Cahier n° 43

l'académie
SCIENCES TECHNIQUES COMPTABLES FINANCIÈRES

# IA GÉNÉRATIVE ET PROTECTION DES DONNÉES : CONFIDENTIALITÉ, RGPD, SECRET PROFESSIONNEL

AVRIL 2025

Sage

ORDRE DES EXPERTS-COMPTABLES

# EDITOS

## Intelligence artificielle, données sensibles et secret professionnel : une responsabilité accrue

La généralisation de l'intelligence artificielle générative dans les organisations représente un tournant technologique majeur. Pour les professionnels du chiffre, ce progrès soulève des interrogations fondamentales, à la croisée du secret professionnel, du RGPD et de la confidentialité des données. Le cadre réglementaire européen – RGPD, AI Act, NIS2 – établit des garde-fous nécessaires, mais il ne peut à lui seul répondre à l'ensemble des défis que posent les traitements automatisés, parfois opaques, de données sensibles.

Dans les cabinets, les données personnelles, sociales, fiscales ou patrimoniales sont omniprésentes. Leur exposition – volontaire ou non – à des systèmes d'IA exige une vigilance renforcée, une gouvernance claire et une déontologie réaffirmée. La conformité ne peut être réduite à une formalité : elle suppose un engagement structurant, de la conception des outils jusqu'à leur usage quotidien, en passant par la formation des équipes et la rédaction des contrats.

Les experts-comptables ont toujours été les garants de la confidentialité. Ils doivent désormais l'être dans un monde numérique, automatisé et parfois fragmenté. Cela implique de ne pas subir l'innovation, mais de l'encadrer avec rigueur et discernement.

**Car on ne confie pas à une machine ce qu'on ne dirait pas en conscience à un confrère.**

**William NAHUM**

Président fondateur de l'Académie des Sciences
et Techniques Comptables et Financières

**Pour rester compétitives, nos entreprises doivent maîtriser l'IA, la donnée et le temps réel — dès maintenant.**

Alors que dans le cahier de l'Académie « L'Intelligence Artificielle générative et les Professions du Chiffre » mettait en avant l'arrivée de l'IAGen, l'apprentissage nécessaire pour l'utiliser et savoir bien l'utiliser et l'idée d'un bouleversement à venir dans toutes les organisations, les signaux que nous recevons du terrain sont sans appel : ce changement est déjà là. Ce ne sont plus seulement les entreprises qui s'interrogent, mais avant tout **leurs utilisateurs**, qui attendent des réponses concrètes, immédiates, et poussent leurs organisations à évoluer. Face à cette pression nouvelle, nous devons répondre aux besoins d'aujourd'hui tout en anticipant ceux de demain.

Notre mission, en tant qu'éditeur de solutions de gestion et de comptabilité, est claire : **accompagner les entreprises françaises et européennes** dans leur transformation pour renforcer leur compétitivité nationale. L'**intelligence artificielle** marque la troisième grande révolution numérique, après l'ordinateur et internet.

Son accès grand public, notamment via des outils comme **Sage Copilot**, rend l'IA générative plus visible, plus concrète — et plus attendue. En assistant les utilisateurs dans leurs tâches quotidiennes, en générant du contenu, en synthétisant l'information ou en facilitant la prise de décision, **Sage Copilot illustre comment l'IA peut transformer profondément la manière dont nous travaillons**.

Mais la technologie seule ne suffit pas. Le rôle des entreprises, des acteurs publics, des experts-comptables et des éditeurs est aussi de **former, d'accompagner**, et de permettre la montée en compétence de l'ensemble des professionnels — qu'ils soient déjà en poste ou qu'ils arrivent sur le marché. **La formation est urgente**, car l'IA transforme le travail dès aujourd'hui, dans tous les secteurs.

Dans le monde de la finance et de l'expertise comptable, nous atteignons enfin un cap essentiel : **l'accès à une donnée en temps réel**. Dans un monde qui va toujours plus vite, nos décisions doivent se baser sur des données à jour — pas sur des chiffres vieux de trois mois. La **data devient un actif stratégique**. Elle doit être juste, fiable, exploitable. Et c'est là que **la facture électronique** va jouer un rôle clé dans les mois à venir.

Enfin, impossible de parler d'intelligence artificielle sans évoquer **l'encadrement, l'éthique et la confiance**. L'IA ne peut être acceptée sans règles claires. L'humain doit rester **au centre des décisions**, et l'éthique doit guider chaque déploiement technologique. Car sans la confiance des utilisateurs, il n'y aura ni usage durable, ni impact réel.

En tant que partenaire de l'Académie des Sciences et Techniques Comptables et Financières, Sage ne peut qu'encourager l'Académie pour toutes les initiatives mises en avant aux bénéfices de la Profession du Chiffre et tous les membres des groupes de travail pour leurs implications.

Des remerciements particuliers aux membres du groupe de travail qui ont permis la réalisation de travaux, d'échanges et la parution de ce cahier « IA générative et protection des données » qui met l'accent important sur la notion de RGPD, confidentialité et protection des données.

**Nicolas Mellin**

Directeur de l'Innovation
Sage

# SOMMAIRE

Cet ouvrage a été écrit par le groupe de travail essentiellement au cours du 4e trimestre 2024.
La matière abordée est évolutive et certaines informations peuvent être obsolètes.
Une version électronique est disponible sur le site lacademie.info.

Cahier achevé en avril 2025

# PRESENTATION DU GROUPE

Chers consœurs, confrères, mémorialistes, collaborateurs, collaborateurs juridiques, comptables ou financiers, étudiants,

Vous avez entre vos mains un cahier qui est né des préoccupations de professionnels du chiffre sur la confidentialité des données (personnelles et professionnelles). Ce sujet souvent abordé a suscité de nombreuses questions dans la profession :

« que deviennent mes données ? »,

« où sont stockées les informations ? ».,

« peut-on déposer des fichiers d'écritures comptables (FEC) sur ChatGPT ? ».

L'expert-comptable est le gardien du temple contenant les données de ses clients. Les sujets de protection des données et de confidentialité ne sont pas nés avec l'IA générative mais sont inhérents à la digitalisation croissante des cabinets. Le passage à internet, le cloud ont introduit des questions sur le stockage des données et des e-mails.

Le RGPD a impacté les cabinets dans leur organisation et dans les flux de données avec les clients.

Les données sont le carburant des IA génératives. Les IA génératives traitent des quantités de données croissantes y compris des données personnelles. Comment concilier ce besoin en données avec les exigences de confidentialité, de protection des données, sans pour autant nuire aux innovations technologiques ?

Face à ces différents risques, des mesures juridiques, telles que le RGPD, les Digital Market Acts (DMA) et Digital Services Acts (DSA) sont-elles suffisantes ? Un rapport de l'assemblée nationale explique que « c*es différents textes ne s'appliquent qu'indirectement à l'intelligence artificielle et ne couvrent ni toutes les problématiques qu'elle pose, ni l'ensemble des usages qu'elle peut avoir. Ils viennent par ailleurs s'ajouter aux règlementations sectorielles applicables* »[1].

[1] Assemblée nationale, *Rapport sur les défis sur les défis de l'intelligence artificielle générative en matière de protection des données personnelles et d'utilisation du contenu généré*, 14/02/2024, page 32

En 2024, l'*IA act*, règlement européen sur l'intelligence artificielle (RIA) est venu en complément de ces mesures juridiques avec toute une série d'obligations qui impactent également les entreprises et professionnels du chiffre.

L'IA générative induit de nouvelles questions pour lesquelles nous espérons vous apporter quelques éléments de réflexion, à défaut de réponses certaines tant la matière est complexe et évolutive.

Un groupe de travail de l'Académie a consacré les sujets de l'éthique et de la confiance (voir cahiers n°38 et 39 de L'Académie). Partant du constat que le sujet suscite de nombreuses questions, ce cahier porte sur les aspects de protection des données, de confidentialité, RGPD et de secret professionnel.

Il a été conçu avec le support de plusieurs professionnels (avocat, juristes, spécialistes des systèmes d'information, professeur d'université…) avec des horizons différents et complémentaires.

Je remercie et salue la qualité des échanges et travaux réalisés par ce groupe et ses différents membres.

Merci également à L'Académie pour l'organisation matérielle et en particulier à William Nahum et Marie-Amélie Calmao.

**Vincent Lacomme**

# INTRODUCTION

*Par Jean-Laurent Heim-Lienhardt*[2]

## 1. Contexte de l'Intelligence Artificielle (IA)

### 1.1. Définition synthétique de l'IA et évolution des cas d'usage

L'Intelligence Artificielle (IA) peut se définir comme « l'ensemble de techniques informatiques visant à réaliser des tâches qui nécessitent habituellement l'intelligence humaine, telles que la reconnaissance d'images, le traitement du langage naturel ou encore l'apprentissage à partir de données » (Commission Nationale de l'Informatique et des Libertés [CNIL], 2020). Concrètement, l'IA repose souvent sur des algorithmes d'apprentissage automatique (machine learning) ou d'apprentissage profond (deep learning), lesquels permettent aux machines de s'adapter et de s'améliorer à mesure qu'elles traitent de larges volumes de données (Goodfellow, Bengio & Courville, 2016).

Initialement cantonnée à la recherche académique et à des applications militaires ou industrielles très spécialisées (p. ex. systèmes d'analyse d'images satellites), l'IA s'est progressivement diffusée dans de nombreux domaines. Aujourd'hui, elle tend à faire partie intégrante de l'entreprise : automatisation de processus, détection de fraudes, assistance à la clientèle via les chatbots, maintenance prédictive dans l'industrie, etc. Au-delà du cadre professionnel, elle pénètre également la sphère privée : applications de recommandations de contenus (musicales, vidéo), reconnaissance vocale (assistants personnels), maisons connectées (domotique), ou encore outils de retouche photo automatisés.

[2] Contenu partiellement rédigé à l'aide d'un système d'IA générative

### 1.2. Enjeux de la généralisation de l'IA : opportunités et risques

La généralisation de l'IA génère de multiples opportunités :

- **Optimisation et automatisation** : l'automatisation de tâches répétitives libère du temps pour des missions à plus forte valeur ajoutée et améliore l'efficacité opérationnelle (European Commission, 2021).
- **Personnalisation** : dans le marketing ou la relation client, les recommandations automatisées permettent d'adapter l'offre aux besoins spécifiques de chaque utilisateur.
- **Innovation** : la capacité d'apprendre à partir de données massives (big data) accélère la recherche et le développement de solutions inédites (ex. médecine de précision).

Cependant, l'usage intensif et parfois non maîtrisé de l'IA soulève également des risques, tant sur le plan technique que juridique :

- **Intrusions et fuites de données** : plus les systèmes d'IA s'interconnectent à des bases de données sensibles, plus ils deviennent des cibles de choix pour les cyberattaques (CNIL, 2020).
- **Biais algorithmique** : la qualité de l'IA dépend de celle des données sources ; des jeux de données incomplets ou non représentatifs peuvent aboutir à des discriminations (biais) (CNIL, 2017).
- **Manque de transparence** : certains algorithmes, notamment en *deep learning*, sont parfois considérés comme des "boîtes noires" rendant difficile l'explication des décisions prises, ce qui pose des questions d'audit et de conformité (European Commission, 2021).

Cet essor massif de l'IA, conjugué à une accélération technologique, justifie la mise en place de mesures de protection solides. Il s'agit autant de maîtriser la complexité technique de ces systèmes (sécurisation des algorithmes et des données) que d'en comprendre la portée juridique (respect des réglementations, gestion des responsabilités, etc.). Ce contexte nous permet ainsi de mieux comprendre la nécessité d'une approche globale : allier mesures techniques et cadre légal pour assurer un usage quotidien à la fois sécurisé et conforme aux réglementations, deux volets qui se veulent adressés par l'entrée en vigueur cadencée du RIA en 2025.

PARTIE 1

# CONFIDENTIALITÉ, RGPD, SECRET PROFESSIONNEL À L'HEURE DE L'IA GÉNÉRATIVE

# RAPPELS SUR LE CADRE REGLEMENTAIRE : RGPD, SECRET PROFESSIONNEL

Le lecteur pourra également consulter les publications du CNOEC sur le site Bibliordre :

✓ **La protection des données personnelles à l'usage des experts-comptables, octobre 2024**

✓ **Exercice professionnel et déontologie, septembre 2024**

***Par Jean-Laurent Heim Lienhardt***[3]

## 1. Rappel des principaux textes légaux et réglementaires

### 1.1. RGPD (Règlement Général sur la Protection des Données)

Le RGPD (Règlement (UE) 2016/679) est le socle de la protection des données personnelles en Europe. Il fixe des principes essentiels :

- Licéité, loyauté et transparence : tout traitement de données doit reposer sur une base légale (consentement, intérêt légitime, obligation légale, etc.), être réalisé de manière honnête, et les personnes concernées doivent être informées de façon claire.
- Minimisation et limitation de la finalité : seules les données strictement nécessaires doivent être collectées, et leur usage doit correspondre à la finalité initialement annoncée.
- Exactitude et conservation limitée : les données doivent être exactes, tenues à jour, et conservées pendant une durée strictement nécessaire aux finalités poursuivies.

#### *A. Obligations spécifiques pour l'IA*

En s'appuyant sur les dispositions du RGPD, lorsque l'IA est employée pour traiter des données personnelles, le respect de la protection des données doit être assuré dès la conception et par défaut (*privacy by design* et *by default*). Les entreprises ont l'obligation :

[3] Contenu partiellement rédigé à l'aide d'un système d'IA générative

- D'informer clairement les utilisateurs sur la nature du traitement IA (logique générale, finalité du modèle, etc.).
- De gérer le consentement ou de justifier le traitement par une autre base légale appropriée (intérêt légitime, exécution d'un contrat, etc.).
- De réaliser une évaluation d'impact sur la protection des données (DPIA) si le projet présente un risque élevé (par exemple, usage de données biométriques pour de la reconnaissance d'images).

### *B. Décisions automatisées et profilage*

Le profilage et la prise de décision automatisée sont particulièrement encadrés. L'article 22 (RGPD) prévoit que « la personne concernée a le droit de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé » si cette décision produit des effets juridiques ou l'affecte de manière significative (rejet d'un crédit, refus d'embauche, etc.).

- Les entreprises doivent donc s'assurer de laisser une place à l'intervention humaine dans les décisions importantes.
- Les droits des personnes incluent la possibilité de demander des informations sur la logique sous-jacente à l'algorithme (article 13 et 14) et de s'y opposer dans certaines conditions.

Concernant l'articulation de l'article 22 du RGPD et le RIA, il est à noter que l'article 3 de ce dernier dispose que les systèmes d'IA doivent pouvoir générer des résultats appartenant à l'une de ces quatre catégories (Prédictions, Recommandations, Contenus, Décisions automatisées).

### *C. Sanctions et responsabilités*

Le RGPD prévoit des sanctions administratives pouvant aller jusqu'à 20 millions d'euros ou 4 % du chiffre d'affaires annuel mondial (article 83 du RGPD), en cas de violation des droits des personnes concernées notamment. Les responsabilités sont partagées entre :

- Le responsable de traitement, qui définit les finalités et les moyens du traitement.
- Le sous-traitant, qui traite les données pour le compte du responsable.

Le RGPD encadre étroitement les traitements de données personnelles, ce qui inclut les projets d'IA. Ses principes (transparence, minimisation, licéité) et ses dispositions spécifiques (article 22 (cf. supra) sur les décisions automatisées, nécessités d'effectuer une DPIA, etc.) imposent aux entreprises de :

- Concevoir leurs solutions IA de manière éthique et respectueuse de la vie privée.
- Informer adéquatement les utilisateurs et leur permettre d'exercer leurs droits.
- Gérer avec rigueur les risques (intrusion, biais algorithmique, atteinte aux libertés) via un DPIA et la supervision d'un DPO.

L'IA, en manipulant parfois de grandes quantités de données sensibles ou en prenant des décisions potentiellement impactantes, constitue un enjeu majeur de conformité au RGPD. Une mise en œuvre responsable et maîtrisée est donc non seulement souhaitable, mais indispensable pour éviter les sanctions et maintenir la confiance des utilisateurs.

---

***Exemple : Une entreprise qui déploie un chatbot d'IA pour répondre aux questions des utilisateurs en collectant des données personnelles (nom, e-mail, historique de conversation)***

*Elle devra s'assurer que le traitement est fondé sur une base légale (souvent l'intérêt légitime ou le consentement), informer les utilisateurs de la finalité du traitement, et prévoir un mécanisme de suppression ou d'anonymisation des données après un certain délai.*

---

## 1.2. Directive NIS (Network and Information Security)

### *A. Origine et objectifs de la Directive NIS (2016/1148)*

Adoptée en juillet 2016, la Directive (UE) 2016/1148, dite « Directive NIS », est le premier texte européen à fixer des exigences communes pour la sécurité des réseaux et des systèmes d'information. Elle impose aux opérateurs de services essentiels (OSE) – tels que ceux de l'énergie, des transports, de la santé, ou encore de la finance – et aux fournisseurs de services numériques (e-commerce, plateformes cloud, moteurs de recherche) :

- La mise en œuvre de mesures techniques et organisationnelles adaptées à la gestion des risques (sécurité des infrastructures, plan de continuité…).
- L'obligation de notification aux autorités compétentes (ou aux CSIRT) en cas d'incident majeur, pouvant affecter la disponibilité ou la confidentialité des services.
  Selon l'article 14, ils doivent prendre « des mesures techniques et organisationnelles appropriées et proportionnées pour gérer les risques qui menacent la sécurité ».

### *B. Champ d'application et obligations*

La Directive NIS a instauré :

- **Des autorités nationales** : chaque État membre doit désigner une autorité ou plusieurs autorités chargées de surveiller la conformité.
- **Des CSIRT (Computer Security Incident Response Teams)** : cellules de réaction rapide aux incidents (attaques, brèches, etc.).
- **Des sanctions** : variables selon les pays, elles peuvent inclure des amendes administratives, voire des sanctions pénales en cas de négligence manifeste.

### *C. NIS2 : nouvelles dispositions et renforcement des obligations*

Publiée le 14 décembre 2022, la Directive (UE) 2022/2555, dite « NIS2 », abroge et remplace la première Directive NIS. Ses principaux apports sont :

- Extension du périmètre : elle couvre davantage de secteurs jugés critiques (fournisseurs d'infrastructures numériques, industries manufacturières essentielles, services postaux et de messagerie, etc.).
- Classification des entités : distinction entre « entités essentielles » (plus critiques) et « entités importantes ». Toutes doivent appliquer des mesures de sécurité proportionnées, mais les exigences en termes de sanctions et de supervision diffèrent.
- Renforcement de la supply chain : NIS2 insiste sur la prise en compte des risques tout au long de la chaîne d'approvisionnement, obligeant les entreprises à évaluer et sécuriser les relations avec leurs sous-traitants.
- Sanctions accrues : la nouvelle directive harmonise et alourdit les sanctions pour non-respect, encourageant une meilleure gouvernance en cybersécurité.

- Notification d'incident : l'article 23(NIS 2) impose aux entités essentielles et importantes de signaler tout incident majeur « sans retard indu » et dans les délais prévus par la législation nationale de transposition.

### *D. Enjeux pour les entreprises*

- **Gouvernance de la cybersécurité** : les structures soumises à NIS2 doivent clarifier la répartition des rôles (direction, RSSI, responsables métiers), mettre à jour leurs politiques et procédures internes, et intensifier leurs audits de sécurité.
- **Renforcement de la réactivité** : l'exigence de notification rapide d'incident implique de disposer d'outils de détection performants et de canaux de communication dédiés.
- **Coopération renforcée** : les entreprises doivent collaborer avec les autorités nationales, les CSIRT et parfois avec leurs pairs sectoriels pour prévenir et gérer les incidents.
- **Calendrier de mise en conformité** : la transposition de NIS2 dans le droit national doit intervenir courant 2025, laissant aux entreprises un délai pour se préparer, mais leur imposant aussi une vigilance rapide pour anticiper ces nouvelles obligations.

La Directive NIS a constitué un tournant majeur en imposant, pour la première fois, un socle commun de règles de cybersécurité aux opérateurs critiques et aux fournisseurs de services numériques en Europe. Avec NIS2, l'Union européenne renforce encore davantage les exigences et élargit le périmètre des entités concernées. Pour les entreprises, il s'agit désormais de mettre en place une gouvernance cyber solide, incluant une analyse de risques approfondie, des plans d'urgence, une politique de notification rapide en cas d'incident et une vigilance étendue à l'ensemble de leur chaîne d'approvisionnement numérique.

Cette évolution législative témoigne de la volonté de l'UE de muscler la cybersécurité collective et la résilience de son tissu économique. Les organisations sont ainsi encouragées à se doter sans attendre des ressources techniques, humaines et organisationnelles pour faire face aux menaces croissantes qui pèsent sur leurs systèmes d'information.

---

*Exemple : une société d'hébergement cloud hébergeant les données de plusieurs milliers de clients*

*Elle doit mettre en place un plan de continuité et de reprise d'activité (PCA/PRA), prévoir des mécanismes de suivi en temps réel (monitoring) et désigner un*

*responsable cybersécurité pouvant orchestrer la notification des incidents dans les délais impartis.*

---

## 1.3. L'IA Act (Règlement européen sur l'Intelligence Artificielle)

### *A. Origine et état d'avancement législatif*

Proposé le 21 avril 2021 par la Commission européenne (COM/2021/206 final), l'IA Act vise à établir des règles harmonisées pour l'usage de l'Intelligence Artificielle dans l'UE. Après l'adoption d'une position par le Conseil de l'UE en décembre 2022 et par le Parlement européen en juin 2023, le texte est entré en vigueur le 1er août 2024. Sa mise en application est échelonnée entre le 2 févier 2025 et le 2 août 2027.

### *B. Classification par niveau de risque*

Le règlement propose une typologie des systèmes IA :

- **Risque inacceptable** : systèmes interdits, car jugés contraires aux valeurs fondamentales ou présentant un risque extrême pour les droits et libertés (par exemple, manipulation cognitive d'utilisateurs vulnérables).
- **IA à haut risque** : systèmes utilisés dans des domaines sensibles (santé, sécurité, infrastructures critiques, justice, ressources humaines, etc.). Les fournisseurs devront respecter des obligations strictes de documentation, d'évaluation de conformité et de supervision humaine.
- **Risque limité ou minimal** : la plupart des applications de chatbot, de recommandation ou de traitement d'images se situent dans cette catégorie. Elles sont principalement soumises à des obligations de transparence (mention claire qu'un utilisateur interagit avec une IA).

Ce point sera développé ci-après au titre de ce cahier.

### C. Obligations clés pour les acteurs

L'*IA Act* différencie :

- **Fournisseurs** : développent ou mettent sur le marché le système IA. Ils doivent mettre en place un système de gestion des risques, garantir la fiabilité des données d'entraînement, rédiger une documentation technique et procéder à une déclaration de conformité (voire à une évaluation ex ante pour les cas les plus sensibles).
- **Utilisateurs** : entités ou personnes qui exploitent l'IA. Ils doivent respecter les consignes d'utilisation, surveiller le fonctionnement du système, signaler toute anomalie ou incident et informer les personnes concernées de leur interaction avec une IA si nécessaire.

Pour les systèmes à haut risque, une surveillance humaine est requise afin de prévenir les dérives ou biais potentiels de l'algorithme. Les obligations incluent aussi la notification d'incident majeur aux autorités compétentes, ainsi qu'un suivi post-commercialisation (assurer la mise à jour et la correction du système en continu).

### D. Impacts pour les entreprises

- **Adaptation organisationnelle** : création d'équipes ou de référents en charge de la conformité IA, mise en place de processus internes (évaluation des risques, validation éthique, test et audit **réguliers).**
- **Contrats et responsabilité** : insertion de clauses spécifiques (droits de propriété intellectuelle, qualité des données, maintenance des modèles, etc.).
- **Articulation avec le RGPD et NIS2** : le respect de l'IA Act ne dispense pas de se conformer aux règles relatives à la protection des données (RGPD) et à la sécurité des réseaux et systèmes (NIS2). Les entreprises doivent donc envisager une vision globale de leur gouvernance numérique pour éviter les doublons ou les lacunes.

L'IA Act est le premier cadre réglementaire complet dédié à l'Intelligence Artificielle au sein de l'UE. Bâti sur une approche par niveau de risque, il impose des exigences renforcées pour les usages potentiellement sensibles (sécurité, santé, droits fondamentaux). Les entreprises, organismes publics et éditeurs de solutions IA sont ainsi invités à préparer en amont leur mise en conformité, en prévoyant :

- Une gestion rigoureuse des risques, incluant la qualité des données et la traçabilité des algorithmes.

- Des process de supervision humaine, particulièrement pour les décisions impactant fortement la vie des personnes.
- Une documentation technique précise et des mécanismes de transparence vis-à-vis des utilisateurs.

---

***Exemple : une entreprise concevant un outil d'analyse d'images médicales (radiographies, IRM) pour aider à diagnostiquer des pathologies***

*Elle relève a priori de la catégorie « haut risque ». Elle devra documenter les algorithmes, prouver la fiabilité des datasets utilisés, organiser la supervision par un personnel médical et mettre à jour ses procédures de test à chaque évolution du modèle.*

---

## 2. Impacts sur les pratiques de l'entreprise

### 2.1. Définition des responsabilités et rôles

L'usage de l'Intelligence Artificielle (IA) implique nécessairement des traitements de données, souvent à caractère personnel. Le RGPD impose donc de préciser, au sein de chaque projet, qui est le responsable de traitement, qui est sous-traitant et dans quels cas plusieurs entités sont co-responsables.

#### *A. Responsable de traitement vs. sous-traitant*

Selon l'article 4 du RGPD, le responsable de traitement (RT) est celui « qui détermine les finalités et les moyens du traitement ». Il décide pourquoi et comment les données seront utilisées. À l'inverse, le sous-traitant (art. 28) agit pour le compte du RT, sans déterminer les finalités essentielles ni les choix structurants. En pratique :

- Responsable de traitement : une entreprise A qui veut analyser des données clients pour optimiser sa stratégie marketing.
- Sous-traitant : un prestataire B qui développe l'outil IA selon les spécifications d'A et n'utilise les données qu'aux fins définies par A.

### B. Co-responsables de traitement

L'article 26 du RGPD instaure la notion de co-responsabilité lorsque plusieurs entités déterminent conjointement la finalité et les moyens. Elles doivent alors conclure un accord organisant leurs rôles respectifs et leurs rapports avec les personnes concernées.

---

***Exemple***

*Deux partenaires conçoivent ensemble un algorithme de scoring pour attribuer des prêts. Ils choisissent ensemble quels critères sont utilisés, quelles données sont collectées et comment la décision est rendue.*

---

### C. Spécificités liées à l'IA

Dans le cadre de l'IA :

- Un développeur d'algorithmes peut être sous-traitant si sa mission se limite à exécuter les instructions du client.
- S'il participe activement à déterminer les finalités (par exemple, l'entreprise impose l'usage d'un modèle de scoring avec certaines variables), il risque d'être considéré comme co-responsable.
- Les fournisseurs de solutions standardisées (SaaS) peuvent parfois être responsables à part entière, surtout s'ils imposent la structure de traitement et gèrent les données de manière autonome.

### D. Conséquences sur la contractualisation

La détermination des rôles influe sur les contrats :

- Contrat de sous-traitance (art. 28 RGPD) : doit préciser l'objet du traitement, la durée, la nature, la finalité, le type de données et les mesures de sécurité.
- Accord de co-responsabilité : établit la répartition des obligations (information des personnes, exercice des droits, gestion des incidents).
- Des clauses spécifiques à l'IA sont souvent nécessaires (droits de propriété intellectuelle sur le modèle, mise à jour, corrections de biais, etc.).

**Point de vigilance**

En cas de violation des données ou d'abus, l'autorité de contrôle peut sanctionner l'entreprise ou le sous-traitant, selon la part de responsabilité de chacun.

Par ailleurs, l'IA Act est entré en vigueur pour les systèmes à haut risque, emportant des obligations de documentation et de supervision humaine qui viennent se cumuler aux obligations du RGPD.

Dans un projet IA manipulant des données personnelles, il est crucial de clarifier dès le départ qui est responsable de traitement, sous-traitant ou co-responsable, afin de répartir correctement les obligations (sécurité, confidentialité, information des personnes). Cette distinction a un impact majeur sur la rédaction des contrats, le registre des traitements et le risque de sanctions en cas de manquements au RGPD. L'articulation avec IA Act renforce encore la nécessité d'une gouvernance juridique solide et d'une collaboration fluide entre tous les acteurs impliqués.

## 2.2. Organisation interne : DPO et référents IA

### *A. Le rôle du DPO à l'ère de l'IA*

Le RGPD (articles 37 à 39) prévoit la désignation d'un Délégué à la Protection des Données (DPO) dans plusieurs cas, notamment lorsque le traitement à grande échelle porte sur des données sensibles ou lorsqu'il s'agit d'un organisme public, tout en précisant le rôle et les missions de ce dernier (Article 39). Avec la montée en puissance de l'IA, le DPO voit son champ d'intervention s'élargir :

- Conseil et information : il aide les équipes (techniques, métiers) à comprendre les obligations RGPD (finalité légitime, minimisation, consentement) et assure la promotion du privacy by design.
- Surveillance de la conformité : il s'assure que les traitements IA (profilage, scoring) respectent les principes de loyauté, de transparence et de maîtrise du risque.
- Analyse d'impact (DPIA) : il conseille et valide la réalisation des DPIA pour les projets IA sensibles (données de santé, reconnaissance faciale, etc.).
- Coopération avec l'autorité de contrôle (la CNIL en France) : il peut être amené à communiquer sur les violations de données ou répondre aux demandes d'information.

### *B. La fonction de référent IA*

En complément du DPO, la mise en place d'un référent IA (ou d'une équipe dédiée) apparaît de plus en plus stratégique :

- Expertise technique : le référent IA maîtrise l'architecture des algorithmes, les méthodes de collecte/traitement des données, l'évaluation de la performance et la détection des biais.
- Interface avec le DPO : il traduit les exigences RGPD en contraintes techniques (ex. : pseudonymisation, suppression automatique des données après un certain délai).
- Suivi des évolutions législatives : il se tient informé des développements autour de l'IA Act (classification des systèmes à haut risque, documents de conformité à fournir, etc.).

### *C. Comités ou instances de gouvernance*

Pour coordonner l'action du DPO, du référent IA et des autres parties prenantes (RSSI, direction juridique, experts métiers), beaucoup d'organisations mettent en place un comité IA (ou comité d'éthique). Son rôle :

- Valider les choix technologiques : s'assurer de la pertinence des algorithmes et de la qualité des jeux de données.
- Contrôler l'éthique et la transparence : vérifier que l'IA n'induit pas de discrimination, que les personnes sont informées de leurs droits, etc.
- Assurer la conformité globale : RGPD, IA Act, mais aussi NIS2 (notamment pour la sécurité et la notification des incidents).

### *D. Articulation avec le futur IA Act*

L'IA Act, impose des obligations additionnelles, en particulier pour les systèmes considérés « à haut risque ». Le DPO et le référent IA devront donc :

- Organiser la documentation (fiches techniques, registre IA, suivi post-commercialisation).
- Veiller à la supervision humaine : s'assurer que l'algorithme n'agit pas de manière autonome sur des décisions lourdes de conséquences (accès à l'emploi, crédit, etc.).
- Gérer la conformité croisée : articuler les exigences IA Act (conformité, tests de performance) avec celles du RGPD (licéité, droits des personnes) et éventuellement la sécurisation demandée par NIS2.

L'adoption d'une gouvernance interne structurée est un levier essentiel pour aligner conformité RGPD, innovation IA et prévention des risques. Le DPO demeure la pierre angulaire de la conformité en matière de protection des données, tandis que les référents IA (ou comités IA) assurent la cohérence technique et éthique des projets. Cette collaboration permet :

- Un pilotage en amont des projets IA, évitant les écueils juridiques (non-conformité, sanctions).
- Une supervision continue, notamment pour adapter les algorithmes et gérer les demandes d'exercice des droits.
- Une anticipation des futures obligations liées à l'IA Act (surveillance humaine, déclaration de conformité), en lien avec le RGPD et la sécurité (NIS2).

Ainsi, DPO et référents IA deviennent des partenaires stratégiques pour encourager une IA responsable, transparente et respectueuse des droits fondamentaux

---

***Exemple :***

*Dans une entreprise de e-commerce, le DPO reçoit régulièrement les rapports du référent IA sur l'algorithme de recommandation produit. Ensemble, ils vérifient que le traitement est conforme aux finalités annoncées, qu'un DPIA a été dûment rempli pour l'analyse des données comportementales, et que les clients peuvent exercer leurs droits (opposition, rectification, effacement).*

---

## 2.3. Rédaction et révision des contrats

La conception et l'utilisation d'outils d'IA s'inscrivent dans un cadre légal exigeant : RGPD pour la protection des données, et l' IA Act pour réguler les risques spécifiques à l'Intelligence Artificielle. Dès lors, la rédaction puis la révision des contrats occupent une place centrale pour clarifier les responsabilités, les obligations de sécurité et la conformité aux différentes réglementations.

### *A. Obligations fondamentales issues du RGPD*

**Contrat de sous-traitance (article 28) :**

- Le responsable de traitement doit formaliser un accord détaillant les instructions données au sous-traitant, les mesures de sécurité exigées, les modalités de notification en cas de violation de données et les obligations de confidentialité.
- Le sous-traitant ne peut agir que sur instruction documentée et doit fournir des garanties suffisantes pour protéger les données personnelles.

**Accord de co-responsables (article 26) :**

- Lorsque plusieurs acteurs déterminent conjointement les finalités et moyens du traitement, ils sont co-responsables. L'accord doit préciser leur répartition des tâches (qui répond aux demandes d'exercice des droits ? Qui gère la notification d'incident ?).
- Cet accord n'empêche pas la personne concernée d'exercer ses droits auprès de chacun des responsables.

### *B. Clauses clés à intégrer pour l'IA*

- **Confidentialité et sécurité** : mention explicite de l'article 32 RGPD, décrivant les mesures techniques (chiffrement, pseudonymisation) et organisationnelles (politique de confidentialité, gestion des accès).
- **Qualité et origine des données d'entraînement** : s'assurer que les données collectées ou achetées proviennent de sources légitimes et respectent les principes de minimisation et de licéité.
- **Maintenance et mises à jour** : prévoir une clause obligeant le sous-traitant ou le co-responsable à corriger les biais algorithmiques identifiés, à améliorer la performance du modèle et à documenter chaque évolution.
- **Transparence et supervision humaine**: pour les systèmes classés « à haut risque », stipuler comment sera mise en place la supervision humaine et comment les logs d'utilisation seront conservés.

### *C. Gestion du cycle de vie contractuel*

La mise en place d'un contrat ou accord n'est pas figée dans le temps :

- **Phase de rédaction** : effectuer une « due diligence » pour vérifier la capacité du partenaire (sous-traitant, co-responsable) à respecter le RGPD. Exiger la signature de clauses contractuelles appropriées si des transferts hors UE ont lieu (clauses contractuelles types).
- **Phase d'exécution** : opérer des audits réguliers pour contrôler la qualité des données et la sécurité de l'algorithme, instaurer une procédure de notification d'incident (cyberattaque, fuite de données).
- **Fin du contrat** : définir les modalités de restitution ou de suppression des données, la réversibilité technique (transfert des modèles IA), et la conservation d'une documentation suffisante pour d'éventuels contrôles ultérieurs.

### *D. Coordination avec les autres volets juridiques*

- Propriété intellectuelle : clarifier qui détient les droits sur l'algorithme, le code source et les éventuelles améliorations.
- Transferts de données hors UE : si le système IA est hébergé ou développé en dehors de l'UE, prévoir des garanties équivalentes au RGPD.
- NIS2 et sécurité : en cas d'opérateur de services essentiels ou d'entité importante (NIS2), intégrer au contrat des obligations de reporting en cas d'incident majeur et de tests de cybersécurité.

La rédaction et la révision des contrats constituent un levier majeur pour assurer le respect du RGPD et anticiper les exigences du futur IA Act. Les entreprises doivent veiller à :

- Identifier correctement les rôles (responsable de traitement, sous-traitant, co-responsables).
- Intégrer des clauses spécifiques à l'IA (origines des données, mise à jour des modèles, supervision humaine).
- Assurer une sécurité renforcée et une bonne traçabilité, notamment via des audits réguliers et la conservation de preuves (logs, documentation).
- Prévoir la fin du contrat (suppression ou restitution des données, transfert des connaissances).

Cette approche contractuelle offre un cadre juridique solide, garantissant la confiance des utilisateurs et la sécurité des traitements IA.

---

***Exemple***

*Une entreprise de e-santé conclut un contrat avec un éditeur IA basé aux États-Unis. Le contrat doit alors mentionner les clauses contractuelles types pour le transfert de données médicales (RGPD), fixer les modalités de supervision (IA Act) et définir la responsabilité de chaque partie en cas de non-conformité ou d'attaque informatique (NIS2).*

---

## 2.4. Évaluation d'impact sur la protection des données (DPIA)

L'article 35 du RGPD dispose que toute organisation doit mener une analyse d'impact (DPIA) lorsqu'un traitement est « susceptible d'engendrer un risque élevé » pour les droits et libertés des personnes physiques. Dans le cadre de l'IA, plusieurs critères déclenchent souvent cette obligation :

- Données sensibles ou biométriques (ex. : reconnaissance faciale).
- Profilage ou prise de décision automatisée ayant un effet significatif (octroi de crédit, embauche, etc.).
- Traitement à grande échelle de données (millions de profils).
- Surveillance systématique (analyse de comportements via caméras intelligentes, géolocalisation en temps réel).

### *A. Modalités de conduite et de validation*

La DPIA se déroule en plusieurs étapes :

- Description du traitement : finalités, acteurs impliqués, données utilisées.
- Évaluation de la nécessité et de la proportionnalité : vérifier la base légale (consentement, intérêt légitime) et la pertinence du volume de données collectées.
- Analyse des risques : identifier les menaces (violation de données, biais discriminants) et les dommages potentiels (atteinte à la vie privée, liberté d'opinion).
- Mesures d'atténuation : mise en place de solutions techniques (chiffrement, pseudonymisation), organisationnelles (formation, gouvernance IA).

Le DPO est un acteur clé, chargé d'évaluer la robustesse de la DPIA et de conseiller sur les options de réduction de risques. Si la DPIA conclut à un risque élevé non atténué, le responsable de traitement doit consulter l'autorité de contrôle (article 36 RGPD) avant de lancer le projet.

### B. Évolutions et archivage

Une DPIA n'est pas un document figé. À chaque évolution majeure du traitement (nouvelle finalité, changement d'algorithme, ajout de données sensibles), la DPIA doit être réexaminée et mise à jour. L'organisation est tenue de conserver la documentation associée (pour démontrer la conformité en cas de contrôle) et de s'assurer que les équipes restent informées des modifications apportées.

### C. Articulation avec L'IA Act

L'IA Act, imposera des obligations spécifiques pour les systèmes considérés « à haut risque ». Parmi celles-ci : la documentation technique, la gestion des risques et l'implémentation d'une supervision humaine. Dans la plupart des cas, une DPIA RGPD bien réalisée servira de base pour répondre aux exigences IA Act, en particulier s'agissant de :

- L'identification des risques (biais, discrimination, sécurité).
- Les mesures de mitigation (contrôles humains, audit algorithmique).
- La traçabilité des opérations et l'évaluation continue.

La DPIA constitue un outil clé pour maîtriser les risques liés au traitement de données à caractère personnel, particulièrement dans le cadre de projets IA où la sensibilité et le volume des informations traitées peuvent être très élevés. Correctement mise en œuvre, elle permet :

- D'identifier précocement les enjeux de confidentialité, de sécurité et d'éthique.
- De définir des mesures techniques et organisationnelles de protection (anonymisation, gouvernance, etc.).
- De préparer la conformité à d'autres réglementations (IA Act, NIS2) en offrant une vue d'ensemble sur la gestion des risques.

