CYBERSECURITY

SEMINAR FRIBOURG

Artificial intelligence

27 mars 2025

oxygen
data pilots

oxygen
data pilots

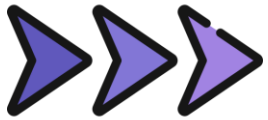
L'IA au service des cyberattaques

- Démonstration -

Par Ruben Terceiro

Ingénieur des données chez
Oxygen Data Pilots et étudiant
Master en Data Science

Démonstration « deepfake »



C’est quoi le « Hacking » ?

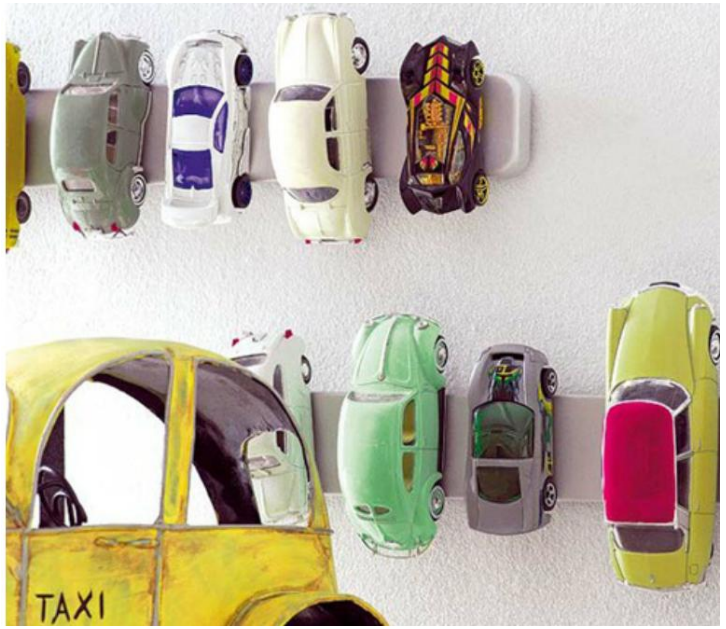
« *Hacking : Pénétrer un système informatique de force.* »

-- Source: Office fédéral de la statistique --

…mais plus largement

C’est quoi le « Hacking » ?

-- BuzzFeed.de © instagram.com --



-- BuzzFeed.de © style-files.com --

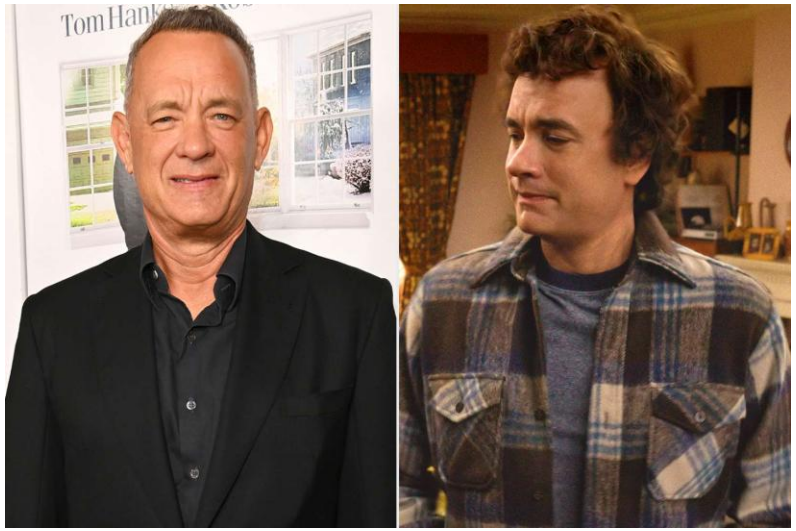
« On peut définir le hacking comme le contournement de l'usage premier ou prévu d'un objet, d'un système ou d'une technologie. »

-- Source: personnelle --

Un cas d'utilisation non malveillant

Tom Hanks et Robin Wright ont été rajeunis "en temps réel" grâce à l'IA dans le nouveau film de Robert Zemeckis

-- Source: *FranceInfo* --



"Nous n'avons pas eu à attendre huit mois de post-production".

"Il y avait deux moniteurs sur le plateau. L'un affichait le flux réel capté par l'objectif et l'autre était juste, une nanoseconde plus lent et affichait notre version 'deep fake'.»

-- Source: BMFTV --

oxygen
data pilots

L’IA au service de la cybersécurité

Tobias Kull

Directeur BU Cyber chez Oxygen Data Pilots et chargé de cours à l’HEIA

Quelles sont les attentes ?

La promesse IA permettrait **l'amélioration** des concepts de cybersécurité **existants,** au travers des outils intégrant de l'intelligence artificielle

- Automatisation face à des comportements inconnus
- Suggestion de réponses en fonction de l'évènement
- Proactivité et surveillance en temps réel

La boîte à outils des cyber consultants 1/3

Les « Frameworks » ou « lignes directrices » apportent :

› Discipline, Structure, Méthodologie…

… pour renforcer les défenses face aux menaces croissantes

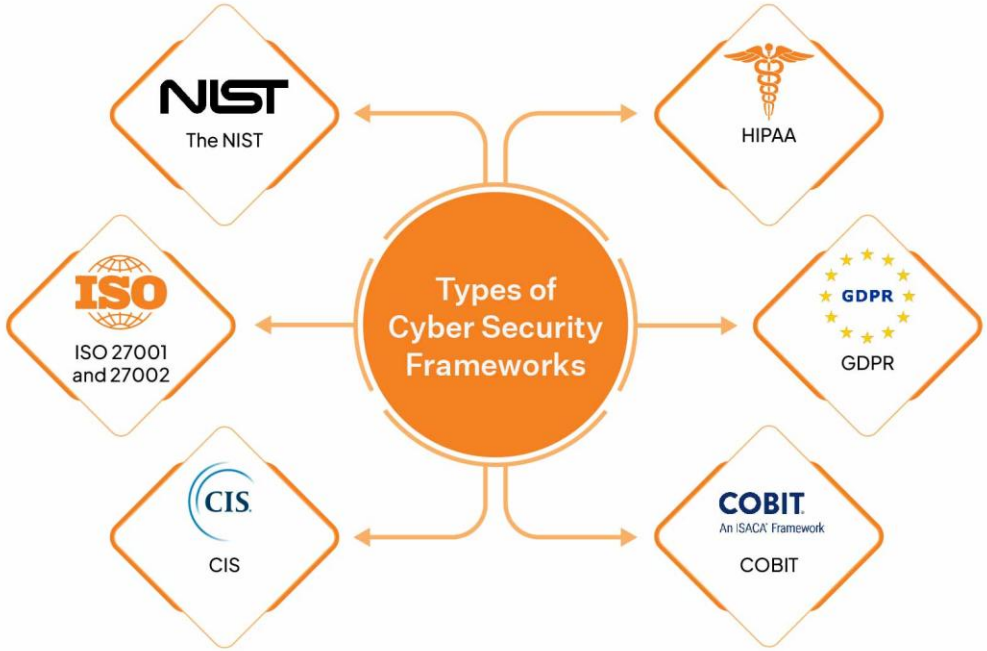
Tout comme la cyber sécurité, l'IA est un sujet transversal et…

