Strasbourg, 26 août 2026 T-(2025)3

COMITÉ CONSULTATIF DE LA CONVENTION POUR LA PROTECTION DES PERSONNES À L'ÉGARD DU TRAITEMENT AUTOMATISÉ DES DONNÉES À CARACTÈRE PERSONNEL

(CONVENTION 108)

Projet de directives sur la vie privée et la protection des données dans le contexte des systèmes basés sur les LLM

(Traduction automatisée)

www.coe.int/dataprotection

SOMMAIRE

I. INTRODUCTION

Ces lignes directrices abordent les risques liés à la vie privée et à la protection des données liés au cycle de vie des systèmes basés sur les LLM. Les grands modèles de langage (LLM) représentent un développement technologique particulièrement influent, combinant des capacités avancées de traitement du langage avec une automatisation à grande échelle dans tous les secteurs, amplifiant ainsi à la fois les opportunités d'innovation et les défis pour la protection des droits et libertés individuels. Ces technologies créent des opportunités significatives pour les individus, les organisations et les autorités publiques, tout en pouvant générer de sérieux risques pour la jouissance effective des droits de l'homme, le fonctionnement de la démocratie et le respect de l'État de droit.

Entraînés sur d'immenses quantités de données et intégrés dans des infrastructures numériques de plus en plus complexes, les systèmes basés sur les LLM peuvent être utilisés pour automatiser un large éventail de tâches, de la génération et synthèse de contenu au soutien des décisions et à l'automatisation des interactions. Parce que leur développement et leur fonctionnement dépendent du traitement à grande échelle des données personnelles, les LLM soulèvent des risques importants pour la vie privée et des considérations de protection des données. Leurs caractéristiques techniques, notamment l'opacité, les résultats probabilistes et l'adaptation continue, remettent en question les garanties traditionnelles de protection des données et compliquent l'application efficace des normes existantes de protection des données.

Les présentes Lignes directrices sont de nature interprétative et ne créent ni de nouveaux droits ni n'imposent d'obligations supplémentaires à celles découlant de la Convention pour la protection des individus en matière de traitement automatique des données personnelles (CETS n° 108, « Convention 108 ») et de son protocole d'amendement CETS n° 223 (désigné dans ce document comme « la Convention » ou « Convention 108+ »). Elles adoptent une approche fondée sur des principes et comprennent deux éléments complémentaires : des orientations interprétatives expliquant comment les principes et obligations de la Convention s'appliquent tout au long du cycle de vie des systèmes basés sur des LLM, et des recommandations soutenant leur mise en œuvre pratique.

Les lignes directrices visent à fournir une orientation globale, de haut niveau et complémentaire sur l'application des principes et obligations de la Convention tout au long du cycle de vie des systèmes basés sur les LLM. Lorsque les caractéristiques techniques des systèmes basés sur les LLM rendent difficile la détermination de la manière dont un principe doit être appliqué en pratique, les recommandations fournissent des illustrations non exhaustives des mesures par lesquelles les obligations existantes peuvent être remplies. Conformément au caractère technologiquement neutre des Lignes directrices, ces mesures ne constituent pas des exigences indépendantes ni ne prescrivent une technologie, une architecture ou une organisation organisationnelle spécifique. Leur sélection et leur mise en œuvre doivent être proportionnelles à la nature, à la portée et au contexte du traitement ainsi qu'aux risques posés pour les droits et libertés fondamentales des individus. Des mesures alternatives peuvent être adoptées lorsqu'elles donnent un effet équivalent aux obligations applicables et offrent un niveau de protection équivalent. La mise en œuvre d'une mesure ne démontre pas, en soi, la conformité à la Convention ; leur

conformité doit être évaluée à la lumière des obligations applicables et de l'efficacité des garanties dans leur ensemble.
Les écosystèmes LLM peuvent impliquer plusieurs [personnes et organisations/acteurs] dont les rôles et responsabilités sont répartis et interdépendants à différentes étapes du cycle de vie. Les recommandations suivent donc une approche multipartite et sont adaptées à des catégories spécifiques d'destinataires, en tenant compte de leurs responsabilités juridiques, techniques et opérationnelles respectives.

Cette approche repose également sur la perspective centrée sur l'humain et l'engagement envers l'innovation responsable reflétés dans la Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit (CETS n° 225), qui garantit que les activités du cycle de vie des systèmes d'IA sont pleinement compatibles avec les droits de l'homme, la démocratie et l'État de droit, tout en favorisant le progrès technologique et l'innovation.

II. PORTÉE ET PRINCIPES CLÉS

[Ces lignes directrices concernent le traitement des données personnelles dans le cadre de la conception, du développement, de la provision, de l'intégration, du déploiement, de l'utilisation, de la modification significative et du démantèlement des LLM et des systèmes basés sur des LLM, y compris les agents IA et les systèmes d'IA agentiques, dans la mesure où ce traitement relève de l'article 3 de la Convention 108+. Elles couvrent le traitement au niveau des modèles et des systèmes ainsi que tout au long de la chaîne de traitement pertinente, y compris les opérations intermédiaires des matériaux et le traitement effectué via des composants de récupération, la mémoire, les outils, les services externes, les communications entre agents et les dossiers opérationnels.

Elles sont adressées, selon le cas, aux Parties, autorités publiques compétentes, aux titulaires de traitement des données et aux traitants et, dans la mesure où cela concerne leurs fonctions en vertu de la loi applicable, à d'autres personnes physiques ou morales impliquées dans les étapes pertinentes du cycle de vie du système. Leurs rôles et responsabilités respectifs doivent être déterminés en fonction de leurs fonctions réelles et de leur pouvoir décisionnel en matière de traitement des données, et non uniquement par des désignations contractuelles ou des descriptions techniques du système. La répartition des composants ou fonctions entre plusieurs personnes ou entre plusieurs juridictions ne doit pas, en soi, empêcher l'identification des responsables et des traiteurs responsables de certaines opérations de traitement.

Leur application dépend du traitement effectivement effectué et des informations concernées, plutôt que de la désignation technique ou commerciale du modèle ou du système. L'absence de données personnelles dans l'entrée initiale ou la sortie finale n'établit pas, en soi, qu'aucun traitement de données personnelles n'a lieu au sein d'opérations intermédiaires ou de composants connectés.

(1) Apprentissage automatique, LLM et systèmes basés sur LLM

L'apprentissage automatique comprend des techniques de calcul par lesquelles les paramètres d'un modèle sont déterminés ou ajustés sur la base des données afin que le modèle puisse identifier des motifs ou des relations et produire des sorties en réponse aux entrées. Un modèle d'apprentissage automatique est la représentation computationnelle résultant de ce processus. L'entraînement fait référence au processus par lequel les paramètres du modèle sont déterminés ou ajustés, tandis que l'inférence modèle fait référence à l'application d'un modèle entraîné à une entrée afin de générer une sortie. La génération d'une nouvelle sortie lors de l'inférence du modèle ne signifie pas, en soi, que le modèle apprend de l'interaction ni que ses paramètres sont modifiés. Dans ce sens technique, l'inférence du modèle doit être distinguée de l'inférence des données personnelles mentionnée ailleurs dans ces Lignes directrices.

Lorsqu'un système est conçu pour mettre à jour le modèle sous-jacent pendant le fonctionnement, y compris par l'apprentissage en ligne ou en continu, cette mise à jour constitue une opération de traitement distincte et doit être identifiée, documentée et évaluée en conséquence. La formation, l'adaptation post-entraînement et l'inférence du modèle peuvent chacune impliquer le traitement de données personnelles, bien que la nature, les objectifs et les risques de ce traitement puissent différer.

Un LLM est un type de modèle d'apprentissage automatique, généralement basé sur une architecture de réseau de neurones profonde et entraîné sur de grandes quantités de texte et, le cas échéant, d'autres formes de données. Il est conçu pour traiter et générer des sorties liées au langage, y compris en prédisant des séquences de jetons. Le LLM peut fournir des capacités polyvalentes pouvant être adaptées à une variété de tâches. Le modèle doit néanmoins être distingué du système basé sur le LLM dans lequel il est intégré.

Un système basé sur un LLM comprend un ou plusieurs LLM et les composants techniques par lesquels leurs capacités sont mises à disposition ou utilisées. Selon sa conception et son objectif, ces composants peuvent inclure des instructions système, des interfaces utilisateur, des interfaces de programmation d'applications, des mécanismes de récupération, de la mémoire, des sources de données externes, des outils et services, des services d'identité et d'accréditation, des fonctions de surveillance et des mécanismes de contrôle humain. Les arrangements organisationnels régissant l'utilisation de ces composants font également partie du contexte dans lequel le traitement doit être évalué.

Le but, le contexte et les conséquences du traitement effectué par un système basé sur un LLM ne peuvent être déterminés uniquement en examinant le modèle sous-jacent ou sa sortie finale. Les risques liés à la protection des données d'entraînement ou aux propriétés du modèle, y compris la mémorisation et la divulgation involontaire, peuvent être réfléchis ou amplifiés au niveau du système. Des risques supplémentaires peuvent découler de la configuration du système, des sources de données, des intégrations, des permissions, des destinataires et de l'environnement opérationnel. Même lorsque le modèle sous-jacent reste inchangé pendant le fonctionnement, les données personnelles peuvent être traitées via des invites, du contexte récupéré, des sorties, de la mémoire à court ou long terme, des outils, des services externes et des enregistrements opérationnels.

Les systèmes basés sur les LLM peuvent, mais ne le font pas, présenter des caractéristiques agentes ; cela dépend de leur configuration fonctionnelle et du degré d'autonomie habilité au niveau système. Les caractéristiques agentiques émergent au niveau système de la manière dont un ou plusieurs modèles sont configurés et autorisés à poursuivre des objectifs spécifiés, formuler ou sélectionner des séquences d'étapes, utiliser des outils ou d'autres agents et effectuer des actions. Le comportement agent n'implique pas nécessairement que le modèle sous-jacent apprend, réentraîne ou modifie ses paramètres pendant le fonctionnement. En particulier, l'utilisation d'instructions système, de mécanismes de récupération ou de mémoire, ou l'ajustement des actions en réponse au contexte, ne doivent pas, en soi, être considérés comme un entraînement ou une modification du modèle sous-jacent.

Ces distinctions sont pertinentes pour l'application de la Convention 108+, et donc pour la portée des présentes Lignes directrices. Le fait que le modèle sous-jacent ne soit pas entraîné ou modifié pendant l'exploitation n'établit pas, comme décrit ci-dessus, qu'aucun traitement de données personnelles n'a lieu. Les données personnelles peuvent néanmoins être collectées, générées, inférées, récupérées, conservées, combinées, divulguées ou autrement traitées via des invites, des contextes récupérés, des sorties, de la mémoire, des outils, des services externes, des communications entre agents et des enregistrements opérationnels. Inversement, lorsque les données obtenues pendant l'opération sont utilisées pour mettre à jour, affiner ou réentraîner le modèle, cette utilisation constitue une opération de traitement supplémentaire dont le but, la base légitime établie par la loi et la compatibilité avec l'objectif initial doivent être évalués séparément.

La distinction entre le modèle et le système est également pertinente pour déterminer les rôles et responsabilités respectifs des personnes physiques et morales impliquées dans la chaîne de traitement des données, notamment l'identification des finalités de traitement applicables et de la base légitime, l'examen de l'impact probable mentionné à l'article 10, paragraphe 2, de la Convention 108+, la sélection des garanties appropriées et l'exercice effectif des droits des personnes concernées en vertu de l'article 9. L'évaluation de l'impact du risque de vie privée doit donc porter sur le système basé sur un LLM tel que déployé et la chaîne de traitement pertinente, et ne doit pas se limiter au modèle sous-jacent, à ses données d'entraînement ou au résultat final. La désignation technique d'un composant, ou la répartition des fonctions entre plusieurs composants ou personnes, ne doit pas obscurcir ni modifier la responsabilité des responsables de traitement et des responsables de traitement concernés.

(2) Agents d'IA et systèmes d'IA agents

Aux fins de ces Directives, un « agent IA basé sur LLM » désigne un système basé sur un LLM qui, dans la poursuite d'un objectif spécifié, peut recevoir et interpréter des informations de son environnement, formuler ou sélectionner une séquence d'étapes, invoquer des outils, services ou autres agents et effectuer des actions capables d'affecter un environnement physique ou virtuel, avec des degrés variables d'implication humaine. Un « système d'IA agentique » peut comprendre un agent ou plusieurs agents interactifs et peut inclure un orchestrateur, une mémoire à court ou long terme, des composants de récupération, des outils et des interfaces de programmation d'applications, des services d'identité et d'accréditation, des

protocoles de communication, des fonctions de surveillance et des points où une revue, une autorisation ou une intervention humaine efficace est requise ou disponible. Ces descriptions fonctionnelles ne nécessitent pas que chaque système affiche les mêmes caractéristiques ou degré d'autonomie.

La portée et la nature des actions qu'un agent est techniquement capable d'accomplir, ainsi que celles qu'il est autorisé à effectuer, doivent être distinguées de son degré d'autonomie. L'espace d'actions de l'agent concerne les systèmes, données personnelles, outils et opérations qui lui sont rendus accessibles, y compris si l'accès est en lecture seule ou permet la création, la modification, la communication, la suppression ou l'exécution. Son degré d'autonomie concerne la mesure dans laquelle il peut sélectionner des objectifs, des étapes, des moments et des moyens sans approbation humaine préalable. Les risques liés à la vie privée et à la protection des données peuvent augmenter soit avec l'étendue et les conséquences potentielles de l'espace d'action, soit avec le degré d'autonomie. Un espace d'action large ou conséquent peut engendrer des risques importants même lorsqu'une étape formelle de supervision ou d'approbation humaine est présente.

Les termes « agent » et « agence » sont utilisés dans ces lignes directrices pour décrire les caractéristiques fonctionnelles. Ils n'impliquent pas la conscience, l'intention, la personnalité juridique ni un rôle juridique indépendant. L'autonomie apparente d'un agent ne doit ni supprimer, ni diluer ni modifier les rôles et responsabilités des responsables de traitement des données et des traitements. Décrire un système comme un agent, ou accroître son autonomie technique, ne doit pas obscurcir ni modifier les obligations des personnes physiques ou morales impliquées dans le traitement. L'IA agente est un moyen de traitement et ne constitue pas, en soi, un but spécifié ni une base légitime établie par la loi pour le traitement des données personnelles.

(3) Données personnelles et applicabilité de la Convention 108+

Du point de vue de la protection des données, les données personnelles peuvent être traitées à chaque étape du cycle de vie d'un LLM ou d'un système basé sur un LLM. La nature et la gravité des risques qui en résultent dépendent, *entre autres*, des contextes d'application, de conception et de développement, ainsi que de déploiement[1]. Conformément à l'article 3 de la Convention 108+, la Convention s'applique au traitement à chaque étape du cycle de vie où ce traitement relève de son champ d'application, dans le but de garantir le droit de chaque individu à la protection des données personnelles et ainsi de contribuer au respect des droits humains et des libertés fondamentales, en particulier le droit à la vie privée.[2]

---

[1] Couvre toutes les activités, de la conception d'un système d'IA à sa retraite, quels que soient les acteurs impliqués. Rapport explicatif, Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit (CETS n° 225), paragraphe 15.

[2] Cette approche est cohérente avec l'évaluation globale et la contextualisation des divers facteurs de risque fournis par le cadre d'évaluation des risques et impacts des systèmes d'IA du Conseil de l'Europe, du point de vue des droits de l'homme, de la démocratie et de l'État de droit. Cadre d'évaluation des risques et de l'impact des systèmes d'IA du Conseil de l'Europe, du point de vue des droits de l'homme, de la démocratie et de l'État de droit (HUDERIA), p. 14.

