

Omar Ali Fdal
Statice GmbH
Eisenacher Str. 1,
10777 Berlin
Germany

Paris, le 13 février 2023

N/Réf. : AGE/BPS/CS231015

Monsieur Ali Fdal,

Vous avez sollicité un avis de la Commission nationale de l'informatique et des libertés (CNIL) concernant les travaux menés par la société Statice sur l'évaluation des risques résiduels de réidentification pour des jeux de données synthétiques. Les méthodes soumises à la CNIL doivent faire l'objet d'une publication sous le nom d'Anonymeter dans le cadre du symposium « Privacy Enhancing Technologies » (PETs), qui se tiendra à Lausanne en 2023.

Premièrement, le service de l'expertise technologique de la Commission accueille favorablement l'approche de Statice consistant en une évaluation pratique des risques de réidentification se basant sur les critères d'*individualisation*, de *corrélation* et d'*inférence* définis dans l'avis 05/2014 du G29 concernant les techniques d'anonymisation. L'analyse de la documentation fournie n'a pas révélé d'éléments suggérant que les méthodes proposées ne pourraient pas permettre de vérifier le respect des trois critères susmentionnés dans le contexte de la production et de l'utilisation de jeux de données synthétiques.

Deuxièmement, le service de l'expertise technologique tient à saluer l'approche adoptée par Statice consistant à se soumettre au processus de publication académique et de revue par les pairs de ses travaux, ainsi que la publication du code source correspondant. En effet, l'évaluation indépendante, la large diffusion et l'amélioration continue des outils présentés sont le meilleur moyen d'assurer leur robustesse et leur adoption.

Troisièmement, bien qu'Anonymeter semble être un outil prometteur pour l'évaluation des risques résiduels de réidentification pour les ensembles de données synthétiques, le service de l'expertise technologique souligne que d'autres indicateurs issus de la littérature scientifique pourraient utilement lui être associés. À titre d'exemple, l'évaluation du risque de réidentification réalisée sur des exemples choisis (tels que les valeurs aberrantes) pourrait s'avérer complémentaire aux méthodes d'évaluation globale proposées par Statice.

RÉPUBLIQUE FRANÇAISE

3 Place de Fontenoy, TSA 80715 - 75334 PARIS CEDEX 07 - 01 53 73 22 22 - www.cnil.fr

Par ailleurs, les services de la Commission souhaitent attirer votre attention sur les points suivants :

- La génération d'un jeu de données synthétiques basé sur des données personnelles constitue un traitement de données à caractère personnelles. Il est donc nécessaire de s'assurer de sa conformité avec la réglementation en vigueur en matière de protection des données, à savoir le règlement général sur la protection des données (RGPD) dans l'Union européenne.
- L'anonymat d'un jeu de données synthétiques ne peut être déterminé qu'au cas par cas, c'est-à-dire pour chaque jeu de données généré, et ne doit donc pas être présumé à partir d'analyses effectuées sur d'autres jeux de données provenant du même fournisseur ou du même outil de synthèse de données.
- Les résultats produits par l'outil Anonymeter ont vocation à être utilisés par le responsable du traitement des données pour décider si les risques résiduels de réidentification sont acceptables ou non, et si l'ensemble de données peut être considéré comme anonyme. Cependant, en tant que fournisseur de solutions, la société Statice devrait également fournir des éléments pratiques décrivant précisément comment utiliser l'outil et interpréter les résultats obtenus (tels que des exemples, des tutoriels, des seuils, etc.).
- Les domaines scientifiques de l'anonymisation et de la réidentification des données sont en perpétuelle évolution. Les méthodologies existantes doivent donc être régulièrement réévaluées pour tenir compte des derniers développements.

En conclusion, le service de l'expertise technologique estime qu'Anonymeter est un outil prometteur et pertinent dans le contexte de la protection des données personnelles. Il recommande donc aux chercheurs et institutions de le tester dans divers contextes, afin de confirmer son utilité et sa fiabilité, et d'affiner les critères d'évaluation et les seuils acceptables.

Je vous prie d'agréer, Monsieur, l'expression de mes salutations distinguées.

Bertrand Pailhès,
Directeur des technologies et de l'innovation

