AI Act et RGPD 2026 : Conformité des Systèmes Agentiques IA

> Guide conformité AI Act et RGPD pour agents IA en 2026 : classification des risques, obligations GPAI, DPIA, registre des activités et sanctions.

En janvier 2026, l'autorité italienne de protection des données (Garante) a prononcé une amende de 15 millions d'euros contre une entreprise de services financiers qui avait déployé un agent IA conversationnel sans avoir réalisé d'analyse d'impact sur la protection des données (DPIA). L'agent collectait des informations personnelles sensibles, prenait des décisions automatisées sur l'éligibilité des clients à des prêts, et conservait un historique illimité des échanges sans base juridique clairement établie. Ce n'est pas un cas isolé : depuis l'entrée en vigueur partielle du règlement européen sur l'intelligence artificielle (AI Act) en août 2025, les autorités de contrôle coordonnent leurs actions sous l'égide du Comité européen de l'IA et du Comité européen de la protection des données (CEPD). En France, la CNIL (recommandations de la CNIL sur l'IA) a publié en mars 2026 ses premières recommandations opérationnelles sur l'usage des grands modèles de langage (LLM) en entreprise, insistant sur la nécessaire articulation entre l'AI Act et le RGPD. Pour les RSSI et DPO, la question n'est plus théorique : les systèmes agentiques IA — agents autonomes capables d'agir dans le monde réel, de mémoriser des contextes, d'enchaîner des outils et de prendre des décisions sans supervision humaine constante — se trouvent désormais à l'intersection de deux régimes réglementaires dont les exigences se cumulent et parfois se contredisent. Ce guide technique fait le point sur l'état du droit en juillet 2026 et propose une feuille de route opérationnelle. (voir EU AI Act officiel)

AI Act — calendrier d'application et ce qui est déjà en vigueur en 2026

Le règlement (UE) 2024/1689, dit AI Act, a été publié au Journal officiel de l'Union européenne le 12 juillet 2024 et est entré en vigueur vingt jours plus tard. Depuis lors, son déploiement suit un calendrier échelonné sur trois ans que tout responsable IA doit maîtriser parfaitement.

Février 2025 — Interdictions absolues. Les premières dispositions à s'appliquer ont été les interdictions de l'article 5. Sont désormais prohibés, avec effet immédiat et sans dérogation possible : les systèmes de manipulation subliminale contournant la conscience des personnes, les systèmes exploitant les vulnérabilités liées à l'âge ou au handicap, la notation sociale par les autorités publiques (social scoring), la reconnaissance biométrique en temps réel dans les espaces publics sauf exceptions policières strictement encadrées, et l'inférence des émotions dans les contextes professionnels et éducatifs. Toute entreprise utilisant un agent IA qui intègrerait l'une de ces fonctionnalités — fût-ce comme composante secondaire — est donc en infraction depuis plus d'un an.

Août 2025 — GPAI et gouvernance. C'est l'étape la plus significative pour les entreprises tech : les obligations relatives aux modèles d'IA à usage général (General Purpose AI, GPAI) sont entrées en vigueur. Tout fournisseur d'un GPAI mis sur le marché européen doit désormais

fournir une documentation technique détaillée, respecter les règles du droit d'auteur, publier un résumé des données d'entraînement, et se conformer aux exigences de transparence de l'article 53. Pour les GPAI à risque systémique — définis par un seuil de calcul d'entraînement supérieur à 10^{25} FLOPs, ce qui concerne GPT-4o, Claude 3 Opus, Gemini 1.5 Pro et leurs successeurs — s'y ajoutent des évaluations adversariales obligatoires, la notification des incidents graves au Bureau de l'IA, et des mesures de cybersécurité renforcées. L'Office of the AI Act (Bureau de l'IA), rattaché à la Commission européenne, est le régulateur compétent pour les GPAI ; il peut prononcer des amendes allant jusqu'à 3 % du chiffre d'affaires mondial.

Août 2026 — Systèmes à haut…

Août 2026 — Systèmes à haut risque. Dans quelques semaines, à compter du 2 août 2026, les obligations relatives aux systèmes d'IA à haut risque (Annexe III) deviendront pleinement exigibles. Cette catégorie comprend notamment : les systèmes d'évaluation du crédit, les outils de tri de CV et d'aide au recrutement, les systèmes d'aide à la décision médicale, les outils d'évaluation des élèves, et les composants de sécurité d'infrastructures critiques. Les entreprises qui déploient de tels systèmes — y compris via des agents IA — doivent impérativement finaliser leurs dossiers techniques, enregistrer leurs systèmes dans la base de données EU AI (article 71), et mettre en place une surveillance post-déploiement.

⚡ **Retour terrain**

Lors d'un accompagnement RGPD pour une chaîne de cliniques privées, le registre des traitements listait 47 traitements. Un audit approfondi en a identifié 89 supplémentaires — dont le système de vidéosurveillance des parkings, les fichiers de résultats biométriques des prestataires de restauration, et les données de géolocalisation des ambulances sous-traitées. L'invisible est souvent le plus risqué.

Août 2027 — Systèmes IA intégrés. La dernière échéance concerne les systèmes d'IA intégrés dans des produits réglementés (dispositifs médicaux, jouets, machines) qui disposaient déjà de marquages CE ; ils bénéficient d'une période transitoire supplémentaire.

En parallèle de ce calendrier réglementaire, les autorités nationales de surveillance de l'IA (désignées par chaque État membre conformément à l'article 70) ont progressivement mis en place leurs structures. En France, la CNIL a été désignée comme autorité nationale compétente pour le contrôle du marché des systèmes IA à haut risque, en coordination avec le Bureau de l'IA pour les GPAI. Cette architecture de gouvernance dual-niveau — nationale pour le haut risque, européenne pour le GPAI — crée une complexité réglementaire que les entreprises doivent gérer simultanément.