La réalisation (et la mise à jour régulière) d'une DPIA est donc non seulement une exigence légale du RGPD, mais aussi une bonne pratique pour renforcer la confiance dans l'usage de l'IA et protéger effectivement les droits fondamentaux des personnes concernées.

***Exemple***

*Une entreprise qui déploie un outil d'analyse sentimentale (IA) sur des e-mails clients doit réaliser une DPIA pour estimer le risque d'atteinte à la vie privée et s'assurer de la licéité du traitement. Si cette IA est considérée « à haut risque » selon l'IA Act, elle devra en outre documenter plus finement les algorithmes et mettre en place un suivi post-commercialisation (rapports de performance, remontée d'incidents).*

# 3. Mise en conformité opérationnelle

## 3.1. Politiques et procédures internes

### *A. Création de politiques de gouvernance de l'IA*

La première étape pour une mise en conformité réussie consiste à formaliser des politiques internes encadrant l'utilisation de l'IA. Il s'agit de :

- **Définir la vision et le périmètre** : préciser les finalités acceptables, les limites d'usage (profilage, biométrie, etc.), et les secteurs d'application (marketing, ressources humaines, maintenance prédictive, etc.).
- **Identifier les rôles et responsabilités** : qui est responsable de la validation des projets IA (référent IA, DPO), qui gère les risques (RSSI), qui rapporte à la direction ?
- **Établir des principes éthiques** : prise en compte de la non-discrimination, de la transparence algorithmique, de la supervision humaine (exigences IA Act).
- Articuler ces politiques avec les cadres légaux (RGPD, NIS2) et anticiper les obligations IA Act (documentation, gestion des risques, qualité des données).

### *B. Mise à jour des processus internes*

Pour appliquer concrètement ces politiques, les entreprises doivent adapter leurs processus existants ou en créer de nouveaux :

- **Onboarding** : tout nouveau projet IA passe par une validation RGPD (DPIA si nécessaire), un check cybersécurité (RSSI), et un examen des biais potentiels (référent IA).
- **Audit et revue de code** : mise en place de points de contrôle réguliers pour vérifier la conformité (sécurité, performance, absence de biais discriminants).
- **Gestion des risques** : déployer des outils de suivi (cartographie des risques IA) et des plans de remédiation en cas d'alerte.
- **Documentation et traçabilité** : archiver les décisions et maintenir un historique de l'évolution des algorithmes (données d'entraînement, logs, correctifs).

### C. Gestion des incidents et notification

Les politiques internes doivent également inclure une procédure de gestion des incidents :

- **Détection et alerte** : mise en place de systèmes de surveillance (monitoring), définition d'indicateurs (temps moyen de détection).
- **Coordination interne** : déclenchement d'une cellule de crise impliquant le DPO, le RSSI, le référent IA et la direction.
- **Notification** : si l'incident concerne des données personnelles (fuite, compromission), la CNIL doit être informée dans les 72 heures (RGPD). En cas d'incident majeur impactant la disponibilité ou l'intégrité des systèmes, la notification peut également relever de la Directive NIS2.

L'établissement de politiques internes et la mise en œuvre de procédures claires constituent le socle d'une gouvernance IA conforme au RGPD, aux exigences de cybersécurité (NIS2) et à l'IA Act. grâce notamment à :

- Une gouvernance explicite (définition des rôles, principes éthiques, périmètre d'usage).
- Des processus adaptés (onboarding des projets IA, audit, documentation).
- Une gestion réactive des incidents (procédure de notification, plan de remédiation).

Les organisations peuvent maîtriser les risques, préserver la confiance de leurs utilisateurs et partenaires, et faire de l'IA un atout dans leur transformation digitale, plutôt qu'une source d'incertitudes juridiques.

### D. Bonnes pratiques

- ✓ **Indicateurs de performance** : la direction peut suivre, par exemple, le taux de DPIA réalisées vs. le nombre de projets IA, ou encore le pourcentage de collaborateurs formés aux enjeux RGPD et IA.
- ✓ **Collaboration avec le DPO et le référent IA** : un comité IA mensuel ou trimestriel peut valider les nouveaux projets, évaluer les risques émergents et actualiser les politiques si nécessaire.
- ✓ **Évaluation de la supply chain** : exiger des sous-traitants ou partenaires qu'ils respectent des standards similaires (obligations RGPD/IA Act, clauses contractuelles de sécurité, audits partagés).

***Exemple***

*Une entreprise du secteur industriel décide d'industrialiser le déploiement de ses algorithmes IA. Elle crée une "checklist IA" dans laquelle figurent : (1) vérification RGPD (DPIA si besoin), (2) validation cybersécurité (tests d'intrusion, chiffrement), (3) revue de la qualité des données d'entraînement (ex. : pas de biais, données licites), (4) documentation technique. Le projet ne peut passer en production sans le feu vert du DPO, du RSSI et du référent IA.*

## 3.2. Formation et sensibilisation juridique

### *A. Évaluation des besoins et ciblage des publics*

La première étape consiste à identifier les différents profils au sein de l'organisation :

- **Équipes IT** : développeurs, data scientists, ingénieurs système (besoins en cybersécurité, RGPD technique, IA Act).
- **Équipes juridiques** : avocats internes, juristes, DPO (besoins sur l'évolution législative, rédaction de contrats IA).
- **Ressources humaines** : gestion du recrutement ou de l'évaluation des salariés via l'IA (article 22 RGPD, biais discriminants).
- **Managers et direction générale** : pilotage stratégique, responsabilité en cas de sanctions, validation des budgets formation.

Un audit ou un questionnaire peut être utilisé pour évaluer les connaissances initiales et définir des objectifs de progression.

### *B. Conception des programmes de formation*

Sur la base de l'évaluation, on définit plusieurs modules :

- **Module RGPD/NIS2** : principes fondamentaux (licéité, minimisation, droits des personnes, obligations de notification d'incident), risques encourus (sanctions CNIL, responsabilité pénale, impact réputationnel).
- **Session IA et éthique** : présentation de l'IA Act, concept de supervision humaine, gestion des biais algorithmiques (ex. : discrimination, opacité), documentation technique exigée.
- **Focus sur la sécurité** : pour l'IT, aspects concrets (chiffrement, gestion des accès, plan de remédiation, tests de vulnérabilité).

Les supports peuvent varier : e-learning, fiches pratiques, conférences interactives. Idéalement, une validation (quiz, attestation) vient clôturer chaque module.

### C. Organisation et déploiement

Pour que la formation soit efficace :

- **Calendrier** : prévoir un cycle de formations récurrent (ex. deux sessions annuelles) et des ateliers ponctuels (ateliers “biais IA”).
- **Formateurs** : le DPO, le référent IA, le RSSI peuvent assurer des sessions internes, ou faire appel à un organisme spécialisé.
- **Suivi et évaluation** : collecter les retours des participants via un questionnaire, mesurer le taux de participation, et évaluer les progrès (tests avant/après formation).

---

***Exemple***

*Une société de e-commerce planifie chaque trimestre une « Semaine RGPD & IA » avec des ateliers ludiques (escape game cybersécurité), des conférences courtes (30 min) sur la transparence algorithmique et un quiz de fin de semaine pour tous les collaborateurs.*

---

### D. Gouvernance et pérennité

La formation et la sensibilisation doivent s’inscrire dans une culture d’entreprise durable :

- **Mise à jour régulière** : adapter le contenu aux nouvelles règlementations (IA Act définitif), aux évolutions de la jurisprudence ou aux incidents survenus dans l’entreprise.
- **Reporting à la direction** : présenter les indicateurs clés (taux de participation, évaluation de la compréhension) pour obtenir un soutien continu.
- **Intégration dans les politiques internes** : la charte “IA & Données” peut exiger un niveau de formation minimal pour chaque collaborateur, voire pour les sous-traitants critiques.

***Exemple***

*Après une attaque de phishing ayant réussi, le RSSI et le DPO incitent la direction à renforcer la formation cybersécurité. Les sessions sont rendues obligatoires pour tout nouvel arrivant, et un rappel est fait dans la newsletter interne sur les bonnes pratiques (vérification d'e-mails douteux, signalement au helpdesk).*

La formation et la sensibilisation représentent un levier essentiel pour réussir la mise en conformité en matière de gouvernance et d'utilisation de l'IA. En investissant dans des programmes adaptés aux différents profils (IT, juridique, RH, management), l'entreprise :

- Renforce sa culture RGPD et cybersécurité (NIS/NIS2).
- Anticipe les exigences de l' IA Act (supervision humaine, gestion des biais).
- Diminue les risques d'incidents, de sanctions ou de litiges liés aux algorithmes IA.

Il s'agit donc d'un investissement stratégique, garantissant à long terme la fiabilité, la sécurité et l'éthique de l'écosystème IA déployé par l'organisation.

## 3.3. Audit de conformité et certification

### *A. Objectifs et portée de l'audit*

L'audit est un outil essentiel pour mesurer le niveau de conformité d'une organisation. Il peut être :

- **Interne** : mené par une équipe dédiée (DPO, référent IA, RSSI) afin d'évaluer la cohérence entre les politiques internes (RGPD, NIS2, charte IA) et la pratique.
- **Externe** : réalisé par un organisme tiers (cabinet d'audit, autorité de contrôle, organisme notifié dans le cadre de l'IA Act) qui apporte un regard indépendant.

Le périmètre d'audit envisageable pourrait être dans un futur proche, le suivant :

- La conformité RGPD (registres de traitement, DPIA, sécurité des données).
- Les exigences NIS2 (analyse des risques, notification d'incident, plan de continuité).
- Les obligations IA Act (documentation, formation, supervision humaine, gestion des biais).

### *B. Modalités d'audit*

Un programme d'audit définit la fréquence (annuelle, semestrielle), le référentiel (ex. ISO 27001, politique interne), et la méthodologie (interviews, revue de documents, tests de sécurité). Les étapes clés :

- **Préparation** : collecte d'informations sur l'organisation, ses politiques, ses procédures IA.
- **Exécution** : analyses, visites sur site, entretiens avec les responsables (DPO, RSSI, référent IA).
- **Rapport d'audit** : synthèse des non-conformités, recommandations de correction.
- **Plan d'action** : priorisation des mesures correctives, suivi de leur mise en œuvre.

***Exemple***

*Lors d'un audit interne IA, l'équipe constate que certaines données d'entraînement ne sont pas documentées (traçabilité insuffisante). Une action corrective est décidée : mise à jour du registre IA et contrôle renforcé de la provenance des données.*

### C. Processus de certification

Les certifications permettent de valoriser l'engagement de l'organisation en matière de sécurité et de protection des données :

- Normes ISO :
  - ISO 27001 : management de la sécurité de l'information (SMSI).
  - ISO 27701 : extension "Privacy" pour le RGPD.
  - ISO 31000 : management des risques.
  - ISO 42001 : management de l'IA

- Labels et référentiels nationaux :
  - SecNumCloud (ANSSI) : pour les fournisseurs de services Cloud sécurisés.
  - HDS (Hébergement de Données de Santé), etc.

Le processus type :

- Diagnostic initial : évaluer l'écart entre l'existant et le référentiel (ISO 27001 par ex.).
- Mise en conformité : déployer les mesures organisationnelles et techniques recommandées.
- Audit de certification : un organisme agréé vérifie la conformité effective.
- Suivi : audits de surveillance, renouvellement périodique.

### D. Spécificités liées à l'IA

Dans le cadre de l'IA Act, les systèmes classés « à haut risque » doivent faire l'objet d'une évaluation de conformité :

- Documentation technique : preuves de la qualité des données, logs de performance, rapport de supervision humaine.
- Gestion des biais : démonstration que l'algorithme a été testé et qu'aucun biais discriminatoire majeur n'apparaît.
- Suivi post-commercialisation : obligation de surveiller l'algorithme en production, d'archiver les incidents et de mettre en place des correctifs si besoin.

---

***Exemple***

*Une entreprise medtech soumet son logiciel d'analyse d'images médicales à un organisme notifié qui évalue la conception, la validité clinique, la sécurité et la supervision humaine (médecin validant la décision). Cette démarche rappelle le principe du marquage CE pour les dispositifs médicaux, et sera exigée par l'IA Act.*

---

L'audit et la certification constituent deux volets cruciaux d'une mise en conformité réussie :

- L'audit (interne ou externe) permet d'évaluer et d'améliorer en continu les pratiques (RGPD, NIS2, IA Act).
- La certification (ISO, SecNumCloud, etc.) ou l'évaluation de conformité (IA Act) offre une garantie de fiabilité auprès des clients, partenaires et autorités de contrôle.

Cette démarche se veut préventive (détecter les vulnérabilités, réduire les risques de sanctions) et valorisante (renforcer l'image de sérieux et la confiance dans les solutions IA). En s'appuyant sur des référentiels reconnus, les entreprises professionnalisent leur gouvernance et sécurisent l'usage de l'IA, tout en anticipant les futurs contrôles et évolutions législatives.

# SECRET PROFESSIONNEL ET DEVOIR DE DISCRETION

Nous rappelons ici les notions de secret professionnel et de devoir de discrétion en reproduisant des extraits de l'ouvrage *Expertise comptable et secret professionnel* édité par l'Ordre des experts-comptables (Collectif, septembre 2021).

## 1. Secret professionnel

Les professionnels de l'expertise comptable soumis au secret professionnel.

L'alinéa 1 de l'article 21 de l'ordonnance n°45-2138 du 19 septembre 1945 précise que « *Sous réserve de toute disposition législative contraire, les experts-comptables, les salariés mentionnés à l'article 83 ter et à l'article 83 quater, les experts-comptables stagiaires et les professionnels ayant été autorisés à exercer partiellement l'activité d'expertise comptable sont tenus au secret professionnel dans les conditions et sous les peines fixées par l'article 226-13 du Code pénal* ».

L'article 226-13 du Code pénal précise que « *la révélation d'une information à caractère secret par une personne qui en est dépositaire soit par état ou par profession, soit en raison d'une fonction ou d'une mission temporaire, est punie d'un an d'emprisonnement et de 15 000 euros d'amende* ».

L'article 226-14 du code précité prévoit quant à lui que « *L'article 226-13 n'est pas applicable dans les cas où la loi impose ou autorise la révélation du secret* ».

Il ressort de ces textes que l'expert-comptable commet le délit d'atteinte au secret professionnel, réprimé à l'article 226-13 du Code pénal, lorsqu'il révèle une information dont il est dépositaire, si aucun texte légal ne prévoit la levée de ce secret.

Cette infraction est constituée lorsque les éléments suivants sont réunis :

• **L'élément matériel** : un acte de révélation portant sur des informations recueillies dans l'exercice de la profession (une appréciation subjective du professionnel ne peut être considérée comme une information).

• **L'élément moral** : l'intention du professionnel de révéler le secret dont il a connaissance, quel que soit son mobile (à distinguer d'une négligence ou d'une erreur de droit)1.

### 1.1. Informations couvertes par le secret professionnel

Les textes applicables aux experts-comptables ne donnent pas de définition précise des informations relevant du secret professionnel.

De manière générale, il est possible de déduire de la jurisprudence et de la doctrine qu'une information est couverte par le secret professionnel dès lors qu'elle a été recueillie par un professionnel es qualité.

C'est parce que certaines informations ont été recueillies par une personne tenue au secret professionnel dans l'exercice de sa profession qu'elles sont couvertes par le secret professionnel.

Le secret professionnel de l'expert-comptable étant absolu, il va au-delà de la protection des intérêts du client, ce qui explique que personne ne peut l'en délier.

Les exemptions possibles doivent être prévues par des dispositions législatives spécifiques.

Par conséquent, la levée du secret par le client n'est pas suffisante pour exonérer l'expert-comptable de son obligation. Cependant, dans les cas où la divulgation de certaines informations participe nécessairement à l'exercice de la mission, il est recommandé de mentionner, dans les lettres de mission ou tous documents contractuels signés avec le client, les informations (listées ou mentionnées en fonction de leur nature ou de leur type) dont la transmission par l'expert-comptable à certains acteurs (organismes agréés, administration fiscale, organismes sociaux, Banque de France par exemple) est inhérente à la mission même confiée à l'expert-comptable, ainsi que les circonstances objectives, justifiant cette transmission.

### 1.2. Personnes soumises au secret professionnel

L'article 21 de l'ordonnance du 19 septembre 1945 prévoit que sont soumis au secret professionnel les experts-comptables, les salariés autorisés mentionnés à l'article 83 ter et à l'article 83 quater, les professionnels ayant été autorisés à exercer partiellement l'activité d'expertise comptable, les dirigeants et administrateurs d'AGC ainsi que les experts-comptables stagiaires.

Il s'agit uniquement de personnes physiques.

Concernant les salariés non experts-comptables employés par les experts-comptables, ceux-ci ne sont pas visés à l'article 21 de l'ordonnance du 19 septembre 1945. Aucun texte ne les soumet au secret professionnel, à la différence des collaborateurs des commissaires aux comptes. Par conséquent, c'est le droit du travail qui leur est applicable. La convention collective nationale des cabinets d'experts-comptables et de commissaires aux comptes prévoit en son article

8.5.2 que : « *Les collaborateurs sont tenus, indépendamment d'une obligation de réserve générale, à une discrétion absolue sur tous les faits qu'ils peuvent apprendre en raison de leurs fonctions ou de leurs missions ainsi que de leur appartenance au cabinet. Cette obligation de réserve concerne exclusivement la gestion et le fonctionnement du cabinet et des entreprises clientes, leur situation financière et les projets les concernant. Ces dispositions ne font pas obstacle à l'application de l'article L 2323-18 du Code du travail. Les documents ou rapports qu'ils établiront ou dont communication leur sera donnée sont la propriété du cabinet ou du client du cabinet. Ils ne pourront ni en conserver de copies ou de photocopies, ni en donner communication à des tiers sans l'accord écrit du membre de l'ordre ou de la compagnie. Toute inobservation à cette stricte obligation constitue une faute lourde, et justifie non seulement un congédiement immédiat, mais en outre, la réparation du préjudice causé* ».

## 2. Devoir de discrétion

L'article 147 du code de déontologie dispose que « *Sans préjudice de l'obligation au secret professionnel, les personnes mentionnées à l'article 141 sont soumises à un devoir de discrétion dans l'utilisation de toutes les informations dont elles ont connaissance dans le cadre de leur activité* ».

Le client qui contracte avec un expert-comptable doit avoir l'assurance que les informations fournies à celui-ci ne seront ni divulguées sans son accord, ni utilisées à des fins étrangères à la mission, notamment dans l'intérêt du professionnel ou d'un tiers.

La discrétion vise toutes les informations, recueillies au cours de la mission, que l'expert-comptable ne doit pas divulguer. Le périmètre de cette obligation est donc identique à celui de l'obligation au secret professionnel.

# DATA ET CABINET D'EXPERTISE COMPTABLE : DE NOMBREUSES DONNEES PERSONNELLES

***Par Vincent Lacomme***

La démultiplication exponentielle du volume de données mesurée par plusieurs études économiques (dont Statista) s'applique également aux cabinets d'expertise comptable et de commissariat aux comptes.

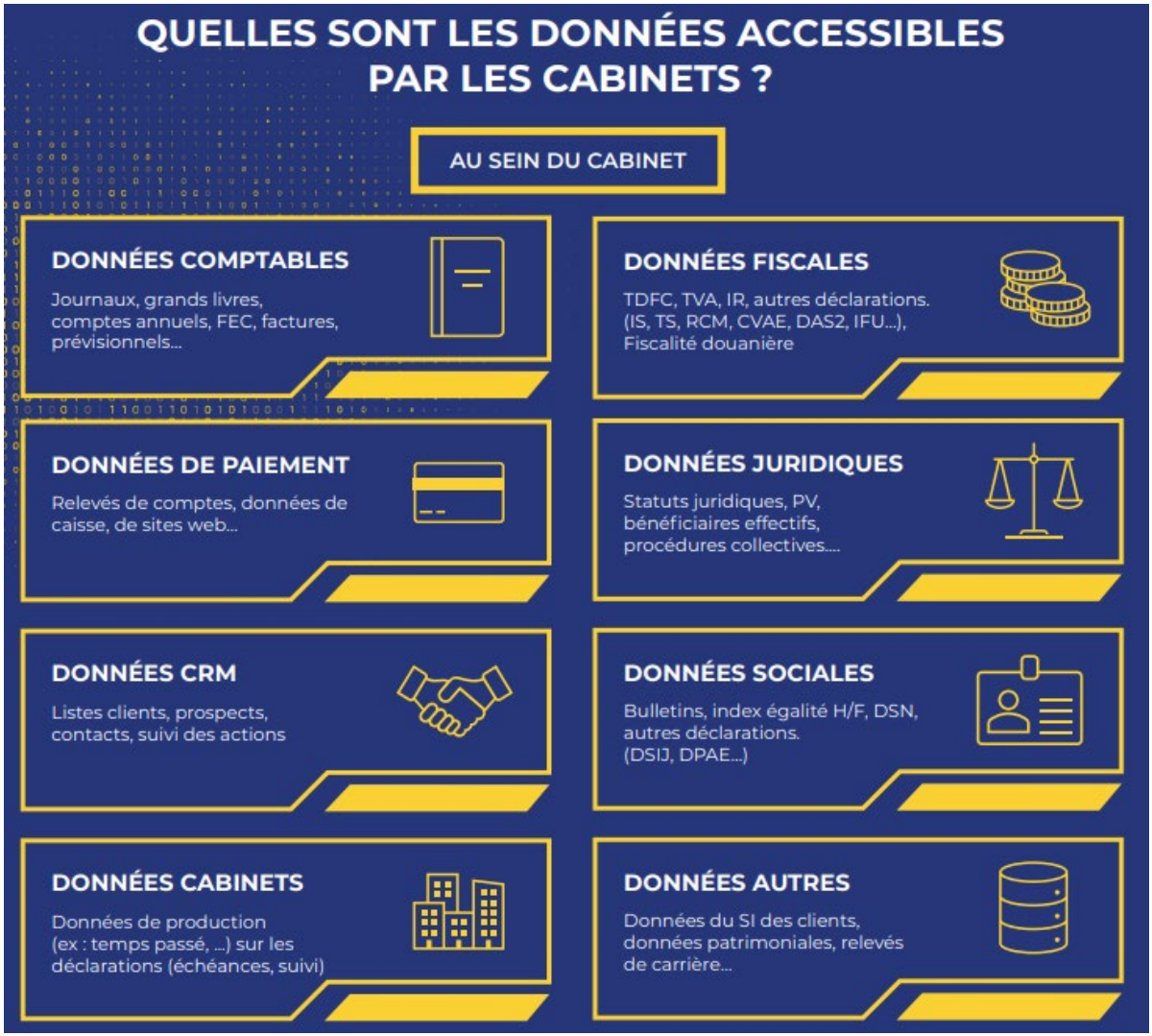
Preuve de l'importance du sujet, le 78e congrès des experts-comptables en 2023 a retenu le titre « *De la facture électronique à la data, le début d'une nouvelle ère* ».

Un baromètre data & IA publié en octobre 2024 dans SIC magazine[4] fait état que 91% des cabinets d'expertise comptable voient dans l'évolution numérique une opportunité et que 40% d'entre eux ont mis en place un processus de gouvernance des données.

La **data** irrigue les cabinets dans leurs missions. Tous les processus d'une entreprise intègrent des données :

- **comptables** : écritures comptables, factures justificatives pouvant contenir des données personnelles…
- **social** : bulletins de paie, cartes d'identité…
- **fiscal** : déclarations d'impôt sur le revenu, donations…
- **commercial** : factures clients, liste de clients…
- **stock** : liste d'articles, coûts de revient…
- …

[4] N°441, https://www.experts-comptables.fr/sic/441




**Extrait du data référentiel du CNOEC, publié en juillet 2023[5]**

[5] Une version complète est disponible sur l'espace privé du Conseil National de l'Ordre des experts-comptables : https://extranet.experts-comptables.org/xtc_recherche?fulltext=data%20r%C3%A9f%C3%A9rentiel?modal=0

Les données personnelles sont d'autant plus présentes dans le domaine des ressources humaines et de la paie mais aussi pour des missions patrimoniales (impôt sur le revenu, IFI).

Parmi les données sensibles conservées par les cabinets, on peut citer plusieurs données à caractère personnel :

- Les données permanentes : dossiers des salariés, dirigeants, bénéficiaires effectifs…
- Au sein de la comptabilité : des noms de clients personnes physiques, de salariés, dirigeants.

L'utilisation de l'IA générative implique dès lors plusieurs risques quant à des fuites des données, à des cyberattaques et *ransomwares*....

Ces risques existaient avant l'IA et sont lié à la dématérialisation des données :

- Stockage ou transmission de fichiers informatiques à des services recourant à des serveurs hors de l'Union européenne ;
- Transmission de bulletins de paie par des moyens non sécurisés (copies de mails) ;
- …

**Lectures recommandées :**

- Espace Parlons Data, CNOEC, https://extranet.experts-comptables.org/dossier/parlons-data-
- Guide de la cybersécurité, CNOEC, septembre 2023, https://bibliotique.com/Record.htm?Record=19334845124911520279
- Kit mission cybersécurité : https://extranet.experts-comptables.org/kit-mission/accompagner-ses-clients-sur-la-cybersecurite-

# L'INTEGRATION DE L'INTELLIGENCE ARTIFICIELLE GENERATIVE DANS LES CABINETS D'EXPERTISE COMPTABLE : VERS UNE UTILISATION RESPONSABLE ET ENCADREE

*Par Sabrina Agrapart*

## 1. IAGen et expertise comptable : entre opportunités et défis

L'utilisation de l'IAGen connaît un essor considérable, et cela est particulièrement vrai dans les métiers du chiffre, notamment dans les cabinets d'expertise comptable. De nombreux professionnels ont déjà intégré ces outils dans leur quotidien en raison des multiples avantages qu'ils offrent, tels que la simplification de certaines tâches répétitives, l'amélioration de la productivité, ou encore la possibilité de se concentrer sur des missions à plus forte valeur ajoutée. Toutefois, certains cabinets ont choisi d'interdire leur utilisation, craignant des risques en termes de sécurité et de confidentialité des données en raison d'une incapacité immédiate de contrôler leurs utilisations.

L'enjeu de prise de conscience et de mesures est primordial puisque certaines études considèrent que près de 25 %, toute tâche et métier confondus pourront faire l'objet d'une automatisation notamment sur les missions administratives[6].

## 2. Interdiction totale de l'IAGen : une solution illusoire

Nous pensons qu'une interdiction totale de ces outils n'est pas une solution viable et surtout une solution à court terme.

En effet, même si l'accès à des outils comme ChatGPT peut être bloqué sur les postes de travail professionnels, il serait difficile de garantir le respect strict d'une telle interdiction. Certains collaborateurs pourraient contourner cette restriction en utilisant leurs appareils personnels, en violation de la charte informatique (si elle existe), et ainsi mettre en danger des informations sensibles de l'entreprise et de ses

[6] Etude de Goldman Sachs du 26 mars 2023 (https://www.key4biz.it/wp-content/uploads/2023/03/Global-Economics-Analyst_-The-Potentially-Large-Effects-of-Artificial-Intelligence-on-Economic-Growth-Briggs_Kodnani.pdf)

clients, en y introduisant des données à caractère personnel ou des informations couvertes par le secret des affaires.

Face à ces risques, il semble essentiel de ne pas interdire l'utilisation de l'IAGen, mais plutôt de (i) former les collaborateurs et (ii) l'encadrer de manière rigoureuse pour éviter les dérives. Les sociétés d'expertise comptable doivent encourager une utilisation responsable de l'IAGen, en veillant à ce que l'éthique et la conformité réglementaire soient au cœur de l'élaboration et de l'implémentation de l'IAGen dans leurs services et processus internes.

Par ailleurs, nos clients vont nous challenger afin de savoir les économies d'échelles que nous réalisons avec l'IA afin que nous en tenions compte aussi dans le cadre de notre facturation.

# L'IA ACT, UN PREMIER PAS VERS LA NORMALISATION

*Par Sabine Marcellin*

## 1. Qu'est-ce que l'AI Act ?

Il s'agit du Règlement européen 2024/1689 du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle, qui a été publié le 12 juillet. Il est désigné indifféremment par la Loi sur l'IA, le RIA ou l'AI Act.

Il s'agit du premier texte au monde régissant l'IA, fruit des travaux de l'Union européenne. Il a pour ambition à la fois de promouvoir l'innovation et de créer les conditions d'une IA de confiance, en garantissant la protection de droits fondamentaux.

## 2. Quel est le contexte ?

L'AI Act s'inscrit dans un large éventail de textes européens applicables à l'IA, directement ou indirectement. Parmi ces textes, figurent la directive relative à la responsabilité du fait des produits défectueux[7] et de la proposition de directive[8] relative à la responsabilité civile dans le domaine de l'IA. Il faut citer également les textes relatifs aux données (RGPD, DSA[9], DGA[10], DMA[11], DA[12]), aux microprocesseurs[13] et à la cybersécurité (NIS2[14]).

L'AI Act est en partie applicable depuis le 2 février 2025 et son calendrier progressif figure ci-après.

Dans chaque État-membre, des lois nationales viendront affiner la transposition de l'AI Act, notamment en désignant l'autorité de contrôle.

[7] Directive (UE) 2024/2853 du 23 octobre 2024 relative à la responsabilité du fait des produits défectueux et abrogeant la directive 85/374/CEE du Conseil.
[8] Proposition de directive relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle.
[9] Règlement 2022/2065 du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE.
[10] Règlement (UE) 2022/868 sur la gouvernance européenne des données.
[11] Règlement 2022/1925 du 14 septembre 2022 relatif aux marchés contestables et équitables dans le secteur numérique et modifiant les directives (UE) 2019/1937 et (UE) 2020/1828.
[12] Règlement 2023/2854 du 13 décembre 2023 concernant des règles harmonisées portant sur l'équité de l'accès aux données et de l'utilisation des données et modifiant le règlement (UE) 2017/2394 et la directive (UE) 2020/1828.
[13] Règlement 2023/1781 du 13 septembre 2023 établissant un cadre de mesures pour renforcer l'écosystème européen des semi-conducteurs et modifiant le règlement (UE) 2021/694.
[14] Directive (UE) 2022/2555 du 14 décembre 2022 concernant des mesures destinées à assurer un niveau élevé commun de cybersécurité dans l'ensemble de l'Union, modifiant le règlement (UE) no 910/2014 et la directive (UE) 2018/1972, et abrogeant la directive (UE) 2016/1148.

## 3. Quelle est la portée de l'AI Act ?

L'AI Act vise à garantir la sécurité des biens et des personnes, ainsi que la protection des droits fondamentaux comme la non-discrimination, la transparence, la responsabilité ou encore le respect des valeurs démocratiques européennes.

Des mesures sont mises en place, avec l'AI Act, pour s'assurer du respect d'un ensemble de critères de gestion des risques, gouvernance des données, documentation technique, traçabilité, supervision humaine, évaluation de la robustesse, exactitude, cybersécurité et mise en place d'un système de management de la qualité.

Cette réglementation est applicable non seulement aux organisations qui conçoivent (fournisseurs) ou utilisent (déployeurs) des systèmes d'IA (ou SIA) au sein de l'Union Européenne, mais aussi à tout opérateur traitant sur le marché européen.

Le Règlement propose une approche par les risques. Les SIA sont catégorisés selon les risques en fonction de l'usage prévu et les obligations y sont graduellement adaptées :

- Risque inacceptable : un ensemble limité de pratiques sont interdites, dès lors qu'elles sont contraires aux valeurs de l'Union européenne et aux droits fondamentaux. *Exemples : notation sociale ou exploitation de la vulnérabilité des personnes.*

- Haut risque : les SIA qui peuvent porter atteinte à la sécurité des personnes ou à leurs droits fondamentaux ce qui justifie que leur développement soit soumis à des exigences renforcées. *Exemples : systèmes biométriques, des systèmes utilisés dans le recrutement.*

- Risque limité : Ces SIA sont soumis à des obligations de transparence spécifiques, notamment en cas de risque manifeste de manipulation.

*Exemples : recours à des chatbots ou à la génération de contenu artificiel.*

- Risque minimal : pour tous les autres systèmes d'IA, l'AI Act ne prévoit pas d'obligation spécifique.

Outre cette classification, une catégorie supplémentaire est identifiée :

- Les modèles d'IA à usage général (GPAI, *Large Language Model* ou *LLM*) : ces technologies, telles que ChatGPT, transforment rapidement la manière dont les SIA sont construits et déployés. Elles sont soumises à des obligations légales et à des exigences adaptées à leurs caractéristiques, ainsi qu'à un système de surveillance du risque systémique.

## 4. Quels sont les impacts du Règlement sur les entreprises fournisseurs ou utilisatrices d'IA ?

Le Règlement s'applique à tous les opérateurs (fournisseurs, déployeurs, distributeurs de SIA) au sein de l'espace européen, en fonction de l'usage des SIA. Les obligations sont spécifiques à chaque niveau de risque de SIA.

### 4.1. Obligations applicables au SIA à risque inacceptable

La fourniture ou le déploiement des SIA présentant un risque inacceptable en raison de leurs risques potentiels pour les valeurs européennes et les droits fondamentaux, sont interdits depuis le 2 février 2025.

La Commission a publié le 4 février 2025, un projet de Lignes Directrices[15] qui fournissent des explications juridiques et des exemples pratiques. Ces Lignes Directrices ne sont pas contraignantes.

[15] https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act

### 4.2. Obligations associées au SIA à haut risque

Les fournisseurs d'IA à haut risque doivent respecter un ensemble d'obligations, notamment :

- Mettre en place un système de gestion des risques tout au long du cycle de vie du SIA à haut risque,
- Assurer la gouvernance des données, en veillant à ce que les ensembles de données de formation, de validation et de test soient pertinents, suffisamment représentatifs et, dans la mesure du possible, exempts d'erreurs et complets,
- Établir une documentation technique pour démontrer la conformité,
- Concevoir leur SIA pour qu'il enregistre automatiquement les événements pertinents pour l'identification des risques au niveau national et les modifications substantielles tout au long du cycle de vie du système,
- Fournir des instructions d'utilisation aux utilisateurs en aval pour leur permettre de se conformer à la réglementation,
- Concevoir leur SIA pour permettre aux déployeurs de mettre en place une surveillance humaine,
- Concevoir leur SIA pour atteindre les niveaux appropriés de précision, de robustesse et de cybersécurité,
- Mettre en place un système de gestion de la qualité.

### 4.3. Obligations applicables au SIA à risque limité

L'utilisation de ces systèmes est soumise à des obligations de transparence spécifiques. Les personnes doivent être informées qu'elles interagissent avec un système d'IA.

Les fournisseurs de systèmes d'IA, y compris les SIA à usage général, qui génèrent des contenus synthétiques (texte, image, audio ou vidéo) veillent à ce que les résultats du système d'IA soient marqués dans un format lisible par machine et détectables comme étant générés ou manipulés artificiellement.

### 4.4. Obligations associées au SIA à risque minimal ou nul

La législation sur l'IA autorise la libre utilisation de l'IA à risque minimal. Cela inclut des applications telles que les jeux vidéo compatibles avec l'IA ou les filtres anti-spam. Il s'agit de la très grande majorité des SIA actuellement utilisés dans l'UE.

Aucune obligation n'incombe aux fournisseurs et déployeurs de SIA dans ce cas.

### 4.5. Obligations associées au modèle d'IA à usage général

Pour cette catégorie de modèles d'IA à usage général (GPAI), l'AI Act prévoit plusieurs niveaux d'obligation, allant de mesures de transparence et de documentation minimales à une évaluation approfondie, pour les modèles GPAI présentant des risques systémiques.

Les modèles GPAI présentent des risques systémiques lorsque la quantité cumulée de calcul utilisée pour la formation est supérieure à 1025 opérations en virgule flottante (FLOP ou *floating-point operations per second*). Dans ce cas, les fournisseurs doivent notifier à la Commission si leur modèle répond à ce critère dans un délai de deux semaines. La Commission peut décider d'elle-même, ou par le biais d'une alerte qualifiée du groupe scientifique d'experts indépendants, qu'un modèle a des capacités d'impact élevées, ce qui le rend systémique.

Dans le cas d'une évaluation approfondie de SIA, notamment en raison de leur puissance, il est nécessaire de mettre en place de mesures d'atténuation des risques systémiques : risques d'accidents majeurs, d'utilisation à mauvais escient pour lancer des cyberattaques, propagation de biais aux effets discriminatoires à l'encontre de certaines personnes, etc.

Tous les fournisseurs de modèles GPAI doivent :

- Rédiger la documentation technique, y compris le processus de formation et d'essai et les résultats de l'évaluation,
- Élaborer des informations et de la documentation à fournir aux fournisseurs en aval qui ont l'intention d'intégrer le modèle GPAI dans leur propre système d'IA, afin que ces derniers en comprennent les capacités et les limites et soient en mesure de s'y conformer,
- Établir une politique de respect de la directive sur le droit d'auteur,
- Publier un résumé suffisamment détaillé du contenu utilisé pour la formation du modèle GPAI.

Outre les quatre obligations susmentionnées, les fournisseurs de modèles GPAI présentant un risque systémique doivent également :

- Effectuer des évaluations de modèles, y compris mener et documenter des tests contradictoires afin d'identifier et d'atténuer le risque systémique,
- Évaluer et atténuer les risques systémiques éventuels, y compris leurs sources,
- Repérer, documenter et signaler les incidents graves et les éventuelles mesures correctives à l'Office AI et aux autorités nationales compétentes dans les meilleurs délais,
- Assurer un niveau adéquat de protection de la cybersécurité.

Les fournisseurs de modèles GPAI sous licence libre et gratuite, dont les paramètres, y compris les poids, l'architecture du modèle et l'utilisation du modèle sont accessibles au public, ce qui permet l'accès, l'utilisation, la modification et la distribution du modèle doivent uniquement :

- Respecter les droits d'auteur et,
- Publier le résumé des données de formation, à moins qu'ils ne présentent un risque systémique.

Tous les fournisseurs de modèles GPAI peuvent prouver qu'ils respectent leurs obligations en adhérant volontairement à un code de bonnes pratiques jusqu'à la publication de normes européennes harmonisées[16], dont le respect entraînera une présomption de conformité. Les fournisseurs qui n'adhèrent pas à des codes de pratique doivent démontrer qu'ils disposent d'autres moyens adéquats pour se conformer à leurs obligations, afin d'obtenir l'approbation de la Commission.

## 5. Quelles sont les sanctions prévues ?

En cas de non-respect des obligations en matière d'IA par les opérateurs, l'AI Act prévoit des sanctions allant jusqu'à 35 millions d'euros ou 7 % du chiffre d'affaires annuel de l'entreprise.

Les autorités nationales chargées de l'application seront désignées avant le 2 août 2025. Le Bureau européen de l'IA (ou Office AI) créé en mai 2024 est l'organisme en charge de la coordination des régulations.

## 6. Quelles sont les dispositions de l'AI Act relatives à la confidentialité ?

Le Règlement ne traite formellement que des obligations de confidentialité applicables à la Commission européenne, aux autorités de surveillance du marché, ainsi que toute autre personne physique ou morale, dès lors qu'elles participent à l'application de l'AI Act.

Concernant les données traitées par les fournisseurs et déployeurs de SIA, ce sont les règles classiques de protection de la confidentialité qui s'appliquent, comme le secret professionnel ou le RGPD.

[16] Un premier projet de Code de bonnes pratiques a été publié le 14 novembre 2024 : https://digital-strategy.ec.europa.eu/fr/news/commission-publishes-first-draft-general-purpose-artificial-intelligence-code-practice

Les experts-comptables et commissaires aux comptes sont soumis au secret professionnel, relatif aux informations confidentielles de leurs clients, selon l'article 226-13 du Code pénal.

L'AI Act et le RGPD sont conçus pour être complémentaires. Le RGPD établit des règles générales pour la protection des données personnelles. L'AI Act se concentre spécifiquement sur les risques et les défis posés par les systèmes d'IA, vis-à-vis d'autres droits fondamentaux.

