EGE Ecole de Guerre Economique

Intelligence artificielle et sûreté des entreprises : du mythe à la réalité

Cédric Jutteau, Frédéric Loisel, Souleymane Meite,
Emilie Seruca-Cau- Aymeric Suchet
Management de la Cybersécurité et Gouvernance des Systèmes d'Information (MaCYB)

● Préface

Dans un monde en constante évolution numérique, l'émergence des dispositifs d'intelligence artificielle transforme en profondeur les dynamiques de sécurité et de sûreté des systèmes critiques, en particulier pour les Opérateurs de Service Essentiel (OSE)[1]. En tant que Directeur des Systèmes d'Information (DSI) d'un OSE, j'ai la responsabilité de garantir la sécurité et la continuité des infrastructures au bon fonctionnement de notre organisme. Cette mission, déjà complexe par nature, se voit aujourd'hui confrontée à de nouveaux défis avec l'essor des technologies d'IA, qui modifient non seulement les menaces, mais aussi les moyens de défense.

L'intelligence artificielle offre des opportunités considérables pour améliorer la résilience et la réactivité des systèmes critiques. En automatisant la détection des anomalies, en renforçant les capacités prédictives et en optimisant la gestion des ressources, l'IA peut devenir une alliée précieuse pour la sûreté des entreprises. Cependant, elle introduit des vulnérabilités nouvelles, créant un paysage de menaces où les attaques sont de plus en plus sophistiquées et imprévisibles. Les systèmes d'IA, en se nourrissant de données massives, sont susceptibles d'être biaisés, manipulés ou détournés à des fins malveillantes, posant des défis inédits pour les responsables de la sécurité.

Le déploiement de l'intelligence artificielle impose une analyse rigoureuse des méthodes et des politiques établies au sein des entreprises. La mise en lumière des vulnérabilités potentielles auxquelles feraient face les infrastructures critiques, qu'elles soient dirigées contre des systèmes de contrôle industriel, des réseaux de communication, ou encore des plateformes logistiques essentielles, se révèle être un modèle vertueux d'amélioration continue au bénéfice de l'ensemble des strates opérationnelles d'une organisation.

Au-delà de l'aspect technologique, les implications éthiques et réglementaires de l'usage de l'IA dans les entreprises doivent être cartographiées.

Comment garantir une transparence et une clarté des décisions prises par des systèmes d'IA autonomes ? Comment assurer que les choix technologiques respectent les principes de sûreté, de protection des données et de résilience face aux cybermenaces ? Ces questions sont au cœur des préoccupations des décideurs dans un contexte où la donnée est devenue le troisième capital productif du monde moderne.

La sûreté des Opérateurs de Service Essentiel est un enjeu stratégique national[2], voire international. Elle requiert une vigilance constante et une adaptation continue aux nouvelles menaces. Dans ce contexte, l'intelligence artificielle ne doit pas être perçue uniquement comme une source de risques, mais également comme un levier pour renforcer la sécurité et la performance des infrastructures critiques. Cependant, cet atout se révélera de son plein potentiel uniquement si nous parvenons à disposer d'une pleine compréhension de ses impératifs techniques, humains, financiers et légaux.

Nicolas Winter - DSI d'un Opérateur de Service Essentielle

[1] « Améliorer les humains plutôt que de les augmenter », Le Monde Diplomatique, n°845, 71ème année, août 2024.

[2] https://cyber.gouv.fr/la-directive-nis-2

Table des matières

● Préface 2
Table des matières 3
Table des illustrations 5
Table des tableaux 6
INTRODUCTION 7
● PARTIE 1. INTELLIGENCE ARTIFICIELLE ET SURETE DES ENTREPRISES : QUELS BENEFICES POUR DEMAIN ? 11
○ **I.1. Les avantages résultant du déploiement de l'intelligence artificielle au sein des entreprises 11**
■ I.1.a. Analyse quantitative ou qualitative du risque ? 11
■ I.1.b. L'intelligence artificielle et le bon quantitatif 13
■ I.1.c. Vers une compréhension prédictive du risque sûreté 14
■ I.1.d. Une vision à 360° des risques pour les entreprises 16
■ I.1.e. Intelligence artificielle, hypervitesse et hybridation par nature 17
■ I.1.f. L'apport de l'IA sur les fonctions primaires d'un système de sûreté 18
○ **I.2. Les opportunités identifiées par le déploiement de l'intelligence artificielle au sein des entreprises 20**
■ I.2.a. Détection et Prévention des risques financiers 23
■ I.2.b. La cybersécurité : Réseaux, Données & Utilisateurs 25
■ I.2.c. La surveillance et la sécurité physique 27
■ I.2.d. La gestion des identités et de l'authentification 30
■ I.2.e. Quels Impacts transversaux au sein de l'entreprise ? 33
● PARTIE II. INTELLIGENCE ARTIFICIELLE ET SÛRETÉ DES ENTREPRISES « UNE ARCHITECTURE À LA COMPLEXITÉ BYZANTINE » 34
○ **II.1. Des limites intrinsèques au déploiement de l'intelligence artificielle au sein des entreprises 34**
■ II.1.a. La gouvernance des données à l'épreuve de l'inflation normative 34
■ II.1.b. Un régime de responsabilité dont les contours restent à éprouver 39
■ II.1.c. Vers une « Intelligence humaine augmentée » en matière de sûreté des entreprises ? 40
■ II.1.d. Des dispositifs d'IA persuasifs et hallucinatoires ? 41
○ **II.2. Les menaces résultant du déploiement de l'intelligence artificielle au sein des entreprises 42**
■ II.2.a. IA et cyberattaque : une menace intrinsèque ? 42
■ II.2.b. Excès de confiance, ingérence économique et souveraineté : le paradigme de l'IA 45
● PARTIE III. Recommandations pour l'intégration de l'intelligence artificielle dans une stratégie de sûreté à 360° 51
○ **III.1. Une organisation intégrée et collaborative 51**

○ **III.2. Les 36 recommandations opérationnelles 53**
■ III.2.a : La planification 53
■ III.2.b : Technologie et infrastructure 55
■ III.2.c : Les ressources humaines 56
■ III.2.d : La sécurité 56
■ III.2.e : PDCA 57
■ III.2.f. Conclusion des recommandations 59
● CONCLUSION 60
● RÉFÉRENCES BIBLIOGRAPHIQUES 61
■ RAPPORTS INSTITUTIONNELS 61
■ OUVRAGES BIBLIOGRAPHIQUES & REVUES 61
■ RECOMMANDATIONS 62
■ RÉGLEMENTATION 62
■ SITES INTERNET 63

Table des illustrations

Figure 1 : Infographie SIA 41
Figure 2 : Système d'IA 43
Figure 3 : Cycle de vie d'un système d'IA 43
Figure 4 : RIA - Approche par les risques 44
Figure 5 : Modèle d'analyse augmentée 48
Figure 6 : Représentation des scénarios d'attaque 52
Figure 7 : Cartographie applicative - Horizon 2028 55

Table des tableaux

Page 26 : Domaines d’application de l’IA au sein des entreprises
Page 27 : Domaines d’application de l’IA dans le domaine de la sûreté
Page 70 : Matrice de recommandations de déploiement de l’IA avec les parties prenantes

INTRODUCTION

"***Ce ne sont pas les murs qui font la cité, mais les Hommes***". Ces quelques mots emplis de la sagesse du philosophe Grec Platon datent du 3ème siècle avant notre ère. Il est difficile de ne pas noter le contraste entre l'ancienneté de ces propos et son omniprésence dans la littérature moderne en lien avec la thématique de la sûreté. Qu'elle soit assurée au niveau de l'Etat ou de celui d'une entreprise privée, la sûreté est depuis toujours considérée comme un besoin fondamental de l'Homme en société. Cette citation illustre également la place qui doit être celle accordée à l'être humain afin de faire face aux risques qui pèsent sur cet ensemble collectif à défendre qu'est la Cité.

Ces quelques mots de Platon ont également la particularité de mettre en lumière la dualité apparente susceptible d'exister entre d'un côté « *les murs* » et de l'autre « *les Hommes* » où, si nous filons aujourd'hui la métaphore, entre les aspects technique, organisationnel et humain considérés comme la pierre angulaire de l'analyse d'un dispositif de sécurité, de sûreté ou de prévention[3] moderne.

S'interroger sur la sûreté suppose tout d'abord de clairement distinguer cette notion de celle de sécurité. Souvent utilisés de manière indifférenciée, il n'en demeure pas moins que chacun de ces termes est strictement défini. Afin de se prémunir de l'écueil qui consisterait à mal penser les choses dès lors qu'elles seraient mal nommées[4], retenons que :

- **La sûreté** peut être définie comme « *l'état de celui qui n'a rien à craindre pour sa fortune ou sa personne*[5] ». Il s'agit en pratique d'une notion large qui englobe l'ensemble des moyens, techniques, humains et organisationnels destinés à **prévenir ou à** réduire les risques de nature **exclusivement malveillante**, comme les vols, les détournements, les sabotages, les actes de terrorisme, l'espionnage, ou les atteintes à l'intégrité physique des individus. La sûreté vise également à assurer la protection des personnes, des biens, des informations et des infrastructures contre ces menaces intentionnelles.

- Contrairement à la **sécurité** (« *safety* » en anglais), qui s'attache principalement à **prévenir les risques accidentels** (comme les incendies, les accidents de travail, etc.), la sûreté se concentre sur les risques délibérés et prémédités.

La sûreté implique donc la mise en œuvre de **mesures préventives** (analyse des menaces, politiques de sensibilisation, dissuasion…) ainsi que **des mesures curatives** (la planification de la réponse aux incidents, détection, réponses…). Ces différentes mesures permettent d'agir soit sur l'occurrence d'apparition d'un événement indésirable (vraisemblance) soit sur sa gravité.

