Strasbourg, le 2 septembre 2025 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 lignes directrices sur la protection de la vie privée et des données personnelles dans le contexte des systèmes basés sur de grands modèles de langage

Schéma

1. Objet et portée

- 1.1 Aider les parties prenantes à identifier, évaluer, atténuer et surveiller (approche IAMM) les risques de confidentialité découlant des systèmes basés sur de grands modèles de langage sur la base des dispositifs et principes de la Convention 108+.
- 1.2 Public visé :
    - États et gouvernements
    - Fournisseurs de LLM (développeurs de modèles et concepteurs de systèmes)
    - Déployeurs (intégrateurs de systèmes et prestataires de services)
    - Industrie
    - Organisations de la société civile
    - Utilisateurs (utilisateurs finaux)
    - Régulateurs / Autorités de surveillance

2. Définitions et terminologie clés

- 2.1 Définitions utilisées dans l'ensemble du document
    - 2.1.1 Différence entre le modèle LLM et le système basé sur LLM :
    - 2.1.2 Définitions de travail (Tech) : modèles de base de l'IA, LLM, petits modèles de langage, agents d'IA, orchestration des LLM, déploiement, inférence, données d'entraînement, risques, etc.
- 2.2 Clarification des concepts essentiels dans ce contexte :
    - 2.2.1 Cycle de vie des LLM et son évolution dynamique (5 étapes fondamentales sélectionnées)
    - 2.2.2 Processus de gestion des risques
    - 2.2.3 Types de risques d'atteinte à la vie privée
    - 2.2.4 Protection des données dès la conception et par défaut
    - 2.2.5 Stratégies d'atténuation

3. Principes de la Convention 108+ pertinents pour les systèmes basés sur la LLM

- 3.1 Décrire les principes fondamentaux de protection des données de la Convention 108+ [Chapitres 2 et 3]
- 3.2 Structure bipartite citant les articles pertinents de la Convention et incluant dans chaque sous-partie a) l'interprétation et l'explication du principe, et b) sa contextualisation.
    - 3.2.1 Légalité et équité
    - 3.2.2 Limitation de la finalité
    - 3.2.3 Minimisation des données
    - 3.2.4 Exactitude et qualité des données
    - 3.2.5 Droits de la personne concernée
    - 3.2.6 Transparence
    - 3.2.7 Sécurité des données
    - 3.2.8 Responsabilisation, etc.

- Expliquez et contextualisez comment ces principes s'appliquent spécifiquement aux grands modèles de langage et aux systèmes basés sur LLM et à leur cycle de vie (par exemple, pendant la formation, la mise au point ou le déploiement).
- Mettre en évidence les tensions et les défis spécifiques que les LLM et les systèmes basés sur les LLM soulèvent en vertu de ces principes (par exemple, la réutilisation des

données, l'hallucination, l'inférence de données personnelles, les risques liés aux données synthétiques).

4. Lignes directrices propres aux intervenants

Pour chaque groupe de parties prenantes (fournisseurs, déployeurs, utilisateurs, régulateurs) :
1/ comment chaque principe se traduit par des obligations ou des responsabilités concrètes pour ce groupe ;
2/ comment chaque acteur doit disposer d'une méthodologie pour identifier, évaluer, atténuer et surveiller les risques d'atteinte à la vie privée ;
3/ avec des exemples de risques pour la vie privée et leurs stratégies d'atténuation ;
Ainsi, fournir (1) une cartographie des principes de la Convention 108+ aux actions, (2) les détails des responsabilités en matière de gestion des risques et (3) des exemples de stratégies d'atténuation et de meilleures pratiques

4.1. Correspondance entre les principes de la Convention 108+ et les actions
- Montrez comment chaque principe se traduit par des obligations ou des responsabilités concrètes pour ce groupe

4.2. Responsabilités en matière de gestion des risques : Identifier, évaluer, atténuer et surveiller l'approche (IAMM)
- Indiquez comment chaque acteur doit suivre l'approche par étapes de l'IAMM :

4.2.1 Déterminer les risques d'atteinte à la vie privée
4.2.2 Évaluer l'impact
4.2.3 Appliquer des stratégies d'atténuation
4.2.4 Surveiller et examiner l'efficacité

4.3. Stratégies d'atténuation et pratiques exemplaires
- Exemple:
  - Fournisseurs : pratiques de conservation des données, confidentialité différentielle, mesures d'évaluation robustes
  - Déployeurs : utilisation de mécanismes de consentement, surveillance rapide et contrôles d'accès
  - Utilisateurs : conception d'invites responsable, cas d'utilisation respectueux de la vie privée
  - Régulateurs : mécanismes de surveillance, renforcement des capacités et critères d'audit

5. Recommandations générales et considérations relatives à la mise en œuvre

5.1 Mettre en évidence les mécanismes de gouvernance et de reddition de comptes
5.2 Favoriser la collaboration interfonctionnelle (équipes techniques, juridiques, éthiques)
5.3 Encourager les évaluations d'impact sur les droits de l'homme et les droits fondamentaux (complémentarité de PIA et HUDERIA)
5.4 Souligner l'interopérabilité avec d'autres cadres réglementaires (p. ex., Loi sur l'IA, RGPD, DSA)

6. Annexes

Annexe 0 : Justifier pourquoi certaines définitions ont été choisies (p. ex., alignement sur la Loi sur l'IA, la Convention 108+, l'OCDE, l'ISO, etc.)