Cependant, l'intelligence artificielle (IA) pose des défis majeurs en matière de confidentialité des données. Les risques d'atteinte à la confidentialité existent lors de l'usage d'un SIA, notamment via la rédaction d'un prompt ou le chargement de documents susceptibles d'être traités par l'IA. L'usage d'un SIA, par exemple pour préparer des annexes comptables, devra être réalisé sans porter atteinte au secret professionnel.

Afin de profiter de la puissance de l'IA, tout en respectant la confidentialité des données, les solutions existent :

- La maîtrise et le contrôle des documents traités par l'IA,
- La vérification des conditions générales des SIA utilisés,
- Différentes solutions techniques (anonymisation, partition des données, etc.)
- La formation des collaborateurs à l'usage de l'IA, qui est rendue obligatoire par l'AI Act depuis le 2 février 2025,
- L'élaboration de charte interne de l'IA, etc.

## 7. Quel est le calendrier d'application de l'*AI Act* ?

Le Règlement sur l'IA est entré en vigueur depuis le 1er août 2024. Certaines dispositions sont applicables depuis le 2 février 2025 et l'application des autres règles s'étalera dans le temps.

Le tableau ci-après présente quelques échéances majeures. Ce calendrier est indicatif, car il est susceptible d'évoluer en fonction des nouvelles dates clés annoncées par les organes officiels de l'Union européenne.

| **Calendrier d'application de l'AI Act** | |
|---|---|
| **Date** | **Dispositions** |
| **2 février 2025** | - Interdictions relatives aux systèmes d'IA présentant des risques inacceptables<br>- Exigences en matière de formation à l'IA |
| **2 août 2025** | - Désignation des Organismes notifiés (évaluant les SIA à haut risque),<br>- Règles applicables aux modèles GPAI,<br>- Gouvernance,<br>- Confidentialité,<br>- Sanctions |
| **2 août 2025** | Délai pour que les États membres désignent les autorités nationales compétentes |
| **2 août 2025** | Délai pour que les États membres fixent les règles relatives aux sanctions et aux amendes, les notifient à la Commission et veillent à ce qu'elles soient correctement mises en œuvre |
| **2 août 2026** | Les autres dispositions de la loi sur l'IA sont applicables |
| **2 août 2026** | Le règlement s'applique aux exploitants de systèmes d'IA à haut risque, mis sur le marché ou en service avant cette date. Toutefois, cela ne s'applique qu'aux systèmes dont la conception est modifiée de manière significative à partir de cette date |
| **2 août 2027** | Les fournisseurs de modèles GPAI mis sur le marché, avant le 2 août 2025, doivent avoir pris les mesures nécessaires pour se conformer aux obligations prévues par le règlement avant cette date |
| **2 août 2029 *(et ensuite tous les quatre ans*)** | La Commission présente au Parlement européen et au Conseil un rapport sur l'évaluation et le réexamen du règlement. |

## 8. Comment se mettre en conformité avec l'AI Act ?

Les personnes morales fournissant, distribuant ou déployant des systèmes ou des modèles d'intelligence artificielle sont concernées par l'AI Act et doivent se mettre en conformité à l'AI Act, selon le calendrier ci-dessus.

Il sera nécessaire de coordonner les travaux de mise en conformité à l'AI Act avec ceux relatifs aux autres réglementations applicables. C'est le cas notamment des obligations liées à la cybersécurité et de la protection des données personnelles. Même si ces obligations semblent se cumuler, un certain nombre d'exigences similaires se retrouvent dans les différents textes, comme l'obligation de sécurité et de transparence.

La conduite du projet de mise en conformité nécessite de :

- Désigner un coordinateur ou une équipe en charge,
- Déterminer le budget adéquat, sachant que la diversité des organisations, leur taille et modèle de fonctionnement rendent difficile de bénéficier d'une estimation générale,
- Mettre en œuvre une cartographie des risques des SIA développés ou déployés,
- Adapter les mesures aux niveaux de risque identifiés.

Les mesures principales à mettre en œuvre s'appliquent selon le calendrier et le niveau de risque et le niveau de risque.

Depuis le 2 février 2025, tous les fournisseurs et déployeurs de SIA doivent s'assurer que leurs collaborateurs disposent d'un niveau suffisant de connaissances en matière d'IA, en fonction de :

- Leur expérience et formation préalable,
- Le contexte dans lequel l'IA est utilisée,
- L'impact de ces systèmes sur les personnes concernées.

Depuis le 2 février 2025 également, l'interdiction de fourniture ou déploiement de SIA présentant des risques inacceptables s'applique.

A compter du 2 août 2025, sont applicables les mesures relatives aux modèles GPAI. Les organisations doivent cartographier leurs systèmes d'IA selon la classification de risque (interdit, élevé, limité, minimal) avant août 2025. Pour les IA à haut risque (recrutement, crédit, santé), des évaluations de conformité, une documentation technique détaillée et des mécanismes de contrôle humain sont exigés. Pour les modèles GPAI mis sur le marché avant le 2 août 2025, les fournisseurs disposent d'un délai jusqu'au 2 août 2027.

A compter du 2 août 2026, toutes les autres mesures doivent être mises en œuvre, notamment celles spécifiques aux SIA à haut risque : information des utilisateurs, surveillance humaine, journalisation, signalisation des incidents graves, etc.

## 9. Quel est l'impact concret pour les cabinets et les entreprises ?

Les cabinets, et leurs entreprises clientes, restent attentifs à l'évolution des lois ayant un effet direct sur leur pratique professionnelle. Le règlement relatif à l'IA fait partie du spectre de cette veille, sachant que certaines de ses dispositions sont déjà applicables et d'autres le seront en août 2026.

Depuis le 2 février 2025, ces structures ont l'obligation d'assurer la formation de leurs collaborateurs à l'IA. Le contenu doit être adapté à leur activité, il peut s'agir d'une formation à l'utilisation de ces systèmes mais aussi à leur contexte, notamment juridique.

Pour maîtriser leur usage et leur responsabilité, les structures sont incitées à compléter les actions de formation stricte par un partage de connaissances plus large. Celle-ci peut prendre la forme d'évolution des chartes internes et de sensibilisations régulières.

# FOCUS SUR L'AI ACT ET LES DIRECTIVES NIS/NIS2 : VERS UNE APPROCHE GLOBALE DE L'IA ET DE LA CYBERSECURITE

*Par Jean-Laurent Heim-Lienhardt*[17]

## 1. AI Act

D'après le site *artificialintelligenceact.eu/fr*, l'AI Act a pour vocation d'établir un « cadre global de régulation de l'intelligence artificielle, en adoptant une approche fondée sur les risques ». Les points clés incluent :

- La classification des systèmes d'IA selon leur niveau de risque (inacceptable, élevé, limité, minimal) et l'interdiction de certains usages jugés trop attentatoires aux droits fondamentaux.
- Des obligations pour les fournisseurs et utilisateurs de systèmes à haut risque (contrôles de conformité, traçabilité, évaluations d'impact, etc.).
- Une transparence renforcée pour les utilisateurs (ex. signaler qu'un contenu ou une interaction est généré(e) par de l'IA).
- La coordination via des autorités nationales et un comité européen de l'IA chargés de superviser l'application de la réglementation.

## 2. Directives NIS (2016) et NIS 2 (2022)

- La première directive NIS, adoptée en 2016 (Directive (UE) 2016/1148), est considérée comme la pierre angulaire de la politique de cybersécurité de l'UE. Elle impose notamment aux opérateurs de services essentiels (OSE) et aux fournisseurs de services numériques (FSN) de mettre en œuvre des mesures techniques et organisationnelles adaptées pour gérer les risques qui menacent la sécurité des réseaux et des systèmes d'information.

[17] Contenu partiellement rédigé à l'aide d'un système d'IA générative

- La directive NIS 2, adoptée en décembre 2022 (Directive (UE) 2022/2555), étend le champ d'application de la directive initiale à davantage de secteurs (ex. plateformes de réseaux sociaux, fournisseurs de services de communication électronique, etc.) et impose des obligations plus strictes en matière de gestion des risques et de notification d'incidents. Elle renforce également les mécanismes de coopération entre les autorités nationales, dans le but d'uniformiser et d'élever le niveau de cybersécurité à l'échelle de l'Union européenne.

En articulant l'IA Act avec les directives NIS/NIS2, l'Union européenne cherche à assurer une protection globale sur deux volets :

1. **La fiabilité et la transparence des applications d'IA**, tout en préservant les droits fondamentaux.
2. **La sécurité de l'infrastructure numérique**, indispensable pour un déploiement serein de l'IA.

Le point de convergence entre ces volets (protection des données, conformité réglementaire, cybersécurité) et ce corpus législatif réside dans la volonté de prévoir et de gérer les risques de manière holistique : la protection quotidienne de l'IA n'est pas uniquement une affaire de pare-feu ou de mots de passe robustes, elle nécessite un cadre juridique cohérent (RGPD, AI Act, NIS/NIS2) et une culture de la cybersécurité au sein des entreprises et chez les utilisateurs.

# COMPLEMENTARITE ENTRE LA SECURITE TECHNIQUE ET LE CADRE JURIDIQUE

*Par Jean-Laurent Heim-Lienhardt*[18]

## 1. L'efficacité des mesures techniques conditionnée par un cadre légal solide

Les mesures techniques (chiffrement, pare-feu, détection des intrusions, segmentation des réseaux, etc.) constituent la première ligne de défense contre les vulnérabilités et les menaces pesant sur l'intelligence artificielle (IA). Toutefois, leur efficacité réelle repose en grande partie sur l'existence d'un cadre légal et contractuel cohérent, garantissant la bonne gouvernance des systèmes et la protection des droits fondamentaux (CNIL, 2020).

- **Respect de la réglementation en vigueur** : sans une prise en compte des dispositions du RGPD, de l'AI Act, ou encore des directives NIS/NIS2, une entreprise, même bien équipée sur le plan technique, risque de se voir exposée à des sanctions (monétaires, réputationnelles) et de faire face à d'éventuels litiges (RGPD 2016/679 ; European Commission, 2021, 2016, 2022).
- **Gestion des responsabilités et obligations contractuelles** : la mise en œuvre d'accords-cadres, de politiques de sécurité internes et de contrats (ex. entre prestataires de services d'IA et entreprises) contribue à clarifier les responsabilités respectives en cas de faille ou de non-conformité.

Par exemple, un système d'IA qui analyserait en continu des données à caractère personnel doit être non seulement sécurisé (chiffrement, anonymisation partielle ou totale) mais également conforme aux principes de minimisation et de finalité imposés par le RGPD et au principe de proportionnalité préconisé par l'AI Act (*artificialintelligenceact.eu/fr*).

[18] Contenu partiellement rédigé à l'aide d'un système d'IA générative

## 2. L’importance du volet juridique pour soutenir et encadrer les dispositifs techniques

Inversement, un cadre légal, même très abouti, ne peut produire ses effets que s’il est réellement appliqué par le biais de mesures techniques adaptées. En effet :

- **Transposition concrète des obligations réglementaires** : par exemple, la Directive NIS 2 exige un niveau élevé de cybersécurité. Pour s’y conformer, les organisations doivent déployer des outils de détection d’incident, procéder à des audits de sécurité réguliers, et respecter des protocoles de notification en cas de cyberattaque (Directive (UE) 2022/2555). Sans ces solutions techniques, la conformité ne serait qu’une formalité administrative sans impact réel sur la sécurité.

- **Prévention des litiges et sécurité juridique** : les preuves de conformité (journaux d’événements, documentation sur la traçabilité algorithmique, etc.) s’obtiennent par l’implémentation de solutions technologiques robustes et traçables. Cela sécurise juridiquement l’entreprise en cas d’enquête ou de contrôle (CNIL, 2020).

- **Flexibilité et adaptation** : un cadre juridique impose des principes (responsabilité, transparence, respect de la vie privée), mais la mise en œuvre technique doit s’adapter aux évolutions technologiques de l’IA. Le Règlement IA (AI Act) prévoit notamment une approche par niveau de risque, qui exige un degré de souplesse et d’évolution permanente pour demeurer pertinent (European Commission, 2021).

## 3. Un tandem indispensable pour un usage quotidien sécurisé de l'IA

Le tandem technique-juridique est donc incontournable pour assurer un usage sécurisé et conforme de l'IA :

- **Synergie des expertises** : les ingénieurs en cybersécurité et les juristes spécialisés en protection des données doivent collaborer afin d'identifier les risques (techniques, légaux) et d'y apporter des solutions globales (processus de "privacy by design", contrats adaptés, cryptographie, etc.).

- **Formation et sensibilisation** : pour être efficaces, les mesures techniques et les obligations juridiques doivent être comprises et appliquées au quotidien par l'ensemble des collaborateurs. La CNIL (2020) et l'ANSSI (2021) insistent sur l'importance de la sensibilisation et de la formation continue pour réduire les risques d'erreurs humaines (phishing, gestion hasardeuse des mots de passe, etc.).

- **Évolution conjointe** : l'environnement législatif (RGPD, AI Act, NIS/NIS2) et les solutions technologiques (IA, blockchains, solutions de cybersécurité) évoluent en permanence. La veille technologique et la veille juridique sont donc essentielles pour maintenir un niveau de protection satisfaisant.

En somme, l'efficacité des mesures techniques dépend de la robustesse du cadre légal, tandis que la réussite du volet juridique repose sur des dispositifs technologiques concrets et régulièrement mis à jour. Cette complémentarité est au cœur de la démarche globale de sécurisation de l'IA et de respect des droits et libertés fondamentaux.

# CONFIDENTIALITE ET INTELLIGENCE ARTIFICIELLE : DEFIS ET RESPONSABILITES POUR LES PROFESSIONS REGLEMENTEES : UN ENTRETIEN AVEC FELICIEN VALLET (CNIL)

Entretien avec Félicien Vallet, chef du service IA à la CNIL.

***Cet entretien a été réalisé en visioconférence lors d'une réunion du groupe de travail le 19/11/2024. Il est reproduit avec l'aimable autorisation de M. Vallet que nous remercions.***

## 1. Les risques autour des données

**Vincent :** Quels sont les enjeux autour de la confidentialité des données pour les métiers du chiffre, notamment sur les défis que posent les outils d'intelligence artificielle comme ChatGPT ?

**Félicien :**
C'est une question évidemment très compliquée. Cela nous intéresse énormément à la CNIL, car ce sont des problématiques qui nous sont régulièrement posées. Nous appréhendons le sujet en le scindant en deux volets : le développement et l'utilisation des systèmes d'intelligence artificielle.

Concernant le développement, sur lequel nous avons principalement travaillé ces derniers mois, il s'agit de garantir que les données à caractère personnel soient correctement utilisées pour entraîner ces systèmes, tout en respectant le RGPD.

Cependant, la majorité des interrogations que nous recevons concernent l'utilisation pratique de ces systèmes, comme dans le cas des experts-comptables. Il est cependant difficile de répondre à ces questions d'utilisation si nous n'avons pas réglé en amont celles relatives au développement. Par exemple, quelles sont les garanties qu'un système soit développé dans les règles et puisse ensuite être exploité en toute conformité ? Nous avons publié une FAQ cet été pour clarifier certaines bases, mais de nombreuses précisions restent à apporter, notamment sur la gestion des cas d'usage spécifiques, comme ceux des métiers du chiffre.

**Vincent :** Dans un environnement numérique où la gestion des données devient critique, pensez-

vous qu'on puisse réellement éliminer le risque lié à l'utilisation de ces outils ?

**Félicien :**
Non, le risque est inhérent au monde numérique. L'enjeu n'est pas de l'éliminer totalement, car c'est impossible, mais plutôt de le réduire à un niveau raisonnable et acceptable. Cela implique d'évaluer les différents cas d'usage et de mettre en place des mesures adaptées.

Par exemple, certains scénarios d'utilisation de l'IA pourraient être considérés comme peu problématiques et gérables. Dans d'autres cas, où les données manipulées sont plus sensibles, il faudra réfléchir à des solutions renforcées pour assurer la confidentialité et la souveraineté des informations traitées.

**« il est essentiel de former [les utilisateurs] à la bonne utilisation de ces outils et de les alerter sur les mauvaises pratiques »**

## 2. Anonymiser ou pseudonymiser les données ?

**Sabrina :** Une des préconisations est **l'anonymisation des données**. Quelle est votre position à ce sujet, notamment pour les données personnelles ou professionnelles utilisées dans des systèmes comme ChatGPT ?

**Félicien : L'anonymisation** est effectivement une bonne pratique, mais il faut être précis sur ce terme. Pour sortir du champ d'application du RGPD, les données doivent être rendues totalement anonymes, ce qui est souvent complexe. Un simple retrait de noms ou prénoms ne suffit pas contrairement à une idée reçue. Les CNIL européennes considèrent que des transformations significatives sont nécessaires pour garantir qu'aucune réidentification ne soit possible.

Dans la pratique, il est souvent préférable de parler de **pseudonymisation**, qui consiste à retirer les informations directement identifiantes, comme un numéro d'identification, un nom ou une date de naissance. Bien sûr, cela dépend des cas d'usage : dans certains scénarios, la pseudonymisation peut apporter une garantie importante sans pour autant nuire au traitement de données poursuivi, mais dans d'autres, elle pourrait poser problème si des informations clés sont supprimées.

*[NDLR : voir aussi cet article de la CNIL :* https://www.cnil.fr/fr/recherche-scientifique-hors-sante-enjeux-et-avantages-de-lanonymisation-et-de-la-pseudonymisation*]*

## 3. Les solutions d'IA génératives

**Vincent :** Nous avons remarqué que la CNIL italienne a suspendu temporairement ChatGPT en 2023 en raison de certains risques. Quelles leçons tirez-vous de cette décision et des suites données à cette affaire ?

**Félicien :** Cette décision de la CNIL italienne, en mars 2023, s'inscrivait dans une procédure d'urgence[19]. Les autorités italiennes avaient estimé que certains risques, notamment liés à la réutilisation des données, justifiaient la suspension du service dans le pays.

Suite à cette suspension, OpenAI a proposé plusieurs modifications, notamment la possibilité pour les utilisateurs d'exclure leurs données de l'entraînement des modèles[20]. Cela a permis de rétablir l'accès à ChatGPT en Italie après environ un mois.

Cependant, des manquements structurels persistent, notamment sur l'utilisation des données personnelles lors de l'entraînement des modèles. Ces questions restent en cours d'examen par plusieurs CNIL européennes. Une task force a été mise en place pour harmoniser les approches des différentes autorités nationales face à ces problématiques.

*[NDLR : Quelques semaines après l'entretien, le 20 décembre 2024, la CNIL italienne a sanctionné OpenAI avec une amende de 15 millions d'euros]*

## 4. Bonnes pratiques

**Vincent :** Pour les professions réglementées, comme les experts-comptables, qui travaillent avec des données sensibles, quels conseils donnez-vous concernant l'intégration d'outils comme ChatGPT dans leur quotidien ?

**Félicien :** La première étape est de clarifier l'usage précis attendu et en quoi un système d'IA générative apporte une réponse adaptée à un problème bien identifié. Il s'agit ensuite d'établir une pédagogie et une sensibilisation rigoureuses des utilisateurs. Cela peut paraître basique, mais il est essentiel de former ces derniers à la bonne utilisation de ces outils et de les alerter sur les mauvaises pratiques, comme injecter directement des documents sensibles dans des systèmes non maîtrisés.

Ensuite, il faut évaluer chaque cas d'usage pour déterminer les outils les plus adaptés aux besoins spécifiques. Fonction de différents critères, comme le cas d'usage ou la sensibilité des données, on pourra préférer le déploiement d'un

[19] NDLR : article 66 du RGPD : https://www.cnil.fr/fr/reglement-europeen-protection-donnees/chapitre7#Article66

[20] NDLR : option pour désactiver l'entrainement des donnés (*opt out*) disponible sur ChatGPT

système on-premise ou dans un cloud auto-administré au recours à un système disponible sous forme d'API.

**Virginie :** Les experts-comptables, tout comme les avocats et notaires, sont des professions réglementées avec des obligations déontologiques. Comment concilier ces règles avec les recommandations de la CNIL en matière de protection des données ?

**Félicien :**
La difficulté ici réside dans l'intersection entre les exigences déontologiques et les obligations réglementaires. Les professions réglementées ont des responsabilités uniques, notamment en termes de confidentialité, qui peuvent compliquer l'adoption d'outils d'IA.

La clé est de bien distinguer les responsabilités : un cabinet d'expertise comptable agit souvent comme responsable du traitement des données et non comme sous-traitant. Cela signifie qu'il a une grande autonomie dans les choix technologiques, mais aussi une responsabilité accrue. Une bonne pratique serait d'informer clairement les clients des outils utilisés, par exemple dans les conditions générales de prestation. En cas de refus d'un client, il faudrait évaluer s'il est possible de proposer des alternatives.

**Sabrina :** Concernant les durées de conservation des données, y a-t-il des recommandations spécifiques à la lumière de l'utilisation de l'intelligence artificielle ?

**Félicien :** Les durées de conservation doivent être définies en fonction des besoins spécifiques. Pour les professions réglementées, cela peut être lié à la responsabilité professionnelle, qui peut s'étendre sur plusieurs décennies. Cela pose des défis en termes de stockage et de compatibilité technologique à long terme.

Il est important de distinguer entre les bases actives, nécessaires au fonctionnement quotidien, et l'archivage intermédiaire, mais cela reste difficile dans certains contextes. Dans tous les cas, les choix doivent être justifiés et documentés pour être conformes au RGPD.

**Vincent :** De nombreux outils logiciels intègrent désormais de l'intelligence artificielle. Par exemple, des solutions SaaS utilisées par les experts-comptables exploitent déjà des algorithmes d'IA pour optimiser les processus. Ces outils nécessitent-ils une communication spécifique envers les clients, ou cette responsabilité incombe-t-elle aux éditeurs des logiciels ?

**Félicien :** C'est une question qui touche à la responsabilité partagée entre l'utilisateur final, ici

le cabinet, et les éditeurs de logiciels. En tant que responsable du traitement des données, le cabinet doit veiller à informer ses clients si des données sont transmises ou utilisées d'une manière nouvelle, même si cela passe par un outil tiers.

Cependant, l'éditeur a également des obligations, notamment en termes de transparence vis-à-vis des fonctionnalités IA intégrées à ses produits. Il est important de s'assurer que les solutions choisies respectent les standards de sécurité et de conformité attendus. Dans ce cadre, le rôle de l'ordre professionnel peut être crucial pour aider à identifier les solutions les plus adaptées et conformes.

**Virginie :** Une des difficultés soulevées est l'utilisation de données sensibles pour entraîner des modèles. Est-ce que la CNIL considère que les éditeurs doivent obtenir un consentement explicite ou s'appuyer sur d'autres bases légales comme l'intérêt légitime ?

**Félicien :** C'est effectivement une question complexe. Il est possible de recourir l'utilisation de l'intérêt légitime comme base légale pour entraîner des modèles à partir de données personnelles. Toutefois, cela n'est pas envisageable dans tous les cas et doit être correctement justifié.[21]

Le RGPD exige une analyse minutieuse des risques pour les droits et libertés des personnes concernées. Si l'intérêt légitime est invoqué, les utilisateurs doivent être clairement informés, et il doit être possible de s'opposer à cette utilisation, ce qui reste difficilement réalisable pour des systèmes d'IA génératifs lorsqu'ils recourent à des techniques de moissonnage des données en ligne (*webscraping*). Nous avons vu des cas où des entreprises n'ont pas pu garantir ce droit d'opposition de manière fonctionnelle, ce qui a conduit à des blocages. Des travaux adoptés récemment au niveau européen viennent préciser ces aspects[22].

**Vincent :** Sur la base de ce que vous avez mentionné, est-ce que la CNIL pourrait envisager des contrôles dans les cabinets d'experts-comptables ou d'autres petites structures pour s'assurer de la bonne utilisation de ces outils ?

**Félicien :** Oui, c'est tout à fait envisageable. Cela pourrait se faire dans le cadre de plaintes déposées par des individus ou dans une

[21] Voir : https://www.cnil.fr/fr/base-legale-interet-legitime-developpement-systeme

[22] https://www.edpb.europa.eu/news/news/2024/edpb-opinion-ai-models-gdpr-principles-support-responsible-ai_en

démarche proactive de la CNIL pour vérifier les pratiques dans les professions réglementées.

Cela dit, ces contrôles ne visent pas à sanctionner systématiquement, mais également à « prendre le pouls » d'un secteur et, le cas échéant, à encourager de bonnes pratiques et sensibiliser les acteurs. La priorité reste d'opérer en amont de la chaîne de valeur, c'est-à-dire chez les éditeurs de logiciels, pour s'assurer que les solutions mises sur le marché respectent les réglementations en vigueur et ainsi maximiser la prise en compte des enjeux de protection des données.

**Sabrina :** Lorsque des outils d'IA intégrés dans nos logiciels métiers exploitent les données de nos clients pour optimiser leurs fonctionnalités, devons-nous inclure des clauses spécifiques dans nos lettres de mission ?

« investissez dans la sensibilisation et la formation de vos équipes »

**Félicien :** Oui, cela semble être une bonne pratique. En tant que responsable de traitement, il est essentiel de garantir la transparence vis-à-vis des clients sur la manière dont leurs données sont utilisées. Cela peut inclure une clause précisant que certaines données pourront être traitées par des outils d'intelligence artificielle, avec des explications claires sur l'objectif et les mesures de sécurité associées.

Il faut également réfléchir à la base légale de ce traitement : l'intérêt légitime pourrait être invoqué, mais cela nécessite une justification rigoureuse et une information préalable aux clients. Si certains refusent, il faudra prévoir des mécanismes pour répondre à leurs attentes sans utiliser l'IA, ce qui peut être complexe mais nécessaire.

**Vincent :** Concernant les durées de conservation, dans des professions comme la nôtre où la responsabilité professionnelle peut être engagée sur plusieurs décennies, peut-on s'appuyer sur ces exigences pour justifier des délais très longs ?

**Félicien :** Oui, les obligations professionnelles, notamment en matière de responsabilité, peuvent justifier des durées de conservation prolongées. Cependant, cela implique de prendre en compte des contraintes techniques : est-il réaliste de conserver des fichiers numériques pendant 20 ou 30 ans tout en garantissant leur lisibilité ?

Une solution consiste à segmenter les données en bases actives et archives intermédiaires, mais je comprends que cela peut être difficile à mettre en place dans certaines professions. L'important est de bien documenter vos choix et de les aligner avec les réglementations en vigueur.

**Virginie :** Nous avons également noté que l'IA génère une instabilité réglementaire, avec des lois en cours d'écriture et des technologies qui évoluent très vite. Comment les professions réglementées peuvent-elles s'adapter ?

**Félicien :** Cette instabilité est en effet un défi majeur. Les réglementations évoluent lentement par rapport à la rapidité des avancées technologiques et la CNIL fait le constat du fort besoin de sécurité juridique des acteurs. Il peut donc apparaître difficile de prendre des décisions stratégiques durables dans ce contexte.

Pour les professions réglementées, il est essentiel de se concentrer sur les bonnes pratiques actuelles, de suivre de près l'évolution des réglementations et d'anticiper les changements en maintenant une certaine flexibilité dans les choix technologiques. Cela peut également inclure des échanges réguliers avec des experts et des régulateurs pour mieux comprendre les implications de ces évolutions.

**Vincent :** Pour finir, si vous aviez un conseil à donner aux professionnels du chiffre qui souhaitent intégrer l'intelligence artificielle, quel serait-il ?

**Félicien :** Mon principal conseil serait de ne pas céder à l'effet de mode sans réflexion. Prenez le temps **d'évaluer les besoins réels de votre activité et les solutions** qui y répondent le mieux. Priorisez toujours la sécurité et la conformité.

Enfin, investissez dans la **sensibilisation et la formation de vos équipes**. Les outils d'IA peuvent être extrêmement puissants, mais leur efficacité et leur sécurité dépendent avant tout de la manière dont ils sont utilisés.

# PARTIE 2

# LES RISQUES LIÉS À L'IA GÉNÉRATIVE

# LES RISQUES GENERAUX

## 1. Panorama introductif des risques

*Par Vincent Lacomme*

L'utilisation des IA génératives soulève de sérieuses questions pour les experts-comptables et métiers du chiffre et notamment :

▶ **Transparence et responsabilité** : l'IA est souvent qualifiée de boîte noire. Il est difficile de comprendre comment l'IA prend des décisions, ce qui soulève des questions sur la responsabilité des résultats générés par ces systèmes. Certaines sociétés conceptrices de *large langage models* (LLM) ne communiquent pas le contenu de leurs bases de données d'entrainement et manquent de transparence sur leurs processus internes[23].

▶ **Flou sur les traitements des données** (articles 12 §3 et 15 de la RGPD[24]) : le manque de transparence de certains acteurs implique des incertitudes sur l'utilisation et le traitement des données :

- Sur le plan de transferts transfrontaliers de données : le règlement européen sur la protection des données (RGPD) encadre précisément les transferts internationaux de données[25].
- Sur la transparence algorithmique : les IA « boîte noire » rendant difficiles l'explicabilité et la traçabilité de la réponse formulée.

▶ **Biais algorithmique et erreurs (« hallucinations »)** : l'IA peut perpétuer des biais présents dans les données de formation, ce qui peut conduire à des résultats injustes et discriminatoires.

Les IA peuvent ainsi violer l'article 5 §1 d de la RGPD et le principe de l'exactitude des données personnelles[26].

[23] Sénat, *Rapport sur les nouveaux développements de l'intelligence artificielle*, page 118, 29/11/2024
[24] https://www.cnil.fr/fr/reglement-europeen-protection-donnees/chapitre3#Article12
[25] https://www.cnil.fr/fr/transferts-de-donnees-hors-ue-le-cadre-general-prevu-par-le-rgpd
[26] https://www.cnil.fr/fr/reglement-europeen-protection-donnees/chapitre2#Article5

▶ **Attaque par empoisonnement** (*data poisoning attack*) : selon la CNIL, elles consistent à introduire des données corrompues en phase d'entraînement » du modèle de langage d'IA[27].

▶ **Confidentialité des données** : l'utilisation de données sensibles, protégées par le secret professionnel, le RGPD.

L'article 226-13 du Code pénal français sanctionne la divulgation d'informations confidentielles par l'expert-comptable ou le commissaire aux comptes. Cet article est rappelé à l'article 21 de l'ordonnance de 1945.

Respect des points suivants de la RGPD :

- **Minimisation des données** : ne collecter que les données strictement nécessaires à la finalité poursuivie.
- **Finalité** : utiliser les données uniquement pour les finalités déclarées.
- **Exactitude** : veiller à l'exactitude des données et les mettre à jour si nécessaire.
- **Sécurité** : mettre en place des mesures de sécurité appropriées pour protéger les données contre la perte, l'accès non autorisé ou la modification.
- **Transparence** : informer les personnes concernées de la manière dont leurs données sont traitées.

Notons également que les éditeurs de solutions informatiques sont soumis à une obligation de moyen et non de résultat quant à la sécurité des données.

▶ **Secret des affaires** : le secret des affaires protège les informations confidentielles d'une entreprise lui conférant un avantage concurrentiel.

[27] https://www.cnil.fr/fr/definition/attaque-par-empoisonnement-data-poisoning-attack

**Exfiltration de données** : l'utilisation de *prompts* malicieux pourrait en effet permettre de récupérer des données d'autres utilisateurs[28]. Les données d'entrainement ou les requêtes d'autres utilisateurs à l'IA pourraient également être récupérées par des techniques d'attaques adverses[29].

**Des risques de cybersécurité liés à l'IA** : les systèmes d'IA pouvant être vulnérables et interconnectés à des systèmes critiques.

Ces différents risques sont assortis de sanctions :

- **Sanctions pénales** : en cas de violation du secret professionnel ou du secret des affaires.
- **Sanctions disciplinaires**
- **Sanctions administratives** : amendes prononcées par la CNIL en cas de non-respect du RGPD.
- **Atteinte à la réputation** : une fuite de données peut ternir l'image de l'expert-comptable et nuire à sa crédibilité.
- **Perte de confiance des clients** : une fuite de données peut entraîner une perte de confiance et la rupture de la relation.

## 2. Risques sur la confidentialité

*Par Sylvain Navers*

La plupart des IAGen utilisées sur le système d'information local ou via Internet, communiquent avec des éléments exogènes, voire stockent des données sur des systèmes cloud dont la frontière n'est pas maitrisée.

Ainsi l'envoi de données internes, même anodines, peut induire des risques sur la confidentialité d'un ensemble de données plus importantes, voire de stratégies d'entreprise.

---

[28] Gregory Schwartzman, *Exfiltration of personal information from ChatGPT via prompt injection,* 06/06/2024

[29] Une attaque adverse (adversarial attack), parfois aussi appelée « attaque antagoniste » ou « attaque par exemples contradictoires » vise à envoyer à un système d'IA une ou plusieurs requêtes malveillantes dans le but de tromper ou d'altérer son bon fonctionnement (ANSSI).

Nous avons identifié les risques suivants :

1. Perte d'information non formalisée mais explicitée lors de l'échange avec l'IAGen
2. Reconstruction d'une information classifiée d'une entreprise par récupération d'un ensemble d'information non classifiées
3. Divulgation d'une information classifiée par divulgation excessive non maitrisée

## 3. Plusieurs faits révélateurs des dangers d'une utilisation non maitrisée de l'IA générative

***Par Vincent Lacomme***

La digitalisation croissante des entreprises a créé de nouveaux risques qui ne datent pas de l'arrivée de l'IA génératives. On peut citer un recours collectif déposé en 2019 à l'encontre de la solution Siri d'Apple qui pouvait s'activer et enregistrer accidentellement des conversations privées sans le consentement de son utilisateur[30]. Les données en question étaient mises à la disposition de sous-traitants chargés du contrôle qualité chez Apple.

En 2023, quelques semaines après la levée d'une interdiction d'utilisation de ChatGPT au sein de Samsung, des salariés de la société ont déposé des données stratégiques (code informatique critique) sur ChatGPT[31]. Ces données ont pu servir pour l'entrainement du modèle de langage de l'IA.

Le 9 novembre 2023, ChatGPT a fait l'objet d'une cyberattaque par déni de service (*Denial of service*) revendiquée par le groupe Anonymous Sudan.

Entre juin 2022 et mai 2023, plus de 100 000 comptes ChatGPT ont été compromis et vendus sur le dark web. En France, près de 3 000 comptes étaient concernés. Ces piratages ont été réalisés via des malwares appelés "info stealers" qui collectent les informations enregistrées dans les navigateurs[32].

En 2024, OpenAI a révélé avoir déjoué une vingtaine d'attaques malveillantes pour développer des malwares[33].

---

[30] https://www.lemonde.fr/economie/article/2025/01/02/apple-accepte-de-payer-pres-de-100-millions-de-dollars-pour-mettre-fin-a-des-poursuites-sur-la-confidentialite-des-donnees_6478887_3234.html
[31] https://gizmodo.com/chatgpt-ai-samsung-employees-leak-data-1850307376
[32] https://www.group-ib.com/media-center/press-releases/stealers-chatgpt-credentials/
[33] *Influence and cyber operations: an update*, Open AI, octobre 2024 ? https://cdn.openai.com/threat-intelligence-reports/influence-and-cyber-operations-an-update_October-2024.pdf

Une étude d'un institut technologique au Japon révèle qu'il est possible d'exfiltrer des données personnelles d'autres utilisateurs via un *prompt* posé à ChatGPT[34]. En d'autres termes, des personnes malveillantes peuvent potentiellement récupérer les données de votre compte ChatGPT[35].

Enfin, des utilisateurs ont pu détourner l'utilisation de ChatGPT pour obtenir des réponses violant les règles de bonne conduite grâce à des *skeleton keys* (clés pour contourner les limitations de l'IA).

## 4. Expliciter les finalités : prévenir les risques techniques et juridiques liés à l'usage de l'IA

***Par Jean-Laurent Heim-Lienhardt[36]***

Avant d'adopter des mesures de protection au quotidien, il est essentiel de préciser les raisons et les enjeux sous-jacents. Les objectifs de ce cahier se concentrent sur la prévention des risques liés à l'usage de l'IA, avec un accent particulier sur :

- **La sécurité technique**
  Assurer que les systèmes d'IA et les données associées soient protégés contre les intrusions (cyberattaques, malwares) et les défaillances (fuites de données sensibles, corruption des algorithmes). Le Règlement Général sur la Protection des Données (RGPD) insiste d'ailleurs sur l'importance de mettre en place des « mesures techniques et organisationnelles appropriées pour garantir un niveau de sécurité adapté » (Article 32, RGPD 2016/679).

- **La conformité juridique**
  Veiller à respecter les textes légaux et réglementaires nationaux et internationaux encadrant la collecte, le traitement et le stockage des données (CNIL, 2020 ; European Commission, 2021). Cela implique, entre autres, la transparence quant aux traitements effectués par des algorithmes d'IA, la limitation de la finalité du traitement des données et la justification des décisions automatisées (CNIL, 2017).

[34] Juin 2024
[35] Gregory Schwartzman, *Exfiltration of personal information from ChatGPT via prompt injection*, juin 2024, https://arxiv.org/abs/2406.00199v2
[36] Contenu partiellement rédigé à l'aide d'un système d'IA générative

- **L’anticipation des nouvelles obligations**
    - AI Act : dont l’entrée en vigueur partielle en février 2025, introduit un cadre réglementaire spécifique pour l’IA, visant à « assurer que l’intelligence artificielle respecte les droits fondamentaux, la sécurité et la protection des utilisateurs » (artificialintelligenceact.eu/fr).
    - Directives NIS et NIS 2 : visant à garantir un niveau élevé de cybersécurité dans l’Union européenne, elles imposent des obligations aux opérateurs de services essentiels et aux fournisseurs de services numériques pour la gestion des risques et la notification des incidents (Directive (UE) 2016/1148 ; Directive (UE) 2022/2555).

L’objectif est également de sensibiliser tous les acteurs (collaborateurs d’entreprises, prestataires de services, grand public) à l’importance de comprendre et d’anticiper les menaces émergentes dans un contexte où l’IA se généralise. L’Agence Nationale de la Sécurité des Systèmes d’Information (ANSSI) rappelle en effet que « l’augmentation du degré d’automatisation et d’interconnexion accroît d’autant la surface d’attaque potentielle » (ANSSI, 2021).

### 4.1. Délimiter le champ d'application : focus sur la protection des données, la conformité réglementaire et la cybersécurité

Pour proposer des recommandations concrètes et opérationnelles, le périmètre de la présentation se concentre sur trois axes fondamentaux :

- **Protection des données**
  Analyser la nature des données exploitées par les systèmes d'IA (données personnelles, données confidentielles ou stratégiques, etc.) et déterminer les mesures adéquates pour en protéger la confidentialité, l'intégrité et la disponibilité (RGPD 2016/679 ; CNIL, 2020).

- **Conformité réglementaire**
  Ce volet porte sur l'ensemble des obligations légales imposées aux entreprises et aux organisations (par exemple, la tenue d'un registre des activités de traitement, l'information des utilisateurs, la mise en place d'un délégué à la protection des données, etc.). Il inclut également la veille juridique pour anticiper les évolutions, notamment celles liées :
  - Au Règlement européen sur l'IA,
  - Aux Directives NIS et NIS 2, qui visent à renforcer la résilience des réseaux et systèmes d'information dans les secteurs critiques et à améliorer la coopération entre États membres (European Commission, 2016 ; 2022).

- **Cybersécurité**
  Au-delà de la seule protection des informations, il convient de s'assurer que les systèmes et algorithmes d'IA eux-mêmes sont conçus et maintenus en respectant les bonnes pratiques de sécurité (gestion des accès, chiffrement, segmentation des réseaux, etc.). Dans son Guide d'hygiène informatique, l'ANSSI (2021) répertorie plusieurs recommandations à appliquer en continu, notamment la mise à jour régulière des systèmes et l'utilisation d'outils de détection et de surveillance.

