# Déployer un LLM : quels risques pour les données personnelles ?

*26 juin 2025*

# Ordre du jour

- Introduction
- Les enjeux
- Les LLM
- La maîtrise des risques
- Conclusion

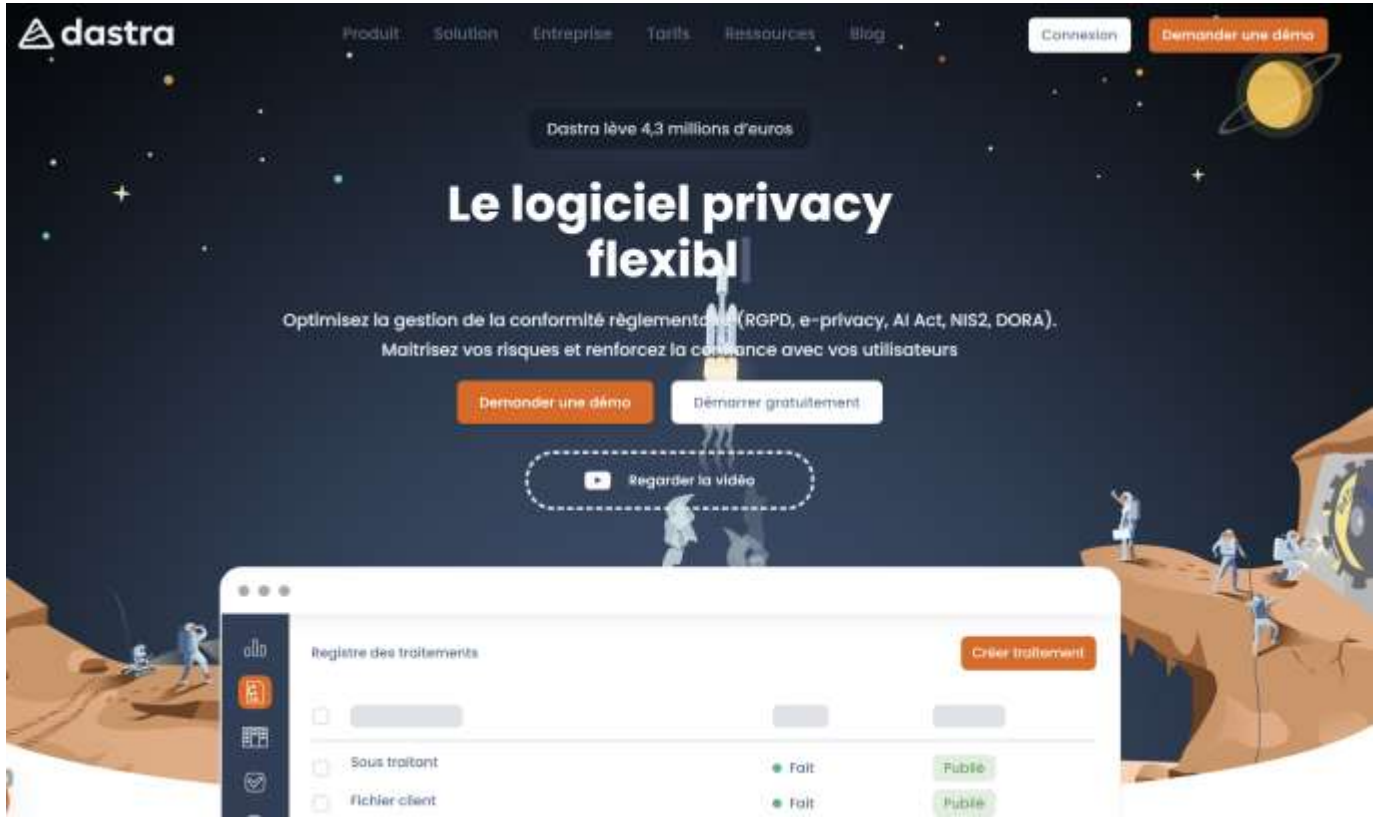
dastra

# Votre intervenant

## Jérôme de Mercey

dastra

Cofondateur et COO de Dastra.
Ancien juriste au service des contrôles de la CNIL, j'ai développé une expertise sur l'application des règles de protection des données




dastra

# Ordre du jour

- Introduction
- Les enjeux
- Les LLM
- La maîtrise des risques
- Conclusion

dastra

# Les LLM en entreprise : promesses et réalité

- 📈 Explosion des cas d'usage : support client, rédaction, RH, analyse documentaire…

