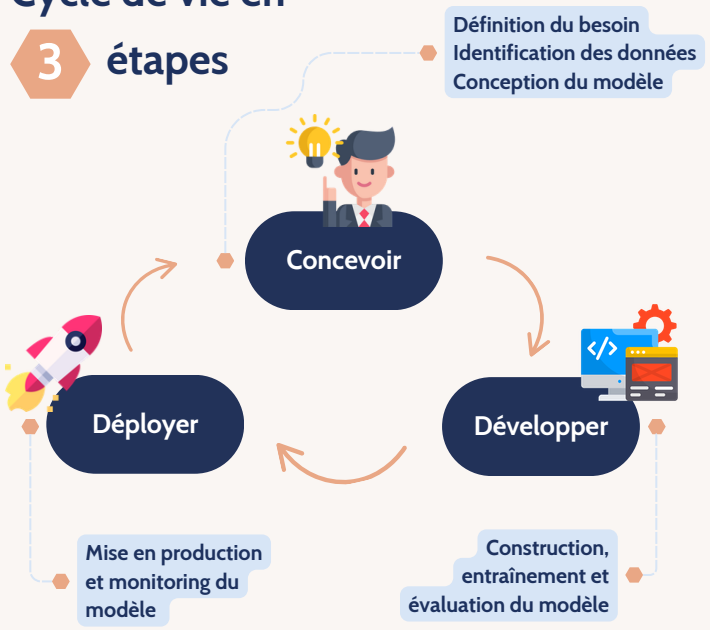
3 Grands types d'IA

ANI – Artificial Narrow Intelligence
IA étroite ou faible conçue pour exécuter des tâches spécifiques

AGI – Artificial General Intelligence
IA générale ou forte capable de comprendre et d'apprendre dans des domaines variés comme l'homme

ASI – Artificial Super Intelligence
IA super intelligente capable de dépasser l'intelligence humaine

Cycle de vie en 3 étapes



3 Grandes catégories d'attaques sur l'IA générative

- **Attaques par manipulation** – Détournement du comportement du système d'IA
- **Attaques par infection** – Contamination du système d'IA pendant son entrainement
- **Attaques par exfiltration** – Vol des informations du système d'IA

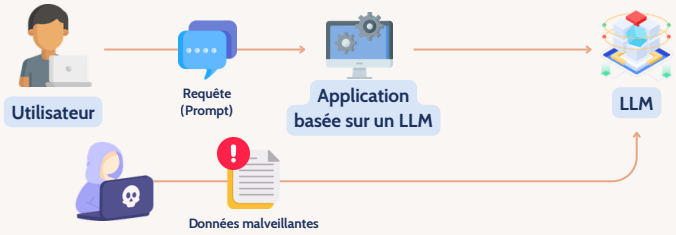
IA, Sécurité et Conformité

OWASP Top 10 for LLM Applications

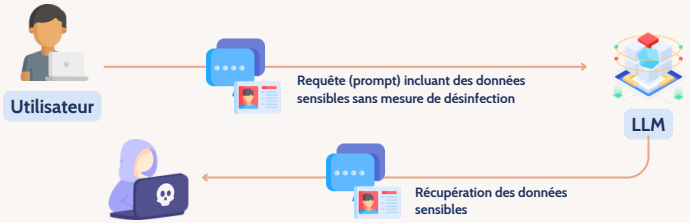
1. Infiltration de requêtes
2. Divulgation d'informations sensibles
3. Chaîne d'approvisionnement
4. Empoisonnement des données et du modèle
5. Mauvaise gestion des sorties du modèle
6. Autonomie excessive
7. Fuite du prompt système
8. Vulnérabilités des représentations vectorielles
9. Désinformation
10. Consommation non bornée

4 Exemples de scénarios d'attaques de systèmes d'IA

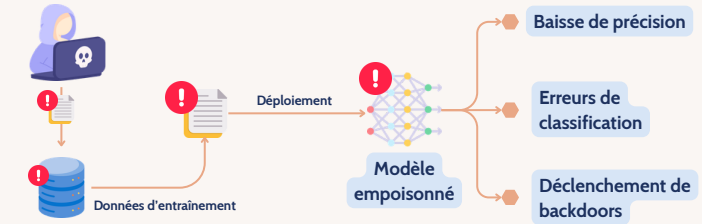
1 - Infiltration indirecte de requêtes



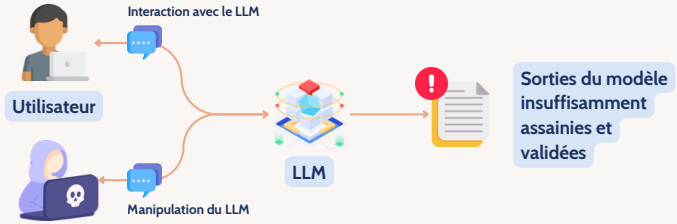
3 - Divulgation d'informations sensibles