# LE VOLET TECHNIQUE : GESTION DES RISQUES CYBER ET CYBERSECURITE

*Par Jean-Laurent Heim-Lienhardt*[37]

## 1. Identifier les menaces spécifiques à l'IA

L'essor de l'Intelligence Artificielle (IA) génère de nouveaux risques pour la sécurité et la confidentialité des données. Selon le rapport de l'ENISA (2020), « les attaques visant l'IA peuvent conduire à des manipulations critiques du système, avec des conséquences potentiellement graves pour la sécurité ». Afin de comprendre ces menaces et d'y faire face, il convient de mettre en évidence quatre catégories principales d'attaques.

### 1.1. Fuites et vols de données

Les solutions d'Intelligence Artificielle (IA) reposent sur des volumes conséquents de données, souvent sensibles ou stratégiques. Une mauvaise gestion de ces informations — que ce soit durant la collecte, l'entraînement du modèle ou la mise en production — expose l'organisation à des risques de fuites et de vols de données, susceptibles d'avoir de lourdes conséquences sur sa réputation, son patrimoine informationnel et sa conformité réglementaire.

#### *A. Vecteurs d'attaque majeurs*

- **Exfiltration de données** : un attaquant exploite des configurations défaillantes (serveurs mal sécurisés, erreurs de droits d'accès) pour s'emparer des fichiers d'entraînement ou d'inférences.
- **Data poisoning** : l'introduction volontaire de données fausses ou malveillantes dans le jeu d'apprentissage, visant à biaiser les résultats du modèle ou à en compromettre la fiabilité.

[37] Contenu partiellement rédigé à l'aide d'un système d'IA générative

- **Vulnérabilités humaines** : le manque de formation ou la négligence (cliquer sur un lien de phishing, stocker des identifiants en clair, etc.) facilite l'accès aux informations sensibles.

### *B. Impacts et conséquences*

- **Atteinte à la confidentialité** : la divulgation de données personnelles ou financières peut conduire à des sanctions (RGPD), une perte de confiance et des coûts de remédiation élevés.
- **Perte de propriété intellectuelle** : l'accès non autorisé à des algorithmes propriétaires ou à des informations techniques peut compromettre l'avantage concurrentiel d'une entreprise.
- **Risques juridiques** : dans le cadre des réglementations en vigueur (notamment le RGPD), les organismes victimes d'une fuite de données peuvent être contraints de notifier les personnes concernées et les autorités compétentes, s'exposant à des sanctions administratives et financières.

### *C. Bonnes pratiques de prévention*

- **Chiffrement systématique** : adopter une approche de chiffrement en transit (TLS) et au repos (chiffrement des disques, conteneurs ou bases de données), conformément aux recommandations du NIST (SP 800-53).
- **Minimisation des données** : la CNIL (2017) souligne l'importance de « n'exploiter que les données strictement nécessaires à l'objectif de l'algorithme ». Réduire la quantité de données sensibles diminue mécaniquement l'impact d'une éventuelle fuite.
- **Contrôle d'accès rigoureux** : mettre en place une gestion fine des droits utilisateurs (ISO/IEC 27005), segmenter les environnements de développement et de production, et utiliser des solutions d'authentification multi-facteur.
- **Surveillance et détection** : recourir à des solutions de monitoring en temps réel (SIEM, IDS/IPS) pour détecter rapidement tout comportement anormal (téléchargement massif, connexions suspectes).
- **Formation du personnel** : sensibiliser les collaborateurs aux bonnes pratiques de sécurité et aux risques d'ingénierie sociale.
- **Clauses contractuelles** : exiger de la part des prestataires et sous-traitants un niveau de sécurité équivalent, avec des audits réguliers et des protocoles de notification en cas d'incident.

La maîtrise des risques de fuites et de vols de données est primordiale pour toute entité souhaitant déployer des projets d'IA. En combinant une gouvernance claire, des contrôles techniques robustes et une sensibilisation permanente du personnel, les organisations peuvent renforcer significativement leur posture de sécurité, tout en respectant leurs obligations réglementaires et éthiques.

---

***Exemple : Fuite de données clients d'une plateforme de e-commerce***

*Une configuration erronée d'un serveur de stockage a rendu accessible, sans authentification, le dossier contenant les logs d'entraînement du modèle IA. Résultat : divulgation d'adresses e-mail, de données de navigation et d'achats, entraînant des répercussions légales et un préjudice d'image.*

---

---

***Exemple : Injection malveillante dans un jeu d'apprentissage***

*Un employé dépose des données altérées dans le répertoire de formation du modèle, rendant ce dernier inapte à détecter certains comportements frauduleux. L'attaque n'a été découverte qu'après la détection d'un volume anormal de fraudes, traduisant un manque de surveillance continue.*

---

### 1.2. Attaques par déni de service (DDoS)

Dans un contexte où les systèmes d'Intelligence Artificielle (IA) sont de plus en plus utilisés pour des fonctions sensibles (diagnostic médical, algorithmes financiers, chatbots de service client, etc.), la disponibilité devient un enjeu majeur. Les attaques par déni de service (DoS) et, plus spécifiquement, par déni de service distribué (DDoS) consistent à submerger un service de requêtes malveillantes afin de le rendre indisponible pour les utilisateurs légitimes.

### A. *Nature des attaques DDoS et vecteurs de menace*

- Les attaquants utilisent des **botnets** ou des **techniques d'amplification** (DNS, NTP, etc.) pour générer des flux massifs de trafic.
- Les IA, souvent déployées via des infrastructures connectées, sont particulièrement exposées si des mesures de limitation de débit (rate-limiting) ou de filtrage ne sont pas mises en place.

### B. *Impact sur la disponibilité et la continuité d'activité*

- Une attaque DDoS visant un modèle IA peut paralyser un processus complet, comme le diagnostic médical à distance ou l'évaluation automatisée de crédits financiers.
- Outre la perte financière, l'indisponibilité du service fragilise la confiance des utilisateurs et partenaires, pouvant conduire à des litiges ou des sanctions réglementaires.

### C. *Prévention et limitations*

- **Cloud computing et redondance** : l'ENISA (2020) recommande une architecture distribuée et élastique, capable d'absorber les pics de trafic malveillant. Les fournisseurs cloud proposent souvent des solutions anti-DDoS (filtrage, équilibrage de charge).
- **Pare-feu applicatif et filtrage avancé** : un web application firewall (WAF) permet de filtrer le trafic en fonction de règles prédéfinies, bloquant les requêtes suspectes.
- **Surveillance continue** : l'utilisation d'outils SIEM (Security Information and Event Management) ou IDS/IPS (Intrusion Detection/Prevention System) aide à détecter rapidement un volume anormal de connexions.
- **Plans de continuité d'activité** : selon ISO/IEC 27005, définir un plan de réponse aux incidents et tester la capacité de l'infrastructure à résister à un DDoS (stress tests) sont des éléments essentiels de la gestion des risques.

### *D. Gestion globale du risque et sensibilisation*

- Au-delà des solutions techniques, il est fondamental de former les équipes IT et de sécurité à repérer les premiers signes d'une attaque DDoS (augmentation soudaine de la latence, saturation du trafic, etc.).
- Le partage d'informations avec les partenaires (fournisseurs, hébergeurs, etc.) est déterminant : coordonner la riposte limite la portée de l'attaque et facilite la restauration du service.
- Sur le plan juridique et réglementaire, des secteurs critiques (finance, énergie, santé) doivent respecter des standards de disponibilité minimum, sous peine de sanctions administratives.

Les attaques par déni de service (DDoS) représentent une menace sérieuse pour les systèmes IA, car elles ciblent directement la disponibilité, un pilier essentiel de la sécurité. Pour limiter les risques, la combinaison d'infrastructures résilientes (cloud, redondance, filtrage) et de démarches organisationnelles (plans de continuité, formation, coordination) demeure la meilleure stratégie. Investir dans la prévention et la détection précoce constitue donc un enjeu incontournable pour toute organisation misant sur l'IA.

---

***Exemple : Attaque volumétrique sur un chatbot de e-commerce***

*En ciblant le point d'accès public de l'API, les attaquants saturent les ressources, rendant le service client indisponible. Résultat : perte de ventes et impact négatif sur l'image de marque.*

---

***Exemple : Menace sur un service de télémédecine***

*La plateforme IA d'une clinique subit une vague de requêtes automatisées. Privés d'accès, les patients ne peuvent plus bénéficier d'un diagnostic rapide, ce qui pourrait engager la responsabilité de la clinique en cas de retard de traitement.*

### 1.3. Manipulation et falsification de la sortie des algorithmes (model hacking)

L'essor de l'intelligence artificielle (IA) s'accompagne d'une intensification des tentatives de manipulation des modèles. Le model hacking vise à fausser ou détourner la sortie d'un algorithme en modifiant ses entrées de manière malveillante. L'ENISA (2020) prévient que « la manipulation des algorithmes d'IA peut avoir un impact direct sur la fiabilité et la sécurité globale d'un système », tandis que la CNIL (2017) souligne « la nécessité d'un encadrement méthodique de la conception et du déploiement des algorithmes ».

#### *A. Modes d'attaque et conséquences*

1. **Adversarial examples :** les attaquants introduisent des perturbations imperceptibles dans les données d'entrée (images, sons, textes) pour tromper le modèle.
2. **Phase d'inférence vulnérable :** des flux de données trafiquées pendant l'exploitation (production) peuvent entraîner des décisions erronées, avec des conséquences potentiellement critiques (erreur de diagnostic, accord de crédit injustifié, etc.).

**Réputation et responsabilité**

**Un modèle compromis remet en cause la confiance envers l'organisation et peut entraîner des litiges si la décision est jugée préjudiciable.**

#### *B. Processus de contrôle et validation robustes*

- **Adversarial training** : inclure des exemples malveillants dans la phase d'entraînement pour renforcer la robustesse du modèle.
- **Test d'intégrité et audit régulier** : appliquer des outils d'évaluation périodique afin de détecter d'éventuelles dérives ou vulnérabilités (ISO/IEC 27005).
- **Authentification des sources de données** : établir un protocole strict de contrôle avant l'ajout ou la mise à jour des jeux de données d'apprentissage.

### C. *Bonnes pratiques de protection*

- **Approche DevSecOps** : intégrer des tests de sécurité, des revues de code et des analyses de risques à chaque étape du cycle de développement de l'IA.
- **Détection d'anomalies** : recourir à des systèmes de monitoring (SIEM, IDS/IPS) capables de repérer des comportements inhabituels du modèle (taux d'erreur anormal, résultats incohérents).
- **Formation des équipes** : sensibiliser les *data scientists* et ingénieurs ML aux tactiques courantes de model hacking et encourager une veille active sur les nouvelles techniques adversariales.
- **Documentation et plan de réponse** : définir des procédures claires (plans d'action, personnes responsables, escalade) en cas de détection d'une falsification.

La manipulation et la falsification de la sortie des algorithmes constituent une menace grandissante pour les organisations qui s'appuient sur l'IA. Des processus de contrôle et de validation robustes permettent de limiter significativement ces risques, en assurant la qualité des données, la sécurisation du cycle de développement et la surveillance en production. À l'heure où la confiance dans l'IA est un enjeu stratégique, la mise en place d'une stratégie de cybersécurité proactive s'impose comme un prérequis indispensable.

---

***Exemple : Escroquerie au scoring de crédit***

*Un attaquant manipule subtilement les données d'entrée (revenus, historique bancaire) pour être classé en « bon payeur ». Le modèle IA octroie alors un prêt injustifié, causant une perte financière pour l'institution émettrice.*

---

***Exemple : Système de radiologie***

*Dans un hôpital, des images médicales sont altérées pour masquer une tumeur dans les radiographies, faussant le diagnostic et retardant la prise en charge du patient.*

### 1.4. Élévation de privilèges et intrusions

La protection des systèmes d'Intelligence Artificielle (IA) ne peut être complète sans tenir compte des menaces liées à l'élévation de privilèges et aux intrusions. L'ENISA (2020) rappelle en effet que « l'exploitation d'API non sécurisées ou de vulnérabilités logicielles peut entraîner une élévation de privilèges », conférant à l'attaquant un accès accru aux données et aux fonctionnalités critiques de l'IA.

#### *A. Vecteurs d'attaque principaux*

- **Interfaces API mal sécurisées** : absence d'authentification robuste ou de contrôle d'accès granularisé.
- **Vulnérabilités logicielles** : failles non corrigées (patch management insuffisant), bibliothèques obsolètes, mauvaise configuration.
- **Comptes à privilèges mal gérés** : identifiants par défaut, privilèges trop étendus, absence de révision périodique des droits.

#### *B. Enjeux et conséquences*

- **Prise de contrôle du modèle** : un intrus peut modifier les paramètres d'entraînement, altérer les résultats ou dérober des secrets de fabrication (propriété intellectuelle).
- **Vol massif de données** : accès non autorisé aux bases de données d'apprentissage ou aux logs contenant des informations confidentielles.

**Risques juridiques et réputationnels**

**En cas de compromission, l'organisation est exposée à des sanctions (RGPD), à une perte de confiance et à de possibles poursuites.**

### *C. Mécanismes de protection et de détection*

- **Contrôle d'accès et authentification multi-facteur (MFA)** : gérer finement les droits utilisateurs et limiter l'accès aux seules ressources nécessaires (principe du moindre privilège).
- **Segmentation des environnements** : cloisonner les environnements de développement, de test et de production pour réduire l'impact d'une intrusion.
- **Tests d'intrusion et audits réguliers** : évaluer la robustesse du système face à des scénarios d'attaque réalistes (cf. ISO/IEC 27005).
- **Systèmes de détection des intrusions (IDS/IPS) et journalisation** : surveiller l'activité du réseau et consigner les événements critiques pour identifier rapidement toute anomalie.

### *D. Bonnes pratiques de validation et de contrôle*

- **Politique de moindre privilège** : aucun utilisateur ou service ne doit disposer de droits supérieurs à ce qui est strictement nécessaire.
- **Suivi et révocation des comptes inactifs** : désactiver systématiquement les accès obsolètes pour réduire la surface d'attaque.
- **Traçabilité et audits** : archiver et analyser les logs d'accès, assurant une vision claire des actions menées sur le système.
- **Plan de réponse aux incidents** : définir des procédures claires (identification, confinement, éradication, restauration) et prévoir l'escalade si nécessaire.

Face aux menaces d'élévation de privilèges et d'intrusions, la mise en place de processus et d'outils techniques de contrôle et de validation robustes est incontournable. Une gestion stricte des droits d'accès, la segmentation des environnements et la surveillance continue des activités réseau s'avèrent

essentiels pour préserver l'intégrité et la confidentialité des données, ainsi que la fiabilité des modèles IA. Dans un contexte réglementaire exigeant (RGPD, recommandations CNIL), ces mesures contribuent également à renforcer la confiance des partenaires et des utilisateurs.

---

**Exemple : Intrusion via une API mal configurée**

Un attaquant obtient un accès administrateur et modifie le modèle IA, entraînant un dysfonctionnement généralisé du service.

---

---

***Exemple : Élévation de privilèges par un collaborateur interne***

*Profitant d'un compte à privilèges non révoqué, un employé exfiltre des données confidentielles, mettant l'organisation en péril.*

---

## 2. Mesures de protection et bonnes pratiques

La sécurisation de l'Intelligence Artificielle (IA) ne se limite pas à identifier les menaces : elle implique également la mise en œuvre d'un ensemble cohérent de mesures de protection et de bonnes pratiques, tout au long du cycle de vie du système IA. L'ENISA (2020) rappelle que « l'adoption de mécanismes de chiffrement et de contrôles d'accès granulaires doit être considérée comme un prérequis à toute mise en œuvre de l'IA dans un contexte sensible ». De plus, la CNIL (2017) insiste sur « des mesures techniques et organisationnelles adaptées à la nature des risques » pour garantir la confidentialité et l'intégrité des données.

### 2.1. Sécurisation des données en entrée et en sortie

La protection des données, qu'il s'agisse de leur collecte ou de leur restitution, est un pilier central de la sécurité des systèmes d'Intelligence Artificielle (IA). Selon l'ENISA (2020), « la mise en œuvre de techniques de chiffrement fortes et la minimisation des données collectées constituent des prérequis majeurs pour tout déploiement sensible d'IA ». Cette recommandation rejoint l'avis de la CNIL (2017) soulignant « la nécessité d'adapter le niveau de sécurité à la sensibilité du traitement ».

#### *A. Principes fondamentaux*

- **Chiffrement** : utiliser des protocoles et algorithmes reconnus (TLS 1.2 ou 1.3 pour le transport, AES-256 pour le stockage). Les clés doivent être stockées de manière sécurisée et faire l'objet d'une rotation régulière.
- **Pseudonymisation et anonymisation** : dissocier les données identifiantes du reste des informations utilisées par le modèle afin de limiter l'impact en cas de compromission.
- **Minimisation des données** : ne collecter et ne conserver que les informations strictement nécessaires au fonctionnement du modèle IA, conformément aux principes de la CNIL.

#### *B. Mise en œuvre technique*

- **Contrôle des flux** : déployer des API gateways et des pares-feux applicatifs pour superviser et filtrer les échanges de données en entrée et en sortie.
- **Gestion des clés** : intégrer des solutions centralisées ou matériellement sécurisées (HSM, Hardware Security Module) afin de protéger l'accès aux clés de chiffrement.

- **Journalisation** : conserver des logs détaillés de chaque transfert de données, avec des mécanismes de détection d'anomalies (SIEM) pour alerter en cas de comportement suspect.

### *C. Bonnes pratiques organisationnelles*

- ✓ **Formation et sensibilisation** : assurer une montée en compétences des équipes techniques et métiers sur la sécurisation des flux de données.
- ✓ **Politiques de gouvernance** : définir clairement qui est responsable de la collecte, du stockage et de la suppression des données, dans le respect des contraintes légales (RGPD).
- ✓ **Audits réguliers** : réaliser des revues périodiques (internes ou externes) pour vérifier la conformité et l'efficacité des dispositifs en place (chiffrement, anonymisation).
- ✓ **Évolutivité des solutions** : réviser régulièrement les protocoles de chiffrement et se tenir informé des nouvelles failles découvertes (mises à jour logicielles, patch management).
- ✓ **Stress tests et tests d'intrusion** : simuler des attaques ciblant les entrées et sorties de données pour identifier les vulnérabilités.
- ✓ **Adaptation aux évolutions réglementaires** : anticiper les nouvelles exigences, notamment en matière de chiffrement post-quantique ou de certifications de conformité européennes.

La sécurisation des données en entrée et en sortie est un facteur-clé pour préserver la confiance des utilisateurs et la fiabilité des modèles IA. Elle repose sur une combinaison de mécanismes techniques (chiffrement, contrôle des flux, journalisation) et de mesures organisationnelles (gouvernance, formation, audits). En mettant en œuvre ces bonnes pratiques, les entreprises réduisent drastiquement leurs risques de fuite, de corruption ou de manipulation de données, tout en restant conformes aux exigences réglementaires et éthiques.

---

***Exemple concret : Chatbot bancaire***

*Dans le domaine bancaire, la sécurisation des données client est critique. Un chatbot IA doit chiffrer toutes les communications (TLS) et anonymiser partiellement les informations financières avant leur traitement par le modèle. Les logs d'accès et*

*d'utilisation sont scrupuleusement analysés via un SIEM pour détecter d'éventuelles anomalies (tentatives de scraping, requêtes massives suspectes, etc.).*

### 2.2. Contrôles d'accès et authentification

La maîtrise des accès aux données et aux modèles d'Intelligence Artificielle (IA) constitue un levier essentiel de sécurité. D'après l'ENISA (2020), « l'identification et la gestion des privilèges sont cruciales pour empêcher tout accès non autorisé aux données et aux modèles IA ». La CNIL (2017) souligne, quant à elle, « l'importance d'un pilotage rigoureux des droits et habilitations, associé à une traçabilité complète des accès ».

#### *A. Principes fondamentaux*

- **Gestion centralisée des identités (IAM)** : regrouper la création, la révocation et la supervision des comptes utilisateurs dans un outil unique, assurant cohérence et traçabilité.
- **Authentification forte (MFA, tokens, certificats)** : renforcer la simple combinaison identifiant/mot de passe par un ou plusieurs facteurs supplémentaires (code SMS, token matériel, certificat numérique).
- **Principe de moindre privilège** : chaque profil ne doit disposer que des droits strictement nécessaires à l'accomplissement de sa mission (Least Privilege Principle).

#### *B. Mise en œuvre technique*

- **Rôles et permissions** : définir des rôles (administrateur, développeur, data scientist, auditeur) et restreindre les opérations (lecture, écriture, exécution) selon les besoins réels.
- **SSO et fédération d'identités** : limiter la multiplication des comptes et mots de passe via des solutions de Single Sign-On (SSO) ou d'authentification fédérée (ex. OAuth, SAML), facilitant aussi la révocation en cas de départ d'un collaborateur.
- **Journalisation et analyse** : enregistrer systématiquement les tentatives d'authentification, les modifications de rôles et les accès aux ressources critiques. Un SIEM (Security Information and Event Management) peut corréler ces événements et émettre des alertes en cas de comportement anormal.

#### *C. Bonnes pratiques organisationnelles*

✓ **Revue périodique des comptes** : vérifier régulièrement (trimestriellement ou semestriellement) si les droits attribués sont toujours appropriés et si des comptes inactifs doivent être supprimés.

- ✓ **Sensibilisation des équipes** : former les utilisateurs à la gestion des mots de passe, à l'activation systématique de la MFA et aux risques de phishing.
- ✓ **Procédures d'escalade** : définir un plan clair en cas d'alerte (tentatives multiples de connexion, accès depuis une zone géographique inhabituelle), incluant la possibilité de bloquer rapidement un compte suspect.
- ✓ **Audits et tests d'intrusion** : réaliser régulièrement des pentests ciblant la compromission de comptes (attaques par force brute, phishing) afin de mesurer la robustesse de l'authentification et des contrôles d'accès.
- ✓ **Mises à jour et ajustements** : suivre les évolutions techniques et normatives (nouveaux protocoles d'authentification, principes Zero Trust) et adapter les politiques de sécurité en conséquence.
- ✓ **Gestion des accès d'urgence** : prévoir un mécanisme de récupération ou d'accès temporaire en cas de défaillance d'un système critique (perte de clés, panne de l'outil MFA), sous contrôle d'un processus validé.

La mise en place de contrôles d'accès rigoureux et d'une authentification robuste constitue un socle incontournable pour la sécurité des systèmes IA. Une gestion centralisée des identités (IAM), l'adoption de mécanismes de MFA et l'application stricte du principe de moindre privilège permettent de limiter drastiquement les risques d'intrusion et de compromission. Associées à une culture de la sécurité et à des processus de surveillance en continu, ces mesures renforcent la confiance dans les environnements de développement et de production, tout en répondant aux exigences réglementaires (RGPD) et aux recommandations des autorités (ENISA, CNIL).

---

***Exemple concret : Plateforme IA en mode SaaS***

*Une société de conseil propose une plateforme d'analyse prédictive accessible à ses clients. Elle utilise un IAM unifié pour gérer les identités clients et internes, active la MFA pour les rôles sensibles (administrateur, data scientist) et surveille en continu les tentatives d'accès. En cas de comportement anormal (grand nombre d'échecs de connexion, accès depuis un pays inhabituel), le SIEM génère une alerte, et un protocole d'escalade est déclenché (vérification du compte, possible suspension préventive).*

---

## 2.3. Détection et réponse aux incidents (SOC, SIEM, IDS/IPS)

L'adoption de l'Intelligence Artificielle (IA) accroît la complexité des infrastructures informatiques et expose à de nouvelles formes de menaces. L'ENISA (2020) souligne « l'importance de disposer d'une visibilité en temps réel sur l'activité des systèmes IA, pour détecter rapidement les anomalies et répondre efficacement aux incidents ». Pour y parvenir, la mise en place d'un SOC (Security Operations Center), l'usage d'un SIEM (Security Information and Event Management) et le déploiement de systèmes de détection et/ou de prévention d'intrusions (IDS/IPS) sont devenus des piliers incontournables.

### *A. Objectifs de la détection et de la réponse aux incidents*

- **Continuité de service** : face aux attaques (DDoS, data poisoning, intrusions), il s'agit de minimiser les interruptions de fonctionnement des modèles IA.
- **Protection des données et du modèle** : en détectant les signaux faibles ou les anomalies dans les logs, les équipes peuvent rapidement circonscrire une tentative d'exfiltration ou de modification malveillante du modèle.
- **Limitation de l'impact et rétroaction** : chaque incident traité permet d'améliorer en continu le dispositif, notamment via des retours d'expérience (post-mortem).

### *B. Mise en place d'un SOC dédié ou mutualisé*

- **Structure et organisation** : un SOC comprend généralement des analystes sécurité, un responsable de la coordination des incidents et des outils de supervision centralisés.
- **In-house ou MSSP** : selon la taille de l'organisation, le SOC peut être internalisé (équipe dédiée) ou externalisé (prestataire spécialisé).
- **Intégration spécifique IA** : les logs d'entraînement, de performance et d'inférence doivent être pris en compte pour pouvoir détecter des anomalies propres à l'activité IA (dégradation subite du modèle, pic de requêtes suspectes).

### *C. Utilisation du SIEM pour la centralisation et l'analyse*

- **Collecte des événements** : le SIEM agrège les logs provenant du réseau, des serveurs, des applications (dont les modèles IA) et des IDS/IPS.

- **Corrélation et détection** : grâce à des règles préétablies et éventuellement des algorithmes d'apprentissage, le SIEM repère les schémas anormaux (trop de requêtes d'une même IP, modifications soudaines de configuration).
- **Alertes et tableaux de bord** : les analystes du SOC reçoivent des alertes temps réel et disposent d'indicateurs clés pour prioriser leur intervention (ex. taux de faux positifs du modèle IA).

### *D. Rôle des IDS/IPS dans l'écosystème IA*

- Détection des attaques : les IDS surveillent les flux et comparent le trafic aux signatures d'attaques connues ou à des comportements suspects (tentatives d'injection, commandes malveillantes, etc.).
- Prévention et blocage : certains systèmes (IPS) peuvent agir immédiatement (fermer une connexion, bloquer une IP).
- Enrichissement du SOC : en remontant des alertes au SIEM, l'IDS/IPS contribue à la vision d'ensemble des menaces et alimente la base de connaissances.

### *E. Gestion des incidents et plan de réponse*

- Identification et confinement : en cas d'alerte, l'objectif est de localiser rapidement l'incident et d'isoler les composants touchés (serveur IA, dataset sensible).
- Éradication et rétablissement : après avoir évité la propagation, il convient de supprimer la cause de l'incident (ex. accès malveillant) et de restaurer le système dans un état sûr (rollback du modèle IA).
- Rétroaction : un compte-rendu complet permet d'ajuster les règles de détection, de renforcer la configuration, voire de préciser les procédures légales (notification aux autorités compétentes en cas de fuite de données).

La détection et la réponse aux incidents forment un élément central de la cybersécurité appliquée à l'IA. L'association d'un SOC, d'un SIEM et d'IDS/IPS, adaptée aux spécificités du cycle de vie des modèles (phase d'entraînement, inférence, mises à jour), offre une vision unifiée des menaces et une capacité de réaction rapide. Cette approche permet de renforcer la confiance dans les systèmes IA et d'honorer les

obligations légales et réglementaires (CNIL, RGPD, ENISA). Par ailleurs, l'analyse post-incident demeure primordiale pour améliorer sans cesse les politiques de sécurité et anticiper l'évolution des attaques ciblant l'IA.

---

***Exemple : Détection d'une exfiltration de données IA***

*Lors d'un entraînement de nuit, un SIEM détecte un volume anormalement élevé de données sortantes vers une adresse IP étrangère. Les équipes SOC reçoivent une alerte critique. En enquêtant, elles constatent qu'une clé d'API compromise était utilisée pour accéder aux logs et extraire un dataset sensible. L'IDS, paramétré pour bloquer toute connexion non autorisée vers certaines destinations, intervient et coupe immédiatement l'accès. Le SOC confine le serveur, mène une analyse forensique et enclenche le plan de réponse.*

---

## 2.4. Sécurisation de la chaîne de développement (DevSecOps)

La mise en œuvre de l'IA au sein des organisations requiert une approche DevSecOps adaptée, intégrant la sécurité et la conformité dès les premières étapes du cycle de vie d'un projet. L'ENISA (2020) souligne « la nécessité d'intégrer la sécurité tout au long du cycle de développement de l'IA, afin d'anticiper et de prévenir les vulnérabilités », tandis que la CNIL (2017) appelle à « inclure la réflexion éthique et réglementaire dès la phase de conception algorithmique ».

### *A. Principes DevSecOps appliqués à l'IA*

- **Collaboration interdisciplinaire** : les équipes de développement, de data science, de sécurité et juridiques doivent travailler de concert pour définir des objectifs communs (conformité RGPD, intégrité du modèle, respect de la vie privée).
- **Sécurité intégrée à chaque étape** : Secure by design implique la prise en compte des menaces potentielles, l'analyse de risques (ISO/IEC 27005) et l'adaptation de l'architecture pour limiter la surface d'attaque.
- **Automatisation** : l'utilisation d'outils de scan (analyse de dépendances, tests d'intrusion automatisés) et de pipelines CI/CD évite que des vulnérabilités ne passent inaperçues d'une version à l'autre.

### *B. Tests d'intrusion et audits de code*

- **Pentests spécialisés IA** : tenter d'injecter des données corrompues ou de contourner les validations logiques pour évaluer la résilience du modèle aux attaques (data poisoning, model hacking).
- **Revue de code** : instaurer un processus systématique de peer review et s'appuyer sur des outils d'analyse statique/dynamique (SAST/DAST) afin de déceler rapidement les failles courantes (injections, dépendances obsolètes).
- **Audit continu** : programmer des audits réguliers, par des ressources internes ou externes, pour valider la pertinence et l'efficacité des mécanismes de sécurité.

### C. *Validation continue et déploiement sécurisé (CI/CD)*

- **Automatisation des builds** : chaque modification du code (ou du modèle) déclenche automatiquement une compilation, des tests fonctionnels, des scans de sécurité et un déploiement éventuel en environnement de test.
- **Gates de sécurité** : si une faille majeure est détectée (bibliothèque vulnérable, test d'injection échoué), le processus de déploiement s'arrête et requiert une action humaine avant de reprendre.
- **Séparation des environnements** : cloisonner les environnements de développement, d'intégration et de production, assurant que les données réelles ne soient jamais exposées dans un environnement trop permissif.

### D. *Approche "Secure by Design"*

- Anticipation des exigences légales : identification des obligations RGPD, des nécessités de minimisation et d'anonymisation des données dès le départ.
- Chiffrement et gestion des secrets : mise en place de solutions de coffre-fort (vault) pour stocker les clés, les identifiants et les jetons API, réduisant le risque de fuite ou de compromission.
- Architecture modulaire : segmenter la solution IA (bases de données, microservices, front-end) pour réduire l'impact d'une éventuelle intrusion et simplifier la traçabilité.

La sécurisation de la chaîne de développement (DevSecOps) est un levier essentiel pour prévenir les vulnérabilités et renforcer la confiance dans les solutions IA. En combinant une culture de la sécurité, des processus outillés (CI/CD automatisée, tests d'intrusion récurrents, audits de code) et une approche "Secure by Design", les organisations peuvent réduire significativement les risques et assurer la fiabilité et la conformité de leurs systèmes IA. Cette démarche, encouragée tant par l'ENISA que par la CNIL, s'inscrit dans une vision proactive où l'innovation technologique va de pair avec une protection rigoureuse des données et des algorithmes.

***Exemple : Plateforme d'apprentissage automatique en microservices***

*Développement : chaque microservice (prétraitement des données, entraînement du modèle, exposition de l'API) fait l'objet d'une revue de code approfondie et de scans de sécurité automatisés.*

*Tests d'intrusion : des attaques ciblées (simulation de bruteforce sur l'API, injection de faux échantillons) sont réalisées à chaque itération majeure pour éprouver la robustesse du système.*

*Déploiement : la pipeline CI/CD déploie en production uniquement si tous les contrôles de sécurité sont validés. Les logs et alertes sont envoyés à un SIEM pour une surveillance en temps réel.*

## 2.5. Approche de gestion des risques

La mise en place d'une démarche structurée de gestion des risques est indispensable pour prévenir et atténuer les menaces pesant sur les projets d'Intelligence Artificielle (IA). L'ENISA (2020) préconise « l'utilisation de cadres de gestion des risques existants (ISO/IEC 27005, EBIOS) pour identifier et prioriser les vulnérabilités spécifiques à l'IA ». La CNIL (2017) ajoute « qu'une démarche systématique de gestion des risques doit prendre en compte la sensibilité des données manipulées par l'IA, ainsi que les conséquences d'un biais ou d'une altération du modèle ».

### *A. Méthodologies de classification et d'évaluation des risques*

- **ISO/IEC 27005** : propose un cadre d'analyse de risques (identification, estimation, traitement, acceptation) adapté aux systèmes d'information, pouvant s'appliquer aux particularités de l'IA (algorithmes, données massives).
- **EBIOS Risk Manager** : développé par l'ANSSI, met l'accent sur l'évaluation des menaces, l'expression des besoins et la définition d'objectifs de sécurité clairs.
- **NIST SP 800-53** : fournit un catalogue de contrôles de sécurité et de confidentialité pouvant s'inscrire dans une politique de gestion des risques IA plus large.

### *B. Cartographie des vulnérabilités*

- **Identification des actifs critiques** : modèles IA, données d'entraînement, interfaces d'inférence (API), pipeline CI/CD.
- **Évaluation des menaces** : scénarios de corruption des données (data poisoning), exfiltration de données sensibles, altération des hyperparamètres, accès non autorisé aux ressources cloud.
- **Impact et probabilité** : estimer le potentiel dommage (financier, réputationnel, opérationnel) et la vraisemblance d'occurrence.

### *C. Plan de remédiation*

✓ **Priorisation** : concentrer d'abord les ressources sur les risques les plus élevés (haute probabilité, fort impact), puis traiter les risques moyens ou résiduels.

- ✓ **Mesures correctives** : déployer des correctifs techniques (patching, durcissement des accès), renforcer l'authentification et la traçabilité, former les équipes aux bonnes pratiques.
- ✓ **Suivi et gouvernance** : nommer des responsables pour mettre en œuvre et vérifier l'efficacité de chaque action, mettre à jour la documentation (cartographie, plans d'actions).
- ✓ **Révision régulière** : la cartographie des risques doit être réévaluée périodiquement, notamment après un incident ou une évolution technologique.
- ✓ **Retour d'expérience** : chaque incident, même mineur, constitue l'occasion d'améliorer les contrôles et la posture de sécurité.
- ✓ **Veille et anticipation** : suivre les publications (ENISA, ANSSI, CNIL, NIST) et la recherche académique sur les attaques émergentes visant l'IA (techniques adversariales, nouveaux vecteurs d'intrusion).

Une approche de gestion des risques solide, fondée sur des référentiels éprouvés (ISO/IEC 27005, EBIOS, NIST SP 800-53), est cruciale pour sécuriser les projets IA et anticiper les menaces. La cartographie des vulnérabilités, la priorisation des actions, le plan de remédiation et le cycle d'amélioration continue constituent les piliers d'une démarche proactive et collaborative, associant experts techniques, juristes et décisionnaires métiers. Au-delà de la simple conformité réglementaire, cet effort coordonné renforce la résilience et la fiabilité des systèmes IA, tout en préservant la confiance des utilisateurs et des partenaires.

***Exemple : Chatbot bancaire***

*Un chatbot servant à gérer des requêtes financières sensibles se trouve exposé à divers risques : vol de données personnelles, accès non autorisé aux fonctionnalités de virement, injection de données malveillantes dans le modèle. L'analyse de risques (basée sur ISO/IEC 27005 ou EBIOS) permet de cartographier les vulnérabilités (mauvaise configuration de l'API, faible authentification pour les clients, absence de logs). Le plan de remédiation comprend alors l'implémentation de l'authentification forte, le chiffrement systématique des flux et la mise en place d'un SOC pour la détection des incidents.*

# PARTIE 3

# BONNES PRATIQUES ET PRÉCONISATIONS

# RECOMMANDATIONS DES INSTANCES PROFESSIONNELLES

## 1. Conseil National de l'Ordre des Experts-Comptables

Dans une publication « Comment utiliser ChatGPT® ? », publiée en octobre 2024, le Conseil National de l'Ordre des Experts-Comptables a préconisé les actions suivantes :

« Ne chargez pas de données personnelles, sensibles et/ou confidentielles : pas de mails clients, de FEC, de DSN… non anonymisés dans des sites non maîtrisés (respect du RGPD et du secret professionnel) »[38].

## 2. Compagnie nationale des commissaires aux comptes (CNCC)

***Un retour sur l'ouvrage « Repenser la Chaîne de Confiance à l'ère de l'Intelligence Artificielle »***

***Par Serge Yablonsky***

### 2.1. La révolution IA, un séisme pas si silencieux

ChatGPT, Midjourney, DALL-E… L'IA générative a débarqué dans notre quotidien sans crier gare. En quelques mois, les discussions autour de l'intelligence artificielle sont passées de la science-fiction à la salle de réunion. Les entreprises sont sur le pont : opportunités de productivité, réduction des coûts, nouveaux business models... L'IA semble tout réinventer.

Mais ce grand saut technologique vient aussi chambouler la façon dont on construit la confiance : entre collaborateurs, vis-à-vis des clients, des régulateurs, et plus largement de la société. C'est le point de départ de ce rapport de la CNCC, que nous avons voulu synthétiser ici pour partager une vision claire, équilibrée et ancrée dans le réel.

[38] Page 7 dudit document

### 2.2. Ce que cache vraiment l'IA (et ce qu'elle ne peut pas faire)

L'IA, ce n'est pas magique. C'est une accumulation de techniques : *machine learning* (les machines apprennent des données), *deep learning* (elles simulent les réseaux de neurones), *large langage models* - LLM (des modèles capables de générer du texte).

Mais attention : les IA ne comprennent pas. Elles probabilisent. Et c'est bien là que le bât blesse : biais culturels, hallucinations (réponses fausses mais plausibles), opacité des modèles... Derrière le vernis techno, l'IA reste une « boite noire ».

Et pourtant, on la retrouve partout. Pourquoi ? Parce qu'elle sait traiter des volumes de données gigantesques, et repérer des signaux faibles là où l'humain fatiguerait.

### 2.3. Entreprises, attention au grand plongeon

Oui, l'IA est puissante. Mais encore faut-il savoir où, quand, comment l'utiliser. Car tous les métiers ne sont pas également exposés. Dans les fonctions support (RH, juridique, compta), l'IA peut automatiser beaucoup. Côté relation client ou production, elle ouvre des perspectives intéressantes mais pose aussi des questions éthiques.

Dans le monde de l'audit, on voit déjà la différence : analyse des données massives, automatisation de tests, détection d'anomalies. Mais attention à ne pas confondre outil et jugement professionnel. L'IA est un levier, pas un pilote.

### 2.4. Gagner la guerre de la confiance

Ce que l'IA bouleverse le plus, c'est la chaîne de valeur. Qui détient les données ? Qui fournit les modèles ? Qui détient l'infrastructure ? Aujourd'hui, les géants du numérique occupent toutes les couches : de la puce à l'application. Un risque de dépendance fort pour les entreprises européennes.

Derrière cette compétition se cache une coopétition : on collabore et on concurrence les mêmes acteurs. La question de la souveraineté technologique devient clé. Et avec elle, celle de la confiance dans les systèmes. Transparence, robustesse, explication : voilà les nouveaux KPI.

### 2.5. Risques vs. opportunités : le funambule sur la ligne de code

Le rapport propose une approche IRO (Incidences, Risques, Opportunités). Ce n'est pas un simple tableau : c'est une grille de lecture concrète pour anticiper les effets de l'IA. Salariés, clients, investisseurs, environnement... Tous sont concernés.

Quelques exemples :

- Opportunité : automatiser les tâches répétitives, créer des services personnalisés.
- Risque : biais discriminants, perte de diversité, augmentation des cyberattaques.
- Incidences : nouvelles compétences à développer, gestion de la réputation, dépendance aux fournisseurs techno.

