Bac à sable

« IA et services publics »

Les recommandations de la CNIL aux lauréats

Ce document constitue une **synthèse des principales recommandations** formulées aux porteurs de projet lors de leur accompagnement « bac à sable ». Ces recommandations, applicables à la date de rédaction du document, s'appuient sur les informations communiquées par les porteurs de projet et ses échanges avec la CNIL.

Elles sont publiées pour en faire **bénéficier les acteurs de la filière et aider les innovateurs à développer leur solution sur des projets similaires**.

Mars 2025

CNIL.
COMMISSION NATIONALE
INFORMATIQUE & LIBERTÉS

Table des matières

Le projet « Conseils Personnalisés » de France Travail : aider les conseillers France Travail à proposer un parcours de formation personnalisé et adapté aux besoins des demandeurs d'emploi

Date de rédaction : décembre 2024

La CNIL a accompagné France Travail de janvier à juin 2024 sur le projet « Conseils Personnalisés »**, un outil de recommandation de formations pour le demandeur d'emploi** sur la base de son profil et des instructions du conseiller. L'offre de formation étant conséquente et les outils multiples, « Conseils Personnalisés » vise à faciliter la sélection des services pertinents et adaptés à la situation et aux besoins des demandeurs pour les professionnels de l'emploi. L'outil peut recommander des formations sur des qualités interpersonnelles (ex : prise de parole en public) ainsi que des évènements locaux du réseau France Travail (ex : un forum de l'emploi local).

Cet outil utilise un système d'intelligence artificielle (IA) composé d'un grand modèle de langage (« *LLM* ») développé par le responsable de traitement sur la base d'un modèle de fondation alimenté en RAG (génération augmentée de récupération)[1]. Ce RAG comprend un catalogue des formations de France Travail, le « profil France Travail » du demandeur d'emploi et les précisions apportées par le conseiller, utilisateur de l'outil, via des invites (ou « *prompts* ») pour affiner les demandes.

Les recommandations de la CNIL sur cet outil se sont inscrites dans le contexte d'une utilisation du modèle Mixtral[2] déployé « *on premise* » (i.e en installation locale), tel que choisi par France Travail après une étude comparative de différents modèles de langage. La CNIL souligne que l'utilisation d'un modèle « *off-premise* » (i.e. service utilisé directement dans l'environnement réseau du fournisseur de licence) présente des risques de perte de confidentialité des données à caractère personnel traitées. À cet égard, elle a récemment publié sur son site des recommandations concernant le déploiement d'une IA générative respectueuse de la vie privée et détaillant ses positions sur les infrastructures à privilégier en fonction des risques que font peser les utilisations envisagées[3].

L'accompagnement s'est principalement matérialisé par la tenue de réunions mensuelles permettant de faire le point sur l'état d'avancement du projet, les évolutions envisagées et traiter les questions des deux parties. L'équipe projet de France Travail était composée de profils juridiques et techniques : délégué à la protection des données, juristes et chargés de mission en protection des données et IA, ingénieurs et chefs de projets IA, experts métiers de l'accompagnement de l'offre d'emploi et des services demandeurs d'emploi et experts des enjeux éthiques.

L'accompagnement – qui a uniquement porté sur le projet « Conseils Personnalisés », sur la base des informations communiquées par France Travail à la CNIL – a principalement porté sur les points suivants.

1. Comment assurer une intervention humaine significative pour éviter les décisions individuelles automatisées ?

L'article 22 du RGPD garantit aux personnes concernées « *le droit de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé, y compris le profilage, et produisant des effets juridiques la concernant ou l'affectant de manière significative de façon similaire* ». Si la réglementation interdit ce type de décision par principe, elle prévoit certaines exceptions sous réserve de la mise en place de garanties pour les droits et libertés des personnes. **Le respect de ce droit est apparu comme un axe structurant du projet « Conseils Personnalisés » en ce que l'outil s'inscrit dans le cadre du service public de l'emploi**

[1] La génération augmentée de récupération (*Retrieval Augmented Generation*) consiste à intégrer un mécanisme de recherche d'informations dans une base de données vectorisée (ou *embedding*). Elle permet de produire des réponses enrichies par des données externes, potentiellement plus spécifiques et plus faciles à actualiser que le modèle lui-même.
[2] https://mistral.ai/fr/news/mixtral-of-experts/
[3] https://www.cnil.fr/fr/comment-deployer-une-ia-generative-la-cnil-apporte-de-premieres-precisions

proposé sur le territoire français, et qu'il appuie les propositions de formations de l'agent France Travail[4].

Sur la base de l'arrêt Schufa de la CJUE[5], l'équipe de la CNIL a analysé si les trois éléments de qualification de l'article 22 étaient réunis en l'espèce :

1. **L'existence d'une décision** : les recommandations produites par l'outil influenceront les formations que le conseiller choisira de proposer au demandeur d'emploi. Aussi, la suggestion de l'algorithme s'apparente à un « avis » ou à une « proposition », définition proposée par l'avocat général dans l'arrêt Schufa. Par ailleurs, la CJUE rappelle que la notion de « décision » n'est pas définie par le RGPD et doit s'entendre de manière large comme étant « *susceptible d'inclure plusieurs actes pouvant affecter la personne concernée de multiples manières* ». Par conséquent, il semble que **la notion de « décision » puisse être retenue**.

   Le profilage : l'outil inclut des paramètres personnalisés, spécifiques à la personne concernée, afin de proposer le service qui lui sera le plus adapté. Il évalue donc certains aspects de sa personnalité mais ne produit pas d'estimation de son comportement futur. Il convient de noter cependant qu'un profilage au sens du RGPD impliquant un traitement automatisé, la qualification du profilage dépend donc de celle de « entièrement automatisé ».