› …a donc pour objectif l'amélioration des concepts IT existants

Notre sélection → NISTv2

› Accessible, Simple, Reconnu et…

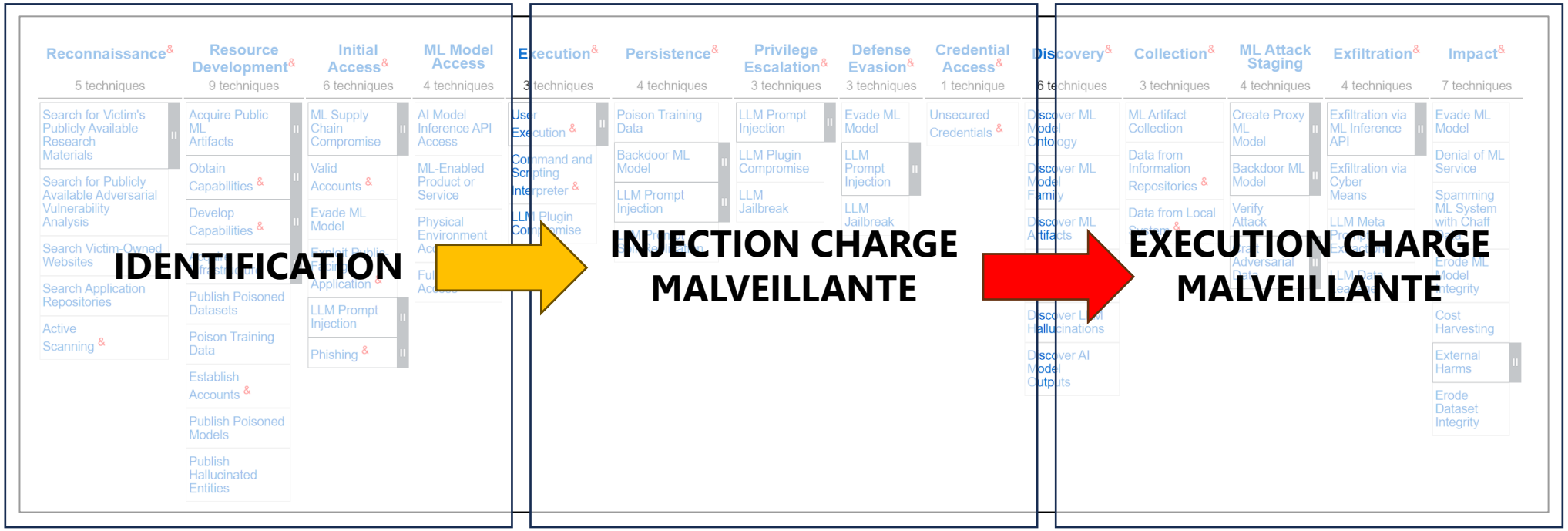
… aligné avec le standard MITRE, plus spécifiquement ATLAS !



-- Source: The Top 10 Cybersecurity Frameworks --

La boîte à outils des cyber consultants 2/3

Le standard MITRE ATLAS décrit le mode opératoire des attaques IA !



-- Source: MITRE --

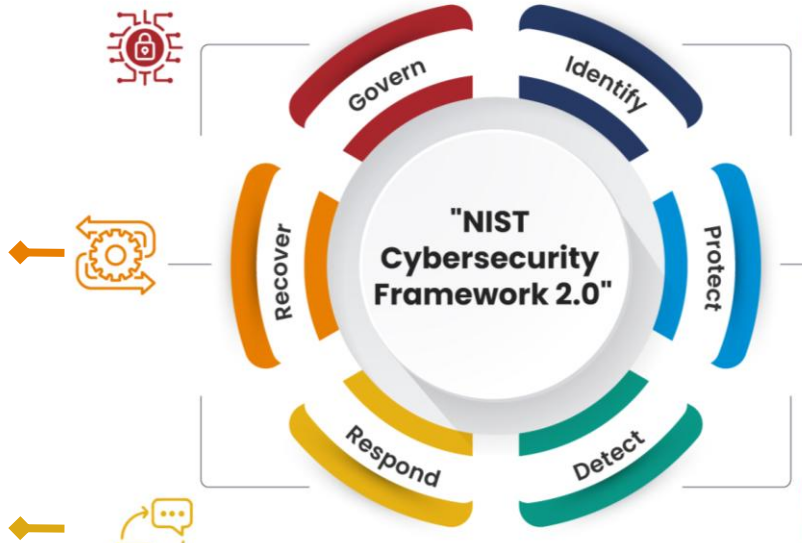
La boîte à outils des cyber consultants 3/3

Quant à NISTv2, il fournit des directives ainsi que de bonnes pratiques en matière de cybersécurité, visant à renforcer la résilience des organisations face aux menaces numériques.

Superviser le développement, le déploiement et l'évaluation des contrôles de cybersécurité

Mettre en place des processus pour rétablir les services et les systèmes après un incident.