Annexe I : Processus de gestion des risques pour les MLA

- Vue d'ensemble de l'identification, de l'analyse, de l'atténuation et de la surveillance des risques
- Intégration avec les analyses d'impact sur la protection des données (AIPD)

Annexe II : Évolution des phases du cycle de vie des systèmes LLM

- Description détaillée des étapes du cycle de vie pertinentes pour les LLM

Annexe III (facultative) : Études de cas ou exemples

- Exemples illustratifs de risques pour la vie privée dans des déploiements LLM réels
- Comment les différentes parties prenantes les ont abordés
- IA agentique

Annexe IV (facultative) : Glossaire

- Liste des termes clés avec de brèves définitions pour une référence facile

Introduction

Faire face aux défis technologiques en matière de confidentialité et de protection des données
Le rythme de l'évolution technologique observé ces dernières années redéfinit progressivement le sens de la vie privée et de la protection des données. À l'ère des modèles de base avancés de l'IA et des systèmes agentiques basés sur le LLM, ces lignes directrices visent à aborder les risques pour la vie privée de ces technologies et la manière dont elles peuvent interférer avec les droits des individus.

Alors que les LLM continuent de façonner l'avenir des applications d'IA orientées utilisateur et gagnent du terrain dans des secteurs tels que le recrutement, l'éducation, la santé et l'administration publique, il devient de plus en plus difficile d'assurer des protections solides de la vie privée et de faire respecter les droits des individus.

Public et structure
L'élaboration de lignes directrices complètes sur la gestion des risques liés à la vie privée et à la protection des données en vertu de la Convention 108+ fournit aux gouvernements, aux responsables de données et aux autorités réglementaires, ainsi qu'aux concepteurs et développeurs, aux déployeurs et aux utilisateurs finaux (partie 1) les conseils et les outils nécessaires pour identifier, évaluer et atténuer ces risques. Dans le même temps, elle promeut le respect des obligations en matière de protection de la vie privée et des données (partie 4) dans le cadre d'un paysage terminologique cohérent et clair (partie 2).

Le rôle du Comité et ses travaux antérieurs dans l'interprétation de la Convention 108+ dans le contexte des technologies émergentes ont été essentiels pour faire progresser la clarté réglementaire, renforcer la coopération internationale et soutenir l'élaboration de politiques fondées sur la recherche (partie 3).

L'une des principales contributions de ces lignes directrices est d'expliquer et de contextualiser comment les principes de la Convention 108 (cf. chapitres 2 et 3) s'appliquent spécifiquement aux LLM (et aux SLM) et aux systèmes basés sur les LLM tout au long de leur cycle de vie – pendant la formation et la création de modèles, l'adaptation et la mise au point après la formation, l'intégration du système, le déploiement opérationnel et l'interaction avec l'utilisateur final – comme indiqué dans la partie 3.

Approche du cycle de vie : distinguer les modèles LLM et les systèmes basés sur LLM
Une approche fondée sur le cycle de vie fondée sur la Convention 108+, la Convention-cadre sur l'IA et les normes relatives aux droits humains garantit que les principes de protection de la vie privée et des données sont intégrés tout au long du développement et de l'utilisation des systèmes d'IA, plutôt que d'être introduits seulement à des stades ultérieurs après coup.

Les risques liés à la protection de la vie privée et des données des LLM peuvent émerger au cours des différentes phases d'un modèle LLM et des phases du cycle de vie du système (partie 2), allant de la formation et de l'inférence du modèle à l'intégration et au déploiement dans des systèmes plus larges, ayant des implications non seulement pour la protection des données, mais aussi pour la vie privée, la dignité et l'autonomie. Les étapes sélectionnées comprennent :

(1) la création de modèles, où les données d'entraînement sont collectées, prétraitées et les modèles sont construits avec des considérations de confidentialité dès la conception ;
(2) L'adaptation post-formation, où les modèles sont formés, affinés et transformés en outils d'assistance adaptés aux tâches ;
(3) l'intégration de systèmes, dans laquelle les LLM sont intégrés dans des applications ou des services (impliquant souvent la mise au point de modèles pré-entraînés) avec des mesures de protection

appropriées ;
(4) Déploiement opérationnel, c'est-à-dire le déploiement en direct du système basé sur LLM avec une surveillance active et des contrôles de gouvernance ; et
(5) L'interaction avec l'utilisateur final, qui couvre la façon dont les utilisateurs interagissent avec les systèmes basés sur LLM et la façon dont les flux de travail autonomes ou les fonctions agentiques sont gérés.