2. **Le fait que la décision soit fondée exclusivement sur un traitement automatisé** : le traitement pourrait être qualifié de traitement entièrement automatisé si le responsable de traitement ne mettait pas en place **des garanties assurant l'autonomie des agents** et/ou si une analyse de leurs décisions montrait qu'ils proposent systématiquement, par exemple, la première proposition de l'algorithme au demandeur d'emploi. **L'outil ne sera pas qualifié de traitement entièrement automatisé si l'intervention de l'agent France Travail après la réception de la suggestion de l'outil est significative.**

   Pour que cette intervention significative soit possible, la CNIL a identifié deux aspects à prendre en compte :

   - **l'intelligibilité du fonctionnement de l'outil**, notamment en veillant à ce que les suggestions du système soient dans un format clair et manipulable ; et

   - **la mise à disposition aux agents FRANCE TRAVAIL des informations nécessaires pour leur permettre de comprendre les principes de fonctionnement de l'outil au moyen de documentation et de formations, ainsi que des conditions de travail adéquates leur permettant d'exercer leur pouvoir de jugement**. Ceci vise à leur permettre d'identifier les erreurs (comme par exemple les hallucinations) et d'encourager la vérification, ce qui renforcerait leur autonomie vis-à-vis des suggestions de l'outil. Ces mesures prendraient la forme de sessions de sensibilisation au fonctionnement et aux enjeux de l'outil, de la possibilité d'obtenir un soutien sur son utilisation (notamment, bénéficier d'un point de contact). Par ailleurs, aucune pression ne saurait être exercée par l'environnement de travail sur l'agent, afin qu'il ait le temps et la liberté de prendre une décision différente de la suggestion du système. Ceci peut être accentué par une information au droit de ne pas suivre les suggestions jugées comme non pertinentes.

   Afin d'assurer l'autonomie des agents, FRANCE TRAVAIL a élaboré un parcours de sensibilisation comprenant une formation continue et générale sur l'IA, un dispositif d'accompagnement spécifique à l'utilisation de « Conseils Personnalisés », ainsi que l'instauration de « correspondants IA » agissant comme canal de communication entre les agences et la direction générale. **Aussi, à ces conditions, la proposition de formations aux demandeurs d'emploi à l'aide de l'outil ne semble pas pouvoir être qualifiée de traitement entièrement automatisé.**

---

[4] Pour en savoir plus sur les enjeux des décisions automatisées et du contrôle humain : https://linc.cnil.fr/13-controle-humain-decisions-hybrides-quels-enjeux

[5] Arrêt du 7 décembre 2023, Schufa, C-634/2, ECLI:EU:C:2023:957

3. **La production d'effets juridiques concernant la personne ou l'affectant de manière significative de façon similaire** : l'outil ne semble pas entrainer de telles conséquences puisqu'une offre de formation est toujours proposée à l'usager. En effet l'outil propose un choix entre différentes formations, mais il ne détermine pas le fait d'avoir accès au service lui-même. Par ailleurs, ces formations aident le demandeur d'emploi à améliorer son profil mais ne régissent pas l'accès à des offres d'emploi (l'outil ne propose pas d'offre d'insertion en milieu professionnel). De plus, « Conseils Personnalisés » permet de proposer des formations auxquelles l'agent France Travail n'aurait pas pensé ou dont il n'aurait pas connaissance, mais n'empêche pas celui-ci de proposer toute autre formation qui lui paraitrait appropriée. Si les formations proposées par l'outil ne sont pas neutres pour le demandeur d'emploi et que le contexte de service public accentue l'importance des actions de France Travail vis-à-vis des usagers, le processus de retour à l'emploi est complexe et dépend de multiples paramètres. Aussi, il est difficile de mesurer le lien de causalité entre ces propositions de l'outil et le retour à l'emploi.

Conclusion

Par conséquent, les services de la CNIL ont considéré, en l'espèce, que l'utilisation de l'outil ne semble pas pouvoir entraîner de décision ayant des effets juridiques ou affectant la personne de manière significative de façon similaire et n'est par conséquent pas interdit par principe par l'article 22.

2. Comment assurer la pertinence des résultats et l'efficacité de l'outil dans le respect du principe de minimisation des données (article 5.1.c RGPD) ?

Afin d'assurer la minimisation des données d'une IA générative, un travail d'identification des données personnelles pouvant être utilisées par l'outil et les différentes sources de ces données est nécessaire.

1. **Délimitation des données nécessaires à la conception et au fonctionnement de l'outil**

- Catalogue de formations : l'ensemble des formations proposées par France Travail et ses partenaires (collectivités, associations, APEC, etc.) sont disponibles dans un catalogue regroupant plus de 20 000 formations. **Afin d'éviter que certaines offres soient invisibilisées par rapport aux autres et de veiller à ce que les bases de données soient nettoyées et à jour, la CNIL recommande d'harmoniser le format des offres issues de différentes bases de données.**

- Dossier personnel du demandeur d'emploi : ce dossier est en partie alimenté par le conseiller France Travail. Les équipes de la CNIL et de France Travail ont passé en revue l'ensemble des données issues du dossier personnel du demandeur d'emploi afin d'analyser leurs sources et leur utilité pour l'outil. Un travail de granularité a été fait afin d'assurer la minimisation des données ce qui a permis d'identifier certaines données qui ne sont pas utiles pour l'outil et pouvant présenter le risque de générer des biais voire des discriminations.

2. **Rédaction d'invites respectueuses de la protection des données personnelles et de la vie privée**

Minimisation du contenu des invites : le contenu des invites d'une IA générative étant par défaut totalement libre, les agents peuvent renseigner de nombreuses informations contenant des données à caractère personnel sur le demandeur d'emploi. Il convient toutefois de limiter celles-ci et d'éviter celles relatives, par exemple, à sa situation familiale, son état de santé ou s'il est en situation de réinsertion sociale.