Il convient également de noter que l'AI Act prévoit des **codes de pratique** élaborés sous l'égide du Bureau de l'IA. Le premier code de pratique pour les GPAI a été finalisé en mai 2025 après une consultation de six mois impliquant plus de 1 000 parties prenantes. Son respect crée une présomption de conformité aux obligations GPAI — les entreprises qui y ont adhéré disposent donc d'un avantage procédural non négligeable en cas de contrôle.

Classification des agents IA selon l'AI Act : haut risque, limité ou minimal

La classification d'un système agentique selon l'AI Act est un exercice juridico-technique non trivial, d'autant que les agents modernes combinent souvent plusieurs composantes (LLM, outils d'action, mémoire, orchestrateur) dont chacune peut relever d'un niveau de risque différent. Voici la méthodologie à appliquer.

Étape 1 : Identifier l'usage effectif, pas l'usage déclaré. L'AI Act classe les systèmes en fonction de leur *intended purpose* (article 3(12)), mais les autorités de contrôle ont rapidement établi qu'elles examinent aussi l'usage réel. Un agent de support client qui, dans les faits, prend des décisions influençant l'accès à un service financier relève du haut risque même si son cahier des charges ne le mentionne pas.

Risque inacceptable (article 5). Comme évoqué ci-dessus, certaines fonctionnalités sont purement et simplement interdites. Un agent IA ne peut pas implémenter de profiling

comportemental visant à manipuler les décisions d'achat par des techniques subliminales, ni inférer automatiquement l'état émotionnel d'un employé pour moduler sa rémunération.

Haut risque (Annexe III). Les agents IA tombent dans cette catégorie s'ils interviennent dans l'un des huit domaines listés à l'Annexe III : infrastructures critiques, éducation, emploi, accès aux services essentiels (crédit, logement, assurance), application de la loi, gestion des frontières, administration de la justice, processus démocratiques. L'autonomie de l'agent est un facteur aggravant : un agent qui non seulement recommande mais prend effectivement une décision (envoi d'un email de refus, blocage d'un compte, génération d'un rapport médical) sera systématiquement classé haut risque dans ces domaines.

Risque limité (articles 50-52). Les agents qui interagissent directement avec des humains sans prendre de décisions substantielles sont soumis à des obligations de transparence allégées : obligation d'informer l'utilisateur qu'il interagit avec une IA (chatbots, assistants vocaux), obligation de marquage des contenus synthétiques (deepfakes, textes générés). Ces obligations sont déjà en vigueur depuis août 2025.

Risque minimal. Les systèmes sans interaction humaine directe et sans impact sur des décisions significatives (filtres anti-spam, moteurs de recommandation de contenu purement informatif) relèvent du risque minimal : aucune obligation spécifique, mais les bonnes pratiques encouragées par le code de pratique volontaire s'appliquent.

Le cas particulier des agents multi-composants…

Le cas particulier des agents multi-composants. Un pipeline agentique typique combine un LLM (GPAI), un orchestrateur (LangChain, CrewAI, AutoGen), des outils d'action (appels API, accès bases de données), et une couche mémoire (vecteurs, base de données). L'AI Act s'applique à chaque maillon de cette chaîne : le fournisseur du GPAI doit respecter les obligations GPAI, le développeur de l'orchestrateur est responsable du système dans son ensemble, et l'entreprise qui déploie l'agent est le *deployer* au sens de l'article 26. Les responsabilités se cumulent et ne s'excluent pas mutuellement.

Critères d'autonomie déterminants. Le Considérant 12 de l'AI Act précise que le degré d'autonomie est un facteur clé pour l'évaluation du risque. On distingue : (1) l'autonomie de décision — l'agent prend-il des décisions sans validation humaine ? ; (2) l'autonomie d'action — l'agent peut-il exécuter des actions dans le monde réel (fichiers, emails, transactions) sans supervision ? ; (3) l'autonomie d'apprentissage — l'agent adapte-t-il son comportement en fonction de ses interactions ? Un agent qui cumule ces trois formes d'autonomie sera presque invariablement classé haut risque si son domaine d'application touche à des intérêts significatifs des personnes.

Obligations GPAI pour les modèles de fondation : Claude, GPT-4o, Gemini

Depuis août 2025, les fournisseurs de modèles d'IA à usage général (GPAI) qui commercialisent leurs modèles dans l'Union européenne sont soumis à un ensemble d'obligations que les entreprises utilisatrices doivent connaître, car elles conditionnent leur propre conformité.

Documentation technique et transparence. L'article 53(1) exige que les fournisseurs GPAI mettent à disposition des entreprises qui construisent sur leurs modèles (les *downstream providers*) une documentation technique suffisante pour évaluer les risques. Concrètement, Anthropic (Claude), OpenAI (GPT), Google (Gemini) et leurs concurrents doivent fournir : les paramètres du modèle et les techniques d'entraînement à un niveau de détail permettant l'audit, les benchmarks de performance, les limitations connues, et une description des données d'entraînement en termes de sources, volumes et procédures de filtrage. Cette documentation est transmise au Bureau de l'IA et doit être partiellement accessible aux entreprises utilisant les modèles via API.