Élaborer des plans pour répondre aux incidents et atténuer les impacts sur les opérations



-- Illustration : Infosectrain --

Comprendre et évaluer les risques en identifiant les actifs, les personnes, les données et les menaces

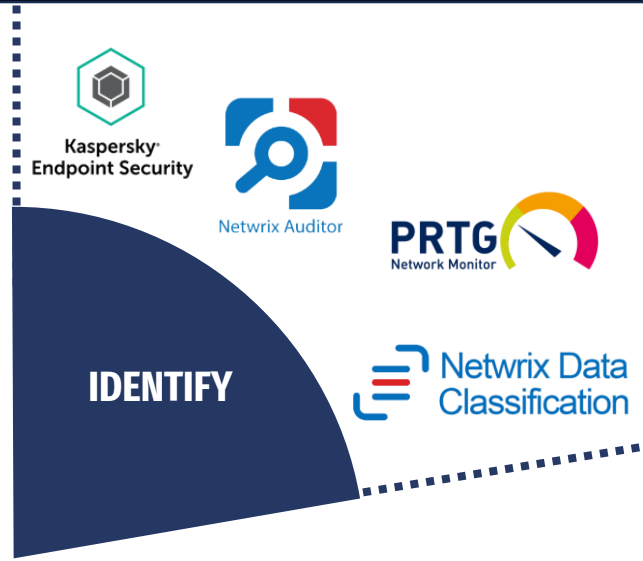
Mettre en œuvre des mesures de sécurité pour protéger les systèmes et les données

Développer des capacités pour détecter les incidents de cybersécurité en temps réel

L’IA au service de la cybersécurité

IDENTIFIER

Comprendre et évaluer les risques en identifiant les actifs, les personnes, les données et les menaces afin de mieux cibler les actions de sécurité



OPTIMISATION GRÂCE À L’IA

Identification des risques

› Identifier les **vulnérabilités** de l'écosystème IT (e.g. Zero-Day)

› Analyser et identifier les **flux applicatifs** (e.g. Model Zero Trust)

› Analyser et identifier les anomalies **déviant de la normalité**

Gestion automatisée

› Inventaire des actifs numériques

› Inventaire des comptes et des privilèges

→ **Plus précis et plus rapide**, facilitant ainsi une **priorisation optimale** des actions de sécurité

→ **Entrainée grâce à >30ans d’historique IT**

L’IA au service de la cybersécurité

PROTÉGER

Mise en œuvre de mesures de sécurité pour protéger les systèmes, les données et les infrastructures contre les menaces répertoriées, comme les non référencées

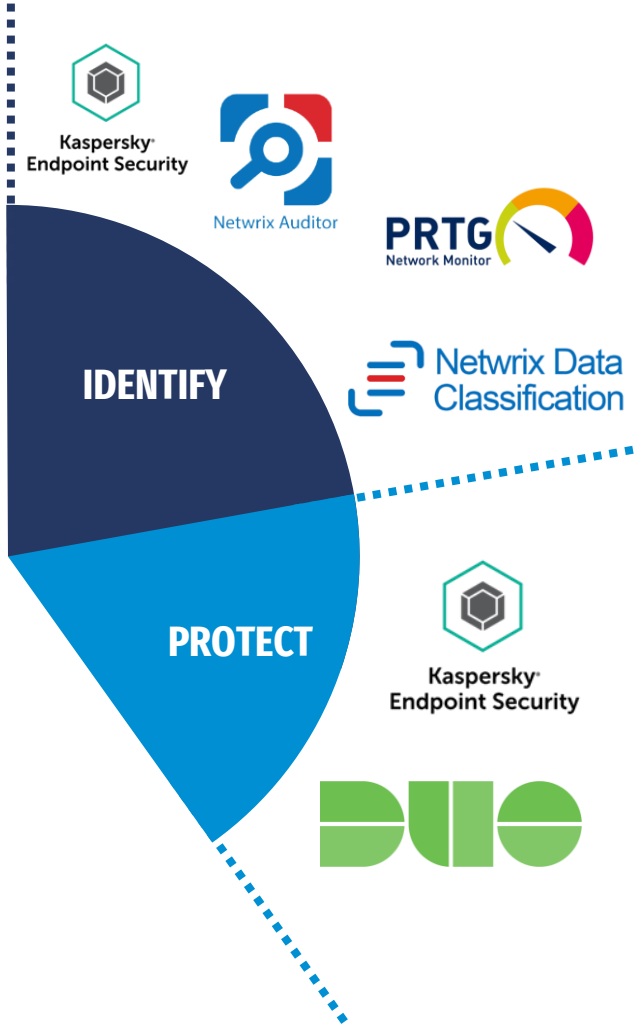
OPTIMISATION GRÂCE À L’IA

Protection des systèmes au travers

- **Adaptation** mesures sécurité en **temps réel**
- Aide à la configuration des **systèmes complexes**
- Analyse **prédictive** des menaces

Gestion automatisée

- Contrôle d'accès basé sur **l'intelligence comportementale**
- Renforcer les **flux de gestion** des identités et des accès
- **Provisionnements** et **révocations** automatisées