- ⚙️ Intégration facile via API → déploiements rapides

- ❗ Problème : ces outils manipulent des volumes massifs de données sans transparence native

Comment garantir le respect du RGPD dans ce contexte ?

dastra

# Enjeu #1 : Alimentation du modèle

🔍 **Problème : ingestion de données personnelles**

- Documents internes, tickets clients, emails…
- Absence de tri = données personnelles + sensibles = danger

dastra

# Enjeu #2 : Sortie du modèle et hallucinations

**Le contenu généré est-il conforme ?**

- Peut-il contenir des données personnelles ? → Oui
- Peut-il inventer des faits faux ? → Oui (hallucinations)

dastra

# Enjeu #3 : Droits des personnes

**Les LLM posent un défi au RGPD :**

- Impossible d'accéder aux « données sources »
- Pas de traçabilité sur les entrées / sorties
- Aucune garantie de suppression effective

→ Droit d'accès, de rectification ou d'effacement difficilement applicables

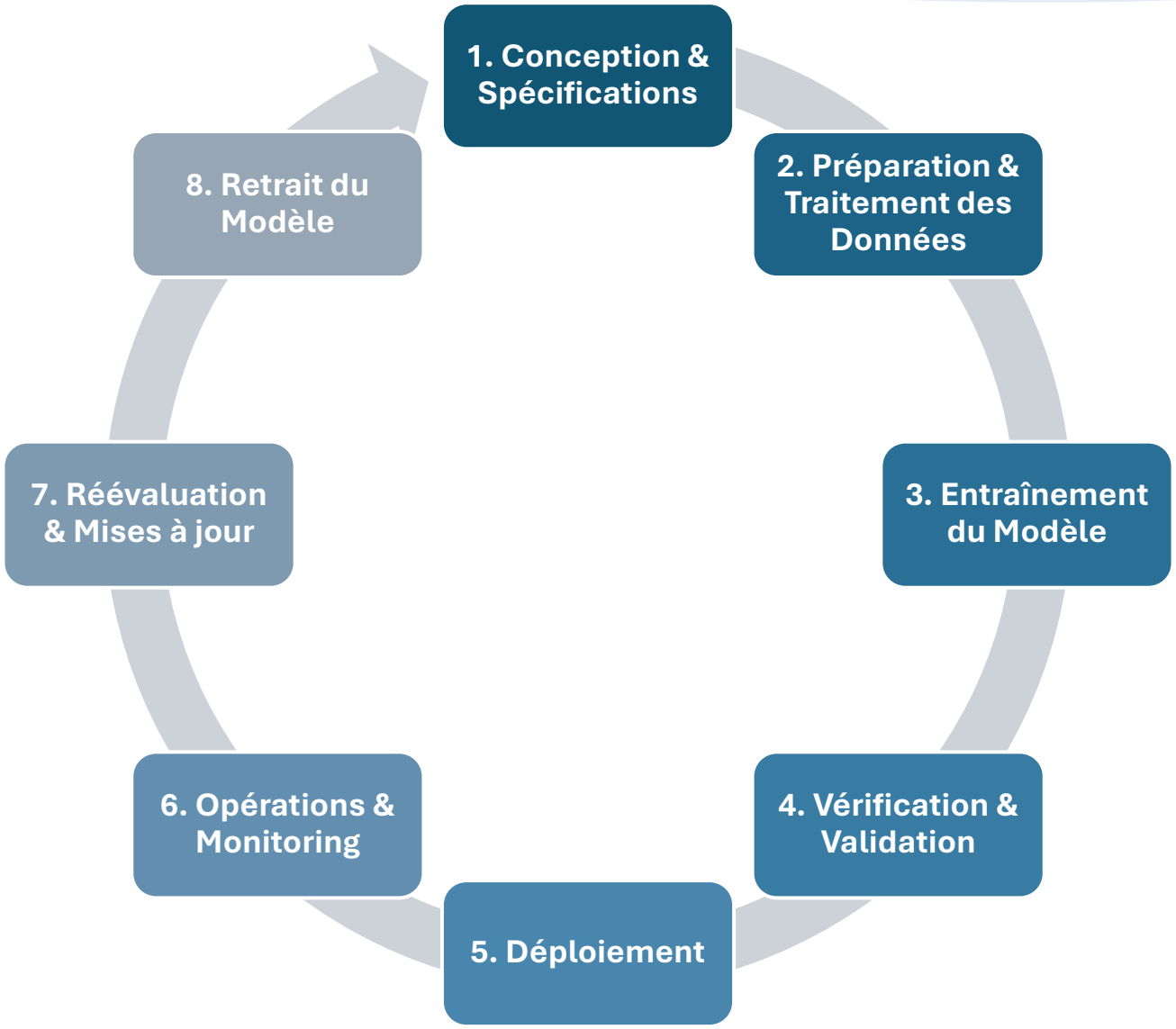
dastra

# Ordre du jour

- Introduction
- Les enjeux
- Les LLM
- La maîtrise des risques
- Conclusion

dastra

# Le cycle de vie d'un système d'IA




dastra

# Trois modèles principaux de déploiement des LLM

| Modèle | Description | Exemples | Maitrise du LLM |
|---|---|---|---|
| **1. LLM as a Service** | Accès au modèle via API cloud, sans accès aux poids. Déploiement rapide mais dépendance complète au provider | GPT-4 API (OpenAI)Claude API (Anthropic) | Faible à modéré |
| **2. LLM Off-the-Shelf** | Le deployer choisit un modèle, personnalise les poids (fine-tuning), souvent via une plateforme cloud ou local | LLaMA, BLOOM, Azure OpenAI, Bedrock | Modéré à élevé |