Historiquement, le triptyque « technique, organisation et humain » qui constitue l'essence même de la sûreté repose en grande partie sur des capacités d'analyse et des expertises subjectives. Force est de constater que ce tryptique est resté relativement stable, sans remise en cause majeure depuis ses premières théorisations dans les années 80. Lorsqu'en 1997, le logiciel édité par la société américaine IBM, Deep Blue bat le champion du monde d'échec Gary KASPAROV, un ordinateur surpasse pour la première fois un humain. Les premières craintes liées à l'utilisation d'une intelligence artificielle émergent au sein du grand public façonnant ainsi l'imaginaire collectif. Ce que l'on appelle intelligence

---

[3] Fondation pour une culture de sécurité industrielle, Méthode ATHOS <https://www.foncsi.org/fr>

[4] WITTGENSTEIN Ludwig, "Mal nommer les choses, nous conduit à mal les penser ». *Remarques sur le Rameau d'or de Frazer,1977, p. 43-54*

[5] ACADEMIE FRANCAISE, Dictionnaire, 8ème édition

artificielle n'est alors qu'une capacité de calcul statistique capable d'évaluer 200 millions de positions par seconde.
Dans le monde de la sûreté c'est au travers de l'analyse de risques que l'on retrouve le plus d'éléments de mathématisation. La conduite d'une analyse de risques apparaît toujours comme étant un préalable à toute politique de sûreté ou de sécurité déployée au sein d'un environnement déterminé. C'est en effet cette approche par les risques qui permet d'identifier et de déployer des moyens proportionnés de prévention ou de défense. De manière pragmatique, l'analyse de risques est une démarche qui repose sur deux méthodologies :

- **Une démarche quantitative** qui repose sur la mathématisation d'éléments, la captation de statistiques et le calcul de probabilités rationnelles
- **Une démarche qualitative** qui repose sur des sentiments et des impressions, laquelle tient compte des référentiels normatifs ou des processus internes dans le contexte auquel elle se rattache

  Si ces deux démarches se distinguent clairement dans leur approche, elles n'en demeurent pas moins complémentaires afin de pouvoir permettre de conduire une analyse de risque considérée comme pertinente. Dans ce contexte, **que dire de l'influence que pourrait avoir la puissance de calcul offerte par un outil d'intelligence artificiel moderne sur l'orientation de la politique de sûreté des entreprises ?**

Ce type d'outil serait capable d'agglomérer une quantité phénoménale de données, bien plus que ce qui ne serait permis par le cerveau humain et de pouvoir apporter des solutions fondées sur des calculs rationnels en un minimum de temps. **Nous pourrions alors observer un basculement de notre approche du risque avec un arc qualitatif qui se verrait bousculer par un flot de données quantitatif sans commune mesure.** Cette approche permettrait alors de faire émerger une vision à 360° de la sûreté qui ne se limiterait plus à la séparation traditionnelle entre le monde physique et le monde virtuel, illustré souvent par des services hermétiques et indépendants de DSI et de sûreté mais qui formerait un véritable pont entre ces deux mondes afin de faire émerger une vision : "la sûreté globale" de l'entreprise.

Si nous avons vu à quel point Platon considérait l'Homme comme important dans la constitution et la protection de la Cité par rapport aux murs qui la soutiennent, les avancées scientifiques dans le domaine de l'intelligence artificielle viennent directement reposer cette question en y apportant la notion "d'intelligence" des "murs" de la cité. Ces mêmes "murs" qui constituaient des cloisons dans le passé seraient ici l'illustration de protections physiques tant que de protections numériques, le tout participant à cette sûreté globale de l'entreprise constitutive d'un tout indissociable.

Les dispositifs de sûreté incluant des composants d'intelligence artificielle ont le vent en poupe, à tel point qu'il en est fait un argument commercial et incontournable pour les entreprises spécialisées dans le domaine des technologies de surveillance et de protection (caméras dites intelligentes, murs connectés ou robots rondiers autonomes…) mais aussi dans le monde des entreprises spécialisées en cybersécurité (solution d'analyse comportemental de sécurité des réseaux, capacités d'analyse poussées des SIEM et SOAR et pare feux intelligents…). A cela s'ajoutent les nombreux fantasmes alimentés par ces technologies quant à la manière dont ils pourraient être mis en œuvre.

Dans ce contexte, une question se pose : **quel pourrait être l'apport de ces dispositifs technologiques équipés d'IA pour assurer la sûreté globale de l'entreprise versus les risques qu'ils pourraient engendrer pour cette même entreprise ?** En effet, l'automatisation de la sûreté tend à effrayer les décideurs et les analystes, cette défiance étant le plus souvent alimentée par la crainte qu'*in fine*, la machine ne supplante l'humain…

Mais revenons à KASPAROV et cette fois, à sa théorie du centaure, laquelle nous offre une philosophie d'approche intéressante.

Dès 1998, un an après sa défaite face à deep blue Gary KASPAROV ne considère pas qu'il y ait d'un côté, une machine supérieure et de l'autre, un Humain désuet d'intérêt et aux capacités de calculs limitées pour le jeu, il voit là des profils complémentaires. Le champion lance alors une nouvelle forme de tournois : "Les échecs avancés, ouverture vers de nouveaux horizons[6]", l'idée étant de combiner l'intuition et l'expérience d'un joueur Humain avec la capacité de calcul d'une machine de manière à offrir le meilleur niveau d'échec possible et de faire s'affronter des binômes Humain/Machine lors de tournois d'échecs. Cette idée de complémentarité est illustrée par l'image mythologique du Centaure, un buste Humain étant porté par un corps de cheval permettant d'aller plus vite et plus loin.

La mise en œuvre de dispositifs d'intelligence artificielle au service de la sûreté nous amène d'emblée à poser les termes de ce que nous entendons par intelligence artificielle.
L'intelligence artificielle, se définit davantage comme une discipline de l'informatique qui vise à créer des systèmes capables d'effectuer diverses tâches incluant l'apprentissage (l'acquisition de connaissances et de règles pour utiliser ces connaissances), le raisonnement (l'utilisation de règles pour atteindre des conclusions approximatives ou définitives) et l'autocorrection. Si le mot "intelligence" peut apparaître comme étant un abus de langage, nous ne débattrons pas ici de son usage ni des multiples débats qui entourent la définition de cette notion[7].

Au-delà de la définition juridique de l'intelligence artificielle telle que proposée par l'OCDE[8], nous avons choisi de retenir la définition proposée par Margaret A.BODEN, ***"L'intelligence artificielle, qui consiste à faire faire aux ordinateurs ce que peut faire l'esprit humain***."[9] A nos yeux, il est important de saisir que l'intelligence artificielle se présente comme une véritable technologie de rupture, une "technologie d'usage général"[10] **pour qualifier toutes les technologies présentant une avancée significative dans le développement de l'humanité.**

S'il est possible de dater les débuts de l'IA aux années 1950[11], ce n'est qu'en 1995 que l'IA est étudiée sous quatre approches principales : **les systèmes qui pensent comme des humains, ceux qui agissent comme des humains, ceux qui pensent rationnellement, et ceux qui agissent rationnellement.**

De nos jours, l'IA peut être scindée en deux catégories principales : **l'IA faible**, également connue sous le nom d'IA étroite, qui est conçue et entraînée pour une tâche particulière (comme la reconnaissance vocale ou la conduite automobile) **et l'IA forte[11]**, qui possède la capacité de comprendre et d'apprendre de nouvelles tâches de manière autonome.
L'évolution de l'IA s'est considérablement accélérée avec l'avènement de l'apprentissage automatique et plus récemment, l'apprentissage profond (*deep learning*). Ces sous-domaines permettent aux

---

[6] CASE Nicky, "How To Become A Centaur"; *Journal of Design and Science*
[7] A titre d'exemple, les chercheurs Shane LEGG et Markus HUTTER, en 2007, ont proposé plus de 70 définitions de cette technologie, la principale difficulté selon eux résidant dans le fait "*qu'un problème fondamental en intelligence artificielle est que personne ne sait vraiment ce qu'est l'intelligence. Le problème est particulièrement aigu lorsque nous devons considérer des systèmes artificiels qui sont significativement différents des humains » Universal intelligence : A definition of machine intelligence; Shane LEGG & Markus HUTTER; 2007.*
[8] OCDE, *Recommandation du Conseil sur l'intelligence artificielle*, OECD/LEGAL/0499, Adoptée le 22 mai 2019 et amendée le 3 mai 2024
[9] BODEN Margaret, *L'intelligence artificielle,* 2016.
[10] BRESNAHAN Timothy F. TRADJENBERG M.; General purpose technologies "Engines of growth", Journal of Econometrics, January 1995, Pages 83-108, https://www.sciencedirect.com/science/article/pii/030440769401598T [11] Alain TURING et le test de turing.
[11] Principalement théorique à ce jour.

machines d'apprendre et de s'adapter à partir de grandes quantités de données, ouvrant ainsi la voie à des avancées rapides et à des applications pratiques dans de nombreux domaines.
De quelle manière dès lors pourrions-nous passer du mythe à la réalité et intégrer pleinement les dispositifs d'intelligence artificielle au service de la sûreté des entreprises sous réserve que cela soit effectivement souhaitable ?

Pour répondre à cette interrogation, nous utiliserons un modèle d'analyse qui a fait ses preuves, la matrice SWOT (strengths, weaknesses,opportunities, threat) afin de décomposer notre raisonnement. La matrice SWOT est un outil d'analyse stratégique largement utilisé pour évaluer les forces, faiblesses, opportunités et menaces d'une organisation ou d'un projet. Cet outil se distingue par sa simplicité et son efficacité dans la formulation de stratégies et la prise de décisions.

Ainsi, le présent exposé a vocation à dresser un état des lieux de la manière dont les dispositifs d'intelligence artificielle pourraient être déployés au sein d'une entreprise en mettant en exergue à la fois les bénéfices et les menaces que représentent ces technologies.

Dans une première partie, nous nous attacherons à cibler les bénéfices que peuvent avoir les technologies d'intelligence artificielle (I) pour la sûreté des entreprises et nous nous attacherons à démontrer en quoi elles peuvent représenter une technologie de rupture en matière de sûreté et de transition en matière d'expertise. Dans une deuxième partie, nous nous intéresserons aux menaces (II) que peut représenter l'utilisation de ces technologies, soit lorsqu'elles sont utilisées par les entreprises sans qu'elles aient au préalable définie un cadre d'utilisation clair, soit directement contre ces dernières à des fins malveillantes. Enfin, Cette analyse nous permettra de proposer des principes directeurs (III) visant en pratique, à mettre ces dispositifs au service de la sûreté des entreprises en limitant les risques associés tout en optimisant l'apport fonctionnel dans l'appréciation et le traitement des risques et menace.

Les propos contenus dans ce rapport ayant été recueillis au travers de recherches variées ou auprès d'interlocuteurs spécialisés sur ces thématiques, ce travail ne prétend pas à l'exhaustivité mais vise à mettre en avant les éléments les plus prégnants identifiés au travers de la pratique professionnelle de chacun ainsi que des échanges avec les différents interlocuteurs rencontrés.

● PARTIE 1. INTELLIGENCE ARTIFICIELLE ET SURETE DES ENTREPRISES : QUELS BENEFICES POUR DEMAIN ?

Afin d'étayer notre vision d'un dispositif de sûreté à 360° au sein des entreprises, il apparaît indispensable de présenter les avantages liés au déploiement de dispositifs d'intelligence artificielle au sein des entreprises (A) ainsi que les opportunités qui découlent de cette mise en œuvre (B).

○ I.1. Les avantages résultant du déploiement de l'intelligence artificielle au sein des entreprises

La mise en œuvre de dispositifs d'intelligence artificielle au service de la sûreté des entreprises doit permettre de mettre la technologie au service de la réalisation de l'analyse de risques en entreprise et par là-même, de changer de paradigme quant à la nature même de l'analyse réalisée.