L’IA au service de la cybersécurité

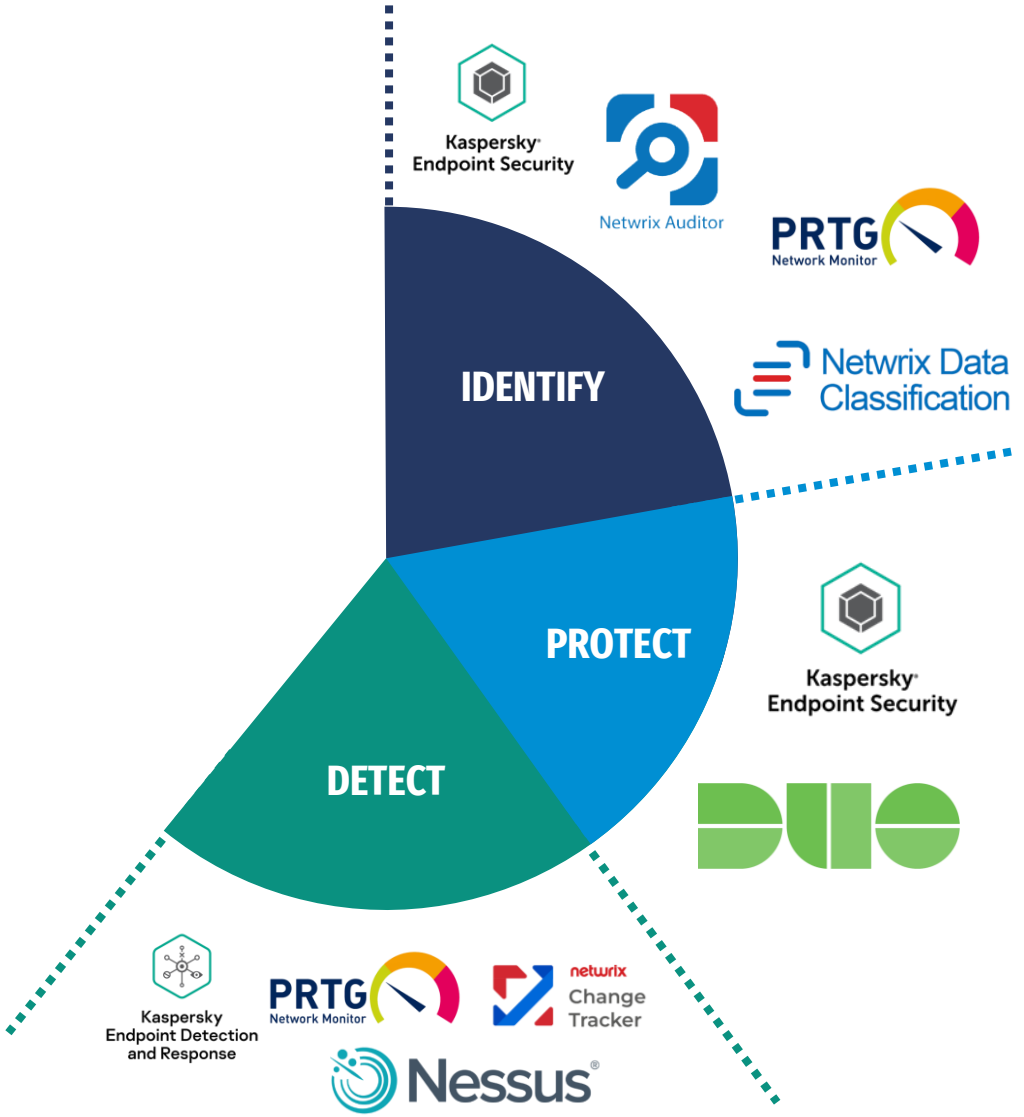
DÉTECTER

Développer des capacités pour détecter les incidents de cybersécurité en temps réel

OPTIMISATION GRÂCE À L’IA

Amélioration **réactivité** grâce à

- Apprentissage auto des **nouvelles menaces**
- Détection **proactive** en **temps réel**
  - Anomalies **inconnues** vs **connues**
  - Anomalies **déviant** de la **normalité**
  - **Exploitations** de vulnérabilité



L’IA au service de la cybersécurité

RÉPONDRE

Réagir efficacement aux incidents en mettant en œuvre des actions correctives pour atténuer leur impact

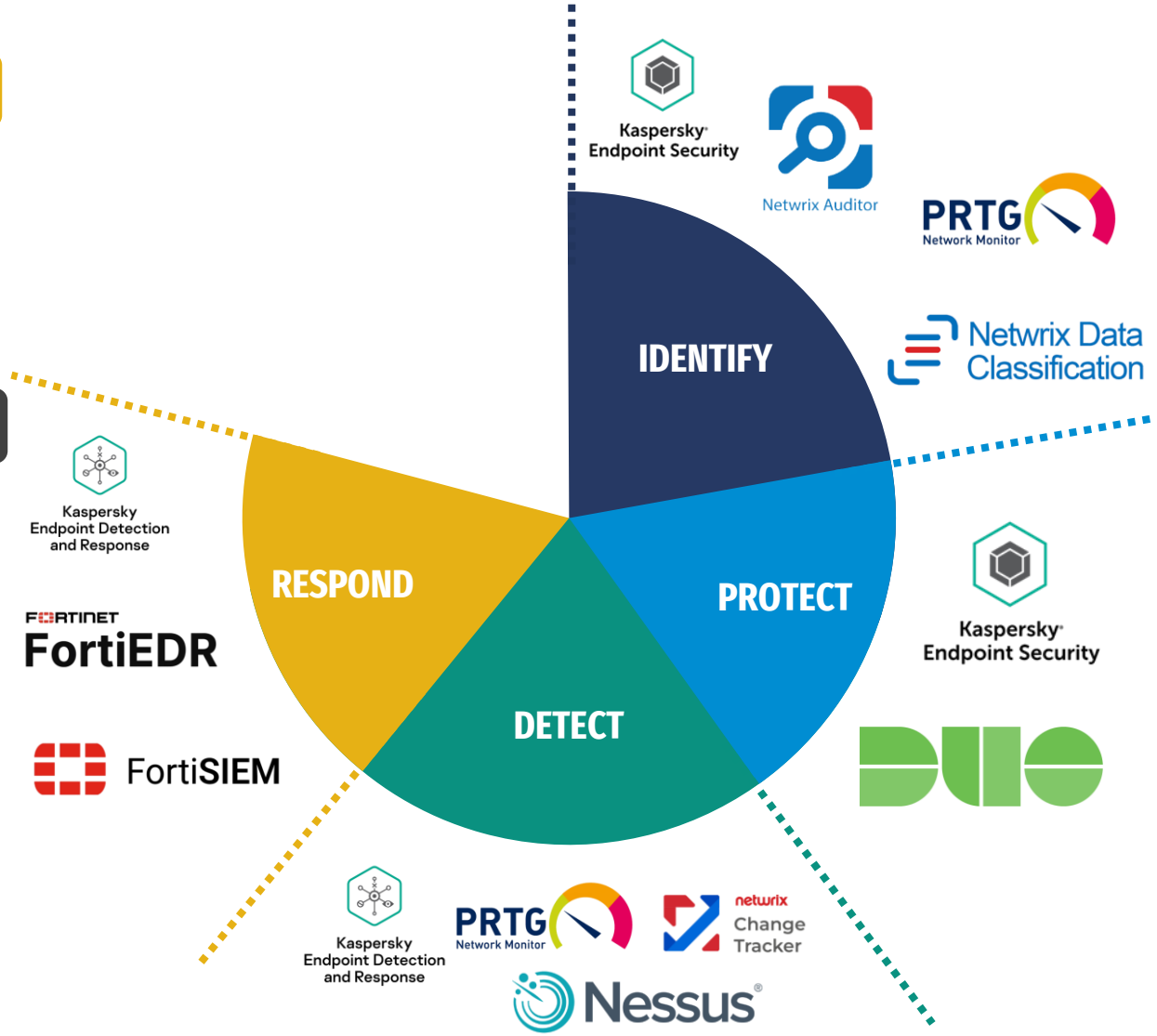
OPTIMISATION GRÂCE À L’IA

Automatiser les processus de réponse aux incidents

› Corrélation des événements et notifications

› Confinement de la menace
  – Isolation des systèmes compromis
  – Mise en œuvre de règles de FW supplémentaires

› Analyse et rapport post-incident
  – Tirer des leçons et adapter les stratégies de prévention futures, identifiant ainsi les points faibles du système



L’IA au service de la cybersécurité

RÉCUPÉRER

Mettre en place des processus pour rétablir les services et les systèmes après un incident