| **3. LLM développé en interne** | Le modèle est entraîné, hébergé et contrôlé par l'organisation elle-même. Autonomie totale mais complexité importante | Cas R&D avancé / grands groupes / start-up IA | Très élevé |

dastra

# Distinction Provider / Deployer (AI Act) :

**Provider** : l'entité qui développe et offre le système d’IA

**Deployer** : l'entité qui intègre et utilise le système auprès des utilisateurs finaux

dastra

# Responsabilités RGPD selon le modèle LLM utilisé

| Modèle | Déployeur comme responsable du traitement | Fournisseur comme responsable du traitement | Fournisseur comme sous-traitant |
|---|---|---|---|
| **LLM as a Service** | Définit les finalités et moyens (requêtes utilisateurs, usages métiers) | Réutilise les données pour entraînement, monitoring ou analytics | Traite les données selon les instructions du déployeur (API, hébergement) |
| **LLM "Off-the-Shelf"** | Personnalise le modèle, contrôle les données (prétraitement, outputs, workflows) | Rétention ou réutilisation des données (debugging, amélioration continue) | Plateforme agissant selon les instructions du déployeur (cloud, API managée) |
| **Self-developed LLM** | Contrôle complet : objectifs, entraînement, hébergement | Non applicable : le fournisseur et le déployeur sont confondus | Non applicable |
| **Agentic AI** | Gère entrées, mémoire, tâches, interactions avec outils externes | Peut conserver les données d'interaction pour améliorer ses modules ou composants intelligents | Fournisseurs d’outils ou d’APIs exécutant des actions à la demande du déployeur (sous-instructions) |