Bref : pas de révolution sans méthode.

### 2.6. Vers une IA responsable… avec des humains aux commandes

Pas de panique : on peut encadrer l'IA. De nombreux acteurs proposent des cadres pour en faire un usage responsable. Données traçables, modèles explicables, supervision humaine, compétences internes, cybersécurité, transparence...

La CNCC s'appuie sur une grille inspirée du NIST américain et du cabinet FTI Consulting. Dix piliers d'une gouvernance IA éthique et robuste. Le message est clair : pas de déploiement sans gouvernance. Pas de transformation sans accompagnement.

Pour conclure, une idée forte : l'IA est un outil. Puissant, certes. Mais encore faut-il lui fixer un cap. Et ce cap, c'est à nous, humains, de le choisir. En conscience. Avec exigence. Et avec confiance.

## 3. CRCC de Paris

Lors d'un webinaire du Lab 50, organisé par la CRCC de Paris, Maud Bodin-Veraldi, vice-présidente de la CRCC de Paris a rappelé que la documentation de la CNCC, qui n'est pas publique, est réservée aux membres de cette instance et ne doit pas être communiquée à des IA génératives[39].

## 4. Conseil National du Barreau des avocats

Un guide du Conseil national du barreau[40] prévoit les dispositions suivantes :

« Il est impératif pour l'avocat de veiller à ne jamais communiquer des données relatives à ses dossiers ou à ses clients à des intelligences artificielles génératives. »

« Afin d'utiliser des outils d'intelligence artificielle générative tout en protégeant ses données, il est conseillé à l'avocat de mettre en œuvre une bonne pratique : la pseudonymisation des données. »

« L'utilisation de l'intelligence artificielle générative doit être conforme au Règlement général sur la protection des données (RGPD), qui impose des obligations strictes en matière de traitement des données personnelles »

[39] https://www.youtube.com/watch?v=kL5r0FS7sg0
[40] *Utilisation des systèmes d'IA générative*, guide pratique, première édition, septembre 2024, https://encyclopedie.avocat.fr/GEIDEFile/CNB_GT_IntelligenceArtificielle_2024.pdf?Archive=132021395020&File=Telecharger_le_guide_ici&verif=480312480317473152469037480312450538469018469029488825481498

# SE PROTEGER AU “QUOTIDIEN” : COMMENT ?

## 1. Principales bonnes pratiques

*Par Romain Mirable et Sylvain Navers*

De manière liminaire, le cabinet d'expertise comptable qui souhaite utiliser des systèmes d'IAGen doit adopter une approche rigoureuse pour protéger les données sensibles. Cela implique :

- d’examiner attentivement les conditions générales du système choisi afin de comprendre la manière dont les informations saisies seront utilisées ;
- de privilégier des solutions d’IAGen garantissant que les données ne seront ni conservées ni exploitées à d’autres fins que celles strictement nécessaires au cabinet, sans aucun partage avec des tiers ;
- d’anonymiser systématiquement toutes les informations pouvant identifier les clients ou leurs dossiers spécifiques. Aucune information concernant un client ne doit être transmise, sans l’accord explicite de ce dernier ;
- de veiller à ne jamais transmettre des données sensibles ou stratégiques, telles que celles liées aux situations financières confidentielles des clients ;
- de ne pas envoyer de document complet pour analyse. A minima vérifier la confidentialité des informations contenues dans le document et expurger les informations non souhaitées ;
- de sélectionner soigneusement une IAGen de meilleure confiance, à savoir Française ou Européenne. Si l’IAGen est installée en local sur l’ordinateur, configurer afin de restreindre les IAGen utilisées ;
- de ne pas sauvegarder les discussions afin de limiter l’analyse a posteriori, bien que cela puisse limiter la pertinence des réponses ;
- de ne pas utiliser de *prompt* généré par autrui sans vérification préalable afin d’en garantir l’innocuité.

**➕ Annexe 1 – Check-list d’audit technique et juridique pour se protéger quotidiennement face à l’usage de l’IA**

## 2. La charte informatique : un outil incontournable pour les cabinets

*Par Sabrina Agrapart*

Chaque cabinet d'expertise comptable devrait adopter une charte informatique incluant une section dédiée à l'utilisation des outils d'intelligence artificielle, et s'assurer que cette charte soit opposable à l'ensemble des collaborateurs[41]. Cette approche permettrait de maximiser les bénéfices de ces technologies tout en minimisant les risques liés à la sécurité des données et à la conformité réglementaire.

**+ Annexe 2 - EXEMPLE DE CHARTE D'UTILISATION DE L'INTELLIGENCE ARTIFICIELLE GENERATIVE**

+ Voir aussi : *Charte interne relative à l'IA générative*, Ministère de la culture, juin 2024,

https://semaphore.culture.gouv.fr/system/files/database/documents/2024-06/charte-intelligence-artificielle-generative-ministere-culture.pdf

+ Voir aussi : Exemple de Charte sur le site du CNOEC (espace privé)

## 3. Former pour prévenir : la sensibilisation des collaborateurs aux enjeux de l'IAGen

*Par Sabrina Agrapart*

Il est primordial de consacrer du temps à informer et sensibiliser les collaborateurs sur le fonctionnement des IAGen et les risques associés. D'ailleurs, cette obligation de formation se retrouve dans l'article 4 de

[41] Un modèle de clause relative à l'utilisation de l'intelligence artificielle est proposé en annexe du présent document à titre indicatif

l'IA Act pour les fournisseurs et déployeurs[42] d'IA. Un accompagnement rigoureux doit donc être mis en place pour garantir que l'utilisation de ces technologies est conforme aux objectifs et aux valeurs de l'entreprise, tout en respectant toujours les principes éthiques tout en tenant compte de (i) leurs expériences techniques, (ii) leurs éducations et (iii) leurs formations ainsi que (iv) de l'utilisation envisagée.

### 3.1. Risques liés à la confidentialité et à la protection des données personnelles

Lorsqu'un utilisateur transmet des données à une IAGen, il existe un risque que ces informations soient réutilisées. Par conséquent, les collaborateurs doivent impérativement s'abstenir de saisir des données à caractère personnel ou des informations couvertes par le secret professionnel dans leurs **instructions génératives**, si la légalité de leur traitement ne peut être garantie.

*A titre d'exemple, le collaborateur ne doit jamais :*

- *Inclure dans **son instruction générative** des informations personnelles telles que des coordonnées, des identités ou toute donnée relative à la vie privée,*
- *Fournir des données contractuelles, juridiques ou financières (c.a.d. confidentielles).*

Pour utiliser efficacement les outils d'IAGen tout en préservant la confidentialité des données, il est recommandé aux cabinets d'expertise comptable d'adopter la pseudonymisation[43] des données

Pour pseudonymiser efficacement :

- Remplacer les informations personnelles identifiables par des pseudonymes ou des données génériques.
- Utiliser des initiales fictives ou des suites de caractères complexes en lieu et place des noms.

Il est important de rester attentif aux possibilités d'identification par recoupement de données, même pseudonymisées. Des éléments tels que les données de comptes annuels (chiffre d'affaires, adresses ...) peuvent potentiellement permettre l'identification d'un client.

[42] Article 3 de l'IA Act, définition de « déployeur » : une personne physique ou morale, une autorité publique, une agence ou un autre organisme utilisant sous sa propre autorité un système d'IA sauf lorsque ce système est utilisé dans le cadre d'une activité personnelle à caractère non professionnel;

[43] La pseudonymisation est un processus de traitement des données personnelles qui consiste à remplacer les identifiants directs (noms, prénoms, etc.) par des identifiants indirects. Contrairement à l'anonymisation, ce procédé est réversible, permettant ainsi l'utilisation des données sans identification immédiate

### 3.2. Former les collaborateurs aux instructions génératives et à contrôler les résultats des IAGen

Il est également important (i) de former les collaborateurs à l'art de formuler des instructions génératives efficaces pour optimiser l'utilisation des IAGen et (ii) de les sensibiliser au contrôle des réponses apportées.
Une bonne maîtrise de cette compétence permet non seulement d'obtenir des résultats plus pertinents, mais aussi de minimiser les risques liés à la confidentialité et à la protection des données et ou aux erreurs et hallucinations de ces IAGen.

Lorsque vous utilisez une intelligence artificielle générative (IAGen), vous devez jouer le rôle de vérificateur et validateur pour détecter les erreurs, appelées "hallucinations". Parfois, l'IA peut générer du contenu qui semble vrai mais est en réalité incorrect. A ce titre, il est recommandé d'utiliser des systèmes d'IAGen qui indiquent leurs sources, mais cela ne garantit pas l'absence d'erreurs.

Pourquoi le contenu peut être faux : les IAGen génère des réponses en se basant sur des modèles probabilistes. Ces modèles sont entraînés sur d'énormes volumes de données textuelles pour prédire le mot ou la phrase le plus probable qui suit une requête donnée. Chaque mot est sélectionné selon sa probabilité statistique de correspondre au contexte, plutôt qu'à partir d'une certitude absolue. Ainsi si le jeu de données indique la réponse la plus probable est A, bien qu'elle soit fausse, la réponse communiquée sera A.

A titre d'information, les IAGen, par défaut, ne réalisent pas de calcul et il convient donc d'être très attentif si vous demandez un quelconque calcul à ces IAGen qui vous trouveront simplement la réponse la plus probable par rapport à sa base de données.

## ETAPE CLES POUR REDIGER UNE INSTRUCTION GENERATIVE EFFICACE

*Avant d'utiliser une IAGen, il est essentiel de tenir compte de la nature de l'IA que vous utilisez lors de la rédaction de votre instruction générative. Les systèmes d'IA peuvent fournir des résultats différents en fonction des données et de leur conception. Pour des tâches générales, comme la rédaction d'un courriel professionnel, une IA généraliste peut suffire. Cependant, il est impératif de ne jamais inclure de données personnelles ou d'informations soumises au secret des affaires dans vos instructions génératives, même pour ces tâches simples. Pour des domaines spécialisés tels que la comptabilité, la fiscalité ou le droit, il est fortement recommandé d'utiliser une IA spécifiquement formée dans ces domaines. Cette précaution est essentielle pour garantir la précision et la pertinence des informations obtenues, particulièrement lorsqu'il s'agit de conseils techniques ou de recherches spécialisées. Assurez-vous donc de choisir l'outil d'IA approprié en fonction de la nature et de la complexité de votre instruction générative, et gardez toujours un œil critique sur les réponses générées, en les vérifiant auprès de sources faisant autorité si nécessaire.*

1. **Définir le rôle de l'IA** : attribuez un rôle spécifique à l'IA pour orienter ses réponses
   *Exemple : agis en tant qu'expert-comptable spécialisé dans la fiscalité des PME*

2. **Préciser l'objectif** : énoncez clairement le but de votre instruction générative en fonction du type de système d'IA que vous utilisez.
   *Exemple : l'objectif est d'expliquer les changements fiscaux récents affectant les PME.*

3. **Définir la mission :** décrivez la tâche spécifique à accomplir**.**

   *Exemple : Rédige un résumé concis des changements majeurs de la loi de finances 2024 affectant les PME, en mettant en avant les points fiscaux"*

4. **Soyez clair et concis** : évitez les phrases longues et complexes qui pourraient entraîner des réponses moins pertinentes. Si votre instruction générative contient plusieurs éléments divisez-la en plusieurs questions pour plus de clarté. Formulez votre instruction générative en commençant par un verbe d'action qui décrit précisément la tâche à réaliser.
   ***E**xemple : "Analyser", "Calculer", "Comparer", "Résumer"*).

5. **Spécifier le format souhaité :** Indiquez la structure ou le type de réponse attendu**.** Souhaitez-vous un paragraphe court ou développé, un langage technique ou au contraire vulgarisé, un ton formel ou informel.
   *Exemple : présente l'information sous forme de liste à puces, avec un maximum de 5 points clés.*

6. **Fournir le contexte** : donnez des informations contextuelles pertinentes.
   *Exemple : ce résumé sera utilisé lors d'une présentation à des dirigeants de PME du secteur manufacturier."*

7. **Définir la tonalité** *:* indiquez le style ou le ton à adopter.
   *Exemple : **U**tilise un ton professionnel mais accessible, adapté à un public non-expert en fiscalité.*

8. **Sources ou références :** indiquez si vous souhaitez des références spécifiques ou des sources d'information.
   *Exemple : cite les articles de loi pertinents pour chaque point mentionné.*

9. **Itération et ajustement :** après avoir obtenu une première réponse de l'IAGen, vous évaluez si elle répond pleinement à vos attentes. Si la réponse n'est pas tout à fait satisfaisante, vous modifiez légèrement votre instruction générative initiale pour obtenir un meilleur résultat. Même un bon prompt, ne vous empêchera pas de devoir poser quelques questions avant d'obtenir une réponse satisfaisante. A noter que les IAGen proposent, selon les modèles, des sauvegardes des requêtes et instructions, vous ne devez donc pas repartir de zéro à chaque échange !

➕ Voir aussi le cahier n°41 de l'Académie qui contient une méthodologie de rédaction de *prompts* et des exemples d'application.

➕ Voir aussi la vidéo « Le *prompt parfait* » du groupe de travail n°63 de L'Académie : https://www.youtube.com/watch?v=jIG2UMiLEp4

### 3.3. L'humain au cœur de la décision : les limites de l'IAGen

L’utilisation d’une IAGen ne doit pas remplacer la prise de décision humaine, ni négliger l’expertise et le raisonnement propres aux professionnels. Comme indiqué, les collaborateurs doivent être conscients que les réponses générées par l’IAGen peuvent comporter des erreurs et qu’elles doivent être vérifiées avec soin. Il est impératif que ces réponses soient validées avant d’être partagées, afin de garantir leur exactitude. Une demande des sources au moment de la requête à l’IAGen simplifiera cette tâche de vérification.

### 3.4. Propriété intellectuelle et IAGen : un défi supplémentaire

Par ailleurs, l’IAGen peut produire des contenus susceptibles de porter atteinte aux droits de propriété intellectuelle des tiers. Chaque collaborateur doit donc s’assurer que les contenus utilisés dans les instructions génératives sont libres de droits et que les informations fournies ne sont pas soumises à des droits d’auteurs.

### 3.5. Sécurité des données et conformité : l'importance de l'audit contractuel

L’utilisation d’une IAGen soulève des défis significatifs en matière de protection des données, notamment concernant les modalités d’hébergement et de manipulation des informations. Un audit approfondi des documents contractuels permet de vérifier que toutes les clauses nécessaires à la protection des données sont présentes, que les responsabilités de chaque partie sont clairement définies et que les mesures de sécurité appropriées sont mises en place. Cela garantit non seulement la conformité aux réglementations en vigueur, mais aussi la sécurisation effective des données sensibles traitées par le cabinet. Sélectionner un prestataire avec une certification ISO 27001 peut être un plus.

### 3.6. Précautions juridiques dans le choix d'une IAGen si vous envisagez de lui communiquer des informations confidentielles ou personnelles (partenaire de confiance, éditeur de logiciel du secteur ...)

Si l'utilisation de données à caractère personnel ou confidentielle est envisagée, le choix d'une IAGen nécessite une vigilance particulière sur le RGPD (accord de sous-traitance RGPD, cf. Article 28 du RGPD), notamment lorsque le fournisseur est situé en dehors de l'Espace Économique Européen (EEE).

De nombreuses solutions d'IAGen sont proposées par des fournisseurs basés hors de l'EEE, ce qui rend essentiel de vérifier le lieu d'implantation du fournisseur et de l'hébergement des données. Cela permet de déterminer si le pays concerné est considéré comme adéquat au sens de la CNIL[44].

Étant donné que beaucoup d'outils d'IAGen sont américains, il est particulièrement important de porter une attention spéciale à ces fournisseurs pour garantir la conformité avec le RGPD. La CNIL considère à ce jour les États-Unis comme un pays partiellement adéquat et précise que seuls les transferts de données personnelles vers les entités certifiées dans le cadre du Data Privacy Framework (DPF)[45] ne nécessitent pas d'encadrement par des outils de transfert supplémentaires. Il est important de consulter régulièrement cette liste, car elle est sujette à des mises à jour fréquentes.

Si le fournisseur de l'IAGen pressenti ne figure pas sur cette liste, d'autres mécanismes de transfert doivent être mis en place, tels que les clauses contractuelles types. Ces mécanismes doivent être complétés par une évaluation rigoureuse des risques liés au transfert de données et la mise en place de mesures de protection supplémentaires si nécessaire. Il est également impératif d'établir un contrat de traitement des données (Data Processing Agreement - DPA) entre le cabinet comptable français et le fournisseur américain. Ce contrat doit détailler les obligations de chaque partie en matière de protection des données. La transparence envers les personnes concernées est également importante : elles doivent être informées du transfert de leurs données vers les États-Unis et des garanties mises en place.

Si les conditions ci-dessus ne peuvent pas être respectées, il est impératif de ne pas saisir de données à caractère personnel ni de données soumises au secret des affaires dans les instructions génératives.

---

[44] liste des pays offrant un niveau de protection des données adéquat : (https://www.cnil.fr/fr/la-protection-des-donnees-dans-le-monde).
[45] Liste des organismes certifiés : Data Privacy Framework

En tout état de cause, une attention particulière devra être apportée dans ce choix d'un éditeur d'IAGen afin d'éviter tout risque de fuite de données notamment en regardant les mesures de sécurité mise en œuvre ou encore en vérifiant la réputation de cet éditeur sur le marché.

### 3.7. Choisir une IAGen : une décision dans un contexte réglementaire et technologique en pleine évolution

À l'heure où l'IAGen révolutionne nos pratiques, de nombreux cabinets envisagent d'utiliser des grands modèles de langage. Cependant, le choix d'une IAGen est une décision complexe qui nécessite une compréhension approfondie du contexte réglementaire et technologique en pleine évolution. Les cabinets d'expertise comptable doivent donc être vigilants et flexibles pour adapter leurs stratégies en fonction des nouvelles directives et évolutions technologiques.

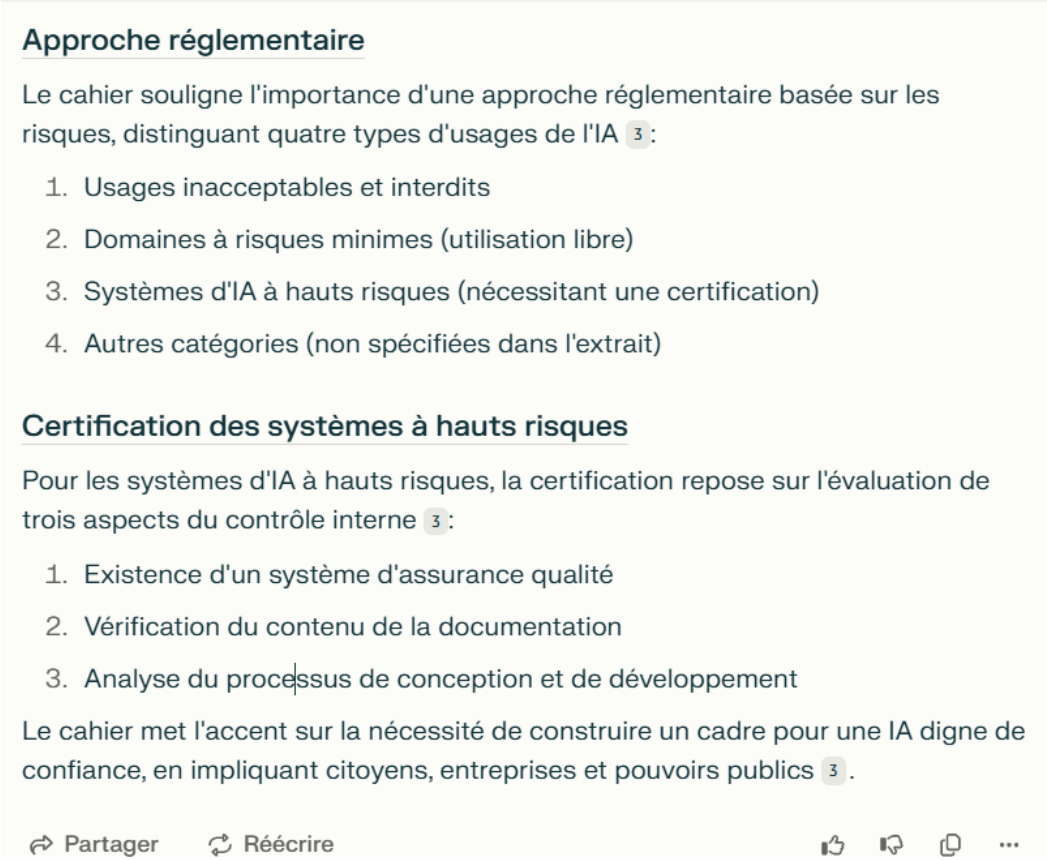
Il est également important de rappeler que la France, avec son riche patrimoine linguistique et culturel, est à l'avant-garde de ce défi technologique, où l'utilisation des LLM requiert une approche judicieuse et informée. Pour autant, la majorité des données d'entraînement de ces modèles proviennent d'anglophones, ce qui soulève des préoccupations quant à leur pertinence et représentativité pour les usages francophones. Cela met en lumière le besoin d'une reconnaissance accrue de la diversité linguistique et culturelle qui doit être intégrée dans les algorithmes pour assurer la pertinence et l'exactitude des réponses fournies.

En conclusion, l'adoption des IAGen nécessite une démarche éclairée, responsable et éthique, soutenue par une collaboration entre les professionnels, les acteurs technologiques et le milieu juridique. Il est à ce titre recommandé que les cabinets d'expertise comptable recherchent l'assistance d'experts juridiques pour guider leur transition vers un avenir où l'IA et le droit se rencontrent harmonieusement.

## 4. Protéger son site internet contre le *web scrapping*

***Par Vincent Lacomme***

La capture ci-après extraite de Perplexity, à partir du prompt « IA et confiance : la réglementation. Résumer le cahier de l'académie » reprend des éléments du cahier n°38 édité par l'Académie et disponible publiquement sur internet. Cela illustre la possibilité par l'IA de s'approprier tout contenu en ligne.



### Approche réglementaire

Le cahier souligne l'importance d'une approche réglementaire basée sur les risques, distinguant quatre types d'usages de l'IA [3] :

1. Usages inacceptables et interdits
2. Domaines à risques minimes (utilisation libre)
3. Systèmes d'IA à hauts risques (nécessitant une certification)
4. Autres catégories (non spécifiées dans l'extrait)

### Certification des systèmes à hauts risques

Pour les systèmes d'IA à hauts risques, la certification repose sur l'évaluation de trois aspects du contrôle interne [3] :

1. Existence d'un système d'assurance qualité
2. Vérification du contenu de la documentation
3. Analyse du processus de conception et de développement

Le cahier met l'accent sur la nécessité de construire un cadre pour une IA digne de confiance, en impliquant citoyens, entreprises et pouvoirs publics [3].

Partager Réécrire

En juillet 2024, une étude[46] constate une augmentation de la quantité de sites internet demandant aux robots de ne pas indexer leurs pages.

Pour protéger le contenu figurant sur vos sites internet, il est possible d'interdire l'utilisation des données du site par les modèles de langage d'IA.

Un fichier robots.txt peut contenir des instructions pour signaler aux robots ou moteurs de recherche qu'ils ne doivent pas s'approprier ces informations.

Une protection plus avancée consiste à limiter le contenu dans des espaces privés nécessitant une authentification.

Les sites d'informations utilisent la notion de « Paywalls » pour protéger leur contenu « premium » uniquement disponible sur abonnement et identification.

Une autre recommandation est d'amender les conditions générales d'utilisation du site.

Voir aussi : https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/10019984

[46] Collectif, *Consent in Crisis: The Rapid Decline of the AI Data Commons, Data provenance, j*uillet 2024, https://arxiv.org/abs/2407.14933

## 5. Pseudonymiser et anonymiser les données

***Par Vincent Lacomme***

Fabrice Heuvrard, expert-comptable et commissaire aux comptes a conçu un outil pour anonymiser un fichier des écritures comptables (FEC). Cet outil est accessible sur ces sites :

- https://anonymiser-fec.com/
- https://github.com/Fabrice-Heuvrard/FEC_Anonymiser

Il subsiste un **risque de réidentification** : même lorsque les données sont anonymisées, des techniques avancées peuvent permettre de réidentifier des individus à partir des données générées par l'IA, compromettant ainsi la confidentialité.

▶ Voir étude : Collectif, *Privacy re-identification atttacks on tabular GANs*, arXiv:2404.00696, mars 2024

➕ Voir aussi fiche CNIL sur l'anonymisation des données, https://www.cnil.fr/fr/technologies/lanonymisation-de-donnees-personnelles, 19 mai 2020

## 6. Confidentialité différentielle

***Par Vincent Lacomme***

Dans un article Dark reading[47], Michael Rinehart, vice-président de l'intelligence artificielle chez Securiti, souligne que les outils avancés d'IA générative, tels que ChatGPT, ne peuvent pas distinguer ce qu'ils doivent ou ne doivent pas mémoriser pendant leur entraînement. Ce défi pose des risques importants pour les organisations qui souhaitent exploiter ces outils pour différents cas d'usage.

Pour atténuer ces risques, il propose deux approches principales :

1. **L'utilisation de méthodes de classification, de masquage ou de tokenisation des données sensibles**, afin de minimiser les risques de fuite d'informations personnelles.
2. **Le recours à la confidentialité différentielle** : méthode qui protège les données en introduisant du "bruit" (ou erreur) dans les ensembles de données réels. Cette technique permet de produire des données synthétiques qui préservent les principales caractéristiques des données réelles tout en masquant l'identité des individus présents dans le jeu de données.

### 6.1. Qu'est-ce que la confidentialité différentielle ?

La confidentialité différentielle est une méthode mathématique conçue pour protéger les informations personnelles dans un ensemble de données. Elle introduit un bruit statistique contrôlé, garantissant que l'ajout ou le retrait d'un individu dans la base n'a pas d'impact significatif sur les résultats des analyses. Cette propriété rend impossible l'identification directe ou indirecte d'une personne à partir des données.

Ce concept a été initialement introduit par Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith dans leur étude : "Calibrating Noise to Sensitivity in Private Data Analysis" en 2006.

[47] https://www.darkreading.com/vulnerabilities-threats/samsung-engineers-sensitive-data-chatgpt-warnings-ai-use-workplace

**Quelques exemples d'applications pratiques**

- **Analyse comportementale** : Les moteurs de recherche, tels que Google, utilisent la confidentialité différentielle pour collecter des informations sur les comportements des utilisateurs. Cela permet d'améliorer les algorithmes tout en assurant la protection de la vie privée.
- **Analyse de tendances** : Apple applique cette technique à grande échelle pour analyser des tendances d'utilisation, comme les emojis les plus utilisés, sans compromettre l'anonymat des utilisateurs.

### 6.2. Pourquoi cette méthode est-elle pertinente ?

Cette méthode permet un équilibre optimal entre :

- **Utilité des données** : les résultats statistiques restent exploitables pour des analyses approfondies.
- **Protection des données personnelles** : les risques d'identification ou de ré-identification des individus sont minimisés.

### 6.3. Comment adopter cette méthode ?

Il existe des bibliothèques logicielles et outils pour intégrer cette méthodologie :

- Google Differential Privacy : Pour intégrer des mécanismes de bruit dans les analyses.
- OpenDP (par Harvard et Microsoft) : Un cadre open source pour mettre en œuvre la confidentialité différentielle.
    - https://opendp.org/
- PySyft (OpenMined) : Pour combiner confidentialité différentielle et apprentissage fédéré.
- Ou utiliser du code Python

## 7. Apprentissage fédéré

***Par Vincent Lacomme***

L'apprentissage fédéré est une méthode d'entraînement de modèles d'intelligence artificielle (IA) qui évite la centralisation des données. Contrairement aux approches traditionnelles où les données sont collectées et stockées sur un serveur centralisé, cette technique permet à plusieurs appareils ou institutions de collaborer sur un même modèle en gardant leurs données localement. Seuls les paramètres mis à jour du modèle (et non les données brutes) sont partagés avec un serveur central pour agrégation.

Pour renforcer la confidentialité des données, des techniques comme l'agrégation sécurisée et la confidentialité différentielle sont souvent combinées à l'apprentissage fédéré. Ces mécanismes empêchent la reconstruction ou l'identification des données d'origine à partir des mises à jour partagées.

Source : Geraldine O Mbah, Data privacy in the era of AI: Navigating regulatory landscapes for global*, International Journal of Science and Research Archive*, 2024, 13(02), 2040–2058

*Businesses*, décembre 2024, § 4.2

## 8. Vérifier la propriété intellectuelle des données

*Par Vincent Lacomme avec la participation de Sabine Marcellin*

**Une question fréquente d'utilisateurs de solutions d'IA est : « Ai-je le droit de déposer un fichier PDF imprimé depuis la documentation professionnelle ? »**

Dans les conditions d'utilisation d'Open AI (ChatGPT), il est indiqué[48] :

**« Propriété du Contenu.** Dans le cadre de votre relation avec OpenAI, et dans la mesure où la loi applicable le permet, vous (a) conservez vos droits de propriété sur les Données d'Entrée et (b) êtes titulaire des droits de propriété sur les Données de Sortie. Par la présente, nous vous cédons tous nos droits, titres et intérêts, le cas échéant, sur les Données de Sortie. »

[…]

Vous déclarez et garantissez que vous disposez de tous les droits, licences et autorisations nécessaires pour fournir des Données d'Entrée à nos Services.

Les conditions générales de Mistral indiquent également[49] :

Vous êtes responsable de toutes les Données d'Entrée que vous utilisez. Vous déclarez et vous nous garantissez que vous disposez de tous les droits, licences et autorisations nécessaires pour fournir les Données d'Entrée aux Services.

Le dépôt d'un document (ouvrage, article, graphique, extrait de documentation professionnelle, etc.), constitue un acte de reproduction au sens de l'article L.122-3 du CPI[50] et implique une mise à disposition à un tiers (le système d'IA tel que ChatGPT).

[48] https://openai.com/fr-FR/policies/terms-of-use/, 1er décembre 2024
[49] Conditions générales au 6 février 2025, https://mistral.ai/static/doc/fr-conditions-generales-20250206.pdf
[50] « La reproduction consiste dans la fixation matérielle de l'œuvre par tous procédés qui permettent de la communiquer au public d'une manière indirecte. »

L’utilisation par le système d'IA se décompose en :

- la génération d'un résumé/d’une analyse, soit la création d'une œuvre dérivée ;
- le potentiel réentraînement : exploitation du contenu par ChatGPT pour améliorer le modèle.

Juridiquement, se posent deux questions :

**1. La légalité du dépôt de fichier dans le système d’IA :**

Déposer un document dans un système d’IA est une action techniquement simple mais qui soulève deux questions juridiques sensibles :

- L’absence d'autorisation de l'ayant-droit ;

- La non-applicabilité de l'exception de copie privée (article L.122-5 2° du CPI) en raison d’un usage non strictement personnel et la mise à disposition à un tiers commercial.

Le dépôt d'un document protégé par le droit d'auteur sur une plateforme d'IA générative pourrait constituer une contrefaçon au sens de l'article L.335-2 du CPI car :

- Il s'agit d'une reproduction non autorisée ;
- Aucune exception légale ne s'applique ;
- L'utilisateur met à disposition l'œuvre à un tiers commercial.

**2. La légalité de l'utilisation par le système d'IA**

- Création d'une œuvre dérivée sans autorisation ;

- Exploitation pour l'entraînement du modèle.

Conclusion :

Cette utilisation constitue une double violation du droit d'auteur :

- Violation directe par l'utilisateur (reproduction et mise à disposition)
- Violation indirecte par le système d'IA (exploitation et création dérivée)

**Recommandations :**

- Obtenir l'autorisation préalable des ayants-droits

- Utiliser uniquement des contenus libres de droits ou sous licence appropriée

*Voir aussi :*

UGGC avocats, *Quand le droit d'auteur rencontre l'IA : ChatGPT et Google BARD vont nous raconter des histoires*, 10/02/2023, https://www.uggc.com/quand-le-droit-dauteur-rencontre-lia-chatgpt-et-google-bard-vont-nous-raconter-des-histoires/

# CRITERES DE CHOIX D'UNE SOLUTION D'IA (EXIGENCES)

*Par Sabrina Agrapart*

**Ce tableau a été préparé au 4e trimestre 2024. En raison des évolutions rapides des conditions, le lecteur sera avisé de consulter les dernières formules d'abonnement.**

| | ChatGPT gratuit | ChatGPT plus | ChatGPT équipe | ChatGPT entreprise | Microsoft 365 Copilot pour les clients professionnels | Mistral |
|---|---|---|---|---|---|---|
| **Réutilisation des données pour l'entrainement** | Désinscription possible[51] | Désinscription possible | Non[52] | Non. Données chiffrées et confidentielles. | Non[53] | |
| **Conservation et contrôle des données** | Pas de contrôle | Pas de contrôle | Contrôle limité, gestion centralisée des équipes. | Contrôle complet, | Contrôle complet[54] | |
| **Sécurité et conformité** | Sécurité basique | Sécurité basique | Sécurité renforcée | Sécurité renforcée[55] | Sécurité renforcée | |
| **Lieu de stockage** | Etats-Unis | Etats-Unis | Etats-Unis<br>Possibilité de mettre en place un addendum sur la protection des données[56] | Etats-Unis<br>possibilité de mettre en place un addendum sur la protection des données [57] | Union européenne | Union européenne |
| **Tarif** | Gratuit | 20$/mois | 25$/utilisateur/mois facturé annuellement<br>ou<br>30 $/utilisateur/mois facturé mensuellement | Sur devis | Sur devis | Technologie \| Mistral AI \| Frontier AI in your hands |

[51] Pour désactiver l'entrainement des instructions génératives : depuis ChatGPT, cliquer sur « Paramètres » puis sur « Gestion des données » et désactiver « Améliorer le modèle pour tous ».
[52] Prix de ChatGPT | OpenAI : Vie privée
[53] Données, confidentialité et sécurité pour Microsoft 365 Copilot | Microsoft Learn
[54] En savoir plus sur la rétention des Microsoft Copilot pour Microsoft 365 | Microsoft Learn
[55] ChatGPT for enterprise | OpenAI
[56] Confidentialité d'entreprise | OpenAI
[57] Confidentialité d'entreprise | OpenAI

# COMPARAISON DES CONDITIONS GENERALES D'UTILISATION (CGU) DES IA GENERATIVES CHATGPT, COPILOT ET MISTRAL

*Par Sabrina Agrapart et Romain Mirabile*

**Ce tableau a été préparé au 4e trimestre 2024. En raison des évolutions rapides des conditions, le lecteur sera avisé de consulter les dernières versions des conditions d'utilisation.**

*Ce détail n'est qu'une analyse et implique pour chaque personne de prendre connaissance de ces documents ou d'autres IA au moment d'une quelconque souscription.*