OPTIMISATION GRÂCE À L’IA

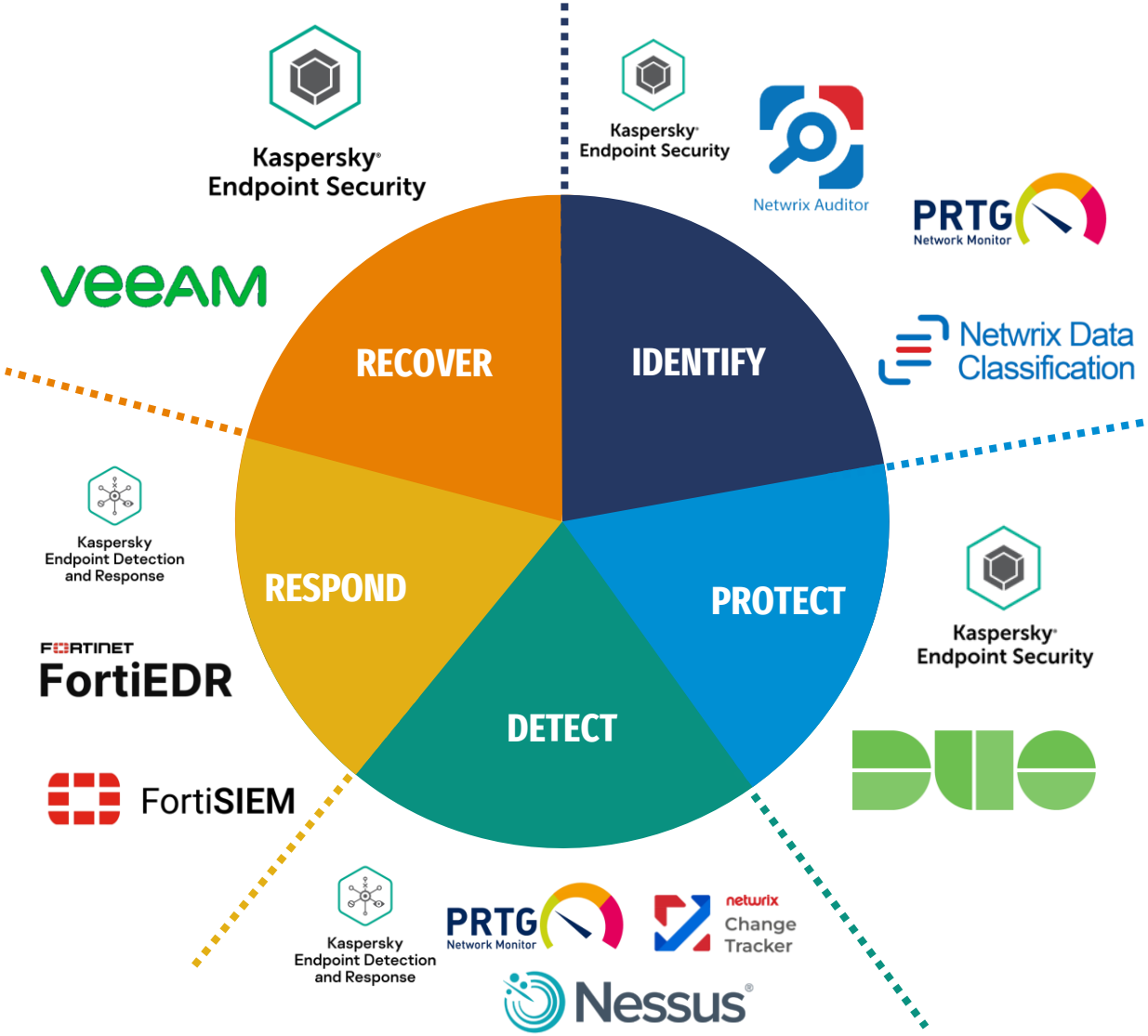
Récupération **accélérée** grâce à

› **L’automatisation** des **processus** de restauration

› **L’optimisation** des plans de reprise pour **minimiser les interruptions**

→ **Sécurisation** des sauvegardes

→ **Simulation** des scénarios de récupération



L’IA au service de la cybersécurité

GOUVERNANCE

Mise en œuvre de politiques et procédures afin de superviser le développement, déploiement et évaluation des contrôles de cybersécurité

OPTIMISATION GRÂCE À L’IA

Pilotage et suivi de la conformité cyber

› La surveillance automatisée des contrôles

› Gestion des politiques de sécurité

→ Aide à la personnalisation/construction des formations à la sensibilisation à la cybersécurité

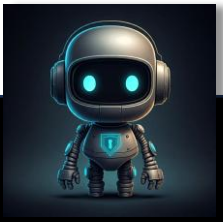
oxygen
data pilots

La cybersécurité au service de l'IA

Tobias Kull

Directeur BU Cyber chez Oxygen Data Pilots et chargé de cours à l'HEIA

Sécurité « by default » et « by design »