■ I.1.a. Analyse quantitative ou qualitative du risque ?

L'analyse des risques constitue le fondement du déploiement d'un système de sûreté pour toute entreprise. L'appréciation des risques sûreté permet la mise en place d'un mécanisme de protection pertinent, en particulier dans un contexte international où la diversité des menaces et des environnements augmente la complexité de la tâche à accomplir.

L'analyse de risques constitue ainsi le point de départ pour les entreprises en matière de sûreté, car elle permet **d'identifier, d'évaluer et de prioriser** les risques auxquels ces entreprises sont exposées, qu'il s'agisse de menaces internes, externes, physiques, numériques ou liées à l'espionnage industriel.

Cette démarche proactive et indispensable offre la possibilité de détecter les risques auxquels une entité est susceptible d'être soumise et, dans un second temps, de confronter ces risques aux vulnérabilités qui auront été identifiées lors d'une analyse approfondie du dispositif de sûreté (aussi dénommée analyse de vulnérabilités) avant qu'elles ne soient exploitées par différentes sources de menaces. En établissant une hiérarchisation des risques fondée sur leur probabilité et leur impact potentiel, l'entreprise peut optimiser l'allocation de ses ressources pour développer des stratégies de remédiation efficaces, incluant par exemple la mise à jour des protocoles de sécurité, la formation du personnel et l'amélioration des infrastructures. Réaliser une telle analyse constitue par ailleurs une étape essentielle pour maintenir une surveillance continue et s'adapter aux nouvelles menaces dans un environnement sécuritaire en constante évolution, tout en respectant les exigences réglementaires applicables en matière de gestion des risques.

Les méthodologies d'analyse de risques ont d'abord vu le jour dans le secteur des assurances. Les assureurs, pionniers en matière d'analyse de risque, ont cherché à quantifier le risque en termes de pertes financières annualisées, il s'agissait alors du produit des pertes potentielles dues à un événement couplé à la probabilité d'occurrence de l'événement en question.

Cette définition s'est vue enrichie depuis la fin des années 1980 jusque dans les années 2000, pour davantage se prêter au cadre de la sûreté. Ainsi, les universitaires Hoyland et Rausand en 2004[12] définissent le risque comme "*une situation incertaine dans laquelle un certain nombre d'événements pourraient se produire, un ou plusieurs de ces événements pourraient être indésirables* ». En général, la notion de risque fait référence à tous les événements qu'une organisation cherchera à éviter, ces derniers faisant référence à la probabilité que chaque événement se produise, avec une ampleur estimée et une occurrence prédéfinie.

L'analyse de risque constitue la base d'une approche de gestion par les risques. Cette approche peut **être heuristique** (*ad hoc*), **inductive ou déductive**. En d'autres termes, certaines analyses se basent sur **une approche quantitative** alors que d'autres s'inscrivent dans **une démarche qualitative.**

Chacune de ces méthodes apporte une perspective unique et complémentaire, permettant de cerner avec précision les risques auxquels les entreprises sont confrontées et de développer ainsi des stratégies efficaces pour les atténuer.

---

[12] Hoyland&Raussand; "Enhancing of Technical Systems Reliability by Implementing of Risk-Oriented Diagnostics"; 2004

L'approche qualitative se concentre sur l'évaluation des risques fondée sur des jugements subjectifs concernant la probabilité et l'impact des risques identifiés. Cette approche s'appuie sur l'expertise, les connaissances et l'expérience des professionnels pour identifier et classer les risques. Cette méthode est dès lors considérée comme particulièrement utile pour analyser des menaces complexes ou nouvelles, pour lesquelles les données quantitatives peuvent être limitées ou absentes.

Cette approche permet également d'intégrer des facteurs humains et organisationnels dans l'analyse des risques, offrant une compréhension plus nuancée des vulnérabilités et des menaces. Cette approche se fonde généralement sur une approche dite de conformité (ou de "compliance") et le dispositif de sûreté déployé, qu'il soit physique ou numérique, sera alors évalué en fonction de son niveau de conformité au regard de standards reconnus.

L'approche qualitative est davantage utilisée dans le cadre **d'une analyse inductive des risques.** Il est alors possible de parler d'approche "*bottom-up*"[13] : les risques sont identifiés au début de l'analyse et ne sont pas le fruit de l'étude réalisée ; la liste des risques identifiés constituant le point de début de l'analyse. Il sera possible d'utiliser des mécanismes et des lignes directrices internes à l'entreprise tout autant que des normes externes reconnues, telles que les normes de type ISO (International Organization for Standardization).

A contrario, **l'approche quantitative** est plutôt utilisée dans le cadre d'une analyse des risques **déductive**. Cette méthode d'évaluation des risques utilise des données numériques pour évaluer les risques, fournissant une méthode plus objective pour estimer la probabilité et l'impact des menaces. Cette méthode s'appuie sur des analyses statistiques, des modélisations et des simulations pour quantifier les risques, facilitant ainsi la comparaison, la priorisation et la prise de décisions fondées sur des critères mesurables. L'approche

---

quantitative est particulièrement précieuse pour évaluer l'efficacité des mesures de remédiation mises en œuvre ainsi que pour allouer de manière optimale les ressources à mobiliser en matière de sûreté.

Dans le domaine de la sûreté, il est courant de déterminer des probabilités à partir de facteurs subjectifs puisqu'il n'est pas toujours possible de bénéficier de données numériques locales récentes et fiables pour alimenter ces modèles de réflexion. Or, l'ajout de subjectivité dans un modèle quantitatif à tendance à nuire aussi bien à l'exactitude des résultats de l'étude, mais aussi à la crédibilité qui pourra être attribuée à cette dernière dans la mesure où une analyse des risques bien menée est un outil à destination de divers acteurs dont la gouvernance de la structure. De plus en plus souvent, il est possible de constater que ces études de risques sont menées à l'aide de logiciel de simulation mathématique, ces derniers permettent d'apporter davantage d'éléments aux modèles logiques préalablement créés.

L'analyse de risque examine les possibilités d'une attaque réussie de l'adversaire au travers de : sa probabilité d'occurrence, son facteur de survenue et sa gravité. Dans cette dynamique d'analyse des risques, l'analyse cherche à répondre, de manière simple, à trois questions[14] :

1. Quel événement indésirable peut se produire ?
2. Quelle est la probabilité que cet événement se produise ?
3. Quelles sont les conséquences de cet événement ?

Dans la perspective d'une analyse quantitative, ces questions sont synthétisées dans l'équation suivante :

R = PA*(1-PE) *C

- **R** serait le risque que l'événement critique intervienne entre 0 (aucune chance) à 1 (événement certain)
- **PA** serait la probabilité que l'événement se produise durant une durée de temps déterminée
- **PE = PI*PN** soit le produit de la probabilité d'interruption et de la probabilité de neutralisation de l'attaque.
- **C** étant la valeur financière des conséquences

---

[13] MARSAN, D., and M. WYSS (2011), Seismicity rate changes, Community Online Resource for Statistical Seismicity Analysis, doi:10.5078/corssa-25837590. Disponible at http://www.corssa.org.

[14] KAPLAN Stanley, GARRICK B.; "*Reflections and Risk Analysis Papers*"; 1981.

La probabilité que l'événement se produise (PA) est traditionnellement le produit d'une fréquence d'exposition à un risque minoré par la probabilité d'interruption ainsi que par les mesures mises en place de réduction de l'impact du risque. L'équation ci-dessus présuppose qu'il s'agit uniquement du calcul du risque résiduel.
Popularisée au sein du secteur pétrolier et gazier depuis la fin des années 1970 et la naissance des éléments de modélisation du risque qui a suivi la catastrophe de la plateforme Piper Alpha en 1976, cette méthode fait aujourd'hui consensus.

Toutefois, comme nous l'avons évoqué précédemment, il faut constater que cette méthodologie :

- S'appuie sur des éléments mathématiques et des données qu'il est possible de quantifier dans un laps de temps donné parfois (trop) important.

- Qu'il peut être difficile voire impossible de calculer la probabilité de certains événements rares à l'aide de méthodes scientifiques (due à la nature même des très faibles probabilités), c'est ce que le statisticien Nassim TALEB qualifie de signes noir[15]. L'exemple le plus parlant de ces phénomènes en matière de sûreté est le terrorisme : Une probabilité très faible (quasi-nulle) mais un impact significatif qui ne peut en aucun cas être ignoré/négligé par les personnes occupant des fonctions de risk management.

■ I.1.b. L'intelligence artificielle et le bon quantitatif

Les outils d'intelligence artificielle ou incluant une fonctionnalité répondant a des caractéristiques comme le machine learning (apprentissage automatique) peuvent jouer un rôle pour combler le fossé entre les approches qualitative et quantitative de l'analyse des risques. Cette synergie entre l'IA et l'analyse des risques permet de répondre de manière plus complète et dynamique aux questions essentielles concernant les événements indésirables, leur probabilité et leurs conséquences. A titre d'exemple, on peut évoquer la société DELOITTE[16] spécialisée dans l'audit et dans le conseil qui utilise des outils d'intelligence artificielle pour la gestion des risques en intégrant des capacités de calcul cognitif. Les entreprises peuvent ainsi analyser de grandes quantités de données internes et externes pour anticiper les risques. L'IA permet de détecter des tendances et des modèles dans les données non structurées, ce qui peut être utilisé pour prédire les incidents futurs et améliorer la précision des évaluations de risques.

Les limites précédemment identifiées des analyses quantitatives pourraient s'amenuir et la compréhension du risque pourrait tendre vers le tout quantitatif.