Pour déterminer si une information constitue des données personnelles, il faut tenir compte des informations traitées et des moyens raisonnablement susceptibles d'être utilisés pour identifier une personne, y compris par combinaison avec d'autres informations. Cette évaluation doit être réalisée tout au long du cycle de vie et doit couvrir les données de formation et post-formation, les prompts et résultats, les données inférées, le contexte récupéré, la mémoire, les paramètres des outils, les dossiers opérationnels et les données divulguées ou mises à disposition d'autres agents ou services. Les désignations techniques telles que « synthétique », « désidentifié » ou « interne » ne déterminent pas, en elles-mêmes, si une information est une donnée personnelle. Les considérations relatives à des catégories particulières de données personnelles et à l'examen des impacts probables sur le risque sont abordées respectivement dans les sections IV(3) et IV(7) des présentes Directives.

III. RISQUES LIÉS À LA VIE PRIVÉE ET À LA PROTECTION DES DONNÉES TOUT AU LONG DU CYCLE DE VIE

La nature, la probabilité et la gravité des risques liés à la vie privée et à la protection des données dépendent des finalités, du contexte et de l'ampleur du traitement, des catégories de données personnelles et des personnes concernées, de l'environnement opérationnel et des garanties appliquées. Une attention particulière doit être portée aux enfants et aux personnes appartenant à des groupes vulnérables. Ces facteurs doivent être pris en compte à travers les étapes suivantes du cycle de vie :

- ***Création du modèle :*** Cette étape comprend la collecte et la préparation des données pour l'entraînement et le développement du modèle. Les ensembles de données d'entraînement peuvent contenir des données personnelles collectées sans base légitime appropriée ou au-delà de ce qui est nécessaire pour l'objectif spécifié. Les risques pertinents incluent une transparence insuffisante, une mauvaise qualité des données, la mémorisation et la divulgation involontaire, les biais et les interférences adversaires.
- ***Adaptation post-formation :*** Cette étape inclut les instructions système et l'alignement, l'ajustement fin et l'adaptation à des tâches ou à des objectifs opérationnels spécifiés en utilisant des ensembles de données ciblés et des entrées spécifiques à chaque tâche. Bien que l'identification et la qualité des données restent l'une des préoccupations principales (en raison de l'amplification du risque potentiel d'inexactitudes et de biais des données), la réaffectation du traitement des données au-delà de l'objectif initial poursuivi, le manque de transparence et de garanties représentent des défis supplémentaires et ne reflètent pas le contexte d'utilisation souhaité.
- ***Intégration système :*** Cette étape concerne l'intégration d'un LLM dans une application, un service ou un système composé, y compris par orchestration, composants de récupération et outils ou services externes. Cela peut augmenter les risques de traitement excessif, de profilage et d'inférence intercontextuels, de réidentification des personnes concernées par

combinaison de données et d'inférence d'identifiants supplémentaires, ainsi que d'accès non autorisé. Les flux transfrontaliers de données personnelles dus aux LLM hébergés dans le cloud (traitement basé sur API) nécessitent des garanties appropriées et un niveau de protection adéquat.

- ***Déploiement opérationnel et interaction utilisateur final*** : Cette étape couvre l'exploitation en direct, les interactions utilisateurs, les flux de travail autonomes, les fonctions mémoire, la surveillance et la gouvernance post-déploiement. Les risques incluent la génération ou la divulgation de données personnelles, la conservation et l'utilisation secondaire des données d'interaction, les décisions ou actions basées sur des informations inexactes, les obstacles à l'exercice des droits des sujets de données, les incidents de sécurité et l'insuffisance de responsabilité.
- ***Modification et redéploiement significatifs :*** Des modifications du modèle, des instructions système, de l'orchestration, des outils, des sources de données, de la mémoire, des permissions, du degré d'autonomie, de la composition de l'agent, des destinataires, du routage géographique ou du contexte de déploiement peuvent modifier les finalités, les moyens ou les risques du traitement. De tels changements doivent être identifiés avant leur mise en œuvre et devraient déclencher une réévaluation des risques liés à la protection des données et à la vie privée lorsqu'ils peuvent affecter la base ou les conclusions d'un examen antérieur.
- ***Désarmement et retraite :*** La fin de l'exploitation doit inclure la fin de l'accès, la révocation des identifiants et des autorisations, ainsi que le retour sécurisé, la suppression ou la conservation légale des données personnelles, mémoire et dossiers opérationnels. Les mesures doivent empêcher le traitement continu par les outils, services ou agents connectés après la mise hors service du système.

Ces caractéristiques peuvent amplifier les risques de collecte excessive, de liaison intercontextuelle et de profilage inférentiel ; mémoire persistante ou inexacte ; divulgation non autorisée via des outils ; utilisation abusive d'identité ou de crédentiel ; propagation d'erreurs entre agents ; et d'actions pouvant produire des effets significatifs ou irréversibles lorsqu'elles reposent sur des résultats probabilistes. L'accès aux comptes, fichiers, communications, capteurs ou services externes peut encore accroître l'asymétrie informationnelle et réduire la capacité pratique des sujets de données à prévoir, comprendre et exercer un contrôle sur le traitement.[3] [4]

Au-delà de cette approche basée sur le cycle de vie, les systèmes basés sur les LLM peuvent avoir des implications plus larges pour les droits humains, la démocratie et l'État de droit, nécessitant un examen approprié et, le cas échéant, une évaluation plus large de l'impact.[5] Les risques liés à la vie privée et à la protection des données

[3] Bureau du Commissaire à l'Information, ICO Tech Futures : IA agente — Protection des données et risques liés à la vie privée, 2026

[4] *Beduschi A.,* Protection des données à l'ère de l'intelligence artificielle agentique, Computer Law & Security Review, vol. 61, 2026, 106342, DOI : 10.1016/j.clsr.2026.106342.

[5] Voir le cadre d'évaluation des risques et impacts des systèmes d'IA du Conseil de l'Europe du point de vue des droits de l'homme, de la démocratie et de l'État de droit (HUDERIA); De plus, pour les implications structurelles décrites dans le Note d'orientation sur les implications de l'IA générative pour la liberté d'expression par le Comité MSI-IA du Conseil de l'Europe.

peuvent également être cartographiés au niveau des modèles et des systèmes.[6] L'intégration des systèmes basés sur les LLM avec des technologies diverses et en évolution rapide peut permettre des interactions persistantes, adaptatives, multimodales et autonomes, une surveillance continue, un profilage et une inférence avancée, augmentant ainsi les asymétries informationnelles et testant l'adéquation des garanties existantes. Ces risques doivent être pris en compte dans le cadre plus large de gestion des risques et d'impacts pour les droits humains, la démocratie et l'État de droit.[7]

/

Le LLM est un modèle d'apprentissage automatique entraîné sur une quantité massive de données et conçu pour traiter et générer des sorties de type langage naturel en prédisant des séquences de jetons. Compte tenu de ses caractéristiques techniques, il sert d'architecture fondamentale pour construire des systèmes basés sur des LLM, créant une structure relativement hiérarchique entre le modèle et le système. Un LLM fournit des capacités linguistiques à usage général pour les systèmes autonomes poursuivant des objectifs et des séquences d'actions complexes. Bien que tous les systèmes basés sur les LLM ne soient pas agentiques, ce dernier repose sur un ou plusieurs modèles d'IA, dirigés vers des objectifs spécifiques et dotés de fonctionnalités autonomes permettant la planification, la prise de décision, la mémorisation (y compris via des outils de ressources externes) et le fonctionnement dynamique en ajustant ses actions[8]. Compte tenu de cette corrélation hiérarchique, les risques liés à la vie privée émergeant au niveau du modèle se reflètent et sont amplifiés dans les systèmes.

Du point de vue de la protection des données, les données personnelles sont traitées à travers les activités des[9] modèles LLM et des systèmes basés sur les LLM, ce qui entraîne des risques liés à la vie privée et à la protection des données en tenant compte de leur contexte d'application, de conception et de développement et de contexte de déploiement[10]. Dans le but de garantir le droit individuel à la protection de ses données personnelles et de contribuer au respect de ses droits humains et libertés fondamentales, en particulier du droit à la vie privée, et conformément à l'article 3 de la Convention 108+, l'applicabilité de celle-ci s'étend à chaque phase du cycle de vie de la technologie IA.]

IV DIRECTIVES GÉNÉRALES SUR L'APPLICATION DES PRINCIPES FONDAMENTAUX DE PROTECTION DES DONNÉES ET DES ARTICLES SPÉCIFIQUES

[6] Voir le rapport d'expert sur « Risques de confidentialité et de protection des données dans les grands modèles de langage (LLM), pp. 11-20.

[7] Article 16 de la Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit (CETS n° 225).

(1) Désignation des rôles et responsabilités

Tout au long de leur cycle de vie, les systèmes LLM et basés sur des LLM fonctionnent sur d'immenses ensembles de données, ce qui crée une chaîne d'activités de traitement des données. L'identification des acteurs pertinents tout au long de cette chaîne est cruciale pour comprendre la concentration des pouvoirs de contrôle des données et l'attribution des pouvoirs décisionnels en vertu de la Convention 108+.

Les rôles et responsabilités des acteurs concernés impliqués dans les activités de traitement dépendent du traitement effectivement effectué et doivent être évalués au cas par cas. Ils peuvent être répartis ou interdépendants entre plusieurs acteurs tout au long du cycle de vie des écosystèmes LLM, du développement du modèle au déploiement et à l'interaction des utilisateurs. Leur détermination doit refléter un pouvoir décisionnel factuel sur les objectifs et les moyens, plutôt que des étiquettes contractuelles ou techniques.

Les arrangements agents peuvent également impliquer un développeur ou orchestrateur d'agents, un déploieur, des fournisseurs d'outils et de services de données, des opérateurs d'autres agents, ainsi que la personne ou organisation pour le nom de laquelle le système agit. Des descriptions telles que « agent », « assistant » ou « autonome » ne déterminent pas le statut juridique au sens de la Convention. L'utilisation d'un agent IA ne modifie pas les rôles et responsabilités des personnes physiques ou morales, des autorités publiques, des services, des agences ou d'autres organismes qui déterminent les finalités et les moyens de traitement, ou traitent les données personnelles en leur nom.

Lorsque plusieurs acteurs déterminent des objectifs ou des moyens, l'existence et la portée du contrôle conjoint doivent être évaluées conformément à la Convention et à la législation applicable. Dans ces évaluations, des distinctions claires doivent être faites entre acteurs spécifiques en fonction de la capacité ou du pouvoir de déterminer les principales caractéristiques et conditions du traitement des données. Les arrangements entre acteurs doivent attribuer en particulier, sans préjudice aux droits des personnes concernées, la responsabilité d'autoriser les outils et autorisations, de définir la mémoire et la conservation, de conserver les dossiers d'actions pertinents, de gérer les droits et violations en amont, de gérer les traités et transferts en aval, et de suspendre l'agent. L'attribution contractuelle ne doit pas être considérée comme concluante lorsqu'elle ne correspond pas aux circonstances factuelles.

La détermination des responsables et des traiteurs de données (lorsque pertinent, co-contrôleurs et sous-traiteurs) ainsi que l'attribution des rôles et responsabilités respectifs entre eux est directement liée à la responsabilité et à la démontrabilité du respect de la Convention (article 10).

Recommandations :

- *Les relations entre les acteurs concernés doivent être délimitées et documentées conformément à la Convention par des termes contractuels clairs et des matrices de responsabilité opérationnelle couvrant tout le cycle de vie du traitement, y compris les instructions des contrôleurs aux processeurs, la gouvernance des outils et des autorisations, la gestion de la mémoire et de la*

rétention, la surveillance de la sécurité, la gestion des incidents, les droits des personnes concernées, les services en aval et les flux transfrontaliers.

(2) Catégories de données personnelles

La gouvernance efficace des données s'engage avec une compréhension claire de ce que constituent les données personnelles et comment les distinguer des informations non personnelles, notamment en tenant compte du stade initial du cycle de vie des écosystèmes LLM. Cela inclut la détermination de la catégorie de données personnelles, ce qui est une exigence essentielle pour déterminer avant le début des activités de traitement les objectifs légitimes du traitement et peut également éclairer la sensibilité de ces activités (article 6) et soutenir en outre des efforts efficaces d'évaluation de l'impact sur la vie privée (article 10). Tout au long des phases de développement et de déploiement, il est nécessaire d'évaluer et de surveiller le potentiel d'identification des données traitées afin d'empêcher l'identification d'un individu.

L'inclusion de données personnelles dans les ensembles de données d'entraînement LLM crée des risques de mémorisation du modèle et de divulgation involontaire dans les résultats. L'anonymisation et la suppression rapide des données personnelles des ensembles de données d'entraînement – lorsque cela n'est pas possible, l'application de pseudonymisation – peut réduire le risque de réidentification tout au long du cycle de vie du système ou de fuite involontaire de données personnelles dans les sorties du système. Les contrôleurs de données et, lorsque pertinent, les processeurs doivent intégrer des conceptions renforçant la confidentialité, [telles que les données synthétiques, les modèles d'apprentissage automatique fédéré, le chiffrement dur, etc.] et faire de meilleurs efforts pour la détection proactive et la suppression (filtrage) des données personnelles avant l'entraînement du modèle.

Tout au long du cycle de vie et en tenant compte du contexte applicatif, les systèmes basés sur les LLM peuvent traiter, générer, accéder ou inférer des catégories particulières de données personnelles au sens de l'article 6 de la Convention 108+[11]. Un tel traitement n'est permis que lorsque des garanties appropriées sont inscrites dans la loi, complétant celles de la Convention. Les inférences nécessitent la même attention lorsqu'elles révèlent des informations couvertes par l'article 6, y compris lorsqu'elles sont probabilistes ou s'avérent ensuite inexactes.

Dans les systèmes agents, les données personnelles peuvent également être contenues dans des plans, des paramètres fournis aux outils, des contextes récupérés, des communications entre agents, des résultats intermédiaires, des observations, des enregistrements d'actions, des dossiers opérationnels et une mémoire à court ou long terme. Une désignation technique ou un traitement comme

[11] Des catégories particulières de données sont : données génétiques ; données personnelles relatives à des infractions, procédures pénales et condamnations, et mesures de sécurité associées ; données biométriques identifiant de manière unique une personne ; données personnelles pour les informations qu'elles révèlent concernant l'origine raciale ou ethnique, les opinions politiques, l'appartenance syndicale, les croyances religieuses ou autres, la santé ou la vie sexuelle.

artefact interne du système ne retire pas ces données du champ d'application de la Convention lorsqu'elles concernent une personne identifiée ou identifiable.

Recommandations :

- *Les acteurs concernés doivent identifier et classifier les catégories de données personnelles ainsi que les phases et composantes dans lesquelles elles sont traitées, intégrer des garanties appropriées dès le plus tôt stade et appliquer des techniques de protection de la vie privée lorsque cela est approprié. L'utilisation de données synthétiques ne doit pas être considérée comme automatiquement anonyme ou sans risque : la lier, la mémorisation, la divulgation et la possibilité de cibler des individus doivent être évaluées dans un contexte donné conformément à la Convention 108+[12].*
- *Les responsables et, le cas échéant, les prestataires doivent exclure par défaut des catégories particulières de données personnelles de la collecte, de la conservation, de la récupération, de l'accès à la mémoire et aux outils, sauf si leur traitement est nécessaire à une fin spécifiée, repose sur une base légitime prévue par la loi et est accompagné de garanties appropriées en vertu de l'article 6 complétant celles mises en place pour des catégories non particulières de données personnelles. Ils doivent également prévenir les inférences sensibles inutiles et vérifier la provenance, l'incertitude, la pertinence et l'exactitude de toute inférence nécessaire avant qu'elle ne serve à une action ou une décision.*

(3) Légitimité du traitement des données

Les systèmes basés sur des LLM reposent sur une formation avec de grands volumes de données pour optimiser leurs performances. Il est essentiel que chaque étape de l'opération — y compris la collecte, le traitement et le déploiement des données — soit réalisée sur la base d'un fondement juridique approprié et en conformité avec les exigences pour le traitement légitime en vertu de la Convention 108+.