Respect du droit d'auteur. L'article 53(1)(c) impose aux fournisseurs GPAI de mettre en place une politique de respect du droit d'auteur et d'informer les détenteurs de droits qui ont exercé un opt-out (conformément à la directive sur le droit d'auteur dans le marché unique numérique) que leurs contenus ne sont pas utilisés pour l'entraînement. Cette obligation a des implications pratiques pour les entreprises qui fine-tunent des modèles à partir de données propriétaires : elles deviennent co-responsables si ces données comprennent du contenu tiers non licencié.

Résumé des données d'entraînement. Le résumé suffisamment détaillé (article 53(1)(d)) des données d'entraînement est une obligation qui a fait l'objet d'intenses débats pendant la phase de code de pratique. Le format retenu prévoit une description par catégories de sources (web crawl, livres, code, données synthétiques) avec les proportions approximatives, sans révéler les données elles-mêmes. Ce résumé est public et accessible dans la base de données EU AI.

Obligations supplémentaires pour les GPAI à…

Obligations supplémentaires pour les GPAI à risque systémique. Pour les modèles dépassant le seuil de 10^{25} FLOPs, s'ajoutent : l'évaluation adversariale obligatoire (red teaming) avant chaque mise à jour majeure, la notification dans les 72 heures de tout incident grave ou violation de la sécurité au Bureau de l'IA, et des mesures de cybersécurité appropriées incluant la protection contre les attaques adversariales. Ces modèles doivent également maintenir un registre des incidents accessible aux autorités de contrôle.

Implications pour les entreprises utilisatrices. Lorsqu'une entreprise construit un agent IA sur la base d'un GPAI, elle ne peut pas se contenter de s'appuyer sur la conformité du fournisseur : elle reste responsable des risques introduits par sa propre implémentation (prompts, outils, données injectées, logique d'orchestration). La règle pratique est la suivante : la conformité GPAI du fournisseur couvre le modèle de base, mais chaque couche ajoutée par l'entreprise crée de nouvelles responsabilités. Un agent qui utilise GPT-4o pour analyser des dossiers médicaux ne bénéficie pas de la conformité d'OpenAI pour cette utilisation spécifique à haut risque — c'est à l'entreprise de réaliser son propre dossier technique et sa propre DPIA.

RGPD et agents IA — les 5 points de friction principaux

Le RGPD n'a pas été conçu en pensant aux agents IA autonomes, et cette inadaptation partielle génère des zones de friction que les DPO doivent impérativement maîtriser.

Point de friction 1 : La base légale des traitements. Un agent IA qui interagit avec des utilisateurs ou traite des données personnelles doit s'appuyer sur une base légale au sens de l'article 6 RGPD. Le problème est que les agents modernes *découvrent* des données personnelles au cours de leur exécution (emails, documents, bases de données auxquels ils accèdent via des outils) sans que ces traitements aient été anticipés lors de la définition de la base légale. Le CEPD a précisé dans ses lignes directrices 02/2025 que la base légale doit être identifiée non pas pour l'agent en général mais pour chaque catégorie de traitement qu'il réalise. Une implémentation conforme doit donc maintenir un catalogue dynamique des traitements effectués par l'agent, avec leur base légale respective.

Point de friction 2 : La minimisation des données. L'article 5(1)(c) RGPD impose le principe de minimisation : seules les données strictement nécessaires à la finalité peuvent être traitées. Or, les agents IA ont tendance à aspirer davantage de données que nécessaire pour améliorer la qualité de leurs réponses (plus de contexte = meilleures performances). Cette tension est structurelle. Les architectures Privacy by Design pour agents IA imposent donc des mécanismes de *data scoping* : l'agent ne doit avoir accès qu'aux données strictement nécessaires à la tâche en cours, via des permissions granulaires qui varient selon le contexte d'exécution.

Point de friction 3 : La…

Point de friction 3 : La mémoire persistante. Les agents conversationnels modernes maintiennent une mémoire de long terme, stockant les préférences, l'historique des interactions et des informations personnelles pour personnaliser leurs réponses futures. Cette mémoire constitue un traitement de données personnelles qui doit respecter les durées de conservation (article 5(1)(e)), le droit à l'effacement (article 17), et le droit à la portabilité (article 20). Les entreprises qui déploient des agents avec mémoire persistante doivent implémenter des mécanismes permettant à l'utilisateur de consulter sa mémoire, de supprimer des éléments spécifiques, et d'exporter l'ensemble. En pratique, peu d'implémentations actuelles respectent ces exigences.

Point de friction 4 : Les transferts hors UE. La majorité des LLM performants sont fournis par des entreprises américaines (Anthropic, OpenAI, Google, Meta). Chaque appel API envoyant des données personnelles à ces modèles constitue un transfert vers un pays tiers. Les transferts vers

les États-Unis sont couverts par le Data Privacy Framework (DPF) depuis juillet 2023, mais le DPF fait l'objet d'un recours devant la Cour de justice de l'UE et son maintien à long terme est incertain. Les entreprises qui adoptent une posture de conformité robuste mettent en place des clauses contractuelles types (CCT) de secours et documentent les *Transfer Impact Assessments* (TIA) pour chaque fournisseur. Pour les données particulièrement sensibles (santé, données financières), certaines entreprises optent pour des déploiements on-premise ou des modèles open-source hébergés en Europe.

Point de friction 5 : Les décisions automatisées et le profilage. L'article 22 RGPD donne aux personnes le droit de ne pas faire l'objet de décisions fondées exclusivement sur un traitement automatisé produisant des effets juridiques ou les affectant de manière significative. Les agents IA autonomes qui prennent des décisions sans intervention humaine entrent directement dans le champ d'application de cet article. Les exceptions prévues (consentement explicite, nécessité contractuelle, autorisation légale) doivent être documentées, et dans tous les cas l'entreprise doit mettre en place des garanties appropriées incluant le droit à l'intervention humaine, à l'expression du point de vue, et à la contestation de la décision.