Les systèmes fondés sur **l'IA connexionniste,** en particulier ceux utilisant des réseaux de neurones et l'apprentissage profond, se révèlent être des outils particulièrement adaptés et puissants. L'approche connexionniste, qui s'inspire du fonctionnement biologique des cerveaux humains et animaux, permet à l'IA de traiter et d'analyser des volumes importants de données non structurées.
Ainsi, le TLN, une sous-branche de l'IA connexionniste, peut examiner d'importantes collections de textes, de rapports d'incidents contenus dans l'historique des activités liées à la sécurité d'un site, des articles de presse, des contenus de médias sociaux, et autres communications pour **détecter des signaux faibles ou des motifs révélateurs de menaces émergentes.** A titre d'exemple, une augmentation des discussions en ligne autour de certaines vulnérabilités logicielles serait susceptible de révéler un risque accru de cyberattaques exploitant ces faiblesses. En analysant des séries temporelles de données

[15] TALEB, Nassim Nicholas, "*The Black Swan: the impact of the highly improbable*", Londres, 2010, p.366
[16] https://www2.deloitte.com/us/en/pages/consulting/articles/ai-risk-management.html

historiques, l'IA permettrait d'identifier des tendances et des cycles qui échapperaient à de « seuls analyses humains », tels que l'identification de périodes de risques accrus en matière de piraterie maritime ou d'attaques logiciels à partir d'interprétation de la réitération de périodes cycles d'actes pris isolément.
Les réseaux de neurones sont particulièrement doués pour identifier des modèles complexes dans les jeux de données**, y compris des anomalies qui deviennent des « tendances normales »**. Cette capacité peut être utilisée pour détecter des activités suspectes ou inhabituelles dans des données transactionnelles, des logs de sécurité, ou des flux de données de capteurs, révélant ainsi une tentative d'intrusion, une défaillance de système, ou d'autres types de risque.
L'une des forces de l'approche connexionniste est sa capacité à intégrer et à analyser des données provenant de sources hétérogènes. En combinant des informations issues de la surveillance des réseaux sociaux, des rapports d'incidents, des données météorologiques, et même des images satellites, l'IA peut fournir une vue d'ensemble des facteurs de risque et des vulnérabilités potentielles applicables à une entreprise.
L'analyse de l'ensemble de ces jeux de données permettrait ainsi de donner une vision plus exacte du champ des risques qui pèsent sur une activité en particulier. Il va de soi que l'ensemble de ces données devra être analysé au prisme des caractéristiques de l'activité pour laquelle l'analyse est menée, de manière à ce que ces caractéristiques agissent comme une variable mathématique d'adaptation dans la recherche de cotation des risques. Si l'œil humain reste indispensable, l'interconnexion des systèmes et l'accessibilité croissante de la donnée vont tendre à accentuer le niveau de connaissance des risques à la lumière des données disponibles sur un territoire donné et dans un secteur d'activité donné et à réduire, ainsi l'apport de cette analyse humaine. "*La mise en donnée du monde*"[17] et l'avènement du "*datamonde*"[18] accélèrent la production de données disponibles à l'exploitation par la multiplication des capteurs et l'interconnexion croissante des solutions disponibles.
Afin d'évaluer la probabilité d'occurrence d'un événement indésirable, étape fondamentale de notre raisonnement méthodologique, il résulte de ce qui précède que **l'intelligence artificielle connexionniste,** spécifiquement à travers des techniques d'apprentissage automatique et d'apprentissage profond, **offre une méthodologie robuste**. Ces techniques d'apprentissage automatique et d'apprentissage profond permettent de construire des modèles prédictifs capables d'analyser des données historiques et actuelles afin de prévoir la survenue d'événements futurs. Les modèles d'apprentissage automatique peuvent être entraînés en utilisant de vastes ensembles de données historiques qui incluent des occurrences passées d'événements indésirables ainsi que des conditions ou des indicateurs qui ont précédé ces événements. Ainsi, un modèle pourrait être formé avec des données sur des cyberattaques passées, incluant des informations sur les techniques d'attaque utilisées, les vulnérabilités exploitées, et le contexte dans lequel ces attaques se sont produites.

■ I.1.c. Vers une compréhension prédictive du risque sûreté

Un exemple de statistiques qui pourrait servir dans l'évaluation de la probabilité d'occurrence d'un événement indésirable pour une entreprise en matière de sûreté pourrait être les statistiques de la criminalité en France. Les travaux précurseurs de la prévention situationnelle menés dans les années 1980 par certains experts de l'urbanisme comme Ronald V Clarke mettent en évidence le lien qui existe entre la criminalité et l'environnement dans laquelle elle se produit. Ainsi, l'environnement d'implantation d'une entreprise aura toujours un impact sur l'appréciation des risques que fera cette dernière et plus spécifiquement sur le risque sûreté (vol, agression, intrusion…).
La compréhension et l'analyse de la criminalité en France s'appuient sur une collecte rigoureuse de données et de statistiques, historiquement orchestrée par l'Observatoire National de la Délinquance et des Réponses Pénales (ONDRP). L'ONDRP, établi en 2004, a joué un rôle dans la compilation, l'analyse

[17] GOMART Thomas, "Guerres invisibles",Paris,Tallandier, 2021
[18] MERZEAU Louise, "https://find.org/wp-content/uploads/2020/02/cahier-d-enjeux-fing-questionsnumeriques-controverses.pdf, p67

et la diffusion d'informations sur diverses formes de criminalité jusqu'à sa dissolution en 2019, où ses missions et activités ont été reprises par le Service statistique ministériel de la sécurité intérieure (SSMSI), entité rattachée au ministère de l'Intérieur.
Issues des services de police et de gendarmerie, les Statistiques des Crimes et Délits Enregistrés (SCDE) détaillent le nombre d'infractions signalées, offrant ainsi une vue d'ensemble des tendances criminelles sur le territoire.

- **Enquêtes de Victimation :** Elles fournissent des apports importants sur les expériences de victimisation directement auprès des citoyens, capturant ainsi le chiffre de la criminalité non rapportée aux autorités.

- **Statistiques Judiciaires :** Ces informations révèlent le traitement des affaires par les systèmes judiciaires, y compris les poursuites et condamnations.

- **Données Thématiques :** Elles se concentrent sur des sujets spécifiques tels que les violences faites aux femmes, les cambriolages, les cyberviolences, etc., permettant d'analyser des tendances particulières au sein de la criminalité globale.
  Ces statistiques sont mises en ligne directement depuis la plateforme de données ouvertes (open data) gouvernementale au format des grands fichiers de données statistiques CSV. Les statistiques issues de ces jeux de données se décomposent à différents niveaux de granularité, communal, départemental et régional. Au total, ce fichier comptabilise, en juillet 2024, plus de 34.000 lignes avec une nomenclature des faits se décomposant en seize catégories de faits (homicides, vol, cambriolage…).
  Une fois un modèle entraîné, il peut être utilisé pour **évaluer la probabilité d'occurrence** d'événements indésirables **en analysant les conditions actuelles et en les comparant** aux motifs appris. **L'analyse de risque basculerait d'un modèle statique à un modèle dynamique à tendance prédictive,** la puissance de calcul permettrait en temps réel de générer une matrice de risque en prenant en compte l'ensemble des données disponible et en la faisant évoluer au fur et à mesure de l'ajout de données dans sa base de d'apprentissage.
  L'analyse de risque complète ne serait alors générée qu'une fois avec des mises à jour en temps réel.
  L'utilisation des sources de données disponibles en matière de criminalité pourrait être combinée avec des données disponibles sur des sources fiables. L'INSEE, gère un très grand nombre d'études et de bases de données statistiques, couvrant une multitude de domaines tels que la démographie, l'économie, le social, et bien d'autres aspects de la société française. Nombre des études menées par l'INSEE couvrent des aspects qui peuvent entrer en compte dans la notion de criminologie environnementale (évolution et mutations sociales, urbanismes, politiques économiques locales…).
  Ces éléments présentent des briques de compréhension élémentaires qui peuvent permettre d'expliquer l'évolution de la criminalité et de manière plus analytique, la potentialité qu'un tel phénomène se produise contre les actifs d'une société qui se sera implantée dans ces espaces urbains.
  Des chercheurs comme le prix Nobel d'économie Gary Becker[19] mettent en avant l'analyse économique du comportement humain pour expliquer certaines notions importantes comme la criminalité. Ainsi, ce dernier a développé un modèle selon lequel des individus prennent des décisions en évaluant les coûts et les bénéfices. D'autres, comme Steven Levitt mettront en avant les dynamiques socio-économiques ainsi que l'impact des politiques de police pour expliquer ce phénomène.
  Il est possible de composer des modèles basiques composés de jeux de données comprenant les éléments suivants :

- Evolution des populations et des bassins d'habitation
- Evolution des investissements dans les politiques de la ville
- Evolution des catégories socio-professionnelles
- Suivi du taux de pauvreté
- Suivi des prix de l'immobilier
- Suivi du taux de chômage
- …

[19] BECKER Gary S., "*The economic approach to Human Behavior*", The University of Chicago Press, 1976.

Ces données sont aujourd'hui déjà disponibles mais en volume trop importants pour pouvoir être compilées par un être humain de manière intelligible, flexible, adaptée et donc exploitable.
C'est dans ce type de configuration que la capacité d'une IA à traiter un volume important de donnée pourrait s'avérer précieux, à la fois pour une meilleure compréhension du phénomène de criminalité, mais aussi pour les entreprises qui seraient à même de mesurer de manière dynamique les risques liés à la criminalité que représente l'implantation de leurs locaux dans une zone donnée.
Il n'est d'ailleurs pas fantaisiste d'imaginer qu'à ce stade, de tels modèles pourraient être utilisés par les compagnies d'assurance en vue d'adapter leurs polices aux risques pris par les entreprises désireuses de s'implanter dans tel ou tel zone géographique en fonction du risque sûreté.

▪ I.1.d. Une vision à 360° des risques pour les entreprises

Le domaine de la cybersécurité est encore aujourd'hui une spécialité bien à part dans de nombreuses entreprises. Cette fonction est rarement intégrée à la direction sûreté, les méthodes d'analyse de risques sont spécifiques bien que souvent empruntées au monde de la sûreté comme c'est le cas de la méthode EBIOS RISK MANAGER qui reprend les standards d'analyse de risque sûreté en précisant toutefois que la méthode est avant tout déstinée à l'appréciation des menaces issues du monde numérique. Ces standards d'analyse de risques s'appuient également souvent sur des analyses inductives et des évaluations qualitatives pour évaluer les probabilités d'attaques ainsi que le niveau de conséquences d'une attaque.
Les analyses de risque cyber souffrent donc du même biais subjectif que les analyses de risques traditionnelles du monde de la sûreté alors même que les données affluent sur les typologies d'attaques, les cibles visées, leur typologies, etc.
Le "*datamonde*" et la "*métastructure*[20]" sont des réalités qui tendent à rapprocher la dominante cyber de la dominante sûreté comme des secteurs stratégiques d'une entreprise. Les activités technologiques liées à l'intelligence artificielle et à la production de données de masse se concentrent de plus en plus autour d'un petit nombre d'acteurs qui forment un goulot d'étranglement important sur les technologies d'avenir. Ainsi qualifiées, ces entreprises "*Bigtech*"[22] ont la force de détenir à la fois une forme d'infrastructure au sens Marxiste du terme qui s'oppose à une superstructure ou métastructure dans le cas présent. Certains auteurs parlent davantage d'un "*infra-système*" pour qualifier "*l'ensemble des outils, infrastructures, solutions de connectivité, de captation et de traitement des milliards de données : Câbles sousmarins, satellites, datacenters, systèmes d'information, logiciels, algorithmes d'intelligence artificiel, supercalculateurs*…". Dans les faits, ces sociétés et leurs réseaux s'interconnectent de plus en plus fréquemment de par leur nature intrinsèque : une visée large, une captation d'information massive… et aucune loi Anti-Trust ne semble pouvoir ralentir cette tendance structurelle.
Comme l'alpha et l'oméga, les bigTech sont capables d'investir en matière de recherche et développement de manière à construire un réseau qui leur est propre, de capter les données qui en découlent et de les utiliser comme carburant pour les technologies qu'ils souhaitent lancer à l'avenir. Les budgets investis par ces quelques entreprises en matière de recherche et développement démontrent pleinement toute l'importance qui est consacrée à la recherche de ces technologies et au rassemblement de moyens financiers importants dans cet objectif.
A titre d'exemple, le budget d'Amazon en R&D est à lui seul de 43 milliards en dollars[21], ce qui le placerait à la neuvième position mondiale si c'était un pays. Alphabet, Apple, Huawei, Méta et Microsoft dépassent tous largement les 20 milliards de dollars par an. A titre de comparaison, pour l'année 2024,

[20] MHALLA Asma, "Technopolitique",Seuil, Février 2024, p41. [22]. MHALLA Asma, "Technopolitique",Seuil, Février 2024, p38
[21] BAJPAI Prableen, "Which companies Spend the Most in Research and Development (R&D)?", 21 juin 2021, disponible at https://www.nasdaq.com/articles/which-companies-spend-the-most-inresearch-and-development-rd-2021-06-21

le budget total alloué au ministère de l'Enseignement supérieur et de la Recherche en France s'élève à 26,6 milliards d'euros[22].