Pour répondre à ce risque de perte de confidentialité, la CNIL a proposé plusieurs solutions potentiellement cumulatives :

- concevoir des **invites harmonisées et standardisées,** c'est-à-dire de fournir des modèles à usage des agents afin de les contraindre par le format à renseigner le moins de données personnelles possibles. Une telle solution serait protectrice mais pourrait réduire la flexibilité de l'usage de l'outil et ainsi diminuer l'intérêt de recourir à une IA générative ;

- mettre en place des **filtres de protection ou une liste noire de mots interdits**, bloquant l'utilisation de termes relatifs à des données sensibles ou hautement personnelles ;

- intégrer dans l'interface **des « alertes » sous forme de notifications** se déclenchant lorsque certains mots sont entrés, signalant par exemple aux agents qu'ils renseignent une information personnelle.

- ***a minima*, diffuser un message d'information sur les bonnes pratiques**, alertant notamment l'agent sur l'utilisation de données sensibles, devrait être intégré à l'outil ainsi qu'une **formation à la rédaction d'invites** ou à tout le moins de **sensibilisation** sur la manière d'obtenir une invite à la fois pertinente et respectueuse de la confidentialité.

La mise en place de filtres de protection ou d'alertes de détection de contenu à proscrire suppose de mettre en place des mesures techniques d'analyse automatisée du contenu d'entrée reposant sur du traitement automatique des langues, et sont donc dépendantes d'un travail technique des équipes de France Travail. Indépendamment de ces mesures, la CNIL a proposé de faire circuler une **charte d'usage et de bonnes pratiques à destination des utilisateurs de l'outil et de proposer une ou plusieurs formations sur la rédaction des invites**, comme garantie minimale de leur sensibilisation à préserver la confidentialité des données d'entrée, c'est-à-dire réduire autant que possible l'entrée de données à caractère personnel dans l'outil, même si elles ne peuvent être complètement exclues par principe pour une utilisation flexible de l'outil.

Conservation des invites :

Un besoin de conserver, par le responsable de traitement, les données d'entrée des invites collectées dans le cadre du déploiement de l'outil a été identifié. Ceci lui permettrait de faire des tests sur son outil et ainsi d'en faciliter l'audit, tant pour analyser sa performance globale que pour tester de potentiels biais dans la recommandation des formations. **La CNIL souligne notamment qu'en cas de conservation des invites pour ces finalités, le responsable de traitement devrait alors se fonder sur une base légale adéquate, mener une analyse de la compatibilité de cette finalité de traitement avec la finalité initiale de recommandation de formation (article 6.4 du RGPD) et, le cas échéant, fixer des durées de conservation adaptées.**

Conclusion

La conception d'une IA générative nécessite de cartographier l'ensemble des catégories de données personnelles qui alimenteraient l'outil afin d'assurer le respect du principe de minimisation et la pertinence des résultats. En outre, la CNIL rappelle que la conservation des données à des fins d'audit et d'amélioration doit être mise en balance avec les mesures de minimisation et de protection des données d'entrée. Cet arbitrage dépend des mesures protectrices que le responsable de traitement serait à même de mettre en place et de la solidité des arguments justifiant la conservation.

3. Quelles garanties mettre en place pour empêcher la survenance de biais pouvant entraîner des discriminations dans les résultats proposés (considérant 75 RGPD) ?

L'utilisation d'un outil de recommandation basé sur des éléments du profil du demandeur d'emploi, tels que son genre ou sa situation familiale, peut entraîner un risque de suggestions différenciées en fonction de ces

éléments, dans l'hypothèse d'un biais de l'outil ou d'une sensibilité accrue du système à certains de ces paramètres.

Un travail d'identification des éléments alimentant l'outil et pouvant mener à un traitement différencié, voire à des discriminations, doit être mené par le responsable de traitement. En l'espèce, certaines données géographiques, des données de genre inférées sur des patronymes ou des données sensibles tirées du parcours et de la vie privée du demandeur d'emploi ont par exemple été identifiées comme des sources de biais possibles.

À cet égard, la CNIL relève que le RGPD ne permet pas de lever l'interdiction de principe de traitement des données sensibles pour des finalités d'audit d'un système d'IA ; ce qui nécessite de tester les outils avec des données non sensibles mais potentiellement vectrices de discriminations. L'entrée en application du règlement européen sur l'intelligence artificielle (RIA) (article 10) pourrait permettre dans des cas bien définis de lever cet obstacle en offrant un fondement juridique à cet audit.

Après analyse des données personnelles traitées par « Conseils Personnalisés », les équipes n'ont pas identifié de risques particuliers de discriminations liés à l'usage prévu pour l'outil.

Conclusion

La question des risques de biais des IA génératives est un enjeu important - particulièrement dans une administration publique - et qui se situe à la croisée de différentes règlementations (RGPD, RIA, code des relations entre le public et les administrations, etc.). Les garanties à mettre en œuvre doivent être indexées sur le niveau de risque associé à la sensibilité des données, les sources des discriminations possibles et l'usage qui est fait de l'outil.

Le projet « Ekonom'IA » de Nantes Métropole : sensibiliser les habitants sur leur niveau de consommation d'eau grâce à un programme d'IA

Date de rédaction : mars 2025

La CNIL a accompagné Nantes Métropole de janvier à juillet 2024 sur son projet « Ekonom'IA ». **Ce projet vise à sensibiliser les abonnés à l'eau potable sur leur niveau de consommation d'eau en le comparant à celui d'un foyer présentant des caractéristiques similaires, grâce à un système d'IA**. L'objectif est notamment de pouvoir fournir aux usagers des recommandations contextualisées, sur leur facture d'eau.

Au sein de la métropole de Nantes, la distribution de l'eau fait l'objet d'une gestion en régie pour partie, et d'une gestion par un titulaire du marché public pour une autre partie.