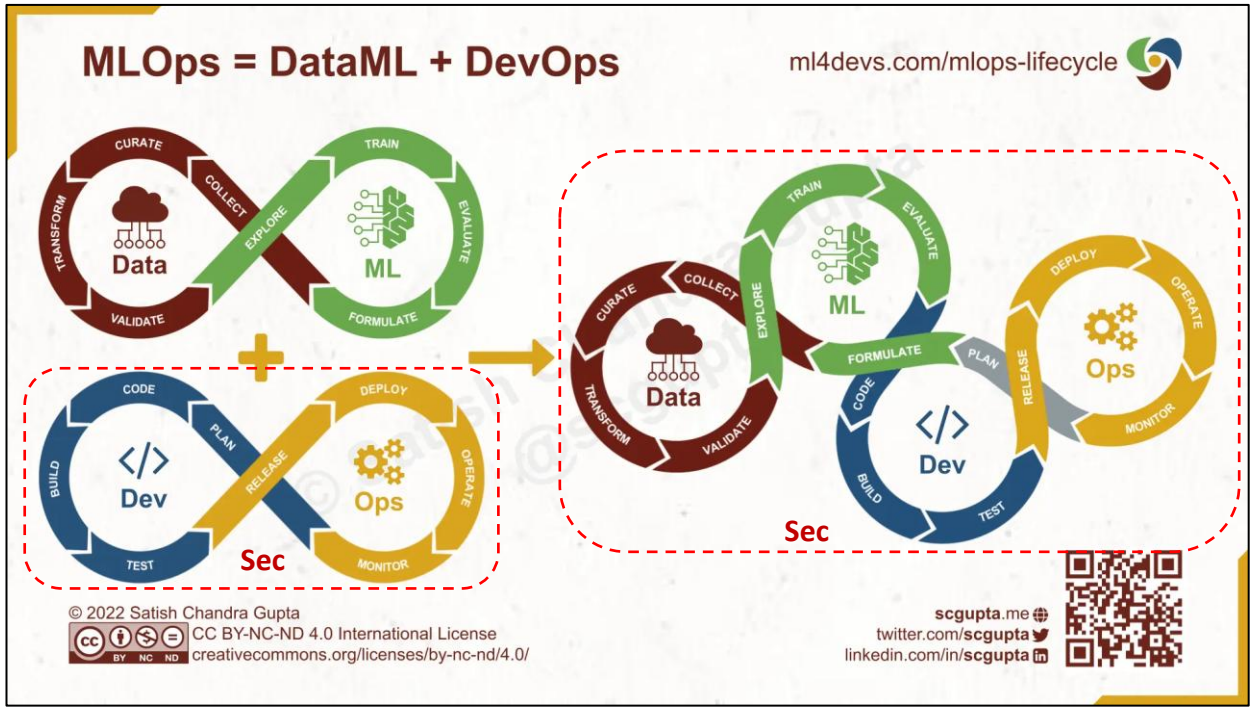
L'intégration des concepts de cybersécurité (DevSecOps), dès la conception des services ou applications, représente la pierre angulaire de tout système qui se veut garant de vos données !

Les méthodes MLOps assurent :

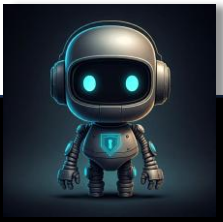
- Sécurité / Conformité
- Automatisation
- Monitorage / Gestion des versions (inc. données)
- Garantie de reproductibilité

Les aspects les plus importants

- Gouvernance
- Sécurité des données
- Fiabilité des modèles

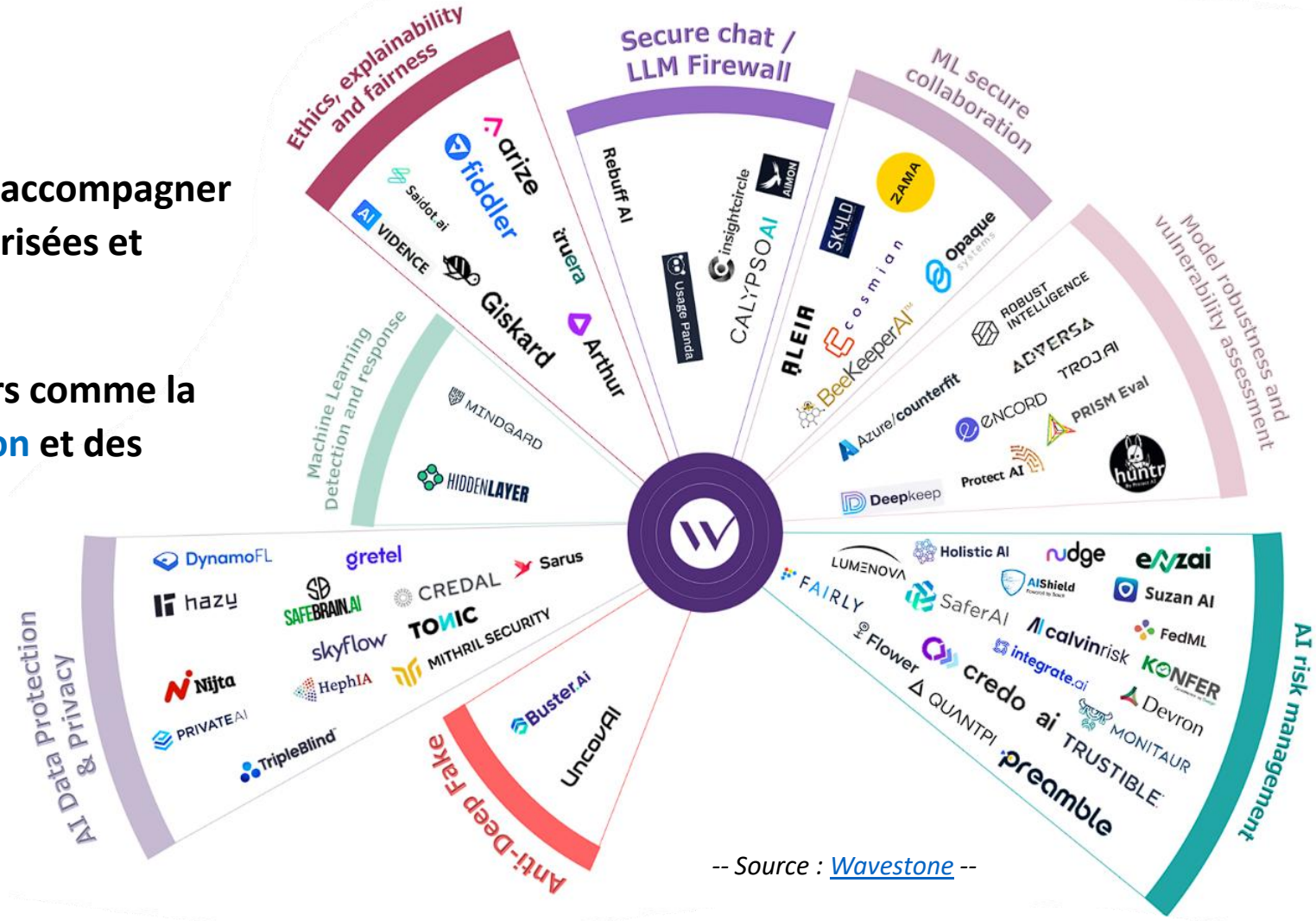


L’outillage !



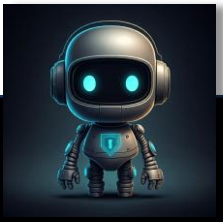
Un grand nombre d’acteurs existent déjà pour accompagner les développements vers des plateformes sécurisées et respectueuses de la vie privée !

…et d’autres sont actifs dans différents secteurs comme la lutte contre la propagation de la désinformation et des contenus frauduleux sur le web



-- Source : [Wavestone](#) --

Gouvernance, un mal nécessaire ?