Si ces entreprises prennent de plus en plus de place dans la « métasphère », c'est qu'elles en sont la composante dominante. Par conséquent, la masse d'information à tendance à se diriger vers ce petit nombre d'acteurs qui seront à même de la traiter *via* leurs infrastructures et leurs capacités de traitement hors de portée d'entreprises intermédiaires. Ainsi, les informations relatives aux cyberattaques sur les systèmes d'informations couverts par les technologies déployées par les *BigTech*, les *modus operandi* et les ressources utilisées sont tant d'informations qui se retrouveront de plus en plus facilement en possession d'entreprises *BigTech*.

Ces dernières seront capables de traiter cette somme de données et de proposer des matrices de risques quantitatives, personnalisées appuyées par des outils d'intelligence artificielle capables d'interpréter les milliers d'attaques réalisées, leurs cibles type et les actifs visés. Ces informations feront également passer la menace cyber dans le champ du tout quantitatif et la méthode de cotation du risque se rapprochera ainsi rapidement de celle des risques traditionnels sûreté auxquels sont confrontés les entreprises.

En synthèse de cette analyse, nous avons cherché à mettre en évidence l'apport que présenterait **l'outil d'intelligence artificielle sur la logique d'appréciation des risques par les entreprises.** Alimentés par une source de donnée importante, variée et pertinente, ces outils d'agglomération et de traitement offriraient **une vision poussée et réaliste des risques** auxquels les entreprises sont confrontées tout en basculant sur une méthode mathématique quantitative et en ostracisant petit à petit les notions de subjectivités qualitatives.

Les données s'agglomèrent dans un outil suffisamment puissant pour pouvoir les interpréter et en tirer les éléments déterminants. Les entreprises pourraient adopter des politiques de prévention efficaces en s'appuyant sur des actions passées menées en rapport avec leur activité et les résultats obtenus dans des cas de figures variés.

Si l'agglomération des données et leur traitement constituent indéniablement une force des dispositifs d'intelligence artificielle, l'hypervitesse en constitue une autre.

■ I.1.e. Intelligence artificielle, hypervitesse et hybridation par nature

"*Lorsqu'on change d'échelle, les phénomènes changent non seulement de grandeur, mais aussi de nature*". Cette citation, tirée des travaux du géographe Français Dollfus en 1970 s'applique parfaitement à la notion d'hypervitesse. En effet, au niveau le plus atomique, celui de l'individu, elle fait référence à la vitesse d'exécution, de traitement et de réponse des outils incluant l'algorithmique ou, de manière plus complète, au travers de l'apprentissage, à l'intelligence artificielle. Cette vitesse défie la capacité cognitive humaine, à la fois dans la captation de données, mais aussi dans son interprétation et dans la réponse qu'elle apporte.

L'hypervitesse permet de réagir **plus efficacement à un événement de sûreté** dont les scénarios auront été identifiés à l'avance et où les données d'entrées correspondent à des schémas type appris et étudiés par les outils de traitement.

Dans un schéma d'hypervitesse où un outil de traitement de l'information est capable de prendre une décision adéquate en une fraction de seconde, il apparaît essentiel de se poser la question de la place de l'Être Humain et de sa sortie progressive de la boucle décisionnelle. Ce passage de la théorie du Centaure vers celle du berger vise à garantir une réaction à la fois rapide et adaptée tout en conservant une supervision humaine au moins dans un premier temps. **La machine n'est pas affectée par des éléments externes comme la fatigue, le stress ou l'inattention et sa vitesse d'analyse et de rédaction est toujours la même quels que soient les facteurs externes.**

[22] https://www.enseignementsup-recherche.gouv.fr/fr/projet-de-loi-de-finances-2024-92670

L'intelligence artificielle n'est pas la seule technologie d'hypervitesse. Les réseaux sociaux, les menaces issues du web en sont aussi un exemple. Ces menaces disposent d'une automatisation croissante et de capacités **de latéralisation accélérées qui en font des menaces de plus en plus prégnantes pour les systèmes d'information des entreprises**. De plus en plus de solutions logicielles de cybersécurité intègrent des capacités de traitement à haute vitesse de manière à pouvoir identifier des modèles ("pattern"), des indices ou des comportementaux susceptibles d'être la source de manipulations malveillantes. Les outils incorporant de l'IA peuvent isoler rapidement un segment de réseau compromis ou mettre en quarantaine un fichier suspect de manière automatique. **Coordonner la réponse et agir de manière automatisée sont des avantages déterminant pour un système de sûreté abouti à la fois dans le monde cyber et dans le monde physique.** Cette capacité à réagir instantanément est essentielle pour limiter les dommages en cas d'attaque. Cet apport d'analyse et de vitesse de réaction a parfaitement été intégré par les sociétés spécialisées qui proposent de plus en plus d'utiliser les capacités de traitement, d'interprétation et de réponse offertes par les technologies d'intelligence artificielle pour maximiser l'efficience de leurs solutions de cybersécurité.

Les qualités d'hypervitesse et d'analyse approfondie de données jumelées à une faculté d'apprentissage automatique fond de l'intelligence artificielle un outil hybride par excellence, capable d'interagir à la fois avec le monde physique et avec le monde numérique. Il existe toutefois un prérequis important sur lequel nous reviendrons plus avant dans cette étude : **les sources et les capteurs de données.**

Outre ce qui précède, qu'en est-il de cette notion d'hypervitesse lorsque l'on change d'échelle et que cette simple notion aujourd'hui, constituera demain un caractère intrinsèque de cette vague technologique à venir allant même jusqu'à pouvoir provoquer la mutation de la société dans son ensemble ? Dans cette optique, on ne parle plus d'hypervitesse pour qualifier la vitesse avec laquelle ces outils vont être capables de lire et d'analyser des données mais plutôt pour qualifier la vitesse de propagation des outils dans la société. Comme nous l'avons évoqué précédemment, **c'est le propre des technologies d'intérêt général que de provoquer de profondes mutations de la société, l'intelligence artificielle n'échappe pas à cette règle de par l'étendu du potentiel des technologies sur laquelle elle repose.**

Les conflits géopolitiques ont une tendance à accélérer le déploiement et l'hybridation de ces technologies (du civil grand public au monde militaire) et à rendre ces adaptations accessibles au plus grand monde, démocratisant ainsi des objets connectés au potentiel de destruction important. Tout comme l'intelligence artificielle, les technologies liées aux drones (aériens ou terrestres) sont en ce sens aussi, des technologies dites d'hypervitesse. Les potentiels d'utilisation de ces dispositifs sont multiples et très variés, ce qui peut être utilisé comme des technologies à usage civile peuvent facilement l'être à usage militaire, devenant ainsi des biens à double usage.

Le risque qu'un drone équipé d'un engin explosif improvisé s'en prenne à une entreprise est bien réel et c'est même le quotidien de nombreuses entreprises russes ou ukrainiennes dans le cadre du conflit qui oppose ces deux Nations. Ce risque, même en temps de paix, tend à s'accentuer à mesure de la multiplication de ces dispositifs en accès libre et au manque de réglementation associé.

■ I.1.f. L'apport de l'IA sur les fonctions primaires d'un système de sûreté

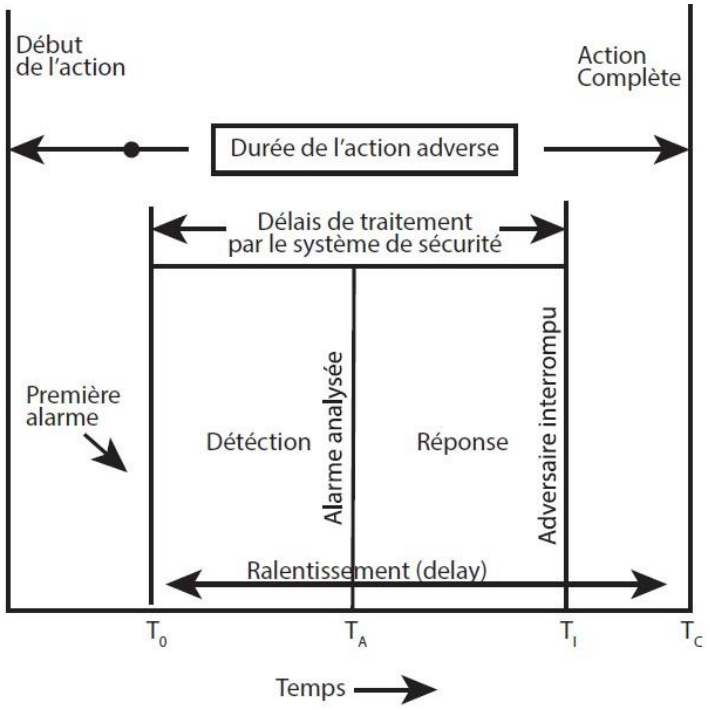
Dans le domaine de la sûreté des entreprises, il est communément admis qu'il existe trois fonctions primaires à un système de protection physique : **la détection, le retardement et la réponse**[23] une fonction secondaire souvent évoquée dans la littérature académique est la **dissuasion**. Pour une entreprise, l'objectif est de maximiser l'efficacité de ces trois dimensions face à des menaces identifiées et évaluées. Nous avons vu la manière dont l'IA était susceptible de jouer un rôle dans l'approche au risque et dans la quantification mathématique du niveau de menace.