L'accompagnement s'est principalement matérialisé par la tenue de **réunions régulières** permettant de faire le point sur les questions posées par les deux parties et les évolutions envisagées du projet. Ces ateliers ont permis de formuler des recommandations à l'attention du porteur du projet, ces dernières étant recensées dans le présent livrable.

Il a aussi donné lieu à une **intervention commune** lors de la journée du 8 février 2024 « IA et territoires », organisée à Paris La Défense, dans le cadre de la masterclass « Traitement des données et évolution de la réglementation en matière d'IA ». Une **seconde intervention commune**, également sous la forme d'une masterclass, dédiée au projet Ekonom'IA, s'est déroulée le 17 septembre 2024 au salon de la data et de l'IA à Nantes.

Les trois questions suivantes ont été traitées lors de l'accompagnement, dans le cadre du « bac-à-sable ».

1. Comment déterminer la base juridique du projet Ekonom'IA et permettre l'accès aux données ?

L'intelligence artificielle permet d'envisager de nombreux cas d'usages dans le domaine de l'action publique, en particulier pour les collectivités territoriales[6]. Le développement d'un dispositif d'intelligence artificielle requiert le plus souvent, pour l'entrainement des modèles, d'avoir accès à de larges bases de données pertinentes par rapport aux usages envisagés.

Le projet Ekonom'IA prévoit ainsi de collecter des données de sources internes (issues des contrats de fourniture d'eau) et externes (notamment des données détenues par l'administration).

La base légale du dispositif Ekonom'IA

À l'occasion des échanges avec la Métropole de Nantes, plusieurs bases légales ont été envisagées pour le projet Ekonom'IA. Cette dernière étant une collectivité, la mission d'intérêt public a été considérée en premier lieu :

- le traitement de données est entièrement tourné vers le service à l'usager en l'aidant à réduire sa consommation d'eau, à mieux identifier une déperdition éventuelle et par conséquent, à réaliser des économies sur sa facture d'eau ;

[6] Etude du Conseil d'Etat à la demande du Premier ministre, « Intelligence artificielle et action publique : construire la confiance, servir la performance », adoptée en assemblée générale plénière le 31 mars 2022 (https://www.conseil-etat.fr/publications-colloques/etudes/intelligence-artificielle-et-action-publique-construire-la-confiance-servir-la-performance).

- le traitement permet d'accomplir une mission en lien avec le service d'eau potable dont est investie Nantes Métropole, en qualité d'autorité publique[7]. Il vise à fournir un service à l'usager, en l'aidant à réduire sa consommation d'eau qui est un objectif directement en lien avec sa mission.

Les travaux ont permis de mettre en évidence que la base légale de la mission d'intérêt public apparaît la plus appropriée et est celle à privilégier.

Nantes Métropole a précisé que l'entraînement de l'algorithme à partir des données des abonnés n'est pas une finalité prévue à ce stade dans leur contrat d'abonnement à l'eau. Aussi, en basant ce traitement sur la mission d'intérêt public, il reviendra à Nantes Métropole de s'assurer de la compatibilité de ce traitement avec la finalité initiale (cf. article 6.4 du RGPD).

La problématique liée à la nécessaire collecte indirecte de certaines données

Le projet Ekonom'IA a été confronté à un obstacle pour l'accès à certaines données de sources externes, compte tenu de la nécessité de lever le secret fiscal et d'autoriser l'échange d'informations entre administrations.

Pour surmonter cette difficulté, la métropole et la CNIL ont organisé des échanges avec les administrations détentrices des données, ainsi qu'avec la Direction interministérielle du numérique (DINUM).

Ces échanges ont permis de dégager quelques pistes : le fondement juridique identifié à ce stade comme étant le plus prometteur serait le II de l'article L. 114-8 du Code des relations entre le public et l'administration (CRPA) relatif à l'échange de données entre administrations dans le cadre du bénéfice éventuel pour une personne à un avantage ou à une prestation.

Dans l'attente de l'avancée des réflexions sur cette question, Nantes Métropole poursuit ses travaux avec un jeu de données fictives qui a été constitué spécifiquement pour le projet Ekonom'IA.

Le projet Ekonom'IA a mis en lumière les obstacles liés au cadre juridique du partage de données entre administrations auxquels les collectivités peuvent être confrontées, compte tenu de la protection de secrets par la loi, en l'espèce le secret fiscal.

2. Comment procéder pour anonymiser et/ou pseudonymiser les données dans le but de mettre en œuvre un dispositif d'IA respectueux de l'éthique et des droits des personnes concernées ?

La pseudonymisation et l'anonymisation de données sont deux procédés qui permettent de limiter les risques pour les personnes. Ils sont toutefois très distincts : les données anonymes ne sont pas des données à caractère personnel alors que les données pseudonymes le sont.

Actuellement, la doctrine de la CNIL et de ses homologues est en large partie définie par l'avis 05/2014 du « groupe de travail Article 29 » (devenu, après l'entrée en application du RGPD, le Comité Européen de Protection des Données – CEPD).

La pseudonymisation

La pseudonymisation, permet de limiter la réidentification possible d'un individu au(x) seul(s) acteur(s) ayant connaissance des moyens et méthodes employés pour modifier la donnée personnelle de son état initial à son état transformé.

[7] Voir notamment les dispositions suivantes : l'article 17-2 de la directive (UE) n° 2020/2184 du 16 décembre 2020 (dite directive « eau potable », les articles L. 2224-7-5 et L. 2224-12-4 III bis du code général des collectivités territoriales (CGCT).

Le cas le plus fréquent est l'emploi d'une table de correspondance : la donnée personnelle initiale est remplacée par un autre identifiant. Une table de correspondance permet de relier ces deux identifiants.

La pseudonymisation ne retire pas son caractère personnel à une donnée et la règlementation en vigueur : la loi « informatique et libertés » et le RGPD continuent de s'appliquer.