Dans un contexte de cycle de vie, chaque responsable doit identifier et documenter la base légitime établie par la loi pour chaque finalité et étape de traitement. Selon les circonstances, le traitement peut se fonder sur le consentement libre, spécifique, éclairé et sans ambiguïté de la personne concernée ou sur une autre base légitime prévue par la loi, conformément à l'article 5, paragraphe 2, de la Convention 108+. La disponibilité d'une base doit être évaluée séparément pour la création du modèle, la post-formation, l'intégration système, le déploiement, l'interaction, la mémoire, l'évaluation et toute amélioration ultérieure du modèle.

Les difficultés à obtenir un consentement valide pour la création, la formation, la post-formation ou l'ajustement fin du modèle ne justifient pas, en soi, le traitement de données personnelles accessibles au public. Pour chaque objectif et étape du cycle

[12] L'article 19 du rapport explicatif de la Convention prévoit que « Les données ne doivent être considérées comme anonymes que tant qu'il est impossible de réidentifier la personne concernée ou si une telle réidentification nécessite un temps, des efforts ou des ressources déraisonnables, compte tenu de la technologie disponible au moment du traitement et des développements technologiques. »

de vie, le responsable doit identifier et documenter une base légitime établie par la loi conformément à l'article 5, paragraphe 2, de la Convention 108+. Les techniques couramment utilisées pour la collecte et la conservation des données — par exemple, le web scraping, [13] l'utilisation de données propriétaires, les données issues des interactions utilisateurs, le traitement secondaire et inférentiel — aux stades initiaux ou avancés du cycle de vie du système peuvent créer des tensions avec les exigences légales fixées par l'article 5 de la Convention 108+.

L'accessibilité publique des données ne crée pas automatiquement une base juridique pour le traitement, tandis que certaines sources de données doivent être exclues du traitement. Compte tenu de la portée, de la gravité et des effets à long terme sur les droits des personnes concernées, tels que la réutilisation de données publiques, les défis liés à la connaissance et au contrôle de l'individu sur leurs données, les attentes raisonnables ainsi que la capacité d'exercer leurs droits à la vie privée,[14] ce traitement peut ne pas démontrer l'intérêt légitime primordial des acteurs des données dans le cadre de l'article 5, paragraphe 2, s'il n'est pas démontré : (a) la légalité de l'intérêt poursuivi[15] ; (b) la nécessité du traitement par rapport à l'intérêt poursuivi ; (c) la proportionnalité de l'intérêt poursuivi par rapport aux droits et libertés en jeu.

Les données personnelles obtenues par le biais d'interactions avec les utilisateurs ou de flux de travail agents ne doivent pas être réutilisées pour l'amélioration des performances, l'amélioration du modèle ou une formation supplémentaire, sauf si le responsable de contrôle a établi un objectif spécifique compatible et une base légitime appropriée en vertu de la Convention et du droit applicable. Lorsque le consentement des personnes concernées est invoqué, il doit être libre, spécifique, informé et sans ambiguïté, exprimé par un consentement affirmatif et pouvant être retiré aussi facilement qu'il a été donné. Le responsable doit s'assurer que les mécanismes de retrait sont efficaces et que le traitement ultérieur est mis en conformité sans affecter le traitement légalement effectué avant le retrait.

Chaque récupération, opération de mémoire, invocation d'un outil, divulgation ou délégation impliquant le traitement de données personnelles doit rester dans la base documentée et légitime. Une instruction de tâche large ne peut légitimer un traitement en aval non lié, l'adoption autonome d'un nouveau but ou destinataire, ni l'accès aux données personnelles relatives à des personnes autres que l'utilisateur au-delà de ce qui est nécessaire et légal.

Ces considérations sont amplifiées en cas de traitement de données personnelles concernant des enfants et des individus appartenant à des groupes vulnérables, où des conceptions appropriées, y compris les pratiques de vérification de l'âge, devraient être activées par défaut.

---

13 EDPB, Directives 03/2026 sur le web scraping dans le contexte de l'IA générative, Version 1.0, 2026.

14 Satakunnan Markkinapörssi Oy et Satamedia Oy c. Finlande [GC], 2017, § 137 ; L.B. c. Hongrie [GC], 2023, § 103.

15 Sur la base des directives de l'EDPB sur l'intérêt légitime, un intérêt légitime doit se conformer aux critères cumulatifs suivants selon lesquels l'intérêt doit être considéré comme légal, l'intérêt doit être clairement et précisément articulé, et réel et présent, Directives de l'EDPB 1/2024 sur le traitement des données personnelles basées sur l'article 6(1)(f) du RGPD, version 1.0, 2024, para. 28-30.

Recommandations :

- *L'activité de traitement des données doit être conçue de manière à garantir la traçabilité de chaque opération, le respect des exigences de base juridique et de légitimité établies par la Convention 108+, et intégrer des garanties appropriées contre les violations des droits des personnes concernées. Tout au long du cycle de vie du système, les responsables de traitement des données et, le cas pertinent, les responsables doivent cartographier, documenter les activités de traitement et s'appuyer sur une base juridique appropriée.*
- *Les responsables de traitement et les traiteurs doivent s'abstenir de pratiques de traitement illégales, y compris le marketing non sollicité, le profilage illégal ou la surveillance comportementale, et doivent évaluer les impacts correspondants sur la vie privée, la protection des données et d'autres droits humains.*

(4) Application des principes fondamentaux pour la protection des données personnelles

a. Principe de nécessité et de proportionnalité

Le traitement doit être nécessaire et proportionné par rapport à un objectif spécifié, explicite et légitime. Les responsables doivent examiner attentivement les intérêts de toutes les personnes susceptibles d'être affectées ; le développement et l'utilisation de nouvelles technologies impliquent une responsabilité particulière de trouver un équilibre équitable entre ces intérêts. Conformément à l'article 5, paragraphe 1 de la Convention 108+, le traitement ne doit pas entraîner une atteinte illégale aux droits et libertés fondamentales des personnes concernées et doit être limité à ce que prévoit l'article 11 de la Convention comme exceptions légales. Les responsables doivent évaluer si le traitement est approprié et essentiel à l'objectif et si l'objectif peut raisonnablement être atteint par des moyens moins intrusifs, et appliquer cette évaluation tout au long du cycle de vie, en étroite coordination avec la minimisation des données.[16] Les traiteurs doivent fournir les informations et l'assistance nécessaires à ces évaluations, lorsque cela est applicable.

Le principe de nécessité et de proportionnalité doit être démontré tout au long du cycle de vie des modèles fondamentaux et des systèmes LLM par une évaluation continue et une réflexion du juste équilibre entre des intérêts concurrents. Cet effort inclut l'articulation claire de l'objectif légitime, l'évaluation d'alternatives moins intrusives pour atteindre des résultats équivalents, la limitation de l'utilisation des données personnelles à ce qui est strictement nécessaire pour l'entraînement, l'ajustement et l'adaptation, et la garantie ─ par défaut ─ que les données personnelles traitées lors des interactions utilisateur ne soient pas plus que nécessaires pour fournir des résultats.

[16] Avis EDPB 28/2024 sur certains aspects de la protection des données liés au traitement des données personnelles dans le contexte des modèles d'IA, 2024, para. 70-75.

b. Équité et transparence

Le traitement équitable et transparent des données personnelles, tel qu'énoncé à l'article 5, paragraphe 4, litterae « a » de la Convention 108+, s'applique tout au long du cycle de vie des modèles LLM et basés sur les LLM. Les acteurs des données doivent veiller à ce que le traitement des données ne conduise pas à une ingérence illégale avec les droits et libertés des personnes concernées, entraînant des conséquences négatives plus larges (telles que des pratiques discriminatoires et des préjugés). Les responsables des traiteurs, et lorsque cela est applicable, les traités doivent trouver un équilibre équitable entre tous les intérêts concernés, qu'ils soient publics ou privés, et les droits et libertés en jeu, et évaluer si les données personnelles ont été collectées et traitées de manière à respecter les attentes raisonnables en matière de vie privée des personnes concernées.

Pour les systèmes agentiques, la nécessité et la proportionnalité doivent être évaluées en fonction de chaque récupération, inférence, ajout d'informations à la mémoire, invocation d'un outil, divulgation et action, ainsi que pour le système dans son ensemble. La généralité ou l'adaptabilité d'un agent ne justifie pas des objectifs ouverts, une opération indéfinie ou un accès aux données personnelles et aux outils qui peuvent simplement être utiles pour des tâches futures. Le degré d'autonomie effectif le moins intrusif et la gamme effective la plus étroite d'actions autorisées doivent être sélectionnés.

Les obligations de transparence prévues à l'article 8 et les exigences de responsabilité prévues à l'article 10 s'appliquent tout au long du cycle de vie d'un système basé sur un LLM. L'opacité des pipelines de formation, de l'adaptation post-formation, des mécanismes de récupération et des intégrations en aval peuvent rendre difficile pour les sujets de données la compréhension de la nature et de la séquence des opérations de traitement. Les contrôleurs et, le cas échéant, les processeurs doivent donc mettre en œuvre des mesures de transparence efficaces qui rendent le traitement pertinent intelligible et soutiennent des explications pertinentes. Dans le contexte plus large de la gouvernance de l'IA, la transparence implique ouverture et clarté concernant la gouvernance, les opérations du système pertinent et les processus décisionnels pour les acteurs et parties prenantes.[17] Cela inclut la fourniture d'informations appropriées sur les données utilisées tout au long du cycle de vie du système ainsi que sur les mesures prises pour protéger les données personnelles.[18] Lorsque pertinent, le contenu généré par l'IA doit être identifié de manière claire et efficace. [19] Le traitement équitable implique, entre autres, que les responsables de traitement des données s'efforcent de concilier les valeurs fondamentales du respect de la vie privée et de la libre circulation de l'information à travers les frontières et entre les peuples, de ne pas s'appuyer sur les déséquilibres de pouvoir pour déterminer les principales caractéristiques et conditions du traitement des données, et de respecter en fin de compte le droit de chacun à l'autodétermination informationnelle.

---

[17] Article 8, Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit (CETS n° 225).

[18] Rapport explicatif, Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit (CETS n° 225), paragraphe 57.

[19] Article 8, Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit (CETS n° 225).

Conformément à l'article 8 de la Convention 108+, la transparence du traitement nécessite une communication claire et conviviale de l'identité des acteurs impliqués dans le traitement, des informations adéquates et pertinentes concernant les étapes des activités de traitement, les ensembles de données, y compris leurs sources, leur base juridique et leurs objectifs de traitement, ainsi que les catégories de données personnelles traitées. Lorsque les interactions avec un LLM peuvent être conservées, analysées ou réutilisées pour améliorer les modèles, celles-ci doivent être communiquées de manière accessible, avant le traitement. Dans le contexte des systèmes basés sur les LLM, les responsables des traitements doivent s'assurer que tous les destinataires sont identifiés et communiqués, ainsi que des informations concernant les processus de raisonnement internes et tout changement des périodes de conservation, la prise de décision automatisée, les flux transfrontaliers de données ou les nouveaux usages de traitement. Les responsables de traitement doivent fournir des explications pertinentes sur le rôle des modèles LLM dans l'activité de traitement, la nature et les objectifs du traitement, les principales caractéristiques et limitations du système, ainsi que les garanties mises en place pour réduire les risques pour les individus. La transparence doit être soutenue par une documentation appropriée visant à informer toutes les catégories de sujets de données (y compris les utilisateurs potentiels).

En vertu de l'article 5, paragraphe 4, litterae « a » et de l'article 8 de la Convention 108+, la transparence doit être opérationnelle lorsqu'elle soutient un traitement équitable et transparent et permet aux personnes concernées une bonne sensibilisation aux risques liés à la vie privée, aux garanties disponibles, aux recours accessibles et à l'exercice de leurs droits. L'avis de confidentialité concernant le traitement des données doit être conçu avec des interfaces conviviales, fournissant des informations contextuelles et des mécanismes pratiques permettant aux individus de comprendre comment les opérations de traitement peuvent les affecter et comment leurs données personnelles sont utilisées tout au long du cycle de vie des systèmes basés sur des LLM.

c. Limitation de but et minimisation des données

La collecte large de données à l'étape initiale de la création du modèle peut entrer en conflit avec les principes de limitation de la finalité et de minimisation des données tels que consacrés à l'article 5 de la Convention 108+. Les responsables des traitements doivent définir et documenter clairement l'objectif spécifique, explicite et légitime pour lequel les données personnelles sont traitées à chaque étape du cycle de vie du système (Article 5, paragraphe 4, litterae « b »). L'évaluation de la légitimité de l'objectif doit être soutenue par un équilibre des droits, libertés et intérêts en jeu dans chaque cas ; le droit à la protection des données personnelles d'une part, et la protection des autres droits d'autre part.[20]

Compte tenu du développement, de l'adaptation et des améliorations du modèle, la compatibilité du traitement ultérieur avec l'objectif initial doit être évaluée en continuelle, en prenant en *compte, entre autres*, la relation entre les objectifs originaux et ultérieurs, le contexte dans lequel les données ont été initialement

[20] Rapport explicatif au Protocole modifiant la Convention pour la protection des individus en matière de traitement automatique des données personnelles, par. 48.

collectées, les attentes raisonnables des personnes concernées, la catégorie des données personnelles et leur sensibilité, ainsi que les impacts négatifs potentiels sur les droits et libertés des personnes concernées. Les données initialement collectées à des fins limitées et spécifiques au contexte peuvent ensuite être réutilisées, agrégées ou réutilisées pour l'optimisation, la prédiction comportementale, la personnalisation ou l'analyse déférentielle de manière à dépasser les objectifs originaux de collecte ou les attentes raisonnables des personnes concernées. Lorsque le système basé sur un LLM fonctionne dans le cadre et l'objectif d'une activité de traitement existante, aucune base juridique distincte n'est requise, cependant, si un traitement supplémentaire est effectué à de nouvelles fins, les responsables des données doivent identifier une base juridique appropriée en vertu de l'article 5, paragraphe 1 de la Convention 108+.

La minimisation des données doit être mise en œuvre dès le stade initial des écosystèmes LLM et observée tout au long de son cycle de vie. Les systèmes LLM doivent être conçus pour accéder et traiter uniquement les données personnelles strictement nécessaires à un objectif spécifique, tout en considérant des alternatives préservant la vie privée (par exemple, par traitement de données synthétiques ou anonymisées, utilisation de données pseudonymisées lorsque possible, filtrage de catégories spécifiques de données personnelles, etc.). Les principes de limitation de but et de minimisation des données doivent être mis en œuvre de manière à répondre non seulement aux pratiques de collecte directe de données, mais aussi à des formes de traitement inférentiel de plus en plus avancées, agrégation multimodale, prédiction comportementale et profilage adaptatif.

Des mesures techniques et organisationnelles appropriées doivent minimiser le traitement afin que les données personnelles soient adéquates, pertinentes et non excessives par rapport à l'objectif légitime poursuivi, conformément à l'article 5, paragraphe 4, litterae « c ». Cela nécessite une évaluation des catégories, sources, volume, granularité, durée et destinataires des données personnelles à chaque étape pertinente du cycle de vie.