DPIA obligatoire pour les agents autonomes — guide pratique

L'analyse d'impact sur la protection des données (DPIA, Data Protection Impact Assessment) est obligatoire dès lors qu'un traitement est susceptible d'engendrer un risque élevé pour les droits et libertés des personnes (article 35 RGPD). Pour les agents IA, cette obligation est quasi-systématique.

Quand la DPIA est-elle obligatoire pour un agent IA ? La liste de la CNIL (délibération de 2018, mise à jour par recommandation de mars 2026) inclut explicitement : les traitements de données sensibles à grande échelle, le profilage systématique, la surveillance systématique, et les décisions automatisées produisant des effets significatifs. Un agent IA qui interagit avec des centaines d'utilisateurs par jour, maintient des profils, et peut prendre des décisions (même recommendations) coche au moins deux de ces critères. La règle prudentielle est donc : toute mise en production d'un agent IA doit être précédée d'une DPIA.

Structure d'une DPIA pour agent IA. La DPIA doit couvrir :

Description du traitement : nature, portée, contexte, finalités de l'agent ; description détaillée du pipeline (LLM utilisé, outils, mémoire, infrastructure) ; flux de données avec identification des données personnelles à chaque étape.

Évaluation de la nécessité et proportionnalité : pourquoi un agent IA plutôt qu'un outil moins intrusif ? Quelles données sont strictement nécessaires ? Quelles sont les durées de conservation prévues ?

Évaluation des risques : risques d'accès non autorisé (prompt injection, jailbreak, exfiltration via outils), risques de biais discriminatoires, risques de décisions erronées et leurs conséquences, risques liés aux transferts hors UE.

Mesures d'atténuation : contrôles d'accès aux outils, sandboxing, logging des actions, mécanismes de supervision humaine, procédures de gestion des incidents.

Consultation préalable : si les risques résiduels restent élevés après les mesures d'atténuation, consultation préalable de la CNIL obligatoire (article 36 RGPD) avant la mise en production.

Questions clés à intégrer dans la DPIA d'un agent IA. Au-delà de la structure standard, une DPIA pour agent autonome doit répondre à ces questions spécifiques :

L'agent peut-il accéder à des données…

L'agent peut-il accéder à des données personnelles au-delà de celles explicitement prévues (via recherche web, accès fichiers, appels API) ?

Comment les actions de l'agent sont-elles loguées et auditées ? Les logs sont-ils eux-mêmes protégés ?

Existe-t-il un mécanisme de *kill switch* permettant d'interrompre immédiatement l'exécution de l'agent ?

Comment les utilisateurs sont-ils informés de l'existence de l'agent et de ses capacités (obligation de transparence AI Act + RGPD) ?

Quel est le délai de réponse aux demandes de droits (accès, effacement, portabilité) pour les données traitées par l'agent ?

Comment les sous-traitants (fournisseur LLM, hébergeur infrastructure) ont-ils été qualifiés et contractualisés (DPA, CCT) ?

La DPIA n'est pas un document statique : elle doit être révisée à chaque modification significative de l'agent (nouveau modèle LLM, nouveaux outils, nouvelle population d'utilisateurs) et au minimum tous les trois ans.

Pour aller plus loin sur la sécurisation technique des agents, consultez notre guide sur la sécurité des agents IA : sandboxing et guardrails, qui détaille les mesures techniques complémentaires à documenter dans la DPIA.

Registre des activités de traitement IA — comment le structurer

L'article 30 RGPD impose à toute organisation traitant des données personnelles de tenir un registre des activités de traitement. L'émergence des agents IA complexifie considérablement cet exercice et exige une approche structurée spécifique.

Pourquoi le registre standard est insuffisant pour les agents IA. Un registre RGPD classique décrit des traitements stables et prédéfinis (paie, recrutement, CRM). Un agent IA réalise des traitements dynamiques dont la nature précise peut varier à chaque exécution selon les instructions reçues et les outils utilisés. Le registre doit donc opérer à deux niveaux : un niveau *système* (l'agent dans son ensemble) et un niveau *activité* (les catégories de traitements que l'agent est susceptible de réaliser).

Structure recommandée pour une fiche registre "agent IA". Chaque agent déployé doit faire l'objet d'une fiche registre comprenant : (1) Identification du système : nom, version, fournisseur LLM, infrastructure, responsable du système ; (2) Finalités et périmètre : cas d'usage autorisés, populations concernées, types de données personnelles accessibles ; (3) Base légale : par catégorie de traitement ; (4) Durées de conservation : des conversations, des logs d'action, de la mémoire persistante ; (5) Destinataires : fournisseur LLM (préciser DPA et base de transfert),

sous-traitants infrastructure ; (6) Mesures de sécurité : chiffrement, contrôle d'accès, monitoring ; (7) DPIA : référence et date ; (8) Droits des personnes : procédure de traitement des demandes.

Articulation avec l'AI Act. Le registre RGPD peut avantageusement être couplé avec la documentation technique requise par l'AI Act pour les systèmes à haut risque (article 11). En pratique, les équipes conformité les plus avancées maintiennent un référentiel unique "IA" qui alimente à la fois le registre RGPD et le dossier technique AI Act, réduisant les doublons et garantissant la cohérence des informations.

Pour approfondir la gouvernance globale des LLM en entreprise, notre article sur la gouvernance des LLM et conformité détaille les cadres organisationnels à mettre en place.