L'anonymisation

L'anonymisation est un processus irréversible qui permet, quant à lui, de **retirer le caractère personnel des données.** Une réidentification de la personne est alors impossible pour quiconque, y compris par le responsable de traitement à l'origine de la collecte.

Pour vérifier si des données sont effectivement anonymes, il suffit de considérer **trois principaux risques** : **l'individualisation**, soit la capacité à isoler les données d'un même individu au sein d'un jeu de données ; la **corrélation**, soit la capacité de rattacher plusieurs données comme appartenant à la même personne ; **l'inférence**, soit la possibilité de déduire des données absentes d'un jeu à partir de celles présentes. Si la vraisemblance associée à chacun de ces risques est négligeable, alors les données peuvent être considérées comme anonymes[8].

Méthodologie du choix entre pseudonymisation et anonymisation

Les dispositifs d'IA nécessitent de larges jeux de données d'entrainement. En application du principe de minimisation du RGPD, il apparaît nécessaire de mettre en balance plusieurs éléments de manière à réduire les risques pour les droits et libertés des personnes concernées, à chaque étape, et d'identifier :

1. Les données à caractère personnel **strictement nécessaires** au traitement (principe de minimisation), qui seront traitées en l'état ;
2. Les données pouvant être **anonymisées** de manière à ce que le dispositif d'IA conserve une pertinence et puisse poursuivre ses finalités ;
3. Les données pouvant être **pseudonymisées** de manière à ce que le dispositif d'IA conserve une pertinence et puisse poursuivre ses finalités.

Nantes Métropole n'a pas pu appliquer pleinement cette méthodologie du fait des problématiques juridiques d'accès aux données et de l'impossibilité de travailler sur des données réelles.

Il a toutefois été possible d'effectuer des tests sur un jeu de données fictives (par exemple issues d'une modélisation). Nantes Métropole a ainsi généré des données fictives et a pu constater que les données conservaient les propriétés statistiques attendues, après anonymisation par deux méthodes différentes (l'une par génération de données synthétiques et l'autre par généralisation). Ces éléments devront être vérifiés dans le cadre de l'utilisation des données réelles[9]. En effet, les données réelles peuvent comporter des informations qui ne seraient pas contenues dans un jeu de données fictives ; par exemple :

- une corrélation possible entre des consommations en eau très largement supérieures à la moyenne ne pouvant correspondre qu'à un ou quelques bâtiments potentiels ;
- une individualisation possible d'un foyer aux nombres d'enfants à charge très largement supérieur à la moyenne ;
- une inférence possible sur des consommations atypiques pouvant être rattachées à des faits extérieurs (vétusté de la plomberie et dégâts des eaux dans certains quartiers, travaux, évènements climatiques sur une zone géographique spécifique, zone identifiée comme composée majoritairement de résidences secondaires, etc.).

[8] Pour plus de détails, voir la page suivante du site internet de la CNIL : https://www.cnil.fr/fr/technologies/lanonymisation-de-donnees-personnelles

[9] À la date de rédaction du présent document, le projet n'est pas en phase de déploiement

3. Transparence du dispositif d'IA : comment informer les personnes concernées et leur permettre d'exercer leurs droits ?

S'agissant de l'information des personnes concernées :

- Dans la mesure où la fourniture d'eau est un traitement distinct du projet Ekonom'IA, l'information des personnes sur le projet Ekonom'IA sera dissociée du contrat d'abonnement au service de l'eau ;
- Nantes Métropole envisage également une information dédiée des personnes concernées s'agissant d'Ekonom'IA sur chaque facture de consommation d'eau (récurrence annuelle), en complément[10].

Ces modalités d'information apparaissent conformes à la réglementation en matière de protection des données.

En complément, la mise en place d'une information sur différents supports (une page internet, des panneaux d'affichage en mairie, etc.) permettrait de maximiser la diffusion de l'information. Il est recommandé de proposer des supports adaptés et rendant l'information la plus compréhensible (ex : vidéos, animations), notamment pour les personnes vulnérables.

Il est aussi recommandé de définir séparément et de façon claire les principaux effets et les conséquences du traitement Ekonom'IA, par exemple en détaillant le fonctionnement du système d'IA.

Pour ce qui est de l'exercice des autres droits :

- Pour permettre la satisfaction d'objectifs importants d'intérêt public, l'article 23 du RGPD autorise l'Union et les États membres à limiter, par des « mesures législatives », la portée des droits qu'il prévoit (art. 12 à 22), en particulier le droit d'opposition des personnes concernées au traitement de leurs données ;
- S'agissant des finalités statistiques[11], il est aussi possible de déroger au droit d'opposition s'il risque de rendre impossible ou d'entraver sérieusement la réalisation des finalités poursuivies, et quand de telles dérogations sont nécessaires pour atteindre ces finalités.

Sous réserve de remplir les conditions pour les deux dérogations envisageables, Nantes Métropole pourrait écarter l'application du droit d'opposition pour le traitement visé. Nantes Métropole a toutefois fait le choix de ne pas écarter le droit d'opposition pour ce traitement.

[10] Depuis la fin de l'accompagnement, Nantes Métropole a poursuivi ses travaux sur cet axe. Il est notamment envisagé une information sur le site institutionnel et d'adapter le modèle de facture.

[11] Selon le RGPD, « *par « fins statistiques », on entend toute opération de collecte et de traitement de données à caractère personnel nécessaires pour des enquêtes statistiques ou la production de résultats statistiques. […] Les fins statistiques impliquent que le résultat du traitement à des fins statistiques ne constitue pas des données à caractère personnel mais des données agrégées, et que ce résultat ou ces données à caractère personnel ne sont pas utilisés à l'appui de mesures ou de décisions concernant une personne physique en particulier* » (cons. 162).