*L'utilisation de toute donnée personnelle ou confidentielle doit être analysée au cas par cas pour chaque outil selon les documents existants et les mesures de sécurité mises en œuvre et la sensibilité des informations.*

| Catégorie | ChatGPT | Mistral | Microsoft 365 Copilot pour les clients professionnels |
|---|---|---|---|
| **Licence** | **Limitation des utilisations des Services pour certaines activités : “**Vous ne pouvez pas utiliser nos Services pour des activités illégales, dommageables ou abusives (...)." (CGU). | **Limitation des utilisations des Services pour certaines activités :** "Vous ne devez pas utiliser nos Services pour des activités illégales, nuisibles ou contraires aux droits d’autrui (...)." (CGU – Terms of Use - 8).<br><br>I**nterdiction de contournement des Services** “Vous ne devez pas utiliser nos Services dans le but de contourner les caractéristiques ; fonctionnalités ou limitations prévues ou pour détourner nos services de leurs objectifs tels qu’ils sont définis dans la présent accord” (CGU – Terms of Use - 8).<br><br>**Possibilité d’un usage commercial des Services :** "Les Services peuvent être utilisés à des fins commerciales sous réserve d’un abonnement approprié et du respect des présentes Conditions (...)." (CGU).<br><br>**Interdiction pour l’utilisateur d’utiliser les Services pour autrui : “**Vous ne devez pas utiliser nos Services pour le bénéfice d’un tiers, sous réserve d’un accord dans un contrat séparé avec nous”. (CGU – Terms of Use - 8)<br><br>**Interdiction pour l’utilisateur d’extraire du contenu des Services :** “Vous ne devez pas | **Limitation des utilisations des Services pour certaines activités :** "Vous devez utiliser les Services en ligne uniquement : (i) de manière légale et conforme à toutes les lois applicables ; (ii) en accord avec les termes de cet accord et du Code de Conduite (...)." (Copilot AI Experiences Terms, Section "Using the Online Services", dernière mise à jour).<br><br>**Sanction en cas de violation du Code de Conduite :** "Microsoft peut suspendre ou résilier votre accès en cas de violation grave ou répétée du Code de Conduite (...)." (Copilot AI Experiences Terms, Section "Suspension and Cancellation", dernière mise à jour). |

| | | | |
|---|---|---|---|
| | | extraire, par transfert permanent ou temporaire, tout ou partie du contenu de Nos Services, par quelque moyen et sous quelque forme que ce soit, y compris par scraping, sauf autorisation contraire aux termes du présent Accord » (CGU – Terms of Use - 8).<br><br>**Interdiction de sous License sauf accord** “Vous ne devez pas accorder de licence, de sous-licence ou d’accès aux Services, ni les vendre, les revendre, les prêter, les louer ou les distribuer, sous quelque forme que ce soit, à des tiers, sauf autorisation contraire en vertu du présent accord” (CGU). | |
| **Modèles** | **Interdiction d’extraction automatique du contenu des Services :** "Vous ne pouvez pas (...) extraire automatiquement ou par programmation des données ou des Données de Sortie (...)." (CGU).<br><br>**Interdiction de découvrir les composants sous-jacents des Services :** "Vous ne pouvez pas (...) découvrir le code source ou les composants sous-jacents de nos Services, y compris nos modèles (...)." (CGU). | **Interdiction de découvrir les composants sous-jacents des Services :** "Vous ne pouvez pas tenter de désassembler, décrypter ou extraire le code source de nos modèles (...)." (CGU – Terms of Use - 8).<br><br>**Interdiction de porter atteinte aux Services pour manipuler le modèle :** « Vous ne devez pas tenter ou vous engager dans des activités qui pourraient compromettre la sécurité, la modération ou le bon fonctionnement des services. En particulier, le client s’abstiendra de toute tentative d’injection de données malveillantes ou d’attaques par injection de données dans le but | **Interdiction d’extraction automatique du contenu des Services :** "Vous ne pouvez pas (...) extraire automatiquement ou par programmation des données ou des Données de Sortie (...)." (Conditions d'Utilisation, Section Ce que vous ne pouvez pas faire).<br><br>**Interdiction de découvrir les composants sous-jacents des Services :** "Vous ne pouvez pas (...) découvrir le code source ou les composants sous-jacents de nos Services, y compris nos modèles (...)." (Conditions d'Utilisation, Section Ce que vous ne pouvez pas faire).<br><br>**Possibilité d’interactions automatisées et manuelles avec les Services :** "Les interactions utilisateur |

| | | | |
|---|---|---|---|
| | **Interdiction d'utiliser le modèle pour le concurrencer :** « Il vous est interdit d'utiliser des données de sorties pour développer des modèles qui concurrencent OpenAI. » (CGU) | de manipuler le comportement du modèle » (CGU – Terms of use - 8). | avec Copilot peuvent inclure à la fois un traitement automatisé et manuel des données (...)." (Copilot AI Experiences Terms, Section "Using the Online Services", dernière mise à jour).<br><br>**Pouvoir discrétionnaire de Microsoft sur les fonctionnalités de son modèle :** "Microsoft se réserve le droit de modifier les fonctionnalités ou de limiter la vitesse des Services en ligne à sa discrétion (...)." (Copilot AI Experiences Terms, Section "No Guarantees; No Representations or Warranties", dernière mise à jour). |
| **Données d'apprentissage** | **Faculté des Services d'entrainer ses modèles avec le contenu de l'utilisateur :** "Nous pouvons utiliser votre Contenu (...) pour entraîner nos modèles (...)." (Politique de Confidentialité).<br><br>**L'utilisateur peut exprimer son refus d'utilisation de ses données pour entrainer les modèles des Services :** "Si vous ne souhaitez pas que nous utilisions votre Contenu pour entraîner nos modèles, vous avez la possibilité de vous y opposer en mettant à jour les paramètres de votre | **Utilisation des données pour entrainer les modèles des Services :** "Vos données sont utilisées pour générer des Outputs, afficher l'historique des Conversations, partager vos Conversations et entraîner nos modèles, sauf opt-out spécifique pour les utilisateurs des Paid Services (...)." (Additional Terms - Le Chat, Section 3, Your User Data).<br><br>**Services gratuits : Utilisation des données d'utilisateur pour entrainer les modèles des Services :** « Si vous utilisez nos Services Gratuits, nous utilisons vos données d'utilisateur pour (…) entrainer nos modèles » (Additionnal Terms – La Plateforme, Section 2.2.2.1). | **Faculté des Services d'entrainer ses modèles avec le contenu de l'utilisateur :** "Nous pouvons utiliser votre Contenu (...) pour entraîner nos modèles (...)." (Politique de Confidentialité, Section 3). |

| | | | |
|---|---|---|---|
| | compte. Pour plus d'informations, consultez cet article du Centre d'Aide. Veuillez noter que, dans certains cas, cela peut limiter la capacité de nos Services à mieux répondre à votre cas d'utilisation spécifique." (CGU). | « Entrainement : Nous n'utilisons pas vos données d'entrée pour entrainer ou améliorer nos modèles, à moins que vous utilisiez nos Services gratuits, auquel cas, vous nous garantissez une licence internationale, sans royaltie, pour utiliser vos données d'utilisation, afin d'entrainer ou améliorer nos modèles pour la durée des droits de propriété intellectuelle applicables » (Additionnal Terms – La Plateforme, Section 2.2.2.1).<br><br>« Nous n'utilisons pas vos données d'utilisations si vous répondez aux conditions suivantes : vous utilisez nos Services payants et vous avez choisi de ne pas participer à l'entrainement en utilisant le paramètre applicable sur votre compte » (Additionnal Terms – La Plateforme, Section 2.2.3).<br><br>**Services payants : Non-utilisation des données d'utilisateur pour entrainer les modèles des Services :** « Services payants (…) : Nous n'utilisons pas vos données d'utilisateur pour entrainer (…) nos modèles. » (Additionnal Terms – La Plateforme, Section 2.2.2.2).<br><br>**Les réclamations relatives à l'entrainement des Modèles sont sujettes à des limites techniques impliquant une procédure technique complexe :** « Nous ferons au mieux afin de répondre à Vos demandes. Cependant, lorsque | |

| | | | |
|---|---|---|---|
| | | vous vous renseignez à propos de l'entrainement de nos Modèles, il est important de noter que vous avez des limites techniques et répondre à vos requêtes pourra impliquer une procédure technique complexe » (CGU – Privacy Policy – 8) | |
| **Confidentialité et sécurité** | **Protection des données à caractère personnel de l'utilisateur :** "Vos données sont protégées par des mesures techniques et organisationnelles (...)." (Politique de Confidentialité, Section 8).<br><br>**Faculté de l'utilisateur de s'opposer à l'utilisation de ses données d'entrée et de sortie pour entrainer les modèles du Service :** "Si vous ne souhaitez pas que nous utilisions votre Contenu pour entraîner nos modèles, vous avez la possibilité de vous y opposer (...)." (Conditions d'Utilisation, Section Refus ; Politique de Confidentialité, Section 3). | **Le Service doit tout mettre en œuvre pour maintenir une sécurité des services (sans pour autant en garantir le résultat) :** Obligations de moyens de maintenir une sécurité des services (CGU, Additional Terms Chat, point 9)<br><br>**Les modèles ne sont en principe pas entrainés à partir de données personnelles mais c'est une possibilité si ces données sont disponibles et publiques sur Internet :** « Nos modèles sont entrainés à partir de données disponibles publiquement sur Internet, ce qui peut contenir des données personnelles, même si nous faisons preuve de bonnes pratiques afin de filtrer de telles données personnelles » (Privacy Policy – 3.3)<br><br>**Aucune utilisation (accès ; traitement) des données personnelles n'est faite lorsque l'utilisateur utilise le service :** « Nous n'accédons, ni ne traitons aucune donnée personnelle contenue dans vos données | **Protection des données personnelles de l'utilisateur :** "Vos données sont protégées par des mesures techniques et organisationnelles (...)." (Politique de Confidentialité, Section 8).<br><br>**Faculté de l'utilisateur de s'opposer à l'utilisation de ses données d'entrée et de sortie pour entrainer les modèles du Service :** "Si vous ne souhaitez pas que nous utilisions votre Contenu pour entraîner nos modèles, vous avez la possibilité de vous y opposer (...)." (Conditions d'Utilisation, Section Refus ; Politique de Confidentialité, Section 3).<br><br>**Les contenus générés peuvent contenir des erreurs ou être imprécis :** "Microsoft ne garantit pas que les contenus générés soient exempts d'erreurs ou adaptés à un usage spécifique (...)." (Copilot AI Experiences Terms, Section "No Guarantees; No Representations or Warranties", dernière mise à jour).<br><br>**Les données d'entrée et de sortie peuvent être bloqués, restreints ou supprimés s'ils violent le Code de Conduite :** "Les prompts ou contenus |

| | | | |
|---|---|---|---|
| | | d'utilisateur lorsque vous utilisez nos services (…) » (Partner Hosted – Deployment Terms – 10) | générés qui violent le Code de Conduite peuvent être bloqués, restreints ou supprimés (...)." (Copilot AI Experiences Terms, Section "Content and Moderation", dernière mise à jour). |
| **Contenus générés** | **L'utilisateur est titulaire de tous les droits sur le contenu généré par le Service :** « Vous êtes titulaires des droits de propriété sur les Données de Sortie (...). Par la présente, nous vous cédons tous nos droits, titres et intérêts, le cas échéant, sur les Données de Sortie » (CGU).<br><br>**La cession des droits évoquée ci-dessus ne s'étend pas aux données de sortie d'autres utilisateurs ou de tiers, quand bien même elles peuvent être identiques :** "En raison de la nature de nos Services et de l'intelligence artificielle en général, l'Output peut ne pas être unique et d'autres utilisateurs peuvent recevoir des résultats similaires de nos Services. Notre cession ci-dessus ne s'étend pas à l'Output d'autres | **L'utilisateur est titulaire de tous les droits sur le contenu généré par le Service (sauf indication contraire dans un contrat spécifique) :** "Vous conservez tous les droits de propriété intellectuelle sur vos Outputs, sauf indication contraire explicite dans le cadre d'un contrat spécifique (...)." (CGU, Section 6, User Data, dernière mise à jour).<br><br>**Les données de sortie peuvent être similaires à d'autres contenus générés :** "Nous ne garantissons pas que les Outputs soient exempts de similitudes avec d'autres contenus générés ou existants (...)." (Conditions d'Utilisation, Section 6, Output Similarity, dernière mise à jour).<br><br>**Par conséquent, aucune indemnisation ne peut avoir lieu au bénéfice de l'utilisateur sous prétexte que ses données de sortie sont similaires à d'autres :** « Vous acceptez qu'en raison de la nature de nos services, si un autre utilisateur utilise une entrée similaire à la vôtre, nos services peuvent générer une sortie similaire ou identique à la vôtre. Nous ne | **L'utilisateur est titulaire de tous les droits sur le contenu généré par le Service :** "Vous conservez vos droits de propriété (...) sur les Données de Sortie (...)." (Conditions d'Utilisation, Section Propriété du Contenu).<br><br>**Les données de sortie peuvent être similaires à d'autres contenus générés :** "Microsoft avertit que les Créations peuvent ne pas être uniques et que des contenus similaires peuvent être générés pour plusieurs utilisateurs (...)." (Copilot AI Experiences Terms, Section "Using the Online Services", dernière mise à jour).<br><br>**Le Service ne garantit pas le respect des droits des tiers par les contenus qu'il génère :** "Microsoft n'offre aucune garantie quant à la non-violation des droits de tiers par les contenus générés (...)." (Copilot AI Experiences Terms, Section "No Guarantees; No Representations or Warranties", dernière mise à jour).<br><br>**Le Service n'a pas de droit sur les contenus générés mais se réserve toutefois des droits de licence pour les réutiliser :** "Microsoft ne revendique pas la propriété des Prompts ou Contenus générés mais se réserve des |

| | | | |
|---|---|---|---|
| | utilisateurs ni à l'Output de tiers." (CGU).<br><br>**Interdiction pour l'utilisateur d'extraire des données de sortie ; de déclarer que les données de sortie ont été générées par un humain ou utiliser ces données pour développer des modèles concurrençant OpenAI :** « Il vous est interdit d'extraire automatiquement ou par programmation des données ou des Données de Sortie, ; déclarer que des données de sortie ont été générées par un humain alors que ce n'est pas le cas ou utiliser des données de sortie pour développer des modèles qui concurrencent OpenAI » (CGU)<br><br>**L'utilisateur est seul responsable de l'utilisation des données d'entrée et de sortie et dispose donc de tous les droits sur les données d'entrée qu'il soumet au Service :** « Vous êtes responsable du Contenu, notamment en vous assurant qu'il n'enfreint | garantissons pas que votre résultat ne soit pas similaire ou identique au résultat d'un autre utilisateur. En conséquence, nous ne sommes pas tenus de vous indemniser en cas de réclamation alléguant que votre contenu généré est similaire à celui d'un autre utilisateur. » (Partner Hosted – Deployment Terms – 6)<br><br>**L'utilisateur est seul responsable de l'utilisation des données d'entrée et de sortie :** « Vous êtes le seul responsable de votre utilisation des données d'entrée et de sortie » (Partner Hosted – Deployment Terms – 6)<br><br>**L'utilisateur est seul titulaire de droits sur les données de sortie, y compris les droits de propriété intellectuelle leur afférent :** « Nous ne revendiquons aucun droit de propriété ou de propriété intellectuelle de quelque nature que ce soit sur les données de sortie. Vous en êtes le seul propriétaire et conservez tous les droits, y compris, mais sans s'y limiter, tous les droits de propriété intellectuelle, sur vos données de sortie » (Partner Hosted – Deployment Terms – 6)<br><br>**Interdiction pour l'utilisateur d'utiliser les données de sortie pour manipuler le Service :** « Vous ne devez pas utiliser les sorties ou toute version modifiée ou dérivée des sorties pour faire de l'ingénierie inverse sur les services | droits de licence pour copier, distribuer, et afficher ces contenus (...)." (Copilot AI Experiences Terms, Section "Ownership of Content", dernière mise à jour). |

| | | | |
|---|---|---|---|
| | aucune loi applicable ni les présentes Conditions. Vous déclarez et garantissez que vous disposez de tous les droits, licences et autorisations nécessaires pour fournir des Données d'entrée à nos Services » (CGU) | » (Partner Hosted – Deployment Terms – 6) | |
| **Responsabilité** | **Indemnisation limitée à 12 mois de services payés (GCU) :** « que vous avez payé pour le service qui a donné lieu à la réclamation au cours des 12 derniers mois et cent dollars (100 $). Les limitations prévues par cet article ne s'appliquent que dans la mesure maximale permise par la loi applicable. » (CGU) | **Irresponsabilité du Service en cas d'utilisation abusive ou non conforme par l'utilisateur :** "Mistral AI décline toute responsabilité liée à l'utilisation abusive ou non conforme des Services par l'utilisateur (...)." (Conditions d'Utilisation, Section 11.1, Warranties and Indemnification, dernière mise à jour).<br><br>**Tout litige sera réglé par arbitrage conformément aux règles de la Chambre de Commerce Internationale :** "Les litiges éventuels seront réglés par voie d'arbitrage selon les règles de la Chambre de Commerce Internationale (...)." (Conditions d'Utilisation, Section 18.2.2, Competent Jurisdiction, dernière mise à jour).<br><br>**Indemnisation limitée à 100 euros pour les services gratuits et 12 mois de services payés ou 100 000 euros pour les services payants :** « Dans la limite de la loi | **Indemnisation fortement limitée :** "Indemnisation limitée à certaines circonstances (...)." (Conditions de Services, Section 1).<br><br>**L'utilisateur doit indemniser le Service pour toute réclamation résultat de l'utilisation des contenus générés ou de la violation des conditions d'utilisation :** "Les utilisateurs doivent indemniser Microsoft contre toute réclamation résultant de l'utilisation des contenus générés ou de la violation des présentes conditions (...)." (Copilot AI Experiences Terms, Section "Indemnification by You", dernière mise à jour). |

| | | | |
|---|---|---|---|
| | | et sous réserve de conditions supplémentaires applicables, la responsabilité totale de Mistral AI ou de nos actionnaires, employés, sociétés affiliées, concédants de licence, agents, fournisseurs et prestataires de services en relation avec ou dans le cadre des services gratuits, ou de votre utilisation ou incapacité à utiliser les services gratuits, n'excédera en aucun cas 100 euros (…)<br><br>Pour les services payants (…) le montant des redevances payées ou dues par le client au cours des 12 mois civils précédant la date à laquelle le premier de ces événements s'est produit ou 100 000 euros » (Partner Hosted – Deployment Terms – 12.2) | |
| **Conformité légale** | **RGPD et lois locales applicables :** "Conformité au RGPD et aux lois locales applicables (...)." (Politique de Confidentialité). | **RGPD et autres réglementations applicables relatives à la protection des données :** "Conformité au RGPD et autres réglementations applicables relatives à la protection des données personnelles (...)." (Conditions d'Utilisation, Section 14.1, Data Controller, dernière mise à jour). | **RGPD et lois locales applicables :** Conformité au RGPD et aux lois locales applicables (...)." (Politique de Confidentialité, Section 1).<br><br>**Code de conduite et lois applicables, notamment en matière de confidentialité et de droits d'auteur :** "Les contenus générés doivent respecter le Code de Conduite et les lois applicables, notamment en matière de confidentialité et de droits d'auteur (...)." (Copilot AI Experiences Terms, Section "Using the Online Services", dernière mise à jour).<br><br>Disponible pour Microsoft 365 Copilot et de Microsoft 365 Copilot Chat |

| | | | |
|---|---|---|---|
| **Droit des utilisateurs** | **Droit d'accès ; rectification ; suppression des données :** "Vous avez le droit (...) d'accéder, de rectifier ou de supprimer vos données (...)." (Politique de Confidentialité)<br><br>Pas de DPA disponible pour tous les niveaux d'abonnement | **Droit d'accès ; rectification ; suppression ; portabilité des données : "**Les utilisateurs disposent du droit d'accéder à leurs données, de les rectifier, de les supprimer ou d'en demander la portabilité (...)." (Privacy policy – 8). | **Droit d'accès ; rectification ; suppression des données :** "Vous avez le droit (...) d'accéder, de rectifier ou de supprimer vos données (...)." (Politique de Confidentialité, Section 6).<br><br>**Faculté pour l'utilisateur de s'opposer à une suspension en cas de désaccord, via un formulaire de préoccupation : "**Les utilisateurs peuvent faire appel d'une suspension en cas de désaccord via le formulaire de préoccupation prévu à cet effet (...)." (Copilot AI Experiences Terms, Section "Access Restriction and User Appeal", dernière mise à jour). |

### 8.1. L'exemple de DeepSeek

***Par Vincent Lacomme***

DeepSeek est un modèle de langage d'IA chinois qui a fait couler beaucoup d'encre en janvier 2025. A la lecture de la politique de confidentialité[58], il apparait que :

- Les informations sont stockées dans des « serveurs sécurités en Chine » ;
- Il n'existe aucun représentant de la société dans l'UE (article 27 de la RGPD).

[58] https://chat.deepseek.com/downloads/DeepSeek%20Privacy%20Policy.html

# PARTIE 4

# EXEMPLES ET TÉMOIGNAGES

Il est pertinent de s'inspirer des exemples d'autres professions et de comparer les pratiques entre pays. Nous proposons au lecteur un tour du monde non exhaustif afin de mettre en perspective les législations, bonnes pratiques et jurisprudences et, tout d'abord, une illustration chez les avocats.

# L'UTILISATION DES OUTILS D'IA GENERATIVE PAR LES AVOCATS ET LA DEONTOLOGIE

***Par Virginie Bensoussan-Brulé***

Les incidences de l'intelligence artificielle pour les professionnels du droit concernent le maintien de la pertinence de la règle de droit dans le temps, ainsi que le suivi des avancées, des limites des IA et de la manière dont elles sont perçues.

## 1. L'utilisation des outils d'intelligence artificielle générative par les professionnels du droit

### 1.1. L'acceptabilité de l'erreur machine

L'essor de l'intelligence artificielle a conduit à une multiplication de ses applications. Cela pose une question essentielle, celle de l'acceptabilité de l'erreur machine, en particulier dans le domaine juridique.

Dans le cas des systèmes de reconnaissance d'image, l'acceptabilité de l'erreur varie en fonction de sa nature et de sa fréquence. Selon une approche qualitative, une erreur grossière, telle que la confusion entre un être humain et un animal, peut entraîner une perte de confiance. Celle-ci est accentuée lorsque l'erreur est perçue comme absurde ou irrationnelle. Selon une approche quantitative, l'occurrence de l'erreur joue un rôle clé dans son acceptabilité.
En outre, les biais anthropomorphiques exercent une influence sur notre perception de l'erreur machine.

Dans le cas des IA génératives, l'acceptabilité de l'erreur réside dans la comparaison par rapport à notre bras droit. L'exigence du *« tout, tout de suite, de manière parfaite »* diminue notre acceptabilité de l'erreur.

### 1.2. L'hybridation des intelligences

Le développement de l'intelligence artificielle transforme la pratique du droit, entraînant ainsi une hybridation croissante des intelligences.

L'intégration de l'IA dans le domaine juridique marque le passage du *« faire »* au *« faire faire »*. Elle permet aux professionnels du droit de déléguer certaines missions à des outils d'IA et de ne pas se consacrer exclusivement à des tâches répétitives et chronophages. Cette réorganisation permet aux professionnels du droit de se concentrer sur des activités à plus forte valeur ajoutée. Ainsi, l'utilisation de l'intelligence artificielle entraîne un phénomène de déplacement de la valeur.

Afin de tirer pleinement profit des bénéfices présentés par l'intelligence artificielle, il est nécessaire de maximiser la coopération entre les avocats et l'IA et de minimiser les confrontations.

## 2. La protection des données personnelles par les professionnels du droit utilisant de l'IA

L'utilisation saine et créatrice de valeur de l'intelligence artificielle par les professionnels du droit repose sur les quatre principes suivants :

- le principe d'adhérence technologique ;
- le principe de compétence ;
- le principe de maîtrise et de non dessaisissement ;
- le principe de transparence ;
- le respect du secret professionnel.

Dans le cadre de la protection des données personnelles par les professionnels du droit utilisant des outils d'IA, l'Union internationale des avocats et le Conseil national des barreaux ont formulé plusieurs recommandations.

### 2.1. Les recommandations de l'Union internationale des avocats

En septembre 2024, à Novi Sad, ont été adoptées les directives de l'Union internationale des avocats sur l'utilisation des systèmes d'intelligence artificielle par les avocats.

L'UIA, en tant que leader mondial de la *« nation des avocats »*, a un rôle important et significatif à jouer dans la formation des avocats sur l'utilisation responsable et éthique de l'IA.

Plus globalement, l'UIA joue un rôle clé dans le débat sur la manière dont l'IA peut servir au mieux la justice, sur la question de savoir dans quelle mesure l'IA doit être réglementée, sur les dangers et les conséquences d'une utilisation irresponsable et non réglementée, ainsi que sur la manière de s'assurer que l'IA n'est pas discriminatoire, pour ne citer que quelques-unes des questions pertinentes.

### 2.2. Les recommandations du Conseil national des barreaux

Le CNB a participé à la consultation préalable du Livre Blanc sur l'intelligence artificielle de la Commission européenne publié le 19 février 2020.

En 2024, le CNB a mis en place un groupe de travail dédié à l'IA, à savoir le comité de pilotage chargé d'accompagner et de coordonner la réflexion et le travail effectués au sein des différentes commissions.

En septembre 2024, le CNB a publié un guide pratique sur l'utilisation des systèmes d'intelligence artificielle. L'objectif de celui-ci est d'aider les avocats à se familiariser avec l'IA générative pour permettre l'intégration de cette technologie dans leur pratique professionnelle.

Les risques de l'utilisation des IA génératives identifiés sont :

- les risques de confidentialité liés à la réutilisation des données transmises par le propriétaire du modèle d'IA générative ;
- les risques pour la protection des données personnelles.

Le guide rappelle que le RGPD s'applique à tous les cabinets d'avocats, *« quels que soit leur taille, leur structure et leur domaine d'activité »*.

A ce titre, le guide fournit plusieurs recommandations, parmi lesquelles figurent les deux suivantes :

- *« Il est impératif pour l'avocat de veiller à ne jamais communiquer des données relatives à ses dossiers ou à ses clients à des intelligences artificielles génératives. »* ;
- *« Afin d'utiliser des outils d'intelligence artificielle générative tout en protégeant ses données, il est conseillé à l'avocat de mettre en œuvre une bonne pratique : la pseudonymisation des données. La pseudonymisation est un traitement de données personnelles qui consiste à « remplacer les données directement identifiantes (nom, prénom, etc.) d'un jeu de données par des données indirectement identifiantes ». Cette pratique, qui ne présente pas de caractère irréversible (contrairement à l'anonymisation), permet ainsi d'utiliser les données sans accorder d'identification. »*

# LES AVOCATS : DIRECTIVES DE L'UIA (UNION INTERNATIONALE DES AVOCATS)

***Par Vincent Lacomme***

L'union internationale des avocats est une association française loi 1901. En octobre 2024, lors d'un congrès à Paris, elle a publié un guide de bonnes pratiques sur l'IA[59] à l'attention des avocats dont nous reprenons les points clés ci-dessous.

## 1. Compréhension et maîtrise des outils

Avant toute utilisation d'un système d'IA, le praticien doit développer une compréhension approfondie de l'outil qu'il compte utiliser. Cela implique, comme souligné au §2 du document, de connaître : *« les sources des données utilisées dans l'entraînement du système, comment les données saisies seront utilisées à l'avenir, et la fiabilité des résultats ainsi que les risques d'hallucinations. »*

## 2. Protection des données et confidentialité

Le document définit les mesures de protection à prendre :

1. Examiner systématiquement les conditions générales d'utilisation du système d'IA
2. Privilégier les systèmes qui ne conservent pas les informations
3. En cas d'impossibilité, procéder à une anonymisation rigoureuse
4. Exclure le partage d'informations particulièrement sensibles (données de santé, mots de passe, informations stratégiques…).

[59] https://www.uianet.org/sites/default/files/uia_guidelines_use_ai_systems_by_lawyers_en_final.pdf

## 3. Supervision et contrôle

L'IA doit être considérée comme un outil d'assistance et non de remplacement. Le document insiste particulièrement (§8) sur la nécessité d'une « vérification systématique des résultats produits par l'IA ». Cela implique notamment de :

- Vérifier l'existence réelle et l'exactitude des sources juridiques citées ;
- Contrôler la pertinence des contenus générés ;
- S'assurer de la conformité avec le droit applicable.

## 4. Responsabilité professionnelle et éthique

L'avocat « reste pleinement et uniquement responsable de toute décision prise, activité réalisée ou contenu diffusé » (§7), même lorsque ceux-ci sont basés sur des résultats produits par l'IA.

## 5. Transparence client

Le guide prescrit une information claire sur l'utilisation de l'IA et l'obtention du consentement pour l'utilisation des données.

Cela implique de demander le consentement de tous les participants à une réunion avant d'utiliser un système d'IA pour transcrire, résumer ou traduire des réunions ou des conversations téléphoniques.

## 6. Traçabilité et documentation

Les bonnes pratiques en matière de traçabilité comprennent :

- L'enregistrement des activités des systèmes d'IA
- La documentation des données d'entrée
- La conservation des logs d'utilisation

## 7. Protection des droits fondamentaux

Le document conclut sur la nécessité d'identifier proactivement les risques pour les droits humains et de maintenir les plus hauts standards éthiques.

# INTERNATIONAL : L'AFRIQUE

***Par Vincent Lacomme***

En Afrique, la protection des données est encadrée par l'African Union (AU) Data Policy Framework adopté en juillet 2022[60]..

La convention de Malabo[61] appelée « **Convention de l'Union Africaine sur la cybersécurité et la protection des données personnelles** (African Union Convention on Cyber Security and Personal Data Protection) », est entrée en vigueur en 2023 (après une adoption en 2014 par l'Union Africaine) afin d'harmoniser les différents textes légaux relatifs à la cybersécurité et la protection des données personnelles.

Un livre blanc « Régulation et adoption responsable de l'IA en Afrique vers la réalisation de l'Agenda 2063 de l'Union africaine », publié en avril 2024 par l'Agence de développement de l'Union africaine, dans le cadre du Nouveau Partenariat pour le développement de l'Afrique (AUDA-NEPAD)[62], à son article 5.9.2, insiste sur les dangers de vols de données, usurpations d'identité, mauvaise utilisation de données et fait référence à l'article 13 de la Convention de Malabo qui établit six principes essentiels :

1. Consentement et légitimité du traitement des données
2. Légalité et juste utilisation des données
3. Objectif, pertinence et durée de conservation
4. Exactitude
5. Transparence dans le traitement des données
6. Confidentialité et sécurité

[60] https://au.int/sites/default/files/documents/42078-doc-AU-DATA-POLICY-FRAMEWORK-ENG1.pdf

[61] https://www.afapdp.org/wp-content/uploads/2018/06/CONV-UA-CYBER-PDP-2014.pdf

[62] AUDA-NEPAD White Paper: Regulation and Responsible Adoption of AI in Africa Towards Achievement of AU Agenda 2063, avril 2024, https://dig.watch/resource/auda-nepad-white-paper-regulation-and-responsible-adoption-of-ai-in-africa-towards-achievement-of-au-agenda-2063

Lecture recommandée :

Laure Tall, Mamadou Niang, Fatou Fofana, Cheikh Faye, Marame Cissé, Ndèye Fatou Mboup, Isac Mingou, Tabara Korka Ndiaye and Seynabou Sall, “Analyse des cadres politique, juridique, institutionnel et éthique pour une IA responsable en Afrique de l’Ouest : les cas du Bénin, du Burkina Faso, de la Côte-d’Ivoire et du Sénégal”, 01/11/2024, http://journals.openedition.org/ctd/12418

# INTERNATIONAL : ITALIE

*Par Vincent Lacomme*

La GPDP, équivalent de la CNIL en Italie, a ouvert une enquête sur OpenAI dès le mois de mars 2023.

Elle fait suite à des préoccupations soulevées par le Comité européen de la protection des données (EDPB) concernant le traitement des données personnelles par les solutions d'IA.

Les infractions reprochées à OpenAI incluent :

- **Absence de notification d'une violation de données** : OpenAI n'a pas signalé une fuite de données en mars 2023.
- **Traitement illégal des données personnelles** : OpenAI a utilisé des données personnelles pour entraîner ChatGPT sans base légale appropriée.
- **Violation du principe de transparence** : l'entreprise n'a pas respecté ses obligations d'information envers les utilisateurs.
- **Manque de vérification de l'âge** : absence de mécanismes empêchant des enfants de moins de 13 ans d'accéder à des réponses potentiellement inadaptées.

Le 20 décembre 2024, la GPDP a prononcé[63] :

- une **sanction financière** de 15 millions d'euros à l'encontre d'Open AI.
- une **campagne de communication obligatoire** d'une durée de six mois à travers divers médias (radio, télévision, journaux, Internet). Cette campagne devra :
  - Informer le public sur le fonctionnement de ChatGPT.
  - Expliquer la collecte des données des utilisateurs et non-utilisateurs.
  - Sensibiliser sur les droits des utilisateurs en matière de protection des données (opposition, rectification, suppression).
- **la transmission du dossier à l'autorité irlandaise** de protection des données (irish Data Protection Authority - DPC) dans le cadre du mécanisme de guichet unique prévu par le RGPD.

[63] https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/10085432#english

Voir aussi les communiqués de presse :

- https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870847
- https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/10019984
- https://www.gpdp.it/web/guest/home/docweb/-/docweb-display/docweb/9997575

# INTERNATIONAL : AUTRICHE

*Par Vincent Lacomme*

L'association autrichienne (ONG) Noyb[64] a attaqué OpenAI sur plusieurs points de la RGPD résumés dans le tableau suivant.

| Motif | Description | Articles du RGPD concernés |
|---|---|---|
| **Données personnelles inexactes** | ChatGPT génère des informations erronées sur des individus, comme des dates de naissance incorrectes. | Article 5 : exactitude des données |
| **Refus de rectification** | OpenAI déclare ne pas pouvoir corriger ou supprimer des données erronées sur une personne. | Article 16 : droit de rectification |
| **Manque de transparence** | OpenAI ne fournit pas aux utilisateurs les données détenues sur eux, leurs sources en réponse à une demande d'accès. | Article 15 : droit d'accès |

A ce jour, le dénouement de cette plainte n'est pas connu.

[64] ChatGPT provides false information about people, and OpenAI can't correct it, 29/04/2024, https://noyb.eu/fr/chatgpt-provides-false-information-about-people-and-openai-cant-correct-it

# INTERNATIONAL : LES ÉTATS-UNIS

*Par Vincent Lacomme*

En Californie, le CCPA (California Consumer Privacy Act) et le CPRA (California Privacy Rights Act) sont les équivalents américains de la RGPD dans l'Union Européenne.

## 1. Le CCPA

Le premier texte a été voté en 2018 et est entré en vigueur le 1er janvier 2020 et vise à réguler l'utilisation de données personnelles.

Contrairement au RGPD, qui repose principalement sur le consentement explicite, le CCPA met l'accent sur les droits des consommateurs, notamment le droit d'accès, de suppression et d'opposition à la vente de leurs données. Une particularité notable de cette loi est l'obligation pour les entreprises de divulguer clairement si elles vendent des données personnelles et de fournir une option d'exclusion via un lien visible intitulé « Ne vendez pas mes informations personnelles ». Toutefois, les pénalités pour non-conformité, plafonnées à 7 500 $ par infraction, sont relativement faibles comparées à celles du RGPD, ce qui en limite parfois l'impact dissuasif.

Source : Geraldine O Mbah, Data privacy in the era of AI: Navigating regulatory landscapes for global, *International Journal of Science and Research Archive*, décembre 2024, 13(02), 2040–2058

## 2. ADMT

En décembre 2023, une nouvelle règlementation sur les technologies automatiques a été actée (ADMT : Automated Decision-Making Technology) avec une mise en application pour la fin d'année (https://cppa.ca.gov/meetings/materials/20231208_item2_draft.pdf). L'objectif est d'adresser les solutions prenant des décisions automatiques telles que certaines IA.

La philosophie du texte est d'encadrer l'utilisation d'informations personnelles dans le cadre de solutions de prises de décisions automatiques. Il s'agit de prévenir en amont les utilisateurs de ces solutions, la possibilité de refuser l'utilisation de ces solutions et le droit d'accès à des informations sur l'utilisation de ces solutions.

Plus d'informations :

- Danielle Ocampo, *CCPA and the EU AI ACT*, https://calawyers.org/privacy-law/ccpa-and-the-eu-ai-act/, juin 2024
- https://cppa.ca.gov/announcements/2023/20231127.html

En décembre 2024, la Californie a adopté une nouvelle loi « AB 2013 » à effet au 1er janvier 2026[65]. L'objectif de ce texte est d'imposer aux sociétés éditant des IA des obligations de transparence sur les données utilisées pour entrainer les modèles de langage d'IA. Le but est de limiter les biais, l'utilisation de données protégées par les droits d'auteur.

Plus d'informations : https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240AB2013&utm_source=chatgpt.com

[65] https://www.jdsupra.com/legalnews/california-s-new-generative-ai-law-what-6336899/

## 3. Guide de bonnes pratiques des barreaux américains

En janvier 2024, un guide de bonnes pratiques sur l'utilisation de l'IA générative dans le domaine juridique a été établi par le barreau d'avocats américains californien et de Floride.

- https://calawyers.org/privacy-law/california-state-bar-releases-guidance-on-use-of-genai-in-practice-of-law/
- https://calawyers.org/california-lawyers-association/ethics-guidelines-for-lawyers-using-generative-ai/
- https://www.floridabar.org/etopinions/opinion-24-1/ (ce lien n'est accessible que depuis les États-Unis).

En avril 2024, le barreau de New York a établi des recommandations détaillées :

https://nysba.org/app/uploads/2022/03/2024-April-Report-and-Recommendations-of-the-Task-Force-on-Artificial-Intelligence.pdf

En juillet 2024, l'American Bar association a publié un avis ("formal opinion 512")[66] qui met l'accent sur plusieurs préconisations en matière d'éthique sur le plan des compétences, de la confidentialité, de la communication et des responsabilités en matière de supervision.

Un avocat au Colorado a été suspendu par le barreau américain et licencié de son cabinet après avoir admis avoir utilisé ChatGPT pour préparer une plaidoirie. Les cas cités étaient fictifs.

https://kygo.com/colorado-lawyer-fired-suspended-from-bar-for-using-ai-in-court/

https://www.cbsnews.com/colorado/news/colorado-lawyer-artificial-intelligence-suspension/

Des avocats américains ont été sanctionnés financièrement par un juge à New York pour avoir utilisé l'IA : https://www.cnbc.com/2023/06/22/judge-sanctions-lawyers-whose-ai-written-filing-contained-fake-citations.html

[66] https://www.americanbar.org/content/dam/aba/administrative/professional_responsibility/ethics-opinions/aba-formal-opinion-512.pdf

Autres exemples :

https://www.lemondedudroit.fr/259-magazine/decryptages/91863-sanction-pour-une-avocate-ayant-cite-des-cas-fictifs-generes-par-chatgpt-2.html

Les recommandations en matière de confidentialité des données sont résumées dans le tableau ci-après :

| Thématique | Recommandations |
|---|---|
| **Confidentialité** | • L'IA générative pourrait utiliser les données saisies pour s'entraîner ou les partager avec des tiers, posant des risques de sécurité et de confidentialité. Il est donc crucial de ne pas saisir de données confidentielles sans garanties de protection adéquates. Un avocat peut anonymiser les informations clients et s'abstenir de communiquer des détails permettant d'identifier le client. (Californie)<br>• La divulgation d'informations est interdite sans le consentement éclairé du client, sauf exception (ex : pour servir l'intérêt du client). (Floride – règle 4-1.6 des règles applicables au barreau de la Floride[67])<br>• Les avocats doivent consulter des experts IT / cybersécurité pour s'assurer que les systèmes d'IA utilisés respectent des protocoles stricts de sécurité et de confidentialité. (Californie)<br>• Les avocats doivent évaluer la réputation des fournisseurs de technologie, leurs mesures de sécurité, les politiques de rétention des données et leurs obligations de préserver la confidentialité des informations. (Floride)<br>• L'utilisation d'IA expose potentiellement les données à des cyberattaques, compromettant les informations confidentielles. Les |

[67] https://www.floridabar.org/rules/rrtfb/

| | |
|---|---|
| | avocats doivent être conscients des risques pour éviter des fuites de données sensibles lors de l'utilisation d'outils d'IA. (New York – page 36) |
| **Conformité légale** | • Les avocats doivent veiller à ce que l'usage de l'IA soit conforme aux lois et réglementations pertinentes, y compris celles sur la confidentialité, les transferts de données à l'étranger, les règles de propriété intellectuelle, et la cybersécurité. (Californie) |
| **Supervision, compréhension** | • Des politiques claires sur l'utilisation de l'IA doivent être établies au sein des cabinets pour garantir le respect des obligations professionnelles. Des formations sur les aspects éthiques et pratiques de l'IA sont recommandées. (Californie)<br>• Les avocats doivent comprendre les bénéfices et les risques des technologies, y compris les programmes d'IA générative, pour remplir leurs obligations éthiques. (Floride – règle 4-1.1 des règles applicables au barreau de la Floride[68])<br>• Les avocats doivent superviser des non-avocats impliqués dans la représentation de clients. Le terme de « non-avocats » recouvre les entités non-humaines telles que l'IA générative. (New York – page 30 – Rules of professional conduct RPC 5.3)<br>• Un cabinet d'avocat californien a été sanctionné pour violation éthique en raison de la présence d'hallucinations dans ses argumentaires.<br>(New York – page 30) |
| **Mesures de précaution** | • Utiliser des solutions IA internes (hébergées sur des serveurs locaux plutôt qu'auprès de tiers) peut réduire les risques de divulgation de données. (Floride) |
| **Conservation des données** | Les informations saisies dans des systèmes d'IA doivent être préservées lorsqu'elles sont pertinentes pour des litiges en cours. Leur nature éphémère complique cette préservation, posant des défis juridiques. (New York page 36) |

[68] https://www.floridabar.org/rules/rrtfb/

Les recommandations et interdictions relatées sont facilement transposables aux métiers du chiffre français.

## 4. Rapport au sujet de confidentialité

La solution américaine Harvey développée par des avocats prend au sérieux les sujets de sécurité (situé en 2e position du ruban de leur site internet).

Les garanties sont annoncées en matière :

- Soc 2 II
- ISO 27001
- CCPA
- RGPD

## 5. Avis du NIST

Le NIST, agence du département du commerce aux États-Unis a publié en juillet 2024 un guide des bonnes pratiques relatives à la gestion du risque autour de l'IA générative[69]. Ce guide contient une cartographie des risques et des pistes d'actions suggérées.

Voici quelques-unes des actions proposées :

*Tableau généré à l'aide de GPT4 et révisé par le groupe de travail :*

| Thème | Propositions de solutions | N° mesure |
|---|---|---|
| **Confidentialité des données** | **Établir des politiques de transparence :** Documenter l'origine et l'historique des données de formation des systèmes d'IA générative pour accroître la transparence tout en respectant les contraintes de confidentialité. | GV-1.2-001 |
| | **Limiter l'utilisation des données personnelles :** Mettre en place des mesures pour minimiser l'inclusion de données sensibles dans les ensembles de formation et éviter les fuites d'informations. | GV-1.2-001 |
| | **Contrôler les fuites de données :** utiliser des techniques telles que le *watermarking* et la gestion des droits numériques pour suivre et contrôler l'utilisation des données personnelles dans les modèles d'IA générative. | GV4.3-001 |
| | **Mettre en place des mécanismes d'évaluation continue :** Évaluer régulièrement les systèmes GAI pour détecter les risques de divulgation d'informations sensibles et ajuster les pratiques de gestion des données en conséquence. | GV-1.5-002 |

[69] https://www.nist.gov/itl/ai-risk-management-framework

| | | |
|---|---|---|
| | **Établir des politiques d'accès restreint :** Limiter l'accès aux modèles et aux données d'entraînement uniquement aux utilisateurs autorisés et formés pour gérer des informations sensibles. | GV-1.6-003 |
| **Propriété intellectuelle** | **Aligner le développement sur les lois en vigueur :** Assurer que le développement et l'utilisation des systèmes GAI respectent les lois sur les droits d'auteur et la propriété intellectuelle, incluant des évaluations légales continues. | GV-1.1-001 |
| | **Mettre en place des contrats clairs :** Rédiger des contrats et des accords de niveau de service (SLAs) avec des clauses spécifiques sur la propriété du contenu, les droits d'usage et les standards de qualité pour les technologies GAI. | GV-6.1-004 |
| | **Évaluer les fournisseurs tiers :** Utiliser un cadre d'évaluation des risques des fournisseurs pour mesurer leur respect des standards de gestion de la propriété intellectuelle et de la provenance du contenu. | GV-6.1-005 |
| | **Établir des politiques de provenance des données :** Documenter et vérifier les sources de données pour s'assurer que le contenu utilisé et généré respecte les droits de propriété intellectuelle des tiers. | GV-1.6-003 |
| | **Surveiller les infractions :** Mettre en place des processus de surveillance et de réponse rapide aux infractions potentielles des droits de propriété intellectuelle identifiées dans les systèmes GAI. | GV-6.2-003 |

## 6. IA générative au DIGITAL CPA (Denver - 2024)

*Par Serge Yablonsky*

Au DCPA, congrès américain pour les experts-comptables, l'intelligence artificielle générative (IAGen) était partout, présentée à chaque table ronde, à chaque démonstration et j'ai noté 3 aspects qui ont particulièrement attiré mon attention :

- Une intervention d'une consultante en IA. Brillante, beaucoup de cas d'usage, très entraînante, très motivante. A la fin de son allocution, une question de la salle : comment conciliez-vous notre devoir de respect du secret professionnel et l'utilisation de l'intelligence artificielle générative avec des données de nos clients ? La réponse : posez-vous en premier lieu la question sur les risques et demain vous serez derrière les autres, vous aurez loupé le train de l'IAGen …..
- Une intervention sur les « outils d'audit » : tous les outils pour les auditeurs sont en train d'intégrer des assistants d'IAGen. Dans les 2 ans qui viennent, il y aura plus de changement d'outils pour les auditeurs que durant les 20 dernières années.
- Une intervention sur le modèle de cabinet « Client Advisory Service » entrevoit une croissance de 17% du fait de l'IAGen dans l'intégration des processus et des systèmes avec les clients des cabinets, amenant à plus de services délégués par les clients aux cabinets d'expertise et une amélioration de la rentabilité du fait de l'augmentation de la sous-traitance par les cabinets à des services spécialisés.