Pour ce qui est de l'impact de l'IA et de son hypervitesse au niveau micro, il convient de s'intéresser à ces trois fonctions primaires et de les étendre au-delà du seul périmètre physique défini par Garcia en

[23] Garcia, 2008, pp.2-6

2008. Puisque l'IA tend à générer un pont entre les mondes physiques et cyber, **il faut voir ces trois dimensions à 360° pour prendre pleinement en considération la sûreté globale d'une entreprise.**

La séquence d'action de l'adversaire ci-contre, décompose les différentes actions dans le temps. On peut y voir ainsi la succession des actions de ralentissement, de détection et d'interruption. Dans cette composition, un outil d'intelligence artificielle permettrait de diminuer grandement le temps de traitement de l'alerte par le système réduisant de fait l'espace-temps qui se situe entre la première alarme et l'interruption de l'adversaire. Le délai obtenu par la fonction ralentissement peut ainsi être plus court et la pertinence de la réponse plus adaptée aux caractéristiques de l'attaquant (typologie d'action, moyens et motivations).



Pour mieux comprendre encore l'apport de l'IA sur ces trois fonctions primaires, il est intéressant de reprendre pour exemple une menace émergente qui prend de plus en plus de place dans le panorama des risques sûreté en entreprise : **la menace des drônes**. En 2022, plus de 2 000 drones ont été observés près des aéroports au cours des six premiers mois[24], ce qui représente un risque majeur pour la sûreté du trafic aérien, mais aussi pour les infrastructures au sol. Ces incidents ont eu un impact financier considérable, comme le montre l'exemple de la fermeture de l'aéroport de Gatwick en 2018, qui a coûté plus de 60 millions de dollars.

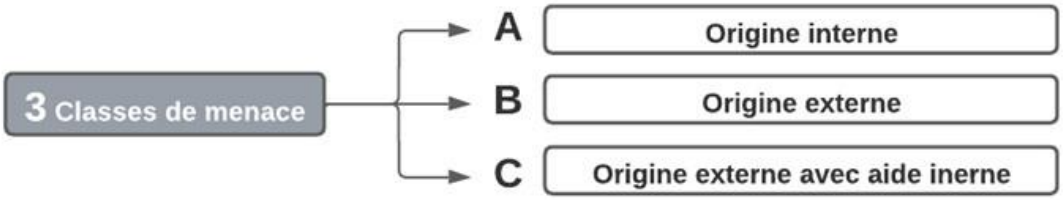
Dans cette configuration, une technologie d'IA munis de capteurs de données efficients pourrait :

- Lancer rapidement une étude sur le type d'onde montante télépilote/pilote de manière à catégoriser le dispositif comme nuisible (**fonction détection**), à en discriminer le modèle dans une base de données constituée de milliers de modèles enregistrés, à en évaluer la trajectoire, voir la position du télépilote.
- Face au risque identifié, le système pourrait aussi au préalable d'une attaque construire le site de manière à maximiser les probabilités d'interception en allongeant l'espace libre devant être parcouru par le drone avant de pouvoir toucher une infrastructure sensible (**fonction ralentissement**).
- Le dispositif pourrait décider de manière automatique de marche à suivre pour répondre (**fonction réponse**) à la menace (interception, neutralisation, simple signalement…).

La boucle Captation/Analyse/Décision/Réponse aurait une vitesse d'exécution dépassant de loin les capacités dévolues à un Être Humain et étant parfaitement compatible avec la neutralisation de la menace dans un temps quasi simultané à celui de l'action de cette dernière.

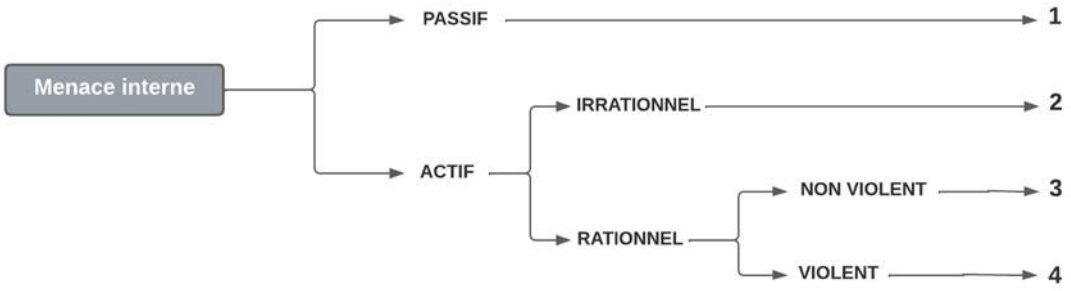
L'optimisation de la boucle d'action dépend aussi du type de menace qui est traité par le système de sécurité. En sûreté, on identifie couramment 3 classes de menaces, ces menaces revêtent des caractéristiques différentes ainsi que des particularités intrinsèques. Il n'est pas rare de retrouver l'ensemble de ces menaces dans le spectre des menaces type qu'il est possible d'identifier pour un site donné. La menace type étant la menace contre laquelle le système de sécurité a été spécialement créé.

[24] QUARLES Gregory ; "*The security implication of drone threat*"; 22 mai 2023.



Les statistiques tendent à montrer que la menace interne est responsable de la plus grande partie des événements de sécurité sur un site. Plusieurs études montrent également que la menace interne fait partie des plus difficiles à appréhender dans la mesure où les mesures de

protection classiques sont inefficaces, ces mêmes études montrent que c'est l'appât du gain financier qui est leur première motivation.



Les systèmes de sécurité physiques sont généralement bien plus efficaces pour lutter contre les menaces d'origine externe, la menace interne est beaucoup plus difficile à traiter de part sa la nature légitime de l'instigateur. Au travers ce prisme de vision, il est possible de distinguer les apports des outils d'intelligence artificielle sur les deux types de menace :

- **Externe** *via* l'intégration des systèmes d'intelligence artificielle dans la protection périmétrique et volumétrique des éléments
- **Interne** *via* le traitement des comportement suspects sur des actifs vitaux de l'entreprise

L'exemple précité illustre par ailleurs l'impact que pourrait avoir l'intelligence artificielle sur chacune des composantes d'un système de sûreté en particulier sur des menaces émergentes qu'un système classique n'arrive pas à traiter. C'est d'ailleurs ces composantes que l'on retrouve *in fine* dans certains systèmes de sécurité informatique comme les EDR (Endpoint detection & response) les plus évolués sans que la littérature académique ne soit venue les conceptualiser alors même qu'il s'agit très clairement des trois dimensions présentées (**détecter, ralentir et répondre) qui doivent permettre d'évaluer le système de sûreté d'une entreprise *a fortiori* lorsqu'une couche additionnelle d'intelligence artificielle est intégrée.**

○ I.2. Les opportunités identifiées par le déploiement de l'intelligence artificielle au sein des entreprises

L'intelligence artificielle est souvent qualifiée d'innovation de rupture tant elle vient modifier en profondeur l'appréhension des individus et des collectivités avec les fondamentaux organisationnels qui les structurent. L'analyse à haute vitesse des grands ensembles de données ouvre aux entreprises la capacité de disposer d'informations intelligibles, interprétables et structurées selon des modèles définis et optimisés aux fins visées.

L'intelligence artificielle appliquée au modèle de l'entreprise redéfinit les modèles d'organisations aussi bien à l'échelle des départements que dans leurs ensembles opérationnels. La Direction Générale dispose d'informations analytiques et prédictives traitant de l'évolution des marchés ou du comportement des consommateurs, tout comme le département IT sera en mesure d'automatiser une partie de ses actions et le département marketing de disposer d'outils analytiques automatisés.
Cependant, les gains d'efficacité et de productivité ne sont pas automatiquement certains dès lors que les prérequis d'intégration de la technologie n'ont pas été pris en considération. Le déploiement de l'IA impose des décisions stratégiques et tactiques qui requièrent une adaptation des processus, une adéquation des compétences, la vigilance portée sur les questions éthiques et de conformité et enfin des investissements financiers conséquents.
Ainsi, la prise de décision d'investissement dans les technologies d'intelligence artificielle doit être **construite et éclairée.** L'intégration de nouvelles méthodes productives grâce à l'essor de nouvelles technologies est jalonnée d'exemples d'accidents industriels, car réalisée sans un travail en amont d'analyse des besoins et d'identification précise des gains recherchés et des investissements nécessaires.
Nous allons essayer d'identifier quels sont les acteurs de la construction de cette stratégie d'intégration, le périmètre de leur engagement opérationnel, les avantages escomptés ainsi que les risques à éviter.
Au sein du tableau ci-dessous, nous pouvons constater que de nombreux domaines d'applications sont éligibles au déploiement des technologies d'intelligence artificielle.
Cependant, il faut comprendre qu'il y aura autant de modèles d'IA que de champs d'application.
A l'échelle décisionnelle, cela implique une priorisation des objectifs selon la criticité des avantages de l'entreprise à sélectionner un champ d'applicabilité de l'IA par rapport à un autre. Si une entreprise possède un outil productif performant, mais identifie une lacune dans la détection et l'orientation de son marché, alors elle devra privilégier un investissement probablement dans le domaine du marketing pour renforcer les efforts de son équipe commerciale.

Domaine d'application	Opportunité	Avantages
Optimisation des opérations	Automatisation des processus métier	Réduction des coûts, augmentation de l'efficacité, précision améliorée
	Gestion de la chaîne d'approvisionnement	Prévision de la demande, optimisation des stocks, réduction des délais
Expérience client	Chatbots et assistants virtuels	Service client 24/7, personnalisation des interactions, réduction des temps de réponse
	Analyse des données client	Segmentation précise, anticipation des besoins, amélioration de la satisfaction client
Innovation et développement	Conception assistée par l'IA	Accélération du cycle de développement, réduction des coûts de prototypage, amélioration de la qualité des produits
	Détection des tendances du marché	Analyse prédictive, innovation guidée par les données, identification rapide des opportunités
Prise de décision	Analyse des Big Data	Extraction d'insights stratégiques, support à la prise de décision, amélioration des stratégies d'affaires
	Modèles prédictifs	Prévision des performances, gestion proactive des risques, optimisation des ressources
Sécurité et gestion des risques	Détection des fraudes	Surveillance en temps réel, réduction des pertes financières, amélioration de la confiance des clients
	Cyber sécurité	Identification des menaces, renforcement des défenses, réduction des risques de cyberattaques
Productivité des employés	Automatisation des tâches répétitives	Libération du temps pour des tâches à valeur ajoutée, réduction de la fatigue des employés
	Outils de collaboration intelligente	Optimisation de la gestion de projets, facilitation du travail à distance, amélioration de la communication
Marketing personnalisé	Publicité ciblée	Campagnes plus efficaces, augmentation du ROI, engagement client amélioré
	Recommandations de produits	Augmentation des ventes croisées et additionnelles, personnalisation des offres, fidélisation client
Ressources humaines	Recrutement assisté par l'IA	Sélection plus rapide et précise des candidats, réduction des biais, amélioration de la qualité des recrutements
	Formation et développement	Programmes de formation personnalisés, suivi des performances, développement des compétences
Modèles économiques	Transformation des modèles d'affaires	Émergence de nouveaux secteurs, modification des chaînes de valeur, création de nouvelles sources de revenus
	Économie de l'IA	Monétisation des technologies IA, exploration de nouvelles opportunités commerciales, innovation continue

Tableau des domaines d'application[25]

Dans le domaine plus spécifique à la sûreté, les opportunités développées par l'IA ne se limitent pas à une seule fonction ou à un seul département de l'entreprise. Elles proposent des avantages transversaux qui influencent divers aspects de celle-ci :

- **Efficacité Opérationnelle :** L'automatisation et l'amélioration des processus de sûreté permettent une meilleure allocation des ressources, réduisant les coûts et augmentant la productivité.
- **Résilience organisationnelle :** La capacité à détecter et à prévenir les menaces améliore la résilience globale de l'entreprise face aux incidents de sécurité.
- **La confiance et la réputation :** Une gestion structurée de la sûreté renforce la confiance des clients, des partenaires et des investisseurs.
- **L'innovation et la compétitivité :** L'utilisation de technologies avancées pour la sûreté positionne l'entreprise comme un leader technologique, augmentant ainsi sa compétitivité sur le marché.