Le projet PRIV-IA de la RATP: étudier un traitement algorithmique d'images issues de nouvelles technologies de captation vidéo

Date de rédaction : janvier 2025

La CNIL a travaillé sur le projet « PRIV-IA » avec la RATP de janvier à avril 2024.

Ce projet a pour objectif d'étudier **une nouvelle forme de captation vidéo,** dénommée « **temps de vol** », destinée à développer **un système d'intelligence artificielle permettant de détecter certains événements à partir de représentations visuelles moins intrusives pour les personnes.**

Cette technologie permet d'obtenir des données via des capteurs différents de ceux utilisés dans des caméras classiques, mais qui peuvent fournir une représentation visuelle des espaces, des éléments présents (personnes ou objets), ainsi que des événements qui s'y produisent (passage d'un train, mouvement d'une personne, etc.).

Contrairement aux images de caméras classiques, les représentations visuelles issues de ces capteurs ne permettent généralement pas de reconnaître le visage des personnes présentes et limitent la possibilité de les identifier. L'objectif de ce projet est de développer un logiciel de traitement algorithmique qui puisse être mis en œuvre sur ces représentations visuelles limitant la possibilité d'identifier les personnes.

Les travaux ont principalement porté sur l'évaluation du caractère anonyme d'un traitement de données issues d'un capteur « temps de vol ». Ils ont également permis de préciser les modalités d'utilisation de ces capteurs à des fins statistiques mais aussi d'étudier l'intérêt de ces capteurs (ainsi que d'autres dispositifs équivalents) pour minimiser les données collectées.

Lors de l'accompagnement « bac à sable », l'architecture technique de cette solution a également été analysée avec la RATP, en parallèle de discussions sur l'utilisation de techniques similaires avec certains fournisseurs de solutions démontrant ainsi l'intérêt du secteur pour ces dispositifs. À la suite de ces discussions, **la CNIL a rappelé les principes et règles applicables à la mise en place de ces technologies.**

1. Les capteurs « temps de vol » : en quoi se distinguent-ils des caméras classiques ?

Le fonctionnement du capteur et l'exploitation algorithmique des données

Un capteur de caméra « classique » mesure les variations de luminosité dans la scène observée, qui sont converties en un signal numérique duquel peut être extraite une image. Dans le cas des capteurs temps de vol (ou *time of flight*), la mesure repose sur un fonctionnement différent, dit « actif ».

Généralement le dispositif fonctionne comme suit :

- composé d'un émetteur (laser, infrarouge) et d'un capteur orientés vers la même scène, le dispositif temps de vol émet tout d'abord un signal sur l'ensemble de la scène ;
- le capteur mesure la distance de chacun des « objets » présents dans la scène en évaluant le temps que met le rayonnement à revenir après ses « rebonds » sur les objets - d'où son appellation « temps de vol » ;
- de la même manière qu'une caméra classique, ces informations peuvent être transformées en signal numérique et peuvent être traduites en une représentation visuelle intelligible.

Comparaison d'images de la même scène issues d'une caméra classique (à gauche) et d'un capteur temps de vol (à droite) – Source : RATP

Membres de la famille des LIDAR (*Light Detection And Ranging*), les capteurs « temps de vol » sont utilisés pour des applications courantes depuis les années 2000 telles que pour l'amélioration de la qualité des prises de vue dans les ordiphones mais aussi dans le domaine de la robotique. Au-delà de ces applications, ces capteurs sont également considérés pour des cas d'usage de vidéo dite intelligente ou « augmentée ».

Grâce à l'ajout d'un traitement d'intelligence artificielle (IA) sur les données issues du capteur, il est possible d'analyser la scène et d'en tirer des informations utiles. Les algorithmes utilisés, issus du champ de la vision par ordinateur, sont similaires à ceux utilisés pour analyser des images « classiques », mais leur application aux données temps de vol est novatrice. Elle permettrait ainsi de tirer des informations utiles d'un point de vue opérationnel, tout en minimisant les données collectées. Les algorithmes d'apprentissage profond utilisés permettent, par exemple, de détecter une silhouette, d'identifier un objet, un animal ou un geste par de la classification. Ils permettent plus encore d'identifier et de décompter l'occurrence d'événements, comme décrit ci-dessous.

Les usages opérationnels envisageables pour les capteurs « temps de vol »

Grâce à la diversité d'objets et d'événements qu'ils permettent d'identifier, les capteurs « temps de vol » offrent un éventail d'usages pour de nombreux organismes. En particulier, ceux qui souhaitent observer une scène à distance afin de pouvoir mesurer ou détecter une action en vue de prendre une décision, ou d'étudier un phénomène sur le long terme.

Ces usages pourraient être les suivants :

- détection de la présence d'une ou plusieurs personnes aux fins d'ouverture d'un accès ou de déclenchement d'une alerte (ex. locaux professionnels ou commerciaux, tentative accès légitime ou non d'une personne ou véhicule) ;
- détection de mouvements, ou de l'absence de mouvements, susceptibles de révéler une anomalie (ex. accident d'une personne à terre, blocage d'un mécanisme utilisé par des personnes) ;
- détection, classification et dénombrement d'un flux de personnes et d'objets (ex. voitures, vélo, animaux, humains).

D'une manière plus générale, d'autres cas d'usage que ceux discutés avec la RATP semblent également envisageables, comme la détection de phénomènes naturels en zone extérieure (ex. phénomènes météorologiques, éboulement, crues, affaissement).

Les données mesurées par les capteurs temps de vol étant, par nature, moins identifiantes que celles issues de caméras classiques pour des cas d'usages similaires, **ces capteurs représentent un réel avantage en termes de minimisation**. Ils permettent par ailleurs de limiter les observations à une portion de la scène, en ne collectant que les points situés au-delà ou avant une certaine distance (par exemple entre 3 et 10 mètres).

