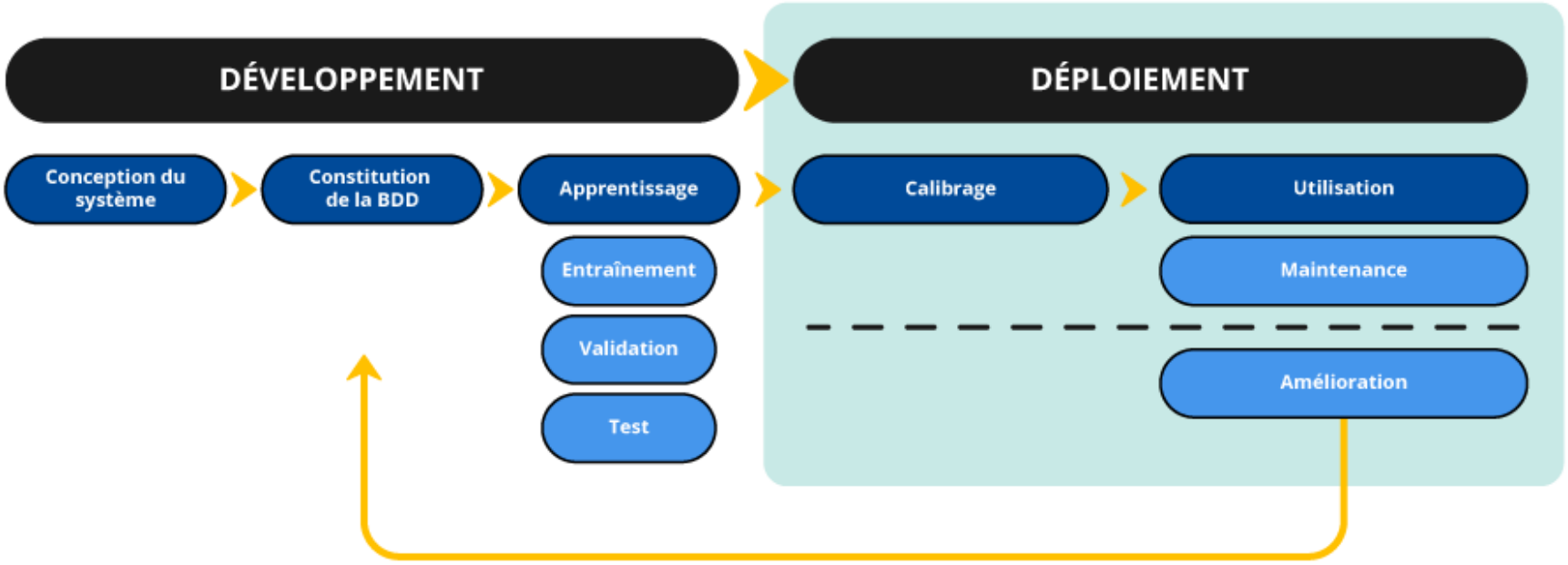
DPO Forum Nantes – 26 juin 2025

RGPD et IA
encadrer l’usage des données pour innover

Matthieu Camus
matthieu.camus@privacyimpact.fr

Systèmes d’IA concernés par le RGPD

- **Systèmes d’IA impliquant un traitement de données à caractère personnel**
  - Attention particulière au cours de la phase d’entraînement



Définition de l’objectif du système d’IA

- **Finalité** = objectif défini pour cadrer et limiter les données utilisées → **déterminé, explicite, légitime**

- Situation 1 : **Système spécifique** avec connaissance de l’usage opérationnel du système
  - Objectif de phase de développement = objectif de phase de déploiement et d’utilisation
- Situation 2 : Système d’**IA à usage général**
  - Objectif spécifique au développement, avec référence au type de système et aux fonctionnalités et capacités techniques (par ex. développement d’un système de vision nocturne avec reconnaissance des formes)
  - Attention : pas de finalité du type « développement et amélioration d’un système d’IA »
- Situation 3 : Système d’**IA à des fins de recherche scientifique**
  - Objectif décrivant la démarche détaillée par exemple à des fins de démonstration de robustesse du système

Détermination des responsabilités

- Qualification du niveau de responsabilité RGPD : **responsable de traitement** (RT) ou **sous-traitant** (ST)
- Le règlement européen sur l'IA définit plusieurs rôles
  - Le **fournisseur du système d'IA** développe ou fait développer le système et le met sur le marché
  - Les **importateurs**, **distributeurs** et **utilisateurs** (déployeurs) du système d'IA
- Le degré de responsabilité est à analyser au cas par cas
  - Le fournisseur agissant pour son propre compte est RT de la partie développement (base de données d'apprentissage)
  - Attention : en cas de plusieurs acteurs, possibilité de qualification de **responsables conjoints** du traitement
  - Si un client détermine **l'objectif**, **les moyens** et **les techniques** du développement alors le fournisseur est ST
  - Tous les prestataires agissant sur instructions sont ST (par ex. pour la constitution de la base de données)
- Responsabilités du ST
  - Vérification de la **conformité du contrat**, agissement uniquement sur **instructions du RT**
  - **Sécurisation** des données, **conformité** au RGPD, **assistance** au RT

