

Intelligence artificielle : le plan d'action de la CNIL

16 mai 2023

Devant les récentes actualités sur l'intelligence artificielle, et en particulier des IA dites génératives telles que ChatGPT, la CNIL publie un plan d'action pour un déploiement de systèmes d'IA respectueux de la vie privée des individus.

L'essentiel :

- La CNIL a engagé depuis plusieurs années des travaux pour anticiper et répondre aux enjeux soulevés par l'IA.
- En 2023, elle va prolonger son action sur les caméras augmentées et souhaite élargir ses travaux aux IA génératives aux grands modèles de langage et aux applications dérivées (notamment les chatbots).
- Son plan d'action s'articule autour de 4 volets :
 - appréhender le fonctionnement des systèmes d'IA et leurs impacts pour les personnes ;
 - permettre et encadrer le développement d'IA respectueuses de la vie privée ;
 - fédérer et accompagner les acteurs innovants de l'écosystème IA en France et en Europe ;
 - auditer et contrôler les systèmes d'IA et protéger les personnes.
- Ces travaux permettront également de préparer l'entrée en application du projet de règlement européen IA, actuellement en cours de discussion.

La protection des données personnelles, un enjeu fondamental dans le développement de l'IA

Le développement de l'IA s'accompagne d'enjeux en matière de protection des données et des libertés individuelles auxquels la CNIL s'attache à répondre depuis maintenant plusieurs années. Depuis la publication en 2017 de son [rapport sur les enjeux éthiques des algorithmes et de l'intelligence artificielle](#), la CNIL prononcée à plusieurs reprises sur les questions soulevées par les nouveaux outils amenés par cette nouvelle technologie.

En particulier, **les intelligences artificielles génératives** (voir encadré ci-dessous) se développent rapidement depuis plusieurs mois, que ce soit dans le domaine du texte et de la conversation, via les grands modèles de langage (*Large Language Models* ou *LLMs* en anglais), tels que GPT-3, BLOOM ou Megatron NLG et les agents conversationnels (« chatbots ») dérivés (ChatGPT ou Bard), mais également dans ceux de l'imagerie (Dall-E, Midjourney, Stable Diffusion, etc.) ou encore de la parole (Vall-E).

Ces modèles de fondation (*Foundation models* en anglais) et les briques technologiques qui se reposent sur eux semblent d'ores et déjà trouver de nombreux cas d'application dans des secteurs variés. Néanmoins, la compréhension de leur fonctionnement, de leurs possibilités et de leurs limites, ainsi que les enjeux juridiques, éthiques et techniques autour de leur développement et leur usage restent encore largement en débat.

Considérant que la **protection des données personnelles est un enjeu majeur pour la conception et l'utilisation de ces outils**, la CNIL publie son plan d'action sur l'intelligence artificielle qui vise – entre autres – à encadrer le développement des IA génératives.

Qu'est-ce qu'une IA générative ?

Une intelligence artificielle générative est un système capable de **créer du texte, des images ou d'autres contenus** (musique, vidéo, voix, etc.) **à partir d'une instruction d'un utilisateur humain**. Ces systèmes peuvent produire des nouveaux contenus à partir de données d'entraînement. Leurs performances sont aujourd'hui proches de certaines productions réalisées par des personnes en raison de la grande quantité de données ayant servi pour leur entraînement. Ces systèmes nécessitent toutefois que l'utilisateur spécifie clairement ses requêtes pour obtenir les résultats attendus. Se développe donc un véritable savoir-faire autour de la composition des requêtes de l'utilisateur (*prompt engineering*).

Par exemple, l'image ci-dessous, intitulée « Théâtre d'Opéra Spatial » a été générée par l'utilisateur Jason M. Allen grâce à l'outil Midjourney sur la base d'une instruction textuelle décrivant ses attentes (décor théâtral, toges, inspirations picturales, etc.).



Crédit : Jason M. Allen (2022), [licence CCo](#)

Un plan d'action en quatre volets

La CNIL a engagé depuis plusieurs années des travaux visant à anticiper et à répondre aux défis posés par l'intelligence artificielle, ses différentes déclinaisons (classification, prédiction, génération de contenus, etc.) et ses différents cas d'usage. Son nouveau **service de l'intelligence artificielle sera dédié à ces questions**, et appuiera les autres services de la CNIL qui sont également confrontés à des utilisations de ces algorithmes dans de nombreux contextes.

Face aux enjeux liés à la protection des libertés, à l'accélération de l'IA et à l'actualité liée aux IA génératives, la régulation de l'intelligence artificielle constitue un axe principal de l'action de la CNIL.