2. L'évaluation du caractère anonyme dépend du contexte d'utilisation de la technologie

Il ressort de la réglementation en matière de protection des données que **l'évaluation du caractère anonyme d'un traitement de données ne peut être fondée sur le fonctionnement technique du dispositif pris isolément, mais doit être menée au cas par cas en tenant compte du contexte dans lequel il intervient et de sa finalité**. Il faut tenir compte à cet égard de l'ensemble des moyens susceptibles d'être raisonnablement mis en œuvre à des fins d'identification[12].

Dit autrement, le **caractère anonyme d'un dispositif de traitement ne peut être reconnu que dans un cadre d'usage suffisamment circonstancié :** il convient de tenir compte d'éléments de contexte, comme l'espace dans lequel est disposé le capteur, les personnes concernées, la finalité d'exploitation des données ainsi que les données tierces à disposition du responsable du traitement (ou d'un tiers destinataire des données dont le caractère anonyme est examiné) permettant la réidentification de la personne.

Les traitements mis en œuvre dans un espace sous vidéoprotection peuvent-ils être anonymes ?

Pour une très large part des cas d'usage envisageables dans le cadre de ce bac à sable, il s'est avéré que la technologie était susceptible d'être installée dans des espaces accessibles aux usagers, couverts par un réseau de caméras de vidéoprotection.

La question s'est donc posée de savoir si, même dans le cas où les données issues des capteurs « temps de vol » ne permettaient pas, seules, d'identifier une personne présente à l'image, l'existence d'un réseau de vidéoprotection dans les mêmes espaces conduisait à exclure le caractère anonyme de ces images.

En effet, la redondance avec un tel réseau de caméras rend possible un croisement d'informations avec les données « temps de vol » provenant d'une même scène et obtenues simultanément (celle du dispositif avec celle des caméras vidéoprotection). Cette possibilité peut d'ailleurs présenter des avantages opérationnels pour certains cas d'usage en permettant une levée de doute (par exemple, effectuer des vérifications à distance pour décider ou non de la nécessité d'une intervention).

Cela confirme toutefois qu'il est possible, dans de tels cas de figure, d'identifier les personnes dont les silhouettes sont représentées par le dispositif temps de vol au moyen d'un croisement avec les images recueillies par son dispositif de vidéoprotection[13]. Il n'importe pas que les données issues des deux flux visuels soient effectivement croisées, la seule possibilité de pouvoir croiser ces informations est suffisante pour considérer que les données ne sont pas anonymes (l'avis 05/2014 du G29[2] exige que le responsable soit « empêché » de réidentifier).

Conclusion

En conclusion, dès lors que les représentations visuelles de personnes captées par le dispositif « temps de vol » peuvent être croisées avec des images issues des caméras de vidéoprotection, **elles ne peuvent pas être considérées comme anonymes.**

[12] Voir le considérant 26 du RGPD et l'avis 05/2014 du G29 sur les techniques d'anonymisation.

[13] L'avis 05/2014 du G29 précise qu'« *une solution d'anonymisation efficace doit empêcher toutes les parties d'isoler un individu dans un ensemble de données, de relier entre eux deux enregistrements dans un ensemble de données (ou dans deux ensembles de données séparés) et de déduire des informations de cet ensemble de données* ». Le Conseil d'Etat (Conseil d'Etat, JC Decaux, 08/02/2017, 393714) a pu également rappeler ce principe en ces termes : « il résulte de la définition de la donnée personnelle [...] qu'une telle donnée ne peut être regardée comme rendue anonyme que lorsque l'identification de la personne concernée, directement ou indirectement, devient impossible [...]. Tel n'est pas le cas lorsqu'il demeure possible d'individualiser une personne ou de relier entre elles des données résultant de deux enregistrements qui la concernent »

Les traitements envisagés peuvent-ils être anonymes lorsqu'ils sont mis en œuvre pour intervenir auprès de la personne concernée ?

D'autres cas d'usages ont été examinés. Il s'agit notamment de l'utilisation de ces dispositifs, dans des espaces qui ne sont pas couverts par des réseaux de vidéoprotection, afin de détecter une intrusion de personnes ou des actes de vandalisme.

Dans ces cas de figure, la question qui s'est posée était de savoir si un **traitement de données ayant pour objet d'intervenir directement auprès d'une personne (dont les données sont traitées) constitue un traitement de données à caractère personnel**.

La définition de la notion de donnée à caractère personnel, éclairée par l'avis du G29 publié en 2007[14], souligne que la finalité du traitement est un facteur déterminant quant au caractère identifiant ou non des données. Cet avis précise que « *lorsque la finalité du traitement implique l'identification de personnes physiques, il est permis de penser que le responsable du traitement ou toute autre personne concernée dispose ou disposera de moyens « susceptibles d'être raisonnablement mis en œuvre » pour identifier la personne concernée. En réalité, prétendre que les personnes physiques ne sont pas identifiables alors que la finalité du traitement est précisément de les identifier serait une contradiction absolue* in terminis. »

L'approche de cet avis a été reprise et appliquée de manière continue depuis sa publication, notamment dans la jurisprudence nationale[15] et européenne[16].

Ainsi, même en l'absence de possibilité d'identifier la personne en croisant les données avec les images de vidéoprotection, l'intervention auprès de la personne détectée (provoquant un « effet » direct sur la personne) est précisément rendue possible par un ensemble de moyens prévus à cet effet. Ce sera ainsi le cas si le traitement algorithmique a vocation à détecter la présence ou le comportement d'une personne humaine et que ce signalement est associé avec la localisation de la personne afin de permettre, le cas échéant, une intervention par les personnes compétentes auprès d'elle.