Dans un flux de travail agentique, la limitation de la finalité et la minimisation des données doivent être techniquement applicables pour chaque recherche, message, opération mémoire, invocation d'un outil et divulgation. L'accès ouvert à des boîtes aux lettres, disques, listes de contacts, bases de données ou comptes ne devrait pas être accordé lorsque l'accès est limité par champ, enregistrement, but, temps ou tâche suffirait. La délégation à un autre agent ou service ne devrait pas élargir silencieusement l'objectif, les catégories de données personnelles, les destinataires ou les actions autorisées.

d. Limitation de stockage

Conformément à l'article 5, paragraphe 4, litterae « e », lorsque les données personnelles ne sont plus requises pour atteindre les objectifs identifiés, elles doivent être supprimées ou préservées de manière sécurisée sous une forme qui ne permet pas l'identification ou la réidentification des personnes concernées. Les responsables des traitements doivent établir des périodes de conservation documentées pour des catégories spécifiques de données personnelles traitées, le(s) but(s) pour lequel les

données sont conservées et les critères régissant la suppression ou la désidentification irréversible, ce qui nécessite un examen et une évaluation itératifs.

Une attention particulière doit être accordée aux risques engendrés par la mémorisation des données, car les modèles fondamentaux peuvent mémoriser et reproduire les données personnelles représentées dans leurs données d'entraînement. De plus, la mémorisation des données remet en cause les normes de qualité des données et affecte la capacité des personnes concernées à exercer efficacement leurs droits en vertu de l'article 9 de la Convention 108+. Ces préoccupations sont amplifiées en lien avec les fonctions mémoire des systèmes basés sur les LLM. Les contrôleurs de données doivent être capables de démontrer que les systèmes de mémoire et les risques de mémorisation sont correctement évalués, traités par des mécanismes d'atténuation et entreprennent une extraction ultérieure des données mémorisées. Lorsque la suppression des données n'est pas possible en raison de limitations techniques, les contrôleurs peuvent envisager la mise en œuvre de garanties supplémentaires (telles que la rééducation des modèles, les techniques de désapprentissage automatique, etc.) en réponse aux impacts potentiels sur les risques.

Dans le contexte des interactions utilisateurs, les responsables doivent veiller à ce que les invites, l'historique des conversations et les résultats générés contenant des données personnelles ne soient conservés que lorsque cela est nécessaire pour des fins spécifiées, telles que le respect d'une obligation légale, la sécurité ou la responsabilité. Le traitement doit être légal, équitable et transparent, et la conservation ne doit pas être conditionnée uniquement à la connaissance présumée de la personne concernée.

Les systèmes agents peuvent utiliser des fenêtres contextuelles, de la mémoire de travail ou de session, des profils utilisateurs persistants, de la mémoire organisationnelle, des mémoires vectorielles, des caches et des enregistrements externes. Les contrôleurs doivent déterminer séparément l'objectif, la base légitime, les conditions d'accès et la période de conservation pour chaque couche mémoire. La mémoire persistante doit être désactivée par défaut, sauf si la continuité est nécessaire pour un objectif spécifié et que les risques résultants sont correctement atténués ; pour les tâches à fort impact, la séparation de but et d'utilisateur ou de cas doit être appliquée par conception.

Lorsque cela est applicable, les personnes concernées doivent pouvoir localiser, inspecter, rectifier et effacer les données personnelles conservées en mémoire. Les mesures de suppression et de rectification doivent traiter les historiques de conversation, les embeddings, les stockages vectoriels, les caches, les répliques, les mémoires partagées et les destinataires en aval, et empêcher que des données obsolètes ou effacées ne soient réintroduites à partir de copies obsolètes. Les entrées mémoire doivent, le cas échéant, contenir des informations sur la provenance, le temps, la finalité, la qualité et l'expiration pertinentes.[21] [22]

---

[21] AEPD, Agentic Artificial Intelligence from the Perspective of Data Protection, version 1.2, 2026, sections concernant la mémoire à court et long terme et l'exercice des droits des sujets de données concernés.

[22] *Wang Bo et al.,* Dévoiler les risques de confidentialité dans la mémoire des agents LLM, Actes de l'ACL 2025, DOI : 10.18653/v1/2025.acl-long.1227, https://aclanthology.org/2025.acl-long.1227/.

La journalisation doit soutenir la responsabilité et la sécurité sans devenir une source supplémentaire de traitement disproportionné. Les événements pertinents, y compris les sources de données accédées, les agents et outils invoqués, les divulgations et modifications, les destinataires, l'autorité déléguée, les approbations humaines, les résultats et erreurs, doivent être enregistrés sous une forme récupérable soumis à des contrôles stricts d'accès et de conservation. Les traces de raisonnement du modèle interne ne doivent pas être conservées ou divulguées par défaut ; lorsqu'elles contiennent des données personnelles, elles restent soumises à la Convention et peuvent elles-mêmes créer des risques de sécurité et de confidentialité.[23]

e. Qualité des données :

Les responsables de traitement doivent s'assurer que les ensembles de données d'entraînement restent conformes aux exigences de qualité des données fixées par l'article 5, paragraphe 4, litterae « d » de la Convention 108+. En particulier, « les données personnelles en traitement doivent être exactes et, si nécessaire, maintenues à jour », sinon, les responsables doivent supprimer les données inexactes et s'abstenir de traitements ultérieurs. Dans le contexte des écosystèmes LLM, cette exigence s'applique à toutes les étapes de son cycle de vie, couvrant les sorties du système, les pratiques de traitement inférentiel et les usages en aval des informations probabilistes concernant des individus identifiables.

Les responsables de traitement et les prestataires de traitement doivent s'appuyer sur des données exactes, fiables et représentatives, en tenant compte de l'impact à long terme des données d'entraînement dépourvues de ces qualités. En particulier, les données d'entraînement des systèmes basés sur les LLM peuvent inclure des biais, qui peuvent se refléter dans des résultats discriminatoires ou inexacts. Les risques associés englobent, sans s'y limiter, des déclarations inexactes concernant des individus identifiables, y compris des allégations fausses, des informations contextuelles fabriquées ou trompeuses, des « hallucinations » ou la génération d'informations personnelles plausibles mais inexactes dans les résultats, pouvant avoir des impacts négatifs à long terme sur les droits et libertés des sujets ; l'amplification des biais, lorsque des associations statistiques sous-jacentes reproduisent ou renforcent des schémas injustes ou discriminatoires concernant des individus ou des groupes.

Par conception, les responsables de données et, le cas échéant, les processeurs doivent mettre en œuvre des mesures efficaces soutenant la vérification du contenu des données et en garantissant la précision ainsi que la représentativité. Cela implique également l'utilisation de techniques de filtration dans les entrées et sorties, l'indexation des ressources de données, l'exclusion des sources non vérifiées des ensembles de données d'entraînement (notamment pertinentes en matière de scaping des données), [24]l'évaluation itérative et le suivi des résultats du système, y compris les inférences. Compte tenu des capacités probabilistes des systèmes basés sur LLM,

---

[23] *Green Tommaso et al.,* Pensées fuyantes : Les grands modèles de raisonnement ne sont pas des penseurs privés, Actes de l'EMNLP 2025, DOI : 10.18653/v1/2025.emnlp-main.1347, https://aclanthology.org/2025.emnlp-main.1347/.

[24] EDPB, Directives 03/2026 sur le web scraping dans le contexte de l'IA générative, Version 1.0, 2026, para. 39–42.

une surveillance significative à cet égard fait référence à la supervision humaine et à la capacité d'intervention proactive, lorsque nécessaire. Un traitement transparent, dans ce contexte, implique de contribuer à la conscience des sujets des données face aux sorties probabilistes du système et à leurs conséquences. Les résultats et inférences doivent être testés et surveillés tout au long du fonctionnement. Lorsque les risques les exigent, les examinateurs humains doivent être capables d'identifier les erreurs et d'intervenir efficacement. Les sujets de données doivent recevoir des informations intelligibles sur la nature probabiliste et les limites pertinentes des résultats du système lorsque cela est nécessaire pour garantir un traitement équitable et transparent.

Lorsqu'une sortie, une inférence, une saisie mémoire ou un enregistrement d'action constitue des données personnelles, les garanties d'exactitude doivent refléter son usage et ses conséquences prévues. Avant qu'un agent n'utilise des informations probabilistes ou récupérées externement pour modifier des données, communiquer avec un tiers ou entreprendre une action affectant une personne, ces informations doivent être vérifiées dans une mesure proportionnelle au risque. Les systèmes doivent être conçus pour détecter, contenir et corriger les inexactitudes en cascade entre les agents et les services en aval.

Recommandations (concernant les principes applicables et les normes de qualité des données) :

- *Une répartition claire des rôles et responsabilités entre les contrôleurs, les prestataires et autres acteurs concernés est essentielle à la légitimité du traitement et à l'application effective des exigences de qualité des données. Lorsque pertinent, les arrangements contractuels et autres doivent spécifier la responsabilité de la collecte, de la curation, de l'assurance qualité et de la correction des données lors de la création du modèle, de l'ajustement fin et du développement ultérieur. De tels arrangements doivent soutenir le respect de la limitation des objectifs, la minimisation des données et la qualité des données tout au long du cycle de vie, sans modifier l'attribution des responsabilités découlant des circonstances factuelles.*
- *Le traitement des données dans les écosystèmes LLM doit respecter strictement un traitement équitable et transparent, où les principes de nécessité, de proportionnalité et de minimisation des données sont respectés, garantissant que les objectifs légitimes sont clairement définis, que le traitement des données personnelles est essentiel, que des alternatives moins intrusives sont envisagées et que l'utilisation des données personnelles est limitée tout au long du cycle de vie du système. Engager les parties prenantes et réfléchir continuellement à l'équilibre entre les droits humains et les libertés fondamentales, qu'il s'agisse du domaine public ou privé, est essentiel pour une gouvernance responsable des données.*
- *Les responsables des traitements doivent garantir un traitement transparent et équitable des données personnelles tout au long du cycle de vie du système LLM en fournissant des avis de confidentialité clairs et ciblés aux utilisateurs, des notifications transparentes et des mesures de*

transparence efficaces, permettant ainsi aux personnes concernées de comprendre, gérer et exercer leurs droits et garanties. Cette approche soutient des attentes raisonnables, atténue les interférences illégales et prévient des conséquences négatives plus larges, telles que la discrimination.

- *Les responsables des traitements doivent fixer et réviser régulièrement les périodes de conservation documentées et les critères de suppression des données personnelles, afin de s'assurer que les risques de mémorisation sont atténués et que les données d'interaction utilisateur ne sont conservées que lorsque strictement nécessaire et sont communiquées de manière transparente, tandis que la réutilisation pour le développement du modèle n'est autorisée que si justifiée par une base légale et des garanties appropriées.*
- *Les responsables de traitement des données doivent veiller, et les traités doivent les aider à s'assurer, que les données personnelles utilisées pour entraîner ou exploiter les systèmes basés sur des LLM sont exactes, fiables et, le cas pertinent, représentatives par rapport à l'objectif du traitement. Les mesures appropriées doivent inclure des vérifications de sources et de provenance, la validation, le filtrage, les tests périodiques et un examen humain significatif proportionnel au contexte et aux risques. Lorsque cela est nécessaire pour garantir un traitement équitable et transparent, les sujets des données doivent recevoir des informations intelligibles sur la nature probabiliste et les limites pertinentes des résultats.*

V. DIRECTIVES GÉNÉRALES SUR LES DROITS DES PERSONNES CONCERNÉES

Les caractéristiques techniques des écosystèmes LLM peuvent entraver l'exercice effectif des droits des personnes concernées. Les responsables doivent donc mettre en œuvre, et les traiteurs doivent les assister, des mesures techniques, organisationnelles et procédurales appropriées pour mettre en œuvre les droits énoncés à l'article 9 de la Convention 108+. La difficulté technique seule ne doit pas être considérée comme un motif de restriction de ces droits. Toute exception ou restriction doit se conformer à l'article 11 de la Convention[25]. Les mesures doivent être établies dès le plus tôt stade du cycle de vie et doivent tenir compte du contexte d'utilisation, des catégories et de la sensibilité des données personnelles, ainsi que des besoins particuliers des enfants et *des individus appartenant à des groupes vulnérables.* .

(1) Droit de ne pas être soumis à la prise de décision automatisée

Les systèmes basés sur les LLM sont de plus en plus intégrés dans des environnements décisionnels automatisés et semi-automatisés. L'article 9, paragraphe 1, litterae « a » garantit le droit des personnes concernées à ne pas être soumis à une décision qui les a significativement affectées uniquement sur la base

[25] Pour plus de détails, veuillez consulter : https://rm.coe.int/t--2021-7rev13-interpretation-of-general-principles-article-11-c108-/1680b6c146

d'un traitement automatisé des données sans prise en compte de leurs opinions. Dans les cadres plus larges d'évaluation et de gestion des risques liés à l'IA, le processus d'engagement de toutes les parties prenantes est un élément d'une importance capitale.[26] Pour les efforts de gouvernance des risques liés à la vie privée, la Convention 108+ impose spécifiquement la prise en compte des intérêts des personnes concernées. Cela implique que les personnes concernées puissent contester, contester et étayer l'inexactitude présumée du traitement ou d'autres facteurs affectant leurs droits et libertés.[27]

Les contrôleurs doivent garantir une supervision humaine significative et opportune des décisions automatisées et des actions agentiques. Les examinateurs doivent disposer de la compétence, de l'autorité, de l'indépendance, du temps et des informations nécessaires pour comprendre les facteurs pertinents, refuser l'approbation, prendre en compte les opinions de l'individu, mettre en pause ou passer outre le système et modifier ou inverser le résultat lorsque cela est approprié. Les mesures doivent traiter le biais d'automatisation, la fatigue des alertes et la surdépendance. Même lorsque l'article 9, paragraphe 1, litterae « a », n'est pas engagé, les autres exigences de la Convention, y compris les articles 5, 7, 8, 9 et 10, restent applicables aux traitements et aux actions affectant les individus.

(2) Droit d'accès

L'article 9, paragraphe 1, litterae « b », garantit à chaque individu le droit d'obtenir, sur demande, à des intervalles raisonnables et sans retard ni frais excessifs, la confirmation du traitement des données personnelles qui le concernent, la communication intelligible des données traitées, toutes les informations disponibles sur leur origine et la période de conservation, ainsi que toute autre information que le responsable doit fournir afin d'assurer la transparence du traitement conformément à l'article 8, paragraphe 1. Dans le contexte des systèmes agents, les mécanismes d'exercice du droit d'accès doivent être capables d'identifier et de récupérer les données personnelles pertinentes contenues dans les consignes, contextes récupérés, mémoire à court et long terme, mémoires vectorielles, enregistrements d'actions et services connectés, ainsi que de retracer les divulgations de ces données vers des outils ou autres agents. Les responsables des traitements doivent établir des arrangements de coordination appropriés avec les traiteurs et autres entités impliquées dans le traitement, conformément à leurs rôles et responsabilités respectifs, afin de garantir que les personnes concernées puissent exercer efficacement leurs droits sans avoir à identifier et contacter chaque entité de la chaîne de traitement séparément.

L'article 9, paragraphe 1, litterae « c », garantit à chaque individu le droit d'obtenir, à la demande, la connaissance du raisonnement sous-jacent au traitement des données lorsque les résultats lui sont appliqués. Dans les systèmes basés sur LLM et agents, la conformité ne doit pas se réduire à la divulgation du code source, des paramètres

---

[26] Voir l'article 16, Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit (CETS n° 225); Cadre d'évaluation des risques et de l'impact des systèmes d'IA du Conseil de l'Europe, du point de vue des droits de l'homme, de la démocratie et de l'État de droit (HUDERIA).