Donc, en synthèse, fonçons avec l'IAGen !

# INTERNATIONAL : CHINE

***Par Vincent Lacomme***

Le **Personal Information Protection Law (PIPL)**, entrée en vigueur en Chine en 2021, marque une avancée significative en matière de protection des données personnelles. Inspirée du RGPD, la PIPL accorde aux citoyens chinois des droits similaires, tels que l'accès, la correction et la suppression de leurs données.

Sur le plan de la souveraineté des données, cette loi exige que les informations sensibles collectées en Chine soient stockées localement, sauf dérogation expresse approuvée par les autorités. Cette **exigence de localisation des données** a contraint de nombreuses entreprises internationales, comme Apple, à ajuster leurs pratiques en investissant dans des infrastructures locales.

Les sanctions prévues par la PIPL sont particulièrement sévères, incluant des amendes allant jusqu'à 5 % des revenus annuels d'une entreprise et la possibilité de suspension de ses opérations en Chine en cas de violation grave.

Source : Geraldine O Mbah, Data privacy in the era of AI: Navigating regulatory landscapes for global*, International Journal of Science and Research Archive*, 2024, 13(02), 2040–2058

*Businesses*, décembre 2024

# INTERNATIONAL : AUSTRALIE

*Par Vincent Lacomme*

## 1. L'évolution des règles d'éthique et de déontologie des experts-comptables australiens

En Australie, une révision du code de déontologie des comptables australiens a pour but d'intégrer les évolutions technologiques récentes et d'accentuer l'importance du jugement professionnel.

Le code de déontologie des experts-comptables australiens, membres de CPA Australia, est revu depuis le 15/12/2024 avec l'ajout de :

- Ligne de conduite dans l'utilisation des technologies
- Bonnes pratiques relatives aux principes de confidentialité

Plus d'informations : la revue australienne In the Black de CPA Australia consacre fréquemment des articles sur le sujet.

Voir : https://intheblack.cpaaustralia.com.au/ethics-and-governance/generative-ai-in-business-navigate-ethics

Cela s'est matérialisée par une nouvelle norme « APES 110 Code of Ethics for Professional Accountants » établie par l'APESB ( organisme indépendant basé en Australie qui élabore, émet et surveille les normes professionnelles et éthiques applicables aux comptables).

Cette norme est applicable au 01/01/2025.

**Détail de la norme : https://apesb.org.au/artificial-intelligence-and-digital-technology/**

https://apesb.org.au/wp-content/uploads/2024/06/APES_110_AS_Technology.pdf

**Détail des modifications :**

https://apesb.org.au/wp-content/uploads/2024/06/TU_2024_1_APES_110_AS_Technology.pdf

## 2. Témoignages

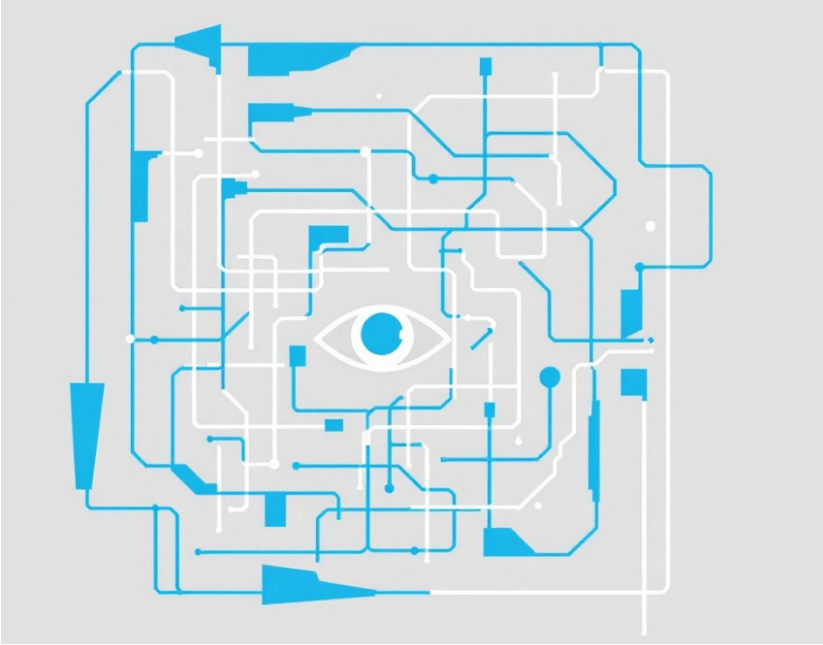
# « Fermer la cage après que les oiseaux se sont envolés. »

Témoignage de Daniel Arnephy - cabinet ACCRU (Melbourne - Australie), directeur d'un cabinet d'environ 90 salariés

*Interviewé par Vincent Lacomme le 14/11/2024*

*Durant une interview accordée en visioconférence, Daniel nous a partagé sa vision. Le point de vue australien semble rejoindre les craintes des professionnels français sur la confidentialité des données. Un regard perspicace et pragmatique.*

*Nous reproduisons avec son accord les points clés de cet entretien. Remerciements à Daniel Arnephy.*



**Les risques autour de l'IA**

**Vincent :** Quels sont, selon vous, les principaux risques liés aux données lorsqu'on utilise l'IA générative ?

**Daniel :** Le principal risque concerne l'intégration d'informations confidentielles dans des systèmes publics, comme ceux d'OpenAI. Une fois les données introduites, elles deviennent une partie d'un système sur lequel vous n'avez plus de contrôle. Concernant les résultats produits, il est possible que l'information ne soit pas correcte à 100 %, on parle souvent d' "hallucinations". Cette boîte noire ne permet pas de reconstituer la piste de la réponse.

**Les bonnes pratiques pour assurer la confidentialité**

**Vincent :** Quelles bonnes pratiques recommandez-vous pour assurer la confidentialité et la sécurité des données ?

**Daniel :** Nous avons une politique stricte : il est interdit d'introduire des données de clients dans ces programmes. Cette politique a été formalisée il y a quelques mois dans une charte. La règle est claire : pas de données clients, même anonymisées. De surcroît, nous préférons que nos employés utilisent l'IA uniquement pour des tâches comme la rédaction ou l'amélioration de textes, mais jamais avec des informations sensibles.

*[NDLR : lors de l'entretien, Daniel a également souligné la nécessité de rappeler ces règles régulièrement et notamment aux nouvelles recrues]*

**Législation**

**Vincent** : Pensez-vous qu'il pourrait y avoir de nouvelles lois à l'avenir qui limiteraient les risques liés à l'utilisation des systèmes d'IA ? Ou est-ce inutile, comme vous l'avez mentionné, parce que ces systèmes sont déjà très avancés ?

**Daniel** : C'est une excellente question, car, d'une certaine manière, ce serait fermer la cage après que les oiseaux se sont envolés. Il est probablement trop tard pour essayer de stopper certains développements. De plus, les entreprises technologiques sont souvent en avance sur les régulateurs, ou bien elles ne se soucient tout simplement pas des réglementations et continuent de développer leurs technologies.

**Vincent :** En Australie, CPA Australia incorpore désormais l'IA dans ses règles d'éthique. Pensez-vous que cela pourrait aider les professionnels à prendre conscience des risques ?

**Daniel :** J'espère bien. Je ne sais pas si cela touchera les bonnes personnes, mais au moins, il y aura des garde-fous. Cela devrait aussi être intégré dans les parcours éducatifs, que ce soit au lycée, à l'université, ou dans le cadre des modules d'éthique pour devenir CPA *[NDLR : équivalent d'expert-comptable en Australie]*.

**Garanties de confidentialité des outils d'IA**

**Vincent :** Pensez-vous que les solutions comme ChatGPT ou Perplexity offrent des garanties suffisantes en termes de confidentialité ?

**Daniel :** Non, je pense que c'est impossible. Si le principe est que toute information introduite devient une partie de cette machine géante, il n'y a pas vraiment de moyen de garantir la confidentialité. La seule solution est de dire aux gens de ne pas introduire de données sensibles.

**Vincent** : Si l'on introduit une grande quantité de données, l'IA pourrait-elle extrapoler des informations, comme deviner qu'un cabinet est spécialisé dans la restauration ou dans une taille d'entreprises, et en comprendre la croissance et la composition de ses clients ?

**Daniel** : Oui, je pense qu'elle peut faire des choses impressionnantes dans ce domaine. Il faut être prudent sur le caractère identifiable des informations. Par exemple, il y a un très bon restaurant français à Melbourne. Si nous entrions des informations indiquant qu'il s'agit d'un restaurant français dans ce code postal, il n'y en aurait que deux ou trois, ce qui rendrait l'identification très facile. Par contre, si l'on se limite à dire que c'est un restaurant à Melbourne, cela devient moins identifiable. Mais, il faut aussi se rappeler que l'on ne sait jamais sur quoi l'IA se base pour faire des comparaisons, ni la qualité des données sur lesquelles elle est entraînée.

**L'anonymisation des données est-elle suffisante ?**

**Vincent :** Pensez-vous que l'anonymisation des données pourrait être une solution ?

**Daniel :** Cela peut aider, mais il faut s'assurer que l'anonymisation est complète. Les grandes organisations peuvent développer leurs propres modèles de langage (LLM), ce qui permet de rester dans un environnement privé. Mais, pour les IA publiques, il faut s'assurer que rien n'est identifiable.

**Vincent :** Votre entreprise pourrait-elle être intéressée par le développement de sa propre IA interne ?

**Daniel :** Probablement pas. Nous n'avons ni développeurs de logiciels, ni la capacité d'expérimenter dans ce domaine. Nous préférons nous appuyer sur des solutions existantes du marché comme Microsoft Copilot. Pour l'instant, nous évaluons combien de licences nous devrions prendre.

**Vincent** : Merci beaucoup pour votre témoignage !

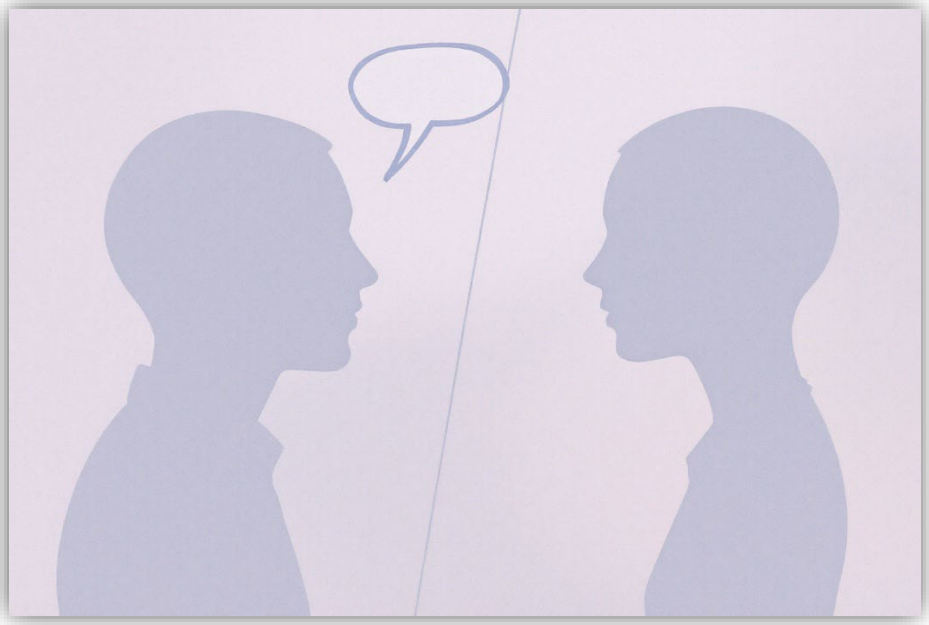
## « Nous avons la chance que la culture de la sécurité des données soit ancrée dans l'entreprise »

Témoignage de Jake Duignan associé du cabinet ACCRU Felsers (Melbourne - Australie), cabinet d'environ 70 salariés

*Interviewé par Vincent Lacomme le 20/11/2024*

Durant une interview accordée en visioconférence, Jake nous a partagé sa vision.

Nous reproduisons avec son accord les points clés de cet entretien. Remerciements à Jake Duignan.



**Vincent :** Quel est votre niveau d'utilisation de l'IA dans votre cabinet ?

**Jake :** Personnellement, j'utilise l'IA tous les jours. Principalement, j'utilise une solution appelée Perplexity[70], et je précise que je n'utilise jamais de données clients avec cet outil. Il m'aide à obtenir des statistiques, des chiffres et à effectuer des recherches fiscales. Cela me donne une bonne base pour commencer mes investigations. J'ai aussi une licence Copilot[71] que j'utilise une fois par semaine environ, particulièrement quand il s'agit de travailler sur des données clients, car c'est la seule IA en laquelle j'ai confiance pour cela.

**Vincent :** Qu'en est-il de l'utilisation de l'IA par l'ensemble de l'entreprise ?

**Jake :** Nous avons une règle très stricte : aucune donnée client ne peut être entrée dans un outil autre que Copilot. Avant Copilot, certains essayaient d'utiliser ChatGPT pour rédiger des emails, mais cela impliquait de sortir les données clients pour les réintégrer ensuite, ce qui était problématique. Je dirais que 80 % de notre personnel n'utilise pas d'IA chaque semaine, et que les 20 % restants l'utilisent principalement pour des besoins spécifiques comme le codage, les macros Excel et Python.

**Confidentialité et sécurité des données**

**Vincent :** Concernant la confidentialité, vous avez mentionné que Copilot est le seul outil dans lequel vous saisissez des données clients. Considérez-vous que Microsoft Azure est suffisamment sécurisé pour cela ?

**Jake :** Oui, exactement. La réalité est que les données de nos clients sont déjà présentes dans nos emails, sur OneDrive, ou dans des feuilles Excel. Tous ces outils font partie de l'écosystème Microsoft. Nous avons donc fait le choix d'utiliser Copilot, car cela ne représente aucun risque additionnel comparé à l'utilisation de ces autres services Microsoft.

En revanche, nous avons une politique stricte concernant les autres IA génératives comme ChatGPT. Notre politique est simple : aucune donnée client ne doit être insérée dans des outils qui ne garantissent pas le même niveau de sécurité que Microsoft.

**Vincent :** Selon vous, quels sont les principaux risques de sécurité des données lorsqu'on utilise l'IA générative aujourd'hui ?

**Jake :** Je pense que plus les gens utilisent l'IA générative pour converser, plus le risque est élevé que vos employés, ainsi que vous-même, deveniez

[70] https://www.perplexity.ai

[71] Microsoft Copilot

vulnérables. Le principal risque est l'ingénierie sociale[72].

Nous n'avons pas de barrière physique empêchant nos employés d'entrer des données sensibles dans un outil comme ChatGPT. Il n'y a que des politiques/chartes, et il y a donc un risque que certains contournent cette politique par inadvertance.

**Construire une culture de sécurité des données**

**Vincent :** Que faites-vous pour sensibiliser vos collaborateurs à ces risques ?

**Jake :** Nous avons des politiques écrites et des formations en cybersécurité. Chaque nouvel employé suit une formation d'une à deux heures avec notre responsable IT, où nous couvrons les politiques de sécurité, y compris celles sur l'IA. Nous avons également des mises à jour régulières en présentiel chaque année, ainsi qu'un suivi trimestriel en ligne. Nous avons la chance que la culture de la sécurité soit ancrée dans l'entreprise, et cela vient des associés qui en font une priorité.

**Perspectives légales et future de l'IA**

**Vincent :** Pensez-vous que les mesures légales actuelles sont suffisantes pour encadrer l'utilisation de l'IA ?

**Jake :** Non, absolument pas. Le cadre juridique est loin d'être suffisant, que ce soit pour l'IA ou même pour les cryptomonnaies. Nous sommes encore au début, et il reste énormément à faire.

**Vincent :** En dehors des aspects légaux, pensez-vous qu'il existe des solutions alternatives pour garantir la confidentialité des données, comme l'anonymisation, le chiffrement, ou des environnements de travail isolés ?

**Jake :** Pour être honnête, je suis plutôt pessimiste à ce sujet. Peu importe l'idée que vous pouvez avoir, il suffit de quelqu'un d'un peu plus malin pour trouver un moyen de la contourner. Encore une fois, cela devrait être encadré légalement. Si la loi obligeait à ce que les données soient chiffrées ou hachées, ce serait une bonne chose. Mais même si ChatGPT ou d'autres outils imposaient ces mesures, il y aura toujours quelqu'un, comme Bob au coin de la rue, qui décidera de ne pas suivre ces règles et qui utilisera son propre modèle de langage. Il y aura toujours quelqu'un pour le contourner. Donc, je suis un peu défaitiste là-dessus. Je préfère soutenir une accroche de culture d'entreprise en disant à mes équipes de ne pas mettre en péril nos données. Rien ne pourrait augmenter ma confiance en ChatGPT.

[72] NDLR : selon IBM, les attaques d'ingénierie sociale visent à manipuler les individus pour qu'ils communiquent des informations confidentielles, qu'ils téléchargent des logiciels malveillants, qu'ils se rendent sur des sites Web dangereux, qu'ils envoient de l'argent à des criminels ou qu'ils commettent d'autres erreurs qui compromettent leur sécurité personnelle ou leur organisation.

**Vincent :** Une solution alternative serait-elle d'investir dans une IA interne, installée sur vos propres serveurs ?

**Jake :** En termes de puissance de calcul, cela n'est pas faisable pour une entreprise de notre taille. Même commercialement, cela n'aurait pas de sens. À moins d'être un cabinet comme KPMG ou Deloitte, qui peuvent se permettre un tel investissement, cela reste une stratégie très coûteuse.

**Vincent** : Merci beaucoup pour votre témoignage !

# CONCLUSION

# RECAPITULATIF DES ENJEUX CLES

*Par Jean-Laurent Heim-Lienhardt*[73]

## 1. Importance de l'approche intégrée (technique et juridique)

L'utilisation quotidienne de l'intelligence artificielle (IA) dans les entreprises exige une approche intégrée qui combine des mesures techniques avancées et un cadre juridique solide. D'un point de vue technique, la sécurisation des systèmes repose sur des pratiques telles que le chiffrement des données, les contrôles d'accès, et l'implémentation de protocoles de surveillance. Sur le plan juridique, les obligations imposées par des régulations comme le RGPD et l'IA Act nécessitent la mise en place de politiques internes et de pratiques conformes.

Les systèmes d'IA, souvent complexes et interconnectés, accentuent les risques de cyberattaques, de biais algorithmique et d'utilisation abusive des données. Par conséquent, les mesures techniques doivent être conçues pour s'aligner avec les principes de gouvernance légale, notamment en matière de protection des données personnelles, de respect des droits fondamentaux, et de minimisation des risques. Une telle complémentarité garantit une protection optimale pour les entreprises et les utilisateurs.

## 2. Nécessité d'une gouvernance solide couvrant tout le cycle de vie de l'IA

Une gouvernance robuste de l'IA ne peut se limiter à des actions ponctuelles ou réactives. Elle doit s'appliquer tout au long du cycle de vie des systèmes d'IA, depuis leur conception jusqu'à leur mise en production et leur exploitation continue. Selon le Règlement européen IA Act, la gestion des systèmes à haut risque, par exemple, exige une documentation technique rigoureuse, des audits réguliers, et un suivi post-commercialisation.

[73] Contenu partiellement rédigé à l'aide d'un système d'IA générative

Dans cette logique, la mise en œuvre d'un cadre de gouvernance repose sur trois piliers :

1. **Conformité réglementaire** : Respect des exigences des régulations telles que le RGPD, NIS2, et l'IA Act.
2. **Sécurité technique** : Intégration des bonnes pratiques de cybersécurité (chiffrement, surveillance en temps réel).
3. **Responsabilité éthique** : Adhésion aux principes d'équité, de transparence et de supervision humaine pour prévenir les biais ou décisions discriminatoires.

La collaboration entre les parties prenantes (juristes, ingénieurs, gestionnaires de risques) est essentielle pour établir des politiques internes, effectuer des évaluations d'impact (DPIA), et maintenir la conformité aux évolutions légales et technologiques.

# TENDANCES REGLEMENTAIRES ET TECHNOLOGIQUES

*Par Jean-Laurent Heim-Lienhardt*[74]

## 1. Évolution probable de l'IA Act et d'autres réglementations spécifiques

L'Union européenne se positionne comme un acteur de premier plan dans la régulation de l'IA à travers l'IA Act, un cadre législatif qui repose sur une approche par niveaux de risque. L'adoption définitive de ce règlement est attendue pour 2025, et ses principaux axes incluent :

- **Classification par risque** : Les systèmes d'IA sont classés en quatre catégories : risque inacceptable (interdits), risque élevé (obligations strictes), risque limité et risque minimal. Par exemple, les systèmes d'IA à haut risque, comme ceux utilisés dans la santé ou la justice, devront se conformer à des exigences renforcées en matière de documentation, de traçabilité, et de supervision humaine.
- **Obligations de transparence** : Les utilisateurs doivent être informés lorsqu'ils interagissent avec une IA (par exemple, un chatbot) afin d'assurer une transparence accrue.
- **Supervision nationale et européenne** : Des autorités nationales superviseront l'application du règlement, tandis qu'un comité européen de l'IA coordonnera les efforts à l'échelle de l'UE.

Outre l'IA Act, la Directive NIS2 adoptée en 2022 mais non encore transposée à ce jour, qui vise à renforcer la cybersécurité dans des secteurs critiques, imposera aux entreprises des obligations supplémentaires, telles que la notification rapide des incidents et la gestion des risques dans la chaîne d'approvisionnement numérique. Ces deux cadres législatifs, combinés au RGPD, créent une base juridique robuste et interconnectée pour gérer les enjeux croissants liés à l'IA.

[74] Contenu partiellement rédigé à l'aide d'un système d'IA générative

## 2. Développement continu des technologies de cybersécurité appliquées à l'IA

Les avancées technologiques accompagnent et renforcent les efforts de régulation en matière d'IA. Plusieurs innovations prometteuses visent à garantir la sécurité et la confidentialité des systèmes :

1. **Chiffrement homomorphique** :
   Cette technique permet de traiter des données chiffrées sans jamais les déchiffrer, réduisant considérablement les risques d'exfiltration ou d'accès non autorisé.

---

***Exemple***

*Les systèmes d'analyse médicale basés sur l'IA pourraient exploiter des données sensibles tout en préservant leur confidentialité.*

---

2. **Confidential Computing** :
   En utilisant des environnements d'exécution sécurisés (TEE, Trusted Execution Environments), cette technologie protège les données en cours de traitement contre les intrusions ou manipulations malveillantes.

   Recommandé par l'ENISA (2020), le confidential computing est de plus en plus intégré dans les infrastructures critiques.

3. **Détection des anomalies via l'IA** :
   Les outils de détection d'anomalies, souvent basés sur l'apprentissage automatique, permettent d'identifier rapidement des comportements inhabituels ou malveillants, comme une tentative d'injection de données dans un modèle IA.

   Ces systèmes, couplés à des solutions de Security Information and Event Management (SIEM), renforcent la résilience des entreprises face aux menaces émergentes.

4. **Certification et standards de sécurité** :

L'introduction de normes comme ISO/IEC 27005 pour la gestion des risques liés à l'IA ou le développement de référentiels spécifiques pour les systèmes à haut risque (IA Act) reflète une volonté de structurer les bonnes pratiques de sécurité.

## 3. Convergence entre régulation et technologie

La combinaison des progrès technologiques et des évolutions réglementaires traduit une approche holistique pour sécuriser l'usage de l'IA. D'une part, les entreprises doivent anticiper et se conformer à des cadres légaux exigeants. D'autre part, elles doivent investir dans des technologies innovantes pour répondre à ces obligations de manière efficace et pragmatique.

En somme, les tendances réglementaires et technologiques actuelles mettent en évidence une synergie essentielle : une régulation robuste soutenue par des avancées techniques, pour assurer une utilisation de l'IA à la fois sécurisée, éthique et performante.

# RECOMMANDATIONS POUR LES ENTREPRISES

*Par Jean-Laurent Heim-Lienhardt*[75]

## 1. Mise en place d'une feuille de route visant la mise en conformité technique et juridique

La conformité technique et juridique en matière d'IA ne peut être laissée au hasard. Les entreprises doivent s'engager dans une démarche structurée et proactive pour intégrer ces exigences dans leurs stratégies organisationnelles. Une feuille de route bien définie devrait inclure les étapes suivantes :

1. **Évaluation initiale des risques et des pratiques existantes** :
    - Réaliser un audit exhaustif des systèmes IA en place pour identifier les vulnérabilités, les non-conformités et les lacunes organisationnelles.
    - S'appuyer sur des référentiels tels que ISO/IEC 27005 ou EBIOS Risk Manager pour cartographier les menaces techniques et juridiques.

2. **Développement d'une gouvernance interne robuste** :
    - Désigner des responsables dédiés, tels qu'un DPO (Délégué à la Protection des Données) et un référent IA, pour superviser la conformité RGPD, IA Act et NIS2.
    - Mettre en place un comité de gouvernance IA intégrant des juristes, des experts techniques, et des représentants de la direction pour valider les projets.

3. **Intégration des exigences de conformité dès la conception** :
    - Appliquer le principe de "privacy by design and by default" pour garantir la protection des données dès la phase de développement.
    - Adapter les processus DevSecOps afin d'intégrer des tests d'intrusion, des revues de code et des audits réguliers pour prévenir les failles de sécurité.

[75] Contenu partiellement rédigé à l'aide d'un système d'IA générative

4. **Documentation et traçabilité** :
    - Maintenir un registre des traitements IA, conformément au RGPD.
    - Produire des rapports techniques détaillés (qualité des données, traçabilité des algorithmes) pour les systèmes classés à haut risque par l'IA Act.

5. **Plan de réponse aux incidents** :
    - Développer une procédure claire pour détecter, contenir et signaler tout incident de sécurité, en conformité avec les exigences de notification imposées par le RGPD et NIS2.

## 2. Adoption d'une démarche de veille active (technique, réglementaire et éthique)

Dans un domaine en constante évolution, une veille active est indispensable pour anticiper les changements, ajuster les pratiques et rester conforme aux nouvelles exigences :

- **Veille réglementaire**
    - Suivre l'évolution des cadres législatifs comme l'IA Act, la Directive NIS2 ou les futures mises à jour du RGPD.
    - Collaborer avec des associations professionnelles, comme la CNIL ou l'ANSSI, pour accéder aux dernières publications, guides pratiques et recommandations.

- **Veille technologique** :
    - Identifier les nouvelles technologies susceptibles de renforcer la sécurité des systèmes IA, comme le chiffrement homomorphique ou le confidential computing.
    - Participer à des groupes de travail ou des conférences spécialisées pour rester informé des innovations et tendances du marché.

- **Veille éthique** :
    - Sensibiliser les équipes internes aux enjeux éthiques, notamment la gestion des biais algorithmiques, la transparence et la supervision humaine.

- Impliquer des comités d'éthique ou des experts indépendants pour évaluer les impacts sociétaux des projets IA.

- **Formation continue** :
  - Intégrer des sessions de formation pour les collaborateurs, adaptées à leur rôle (technique, juridique, managérial) afin d'assurer une compréhension globale et partagée des enjeux liés à l'IA.

L'élaboration d'une feuille de route structurée et l'adoption d'une veille active permettent aux entreprises non seulement de respecter leurs obligations légales, mais aussi de transformer la conformité en un levier stratégique. En investissant dans des mesures techniques avancées, en anticipant les évolutions réglementaires et en intégrant des principes éthiques à leurs pratiques, les organisations renforcent leur résilience et leur compétitivité dans un environnement de plus en plus exigeant.

# CONCLUSION

*Par Jean-Laurent Heim-Lienhardt[76] & Vincent Lacomme*

## 1. Insister sur la nécessité d'un engagement au plus haut niveau (direction)

La sécurité et la conformité dans l'utilisation de l'intelligence artificielle (IA) ne sont pas uniquement des problématiques techniques ou juridiques. Elles relèvent d'une véritable stratégie d'entreprise qui doit être portée par la direction au plus haut niveau. Cet engagement est essentiel pour :

1. **Établir une culture de la gouvernance** :
    - La direction doit définir une vision stratégique intégrant la sécurité et la conformité comme des priorités. Cela implique de communiquer clairement sur l'importance des réglementations (RGPD, IA Act, NIS2) et de soutenir leur mise en œuvre.
    - Une gouvernance solide repose sur des politiques internes claires, des processus alignés avec les standards réglementaires et des rôles bien définis au sein des équipes.
2. **Allouer des ressources adaptées** :
    - Les investissements en matière de cybersécurité et de conformité nécessitent des moyens financiers et humains. Qu'il s'agisse d'équipes techniques, juridiques ou éthiques, l'entreprise doit garantir des ressources suffisantes pour soutenir l'intégration des meilleures pratiques.
    - La formation des collaborateurs à tous les niveaux (direction, technique, juridique, opérationnel) est également un levier clé pour réduire les risques et renforcer la résilience organisationnelle.

[76] Contenu partiellement rédigé à l'aide d'un système d'IA générative

- 
3. **Mesurer et piloter les progrès** :
    - La direction doit suivre régulièrement des indicateurs de performance, comme le taux de conformité aux réglementations, le nombre de DPIA réalisées ou encore la rapidité de réponse aux incidents.
    - Ces données permettent de piloter les efforts et d'identifier les axes d'amélioration, tout en démontrant l'impact des investissements aux parties prenantes internes et externes.

## 2. Ouverture sur les futurs défis

L'engagement dans une stratégie robuste de sécurité et de conformité est un fondement essentiel, mais les défis futurs imposent une vigilance continue et une capacité d'adaptation :

- **Gestion des biais de l'IA** :
  Les biais algorithmiques restent un défi majeur, en particulier pour les systèmes utilisés dans des domaines sensibles (santé, justice, recrutement). Il sera crucial de renforcer les méthodologies de détection et de correction des biais, tout en intégrant des principes éthiques dès la conception.

- **Collaboration internationale** :
  Les données et les systèmes d'IA dépassant souvent les frontières, une collaboration accrue entre les régulateurs, les entreprises et les experts internationaux sera nécessaire. La convergence des normes européennes avec celles d'autres régions pourrait réduire les conflits réglementaires et promouvoir des standards globaux.
- **Nouveaux standards de sécurité et de transparence** :
  Avec l'évolution des menaces cyber, les technologies comme le chiffrement post-quantique ou les systèmes de supervision avancée deviendront des standards de facto. Par ailleurs, l'exigence croissante de transparence obligera les entreprises à fournir des explications claires sur les décisions prises par leurs systèmes IA.

## 3. Une vision prospective pour une IA éthique et sécurisée

Protéger l'utilisation de l'IA au quotidien repose sur une combinaison de mesures techniques, juridiques et éthiques, renforcée par un leadership engagé et visionnaire. Les entreprises doivent voir dans ces obligations non pas des contraintes, mais des opportunités d'instaurer la confiance avec leurs clients, partenaires et collaborateurs.

**Cette dynamique, soutenue par l'innovation et une anticipation proactive des défis à venir, permettra d'inscrire l'IA dans une démarche performante, conforme et éthique, tout en consolidant la résilience face aux risques émergents.**

Si le sujet de la protection des données est aujourd'hui une préoccupation majeure pour les professionnels du chiffre et du droit, nul doute qu'il ne sera pas résolu en quelques mois.

Une position trop prudente consistant à interdire totalement l'IA risquerait d'être contreproductive (*shadow IA* : avec des utilisations détournées) et nuirait à l'innovation.

Une position au contraire trop dérèglementée pourrait entrainer des risques majeurs pour les cabinets.

Un juste milieu parait nécessaire en instaurant une culture dans l'entreprise pour sensibiliser et transmettre les bonnes pratiques. Cette méthodologie doit être projetée sur le long terme.

En parallèle, des actions pour pseudonymiser voire anonymiser les données, développer des modèles de langage exécutés localement sont des pistes pour réduire le risque.

Nous ne saurions que recommander le cadre suivant défini par une étude universitaire américaine[77].

[77] Geraldine O Mbah, Data privacy in the era of AI: Navigating regulatory landscapes for global*, International Journal of Science and Research Archive*, décembre 2024, 13(02), 2040–2058

**Étapes pour construire un cadre de conformité à la protection des données**

| Étape | Description |
|---|---|
| **1. Évaluer les risques liés aux données** | Réaliser des analyses d'impact sur la protection des données pour identifier les vulnérabilités potentielles. |
| **2. Définir des politiques claires** | Établir des politiques pour la collecte, le stockage, l'accès et la conservation des données, en respectant les réglementations. |
| **3. Assigner des rôles de responsabilité** | Désigner des responsables de la protection des données (ex. : DPO, stewards) pour superviser et garantir la conformité. |
| **4. Mettre en place des contrôles techniques** | Utiliser des mesures telles que le chiffrement, le contrôle d'accès basé sur les rôles (RBAC) et des outils de surveillance automatisée. |
| **5. Former les collaborateurs** | Fournir une formation continue sur les meilleures pratiques en matière de protection des données et les réglementations émergentes. |
| **6. Surveiller et auditer régulièrement** | Réaliser des audits réguliers pour s'assurer de la conformité avec les normes en vigueur et ajuster les pratiques si nécessaire. |

# ANNEXES

# ANNEXE 1 – CHECK-LIST D'AUDIT TECHNIQUE ET JURIDIQUE POUR SE PROTEGER QUOTIDIENNEMENT FACE A L'USAGE DE L'IA

*Conçue par Jean-Laurent Heim Lienhardt*[78]

## 1. Introduction

L'usage croissant de l'intelligence artificielle (IA) dans la vie quotidienne et professionnelle soulève des enjeux importants en termes de sécurité technique et de conformité juridique. Cette check-list fournit des étapes essentielles pour auditer efficacement vos pratiques et renforcer votre protection.

## 2. Identification des actifs critiques et des systèmes utilisant l'IA

☐ **Évaluer les données collectées et traitées** :
Identifier les types de données sensibles (données personnelles, stratégiques, ou financières).

☐ **Lister les systèmes basés sur l'IA** :
Cartographier les outils et applications d'IA utilisés (internes ou fournis par des tiers).

## 3. Sécurité technique

☐ **Mises à jour régulières** :
Assurez-vous que tous les logiciels d'IA et systèmes connexes sont mis à jour pour corriger les vulnérabilités connues.

☐ **Vérification des accès** :

[78] Contenu partiellement rédigé à l'aide d'un système d'IA générative

- Implémenter des contrôles d'accès basés sur les rôles (RBAC).
- Activer l'authentification multi-facteur (MFA) pour accéder aux systèmes sensibles.

☐ **Surveillance continue** :
Installer des systèmes de détection des intrusions (IDS) et analyser les journaux pour identifier toute activité anormale.

☐ **Tests de pénétration** :
Réaliser régulièrement des audits de cybersécurité et des simulations d'attaques.

☐ **Validation des modèles** :
Évaluer les modèles d'IA pour détecter des biais ou des comportements imprévus.

## 4. Conformité juridique et éthique

☐ **Respect des règlements en vigueur** :
- RGPD (Union européenne) : Assurer la conformité des traitements de données personnelles.
- NIS/NIS 2 (non encore transposée) : Assurer la sécurité des systèmes informatiques
- IA Act : Assurer la conformité des systèmes à base d'IA
- LIL : Loi Informatique et Liberté.

☐ **Clause contractuelle** :
Vérifier que vos contrats avec les fournisseurs d'IA incluent des garanties sur la sécurité des données et les responsabilités.

☐ **Audit de la transparence** :
Exiger une documentation claire sur les algorithmes et la provenance des données utilisées.

☐ **Impact sociétal** :
Évaluer les conséquences de l'usage de l'IA (biais, discriminations potentielles, etc.).

## 5. Sensibilisation et formation

☐ **Formation continue** :
- Sensibiliser les équipes aux risques spécifiques à l'IA et à la cybersécurité.
- Intégrer des scénarios d'usage et des exercices pratiques.

☐ **Politique interne** :
Développer des politiques claires sur l'usage de l'IA et des données associées.

## 6. Suivi et amélioration continue

☐ **Revue périodique** :
Organiser des audits internes et externes à intervalles réguliers.

☐ **Tableaux de bord** :
Mettre en place des indicateurs de performance pour suivre la conformité et la sécurité.

Cette check-list (non-exhaustive) est un outil pragmatique pour permettre un usage sécurisé et éthique des systèmes d'IA. Une vigilance constante et une collaboration entre équipes techniques et juridiques demeure essentielle pour rester protégés dans cet environnement en constante évolution.

# ANNEXE 2 - EXEMPLE DE CHARTE D'UTILISATION DE L'INTELLIGENCE ARTIFICIELLE GENERATIVE

*Conçu par Sabrina Agrapart*

*Il est important de souligner que cet exemple doit être considéré comme un point de départ et qu'il devra être soigneusement adapté en fonction des besoins spécifiques, des usages particuliers et du contexte propre à chaque cabinet.*

## 1. Objectifs de la présente charte

La présente charte établit un cadre réglementaire interne visant à régir et optimiser l'utilisation de l'Intelligence artificielle générative au sein de notre entreprise.

Ses objectifs principaux sont :

- Fournir des directives claires pour encadrer l'utilisation de l'IAGen.
- Contrôler et minimiser les risques associés à son utilisation.
- Garantir le respect du cadre légal et réglementaire en vigueur.
- Assurer un niveau optimal de sécurité des données et du Système d'information.
- Promouvoir une utilisation éthique et responsable de l'IAGen.
- Définir les principes fondamentaux guidant chaque Utilisateur dans son interaction avec ces outils.
- Sensibiliser les Utilisateurs aux risques inhérents à l'utilisation de l'IAGen, tels que la divulgation involontaire d'informations confidentielles, l'exécution potentielle de commandes dangereuses, les difficultés de conformité au RGPD ou encore les nécessites de contrôles des résultats de ces IAGen.

Cette charte encourage une vigilance constante face à l'évolution rapide de ces technologies et de leurs risques associés.