Cette régulation se structure autour de **quatre objectifs** :

- Appréhender le fonctionnement des systèmes d'IA et leurs impacts pour les personnes
- Permettre et encadrer le développement d'IA respectueuses des données personnelles
- Fédérer et accompagner les acteurs innovants de l'écosystème IA en France et en Europe
- Auditer et contrôler les systèmes d'IA et protéger les personnes

1. Appréhender le fonctionnement des systèmes d'IA et leurs impacts sur les personnes

Les techniques innovantes utilisées pour la conception et le fonctionnement des outils d'IA posent des questions nouvelles sur la protection des données, en particulier :

- **la loyauté et la transparence** des traitements de données sous-jacents au fonctionnement de ces outils ;
- **la protection des données publiquement accessibles sur le Web** face à l'utilisation du moissonnage, ou *scraping*, de données pour la conception des outils ;
- **la protection des données transmises par les utilisateurs** lorsqu'ils utilisent ces outils, allant de leur collecte (via une interface) à leur éventuelle réutilisation, en passant par leur traitement par les algorithmes d'apprentissage automatique ;
- **les conséquences sur les droits des personnes sur leurs données**, tant en ce qui concerne celles collectées pour l'apprentissage de modèles que celles qui peuvent être fournies par ces systèmes, telles que les contenus créés dans le cas d'IA génératives ;
- **la protection contre les biais et les discriminations** susceptible de survenir ;
- **les enjeux de sécurité inédits** de ces outils.

Ces aspects constitueront un des axes de travail prioritaires pour le service de l'intelligence artificielle et le laboratoire d'innovation numérique de la CNIL (LINC).

Dossier du LINC

Afin de souligner certains de ces enjeux spécifiques aux IA génératives, le Laboratoire d'innovation numérique de la CNIL (LINC) a publié [un dossier qui leur est consacré](#). Constitué de quatre volets, ce dossier :

- détaille le fonctionnement technique des agents conversationnels récents et rappelle la place centrale des données pour la constitution des modèles de fondation sous-jacents ;

- expose différentes questions juridiques posées par la conception de ces modèles, tant pour la propriété intellectuelle que pour la protection des données ;
- précise les enjeux éthiques des IA génératives pour la fiabilité de l'information, les utilisations malveillantes ainsi que les pistes de la détection et de l'avertissement du public quant à la présence de contenus ainsi générés ;
- illustre par différentes expérimentations les usages positifs ou négatifs qui peuvent être faits de ces outils.

Ce dossier complète [les ressources proposées par la CNIL](#) sur son site web pour les professionnels et le grand public.

2. Permettre et encadrer le développement d'IA respectueuses des données personnelles

De nombreux acteurs ont fait part à la CNIL de l'incertitude entourant l'application du RGPD à l'IA, notamment pour l'entraînement des IA génératives.

Afin d'accompagner les acteurs du domaine de l'intelligence artificielle et pour préparer l'entrée en application du règlement européen sur l'IA (en cours de discussion au niveau européen et [sur lequel la CNIL et ses homologues européennes avaient publié un avis en 2021](#)), la CNIL propose déjà :

- de [premières fiches sur l'IA, publiées en 2022](#) sur [cnil.fr](#), comprenant notamment des contenus pédagogiques sur les grands principes de l'IA et un guide pour accompagner les professionnels dans leur mise en conformité ;
- [une position, également publiée en 2022, sur l'usage de la vidéosurveillance « augmentée »](#) qui utilise l'IA sur des images de l'espace public.

Elle poursuit ses travaux doctrinaux et publiera prochainement plusieurs documents. Ainsi :

- la CNIL soumettra bientôt à une consultation un **guide sur les règles applicables au partage et à la réutilisation de données**. Ces travaux incluront notamment la question de la réutilisation de données librement accessibles sur internet et aujourd'hui utilisées pour l'apprentissage de nombreux modèles d'IA. Ce guide sera donc pertinent pour toute une partie des traitements de données nécessaires à la conception des systèmes d'IA, dont les IA génératives.
- elle poursuivra également ses travaux sur la conception de systèmes d'IA et la constitution de bases de données pour l'apprentissage automatique. **Ceux-ci donneront lieu à plusieurs publications à partir de l'été 2023**, à la suite de la concertation qui a déjà été organisée avec plusieurs acteurs, afin d'apporter des recommandations concrètes, notamment en ce qui concerne la conception des systèmes d'IA comme ChatGPT. Les thématiques suivantes seront progressivement abordées :
 - l'utilisation du régime de la recherche scientifique pour la constitution et la réutilisation des bases de données d'entraînement ;
 - l'application du principe de finalité aux IA à usage général et aux modèles de fondation que sont par exemple les grands modèles de langage ;
 - l'explicitation du partage des responsabilités entre les entités qui constituent les bases de données, celles qui élaborent des modèles à partir de ces données et celles qui utilisent ces modèles ;
 - les règles et bonnes pratiques applicables à la sélection des données pour l'entraînement, au regard des principes d'exactitude et de [minimisation](#) des données ;
 - la gestion des droits des personnes et notamment les droits d'accès, de rectification et d'opposition ;
 - les règles applicables concernant la durée de conservation, notamment pour les bases d'entraînement et les modèles les plus complexes à constituer ;
- enfin, consciente que les problématiques soulevées par les systèmes d'intelligence artificielle ne s'arrêtent pas à leur conception, **la CNIL poursuit également ses réflexions éthiques** sur l'utilisation et le partage des modèles d'apprentissage automatique, la prévention et la correction des biais et discriminations, ou encore sur la certification des systèmes d'IA.