[27] Rapport explicatif au Protocole modifiant la Convention pour la protection des individus en matière de traitement automatique des données personnelles, pour 75.

du modèle ou des traces internes de raisonnement du modèle. Les responsables doivent fournir des informations significatives, intelligibles et spécifiques au contexte sur les principaux facteurs, les données personnelles et les sources sur lesquelles on se fie, le rôle du modèle, les outils et l'intervention humaine, ainsi que la séquence des actions pertinentes conduisant au résultat, ainsi que sa signification et ses conséquences prévisibles. Les explications générées par le même système doivent être validées et ne doivent pas remplacer les preuves documentées du traitement.

Pour les flux de travail agents, les informations doivent couvrir les utilisations pertinentes des outils, délégations, autorisations, sources de données externes et approbations humaines lorsqu'elles ont influencé le résultat, et doivent identifier l'acteur responsable à chaque étape pertinente. Toute limitation nécessaire à la protection des droits et libertés d'autrui, sécurité ou secrets légalement protégés, doit être légale, nécessaire et proportionnée et ne doit pas rendre ce droit inefficace. Les dossiers opérationnels et autres documents documentant la provenance et le mouvement des données personnelles doivent être distingués des traces de raisonnement modèles internes. Ces traces ne doivent pas être considérées comme un enregistrement d'audit fiable ni comme suffisantes, en soi, pour satisfaire au droit prévu à l'article 9, paragraphe 1, litterae « c » ; leur conservation ou divulgation doit être limitée à ce qui est légal, nécessaire et proportionné.[28]

(3) Droit à la rectification, à l'effacement et à l'objection

L'article 9, paragraphe 1, litterae « d », de la Convention 108+ garantit à toute personne le droit de s'opposer à tout moment, pour des motifs liés à sa situation, au traitement des données personnelles qui le concernent, sauf si le titulaire de votre mandat démontre des motifs légitimes pour le traitement qui dépassent sur ses intérêts, droits et libertés fondamentales. L'article 9, paragraphe 1, litterae « e », garantit en outre le droit d'obtenir, sur demande, gratuitement et sans délai excessif, de rectification ou d'effacement, selon le cas, des données personnelles relatives à elle lorsque ces données sont traitées ou ont été traitées en violation des dispositions de la Convention. Les caractéristiques techniques des systèmes basés sur des LLM peuvent rendre difficile la localisation, la rectification ou l'effacement de données personnelles particulières dans les ensembles de données d'entraînement, les archives de récupération et mémoire, les sorties stockées, les journaux et les systèmes en aval. Une telle difficulté technique ne dispense pas, en soi, les responsables de l'obligation de donner effet aux droits applicables au traitement dont ils sont responsables. Les responsables des données doivent établir des procédures capables d'identifier les données personnelles pertinentes et de déterminer des mesures efficaces pour répondre aux demandes. Les personnes concernées ne devraient pas être tenues de comprendre l'architecture du système ni d'identifier le composant dans lequel leurs données personnelles sont conservées.

Les responsables des traitements doivent prendre des mesures efficaces pour donner effet aux demandes de rectification ou d'effacement, aux objections au traitement et, si prévu par la loi, aux demandes de restriction du traitement. Les traiteurs doivent

[28] *Turpin M., Michael J., Perez E., Bowman S. R.,* Les modèles de langage ne disent pas toujours ce qu'ils pensent : explications infidèles dans le déclenchement de la chaîne de pensée, avancées dans les systèmes de traitement de l'information neuronale 36 (NeurIPS 2023), DOI : 10.52202/075280-3275 ;

assister les responsables à cet égard, conformément à la loi applicable et aux instructions du contrôleur. Les mesures doivent être choisies en tenant compte du lieu et de la manière dont les données personnelles sont traitées. Dans le cadre du traitement dont le responsable est responsable, ces mesures peuvent inclure la correction ou la suppression de données personnelles des ensembles de données sources, des systèmes de récupération, des magasins vectoriels, de la mémoire et des caches ; la suppression des sorties stockées contenant les données ; l'empêchement de leur réintroduction dans le système ou leur divulgation ultérieure ; la communication de la rectification ou de l'effacement aux destinataires lorsque la loi applicable l'exige ; et l'édition du modèle, l'ajustement ou la rééducation lorsque ces mesures sont nécessaires, proportionnelles et démontrées efficaces. Ni le coût technique d'une mesure particulière ni l'intérêt des contrôleurs à maintenir la performance du modèle ne doivent en soi rendre un droit applicable inefficace.
Lorsqu'un agent a déjà divulgué des données personnelles inexactes ou traitées illégalement, modifié des dossiers sur la base de ces données ou engagé une action affectant un individu, le responsable doit veiller à un examen humain rapide, prendre des mesures pour prévenir ou atténuer d'autres effets négatifs, informer les destinataires lorsque la loi applicable l'exige et, lorsque cela est possible, annuler l'action ou en remédier autrement. Les procédures doivent également empêcher que des données personnelles inexactes, effacées ou supplantées ne soient réintroduites dans le système à partir de mémoire persistante, de copies obsolètes ou d'autres agents.

(4) Accès aux recours et à l'assistance

Conformément à l'article 9, paragraphe 1, litterae « f » et litterae « g », les responsables de traitement des données et, le cas échéant, les traiteurs de données déployant ou exploitant des systèmes basés sur des LLM doivent faciliter l'exercice effectif des droits des personnes concernées en fournissant des procédures accessibles pour la soumission et le traitement de leurs demandes, en coopérant avec les autorités compétentes au sens de l'article 15, et en veillant à ce que les personnes concernées ne soient pas privées d'un recours effectif en raison de défis opérationnels ou de complexités techniques. À cet égard, les personnes concernées jouissent du droit de déposer une plainte, de demander l'assistance et de solliciter l'intervention de l'autorité compétente. Elles ont également accès aux recours administratifs et judiciaires conformément à l'article 12.

Recommandations :

- *Sans préjudice de tout autre droit ou recours prévu par les lois applicables, les personnes concernées devraient pouvoir exercer leurs droits en matière de traitement des données personnelles dans le contexte des systèmes basés sur des LLM.*
- *La mise en œuvre correcte de l'article 9 de la Convention 108+ exige la disponibilité de garanties techniques et organisationnelles, de mesures procédurales et de mécanismes accessibles pour l'exercice des droits des personnes concernées, une amélioration des pratiques de traçabilité et de documentation, une répartition claire des responsabilités entre les acteurs des données, des procédures de supervision humaine efficaces et*

significatives, des garanties contre le profilage inférentiel excessif, ainsi que des mécanismes de plainte et de réparation.

- *Les acteurs des données doivent faire des efforts appropriés et raisonnables pour faciliter et donner exécution aux demandes formulées par les personnes concernées dans l'exercice de leurs droits. À cette fin, ils doivent tenir compte des besoins spécifiques et des intérêts supérieurs des mineurs, ainsi que des circonstances particulières des individus appartenant à des groupes vulnérables, et doivent mettre en œuvre des solutions efficaces garantissant la jouissance accessible et non discriminatoire de leurs droits.*

VI. DIRECTIVES GÉNÉRALES SUR LA SÉCURITÉ DES DONNÉES, LA RESPONSABILITÉ ET LES FLUX TRANSFRONTALIERS DE DONNÉES

(1) Sécurité des données

L'article 7, paragraphe 1 de la Convention 108+ exige que chaque Partie stipule que les responsables de traitement et, le cas échéant, les traités prennent des mesures de sécurité appropriées contre des risques tels que l'accès accidentel ou non autorisé, la destruction, la perte, l'utilisation, la modification ou la divulgation de données personnelles. Dans le contexte des systèmes basés sur des LLM, de telles mesures doivent couvrir tous les composants et opérations pertinents au sein du traitement dont elle est responsable, y compris la formation, la post-formation et les données opérationnelles ; les systèmes de récupération et la mémoire ; les outils et services externes ; les mécanismes de gestion de l'identité et des accès ; les communications entre agents ; et les dispositifs d'intervention et de supervision humains. Lorsque un système est capable d'initier des actions, les mesures doivent également s'étendre aux composants et communications par lesquels ces actions sont autorisées, réalisées et enregistrées.

Lors de la détermination des mesures appropriées, les contrôleurs et, le cas échéant, les processeurs doivent tenir compte de la nature et de la sensibilité des données personnelles, du volume, de la portée, des objectifs et de la durée du traitement, de la longueur et de la complexité de la chaîne de traitement ainsi que du degré d'interconnexion entre les composants concernés. Une attention particulière doit être portée aux vulnérabilités résultant de l'architecture et de l'utilisation des systèmes basés sur LLM, y compris celles affectant les données d'entraînement, la mémorisation des modèles, les mécanismes de récupération et les interfaces de programmation applicative, ainsi que les risques de divulgation involontaire ou de mélange de données personnelles entre contextes, utilisateurs ou sessions. Ces risques peuvent être accrus lorsque les systèmes sont interconnectés ou fonctionnent avec un haut degré d'autonomie et bénéficient d'un accès à des services externes, sources de données ou environnements opérationnels.

Pour donner une application pratique à l'article 10, paragraphes 2 et 3 de la Convention 108+, l'examen de l'impact probable du traitement prévu doit aborder les risques généraux pour la sécurité de l'information ainsi que les vulnérabilités propres

au modèle et au système tel que déployé. Des mesures techniques et organisationnelles appropriées doivent être intégrées dès les premiers stades de conception et maintenues tout au long de la période d'exploitation. Lorsque cela est pertinent, les configurations par défaut doivent limiter l'accès aux données personnelles, aux outils, aux comptes, aux services externes et aux fonctions système à ce qui est nécessaire pour les objectifs spécifiés. L'évaluation, les mesures sélectionnées et les raisons de leur sélection doivent être documentées afin que la conformité puisse être démontrée conformément à l'article 10, paragraphe 1. L'adéquation et l'efficacité des mesures doivent être testées à des intervalles appropriés et maintenues en examen à la lumière des risques et vulnérabilités évolutifs.

L'examen doit également prendre en compte les risques liés à l'autorité et à l'accès accordés au système. Ces risques peuvent inclure[29] :

(a) Risques pour la confidentialité, l'intégrité, la disponibilité et la qualité de la formation, des données post-formation et opérationnelles, y compris l'utilisation de sources peu fiables ou compromises, l'empoisonnement des données et les attaques dérobées pouvant affecter le comportement du modèle, les données personnelles ou les sorties système ;

(b) l'inversion de modèle, l'inférence d'appartenance, l'extraction de données mémorisées et d'autres attaques capables de révéler des données personnelles ou de permettre d'inférences sur des individus à partir de données d'entraînement, de dépôts de récupération, de mémoire ou d'enregistrements d'interactions ;

(c) Attaques d'évasion, compromis des restrictions de sécurité ou des instructions et des injections rapides[30], manipulant le modèle pour accéder et révéler les données et l'injection directe ou indirecte de prompt, incluant des instructions intégrées dans des pages web, fichiers, messages, contenus ou sorties récupérés d'outils, qui peuvent contourner les garanties applicables, détourner le système de l'objectif ou de la tâche autorisée, ou entraîner un traitement ou une action non autorisés ;

(d) L'empoisonnement de la mémoire, les dépôts de récupération, les sorties d'outils ou de communications entre agents, y compris des attaques conçues pour être activées ultérieurement ou pour propager des informations manipulées entre agents ;

(e) Le vol ou l'abus de crédibilités ; la fausse représentation ou l'abus de l'identité des utilisateurs, agents ou services ; l'abus de l'autorité déléguée ; l'augmentation non autorisée des autorisations ; et l'octroi ou la délégation d'un accès ou d'une autorité plus large que nécessaire ou qui n'a pas été documentée ;

---

[29] Pour la classification générale des traités de confidentialité par rapport aux LLM et agents LLM, voir : *Yan B., Xu M., Dong Y., Ren Zh., Cheng X.*, Sur la protection de la confidentialité des données des grands modèles de langage (LLM) et des agents LLM : une revue de littérature, High-Confidence Computing Volume 5, numéro 2, 2025.

[30] Cela inclut les injections indirectes de prompts, où des systèmes interconnectés sont impliqués (dans le cas du RAG et des API externes), conduisant à des fuites de données.

(f) L'utilisation non autorisée ou non intentionnelle d'outils ou de communications externes, y compris la modification, la suppression, la transmission, l'extraction ou la divulgation de données personnelles via des outils, services ou composants de la chaîne d'approvisionnement compromis ; et

(g) La divulgation ou le mélange involontaire de données personnelles entre utilisateurs ou sessions, la propagation d'erreurs ou d'attaques au sein des systèmes multi-agents, et l'insuffisance des moyens de contenir, interrompre ou inverser des actions capables d'affecter significativement les individus.

Les mesures doivent être choisies sur la base des risques identifiés lors de l'examen mené en vertu de l'article 10 et doivent être proportionnelles au contexte particulier. Selon ce contexte, elles peuvent inclure, *entre* autres :

(i) Faire fonctionner les fonctions pertinentes dans un environnement suffisamment isolé et restreindre le système aux modèles, agents, outils, sources de données et points de terminaison approuvés qui, le cas échéant, sont soumis au contrôle des versions et des modifications ;

(ii) Attribuer des identifiants uniques et vérifiables aux agents et autres composants du système à des fins d'authentification, d'autorisation et de responsabilité ; utiliser des identifiants non partagés et, lorsque techniquement faisable, ne sont pas transférables, limités aux autorisations nécessaires à l'objectif spécifié et valables uniquement pour une période limitée ; et établir des autorisations distinctes pour lire, créer, modifier, transmettre, supprimer ou exécuter ;

(iii) Valider les structures de données, paramètres et sorties utilisés lors de l'invocation des outils ; séparer le contenu non fiable des instructions et autres informations de contrôle ; et enregistrer la provenance des données et des indicateurs pertinents pour leur fiabilité ;

(iv) Application de contrôles aux communications sortantes, aux flux transfrontaliers et autres divulgations de données personnelles, ainsi que des mesures contre la perte de données ; et fixant des limites proportionnelles sur le nombre d'étapes ainsi que sur la durée, le volume et la fréquence des opérations, ainsi que sur les destinataires concernés ;

(v) Prévoyant des contrôles indépendants de l'agent ou du modèle concerné et une révision humaine efficace ou une autorisation préalable d'actions pouvant affecter significativement un individu ; et

(vi) Surveiller le système pendant le fonctionnement ; révoquer les identifiants sans délai si nécessaire ; fournir des mécanismes permettant d'interrompre ou terminer l'opération en toute sécurité ; et, lorsque possible, inverser les actions ou en remédier aux conséquences.

Les mesures et technologies améliorant la vie privée, y compris celles destinées à assurer l'anonymisation, la confidentialité différentielle, l'apprentissage fédéré, le

chiffrement homomorphe et d'autres formes de calcul sécurisé, peuvent contribuer à prévenir ou minimiser certains risques lorsque leur adéquation et leur efficacité ont été établies dans le contexte spécifique. Aucune mesure ou technologie ne doit être présumée, en vertu de sa désignation ou de ses caractéristiques techniques, comme rendant les données personnelles anonymes, éliminant le risque de réidentification ou de divulgation, ou garantissant que le traitement soit conforme à la Convention. Son efficacité et ses limites doivent être documentées et mises en examen.

Le respect des exigences de sécurité de la Convention ne doit pas dépendre uniquement des instructions fournies au modèle, des contraintes comportementales générales ou de la capacité du modèle à refuser les demandes. De tels mécanismes doivent, lorsque nécessaire, être complétés par des contrôles techniques et organisationnels dont le fonctionnement ne dépend pas du modèle.

Les mesures organisationnelles doivent inclure des dispositions pour la prévention, la détection et la gestion des incidents ; des responsabilités clairement définies, des procédures d'escalade et de prise de décision ; l'implication effective du responsable de la protection des données lorsqu'il a été désigné ; la formation et la sensibilisation du personnel concernant la protection des données et les risques de sécurité ; et des tests périodiques du système tel que déployé et de toute la chaîne de traitement pertinente.