dastra

# Ordre du jour

- Introduction
- Les enjeux
- Les LLM
- La maîtrise des risques
- Conclusion

dastra

# Risques privacy d’un LLM

| Risque | Description | Articles RGPD |
|---|---|---|
| **1. Protection insuffisante des données** | Manque de chiffrement, accès non autorisé, failles API, logs non sécurisés → exposition aux fuites | Art. 32, Art. 5(1)(f), Art. 9 |
| **2. Mauvaise anonymisation** | Des données mal pseudonymisées sont utilisées pour l'entraînement → inférences possibles | Art. 5(1)(a)(b), Art. 25 |
| **3. Traitement illicite dans l'entraînement** | Aucune base légale pour utiliser des données personnelles dans les jeux d'entraînement | Art. 5(1)(a)(c), Art. 6(1), Art. 7 |
| **4. Traitement de données sensibles ou judiciaires** | Utilisation de données santé, religion, casier judiciaire sans cadre légal clair | Art. 9(1)(2), Art. 10 |
| **5. Impact négatif sur les personnes** | Sorties biaisées ou erronées utilisées pour prendre des décisions | Art. 5(1)(d)(a), Art. 22, Art. 25 |
| **6. Absence d'intervention humaine** | Décisions automatisées (prêt, recrutement) sans possibilité d’intervention humaine | Art. 22(1)(3), Art. 12 |

dastra

# Risques privacy d’un LLM

| Risque | Description | Articles RGPD |
|---|---|---|
| **7. Non-respect des droits des personnes** | Impossible de corriger, supprimer ou restreindre les données contenues dans un modèle | Art. 12–14, 16–18, 21 |
| **8. Réutilisation illicite des données** | Utilisation des prompts ou outputs à des fins autres que prévues sans information | Art. 5(1)(b)(a), Art. 28(3)(a), Art. 29 |
| **9. Stockage illimité** | Conservation de prompts, logs ou outputs au-delà de ce qui est nécessaire | Art. 5(1)(e), Art. 25 |
| **10. Transferts illicites hors UE** | Traitements effectués dans des pays non adéquats sans garanties (ex : cloud LLM Chine) | Art. 44–46 |
| **11. Non-respect de la minimisation** | Collecte de très grandes quantités de données pour entraîner ou ajuster le modèle | Art. 5(1)(c), Art. 6(1)(f), Art. 25 |

dastra

# Risques & mesures selon les phases d'un LLM as a Service

**Chaque phase du système doit être étudiée sous l'angle du risque privacy**

| Phase | Risques majeurs | Mesures d'atténuation clés |
|---|---|---|
| **1. Entrée utilisateur** | - Fuite de données sensibles<br>- Accès non autorisé<br>- Attaques par injection / jailbreaking<br>- Manque de transparence | ✓ Filtres et alertes pour prompts sensibles<br>✓ Anonymisation automatique<br>✓ Chiffrement TLS + repos<br>✓ MFA + mot de passe robuste<br>✓ Politique de confidentialité claire |
| **2. Interface / API** | - Interception réseau<br>- Exploitation API<br>- Interfaces vulnérables / phishing | ✓ Chiffrement de bout-en-bout<br>✓ Authentification forte (OAuth, API keys)<br>✓ Sécurité OWASP<br>✓ Anti-phishing, branding protection<br>✓ Audit & journalisation |
| **3. Traitement LLM** | - Inférences sensibles / hallucinations<br>- Logging excessif<br>- Attaques par empoisonnement (poisoning)<br>- Accès non autorisé aux logs | ✓ Filtrage de contenu + revues humaines<br>✓ Journaux minimisés et chiffrés<br>✓ Accès restreint et surveillé<br>✓ Anonymisation avancée<br>✓ Gouvernance fournisseurs cloud |
| **4. Sortie générée** | - Réidentification possible<br>- Contenu inexact ou sensible<br>- Mauvais usage des réponses | ✓ Filtrage post-processing<br>✓ Redaction / limitation contexte<br>✓ Politique d'usage des outputs<br>✓ Revue humaine pour usages critiques<br>✓ Formation des utilisateurs |

dastra

# Ordre du jour

- Introduction
- Les enjeux
- Les LLM
- La maîtrise des risques
- Conclusion

dastra

# Comment encadrer un projet LLM ?

**Approche recommandée**

- Cartographier les cas d’usage
- Restreindre les données personnelles utilisées
- Contrat clair avec les fournisseurs (pas de réutilisation)
- Définir ses responsabilités
- Réaliser une AIPD

dastra

# Merci

jerome.demercey@dastra.eu

www.dastra.eu