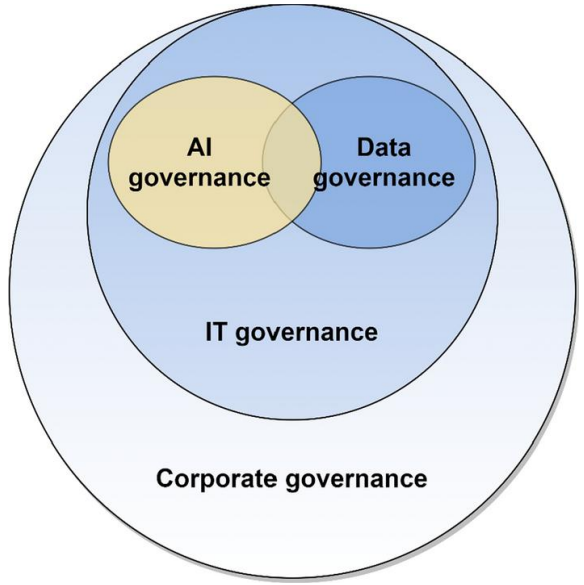
Se soumettre à une gouvernance ou suivre des règles n'est pas toujours plaisant, mais le but est d'apporter un cadre solide, de fixer des objectifs précis et de répartir les responsabilités selon les différents domaines ou axes de travail.

Gouvernance des données

- Europe : RGPD
- Suisse : nLPD
- USA : California Consumer Privacy Act, mais diffère entre les états !

Gouvernance IA

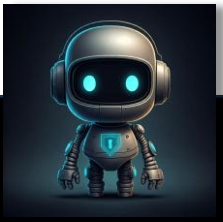
- Europe : AI Act & la Convention-cadre sur l'IA du Conseil de l'Europe
- Suisse : Guidelines du SEFRI & Le DFJP, en collaboration avec le DETEC et le DFAE, élaborera d'ici fin 2026 un projet de consultation qui met en œuvre la Convention sur l'IA du Conseil de l'Europe
- USA : des mesures prises sur les menaces IA, mais pas encore d'accord de lois concrètes en Californie (rejeté le 29 septembre 2024)

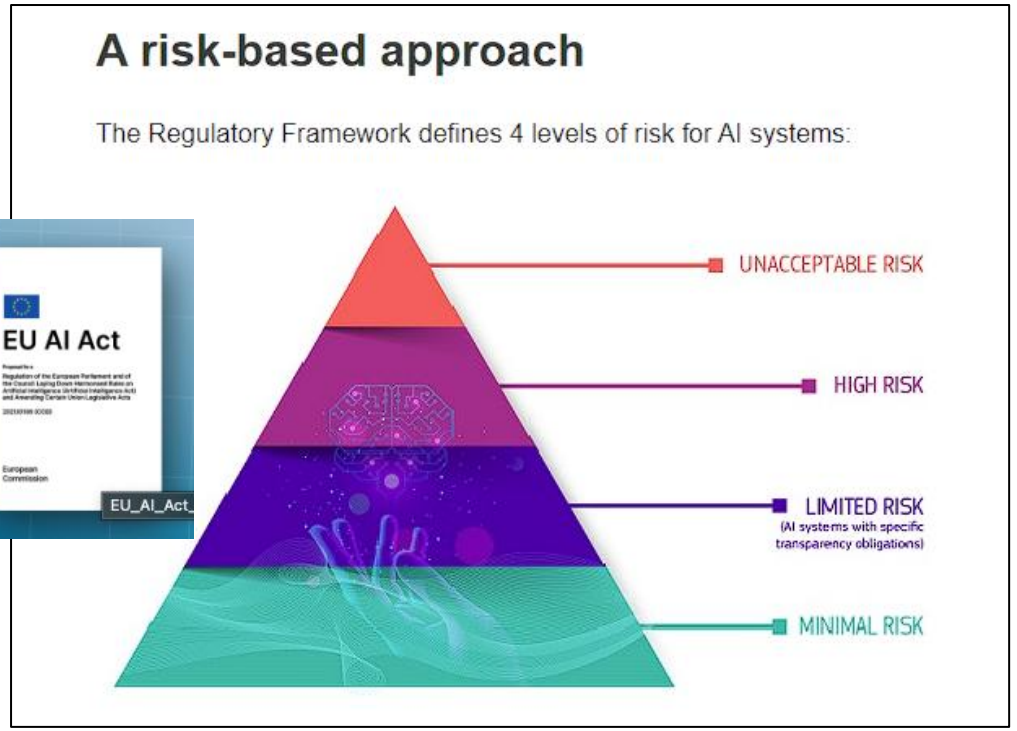


-- Source : ICT --

Rappel nLDP → la personne physique reste responsable et non l'entreprise (personne morale)

EU Artificial Intelligence Act

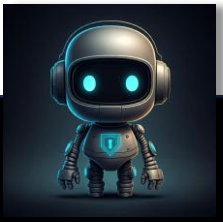




-- Source : [Commission européenne] --

Niveau de risque du système IA	Contraintes	Exemples
Inacceptable	**Interdit** - car menace sécurité ou les droits fondamentaux	• Attribution d'un score social aux citoyens basé sur leur comportement. • Manipulation et tromperie ou exploitation de vulnérabilités fondées sur l'IA • Reconnaissance des émotions sur les lieux de travail et dans les établissements d'enseignement
Élevé	Obligations **strictes** comme l'**évaluation des risques**, la **supervision humaine**, la **qualité des données** et la **documentation complète**. Nécessite **évaluation avant mise sur le marché**.	• Logiciel autonome dans les infrastructures de transport • Chirurgie assistée par un robot
Limité	**Transparence** obligatoire	• Chatbot d'assistance où l'utilisateur doit être informé qu'il communique avec une machine
Minimal	**Libre utilisation** sans restriction spécifique.	• Filtres anti-spam • Jeux vidéo

Aspect éthique « AI Ethics »