3. Fédérer et accompagner les acteurs innovants de l'écosystème IA en France et en Europe

La régulation de l'IA de la CNIL vise à faire émerger, promouvoir et aider à prospérer des acteurs dans un cadre fidèle aux valeurs de protection de droits et libertés fondamentaux françaises et européennes. Cet accompagnement, déjà engagé, prend trois formes :

- la CNIL a lancé depuis 2 ans un « **bac à sable** » pour accompagner les projets et acteurs innovants, ce qui l'a notamment conduite à se pencher sur des projets reposant sur l'IA. Les « bacs à sable » [sur la santé en 2021](#) (12 projets accompagnés) et [sur l'éducation en 2022](#) (10 projets accompagnés) ont ainsi permis de fournir des conseils adaptés à des acteurs innovants de l'IA dans ces domaines. La CNIL ouvrira bientôt un nouvel appel à projet pour l'édition de 2023, qui concernera notamment l'usage de l'intelligence artificielle dans le secteur public ;
- elle a lancé un programme d'accompagnement spécifique des fournisseurs de vidéosurveillance « augmentée » dans le cadre de l'expérimentation prévue par la **loi relative aux Jeux olympiques et paralympiques de 2024** ;
- enfin, la CNIL a ouvert en 2023 [un nouveau programme « d'accompagnement renforcé »](#) pour assister des entreprises innovantes dans leur conformité au RGPD : les premiers lauréats de cet accompagnement renforcé sont des entreprises innovantes dans le domaine de l'IA.

Plus généralement, la CNIL souhaite engager un dialogue nourri avec les équipes de recherche, centres de R&D et entreprises françaises développant, ou souhaitant développer, des systèmes d'IA dans une logique de conformité aux règles de protection des données personnelles.

Ces équipes et entreprises peuvent prendre contact avec la CNIL à l'adresse ia@cnil.fr.

4. Auditer et contrôler les systèmes d'IA et protéger les personnes

La définition du cadre permettant le développement des systèmes d'intelligence artificielle dans le respect des droits et libertés individuelles implique, en aval, que la CNIL en contrôle le respect. Il est donc essentiel

pour la CNIL de développer un outillage permettant d'auditer les systèmes d'IA qui lui sont soumis et cela tant de manière à priori qu'à postériori.

L'action de contrôle de la CNIL portera notamment en 2023 sur :

- le respect de la [position sur l'usage de la vidéosurveillance « augmentée »](#), publiée en 2022, par les acteurs publics et privés ;
- l'usage de l'intelligence artificielle pour la lutte contre la fraude, par exemple pour la lutte contre la fraude à l'assurance sociale, au regard des enjeux liés à l'usage de tels algorithmes ;
- l'instruction de plaintes déposées auprès de la CNIL. S le cadre juridique de l'entraînement et de l'utilisation des IA génératives nécessite d'être clarifié, ce à quoi la CNIL va s'employer, des plaintes ont d'ores et déjà été déposées. La CNIL a, en particulier, reçu plusieurs plaintes à l'encontre de la société OpenAI qui gère le service ChatGPT, et a ouvert une procédure de contrôle. En parallèle, [un groupe de travail dédié a été créé au sein du Comité européen de la protection des données ou CEPD \(en anglais\)](#), en vue d'assurer une démarche coordonnée des autorités européennes et une analyse harmonisée des traitements de données mis en œuvre par l'outil d'OpenAI.

La CNIL sera particulièrement attentive à ce que les acteurs traitant des données personnelles afin de développer, d'entraîner ou d'utiliser des systèmes d'intelligence artificielle aient :

- réalisé une [analyse d'impact relative à la protection des données \(AIPD\)](#), pour documenter les risques et pris des mesures permettant de les diminuer ;
- pris des mesures d'information des personnes ;
- prévu des mesures d'exercice des droits des personnes adaptées à ce contexte particulier.

Grâce à ce travail collectif et essentiel, la CNIL souhaite instaurer **des règles claires, protectrices des données personnelles des citoyens européens afin de contribuer au développement de systèmes d'IA respectueux de la vie privée.**

Texte reference

Pour approfondir

> [Intelligence artificielle \(IA\)](#)

> [Création d'un service de l'intelligence artificielle à la CNIL et lancement des travaux sur les bases de données d'apprentissage](#)

> [\[EN\] Artificial intelligence: the action plan of the CNIL](#)