Protection des données personnelles et nouvelles réglementations RGPD

Tsanta Randriatsitohaina

29 octobre 2025

Objectifs pédagogiques

- Identifier une **donnée personnelle** et les risques associés à son traitement ;
- Expliquer les principes clés de la régulation *Informatique & Libertés* et du **RGPD** ;
- Décrire le rôle et les prérogatives de la **CNIL** ;
- Appliquer ces principes dans un **projet d'IA / data** (collecte, entraînement, déploiement) ;
- Situer la protection des données dans une **perspective internationale**.

Table des matières

1 Partie 1 — Introduction et enjeux

Théorie

Définition. Une **donnée personnelle** est toute information se rapportant à une personne physique identifiée ou identifiable, directement ou indirectement (ex. nom, e-mail, IP, identifiant cookie, voix, visage, données de santé).

Enjeux actuels. Explosion des sources (web, mobiles, objets connectés, logs), généralisation du ML/IA, corrélations massives, risques de *profilage*, *discrimination algorithmique*, *fuites* et *atteintes à la vie privée*.

Exemple

Exemples de données personnelles :

- Directes : nom, prénom, n° de téléphone ;
- Indirectes : adresse IP, identifiant publicitaire, géolocalisation ;
- Sensibles (au sens RGPD) : santé, biométrie, opinions politiques, convictions religieuses, orientation, origine.

Exercice

Ex. 1 — Donnée personnelle ? Indiquez O/N et justifiez en 1 phrase :

1. Numéro de série d'un smartphone d'entreprise ;
2. Photo de groupe publiée sur LinkedIn ;
3. Adresse `contact@entreprise.com` ;
4. Échantillons vocaux pour un modèle d'analyse d'émotions ;
5. Résultats d'un sondage *anonymisés de manière irréversible*.

Mini-TP

Mini-TP 1 — Cartographier un dataset.

1. Prenez un dataset public (p.ex. Titanic ou Adult Income).
2. Classez chaque colonne : *personnelle*, *sensibles*, *non-identifiantes*.
3. Proposez une stratégie d'**anonymisation** ou **pseudonymisation** (suppression, agrégation, hachage, bruit différentiel).

Livrable : un tableau *colonne* → *catégorie* → *mesure de protection*.

2 Partie 2 — Régulation « Informatique & Libertés » et RGPD

2.1 Origines et articulation

Théorie

— **1978** : Loi française *Informatique & Libertés* (création de la CNIL) ;
— **2018** : Entrée en application du **RGPD** (cadre européen harmonisé) ;
— La loi française complète et précise l'application du RGPD en France.

2.2 Principes fondamentaux du RGPD

Théorie

Principe	Description / Illustration
Licéité, loyauté, transparence	Base légale claire (consentement, contrat, intérêt légitime, etc.), information compréhensible.
Finalité déterminée	Les données ne servent qu'à l'objectif déclaré.
Minimisation	Collecter uniquement le nécessaire.
Exactitude	Données à jour, rectification possible.
Limitation de conservation	Durées définies, purge/archivage.
Intégrité et confidentialité	Sécurité technique & organisationnelle (chiffrement, MFA, cloisonnement).
Responsabilité (*accountability*)	Capacité à prouver la conformité (registre, DPIA, politiques).

2.3 Acteurs et droits

Théorie

Acteurs. Responsable de traitement, sous-traitant, **DPO**, personnes concernées.
Droits. Accès, rectification, effacement, limitation, opposition, portabilité ; information & recours.

Exemple

Ex. IA vocale. Une app collecte des échantillons audio pour *détecter l'humeur*.
— Base légale probable : *consentement explicite*.
— Minimisation : pas de métadonnées superflues (géoloc fine, identifiant durable).
— Transparence : notice claire, finalités, durées, droits.
— Droits : effacement à la demande, retrait du consentement.

Exercice

Ex. 2 — Associer principe ↔ mesure
Associez chaque principe à *deux* mesures concrètes dans un pipeline data/ML (collecte, feature store, entraînement, déploiement).

Corrigé (enseignant)

Exemples : Minimisation → schéma restreint, agrégation ; Limitation de conservation → TTL sur objets, tâches d'auto-purge ; Intégrité/confidentialité → chiffrement au repos/en transit, RBAC ; Responsabilité → registre traitements, DPIA pour systèmes à risques ; Transparence → privacy notice versionnée, traçabilité des versions de modèles.

Mini-TP

Mini-TP 2 — *Privacy by Design*

Choisissez un mini-projet par groupe. Rédigez une **fiche de conception** intégrant :

1. **Finalités & base légale** : précisez les objectifs du traitement et sa justification RGPD.
2. **Données collectées** : avant / après minimisation.
3. **Durées de conservation & mécanismes de purge**.
4. **Mesures techniques** : chiffrement, journaux, contrôle d'accès, tests de *membership inference*.
5. **Processus droits RGPD** : accès, effacement, portabilité via endpoint.

Livrable : 1 page + un schéma simple du flux de données (collecte → traitement → stockage → restitution).