L'application de l'IA à destination de la sûreté des entreprises permet de couvrir un large panel d'applications défensives tout synchronisant les différents acteurs de la chaîne de valeur de l'entreprise

[25] Tableau des domaines d'application de l'IA au sein des entreprises – Cédric Jutteau 07/2024

autour de lignes directrices (Gouvernance). Il est démontré que prendre “soin” de la sûreté des acquis d’une entreprise est un cercle vertueux qui engage tous les membres d’une entreprise vers l’excellence. Au sein du tableau ci-dessous, nous pouvons identifier que les problématiques de sûreté ne sont pas l’apanage de la Direction Sûreté, même si cette dernière en définit les règles et en valide les processus, mais s’égrènent au sein de toutes les fonctions de l’entreprise. Par une meilleure surveillance des opérations, nous pouvons escompter une expérience utilisateur plus encadrée, un processus client optimisé, des risques de pertes de compétences maîtrisés et enfin une gestion du personnel qui tend vers une plus grande fidélisation.

Domaine d'application	Opportunité	Avantages
Détection et prévention des fraudes	Analyse en temps réel des transactions	Identification rapide des comportements suspects, réduction des pertes financières, amélioration de la confiance des clients
	Algorithmes de détection des anomalies	Précision accrue dans la détection des fraudes, réduction des faux positifs, réponse rapide aux menaces
Cybersécurité	Identification des menaces potentielles	Surveillance proactive des cybermenaces, prévention des attaques, renforcement des mesures de sécurité
	Réponse automatique aux incidents	Réduction du temps de réponse aux incidents, limitation des dommages, continuité des opérations
	Protection des données sensibles	Chiffrement avancé, détection des accès non autorisés, protection contre les violations de données
Surveillance et sécurité physique	Reconnaissance faciale et analyse vidéo	Surveillance en temps réel, amélioration de la sécurité des installations, prévention des intrusions
	Détection des comportements anormaux	Identification des comportements suspects, prévention des incidents, amélioration de la sécurité sur le lieu de travail
Gestion des accès et des identités	Authentification biométrique	Sécurisation des accès, réduction des risques de vol d'identité, amélioration de l'expérience utilisateur
	Gestion intelligente des accès	Contrôle d'accès basé sur l'IA, ajustements dynamiques des niveaux d'accès, suivi des activités suspectes
Sécurité des réseaux	Surveillance des réseaux en temps réel	Détection proactive des anomalies, protection contre les cyberattaques, maintien de l'intégrité des réseaux
	Analyse prédictive des menaces	Anticipation des attaques, renforcement des défenses, adaptation des stratégies de sécurité en temps réel
Conformité et gestion des risques	Analyse de conformité automatisée	Surveillance continue des régulations, réduction des risques de non-conformité, amélioration de la gestion des audits
	Gestion proactive des risques	Identification des vulnérabilités, évaluation continue des risques, mise en place de stratégies d'atténuation

Tableau des domaines d’application à la sûreté de l’IA [28]

[28] Domaines d’application de l’IA dans le domaine de la sûreté – Cédric Jutteau – 07/2024

■ I.2.a. Détection et Prévention des risques financiers

Internet et les technologies de communication ont redéfini la vitesse des transactions financières informatisées. Aujourd’hui, nous parlons de transactions boursières réalisées à la milliseconde, engendrant ainsi une croissance exponentielle des volumes de transactions. L'intelligence artificielle se positionne comme l’une des solutions afin de répondre aux problématiques de qualification de véracité des transactions (lutte contre la fraude) ainsi qu’aux enjeux de confiance exprimés par les clients.

- **Détection et Prévention des Fraudes**

  L'une des principales opportunités offertes par l'analyse en temps réel des transactions est **la détection et la prévention des fraudes**. Les algorithmes d'apprentissage automatique peuvent analyser des milliers de transactions par seconde, identifier des modèles et des anomalies qui seraient impossibles à détecter pour les humains. Par exemple, des comportements inhabituels tels que des transactions provenant de lieux géographiques différents en peu de temps ou des achats anormalement élevés peuvent être instantanément signalés pour une vérification supplémentaire.

- **Réduction des pertes financières :** En détectant les fraudes avant qu'elles ne causent des dommages significatifs, les entreprises peuvent économiser plusieurs millions de dollars.
- **Amélioration de la confiance des clients :** Les clients sont rassurés de savoir que leurs transactions sont surveillées et protégées, ce qui renforce leur fidélité.
- **Conformité accrue :** Les entreprises peuvent se conformer plus facilement aux réglementations financières et de sécurité, évitant ainsi des amendes et des sanctions.

  Pour aller plus loin dans cette logique et donner un exemple qui illustre à la fois l'utilisation d'algorithmes d'apprentissage automatiques et le traitement de données financières importantes, raisonnons autour d'une société particulière. La société britannique INTELLI-Q spécialisée dans la prévention des pertes basées sur les transactions de caisse dans le secteur du retail déploie pour ses clients une solution logiciel particulièrement prometteuse.

  En effet, cette solution propose de faire remonter automatiquement les données de caisses vers leur solution logicielle et de procéder à l'analyse de l'ensemble des données. L'outil est ainsi capable de traiter des milliers d'opérations de caisses en temps réel et d'analyser les comportements suspects. La solution se concentrera ainsi sur la démarque inconnue interne en identifiant certains modes opératoires redondants chez les opérateurs de caisse :
- Remboursements importants et réguliers d'un caissier sur une même carte de crédit
- Annulation de ventes redondantes
- Ventes à bas prix régulières
- Editions manuelles des prix redondants
- Remboursements systématiques en espèce
- …

  Si cette solution n'est pas encore optimisée en incluant une notion d'apprentissage automatique, on imagine très bien la plus value d'une telle mise à niveau :
- Apprentissage des nouvelles techniques de fraudes d'un client à l'autre
- Préconisations dynamiques de lutte contre la fraude
- Prise en compte d'un contexte sécuritaire plus large incluant les tendances économiques et sociales au niveau national et/ou international

  L'interconnexion des outils (dans cet exemple les caisses enregistreuses de divers établissements de vente) et l'hypervitesse de traitement des données à grande échelle permettent ainsi de sécuriser un périmètre très large au niveau national et international. Si cet exemple illustre la lutte contre les menaces financières d'origine interne, il est aussi possible de traiter les menaces d'origine externe avec le même processus d'analyse dynamique.
- **Optimisation des opérations financières**

  L'analyse en temps réel des transactions permet d'optimiser les opérations financières. Les entreprises peuvent obtenir des aperçus instantanés sur leurs flux de trésorerie, identifier des inefficacités et prendre des décisions éclairées pour améliorer leur gestion financière. Par exemple, en analysant les transactions en temps réel, une entreprise peut ajuster ses stratégies de prix, optimiser ses inventaires et améliorer ses processus de facturation et de recouvrement.
- **Réactivité accrue :** Les entreprises peuvent réagir rapidement aux fluctuations du marché et aux changements dans les comportements des clients.
- **Efficacité opérationnelle :** En automatisant l'analyse des transactions, les entreprises réduisent la charge de travail des équipes financières, leur permettant de se concentrer sur des tâches à plus forte valeur ajoutée.
- **Réduction des erreurs :** Les analyses automatisées réduisent les erreurs humaines, améliorant ainsi la précision des rapports financiers.
- **Préserver le capital Confiance**

  L'intégrité et la transparence des transactions sont essentielles pour maintenir la confiance des partenaires commerciaux et des investisseurs. Les capacités d'analyse en temps réel de l'IA permettent de garantir que toutes les transactions sont surveillées et vérifiées de manière continue, offrant une transparence totale. Cela est particulièrement important pour les entreprises opérant dans des secteurs

hautement réglementés ou pour celles ayant des relations complexes avec des partenaires commerciaux multiples.

- **Transparence :** Les partenaires et les investisseurs ont une visibilité accrue sur les opérations financières de l'entreprise, renforçant la confiance mutuelle.
- **Réputation :** Une surveillance rigoureuse et une gestion proactive des transactions améliorent la réputation de l'entreprise en tant qu'entité fiable et sécurisée.
- **Financier :** Les entreprises capables de démontrer une gestion financière solide et transparente attirent plus facilement les investisseurs et les partenaires stratégiques.

L'analyse en temps réel des transactions assure la sécurité financière des entreprises et les dote d'une nouvelle capacité de prise de décision. Les entreprises peuvent non seulement se protéger contre les fraudes et les inefficacités, mais aussi renforcer leur position sur le marché et gagner la confiance des clients, des partenaires et des investisseurs.

■ I.2.b. La cybersécurité : Réseaux, Données & Utilisateurs

Les cyberattaques devenant de plus en plus sophistiquées et fréquentes, la cybersécurité constitue une priorité absolue pour les entreprises. La capacité à détecter et à répondre rapidement aux menaces est essentielle pour protéger les données sensibles, les
infrastructures critiques et la réputation de l'entreprise. L'intelligence artificielle peut jouer un rôle déterminant dans cette lutte en offrant des solutions avancées pour l'identification en temps réel des menaces de cybersécurité.

- **Surveillance continue et proactive**

L'une des principales opportunités de l'IA en cybersécurité est la capacité à surveiller en continu les réseaux et systèmes pour détecter les menaces potentielles. Les algorithmes d'IA peuvent analyser des volumes massifs de données en temps réel, identifier des anomalies et des comportements suspects, déclencher des alertes avant que les menaces ne se transforment en incidents majeurs.