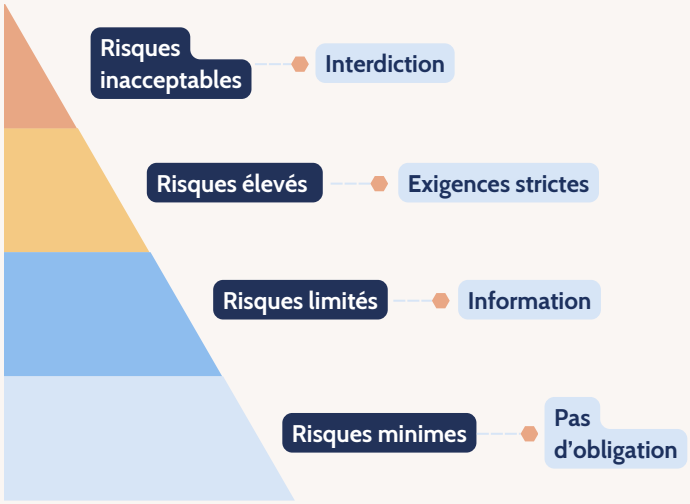
2 - Empoisonnement des données



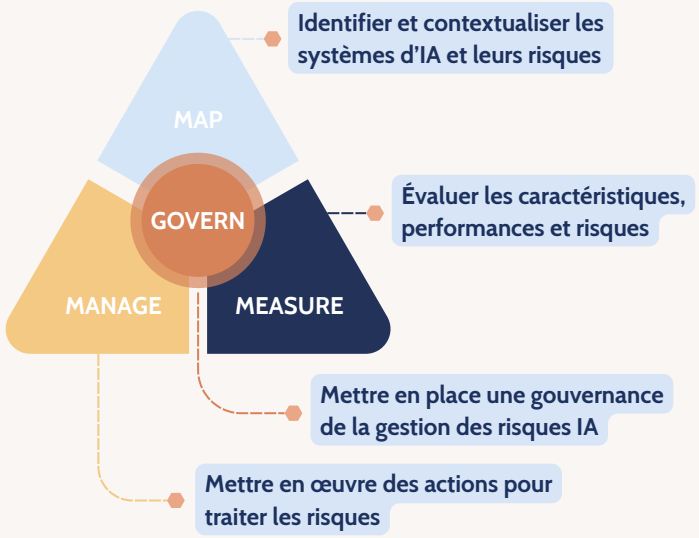
4 - Mauvaise gestion des sorties du modèle



4 Niveaux de risques (AI Act)



4 Piliers pour identifier, évaluer et gérer les risques (NIST)



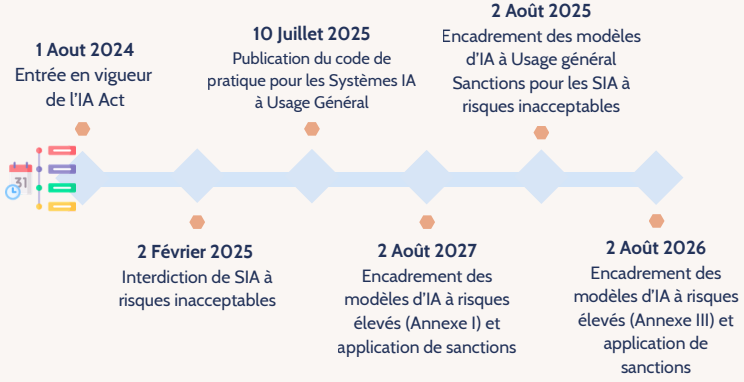
9 Acronymes incontournables

LLM	Large Language Model
RAG	Retrieval-Augmented Generation
NLP	Natural Language Processing
GPU	Graphics Processing Unit
LoRA	Low-Rank Adaptation
RLHF	Reinforcement Learning from Human Feedback
GAN	Generative Adversarial Network
TTS	Text-to-Speech
ASR	Automatic Speech Recognition

WLF IA, Sécurité et Conformité WLF

Garantir les valeurs fondamentales de l'UE en établissant un cadre harmonisé sur l'IA

- Renforcer la sécurité et la transparence des SIA
- Promouvoir une IA étique et digne de confiance
- Harmoniser des règles dans l'UE
- Soutenir l'innovation en créant un cadre stable



8 Systèmes d'IA interdits (AI Act)

- Techniques subliminales
- Exploitation des vulnérabilités de personnes fragiles
- Notation sociale
- Prédiction policière
- Constitution de base de données de reconnaissance faciale
- Evaluation des émotions
- Catégorisation biométrique sensible
- Identification biométrique en temps réél

"Chez WLF, nous combinons expertise en cybersécurité et maîtrise de l'IA pour protéger vos systèmes, vos données et vos utilisateurs."

Contactez-nous !

www.wlf.fr
contact@wlf-services.fr
www.linkedin.com/company/wlf-services

WLF · CYBERSÉCURITÉ · CABINET DE CONSEIL