FICHE PRATIQUE

CYBERSECURITE DE L'INTELLIGENCE ARTIFICIELLE : LEVIER STRATEGIQUE POUR LES SOCIETES DE GESTION

L'IA, un risque à maîtriser face à un impératif stratégique

L'Intelligence Artificielle (IA) offre des opportunités de croissance et d'optimisation majeures pour votre société de gestion, notamment par l'automatisation, l'analyse avancée et l'aide à la décision tout en introduisant des risques significatifs.

Au-delà des problématiques posées en termes de gouvernance et d'éthique (opacité des décisions, absence de supervision), l'IA amplifie les risques existants et en introduit de nouveaux, liés notamment au caractère extrêmement accessible généré par l'utilisation du langage naturel.

Notre approche vise à transformer ces défis de sécurité pour les SGP en **avantages compétitifs** en structurant votre sécurité autour de 5 piliers fondamentaux :

1. **Maîtrise des vulnérabilités spécifiques à l'IA**
2. **Intégration du facteur humain**
3. **Stratégie de gestion des risques cyber liés à l'IA : Définir votre posture**
4. **Gouvernance des risques & Shadow AI**
5. **Protection de la Donnée : Sécuriser l'actif critique**

1. Maîtrise des vulnérabilités spécifiques à l'IA

Les systèmes d'IA ouvrant à des vulnérabilités, il est essentiel de maîtriser les vecteurs d'attaque spécifiques liés à. Pour ce faire, **comprendre les techniques d'exploitation est la première étape pour bâtir une défense robuste**.

ALTERATION / INJECTION DE REQUÊTE

Prompt Injection

MECANISME :

- L'attaquant insère des instructions cachées (ex : texte invisible dans un document Word) pour manipuler la réponse de l'IA

RISQUE POUR LA SGP :

- Fuite d'informations sensibles ou exécution d'actions contraires aux règles de sécurité

EMPOISONNEMENT DES DONNEES

Data poisoning

MECANISME :

- Introduction de données malveillantes dans les jeux d'entraînement. (Ex : Faux emails "propres" pour tromper un modèle de détection de spam)

RISQUE POUR LA SGP :

- Perte de fiabilité et d'intégrité des modèles IA, qui apprennent de mauvais comportements

CONTOURNEMENT DE FILTRES

Jailbreaking

MECANISME :

- Utilisation de ruse linguistique et de commandes détournées pour outrepasser les garde-fous des modèles

RISQUE POUR LA SGP :

- Obtention d'informations confidentielles ou génération de contenu interdit

2. 👥 Intégration du facteur humain

Sensibilisation et Charte IA

L'utilisation de l'IA au sein de la SGP rend indispensable une **vigilance collective**.
Qui plus est la formation n'est pas une option, c'est une **obligation réglementaire**.

Former et sensibiliser

Chaque utilisateur doit comprendre le fonctionnement, les limites et les risques associés aux outils d'IA. Pour une SGP, accompagner ses salariés en diffusant de bonnes pratiques constitue un pilier de sécurité. Parmi les axes de sensibilisation à prévoir, la vérification des sources, le réflexe de prudence à intégrer dans le partage d'informations sensibles, le signalement de toute anomalie.

Structurer l'usage de l'IA dans une charte

Afin d'offrir un cadre sécurisant à la SGP, une charte, qu'elle soit entièrement dédiée à l'IA ou s'inscrive dans une extension d'une autre charte (informatique…) est indispensable **encadrer l'usage de l'IA** et définir ses règles d'utilisation en intégrant :
Un double engagement de **cybersécurité et de confidentialité des données** (outre l'éthique et la conformité).
Les modalités de contrôle et de sanctions proportionnées en cas de non-respect.
Ce cadre servira de socle de confiance à l'ensemble des collaborateurs.

3. ⚖️ Stratégie de gestion des risques cyber liés à l'IA : Définir votre posture

Adopter une stratégie - A votre SGP de choisir la posture qui lui convient le mieux dans l'équilibre sécurité et innovation, entre les deux options qui s'offrent à elle :

Option 1 : Stratégie Restrictive

(Privilégier la sécurité maximale - *Whitelist)*

- **Protection forte :** Blocage systématique pour prévenir la fuite d'informations confidentielles.
- **Contrôle total :** Fermeture des grandes plateformes d'IA publiques non contrôlées (ChatGPT, Gemini, etc.).
- **Gouvernance stricte :** Mise en place d'une **Whitelist** d'outils IA autorisés, validés selon des critères stricts de sécurité, confidentialité et conformité (RGPD, normes sectorielles).
- **Solution Interne Sécurisée :** Développement d'IA internes sécurisées (Secure GPT, solutions *on-premise* ou *cloud* privé) pour garantir une maîtrise totale des données.