Définition de la « base légale »

- Choix d'une base légale au sens du RGPD : consentement, obligation légale, contrat, intérêt public, intérêts vitaux, intérêt légitime
- Si les données sont collectées directement auprès des personnes libres d'accepter ou de refuser → consentement
  - Consentement **libre**, **spécifique**, **éclairé** et **univoque**
- Parfois difficile d'obtenir le consentement → possibilité de l'intérêt légitime
  - Besoin de justifier la légitimité → **légal**, **défini** de manière précise et **réel**
  - Justifier la nécessité des données personnelles pour l'entraînement du système (usage de **données anonymes** ?)
  - Usage qui ne doit pas porter une « **atteinte disproportionnée** » à la vie privée des personnes
- Cas spécifique des acteurs publics → traitement relatif à la mission d'intérêt public ?
- Autres cas plutôt marginaux : contrat, obligation légale

Possibilité de réutilisation des données

- Attention à la légalité de la réutilisation de données → responsabilité du RT
- Cas 1 : données directement collectées par le fournisseur
  - Vérification de la **compatibilité avec l'objectif initial**, sauf si **consentement** des personnes ou **autorisation** réglementaire
  - **Test de compatibilité** : existence d'un lien entre objectifs, contexte de collecte, type et nature des données, conséquences sur les personnes, garanties appropriées (pseudonymisation ?)
  - En cas de réutilisation des données à des fins statistiques ou de recherche scientifique, alors présomption de compatibilité
- Cas 2 : données publiquement accessibles (*open source*)
  - Démarche à **documenter** : description de la source des données, licéité initiale, absence de données sensibles ou d'infraction
- Cas 3 : données acquises auprès d'un tiers (*data broker*)
  - Si collecte initiale pour apprentissage système d'IA alors vérification de conformité au RGPD (objectif, licéité, information, droits…)
  - Si collecte initiale avec objectif différent, nécessité de réaliser le test de compatibilité

Principe de minimisation

- Données collectées **adéquates**, **pertinentes** et **limitées** au strict nécessaire au regard de l'objectif
- Recours au *deep learning* non systématique
- Validité des **choix de conception**
  - **Étude pilote à petite échelle**, si possible avec données fictives, synthétiques, anonymisées
  - Appel à un comité éthique pour prise en compte des enjeux de protection des droits et libertés des personnes
- Vérification de la pertinence des données au regard de l'objectif
  - **Nettoyage** des données → amélioration de la qualité des données
  - Identification des **données pertinentes** → optimisation du système pour éviter tout sous- / sur- apprentissage
  - Mesures de **Privacy by-design** → application de transformation sur les données (random, anonymisation…)
  - Suivi et **mises à jour** des données → éviter toute dérive des données et dégradation des performances
  - **Documentation** des données → garantie de traçabilité des jeux de données

Choix d’une durée de conservation

- **Durée de conservation des données initiales en fonction de l’objectif** du traitement
- Phase de développement → **planification** en amont et **suivi** dans le temps
  - Information nécessaire auprès des personnes
- Phase de maintenance ou d’amélioration → principe de la **suppression**
  - Conservation possible avec garanties nécessaires (support cloisonné, restriction des accès…)

- Argument : données d’apprentissage utiles pour **réalisation d’audits et mesure des biais**
  - Justification adéquate en cas d’insuffisance d’informations générales sur les données
  - Vérification du principe de minimisation
  - Besoin de mesures de sécurité renforcées

Réalisation d’une AIPD

- **AIPD → Analyse d’impact sur la protection des données**
  - Démarche de cartographie et d’évaluation des risques
  - Établissement d’un plan d’action pour réduire les risques à un niveau acceptable
- Réalisation si **données sensibles**, **large échelle**, **personnes vulnérables**, **croisement**, **solution innovante**
- Règlement européen sur l’IA : **système à haut risque = AIPD**
- Si développement ↔ usage opérationnel → AIPD sur l’ensemble du cycle de vie du traitement (attention RT/ST...)
- Cas spécifique d’un système d’IA à usage général → AIPD pour phase de développement uniquement
  - Risques spécifiques : atteinte à la confidentialité, mésusage des données, discrimination automatisée, production de contenu fictif sur une personne réelle, décision automatisée, perte de contrôle des données utilisateurs librement accessibles, attaques spécifiques, risques éthiques systémiques
  - Mesures de sécurité, principe de minimisation, anonymisation et/ou pseudonymisation, Privacy by-design, facilitation des droits ou recours, mesures d’audit et de validation

MERCI

PrivacyImpact
www.privacyimpact.fr