En fin de compte, l'adoption d'une approche axée sur le cycle de vie permettra de s'assurer qu'à mesure que les technologies LLM progressent, elles le font parallèlement à une gestion robuste des risques d'atteinte à la vie privée (partie 4), en atteignant l'équilibre nécessaire entre l'innovation et la conformité au profit des individus et de la société dans son ensemble.

Gestion des risques : méthodologie basée sur le cycle de vie pour évaluer et gérer les risques d'atteinte à la vie privée associés aux systèmes basés sur la gestion des droits de l'information
S'appuyant sur le rapport d'experts intitulé « Privacy and Data Protection Risks in Large Language Models (LLM) », les lignes directrices s'appuient sur des preuves préliminaires et un raisonnement juridique consolidés et fondés sur la science pour soutenir le leadership du Comité dans la promotion d'un cadre normatif à l'épreuve du temps et la promotion d'une approche cohérente et prospective de la gouvernance de la protection de la vie privée dans les systèmes basés sur le LLM.

En mettant l'accent sur les meilleures pratiques en matière de gestion des risques liés à la vie privée, les lignes directrices veillent à ce que les principes de la Convention 108+ soient traduits en normes pratiques et à ce que les technologies émergentes soient alignées sur les valeurs démocratiques, les droits humains et l'État de droit dans le cadre de la mission mondiale du Conseil de l'Europe.

Alors que l'identification, l'évaluation, l'atténuation et la surveillance des risques de confidentialité et de protection des données associés à l'utilisation de grands modèles de langage (LLM) doivent évoluer en même temps que les progrès technologiques, ces lignes directrices répondent également à la demande croissante des organisations qui déploient des systèmes basés sur des LLM et des flux de travail agentiques pour des cadres pratiques et interopérables qui couvrent le cycle de vie de l'IA. À cette fin, ces lignes directrices intègrent les meilleures pratiques de recherche et de l'industrie, abordent les défis communs et mettent en évidence des stratégies d'atténuation efficaces pertinentes pour les principales parties prenantes, notamment les développeurs, les déployeurs, les chercheurs, les décideurs, les organisations de la société civile et les régulateurs.

Mécanismes de gouvernance
Les présentes lignes directrices promouvront également une approche proactive de l'innovation fondée sur les droits, tout en préservant les principes de transparence, de responsabilité et de dignité humaine qui sont au cœur de la Convention 108+ et de la nouvelle Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle. Ils fournissent des orientations transversales sur les mécanismes de gouvernance et de responsabilisation qui favorisent la collaboration interfonctionnelle entre les équipes techniques, juridiques et éthiques et encouragent l'utilisation d'évaluations d'impact sur les droits humains et les libertés fondamentaux qui sont interopérables dans tous les cadres réglementaires (partie 5). Ce faisant, les lignes directrices contribuent aux efforts normatifs conjoints des comités compétents du Conseil de l'Europe visant à promouvoir une approche intégrée de la gouvernance de l'IA et de la protection des données, complétant les méthodologies d'évaluation et de gouvernance existantes telles que les évaluations d'impact sur la vie privée (EFVP) et HUDERIA, et offrant une perspective plus large centrée sur les impacts systémiques sur les droits humains, la démocratie, la protection de la vie privée et l'État de droit.

Le Conseil de l'Europe est particulièrement bien placé pour veiller à ce que la gouvernance internationale de l'IA et la protection des données évoluent en parallèle, et à ce que la protection de la dignité humaine, de la vie privée et du contrôle démocratique reste au cœur du progrès technologique mondial grâce à ses trois conventions complémentaires (Convention 108+, Convention-cadre sur l'IA et Convention sur la cybercriminalité). Ces mesures soutiendront un cadre solide et évolutif qui permettra à la fois l'innovation et la responsabilisation axées sur les droits humains, tout en assurant la convergence réglementaire et en évitant la fragmentation.

Alors que les systèmes basés sur la LLM continuent de remodeler le paysage numérique, ces lignes directrices offrent des conseils spécifiques à toutes les parties prenantes sur une méthodologie commune de gestion des risques de confidentialité (partie 4), offrant une interprétation exploitable des principes inscrits dans la Convention 108+ (partie 3) et un cadre de gouvernance international coordonné (partie 5), qui peut finalement servir de pierre angulaire pour l'alignement mondial sur la vie privée à l'ère de l'IA générative en conformité aussi avec la Convention-cadre sur l'IA et capable de guider l'innovation responsable au-delà des frontières.

Questions ouvertes à la discussion :

1. En ce qui concerne le titre : « Lignes directrices sur les [risques] de confidentialité et la protection des données dans le contexte des systèmes basés sur la LLM »

OU : « Lignes directrices sur la gestion des risques liés à la vie privée et à la protection des données dans le contexte des systèmes basés sur la LLM »

Toute autre suggestion pour le titre est également la bienvenue.

2. À la section 1.1.2. des experts ont proposé d'inclure le terme d'agents d'IA, et nous aimerions discuter de la façon d'encadrer cette définition.

3. La justification que les lignes directrices fourniront pour le choix des définitions devrait-elle faire partie de la section 2. ou être inclus dans une annexe ?