Avantage : Sécurité maximale, conformité garantie.
Contrainte : Risque de limiter l'innovation et nécessiter des investissements importants.

Option 2 : Stratégie Ouverte

(Favoriser l'agilité et l'innovation - *Blacklist)*

- **Innovation facilitée :** Pas de blocage systématique, favorisant l'accessibilité et l'innovation.
- **Culture de sécurité nécessaire :** Une ouverture large doit s'accompagner d'une sensibilisation renforcée sur les bonnes pratiques et la responsabilité individuelle.
- **Contrôle réactif :** des audits et contrôles réguliers sont nécessaires pour détecter les comportements à risque (partage de données sensibles, non-respect des politiques).
- **Définition claire :** Mise en place d'une charte d'utilisation définissant les cas d'usage autorisés et interdits.
- **Filtrage ciblé :** Blocage des sites d'IA publics potentiellement malveillants ou non conformes.

Avantage : Agilité et productivité accrues.
Contrainte : Exige une culture de sécurité forte et une gouvernance rigoureuse.

Points clés à vérifier pour la stratégie d'ouverture

- Identification claire et documentée des outils IA autorisés (Whitelist) ou interdits (Blacklist), avec justification des choix pour assurer la traçabilité des décisions.
- Critères de sélection rigoureux basés sur la sécurité (chiffrement, authentification, hébergement des données), la conformité réglementaire (RGPD, certifications ISO, SOC2) et la fiabilité des outils (réputation du fournisseur, historique de sécurité, SLA).
- Processus de mise à jour régulière des listes (au minimum trimestriel) en fonction des évolutions technologiques, des nouvelles vulnérabilités identifiées, des changements réglementaires et des retours d'expérience terrain.
- Communication claire, transparente et accessible aux utilisateurs sur les règles d'utilisation via l'intranet, des sessions d'information et des rappels réguliers par mail.
- Surveillance continue des usages avec collecte de feedback utilisateurs, analyse des incidents de sécurité et ajustements agiles des stratégies en fonction des retours, des nouvelles menaces et de l'évolution des besoins métiers.
- Gouvernance collaborative impliquant les équipes IT, sécurité, juridique et métiers pour garantir l'alignement entre contraintes techniques et besoins opérationnels.

Adopter un déploiement adapté à ses risques

- **Déploiement en interne uniquement pour les cas sensibles :** Restriction des IA publiques pour les départements traitant des données critiques (finance, juridique, R&D) et mise à disposition d'alternatives internes sécurisées.
- **Ouverture progressive avec filtrage des usages :** Accès limité à certains services IA selon les profils et besoins, avec extension graduelle après évaluation des risques.
- **Ouverture large avec formation continue :** Programmes de formation obligatoires incluant protection des données, identification des informations sensibles et conséquences des violations.

Surveiller et contrôler l'usage d'IA publiques

Quelle que soit la stratégie choisie par la SGP, il est utile de prévoir la surveillance et le contrôle de l'usage des IA publiques, avec les points d'attention suivants :

- **Surveillance Proactive :** Utilisation de solutions de Proxy, DLP *(Data Loss Prevention)* et monitoring réseau pour identifier les flux et connexions suspects vers les plateformes IA publiques.
- **Analyse d'Usage :** Analyse approfondie des logs pour détecter les usages non conformes, tentatives d'exfiltration ou comportements anormaux (volumes de requêtes inhabituels, accès non conventionnels).
- **Réactivité :** Blocage/restriction immédiate en cas d'usage non autorisé, avec notification aux équipes de sécurité et déclenchement d'une procédure d'investigation.
- **Indicateurs de Performance (KPI)** : taux de détection, temps de réponse, conformité aux politiques.