Sanctions et cas d'enforcement 2025-2026

L'enforcement autour de l'IA n'est plus théorique. Voici les principaux précédents qui dessinent la doctrine des autorités de contrôle.

Clearview AI — cumul RGPD + AI Act (2025). La Commission irlandaise de protection des données (DPC) a prononcé en septembre 2025 une amende de 30,5 millions d'euros contre Clearview AI, confirmant et augmentant les sanctions précédentes d'autres autorités européennes. L'affaire illustre le cumul possible : violation du RGPD (absence de base légale pour la collecte biométrique) et infraction à l'AI Act (système de reconnaissance faciale en temps réel sans autorisation). Les deux régimes ont été appliqués simultanément par coordination inter-régulateurs.

Amende italienne pour agent financier sans DPIA (janvier 2026). Mentionnée en introduction, cette affaire de 15 millions d'euros est la première amende spécifiquement liée au déploiement d'un agent IA autonome sans DPIA préalable. Le Garante a insisté sur trois manquements : absence de DPIA, décisions automatisées sur l'éligibilité au crédit sans supervision humaine (article 22 RGPD), et conservation illimitée des historiques de conversation.

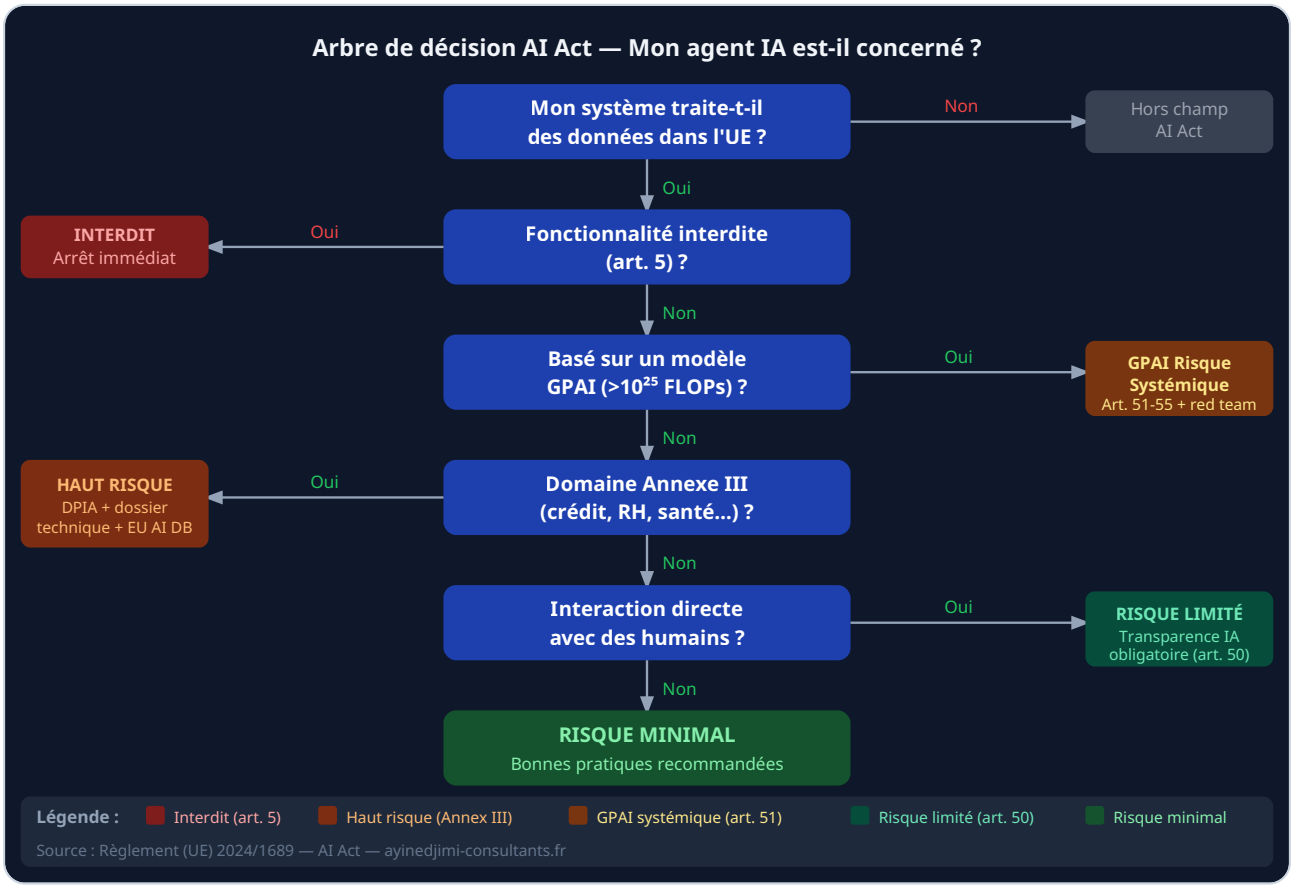
Avertissements CNIL sur les LLM en entreprise (mars 2026). La CNIL a adressé des mises en demeure à six entreprises françaises ayant déployé des assistants IA internes sans base légale

établie pour le traitement des données RH et clients. Ces mises en demeure, assorties d'un délai de trois mois pour mise en conformité, concernent des outils construits sur des API de fournisseurs américains sans DPA signé ni TIA réalisé.

Amendes AI Act Bureau de l'IA (printemps 2026). Le Bureau de l'IA a prononcé ses premières sanctions contre des fournisseurs GPAI n'ayant pas respecté leurs obligations de documentation technique (article 53). Les montants, non publiés dans leur intégralité pour raisons de confidentialité commerciale, seraient compris entre 500 000 euros et 2 % du chiffre d'affaires mondial selon des sources proches des procédures.