Les dispositions pour la gestion et la notification des violations de données doivent permettre au responsable de la loi de se conformer à l'article 7, paragraphe 2 de la Convention 108+. Les procédures doivent permettre la détection, la contenance et l'évaluation sans délai des violations de données et autres incidents de sécurité ; conserver uniquement les informations nécessaires et proportionnelles à leur enquête ; prévoir l'examen effectué en vertu de l'article 10 et les garanties à révision ; et permettre au responsable de déterminer sans délai si une violation peut gravement porter atteinte aux droits et libertés fondamentales des personnes concernées. Lorsque ce seuil est atteint, au moins l'autorité de surveillance compétente doit être informée sans délai. Les personnes concernées et les autres responsables doivent également être informés lorsque la loi applicable l'exige.

Recommandations :

- *Les responsables de traitement des données et, le cas échéant, les traités doivent mettre en œuvre des mesures de sécurité techniques et organisationnelles appropriées et adaptatives tout au long du cycle de vie d'un système basé sur un LLM. Les mesures doivent traiter les vulnérabilités existantes et émergentes, y compris les violations de données personnelles, l'empoisonnement des données et de la mémoire, l'injection rapide, les attaques d'inférence et l'utilisation non autorisée d'outils. Elles doivent être intégrées par conception et par défaut, informées par l'évaluation des risques et impacts probables, documentées, régulièrement testées et soutenues par des dispositifs efficaces de prévention et de réponse aux incidents, conformément aux articles 7 et 10 de la Convention 108+.*

- *Les autorités compétentes peuvent adopter et mettre en œuvre des initiatives préventives, notamment des initiatives de sensibilisation, des politiques et directives sectorielles et d'autres mesures appropriées, visant à favoriser le respect des normes de la Convention 108+ et des lois applicables.*

(2) Évaluation des risques liés à la vie privée et vie privée dès la conception

L'article 10 de la Convention 108+ exige que les contrôleurs et, le cas échéant, les traités prennent les mesures appropriées pour se conformer à la Convention et être capables de démontrer leur respect. Les responsables doivent examiner l'impact probable du traitement prévu sur les droits et libertés fondamentales des personnes concernées avant de le commencer et concevoir le traitement afin de prévenir ou de minimiser le risque d'interférence avec ces droits. Des mesures doivent être maintenues tout au long des étapes de traitement concernées. Compte tenu des risques liés à la vie privée introduits par l'architecture opérationnelle et la nature adaptative et interconnectée du système basé sur les LLM, qui entraîne la sensibilité de l'activité de traitement, l'*évaluation ex ante* des risques de vie privée, quel que soit le niveau de risque potentiel[31], doit être considérée comme l'une des obligations principales, imposées par la Convention 108+ et à entreprendre avant le début effectif du traitement. À cet égard, l'évaluation d'impact sur la protection des données doit être comprise comme des efforts contributifs, capables de conclure des résultats probants et de responsabilité réutilisables pour la gouvernance d'un système d'IA plus large, à la lumière des droits humains, de la démocratie et de l'État de droit[32], et sans préjudice des cadres d'évaluation d'impact développés à leur base.

Dans le contexte d'un système basé sur les LLM, l'évaluation des risques de vie privée doit identifier, évaluer et atténuer les risques de confidentialité au niveau du modèle et au niveau du système, en capturant la chaîne de traitement des données tout au long du cycle de vie. Selon l'article 10, l'évaluation du paragraphe 2 doit être réalisée dès le plus tôt stade de la création du modèle. Dans le contexte des écosystèmes LLM, cette obligation doit être comprise comme un effort continu et itératif menant à une évaluation à jour des risques et de l'efficacité de la vie privée, ainsi qu'à la proportionnalité des mesures d'atténuation correspondantes. Les responsables de traitement des données, et lorsque cela est applicable, les traités, doivent concevoir et mettre en œuvre des mesures organisationnelles pour l'examen itératif des évaluations menées et des solutions d'atténuation déployées.

Les efforts d'évaluation des risques doivent commencer par cartographier les opérations de traitement et identifier au cas par cas les rôles et responsabilités entre les acteurs de données impliqués tout au long du cycle de vie du système. Les acteurs

---

[31] Il est précisé que le traitement des données n'a pas besoin de présenter un risque élevé pour les droits et libertés individuels afin de donner lieu à une obligation d'évaluation des risques de vie privée en vertu de l'article 10 de la Convention 108+.

[32] Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit (CETS n° 225); Cadre d'évaluation des risques et de l'impact des systèmes d'IA du Conseil de l'Europe, du point de vue des droits de l'homme, de la démocratie et de l'État de droit (HUDERIA).

de données respectifs (par exemple, développeurs système, fournisseurs, déployateurs) ─ dans le cadre de leur rôle de contrôleur, doivent effectuer une évaluation des risques indépendante dont ils sont responsables. La prise en compte ex ante du ou des contextes applicatifs du système (par exemple, domaines et secteurs) permettra d'identifier à l'avance la sensibilité et la gravité de l'activité de traitement. Lors de l'évaluation des impacts sur les risques, les responsables de traitement et, le cas échéant, les prestataires doivent analyser contextuellement la portée, la nature, le volume et l'objectif des opérations de traitement[33], en respectant un équilibre équitable entre tous les intérêts concernés (principe de proportionnalité) et en tenant compte des points de vue des parties prenantes concernées (par exemple, l'engagement des sujets de données en cas de prise de décision automatisée).

Les responsables des traiteurs doivent documenter les risques identifiés, les mesures adoptées et le risque résiduel qui en résulte, et être capables de démontrer que le traitement est compatible avec la Convention. L'acceptation seule du risque ne prouve pas la conformité. Lorsque cela est approprié, les responsables doivent consulter le responsable de la protection des données et l'autorité de surveillance compétente, conformément à la législation applicable. Les autorités compétentes relevant de l'article 15 de la Convention 108+ sont encouragées à développer des méthodologies et modèles d'évaluation des risques spécifiques à chaque secteur afin de faciliter l'identification et la gestion des risques de vie privée dans divers contextes applicatifs des systèmes basés sur les LLM.

Selon l'article 10, paragraphe 2, les responsables de traitement et, le cas échéant, les processeurs doivent intégrer la protection des données par conception et par défaut tout au long du cycle de vie du système, y compris dans l'architecture, les interfaces, les paramètres par défaut et la gouvernance. Pour les systèmes agents, des contrôles indépendants du modèle doivent techniquement appliquer les limites de tâche et de but ; limiter l'accès et les autorisations à ce qui est nécessaire pour la tâche et la tâche spécifiées ; assurer la séparation et l'expiration de la mémoire ; restreindre l'exploitation aux outils et destinations approuvés ; établir des points d'examen humain efficace ou d'autorisation préalable ; limiter les actions autorisées ; prévoir des moyens de suspension ou de résiliation sécurisée et, lorsque cela est possible, annuler ou remédier ; soutenir une gestion efficace des droits et de la traçabilité des sujets ; et garantir un démantèlement sécurisé. Les mécanismes d'atténuation doivent proposer des mesures techniques et organisationnelles, permettant par défaut de réduire les risques identifiés et intégrer des solutions favorisant la confidentialité, telles que la filtration des données, les mécanismes de minimisation des données et les technologies d'amélioration de la vie privée, dès le stade initial du traitement. Les acteurs des données doivent prendre en compte la nature, la probabilité et la gravité des risques identifiés lors de l'élaboration des stratégies d'atténuation respectives. Cela inclut la mise en œuvre de mesures pour faciliter l'exercice des droits des sujets de données et des mécanismes de réparation pour les individus potentiellement affectés, dès le départ.

[33] Rapport explicatif au Protocole modifiant la Convention pour la protection des individus en matière de traitement automatique des données personnelles, pour 85, 88, 89.

Lorsque les risques identifiés ne peuvent être correctement évités ou atténués, les responsables de traitement des données, et lorsque cela est pertinent, les prestataires de traitement, doivent modifier, redessiner, restreindre, suspendre ou abandonner certaines opérations de traitement des données, fonctionnalités, environnements de déploiement ou architectures système afin d'assurer la conformité à la Convention 108+. Des considérations spécifiques doivent être prises en compte pour les garanties appropriées pour le traitement des données sensibles en vertu de l'article 6. En soutien à une obligation plus large de responsabilité et à la démontrabilité du respect de la Convention, les acteurs des données devraient être encouragés à publier périodiquement des rapports d'évaluation des risques, qui, *entre autres*, contribuent aux obligations de transparence en vertu de l'article 8 de la Convention 108+.

Les contrôleurs et les processeurs doivent utiliser une évaluation, une supervision, un audit et une réévaluation continus pour surveiller les risques résiduels et émergents tout au long du cycle de vie. La réévaluation doit avoir lieu lorsqu'un changement significatif du modèle, des instructions système, de l'orchestration, des outils, des sources de données, de la mémoire, des permissions, du degré d'autonomie, de la composition des agents, des destinataires, du routage géographique ou du contexte de déploiement affecte les finalités, les moyens ou les risques du traitement.

Recommandations :

- *Les contrôleurs doivent évaluer l'impact probable du traitement prévu dès le plus tôt stade et avant le début du traitement, conformément à l'article 10 de la Convention 108+. L'évaluation doit identifier et évaluer les risques tant au niveau du modèle que du système, et définir des mesures pour prévenir ou minimiser l'interférence avec les droits et libertés fondamentales des personnes concernées. Elle doit être documentée, mise en examen tout au long du cycle de vie et renouvelée lorsque des changements importants surviennent. Les responsables doivent fournir les informations et l'assistance nécessaires dans le cadre de leur rôle et de leurs instructions.*
- *Les responsables de traitement des données, et le cas échéant, les traités doivent déterminer le niveau d'impact sur les risques identifiés en tenant compte de l'analyse contextuelle de la portée, de la nature, du volume et de l'objectif des opérations de traitement, en veillant à un équilibre équitable entre tous les intérêts concernés et en tenant compte des points de vue des personnes concernées.*
- *Des stratégies d'atténuation correspondantes devraient être élaborées pour offrir des mesures techniques et organisationnelles appropriées dès la conception et par défaut, permettant de traiter les risques identifiés et d'intégrer des solutions favorisant la vie privée dans les architectures système basées sur des LLM. Un cadre approprié d'évaluation de l'impact des risques implique un examen périodique de l'adéquation des solutions mises en œuvre en tenant compte de la nature, de la portée, de la sensibilité et des impacts négatifs potentiels des activités de traitement. Dans le cadre de la responsabilité et de la transparence, les acteurs des données devraient être encouragés à publier*

une évaluation initiale des risques, ainsi que des rapports de réévaluation ultérieurs.

- *Les autorités compétentes sont encouragées à procéder à des consultations et à adopter des méthodologies et modèles d'évaluation des risques spécifiques à chaque secteur afin de faciliter l'identification et la gestion des risques de vie privée dans les divers contextes applicatifs des systèmes basés sur les LLM.*

(3) Flux transfrontaliers de données personnelles

Les flux transfrontaliers de données personnelles dans les écosystèmes LLM peuvent survenir à plusieurs niveaux du cycle de vie, notamment via des infrastructures distribuées, des services cloud externes, des composants de génération augmentés par récupération hébergée, des bases de données vectorielles, des services outils, des fournisseurs de surveillance ou des agents opérant dans une autre juridiction. L'accès à distance à un outil, le routage dynamique et la délégation entre agents peuvent constituer un transfert ou une divulgation même lorsque l'utilisateur expérimente une interface unique. Chaque opération pertinente doit donc être identifiée et évaluée selon son flux de données factuel. Une dimension de risque supplémentaire en ce qui concerne le niveau de protection approprié est créée lorsque les transferts de données ont lieu dans des juridictions d'un État qui n'est pas Partie à la Convention 108+. Au sens de l'article 14 de la Convention 108+, dans de tels contextes, le flux transfrontalier doit être fondé sur un niveau de protection approprié, et seulement dans des situations limitées, conformément aux dérogations prévues par l'article 14, paragraphe 4.

Sous réserve des garanties prévues à l'article 14 de la Convention 108+, chaque transfert au sein d'une chaîne de système LLM doit être analysé séparément et de manière distincte. Les risques pertinents doivent être abordés par une évaluation des risques de vie privée, en informant sur les impacts négatifs potentiels de tels transferts. Les acteurs des données doivent évaluer et considérer l'adéquation du niveau de protection approprié garanti par la loi ou par *des garanties contractuelles standardisées ad hoc ou approuvées afin d'assurer la continuité d'une protection adéquate garantie par la Convention 108+, indépendamment des chaînes d'approvisionnement, de l'attribution des rôles et des responsabilités, tout au long du cycle de traitement des données.*

Pour les exportations de données hors de la juridiction des Parties, l'utilisation des clauses contractuelles types de la CoE[34] comme outil pour garantir un niveau de protection approprié reconnu par les Parties à la Convention pourrait être envisagée.

Recommandations :

- *Les responsables de traitement des données, et lorsque cela est applicable, les transformateurs doivent mettre en œuvre et maintenir des dispositifs de gouvernance des données permettant une identification précise de la chaîne de traitement des données et des acteurs impliqués*

[34] Comité de la Convention 108 (T-) - Protection des données

tout au long du cycle de vie du système. Cela inclut la cartographie et la documentation des informations relatives aux flux transfrontaliers et à la chaîne de traitement associées au fonctionnement du système. Les acteurs des données doivent effectuer une évaluation appropriée des risques du transfert prévu afin de déterminer si les environnements juridiques et techniques des destinataires de données peuvent affecter le niveau de protection approprié, notamment par non-respect des normes internationales de protection des données telles que définies par la Convention ou par l'accès tiers aux données dans une juridiction donnée.

- *Les responsables de traitement des données et, lorsque pertinent, les prestataires doivent mettre en œuvre des mécanismes contractuels pour établir des obligations exécutoires et un niveau de protection approprié pour démontrer leur conformité à la Convention 108+. Dans ce contexte, les acteurs des données sont encouragés à s'appuyer sur les clauses contractuelles types du Conseil de l'Europe (MCC CoE).*
- *Les responsables des traiteurs et, lorsque cela est pertinent, les prestataires doivent veiller à ce que les personnes concernées soient dûment informées, de manière claire et accessible, lorsque leurs données personnelles peuvent être transférées vers d'autres juridictions, ainsi que les garanties applicables à ces transferts et autres informations pertinentes. Des considérations spécifiques doivent être prises en compte pour les transferts impliquant des catégories particulières de données personnelles ou touchant certaines catégories de personnes concernées (comme les enfants et les individus appartenant à des groupes vulnérables).*

VII. RECOMMANDATIONS SUPPLÉMENTAIRES CONCERNANT LES SYSTÈMES BASÉS SUR DES LLM AGENTS

Les systèmes agents peuvent transformer une seule requête en opérations multiples de traitement, incluant la récupération, l'inférence, le stockage en mémoire, l'invocation d'outils, la communication avec d'autres agents et les actions affectant les personnes n'ayant pas interagi avec le système. Les communications entre agents, les mémoires partagées, les paramètres fournis aux outils, les observations et les sorties intermédiaires peuvent constituer un traitement de données personnelles et ne doivent pas être traités comme de simples opérations techniques internes. Les catégories de données personnelles, de destinataires et de juridictions concernées peuvent changer lors de l'exécution. L'instruction d'un utilisateur à un agent, ou l'acceptation d'une caractéristique agentique, ne doit pas en soi être considérée comme un consentement à tout traitement ultérieur et ne constitue pas, en soi, un consentement au traitement de données personnelles relatifs à d'autres personnes.

Ces caractéristiques peuvent amplifier les risques de collecte excessive, de liaison intercontextuelle et de profilage inférentiel ; mémoire persistante ou inexacte ; divulgation non autorisée via des outils ; utilisation abusive d'identité ou de crédentiel ; propagation d'erreurs entre agents ; et d'actions pouvant produire des effets significatifs ou irréversibles lorsqu'elles reposent sur des résultats probabilistes.