Conclusion

Les services de la CNIL considèrent que, même en l'absence de croisement avec des images de vidéoprotection, les traitements de données réalisés sur des représentations issues des capteurs « temps de vol » **lorsqu'ils ont pour objet l'identification indirecte d'une personne dont les données sont recueillies, et sont susceptibles de produire un effet sur elle, sont des traitements de données à caractère personnel.**

Apport en matière de minimisation des données

Si l'utilisation de ces dispositifs dans les cas d'usage décrits ci-dessus ne conduit pas à écarter l'application de la règlementation en matière de protection des données, cela peut constituer en revanche une mesure forte pour la minimisation des données et la prise en compte de la protection de la vie privée dès la conception (articles 5.1.c et 25 du RGPD).

En ce sens, le recours à ces dispositifs, en lieu et place d'une caméra classique, peut favorablement peser dans l'analyse de la nécessité et de la proportionnalité du traitement, sous réserve de leur conformité aux autres exigences en matière de protection des données (notamment du droit d'opposition).

---

[14] G29, Avis 4/2007 sur le concept de données à caractère personnel, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_fr.pdf.

[15] Conseil d'Etat, décision n°441065 du 26 juin 2020, le juge a considéré, à propos des caméras dites thermiques, qu'un traitement porte sur une donnée personnelle même si la donnée collectée n'est pas la source ou le fondement de l'identification et si la mise en œuvre du traitement implique nécessairement que la personne concernée soit connue ou identifiée à cette occasion.

[16] CJUE, arrêt Nowak du 20 décembre 2017, C-434/16. La CJUE considère que constitue une donnée à caractère personnel toute sorte d'information, à condition que celle-ci « concerne » la personne identifiée ou identifiable en cause. Cette dernière condition « *est satisfaite lorsque, en raison de son contenu, sa finalité **ou son effet**, l'information est liée à une personne déterminée* »

3. Les traitements de données permettant d'atteindre une finalité statistique

La réflexion avec la RATP s'est portée sur deux catégories de traitements :

1. Ceux ayant pour finalité **l'amélioration des politiques de régulation des flux de transport et de gestion des stations ;**
2. Ceux ayant pour finalité **la régulation des flux de transport et la gestion des stations en temps réel.**

Ces catégories ont été analysées afin de déterminer si la qualification de finalité statistique leur était applicable, au regard des critères prévus par le RGPD.

Focus : le cadre spécifique aux traitements à des fins statistiques.

Comme indiqué par la CNIL dans sa position de 2022 sur les conditions de déploiement des caméras augmentées[17], les utilisations pour des finalités statistiques peuvent bénéficier d'un régime dérogatoire au titre duquel il est notamment permis, sous conditions, d'exclure le droit d'opposition des personnes. Afin de déterminer si le traitement peut se prévaloir de ce régime, les acteurs pourront utiliser les critères de définition d'un traitement statistique, détaillés au sein de la position « caméras augmentées » publiée en juillet 2022.

Toutefois, même s'il ne s'agit que de produire une information agrégée et statistique, le fait de construire cet indicateur par des images filmées dans des lieux publics n'est pas sans risque pour les droits. Ces traitements doivent ainsi faire l'objet d'une analyse pour s'assurer de leur licéité notamment en évaluant le caractère nécessaire et proportionné de leur usage, comme tout traitement de données à caractère personnel.

Tout d'abord**,** le résultat d'un traitement à des fins statistiques **ne peut pas, par définition, constituer des données à caractère personnel.** Les données sont agrégées et ce résultat ne peut pas être utilisé à l'appui de mesures ou de décisions concernant une personne physique en particulier[18]. La position de la CNIL sur les caméras « augmentées » précise cette condition en indiquant que le traitement ne peut être regardé comme uniquement statistique lorsqu'il tend par lui-même à une prise de décision immédiate. L'existence d'une telle décision peut conduire :

- au déclenchement immédiat d'une alerte ou d'une intervention (ex. intervention d'un agent en cas de densité trop importante de personnes ou de la présence non autorisée d'une personne dans un lieu spécifique ou du nombre trop élevé de détritus sur une zone) ;
- à l'affichage immédiat d'une publicité (ex. affichage d'une publicité vantant un article de sport lorsqu'une personne court devant le panneau) ;
- au blocage immédiat d'un accès à un lieu (ex. salle de concert, quai de gare, magasin, etc.) lorsqu'une certaine jauge est atteinte.

Un traitement n'a une finalité statistique que **s'il tend à la production de données agrégées pour elles-mêmes**. Le traitement doit avoir pour unique objet le calcul des données, leur affichage ou publication, leur éventuel partage ou communication.
Ainsi, l'information statistique produite ne doit pas permettre de remonter aux individus, de telle sorte que les mesures usuellement mises en œuvre pour anonymiser un traitement s'appliquent ici aussi :

- pour les cas d'usage nécessitant le décompte de personnes et la mesure des flux, assurer l'atteinte d'un nombre minimum de personnes avant d'enregistrer ce décompte est une condition nécessaire, bien que rarement suffisante, pour considérer qu'il s'agit d'une information anonyme ;

[17] Caméras dites « intelligentes » ou « augmentées » dans les espaces publics : position sur les conditions de déploiement, juillet 2022 : https://www.cnil.fr/fr/cameras-dites-augmentees-dans-les-espaces-publics-la-position-de-la-cnil
[18] Considérant 162 du RGPD.

- pour les cas d'usages nécessitant le recensement d'événements, il ne doit pas être possible de réidentifier les personnes concernées grâce aux informations remontées, notamment par leur recoupement avec des informations tierces telles que les images issues des dispositifs de vidéoprotection.

Les mesures de protection des données additionnelles restent toutefois à apprécier pour chacun des cas d'usages spécifiques.

Conclusion

L'analyse a conclu que **certains des cas d'usage des capteurs « temps de vol » envisagés pouvaient être qualifiés de traitements à des fins statistiques dans la mesure où ils visent à la production et l'exploitation de statistiques anonymes et que les conséquences du traitement ne ciblent pas directement les voyageurs dont les données ont été collectées.**