Leçons à retenir. Les autorités coordonnent activement leurs contrôles. La DPIA manquante est le premier point de contrôle. Les décisions automatisées sans supervision humaine constituent un risque élevé d'amende. Les transferts hors UE sans DPA ni CCT sont systématiquement sanctionnés.

Arbre de décision AI Act —…



Roadmap conformité AI Act pour RSSI et DPO

Face à la complexité réglementaire, une feuille de route structurée en cinq phases permet d'avancer méthodiquement vers la conformité sans paralyser les équipes techniques.

Phase 1 — Inventaire (J+0 à J+30). Cartographier tous les systèmes IA utilisés dans l'organisation, qu'ils soient déployés par la DSI ou adoptés en *shadow AI* par les métiers. Cet inventaire doit préciser pour chaque système : le fournisseur, le modèle de base, les données personnelles traitées, le cas d'usage, et les utilisateurs. Notre guide sur le shadow AI en entreprise détaille les méthodes de détection des usages non déclarés. L'inventaire alimente directement l'étape de classification.

Phase 2 — Classification et priorisation (J+30 à J+60). Appliquer l'arbre de décision AI Act à chaque système identifié. Prioriser les systèmes haut risque et ceux impliquant des GPAI à risque systémique. Pour chaque système classifié haut risque, déclencher immédiatement la procédure DPIA.

Phase 3 — DPIA et mise en conformité technique (J+60 à J+120). Réaliser les DPIA pour tous les systèmes qui le nécessitent. Mettre en place les mesures techniques de conformité : logging des actions agent, mécanismes d'intervention humaine, contrôles d'accès aux outils, sandboxing. Pour les transferts hors UE, signer les DPA avec les fournisseurs et réaliser les TIA. L'article sur la confidential computing pour LLM présente des solutions techniques pour les données les plus sensibles.

Phase 4 — Gouvernance et registre (J+90 à J+150). Structurer la gouvernance IA : désigner un référent IA (distinct du DPO mais en coordination étroite), mettre en place un processus de revue des nouveaux projets IA intégrant la DPIA dès la phase de conception (Privacy by Design), et enrichir le registre des activités de traitement avec les fiches "agent IA". Consulter également notre article sur la conformité RGPD des modèles IA pour les aspects spécifiques aux données d'entraînement.

Phase 5 — Surveillance continue et…

Phase 5 — Surveillance continue et mise à jour (permanent). La conformité AI Act n'est pas un projet avec une date de fin : c'est un processus continu. Mettre en place une veille réglementaire (évolutions du Bureau de l'IA, recommandations du CEPD, décisions des autorités nationales), réviser les DPIA lors de chaque modification significative des agents, et former régulièrement les équipes aux nouvelles obligations.

FAQ — Questions fréquentes

ChatGPT ou Claude en entreprise sont-ils conformes au RGPD et à l'AI Act ?

La réponse courte est : cela dépend entièrement de l'usage que vous en faites et des mesures que vous avez mises en place. OpenAI et Anthropic proposent des offres entreprise (ChatGPT Enterprise, Claude for Work/API) qui incluent un Data Processing Agreement (DPA) conforme au RGPD et qui s'engagent à ne pas utiliser vos données pour entraîner leurs modèles. Si vous utilisez ces offres avec les garanties contractuelles appropriées et que vous n'envoyez pas de données particulièrement sensibles sans analyse préalable, vous êtes dans une meilleure posture. En revanche, si vos collaborateurs utilisent les versions grand public sans DPA, chaque requête incluant des données personnelles constitue un transfert non encadré. Sur l'AI Act : ces outils sont des GPAI, les fournisseurs sont responsables de leurs obligations GPAI, mais vous restez responsable de l'usage que vous en faites — notamment si vous construisez un agent autonome par-dessus.

Un agent autonome est-il automatiquement classé "haut risque" au sens de l'AI Act ?

Non, l'autonomie seule ne suffit pas à classer un système en haut risque. Ce qui détermine la classification, c'est le domaine d'application et l'impact potentiel sur les droits des personnes. Un agent autonome qui gère un calendrier de réunions ou rédige des résumés de documents internes relève au mieux du risque limité. En revanche, un agent autonome qui décide de l'éligibilité d'un client à un crédit, trie des candidatures de recrutement, ou prend des décisions médicales sera invariablement classé haut risque — quelle que soit la sophistication de l'interface ou la précision des instructions données. Le critère déterminant est : l'agent prend-il (ou influence-t-il de

manière substantielle) des décisions qui affectent de manière significative des personnes dans un domaine listé à l'Annexe III ?

Quels sont les délais pour être en conformité AI Act en 2026 ?

En juillet 2026, voici l'état des obligations : les interdictions absolues (art. 5) s'appliquent depuis février 2025 — si vous n'êtes pas encore en conformité, vous êtes en infraction depuis plus d'un an. Les obligations GPAI (art. 51-55) s'appliquent depuis août 2025 pour les fournisseurs de modèles. Les obligations pour les systèmes à haut risque (Annexe III) deviennent pleinement exigibles le 2 août 2026 — soit dans quelques semaines. Pour les entreprises qui déploient des agents IA à haut risque, il est urgent de finaliser le dossier technique, d'enregistrer le système dans la base EU AI, et de mettre en place la surveillance post-déploiement. Les systèmes d'IA intégrés dans des produits certifiés CE ont jusqu'en août 2027. Aucun délai supplémentaire ne peut être espéré : le Bureau de l'IA a confirmé qu'il n'y aurait pas de période de grâce au-delà des transitions prévues par le règlement.