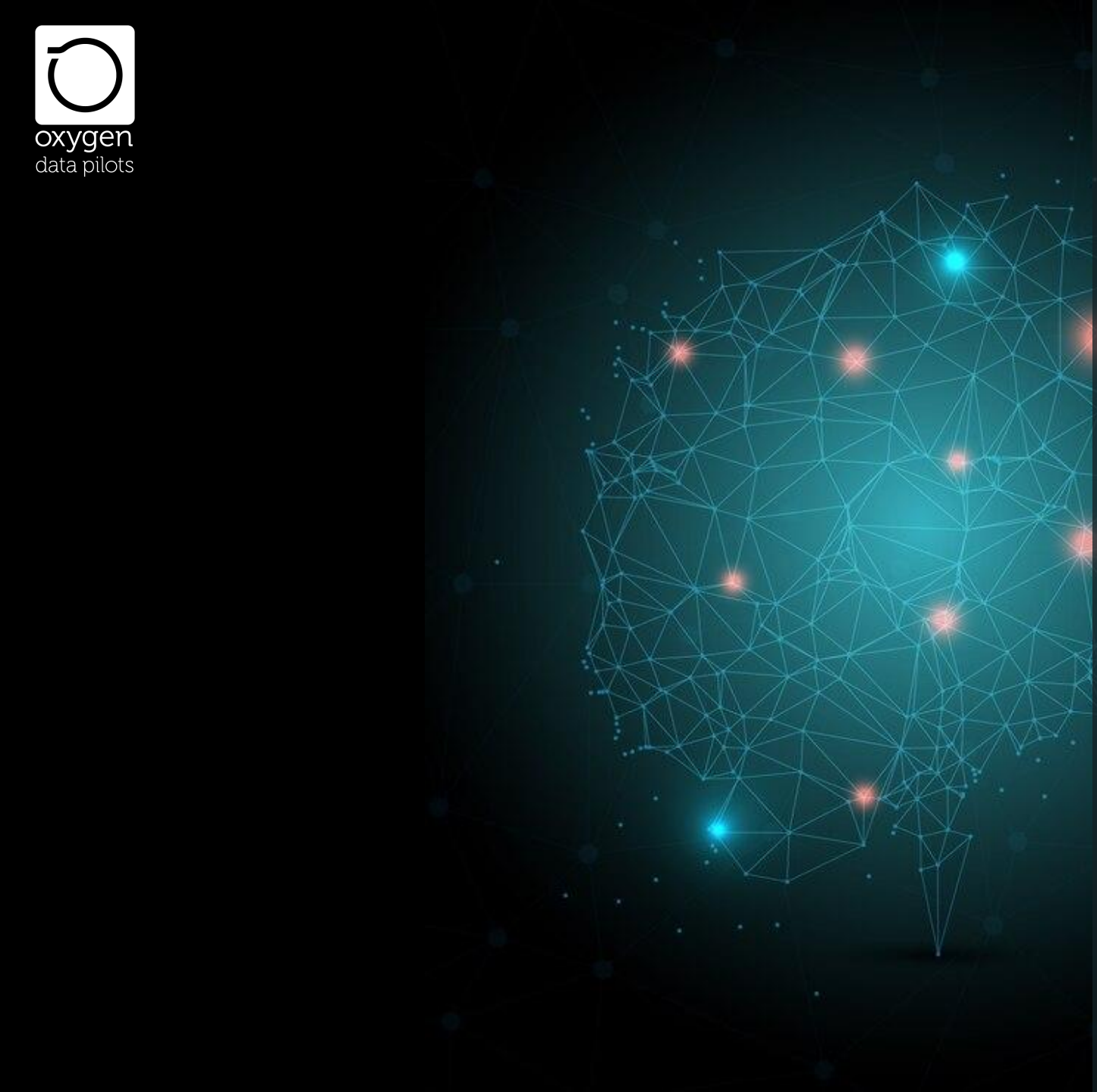
*« Les systèmes d'IA soulèvent des **enjeux éthiques majeurs**, car ils prennent des **décisions autonomes** avec un **impact direct sur les individus**, ils risquent de **reproduire des biais**, ils **manquent** de **transparence** et peuvent **influencer des comportements.***

*Ceci les **distingue des systèmes informatiques conventionnels** et ils nécessitent donc une **vigilance accrue**. »*

Défis éthiques de l'IA

- **Hallucinations** → Quand dire « Je ne sais pas » ? Comment déceler vrai du faux ?
- **Biais** → Représentation dans données inégales, préjugés, genrés (e.g. doctoresses…)
- **Privacité** → Consentement, données d'entraînements, traçabilité,…
- **Transparence** → Détails raisonnement, reproductibilité, XAI,...
- **Influence des comportements** → Effet de bulle, manipulation, polarisation d'opinions,…
- **Responsabilité** → Qui paie pour les erreurs du système ?
- **Environnement** → Modèles énergivore, impact climatique, durabilité,…

oxygen
data pilots

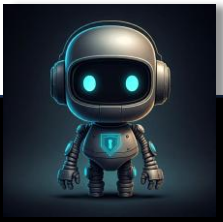


Conclusion

Tobias Kull

Directeur BU Cyber chez Oxygen Data Pilots et chargé de cours à l'HEIA

Bonnes pratiques pour l'IA générative



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

18 janvier 2024 V1.2[1]

Fiche technique sur l'utilisation d'outils d'IA générative au sein de l'administration fédérale

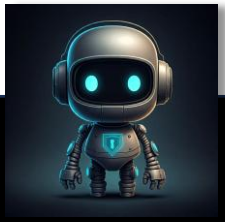
Référence: 822.1-1/8/5/1

Exemples d'utilisation d'outils d'IA générative

Vous pouvez recourir à un outil d'IA générative pour les tâches suivantes 👍	INTERDICTION de recourir un outil d'IA générative pour les tâches suivantes 🚫
Résumer des textes publiés (rapports, etc.)	Résumer les documents de la procédure de co-rapport (→ non publics; les documents sont peut-être classés)
Se familiariser avec un sujet (analogue à Google ou Wikipédia)	Traduire des CV (→ données personnelles)
Formuler du texte pour une présentation Power-Point	Saisir telle quelle la demande concrète reçue de Monsieur Pierre Exemple (→ données personnelles)
Trouver de l'inspiration pour générer des code	Utiliser des réponses avec *copier / coller* (→ contrôle)
Créer des images pour une présentation	Saisir un code logiciel existant afin de le déboguer (→ violation du droit d'auteur)

-- Source : [Fiche technique sur l'utilisation d'outils d'IA générative au sein de l'administration fédérale] --

Equilibre liberté vs régulations?



Liberté des modèles IA Open-Source

→ Liberté d'innovation dans les systèmes qui ne tiennent pas compte de l'aspect éthique au détriment de l'efficacité

Permettre uniquement les modèles IA propriétaires

→ Application stricte de règles et contrôles pour la commercialisation / utilisation de chaque modèle IA

Question subsidiaire → **quid de la sobriété énergétique ?**

Merci pour votre attention !

SEMINAR FRIBOURG

Artificial intelligence