4. 🔗 Gouvernance des risques & Shadow AI

Le **Shadow AI**, qui correspond à l'utilisation de l'IA par des collaborateurs de la SGP de façon non autorisée, peut être le reflet d'un besoin interne non couvert.
L'objectif pour la SGP en contrôlant ces usages sera d'orienter et de **canaliser ces usages de manière vertueuse afin de limiter les risques en termes de cybersécurité.**

Structurer l'organisation

Quelque soit sa taille, il est utile que la SGP mette en place une **gouvernance collaborative** intégrant les personnes en charge de la cybersécurité, conformité, risques, données personnelles, IT, juridique, et Direction Générale, pour établir :

- Des définitions claires des rôles et responsabilités.
- Un suivi régulier des usages et des risques.
- Une intégration de la sécurité **dès la conception** des projets IA.

Cartographier les usages de l'IA

Dresser un inventaire pour répertorier les **usages de l'IA au sein de la SGP** est un prérequis pour une SGP qui n'a pas nécessairement une vision des utilisations de l'IA dans son entité ;
La cartographie réalisée doit permettre :

- Une analyse des risques associés à chaque usage.
- L'identification des cas d'usage pertinents et évaluation des risques spécifiques.
- La mise en place de mesures adaptées.

Cette cartographie peut conduire à constituer un registre documentant l'ensemble des cas d'usage de l'IA (autorisés ou mis à disposition).

Prendre des mesures préventives essentielles

- **Cohérence des prompts :** Mettre en place des filtres et garde-fous pour contrôler la cohérence des requêtes soumises (prompts).
- **Intégrité des données :** Valider et nettoyer les données d'entraînement pour limiter l'empoisonnement.
- **Détection d'anomalies :** Surveillance continue des interactions pour détecter les comportements inhabituels (requêtes répétitives, réponses incohérentes, tentatives de fuite).
- **Alerte utilisateur :** Former les utilisateurs à reconnaître et signaler rapidement un comportement suspect de l'IA (réponses incohérentes, demandes de données sensibles).

5. Protection de la donnée : sécuriser l'actif critique

La sécurité des données est la garantie de la confiance et de la conformité pour votre SGP.

Sécuriser les environnements et l'accès

- **Environnements contrôlés :** Privilégier un hébergement sécurisé, conforme aux normes, et séparer physiquement/logiquement les modèles et les données sensibles.
- **Classification des données :** S'assurer de la bonne classification en amont de leur utilisation dans le moteur IA.
- **Contrôle d'accès robuste :** Mettre en place la segmentation et le **principe du moindre privilège**. Ex : Les bases de connaissances sensibles ne doivent pas être accessibles par défaut à tous les collaborateurs.
- **Gérer les habilitations :** Imposer une gestion rigoureuse des habilitations (création, révision, révocation) et une authentification forte pour ceux qui pourraient altérer le moteur IA.

Protéger contre la fuite d'information (exfiltration)

- **Identifier les vecteurs de fuite :** Maîtriser les points de sortie des données (modèles, prompts, journaux et autres artefacts produits par l'IA).
- **Sécuriser l'infrastructure :** Sécuriser les données d'entraînement et d'inférence (chiffrement, stockage en environnements contrôlés).
- **Contrôle des prompts :** Contrôler les requêtes (prompts) pour éviter la divulgation d'informations sensibles par inadvertance.
- **Surveillance :** Surveiller et journaliser les accès, les requêtes et les usages pour détecter toute anomalie ou abus (exfiltration de données).

Maîtriser la relation fournisseur

- **Propriété contractuelle :** Encadrer contractuellement l'exploitation des données par le fournisseur de service pour protéger la propriété intellectuelle et éviter tout usage non autorisé.
- **Maîtrise des traces :** Restreindre l'utilisation potentielle des traces (logs) par le prestataire.
- **Protection des actifs :** Mettre l'accent sur la protection de vos données critiques (financières, stratégie d'investissement, documents M&A) et de vos algorithmes propriétaires.

L'AFG remercie Wilfried Lauber (Amundi Asset Management), René Amirkhanian (DNCA Investments), Clément Civeit (Moneta), Walif El Hitti (Comgest), Mohamed Ghayati (Tikehau), Frédéric Gleizer (BNPP AM), Alexandre Joachim (LBP AM), Stanislas Perney (BDL Gestion), Olivier Pallany (Sienna AM France), Olivier Tomatis (Groupama AM), Aurélie Assegon (Ostrum Asset Management), Jean-Baptiste Wuillamier (Natixis investment Managers) et William Techer (AXA IM) qui ont activement participé à l'écriture de cette fiche.