Récapitulatif AI Act : catégories, obligations, délais et sanctions

Catégorie	Exemples d'agents IA	Obligations principales	Applicable depuis / dès	Sanctions max
Risque inacceptable	Agent de notation sociale, manipulation subliminale, inférence émotionnelle au travail	Interdiction totale de mise sur le marché	Février 2025	35 M€ ou 7 % CA mondial
GPAI risque systémique	Claude Opus, GPT-4o, Gemini 1.5 Pro utilisés comme base d'agent	Évaluation adversariale, notification incidents, cybersécurité renforcée, documentation technique complète	Août 2025 (fournisseurs)	15 M€ ou 3 % CA mondial
Haut risque (Annexe III)	Agent de scoring crédit, tri CV automatisé, agent d'aide au diagnostic médical	Dossier technique, enregistrement EU AI, DPIA, logs, supervision humaine, droits contestation	Août 2026	15 M€ ou 3 % CA mondial
Risque limité	Chatbot support client, assistant virtuel, générateur de contenu	Information obligatoire "vous interagissez avec une IA", marquage contenus synthétiques	Août 2025	7,5 M€ ou 1,5 % CA mondial
Risque minimal	Agent de filtrage anti-spam, recommandation de contenu non personnalisé	Aucune obligation spécifique (bonnes pratiques encouragées)	N/A	N/A

À RETENIR

Points clés à retenir

L'AI Act s'applique par couches calendaires : interdictions (fév. 2025), GPAI (août 2025), haut risque (août 2026) — les deux premières sont déjà obligatoires.

Un agent IA autonome n'est pas automatiquement haut risque : c'est le domaine d'application qui détermine la classification, pas le degré d'autonomie.

RGPD et AI Act se cumulent sans s'exclure : un agent IA doit simultanément respecter les deux réglementations, avec des obligations parfois redondantes (DPIA + dossier technique AI Act).

La DPIA est quasi-systématiquement obligatoire pour tout agent IA interagissant avec des données personnelles : ne pas l'avoir réalisée est le premier motif de sanction identifié par les autorités.

La mémoire persistante des agents crée des obligations spécifiques : durées de conservation, droit à l'effacement, portabilité — peu d'implémentations actuelles y satisfont.

Les transferts hors UE (appels API vers fournisseurs américains) nécessitent un DPA signé, un TIA documenté, et idéalement des CCT de secours en cas de remise en cause du DPF.

La conformité n'est pas un projet ponctuel mais un processus continu : révision des DPIA à chaque évolution significative, veille réglementaire, formation des équipes.

Privacy by Design pour agents IA — principes et implémentation technique

Le principe de Privacy by Design, consacré à l'article 25 du RGPD, impose que la protection des données soit intégrée dès la conception des systèmes et non ajoutée en fin de projet. Pour les agents IA, ce principe prend une dimension particulièrement critique car les architectures agentiques sont, par nature, très difficiles à corriger après déploiement : les flux de données sont enchevêtrés, les actions peuvent être irréversibles, et les modifications du comportement d'un agent en production présentent des risques de régression fonctionnelle.

Principe 1 : Minimisation proactive des accès. Un agent IA conforme ne doit jamais disposer d'un accès global aux systèmes de l'organisation. L'architecture doit implémenter le principe du moindre privilège à chaque couche : le LLM ne reçoit que les fragments de données pertinents pour la tâche en cours (et non l'intégralité d'une base de données), les outils disponibles sont limités à ceux nécessaires au cas d'usage, et les permissions sont accordées dynamiquement selon le contexte d'exécution plutôt que de manière statique à l'initialisation. En pratique, cela implique de concevoir un système de permission granulaire qui réévalue les droits d'accès à chaque étape du raisonnement de l'agent.

Principe 2 : Séparation des données personnelles et de la logique de traitement. Les données personnelles ne doivent pas être stockées dans le contexte de l'agent plus longtemps que nécessaire à l'exécution de la tâche. Les architectures modernes utilisent des techniques de *data masking* et de *pseudonymisation* dans le contexte injecté dans le LLM : les identifiants réels (noms, numéros client, emails) sont remplacés par des pseudonymes temporaires lors de l'injection dans le prompt, et la correspondance est maintenue dans une couche séparée qui réalise le mapping au moment de l'action. Cette approche réduit considérablement le risque d'exfiltration involontaire via les sorties du modèle.

Principe 3 : Auditabilité complète des…

Principe 3 : Auditabilité complète des actions. Toute action réalisée par un agent autonome doit être loguée de manière immuable avec suffisamment de contexte pour permettre un audit a posteriori. Le log doit inclure : l'identifiant de la session, l'instruction reçue, le raisonnement de l'agent (chain of thought si disponible), les outils appelés avec leurs paramètres, les données personnelles accédées (catégorie, pas les valeurs), et le résultat. Ces logs sont eux-mêmes des données personnelles et doivent être protégés et soumis à des durées de conservation définies. Un log de six mois est généralement considéré comme proportionné pour les agents à usage interne ; douze mois pour les agents ayant un impact sur des tiers.

Principe 4 : Mécanisme de supervision humaine (*Human in the Loop*). L'article 22 RGPD et les exigences AI Act pour les systèmes à haut risque convergent vers la même exigence : les décisions significatives doivent pouvoir faire l'objet d'une revue humaine. Architecturalement, cela se traduit par l'implémentation de points de contrôle (*checkpoints*) dans le workflow de l'agent, où les actions à fort impact (envoi d'emails, modification de données clients, déclenchement de processus financiers) sont suspendues en attente d'une validation humaine explicite. Ces checkpoints doivent être configurables selon le niveau de risque de l'action : un email de prospection peut être envoyé automatiquement, mais une résiliation de contrat doit toujours passer par un opérateur humain.