L'accès aux comptes, fichiers, communications, capteurs ou services externes peut encore accroître l'asymétrie informationnelle et réduire la capacité pratique des sujets de données à prévoir, comprendre et exercer un contrôle sur le traitement.[35 ;36]

Les contrôleurs et, le cas échéant, les processeurs doivent définir à l'avance et restreindre par conception la portée de la tâche d'un agent, son accès aux données personnelles, ses autorisations et outils, les formes de mémoire qu'il peut utiliser et la durée de son fonctionnement. Ils doivent limiter l'accès à ce qui est nécessaire pour l'objectif spécifié et la tâche assignée ; distinguer les permissions de lecture des autorisations de création, modification, communication, suppression ou exécution ; exiger un examen humain efficace ou une autorisation préalable avant des actions pouvant produire des effets significatifs ou irréversibles ; et fournir des moyens efficaces pour suspendre ou mettre fin en toute sécurité au système et, lorsque possible, pour annuler des actions ou en remédier les conséquences. Les garanties ne doivent pas dépendre exclusivement des instructions exprimées en langage naturel ou de la capacité du modèle à refuser une demande.

Recommandations pour les contrôleurs et, le cas échéant, les processeurs concernant les systèmes basés sur des LLM agents :

a) Les contrôleurs et, le cas échéant, les responsables doivent établir, maintenir et vérifier périodiquement un inventaire documenté du système ainsi qu'une carte des flux de données personnelles tout au long de la chaîne de traitement.[37]

- *L'inventaire et la carte doivent identifier les modèles et leurs versions pertinentes ; les agents et les composants d'orchestration ; les composants de récupération et les sources de données ; les formes temporaires, persistantes et partagées de mémoire ; les outils et services externes ; les identifiants techniques uniques et vérifiables, les identifiants et les autorisations utilisés à des fins de contrôle d'accès ; les communications et délégations entre agents ; les destinataires ; les lieux de traitement ; les flux transfrontaliers de données personnelles ; et les points où un examen, une autorisation ou une intervention humaine efficace est requis ou disponible.*
- *Chaque composant doit être lié à la personne physique ou morale responsable de son fonctionnement et à son rôle dans le traitement.*
- *La documentation doit couvrir les opérations de traitement intermédiaires pertinentes ainsi que la portée et la nature des actions que le système est autorisé, ou peut raisonnablement être attendu, à exécuter. Elle ne doit pas se limiter à l'entrée initiale et à la sortie finale.*

---

35 Bureau du Commissaire à l'Information, ICO Tech Futures : IA agente — Protection des données et risques liés à la vie privée, 2026

36 *Beduschi A.,* Protection des données à l'ère de l'intelligence artificielle agentique, Computer Law & Security Review, vol. 61, 2026, 106342, DOI : 10.1016/j.clsr.2026.106342.

37 Cyber Security Agency of Singapore, Sécuriser l'IA Agentique – Un Addendum aux Directives et Guide Compagnon pour Sécuriser les Systèmes d'IA, 17 juin 2026, Chapitres 2 et 3.

b) Avant d'autoriser une opération de traitement pertinente, les responsables doivent déterminer et documenter son objectif spécifié, explicite et légitime ainsi que la base légitime applicable établie par la loi, conformément à l'article 5 de la Convention 108+.

- *Un objectif technique attribué à un agent, une instruction d'un utilisateur ou la disponibilité de données personnelles via un outil autorisé ne doivent pas, en soi, être considérés comme un but ou une base légitime pour un traitement ultérieur.*
- *La limitation de l'objectif, la nécessité et la proportionnalité, la minimisation des données et la précision doivent être évaluées et appliquées, selon le cas, en ce qui concerne les entrées, les informations récupérées, les inférences, les informations envoyées et reçues par des outils, les communications entre agents, les opérations mémoire, et toute divulgation ou autre action.*
- *Une attention particulière doit être portée aux données inférées, aux catégories particulières de données personnelles et aux données personnelles relatives aux personnes n'ayant pas interagi avec le système.*
- *Les responsables doivent fournir aux contrôleurs les informations nécessaires pour réaliser et démontrer ces évaluations.*

c) Les contrôleurs et, le cas échéant, les responsables doivent définir, documenter et appliquer des limites vérifiables pour le fonctionnement de chaque agent.[38]

- *Ces limites doivent préciser sa tâche, la durée de l'opération, les données personnelles autorisées, les outils, les destinataires, les lieux de traitement ainsi que la portée et la nature des actions qu'il est autorisé à accomplir.*
- *L'accès de chaque agent aux données personnelles, systèmes et outils doit être attribuable, refusé par défaut, et limité à ce qui est nécessaire pour l'objectif spécifié et la tâche assignée. Les autorisations de lecture doivent être distinguées des autorisations de créer, modifier, transmettre, supprimer ou exécuter, et doivent être accordées séparément. Cela peut être réalisé, par exemple, en attribuant des identifiants techniques distincts et vérifiables à des agents individuels.*
- *Les identifiants utilisés par ou au nom d'un agent doivent être limités à ce qui est nécessaire pour la tâche assignée et pouvant être révoqués sans délai. Ils ne doivent pas être inclus dans les invites, dans la*

[38] Agence espagnole de protection des données (AEPD), Intelligence artificielle agente du point de vue de la protection des données, version 1.2, février 2026, sections VII.E et VII.F ; Infocomm Media Development Authority, cadre de gouvernance de l'IA modèle pour l'IA agentique, version 1.5, publiée le 20 mai 2026 et mise à jour le 5 juin 2026, sections 1.1.3, 2.1.2, 2.2.2 et 2.3.1 ; Eugene Bagdasarian et al., « AirGapAgent : Protecting Privacy-Conscious Conversational Agents », Actes de la conférence ACM SIGSAC 2024 sur la sécurité informatique et des communications, pp. 3868-3882, DOI : 10.1145/3658644.3690350.

mémoire générale ou dans les archives qui ne sont pas soumises à une protection appropriée.

- *Un agent ne devrait pas pouvoir augmenter ses propres autorisations ni accorder des autorisations plus larges à un autre agent uniquement sur la base d'une requête générée par un modèle basé sur un LLM.*
- *Les exigences d'autorisation, les points d'examen humain effectif ou d'approbation préalable, les moyens de suspendre ou de mettre fin en toute sécurité au système, et les moyens d'annuler les actions ou de remédier à leurs conséquences lorsque cela est possible doivent être proportionnés à l'étendue et à la nature des actions autorisées ; au degré d'autonomie ; à la durée et au contexte de fonctionnement ; et à la probabilité, la gravité, la nature cumulative et la réversibilité des effets possibles.*

d) Les contrôleurs doivent adopter et implémenter des règles documentées pour chaque forme de mémoire utilisée par le système, y compris le contexte conversationnel, la mémoire utilisateur persistante, la mémoire partagée, les bases de données utilisées pour récupérer l'information, y compris les mémoires vectorielles, les caches et les enregistrements opérationnels.[39]

- *Les règles doivent préciser quelles informations peuvent être stockées, récupérées, combinées, corrigées ou divulguées ; sa source et sa fiabilité ; les personnes ou agents autorisés à y accéder ; ainsi que les périodes de conservation et de suppression applicables.*
- *Les mesures doivent prévenir l'ajout involontaire d'informations récupérées ou générées à la mémoire, le mélange involontaire de données personnelles liées à différents utilisateurs ou contextes, ainsi que toute utilisation ultérieure à des fins incompatibles.*
- *Les responsables doivent pouvoir localiser les données personnelles dans toutes les formes pertinentes de mémoire et donner un effet pratique aux droits d'accès, de rectification, d'effacement et d'objection applicables. Lorsque nécessaire, les corrections ou suppressions doivent s'étendre aux copies, caches, enregistrements dérivés, agents en aval et processeurs.*
- *Les données personnelles ne doivent pas être conservées uniquement parce qu'elles pourraient améliorer la performance future de l'agent.*

e) Conformément à l'article 8 de la Convention 108+, les responsables doivent fournir aux personnes concernées des informations claires, intelligibles et facilement accessibles sur l'utilisation des systèmes basés sur des LLM agentiques.

---

[39] *Wang B. et al.,* Dévoiler les risques de vie privée dans la mémoire des agents LLM, Actes de la 63e réunion annuelle de l'Association for Computational Linguistics, 2025, pp. 25241-25260, DOI : 10.18653/v1/2025.acl-long.1227 ; *Liu J. et al.,* La topologie importe : Mesurer la fuite de mémoire dans les LLM multi-agents, Findings of the Association for Computational Linguistics : ACL 2026, pp. 39728-39746, DOI : 10.18653/v1/2026.findings-acl.1980.

- *Les informations doivent traiter, selon le cas, de l'identité du responsable du traitement ; de la base juridique et des finalités du traitement ; des principales catégories et sources de données personnelles ; de l'utilisation et de la durée de la mémoire ; des types d'outils, services et autres agents impliqués ; des destinataires et des flux transfrontaliers de données personnelles ; des types d'actions que le système peut initier ; des points où une intervention humaine efficace est disponible ; et des moyens d'exercice des droits des personnes concernées.*
- *Des informations et, le cas échéant, une confirmation doivent être fournies au moment où un agent cherche un accès supplémentaire ou propose une action à laquelle la personne ne s'attendrait raisonnablement pas.*
- *Lorsqu'un flux de travail produit une décision affectant significativement une personne uniquement sur la base d'un traitement automatisé des données personnelles, les garanties de l'article 9 doivent être garanties. En particulier, la personne doit avoir une véritable opportunité d'exprimer ses opinions et de contester la décision.*
- *Toute personne chargée de l'examen doit avoir la compétence, l'information, le temps et l'autorité nécessaires pour refuser, suspendre ou annuler l'action proposée. Une demande de confirmation purement formelle ou générale ne doit pas, en soi, être considérée comme un contrôle humain effectif.*

f) Les contrôleurs et, le cas échéant, les processeurs doivent appliquer des mesures techniques et organisationnelles appropriées à toutes les interactions avec les outils et services externes, les opérations mémoire, les communications et les délégations entre agents.[40]

- *L'agent demandeur et la destination prévue doivent être authentifiés et autorisés. Le but et les paramètres d'une demande, ainsi que tout résultat retourné par un outil ou un service, doivent être validés avant qu'une action susceptible de produire des effets significatifs ne soit effectuée.*
- *Le contenu externe, les informations et messages récupérés auprès d'autres agents doit être évalué à la lumière de leur origine et de leur fiabilité.*
- *Les mesures doivent traiter la manipulation directe ou indirecte des instructions du modèle, y compris l'injection rapide ; l'empoisonnement*

[40] *Greshake K. et al.,* Ce n'est pas ce à quoi vous vous êtes engagé : compromettre les applications intégrées LLM réelles avec l'injection indirecte de prompts, Actes du 16e atelier ACM sur l'intelligence artificielle et la sécurité, 2023, pp. 79-90, DOI : 10.1145/3605764.3623985 ; *Zhan Q. et al.,* InjecAgent : Benchmarking des injections indirectes de prompts dans des agents de grands modèles de langage intégrés par outils, Findings of the Association for Computational Linguistics : ACL 2024, pp. 10471-10506, DOI : 10.18653/v1/2024.findings-acl.624 ; *Debenedetti E. et al.,* AgentDojo : Un environnement dynamique pour évaluer les attaques par injection prompte et les défenses pour les agents LLM, NeurIPS 2024, DOI : 10.52202/079017-2636 ; *Chen Zh. et al.,* AgentPoison : Red-teaming des agents LLM via empoisonnement des bases de mémoire ou de connaissances, NeurIPS 2024, DOI : 10.52202/079017-4136.

des sources de récupération ou de la mémoire ; l'accès non autorisé, la divulgation ou l'extraction de données personnelles ; l'augmentation non autorisée des permissions ; et la propagation d'erreurs entre agents ou composants.

- *Lorsqu'un outil peut produire des effets significatifs ou difficiles à inverser, les responsables doivent limiter les systèmes, les données personnelles et les destinations accessibles par son intermédiaire, et rester en mesure de contrôler les divulgations et transferts qu'il peut entraîner. De telles mesures doivent être proportionnelles à la gravité et à la réversibilité des effets possibles.*
- *La délégation à un autre agent ne doit pas prolonger l'objectif initial, l'accès aux données personnelles, les permissions, les périodes de conservation, les lieux de traitement ou les exigences pour un examen humain.*
- *La sécurité ne doit pas dépendre uniquement des instructions exprimées en langage naturel, des contraintes comportementales générales du modèle ou de la capacité du modèle à refuser une demande.*

g) Les contrôleurs et, le cas échéant, les processeurs doivent maintenir des enregistrements opérationnels sécurisés et proportionnés permettant de retracer et, si nécessaire, de reconstruire des séquences pertinentes d'opérations et d'actions de traitement, et qui soutiennent la démonstration de la conformité.

- *Le contenu et le niveau de détail des enregistrements doivent être proportionnels aux risques liés au traitement et à ce qui est nécessaire pour retracer et, lorsque nécessaire, reconstruire les séquences pertinentes d'opérations et d'actions de traitement. Selon le contexte, les enregistrements peuvent concerner, entre autres, l'agent, le modèle et la configuration concernés ; la tâche assignée ; les sources de données et les autorisations utilisées ; les récupérations, utilisations d'outils et communications pertinentes entre agents ; les destinataires et les lieux de traitement ; les actions tentées, effectuées, empêchées ou annulées ; et les approbations ou interventions humaines. Les enregistrements eux-mêmes doivent être soumis à des limitations de finalité, à la minimisation des données, aux contrôles d'accès, aux garanties assurant leur intégrité et à des périodes de conservation définies. Ils ne doivent pas reproduire systématiquement l'intégralité des données personnelles lorsque des références ou des métadonnées limitées sont suffisantes.*
- *Les traces de raisonnement modèle internes ne doivent pas être considérées comme un enregistrement d'audit fiable ni, en elles-mêmes, comme suffisantes pour fournir la connaissance du raisonnement sous-jacent au traitement des données mentionné à l'article 9, paragraphe 1, sous-alinéa « c ». Toute conservation ou divulgation de telles traces doit avoir une base légitime établie par la loi, servir un objectif spécifié, et être nécessaire et proportionnée.*

- *Les dossiers opérationnels doivent soutenir l'exercice des droits des personnes concernées, la responsabilité interne, l'audit indépendant et l'enquête par les autorités de surveillance.*

h) Avant le déploiement, à des intervalles appropriés et après des changements significatifs, les contrôleurs et, le cas échéant, les processeurs doivent tester l'ensemble du système tel qu'il est déployé et les séquences représentatives de ses opérations et actions de traitement. L'évaluation du modèle sous-jacent seule n'est pas suffisante.

