exigences particulières, notamment s'agissant des données de santé) ; soit les données utilisées par ces algorithmes ont été anonymisées et, dans ce cas, il convient de s'assurer que les techniques utilisées permettent de rendre impossible toute identification des personnes concernées par quelque moyen que ce soit et ce, de manière irréversible.

### Alexandre FIEVEE

Avocat associé  
Derriennic Associes

#### Notes

- (1) <https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles>
- (2) <https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles>
- (3) Exemple donné par la Cnil : « Une base de données de CV où seuls les nom et prénoms d'une personne auront été remplacés par un numéro (qui ne correspond qu'à elle) permet d'individualiser cette personne. Dans ce cas, cette base de données est considérée comme pseudonymisée et non comme anonymisée. »
- (4) Exemple donné par la Cnil : « Une base de données cartographique renseignant les adresses de domiciles de particuliers ne peut être considérée comme anonyme si d'autres bases de données, existantes par ailleurs, contiennent ces mêmes adresses avec d'autres données permettant d'identifier les individus. »
- (5) Exemple donné par la Cnil : « Si un jeu de données supposément anonyme contient des informations sur le montant des impôts de personnes ayant répondu à un questionnaire, que tous les hommes ayant entre 20 et 25 ans qui ont répondu sont non imposables, il sera possible de déduire, si on sait que M. X, homme âgé de 24 ans, a répondu au questionnaire, que ce dernier est non imposable. »



Vous avez envie de vous exprimer sur un sujet qui vous tient à cœur, de partager votre analyse avec la communauté des lecteurs d'Expertises, d'exposer un point de vue différent sur un article déjà publié, de lancer un débat sur un thème émergent, ou simplement de commenter l'actualité du droit du numérique ?

Contactez la rédactrice en chef d'Expertises Sylvie Rozenfeld [sr@expertises.info](mailto:sr@expertises.info)