Principe 5 : Gestion du cycle de vie des données de l'agent. La décommission d'un agent IA crée des obligations de suppression de données souvent sous-estimées. Outre la suppression des

logs et de la mémoire persistante, il faut considérer : les données potentiellement absorbées dans un fine-tuning du modèle de base, les embeddings vectoriels qui peuvent encoder des informations personnelles de manière non directement lisible mais potentiellement récupérable, et les données cachées dans les systems prompts conservés pour le debug. Un processus de décommission d'agent doit prévoir une procédure de suppression complète de toutes ces traces, documentée et vérifiable.

Implémentation technique : le pattern "Privacy…

Implémentation technique : le pattern "Privacy Wrapper". Une architecture concrète pour implémenter ces principes consiste à encapsuler l'agent dans un *Privacy Wrapper* — une couche middleware qui intercepte toutes les entrées et sorties de l'agent. Ce wrapper réalise : (1) la pseudonymisation des données personnelles avant injection dans le LLM, (2) la vérification que les outils appelés correspondent aux permissions autorisées, (3) l'enregistrement de chaque action dans le journal d'audit, (4) la dépseudonymisation des sorties pour les actions légitimes, et (5) la détection des tentatives d'exfiltration (outputs contenant des patterns de données personnelles non autorisées). Cette approche permet de développer la logique métier de l'agent indépendamment des contraintes de conformité, tout en garantissant que celles-ci sont systématiquement appliquées.

Sous-traitance et responsabilité partagée dans les chaînes agentiques

Les agents IA modernes reposent invariablement sur une chaîne de sous-traitants : fournisseur du modèle de base, plateforme d'orchestration, infrastructure cloud, services d'embedding, bases de données vectorielles. Chacun de ces intervenants traite potentiellement des données personnelles et doit donc être qualifié au sens du RGPD et contractualisé en conséquence.

La qualification RGPD des acteurs de la chaîne IA. L'entreprise qui déploie l'agent est presque toujours le responsable de traitement au sens de l'article 4(7) RGPD : c'est elle qui détermine les finalités et les moyens du traitement. Les fournisseurs de services (LLM API, hébergeur) sont des sous-traitants au sens de l'article 4(8), à condition qu'ils traitent les données *uniquement sur instruction* du responsable de traitement. C'est là que des difficultés apparaissent : les conditions générales de certains fournisseurs LLM prévoient une utilisation des données pour l'amélioration du modèle, ce qui les qualifie alors de co-responsables de traitement, avec des implications contractuelles et de responsabilité différentes. Vérifier les conditions contractuelles et s'assurer d'options "no training on your data" est donc une priorité de due diligence.

Les Data Processing Agreements (DPA) : ce qu'ils doivent couvrir. Tout sous-traitant qui traite des données personnelles pour le compte de votre organisation doit signer un DPA conforme à l'article 28 RGPD. Pour les fournisseurs LLM, le DPA doit spécifiquement couvrir : l'interdiction d'utiliser les données pour entraîner ou améliorer les modèles, les mesures de sécurité techniques et organisationnelles, les modalités de suppression des données après chaque requête (ou la confirmation que le modèle est stateless), les conditions de sous-traitance ultérieure (le fournisseur LLM utilise-t-il lui-même des sous-traitants ?), et les procédures de notification en cas de violation de données. La plupart des grands fournisseurs proposent des DPA standardisés ; vérifiez qu'ils couvrent tous ces points avant signature.

La responsabilité spécifique de l'orchestrateur. Dans…

La responsabilité spécifique de l'orchestrateur. Dans une architecture multi-agents ou pipeline complexe, l'orchestrateur (la couche qui coordonne plusieurs agents spécialisés) joue un rôle central en matière de conformité. C'est lui qui agrège les résultats de plusieurs agents, potentiellement en combinant des données issues de sources différentes — créant ainsi des traitements de données personnelles nouveaux qui n'existaient pas au niveau de chaque agent individuel. L'agrégation peut créer un risque de ré-identification : des données pseudonymisées à chaque niveau peuvent, une fois combinées par l'orchestrateur, permettre l'identification des

personnes. Une DPIA spécifique au niveau de l'orchestrateur est recommandée pour les pipelines multi-agents traitant des données personnelles.

Incidents et notifications. Les agents IA créent de nouveaux vecteurs d'incidents de sécurité : prompt injection entraînant une exfiltration de données, hallucination produisant des données personnelles fictives mais plausibles communiquées à des tiers, accès non autorisé via des outils détournés. L'article 33 RGPD impose de notifier la CNIL dans les 72 heures pour tout incident de sécurité affectant des données personnelles. L'article 20 de l'AI Act impose des notifications au Bureau de l'IA pour les incidents graves impliquant des systèmes à haut risque. Ces deux obligations se cumulent et nécessitent une procédure de gestion des incidents spécifiquement conçue pour les agents IA, avec des critères de qualification clairs et des canaux de notification distincts.

Pour aller plus loin dans la conformité de vos systèmes IA, consultez nos guides spécialisés : gouvernance des LLM en entreprise, RGPD et données des modèles IA, sécurité des agents IA, et confidential computing pour LLM sensibles. Pour la détection du shadow AI dans votre organisation, notre article dédié au shadow AI en entreprise vous donnera les outils nécessaires.

Sources et références

ANSSI — Référentiel de sécurité

NIST SP 800-53 — Contrôles de sécurité

ISO/IEC 27001:2022 — Norme de sécurité de l'information