## 2. Termes et définitions

**Pour plus de détail, une liste des vocabulaires associés à l'IAGen a été publiée le 6 septembre 2024 au JOF**

| Termes | Définitions |
|---|---|
| **Biais** | Tendance systématique à produire des résultats qui favorisent ou défavorisent injustement certains groupes ou idées. |
| **Données à caractère personnel** | Toute information se rapportant à une personne physique identifiée ou identifiable au sens de l'article 4 point 1 du RGPD. |
| **Grands modèles de langage**<br>**Abréviation GML** | Ou plus communément appelé suivant le terme anglicisme suivant : large language model abrégé LLM est une forme avancée d'intelligence artificielle qui est entraînée sur de grands volumes de données textuelles pour apprendre des modèles et des connexions entre les mots et les phrases. Ce GML, à partir de grands volumes de données textuelles, calcule des probabilités des enchaînements de jetons textuels en vue de la génération automatique de texte ou de code informatique. Cela permet aux LLM de comprendre et de générer des textes semblables à ceux des humains, avec un degré élevé de fluidité et de cohérence[79]. |
| **Instructions génératives** | Ou plus communément appelé prompt. Il s'agit d'une instruction ou une requête donnée à une IAGen dans un langage naturel. Il sert de point de départ pour générer une réponse ou un contenu. |
| **Intelligence artificielle générative**<br>**Abréviation IAGen** | Technologie d'intelligence artificielle capable de créer du contenu original (texte, image, audio, vidéo) en s'appuyant sur des modèles d'apprentissage automatique. L'IAGen peut produire des résultats nouveaux et créatifs à partir des données sur lesquelles elle a été entraînée |

[79] Définition de Proofpoint Qu'est-ce qu'un grand modèle de langage (LLM) ? | Proofpoint FR

| | |
|---|---|
| **RGPD** | Règlement (UE) 2016/679 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 |
| **Secret des affaires** | Toutes les informations non publiques relatives à une entreprise, qui revêtent une valeur économique ou stratégique. Il peut s'agir de données techniques, financières, commerciales ou encore organisationnelles. |
| **Système d'information** | Ensemble de composants interconnectés, d'outils et de méthodes qui permettent de recueillir, stocker, gérer et partager des informations au sein de son entreprise. |
| **Utilisateur** | Collaborateur, personne ou entité qui utilise le Système d'information de l'entreprise |

## 3. Utilisation responsable de l'outil

L'utilisation de l'IAGen ne doit pas remplacer la prise de décision humaine ni négliger l'expertise humaine et le raisonnement associé. Les Utilisateurs doivent veiller à ne pas développer une dépendance excessive à l'IAGen, qui pourrait conduire à une perte de compétences humaines. L'outil doit être utilisé comme un complément à l'expertise humaine, et non comme un substitut.

Bien que les IAGen puissent être des accélérateurs vers le principe du collaborateur augmenté et permettre un gain important de productivité, elles doivent être utilisées avec une extrême prudence.

## 4. Gestion des risques liés à l'exactitude des informations

Les Utilisateurs doivent être conscients que les réponses générées par l'IAGen peuvent être sujettes à des erreurs ou hallucinations, en raison des données de l'IAGen, et doivent être évaluées avec soin.

Une vigilance particulière est requise face au risque de génération de contenu fictif par l'IAGen. L'IAGen peut parfois créer des informations qui semblent vraies mais sont en réalité totalement inventées.

Les réponses générées par une IAGen doivent être vérifiées et validées avant d'être partagées aussi bien en interne qu'en externe.

Chaque réponse ou génération de résultats doit être évaluée au regard de l'expérience et de la compétence de l'Utilisateur.

Une recommandation peut être de n'avoir recours à l'IAGen seulement sur les sujets où vous êtes en mesure de vérifier la réponse.

## 5. Prévention des Biais (erreurs) et de la discrimination

Les Utilisateurs doivent être conscients qu'une IAGen peut reproduire les Biais présents dans les données d'entraînement.

Il est important de rester vigilant face au risque d'altération des données d'apprentissage de l'IAGen.

Il est de la responsabilité des Utilisateurs de surveiller et de corriger les réponses générées par l'IAGen pour éviter tout contenu biaisé, discriminatoire ou offensant.

## 6. Sécurité et confidentialité des données

Les données suivantes ne doivent jamais être saisies dans les Instructions génératives adressés aux outils d'IAGen :

- Données à caractère personnel (vie privée, coordonnées, identités etc…). Une pseudonymisation des Instructions génératives devra être appliquées au préalable.
- Données contractuelles, juridiques ou financières de l'entreprise, de ses fournisseurs ou de ses clients et de manière générale, des données soumises au Secret des affaires.
- Secrets informatiques comme des mots de passe,

## 7. Propriété intellectuelle

L'IAGen peut produire des contenus (textes, images, vidéos, etc.) qui peuvent porter atteinte aux droits de propriété intellectuelle des tiers. Chaque Utilisateur doit vérifier que les contenus soumis aux Instructions génératives sont libres de droits.

Il est déconseillé de demander à l'IAGen de s'inspirer du style d'un auteur spécifique pour éviter des risques de contrefaçon et de concurrence déloyale.

Les contenus générés doivent être retravaillés pour éviter leur réutilisation "tels quels" et minimiser les risques de contrefaçon.

L'Utilisateur doit éviter de citer une source d'inspiration lors des générations et retravailler les documents pour limiter la violation de la propriété intellectuelle.

## 8. Surveillance et encadrement.

L'IAGen est encore un produit en développement, tant dans ses usages que dans ses possibilités. Notre entreprise pourra, au fil de l'utilisation, définir les usages qui sont tolérés et ceux qui seront totalement prohibés, sur la base de l'analyse des risques et des retours d'expérience.

Ces règles seront mises à disposition des Utilisateurs et devront être respectées.

Il est impératif de respecter les consignes sur l'utilisation de l'IAGen établies par l'entreprise.

Une liste d'utilisation (i) encouragée, (ii) avec de fortes précautions ou (iii) alors à éviter peut-être mise en place[80].

[80] ES250008_PREMS 005519 FRA Charte ethique CEPEJ WEB A5.pdf

# GLOSSAIRE

*Par Virginie Bensoussan-Brulé et Jean-Laurent Heim-Lienhardt*[81]

Une partie des définitions ci-dessous sont issues du site de l'ANSSI et du site la CNIL.

**A**

***AI Act* (Règlement sur l'Intelligence Artificielle)**

Proposition de règlement européen qui établit un cadre légal pour l'utilisation de l'IA, en classifiant les systèmes selon leur niveau de risque.

**Algorithme**

Description d'une suite d'étapes permettant d'obtenir un résultat à partir d'éléments fournis en entrée.

**Annotation**

Procédé par lequel les données sont décrites manuellement afin d'être caractérisées.

**Apprentissage actif**

Cette technique fait intervenir un opérateur pendant le processus d'apprentissage pour lui demander de qualifier certaines données. Il s'agit d'une méthode d'apprentissage semi-supervisée.

[81] Contenu partiellement rédigé à l'aide d'un système d'IA générative

**Apprentissage auto-supervisé**

Méthode d'apprentissage automatique où un modèle extrait de l'information à partir de données non étiquetées, en créant ses propres tâches de supervision : l'algorithme sépare les données en différentes parties, utilisant certaines pour créer des prédictions et d'autres pour évaluer ces prédictions, s'améliorant sans supervision initiale.

**Apprentissage automatique**

Champ d'étude de l'intelligence artificielle qui vise à donner aux machines la capacité d'« apprendre » à partir de données, via des modèles mathématiques. Plus précisément, il s'agit du procédé par lequel les informations pertinentes sont tirées d'un ensemble de données d'entraînement. Le but de cette phase est l'obtention des paramètres d'un modèle qui atteindront les meilleures performances, notamment lors de la réalisation de la tâche attribuée au modèle. Une fois l'apprentissage réalisé, le modèle pourra ensuite être déployé en production.

**Apprentissage continu**

Capacité d'un système à s'améliorer et à s'adapter à mesure qu'il intègre de nouvelles données, y compris pendant sa mise en service. Dans le cas de l'apprentissage continu, la phase d'apprentissage du système se poursuit pendant le déploiement du modèle.

**Apprentissage fédéré**

Paradigme d'apprentissage dans lequel plusieurs entités entraînent collaborativement un modèle d'IA sans mise en commun de leurs données respectives. En pratique, les entités impliquées dans l'apprentissage envoient les modèles appris sur leurs données locales à un centre orchestrateur afin de consolider le modèle global.

**Apprentissage non supervisé**

Procédé d'apprentissage automatique dans lequel l'algorithme utilise un jeu de données brutes et obtient un résultat en se fondant sur la détection de similarités entre certaines de ces données.

**Apprentissage par renforcement**

Procédé d'apprentissage automatique consistant, pour un système autonome, à apprendre les actions à réaliser, à partir d'expériences, de façon à optimiser une récompense quantitative au cours du temps.

**Apprentissage par renforcement et rétroaction humaine**

Approche d'apprentissage par renforcement qui utilise les commentaires et les évaluations d'utilisateurs humains pour guider l'apprentissage d'un modèle d'intelligence artificielle. Ce type d'apprentissage est utilisé dans les générateurs de texte fondés sur les grands modèles de langue.

**Apprentissage par transfert**

En apprentissage automatique, l'apprentissage par transfert consiste à utiliser les connaissances acquises lors de l'apprentissage d'une tâche pour améliorer les performances sur une tâche analogue, généralement lorsque les données d'apprentissage sont limitées pour cette nouvelle tâche.

**Apprentissage profond (deep learning)**

Procédé d'apprentissage automatique utilisant des réseaux de neurones possédant plusieurs couches de neurones cachées. Ces algorithmes possédant de très nombreux paramètres, ils demandent un nombre très important de données afin d'être entraînés.

**Attaque adverse**

Une attaque adverse (adversial attack), parfois aussi appelée « attaque antagoniste » ou « attaque par exemples contradictoires » vise à envoyer à un système d'IA une ou plusieurs requêtes malveillantes dans le but de tromper ou d'altérer son bon fonctionnement.

**Apprentissage supervisé**

L'apprentissage supervisé est un procédé d'apprentissage automatique dans lequel l'algorithme s'entraîne à une tâche déterminée en utilisant un jeu de données assorties chacune d'une annotation indiquant le résultat attendu.

**Attaque par empoisonnement (data poisoning attack)**

Les attaques par empoisonnement visent à modifier le comportement du système d'IA en introduisant des données corrompues en phase d'entraînement (ou d'apprentissage). Elles supposent que l'attaquant soit en mesures de soumettre des données à utiliser lors de l'entraînement du système d'IA.

**Attaque par exemples contradictoires (adversarial examples attack)**

Les attaques par exemples contradictoires visent à soumettre des entrées malicieuses ou corrompues au système d'IA en phase de production. Ces entrées apparaissent, pour un humain, quasiment identiques à leurs copies non altérées. À la suite de cette attaque, qui peut être vue comme l'équivalent d'une illusion d'optique, le comportement du système d'IA est profondément altéré.

**Attaque par exfiltration de modèle (model evasion attack)**

Les attaques par exfiltration de modèle visent à permettre le vol d'un modèle d'IA et/ou de ses paramètres et hyperparamètres. Le modèle constitue un actif de grande valeur pour un système d'IA.

**Attaque par inférence d'appartenance (membership inference attack)**

Les attaques par inférence d'appartenance visent à permettre à un attaquant d'acquérir des connaissances sur les données utilisées pour la production du modèle d'IA. En pratique, il s'agit de déterminer si des données relatives à un individu ont été utilisées lors de la phase d'entraînement (ou d'apprentissage). Cette connaissance peut permettre à l'attaquant de déduire des informations concernant une personne.

**Attaque par inversion de modèle (model inversion attack)**

Les attaques par inversion visent à reconstruire les données ayant servi pour l'apprentissage du système. En pratique, les attaques par inversion sont menées en soumettant un grand nombre d'entrées au système d'IA et en observant les sorties produites. On utilise, de façon équivalente, le terme d'attaque par extraction de données (data extraction attacks).

### Augmentation de données (IA)

Dans le domaine de l'intelligence artificielle, le processus d'augmentation de données accroît la quantité de données d'entraînement par la création de nouvelles données à partir des données existantes. Cette augmentation peut être réalisée par différentes opérations, par exemple, dans le cas d'images, par translation, rotation, ajout de bruit, etc.

## B

### Bulle de filtre

Phénomène principalement observé sur les réseaux sociaux où les algorithmes de recommandation – qui alimentent par exemple les fils d'actualité des publications susceptibles d'intéresser les utilisateurs– peuvent parfois ne proposer que des contenus similaires entre eux. Ce phénomène intervient lorsqu'un algorithme est paramétré pour ne proposer que des résultats correspondant aux goûts connus d'un utilisateur, il ne sortira alors jamais des catégories connues.

## C

### Calcul multipartite sécurisé

Le calcul multipartite sécurisé (en anglais, « secure multi-party computation ») est une branche de la cryptographie permettant à plusieurs entités (ou parties) de calculer conjointement une fonction sur leurs données respectives, sans que celles-ci ne soient divulguées aux autres participants et tout en étant assuré que le résultat est exact.

**Caractéristique (IA)**

Dans le domaine de l'intelligence artificielle, la caractéristique (feature en anglais) est la variable utilisée pour représenter une propriété définie d'une entité ou d'un objet. Il peut s'agir d'informations relatives à la forme, la texture, ou encore à la couleur d'une image. Dans le cas d'un fichier audio, à la hauteur des sons, au timbre ou au tempo.

**Chiffrement homomorphe**

Le chiffrement homomorphe est une technique de cryptographie permettant de réaliser des opérations sur des données chiffrées sans que celles-ci aient à être déchiffrées. Le résultat de ces opérations reste sous forme chiffrée et ne peut être déchiffré que par les destinataires autorisés (détenant la clé de déchiffrement). Cette technique permet ainsi aux participants d'un calcul de garder leurs données confidentielles au cours d'un calcul.

**Classification (IA)**

La classification est une méthode de catégorisation qui consiste à attribuer une classe ou catégorie à une entrée qui lui est soumise en fonction de sa proximité à la classe en question selon des critères bien choisis.

**CNIL (Commission Nationale de l'Informatique et des Libertés)**

Autorité française de régulation et de contrôle des données personnelles.

**Confidentialité différentielle**

La confidentialité différentielle désigne une propriété mathématique assurant que la présence ou l'absence d'un individu dans une base de données n'affecte pas le résultat obtenu par un processus d'anonymisation appliqué à celle-ci. Pour l'atteindre, l'ajout d'un bruit spécifique est généralement nécessaire (c'est-à-dire une perturbation des données). Cela implique de détériorer l'utilité et la qualité des données.

**Contrôle itératif de l'apprentissage**

Cette technique vise à étudier l'impact de chaque donnée sur le fonctionnement du modèle. On parle également de défense RONI (pour Reject On Negative Impact en anglais) permettant de supprimer du jeu d'apprentissage les données ayant un impact négatif sur la précision du modèle.

**Couche de neurones**

Organisation des neurones dans un réseau. Il n'y a pas de connexion entre les neurones d'une même couche : les connexions ne se font qu'avec les neurones de la couche suivante. Généralement, chaque neurone d'une couche est lié avec tous les neurones de la couche en aval et celle-ci uniquement.

On appelle couche d'entrée l'ensemble des neurones d'entrée et couche de sortie l'ensemble des neurones de sortie. Les couches intermédiaires n'ont pas de contact avec l'extérieur et sont donc nommées couches cachées.

**Critère d'arrêt (IA)**

Élément de contrôle de l'évolution d'un algorithme d'apprentissage automatique qui permet, s'il est atteint, d'arrêter le processus itératif.

**D**

**Déployeur**

Une personne physique ou morale, une autorité publique, une agence ou un autre organisme utilisant sous sa propre autorité un système d'IA sauf lorsque ce système est utilisé dans le cadre d'une activité personnelle à caractère non professionnel.

**Dérive des données**

Variation des données utilisées à l'étape de production par rapport aux données qui ont été utilisées pour tester et valider le modèle avant son déploiement. Plusieurs facteurs peuvent entraîner cette dérive : des modifications de processus en amont, des problèmes de qualité des données, de soudains changements dans les données, etc.

**Dérive du modèle**

La dérive du modèle est la perte d'adéquation entre un modèle et la tâche qu'il doit accomplir. Cette dérive peut résulter d'un réapprentissage du modèle ou d'une évolution de l'environnement dans lequel il s'applique (domaine d'emploi).

**Désapprentissage machine**

Le désapprentissage machine est une technique liée à l'apprentissage automatique visant à faire disparaître une information des connaissances apprises par un modèle d'IA. L'objectif de ces techniques est de supprimer une information contenue dans un modèle sans avoir à ré-entraîner ce dernier.

**Destination**

L'utilisation à laquelle un système d'IA est destiné par le fournisseur, y compris le contexte et les conditions spécifiques d'utilisation, telles que précisées dans les informations communiquées par le fournisseur dans la notice d'utilisation, les indications publicitaires ou de vente et les déclarations, ainsi que dans la documentation technique.

**Domaine d'emploi (IA)**

Dans le domaine de l'intelligence artificielle, le domaine d'emploi est la description de l'environnement et de la population visée par le procédé d'apprentissage automatique.

**Donnée brute (IA)**

Dans le domaine de l'intelligence artificielle, une donnée brute est une donnée n'ayant subi aucune transformation depuis son observation initiale.

**Données à caractère non personnel**

Les données autres que les données à caractère personnel au sens de l'article 4, point 1), du règlement (UE) 2016/679.

**Données à caractère personnel**

Les données à caractère personnel définies à l'article 4, point 1), du règlement (UE) 2016/679.

**Données biométriques**

Les données à caractère personnel résultant d'un traitement technique spécifique, relatives aux caractéristiques physiques, physiologiques ou comportementales d'une personne physique, telles que des images faciales ou des données dactyloscopiques.

**Données d'entraînement**

Les données utilisées pour entraîner un système d'IA en ajustant ses paramètres entraînables.

**Données d'entrée**

Les données fournies à un système d'IA ou directement acquises par celui-ci et à partir desquelles il produit un résultat.

**Donnée de sortie (IA)**

Dans le domaine de l'intelligence artificielle, une donnée de sortie est une valeur représentant tout ou partie de l'opération effectuée par le système d'IA à partir des données d'entrée.

**Données de test**

Les données utilisées pour fournir une évaluation indépendante du système d'IA afin de confirmer la performance attendue de ce système avant sa mise sur le marché ou sa mise en service.

**Données de validation**

Les données utilisées pour fournir une évaluation du système d'IA entraîné et pour régler ses paramètres non entraînables ainsi que son processus d'apprentissage, afin, notamment, d'éviter tout sous-ajustement ou surajustement.

**DPIA (Data Protection Impact Assessment)**

Analyse d'impact sur la protection des données, obligatoire pour les traitements présentant un risque élevé pour les droits des personnes.

**E**

**Échantillon (IA)**

Fraction représentative d'une population ou d'un univers statistique.

**ENISA (European Union Agency for Cybersecurity)**

Agence européenne chargée de renforcer la cybersécurité dans l'Union européenne.

**Ensemble de test (IA)**

Jeu de données utilisé lors de la phase de test.

**Ensemble de validation (IA)**

Jeu de données utilisé lors de la phase de validation.

**Ensemble d'entraînement/d'apprentissage**

Jeu de données (texte, sons, images, listes, etc.) utilisé lors de la phase d'entrainement / d'apprentissage : le système s'entraîne sur ces données pour effectuer la tâche attendue de lui.

**Entraînement (ou apprentissage)**

L'entraînement est le processus de l'apprentissage automatique pendant lequel le système d'intelligence artificielle construit un modèle à partir de données.

**Environnement d'exécution de confiance**

Les Trusted Execution Environnement, ou environnement d'exécution de confiance, sont des zones sécurisées et isolées des autres environnements d'exécution situés dans un processeur. Les TEE garantissent que des données confidentielles restent stockées, traitées et protégées dans un environnement de confiance.

**Estimation bayésienne**

L'estimation ou inférence bayésienne s'appuie sur un théorème énoncé par le mathématicien Thomas Bayes. Ce théorème donne une méthode pour calculer la probabilité d'un phénomène grâce à la connaissance de certaines informations. L'estimation bayésienne est donc la méthode qui s'appuie sur ce raisonnement.

**Estimation de poses**

Technique de vision par ordinateur (« computer vision ») qui permet de détecter la posture d'une personne sur une image ou une vidéo, afin d'en extraire un modèle biomécanique.

**Evaluation de la conformité**

La procédure permettant de démontrer que les exigences relatives à un système d'IA à haut risque énoncées au chapitre II, section 2, ont été respectées.

**Explicabilité (IA)**

Dans le domaine de l'intelligence artificielle, l'explicabilité est la capacité de mettre en relation et de rendre compréhensible les éléments pris en compte par le système d'IA pour la production d'un résultat. Il peut s'agir, par exemple, des variables d'entrée et de leurs conséquences sur la prévision d'un score, et ainsi sur la décision. Les explications doivent être adaptées au niveau de compréhension de la personne auxquelles elles sont destinées.

**Extraction de caractéristiques (feature extraction)**

Etape au cours de laquelle sont induites depuis des données brutes (fichier son, image, document textuel, tableau numérique, etc.) des caractéristiques (features) sur lesquelles le système d'IA doit se reposer pour effectuer la tâche pour laquelle il est programmé. La définition de ces caractéristiques et leur nature discriminante sont essentielles.

**F**

**Fonction de perte ou de coût (loss function)**

Dans le domaine de l'intelligence artificielle, la fonction de perte ou de coût est la quantification de l'écart entre les prévisions du modèle et les observations réelles du jeu de donnée utilisé pendant l'entraînement. La phase d'entraînement vise à trouver les paramètres du modèle qui permettront de minimiser cette fonction.

**Fonction d'activation**

Dans le domaine de l'intelligence artificielle, la fonction d'activation peut être vu comme l'équivalent du « potentiel d'activation » qu'on retrouve dans les neurones biologiques. Cette fonction détermine si un neurone artificiel doit être activé ou pas et, dans le premier cas, le degré de cette activation. Il existe plusieurs fonctions apportant chacune des comportements différents (sigmoïde, tangente hyperbolique, ReLU, etc.).

**Forêts aléatoires (random forests)**

Les forêts aléatoires sont une méthode d'apprentissage automatique ensembliste, se basant sur de multiples arbres de décision entraînés sur des sous-ensembles de données légèrement différents.

**Fournisseur**

Une personne physique ou morale, une autorité publique, une agence ou tout autre organisme qui développe ou fait développer un système d'IA ou un modèle d'IA à usage général et le met sur le marché ou met le système d'IA en service sous son propre nom ou sa propre marque, à titre onéreux ou gratuit.

**Fournisseur en aval**

Un fournisseur d'un système d'IA, y compris d'un système d'IA à usage général, qui intègre un modèle d'IA, que ce modèle soit fourni par le même fournisseur ou non, et verticalement intégré ou fourni par une autre entité sur la base de relations contractuelle.

**G**

**Gabarit facial**

En matière de reconnaissance faciale, un « gabarit » désigne les mesures qui sont mémorisées lors de l'enregistrement des caractéristiques d'un visage. Les données extraites pour constituer ce gabarit sont des données biométriques au sens du RGPD (article 4-14).

**Gradient**

Le gradient est, en mathématiques, un vecteur représentant la variation d'une fonction au voisinage d'un point donné (en pratique, lorsqu'on dessine une courbe, plus le gradient est élevé, plus la « pente » de la courbe est forte). Appliqué au cas de l'apprentissage d'un modèle d'IA, le gradient est utilisé pour mettre en œuvre l'algorithme de descente de gradient (ou algorithme de la plus forte pente). Ce dernier permet d'obtenir un résultat optimal selon certains critères (par exemple : minimisation d'une fonction de perte), de manière itérative, c'est-à-dire par une succession d'étapes. Différentes stratégies existent pour réaliser la descente de gradient qui mobilisent des ensembles de données différents (batch gradient descent, mini-batch gradient descent, stochastic gradient descent, etc.).

**H**

**Hyperparamètre**

Élément indépendant de l'apprentissage tels que le nombre de nœuds et la taille des couches cachées du réseau de neurones, l'initialisation des poids, le coefficient d'apprentissage, la fonction d'activation, etc.

**Hypertrucage**

Une image ou un contenu audio ou vidéo généré ou manipulé par l'IA, présentant une ressemblance avec des personnes, des objets, des lieux ou d'autres entités ou événements existants et pouvant être perçu à tort comme authentique ou véridique.

**I**

**IDS/IPS (Intrusion Detection/Prevention Systems)**

Systèmes permettant de détecter et prévenir les intrusions dans les réseaux informatiques.

**Intelligence artificielle**

L'intelligence artificielle est un procédé logique et automatisé reposant généralement sur un algorithme et en mesure de réaliser des tâches bien définies. Pour le Parlement européen, constitue une intelligence artificielle tout outil utilisé par une machine afin de « reproduire des comportements liés aux humains, tels que le raisonnement, la planification et la créativité ». Plus précisément, la Commission européenne considère que l'IA regroupe :

- les approches d'apprentissage automatique ;
- les approches fondées sur la logique et les connaissances ; et
- les approches statistiques, l'estimation bayésienne, et les méthodes de recherche et d'optimisation.

**IA générative (IAGen)**

L'IA générative est un sous-ensemble de l'intelligence artificielle, axé sur la création de modèles qui sont entraînés à générer du contenu (texte, images, vidéo, etc…) à partir d'un corpus spécifique de données d'entraînement.

**J**

**Jeu de données de validation**

Un jeu de données distinct ou une partie du jeu de données d'entraînement, sous la forme d'une division variable ou fixe.

**L**

**Large language model**

Catégorie de modèles d'IA générative qui peuvent générer du texte proche du langage naturel d'un être humain, et qui sont généralement entraînés sur un large ensemble de données.

## M

### Maîtrise de l'IA

Les compétences, les connaissances et la compréhension qui permettent aux fournisseurs, aux déployeurs et aux personnes concernées, compte tenu de leurs droits et obligations respectifs dans le contexte du présent règlement, de procéder à un déploiement des systèmes d'IA en toute connaissance de cause, ainsi que de prendre conscience des possibilités et des risques que comporte l'IA, ainsi que des préjudices potentiels qu'elle peut causer.

### MFA (Multi-Factor Authentication)

Authentification multi-facteurs, méthode de sécurisation des accès reposant sur plusieurs éléments d'identification (exemple : mot de passe + code SMS).

### Modèle de fondation

Catégorie de modèles d'IA sur un ensemble de données dont la quantité et la diversité sont particulièrement importantes, dont les capacités sont générales et qui peut être adapté à une grande diversité de tâches distinctes.

### Modèle de langage

Modèle statistique de la distribution d'unité linguistiques (par exemple : lettres, phonèmes, mots) dans une langue naturelle. Un modèle de langage peut par exemple prédire le mot suivant dans une séquence de mots. On parle de modèles de langage de grande taille ou « Large Language Models » (LLM) en anglais pour les modèles possédant un grand nombre de paramètres (généralement de l'ordre du milliard de poids ou plus) comme GPT-3, BLOOM, Megatron NLG, Llama ou encore PaLM.

### Modèle discriminatif

Modèle capable de réaliser une prédiction quant à l'appartenance à une classe pour des données nouvelles sur la base d'un apprentissage réalisé auparavant sur un jeu de données d'entraînement.

### Modèle d'IA à usage général

Un modèle d'IA, y compris lorsque ce modèle d'IA est entraîné à l'aide d'un grand nombre de données utilisant l'auto-supervision à grande échelle, qui présente une généralité significative et est capable d'exécuter de manière compétente un large éventail de tâches distinctes, indépendamment de la manière dont le modèle est mis sur le marché, et qui peut être intégré dans une variété de systèmes ou d'applications en aval, à l'exception des modèles d'IA utilisés pour des activités de recherche, de développement ou de prototypage avant leur publication sur le marché.

**Modèle génératif**

Modèle défini par opposition à un modèle discriminatif. Il permet à la fois de générer de nouveaux exemples à partir des données d'entraînement et d'évaluer la probabilité qu'un nouvel exemple provienne ou ait été généré à partir des données d'entraînement.

**Modèle pré-entraîné**

Tout modèle ayant déjà été entraîné, souvent pour une tâche générale (on parle de modèles de fondation) ou pour une tâche ne correspondant pas exactement à celle visée, et utilisé comme base pour le paramétrage ou pour l'apprentissage par transfert.

**N**

**Neurone artificiel**

Un neurone artificiel fonctionne d'une manière inspirée de celle d'un neurone biologique : un nœud d'un réseau de plusieurs neurones reçoit généralement plusieurs valeurs d'entrée et génère une valeur de sortie. Le neurone calcule la valeur de sortie en appliquant une fonction d'activation à une somme pondérée des valeurs d'entrée.

**NIS/NIS2 (Network and Information Security Directives)**

Directives européennes visant à renforcer la cybersécurité des opérateurs de services essentiels et des fournisseurs de services numériques.

**O**

**Ontologie**

Modélisation d'un ensemble de données par des concepts et relations issues de connaissances dans un domaine donné (par exemple, géographie, médecine, agriculture, etc.).

**P**

**Paramètre**

Propriété apprise des données utilisées pour l'entraînement (par exemple le poids de chaque neurone d'un réseau).

**Partitionnement de données**

Le partitionnement de données (clustering en anglais) est une méthode ayant pour but de diviser un ensemble de données en différents sous-ensembles homogènes, c'est-à-dire partageant des caractéristiques communes. Ces caractéristiques reposent sur des critères de proximité définis en introduisant la notion de distance entre les objets.

**Performance d'un système d'IA**

La capacité d'un système d'IA à remplir sa destination.

**Profilage**

Le profilage au sens de l'article 4, point 4, du règlement (UE) 2016/679 ou, dans le cas des autorités répressives, au sens de l'article 3, point 4, de la directive (UE) 2016/680 ou, dans le cas des institutions, organes ou organismes de l'Union, au sens de l'article 3, point 5, du règlement (UE) 2018/1725.

**R**

**Reconnaissance d'entités nommées**

En anglais « Named-entity recognition » (NER), sous-tâche d'extraction d'informations qui cherche à localiser et classifier les mentions d'entités nommées dans du texte non structuré en catégories prédéfinies, emplacements, codes médicaux, expressions de temps, quantités, valeurs monétaires, pourcentages, etc.

**Réduction de dimension ou dimensionnalité**

Méthode permettant de diminuer la quantité d'information en ne conservant que le strict nécessaire, permettant ainsi d'obtenir plus d'efficacité en termes de résultats et de temps d'analyse. Cette réduction de l'information utile peut se faire par sélection des caractéristiques les plus pertinentes ou par création de nouvelles caractéristiques plus appropriées que celles de départ.

**Régression**

La régression est un ensemble de méthodes d'analyse statistique permettant d'approcher une variable à partir d'autres qui lui sont corrélées. En apprentissage automatique, on distingue les problèmes de régression des problèmes de classification. Ainsi, on considère que les problèmes de prédiction d'une variable quantitative sont des problèmes de régression tandis que les problèmes de prédiction d'une variable qualitative sont des problèmes de classification.

**Requête**

Une requête (ou prompt) désigne l'instruction sous forme de texte envoyée par l'utilisateur au système d'IA.

**Réseau de neurones artificiels (artificial neural network)**

Ensemble organisé de neurones interconnectés permettant la résolution de problèmes complexes tels que la vision par ordinateur ou le traitement du langage naturel. Il s'agit d'un type particulier d'algorithmes d'apprentissage automatique (comme les machines à vecteur de support (SVM en anglais), arbres de décision, K plus proches voisins, etc.) caractérisés par un grand nombre de couches de neurones, dont les coefficients de pondération sont ajustés au cours d'une phase d'entraînement (apprentissage profond). Il existe de nombreux type de réseaux de neurones artificiels tels que les réseaux de neurones récurrents, les auto-encodeurs, les réseaux transformeurs ou encore les réseaux antagonistes génératifs (generative adversarial networks).

**RGPD (Règlement Général sur la Protection des Données)**

Règlement européen qui encadre la collecte, le traitement et la protection des données personnelles au sein de l'Union européenne.

**Risque systémique**

Un risque spécifique aux capacités à fort impact des modèles d'IA à usage général, ayant une incidence significative sur le marché de l'Union en raison de leur portée ou d'effets négatifs réels ou raisonnablement prévisibles sur la santé publique, la sûreté, la sécurité publique, les droits fondamentaux ou la société dans son ensemble, pouvant être propagé à grande échelle tout au long de la chaîne de valeur.

**Robustesse**

Dans le domaine de l'intelligence artificielle, la résilience est la capacité du système à maintenir sa conformité à des exigences de performance et/ou de sécurité en présence de données d'entrée extérieures à son domaine d'emploi (par exemple en raison d'un défaut sur un capteur).

## S

### Segmentation des données

Méthode permettant la division d'un corpus de données en plusieurs ensembles (par exemple d'entraînement, de validation et de test), soit à partir de critères objectifs (date, âge, etc.) soit de manière aléatoire.

### SIEM (Security Information and Event Management)

Outil permettant de centraliser et analyser les journaux d'événements pour détecter des menaces et anomalies.

### Surapprentissage (overfitting)

Le surapprentissage entraîne un modèle qui correspond trop précisément à une collection particulière de données utilisées pour l'entrainement. Cette analyse risque de ne pas correspondre à des données utilisées en phase de production et donc de ne pas permettre une utilisation fiable du système d'IA.

### Système d'IA

Un système automatisé conçu pour fonctionner à différents niveaux d'autonomie, qui peut faire preuve d'une capacité d'adaptation après son déploiement et qui, pour des objectifs explicites ou implicites, déduit, à partir des données d'entrée qu'il reçoit, la manière de générer des résultats tels que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer les environnements physiques ou virtuels.

### Système d'IA à usage général

Un système d'IA fondé sur un modèle d'IA à usage général qui a la capacité de répondre à diverses finalités, tant pour une utilisation directe que pour une intégration dans d'autres systèmes d'IA.

### Système de reconnaissance des émotions

Un système d'IA permettant la reconnaissance ou la déduction des émotions ou des intentions de personnes physiques sur la base de leurs données biométriques.

## T

### Taux d’apprentissage (learning rate)

Facteur multiplicatif appliqué au gradient. À chaque itération, l'algorithme de descente de gradient multiplie le taux d'apprentissage par le gradient. Le taux d'apprentissage est un hyperparamètre qui joue sur la rapidité de la descente de gradient : un nombre d’itérations plus ou moins important est nécessaire avant que l’algorithme ne converge, c’est-à-dire qu’un apprentissage optimal du réseau soit réalisé.

### Test

Processus consistant à évaluer les performances d’un système et à rechercher des erreurs liées à l’exécution d’un algorithme ou d’un programme en s’appuyant sur des jeux de données d’entrée n’ayant pas été utilisés lors de la phase d’entraînement.

### Traitement automatique de la parole

Ensemble de disciplines dont l'objectif est la captation, la transmission, l'identification et la synthèse de la parole. Ces disciplines rassemblent notamment la reconnaissance de la parole, la synthèse de la parole, l'identification du locuteur ou encore la vérification du locuteur.

### Traitement automatique du langage naturel (natural language processing ou NLP)

Le traitement automatique du langage naturel est un domaine multidisciplinaire impliquant la linguistique, l'informatique et l'intelligence artificielle. Il vise à créer des outils de capable d’interpréter et de synthétiser du texte pour diverses applications.

## V

### Validation

Processus consistant à expérimenter, observer et optimiser (en modifiant les hyperparamètres notamment) le comportement du système lors de son exécution. Elle permet de s'assurer, dans le domaine d'emploi, de l'adéquation des données de sortie avec les résultats attendus.

### Vision par ordinateur (computer vision)

Branche de l'intelligence artificielle dont le principal but est de permettre à une machine d'analyser et traiter une ou plusieurs images ou vidéos prises par un système d'acquisition.

# CONCEPTS CLES

*Par Jean-Laurent Heim-Lienhardt*[82]

- **Approche DevSecOps**
  Méthodologie intégrant la sécurité dès les premières étapes du développement logiciel, particulièrement dans les projets d'IA.

- **Biais algorithmique**
  Distorsions dans les décisions ou prédictions d'un modèle d'IA, souvent causées par des données d'entraînement non représentatives ou biaisées.

- **Chiffrement homomorphique**
  Méthode permettant de traiter des données chiffrées sans avoir besoin de les déchiffrer, garantissant ainsi leur confidentialité.

- **Confidential Computing**
  Technologie sécurisant le traitement des données sensibles en protégeant leur exécution dans des environnements dédiés.

- **Data Poisoning**
  Insertion de données corrompues ou malveillantes dans le processus d'entraînement d'un modèle pour en altérer la fiabilité.

- **Évaluation d'impact réglementaire (IA Act)**
  Processus obligatoire pour les systèmes d'IA à haut risque visant à analyser les risques juridiques, éthiques et techniques associés à leur déploiement.

[82] Contenu partiellement rédigé à l'aide d'un système d'IA générative

- **Model Hacking**
  Tentative de manipulation ou de sabotage d'un modèle d'IA via des entrées malveillantes ou une modification de son fonctionnement.

- **Privacy by Design (« confidentialité dès la conception »)**
  Principe selon lequel la protection des données personnelles doit être intégrée dès la conception des systèmes et des processus.

- **Supervision humaine**
  Nécessité de maintenir une intervention humaine dans les processus d'IA pour éviter des décisions automatisées non conformes ou éthiquement problématiques.

- **Traçabilité des algorithmes**
  Documentation et enregistrement des décisions et évolutions des modèles d'IA pour assurer leur transparence et conformité.

# COMPOSITION DU GROUPE DE TRAVAIL

Nous remercions les contributeurs suivants au groupe de travail et au cahier :

**Sabrina Agrapart**, juriste en droit des affaires et contentieux FCN, DPO

**Virginie Bensoussan-Brulé**, avocate au barreau de Paris, cabinet Lexing

**Jean-Laurent Heim-Lienhardt**, expert-comptable et commissaire aux comptes

**Vincent Lacomme**, expert-comptable et formateur

**Sabine Marcellin**, consultante en conformité de l'intelligence artificielle, chargée de cours de droit de l'IA

**Romain Mirabile**, avocat

**Sylvain Navers**, RSSI Responsable de la Sécurité des Systèmes d'Information Sage

**Serge Yablonsky**, expert-comptable et commissaire aux comptes, CISA, CGEIT, CRISC

Avec la participation de :

**Mickaël Mina**, Directeur de l'IA Sage

**Camille Salinesi**, professeur des universités Systèmes d'Information, Génie Logiciel, Intelligence Artificielle, Conception

Les travaux ont été coordonnés par **Marie-Amélie CALMAO**, chargée administrative, sous la direction d'**Éric FERDJALLAH-CHEREL**, directeur de la stratégie métiers et du département des études métiers au Conseil national de l'Ordre des experts-comptables.

l'académie
SCIENCES TECHNIQUES COMPTABLES FINANCIÈRES

L'Académie est une initiative de l'Ordre des experts-comptables

ORDRE DES EXPERTS-COMPTABLES

## À l'ère de l'intelligence artificielle générative, quels garde-fous pour nos données ?

La démocratisation de l'IA générative introduit de nombreux enjeux pour protéger les données personnelles et professionnelles et en particulier pour les métiers du chiffre et du droit.

Simples d'utilisation, ces nouvelles solutions sont lourdes de conséquences en cas d'utilisation sans précautions.

Rappelant l'environnement juridique (RGPD, règlement européen sur l'IA), ce cahier répertorie les risques et enjeux lié à une utilisation des systèmes d'intelligence d'artificielle en cabinets tout en apportant des pistes de réflexion pour réduire ces risques.

De l'analyse juridique approfondie aux conseils pratiques d'implémentation, ce cahier vous offre les clés pour :

- Évaluer les risques liés à l'utilisation de l'IA générative dans votre cabinet ;
- Mettre en œuvres les bonnes pratiques pour réduire ces risques ;
- S'inspirer des pratiques dans d'autres pays.


**CONTACTS**

Académie des Sciences et Techniques Comptables et Financières

200-216, rue Raymond Losserand,
75014 Paris | Tél. +33 (0)1 44 15 64 24

www.lacademie.info

**William NAHUM**
Président fondateur

**Éric FERDJALLAH-CHEREL**
Directeur de la stratégie métiers et du département des études métiers

**Marie-Amélie CALMAO**
Chargée administrative