Exemples de mini-projets à choisir :

1. **API de recommandation de films (IA collaborative)** : moteur qui recommande des films à partir des historiques d'utilisateurs (collaborative filtering).
2. **Système de détection de churn (abandon client)** : modèle ML prédisant la probabilité qu'un client quitte le service.
3. **Moteur de recommandation e-commerce** : prédiction d'articles susceptibles d'intéresser un utilisateur selon son historique.
4. **Pipeline de détection d'anomalies financières (fraude)** : détection d'opérations bancaires suspectes.
5. **Analyse de sentiments sur avis clients** : modèle NLP qui analyse des textes pour en extraire l'opinion.
6. **Système de recommandation de cours (edtech)** : moteur qui recommande des formations selon le profil de l'apprenant.
7. **Tableau de bord IA de santé publique** : prévision des pics d'hospitalisation à partir de données agrégées.
8. **API de recommandation de playlists audio** : moteur IA qui suggère des morceaux similaires à ceux écoutés.
9. **Assistant conversationnel (Chatbot IA)** : chatbot NLP qui répond automatiquement à des questions clients.
10. **Moteur de recommandation d'emplois (Job Matching)** : algorithme qui recommande des offres selon le profil du candidat.

3 Partie 3 — La CNIL et ses prérogatives

Théorie

Nature : Autorité administrative indépendante (depuis 1978).
Missions : informer, accompagner, *contrôler*, *sanctionner*, traiter les plaintes, publier lignes directrices.
Pouvoirs : mise en demeure, amendes (jusqu'à 20 M€ ou 4% du CA mondial), injonctions, publication.

Exemple

Cas typiques :

- *Profilage publicitaire* sans base légale valable ;
- *Reconnaissance faciale* sur images publiques sans fondement ni information ;
- *Données de santé* hébergées / utilisées hors cadre adéquat.

Exercice

Ex. 3 — Pré-audit CNIL (checklist) Cochez ☐ si en place :

☐ Registre des traitements à jour ;
☐ DPIA pour traitements à risque ;
☐ Politique de conservation et purge automatisée ;
☐ Clauses RGPD avec sous-traitants (DPA), transferts hors UE encadrés ;
☐ Processus droits (portabilité JSON/CSV, effacement, opposition) documenté ;
☐ Sécurité : chiffrement, RBAC, MFA admin, journalisation, tests réguliers.

Mini-TP

Mini-TP 3 — Simuler une réponse à contrôle
En groupe, préparez une *réponse courte* (1 page) à une demande de la CNIL :
« Justifiez la base légale, la minimisation, les durées de conservation et les mesures de sécurité pour votre moteur de recommandation. »

4 Partie 4 — Protection des données dans le monde & IA

4.1 Panorama international

Théorie

Zone	Texte principal	Particularités
UE	RGPD	Cadre global, extraterritorial, droits forts.
USA	CCPA/CPRA (Calif.), HIPAA (santé)	Approche sectorielle/étatique, obligations variables.
Chine	PIPL (2021)	Transferts transfrontières fortement encadrés.
Canada	LPRPDE (PIPEDA)	Proche logique RGPD.
Brésil	LGPD	Inspirée du RGPD, ANPD comme autorité.

4.2 IA, Big Data et risques spécifiques

Théorie

Défis : anonymisation difficile (ré-identification), biais de datasets, opacité des modèles, fuites d'entraînement (*memorization*).
Pistes : *Privacy by Design/Default*, *differential privacy*, *federated learning*, contrôle d'accès au feature store, *model cards*, *data sheets for datasets*.

Exemple

Reconnaissance faciale sur images web « publiques » :
Problèmes de base légale, d'information des personnes, de finalité. Risques de biais ethniques, d'usages détournés. Nécessité d'un fondement juridique solide et d'évaluations d'impact (DPIA).

Exercice

Ex. 4 — Anonymisation vs. pseudonymisation
Expliquez la différence et proposez pour chacun *une* technique et *une* limite dans un contexte d'IA.

Corrigé (enseignant)

Anonymisation : irréversible (agrégation, k-anonymity, bruit DP), mais utilité moindre.
Pseudonymisation : réversible via table de correspondance (hash + sel), réduit le risque mais reste personnelle.

Mini-TP

Mini-TP 4 — Pipeline d'entraînement conforme
Concevez un pipeline (schéma + 8–10 étapes) intégrant : sélection de variables minimisées, consentement/contrat, journalisation des accès, séparation des environnements, *data retention*, *DP noise* sur agrégats, endpoint d'effacement (droit à l'oubli), *model card*.

5 Partie 5 — Synthèse, quiz, ressources

Théorie

À retenir :

- 7 principes RGPD : licéité, finalité, minimisation, exactitude, conservation, sécurité, responsabilité ;
- La CNIL : contrôle, sanction, accompagnement ;
- En IA, la conformité se conçoit *dès le design* : données, modèles, processus.

Quiz rapide (5 questions)

Exercice

1. Définissez « donnée personnelle ». Donnez 2 exemples indirects.
2. Citez 3 bases légales possibles.
3. Donnez 3 droits RGPD d'une personne concernée.
4. Quel est le rôle central du DPO ?
5. Donnez 2 mesures concrètes de minimisation dans un projet de recommandation.

Corrigé (enseignant)

(1) Info liée à une personne identifiée/identifiable ; ex. IP, cookie ID.
(2) Consentement, contrat, intérêt légitime (entre autres).
(3) Accès, rectification, effacement, portabilité, opposition, limitation.
(4) Conseiller, contrôler, point de contact autorité/personnes.
(5) Réduire les champs, agréger catégories, supprimer PII inutiles.

Ressources

- CNIL — guides RGPD, IA & données personnelles : https://www.cnil.fr
- EDPB (Comité européen de la protection des données) : lignes directrices
- NIST Privacy Framework (complément méthodologique)

Ce polycopié est fourni à des fins pédagogiques et doit être adapté au contexte de chaque traitement de données.

6 Annexe

Corrigé (enseignant)

(1) O (identification indirecte possible) ; (2) O (visages) ; (3) N (générique, pas une personne) ;
(4) O (voix = biométrie potentielle) ; (5) N si anonymisation *irréversible*.