- *Les tests doivent couvrir le fonctionnement normal, l'utilisation abusive raisonnablement prévisible et les conditions adverses pertinentes. Ils doivent inclure la récupération excessive de données personnelles ; les tentatives de manipulation des instructions modèles ; l'extraction ou la divulgation non autorisée des données ; l'utilisation non autorisée d'outils ou l'augmentation des permissions ; la divulgation entre agents ou utilisateurs ; la divulgation entre agents ou utilisateurs ; la mémoire inexacte ou empoisonnée ; la propagation d'erreurs ; l'échec de la révision ou de l'intervention humaine ; et les actions affectant les personnes n'ayant pas interagi avec le système.*
- *Des critères mesurables pour une performance et un risque acceptables, des seuils nécessitant une escalade et des conditions nécessitant une suspension ou un licenciement doivent être définis à l'avance.*
- *Lorsque le traitement prévu est susceptible d'avoir un impact négatif significatif sur les droits et libertés fondamentales des individus, l'examen de son impact probable mentionné à l'article 10, paragraphe 2, de la Convention 108+ ne doit pas être effectué uniquement par des personnes directement responsables de la conception ou du déploiement du système. Les contrôleurs et, le cas échéant, les prestataires de traitement doivent impliquer des personnes disposant d'une expertise appropriée et d'une indépendance de jugement suffisante, et doivent documenter les dispositions mises en place pour garantir l'objectivité de l'évaluation et traiter les conflits d'intérêts potentiels.*
- *Après le déploiement, la surveillance doit être capable de détecter des changements significatifs de comportement, des délégations ou d'une utilisation inattendue d'outils, des violations d'accès, des erreurs récurrentes et une revue humaine inefficace. La surveillance ne doit pas impliquer un traitement supplémentaire inutile ou incompatible.*

i) Les contrôleurs et, le cas échéant, les responsables doivent établir et tester régulièrement des procédures pour gérer les incidents impliquant des systèmes agents basés sur des LLM.[41]

[41] *Shao Y. et al.*., PrivacyLens : Évaluer la conscience des normes de confidentialité des modèles de langage en action, NeurIPS 2024, DOI : 10.52202/079017-2837 ; *Zharmagambetov A.* et al., AgentDAM : Évaluation des fuites de confidentialité pour les agents Web autonomes, avancées dans les systèmes de traitement de l'information neuronale 38, NeurIPS 2025, Datasets et

- *Les procédures doivent permettre au contrôleur, sans délai, de suspendre l'opération de tout agent impliqué, de contenir les effets de l'incident, d'empêcher leur propagation à d'autres agents ou destinataires, et de préserver les informations nécessaires pour évaluer et enquêter sur l'incident. Les mesures prises doivent être proportionnelles à la nature, à la gravité, à l'ampleur et à la réversibilité des effets réels ou raisonnablement prévisibles.*[42]
- *Les procédures doivent également permettre au responsable de la part de reconnaître les individus et les données personnelles concernés, les systèmes concernés et les juridictions dans lesquelles les données ont été traitées ou transférées ; de rectifier les données personnelles inexactes conservées en mémoire ou dans les archives en aval ; d'inverser les actions ou, lorsque cela n'est pas possible, d'en remédier les conséquences ; d'identifier et de traiter la cause sous-jacente ; et d'informer les autres responsables de traiteur, autorités de surveillance et les personnes concernées lorsque la Convention ou la loi applicable l'exige.*[43]
- *Les traitants doivent informer le responsable sans délai de toute violation de données personnelles ou autre incident de sécurité, ainsi que de toute déviation significative par rapport au fonctionnement spécifié ou raisonnablement attendu du système, pouvant affecter la protection des données personnelles ou la capacité du responsable de se conformer à la Convention 108+ et à la législation applicable.*[44]
- *L'évaluation du système tel que déployé et des garanties appliquées doit être examinée à intervalles appropriés et mise à jour si nécessaire. Une nouvelle révision doit être effectuée après un incident, un écart significatif par rapport à l'exploitation prévue ou attendue du système ou la découverte d'une faiblesse grave, et avant qu'un changement susceptible d'affecter la base ou les conclusions de l'évaluation ne soit introduit. De tels changements peuvent concerner le modèle, les instructions régissant le fonctionnement du système, les objectifs, les sources de données, la mémoire, les outils, les autorisations, les destinataires, les emplacements de traitement ou le degré d'autonomie.*
- *Lorsque les risques identifiés ne peuvent être empêchés ou réduits à un niveau compatible avec les exigences de la Convention 108+ et la législation applicable, le traitement concerné ne doit ni commencer ni reprendre et, s'il est déjà en cours, doit être suspendu ou mis fin à cela.*

---

benchmarks Track ; Infocomm Media Development Authority, Cadre de gouvernance de l'IA modèle pour l'IA agentique, version 1.5, publiée le 20 mai 2026 et mise à jour le 5 juin 2026, sections 2.3.2 et 2.3.3 ; NIST, Recommandations et considérations pour la gestion des risques en cybersécurité, SP 800-61 Rev. 3, avril 2025, DOI : 10.6028/NIST.SP.800-61r3.

42 Selon l'architecture du système et la nature de l'incident, ces mesures peuvent inclure l'isolement des composants affectés, la révocation des identifiants d'accès, la suspension d'accès aux outils ou la restriction des communications avec d'autres composants ou services externes.

43 Convention 108+, article 7, paragraphe 2, et article 9, paragraphe 1, litterae « e ».

44 Convention 108+, article 7, paragraphes 1 et 2, et article 10, paragraphe 2.

Recommandations pour les décideurs et les décideurs :

a) *Les États liés par la Convention européenne des droits de l'homme devraient, conformément à leurs obligations positives en vertu de l'article 8, et les parties à la Convention 108+ devraient, dans le cadre de cette Convention, établir, maintenir et examiner périodiquement un cadre juridique et institutionnel clair, accessible et prévisible assurant une protection efficace contre les interférences illégales dans la vie privée et le droit à la protection des données personnelles issues des LLM et des systèmes basés sur des LLM. Le cadre doit couvrir l'ensemble du cycle de vie, y compris les arrangements agents développés ou exploités par des acteurs privés, et doit prévoir des garanties exécutoires, une supervision indépendante et des recours accessibles. Les mesures adoptées doivent être adaptées à la nature, à la gravité et à la prévisibilité des risques et ne doivent pas être comprises comme imposant une obligation de prévenir tout résultat négatif.*[45]

b) *Les parties doivent veiller à ce que les cadres juridiques et politiques nationaux permettent de déterminer les rôles et responsabilités respectifs des personnes physiques et morales impliquées dans un système basé sur un LLM selon leurs fonctions réelles et leur pouvoir décisionnel, et non uniquement par des désignations contractuelles ou techniques. Décrire un système comme agent IA, ou accroître son autonomie technique, ne doit ni conférer une personnalité juridique distincte aux fins des présentes lignes directrices ni interrompre, obscurcir ou diluer la responsabilité des contrôleurs et des processeurs. La délégation au sein d'un système multi-agents doit rester traçable jusqu'aux personnes ou organismes responsables du traitement concerné.*[46]

c) *Les parties doivent veiller à ce que les contrôleurs et, le cas échéant, les traiteurs examinent l'impact probable du traitement prévu et le concevent de manière à prévenir ou minimiser toute interférence avec les droits des personnes concernées, conformément à l'article 10 de la Convention 108+. Dans le cas des systèmes agents, l'évaluation doit concerner le système tel que déployé et les séquences complètes des opérations et actions de traitement pertinentes, et non seulement le modèle sous-jacent ou la sortie finale. Elle doit aborder, selon le cas, les objectifs, les sources de données, les mécanismes de récupération et les supports vectoriels, la mémoire, les outils et services externes, les autorisations et identifiants, la délégation et les communications entre agents, destinataires et flux transfrontaliers de données*

---

[45] En ce qui concerne les obligations positives en vertu de l'article 8 de la Convention européenne des droits de l'homme, voir X et Y c. Pays-Bas, 26 mars 1985, alinéa 23 ; Söderman c. Suède [GC], n° 5786/08, 12 novembre 2013, alinéa 78 ; et K.U. c. Finlande, n° 2872/02, 2 décembre 2008, paragraphes 42-49. Voir aussi la Convention 108+, articles 4, 12 et 15, et la Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit (CETS n° 225), articles 1, 3, 4 et 11.

[46] Convention 108+, articles 2(d) et (f), et Rapport explicatif, paragraphe 22 ; Agence espagnole de protection des données (AEPD), Intelligence artificielle agente du point de vue de la protection des données, version 1.2, février 2026 ; *Beduschi A.,* Protection des données à l'ère de l'intelligence artificielle agentique, Computer Law & Security Review, vol. 61, 2026, 106342, DOI : 10.1016/j.clsr.2026.106342.

personnelles, les points d'examen humain effectif, d'autorisation ou d'intervention, la réversibilité, les effets cumulatifs et en cascade, les abus raisonnablement prévisibles et les personnes n'ayant pas interagi avec le système. L'évaluation doit être renouvelée avant des changements significatifs. Le traitement ne doit pas commencer ou continuer lorsqu'il ne peut être compatible avec les exigences de la Convention 108+.[47]

d) *Les parties doivent veiller à ce que les autorités de surveillance établies en vertu de l'article 15 de la Convention 108+ restent pleinement indépendantes et disposent des ressources financières, humaines et techniques, de l'expertise multidisciplinaire et des pouvoirs nécessaires pour examiner à la fois le traitement au niveau modèle et au niveau système. Cela doit inclure, lorsque nécessaire et proportionné, l'accès à la documentation pertinente, aux documents concernant les flux de données, la provenance et les opérations de traitement pertinentes, les inventaires d'agents et d'outils, les évaluations des risques, les résultats d'évaluation et de tests de sécurité, les dossiers opérationnels, les arrangements contractuels, les informations sur les flux transfrontaliers de données personnelles et les dossiers d'incidents. La coopération avec les autorités compétentes en IA, surveillance des marchés, cybersécurité, protection des consommateurs, concurrence, égalité et droits de l'homme doit être organisée sans réduire les compétences ou l'indépendance des autorités de surveillance de la protection des données.*[48]

e) *Les autorités publiques doivent documenter la nécessité et la proportionnalité d'un système basé sur un LLM, ses avantages attendus et toute alternative moins intrusive avant de l'acheter, de le piloter ou de le déployer. Les documents et contrats d'approvisionnement doivent traiter, selon le cas, des objectifs de traitement ; de l'attribution des responsabilités ; des sources de données autorisées, agents et outils ; des restrictions sur l'utilisation secondaire et la formation supplémentaire ; les sous-traiteurs et services tiers ; la conservation et la suppression ; la mémoire et les autorisations ; les droits de test et d'audit de sécurité ; la notification d'incidents ; les flux transfrontaliers de données personnelles ; les modifications significatives ; le soutien à l'exercice des droits des personnes concernées ; et la restitution ou la suppression des données lors de la résiliation. Les réclamations de confidentialité contractuelle ou de secret commercial ne doivent pas empêcher un accès de supervision efficace ou la fourniture d'informations nécessaires à l'exercice de leurs droits.*

---

[47] Convention 108+, article 10 ; Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit (CETS n° 225), article 16 ; *Barberá I.*, Risques et atténuations de la vie privée de l'IA - Grands modèles de langage (LLM), Pool d'experts du Conseil européen de protection des données, avril 2025 ; *Shao Y. et al.,* « PrivacyLens : Évaluer la conscience des normes de confidentialité des modèles de langage en action », NeurIPS 2024, DOI : 10.52202/079017-2837 ; *Zharmagambetov A.* et al., « AgentDAM : Évaluation des fuites de confidentialité pour agents Web autonomes », NeurIPS 2025, Datasets and Benchmarks Track. Les résultats empiriques concernent les systèmes et tâches évalués et ne doivent pas être généralisés à tous les systèmes agentiques.

[48] Convention 108+, article 15, et rapport explicatif, paragraphes 126-130 ; Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit (CETS n° 225), article 26.

Les systèmes ne doivent pas être acquis ou renouvelés lorsque la conformité ne peut être démontrée.[49]

f) *Les parties doivent garantir la disponibilité de recours accessibles et efficaces ainsi que de garanties procédurales pour les personnes concernées par les systèmes basés sur les LLM et les agents. Cela doit inclure des informations rapides et intelligibles, l'exercice effectif des droits des personnes concernées tout au long de la chaîne de traitement, des moyens de contester les résultats ou actions pouvant affecter significativement un individu, ainsi que la révision, la suspension, la correction et, lorsque cela est possible, la remise en question ou la réparation humaine efficace. Une personne ne doit pas porter la charge d'identifier quel modèle, agent, outil ou service au sein d'une chaîne complexe a causé le traitement contesté. Lorsque la supervision humaine est comptée, elle doit être efficace plutôt que nominale : le réviseur doit disposer de la compétence, de l'autorité, des informations et du temps nécessaires pour refuser, suspendre ou passer outre une action, et la responsabilité des défaillances systémiques de conception ne doit pas être transférée à un utilisateur final ou à un superviseur nominal.*[50]

g) *Les États doivent veiller à ce que les autorités compétentes disposent de pouvoirs juridiques effectifs pour restreindre, suspendre ou interdire les usages des systèmes basés sur des LLM ou agents incompatibles avec les droits de l'homme et les libertés fondamentales ou avec les exigences de la Convention 108+. Conformément à l'article 16, paragraphe 4, de la Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit, les parties à cette Convention doivent évaluer la nécessité d'un moratoire, d'une interdiction ou d'une autre mesure appropriée pour des usages particuliers. Les systèmes agents capables d'initier des actions significatives ou irréversibles ne doivent pas être autorisés lorsque leur autorité et leur accès ne peuvent être efficacement contraints, surveillés et interrompus, ou lorsque les personnes concernées ne peuvent pas obtenir un recours efficace.*[51]

h) *Les parties et les autorités de surveillance doivent renforcer la coopération internationale et l'entraide concernant les systèmes basés sur les LLM et agents dont les modèles, infrastructures, sources de données, outils ou opérations couvrent plusieurs juridictions. La coopération doit inclure l'échange rapide d'informations sur les incidents graves et les risques émergents, des*

---

[49] Conseil de l'Europe, Lignes directrices sur l'intelligence artificielle et la protection des données, T-(2019)01, section III, paragraphes 1-3 ; Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit (CETS n° 225), article 16.

[50] Convention 108+, articles 9 et 12 ; Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit (CETS n° 225), articles 14 et 15 ; *Saar Alon-Barkat S., Busuioc M.,* Interactions humain-IA dans la prise de décision du secteur public : « biais d'automatisation » et « adhésion sélective » aux conseils algorithmiques, Journal of Public Administration Research and Theory, vol. 33, n° 1, 2023, pp. 153-169, DOI : 10.1093/jopart/muac007 ; *Green B.,* Les failles des politiques exigeant la surveillance humaine des algorithmes gouvernementaux, Computer Law & Security Review, vol. 45, 2022, 105681, DOI : 10.1016/j.clsr.2022.105681.

[51] Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit (CETS n° 225), article 16, paragraphe 4.

enquêtes coordonnées ou des actions conjointes lorsque cela est approprié, ainsi que l'application effective des garanties régissant les flux transfrontaliers de données. La répartition d'un système agent entre services ou juridictions ne doit pas conduire à un niveau de protection inférieur ni empêcher l'identification des contrôleurs et traiteurs responsables de certaines opérations de traitement.[52]

Recommandations pour les responsables de traitement des données et, le cas échéant, pour les traités :

a) *Dans le but de respecter les exigences légales inscrites dans la Convention 108+, les responsables de traitement et les traiteurs, et le cas échéant les contrôleurs conjoints, devraient entreprendre des actions d'évaluation et de gestion des risques de vie privée ex ante, basées sur l'approche du cycle de vie du système basé sur les LLM, permettant la mise en œuvre d'un plan d'atténuation approprié axé sur les mesures techniques et organisationnelles, ainsi que des conceptions améliorant la vie privée dans tout son contexte d'application, de développement, de conception et de déploiement.*
b) *Pour les systèmes agentiques, l'évaluation doit porter sur la portée et la nature des actions que chaque agent est autorisé à effectuer ainsi que son degré d'autonomie ; la durée et la possible auto-initiation des tâches ; les outils et accréditations ; la persistance et le partage de la mémoire ; l'accès à des sources externes ou non fiables ; la délégation, la configuration et les voies de communication multi-agents ; les divulgations cumulatives ; les points d'intervention humaine ; la réversibilité ; et les effets en cascade ou en aval.*

[52] Convention 108+, articles 16 et 17 ; Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit (CETS n° 225), article 25.