- **Réactivité accrue :** La surveillance proactive permet de détecter les menaces dès leur apparition, réduisant ainsi le temps de réaction.
- **Supervision 24/7 :** Les systèmes d'IA peuvent fonctionner sans interruption, offrant une protection constante contre les cybermenaces.
- **La qualification :** Les algorithmes d'apprentissage automatique améliorent continuellement leur précision, réduisant les alertes inutiles et permettant aux équipes de se concentrer sur les menaces réelles.
- **Détection des comportements anormaux**

Les cyberattaques modernes exploitent souvent des comportements atypiques et subtils qui échappent aux systèmes de sécurité traditionnels, que nous caractérisons comme le traitement des signaux faibles. L'IA excelle dans la détection de ces anomalies en analysant les comportements habituels des utilisateurs et des systèmes, et en signalant toute déviation significative.

- **Les nouvelles menaces :** Les attaques de type « zero-day » et les menaces internes peuvent être détectées grâce à l'analyse comportementale, même si elles n'ont pas été précédemment répertoriées.
- **Protection anticipée :** La détection précoce des comportements anormaux permet de prendre des mesures correctives avant que des dommages ne surviennent.
- **Prévention proactive :** En identifiant les comportements suspects, les entreprises peuvent prévenir les fuites de données sensibles et les violations de sécurité.

La concordance des signaux faibles grâce à l'intelligence artificielle est un nouvel atout dans le domaine de la sûreté, en permettant aux entreprises et aux organisations de détecter et d'anticiper des menaces avant qu'elles ne deviennent critiques. Les signaux faibles sont des indices discrets et souvent difficiles à identifier qui, pris individuellement, ne semblent pas alarmants, mais qui, combinés, peuvent révéler des tendances ou des risques émergents. L'IA, notamment à travers « *le machine learning »* et l'analyse prédictive, offre des outils puissants pour repérer ces signaux épars au sein de vastes volumes de données, les corréler et en extraire des informations pertinentes pour la sûreté.

Cette capacité de l'IA à traiter des données en temps réel et à détecter des corrélations invisibles à l'œil humain ou émaillés dans le temps, est particulièrement utile pour anticiper des incidents complexes.

Dans les environnements industriels, les capteurs IoT/OT couplés à l'IA surveillent continuellement les opérations pour détecter des signaux faibles comme des variations de température anormales ou des vibrations inattendues dans les machines, indicateurs potentiels de défaillances techniques ou de tentatives de sabotage. La concordance de ces signaux permet de déclencher des alertes précoces et d'engager des mesures correctives avant que les anomalies ne se transforment en pannes coûteuses ou en accidents graves.

Dans le domaine de la sûreté physique, des systèmes de vidéosurveillance intelligente utilisent des modèles de reconnaissance d'image pour analyser les flux vidéo en temps réel et repérer des comportements suspects, tels qu'une présence prolongée dans des zones restreintes ou des mouvements anormaux de personnes ou de véhicules.

La gestion de la concordance des signaux faibles grâce à l'IA s'avère également déterminante pour la protection contre les menaces internes, un risque souvent sous-estimé mais potentiellement dévastateur pour les entreprises. En analysant des signaux discrets tels que des changements de comportement des employés, des accès inhabituels aux données sensibles, ou des communications anormales, l'IA peut aider à identifier des situations à risque, comme des intentions de fuite d'informations ou des comportements frauduleux, avant qu'ils ne causent des dommages significatifs.

Cependant, le défi de ces promesses de détection des signaux faibles réside dans la capacité de l'entreprise à disposer de données de qualité et à tester ces algorithmes sur des ensembles de données diversifiés et représentatifs des menaces potentielles. L'interprétation des résultats doit être également qualifiée et soumise à un volume conséquent afin d'éviter des prises de décisions erronées et sur la base d'un faible volume de traitement.

- **Réponse automatisée**
  L'IA ne se contente pas de détecter les menaces, elle peut également automatiser la réponse aux incidents de cybersécurité. Les systèmes d'IA peuvent isoler les menaces, bloquer les accès non autorisés et exécuter des protocoles de réponse en temps réel, minimisant ainsi les impacts des attaques (Zero Trust Network Access)**.**
- **Réduction du temps de réponse :** Les réponses automatiques permettent de neutraliser les menaces en quelques secondes, réduisant les risques de propagation.
- **Optimisation des ressources humaines :** En automatisant les tâches répétitives et urgentes, les équipes de cybersécurité peuvent se concentrer sur des analyses plus complexes et des stratégies à long terme.
- **Minimisation des Dommages :** Une réponse rapide et efficace limite les dommages potentiels et accélère la reprise après incident.
- **Anticiper les menaces**
  L'IA permet également d'anticiper les attaques avant qu'elles ne surviennent grâce à l'analyse prédictive. En utilisant des modèles prédictifs fondés sur des données historiques et des tendances actuelles, les entreprises peuvent identifier les menaces potentielles et renforcer leurs défenses proactivement.
- **Anticipation des attaques :** Les entreprises peuvent prendre des mesures préventives basées sur les prédictions, renforçant ainsi leur posture de sécurité.
- **Planification Stratégique :** L'analyse prédictive permet d'optimiser les ressources et de planifier des stratégies de sécurité plus efficaces.
- **Réduction des coûts :** En prévenant les attaques avant qu'elles ne causent des dommages, les entreprises peuvent économiser des coûts importants associés aux violations de sécurité.
- **Protéger les données sensibles**
  La protection des données sensibles est une priorité pour les entreprises, et l'IA offre des solutions avancées pour sécuriser ces informations. Les systèmes d'IA peuvent détecter les accès non autorisés, surveiller les transferts de données et appliquer des politiques de sécurité strictes en temps réel.

- **Sécurisation des Données :** Les entreprises peuvent protéger efficacement les données sensibles contre les accès non autorisés et les fuites.
- **Conformité Réglementaire :** Les solutions de sécurité fondées sur l'IA aident les entreprises à respecter les réglementations en matière de protection des données, réduisant ainsi les sanctions et les amendes.
- **Confiance des Clients :** En garantissant la sécurité des données, les entreprises renforcent la confiance des clients et des partenaires.

L'intelligence artificielle représente une avancée significative dans le domaine de la cybersécurité, offrant des capacités de détection et de réponse en temps réel qui surpassent largement les approches traditionnelles. En surveillant continuellement les réseaux, en détectant les comportements anormaux, en automatisant les réponses aux incidents, en anticipant les attaques et en protégeant les données sensibles, l'IA permet aux entreprises de se défendre efficacement contre les menaces cybernétiques modernes. L'adoption de ces technologies avancées est essentielle pour toute entreprise cherchant à renforcer sa sécurité et à maintenir sa résilience face aux cybermenaces en constante évolution.

Exemples de fonctionnement :

Analyse du comportement : L'IA analyse le comportement habituel des utilisateurs et des systèmes, en étudiant des schémas d'accès aux données, les heures de connexion, les types de fichiers consultés, etc.

Détection des anomalies : Un employé qui accède à des données sensibles à des heures inhabituelles ou un volume de données inhabituellement élevé transféré vers un périphérique externe.

Réponse automatisée : Bloquer l'accès à certaines données, alerter les équipes de sécurité ou activer des protocoles de sécurité renforcés en adéquation avec les règles d'habilitation.

Analyse prédictive : En analysant des millions de points de données, l'IA peut identifier les tentatives de violation en identifiant les schémas relatifs à une attaque.

■ I.2.c. La surveillance et la sécurité physique

La sécurité physique est une composante essentielle de la protection des actifs d'une entreprise, incluant les biens matériels, les informations sensibles et les personnes. Avec les avancées en intelligence artificielle, la surveillance et la sécurité physique ont connu des transformations significatives, offrant de nouvelles opportunités pour renforcer la protection et l'efficacité opérationnelle. Cependant, la sûreté périmétrique est également une source de défis pour les technologies d'intelligence artificielle au regard des méthodes de tromperies / obfuscation dont l'efficacité ont déjà été reconnues.

● Reconnaissance faciale et analyse vidéo

L'une des applications les plus puissantes de l'IA en matière de sécurité physique est la reconnaissance faciale et l'analyse vidéo. Les systèmes de vidéosurveillance intelligents équipés d'IA peuvent identifier des individus en temps réel, détecter des comportements suspects et alerter les équipes de sécurité en cas de menace potentielle.

Amélioration de la Sécurité : La reconnaissance faciale permet d'identifier rapidement les intrus et les personnes interdites d'accès.

Réduction des temps de réaction : Les alertes en temps réel permettent aux équipes de sécurité d'intervenir rapidement pour prévenir les incidents.

Efficacité des enquêtes : Les enregistrements vidéo analysés par l'IA facilitent les enquêtes post-incident en fournissant des preuves claires et des pistes d'investigation.

Depuis les attentats du 11 septembre 2001, la Chine occupe une position clé dans la course aux technologies de surveillance. Ayant introduit la reconnaissance faciale via des caméras lors des Jeux Olympiques de 2008, elle rivalise désormais avec les États-Unis en termes d'investissements dans l'intelligence artificielle. L'ambition de Pékin est double : d'une part, consolider sa sécurité intérieure et, d'autre part, prendre la tête mondiale de l'intelligence artificielle d'ici 2030, avec un budget annuel dépassant les 60 milliards de dollars. À titre de comparaison, la Commission européenne a annoncé un budget annuel de 20 milliards d'euros.

Les caméras de reconnaissance faciale fonctionnent en créant une représentation mathématique unique de chaque visage, qui peut comporter jusqu'à 500 millions de données.

L'année 2020 marque un tournant pour la Chine avec le déploiement à l'échelle nationale du

Système de crédit social (SCS). Pour appuyer ce système de notation des citoyens, plus de 600 millions de caméras de surveillance sont prévues. À ce jour, environ 400 millions d'entre elles seraient déjà en place, installées dans les espaces publics tels que les rues, les gares, les transports en commun, etc. La plupart de ces dispositifs sont des caméras intelligentes, équipées de la reconnaissance faciale, dont les algorithmes modélisent chaque visage. Certaines caméras sont capables d'identifier les personnes de dos en analysant leur démarche, tandis que d'autres surveillent les comportements, tels que les mouvements brusques ou les changements soudains de température corporelle, pour détecter d'éventuels comportements suspects. Cette dynamique macroéconomique de sécurité publique se décline également au niveau micro-économique avec des investissements accrus des sociétés chinoises dans ce type d'outil en coopération étroites avec les pouvoirs publics. On assiste ainsi à un véritable continuum sécuritaire public/Privé.

L'utilisation de l'intelligence artificielle dans la surveillance vidéo et la reconnaissance faciale présente des avantages considérables, mais **elle comporte également plusieurs limites techniques et défis importants**, tels que :

- **Précision et biais des modèles**

  Biais dans les données d'entraînement : Les systèmes de reconnaissance faciale sont souvent entraînés sur des ensembles de données qui peuvent être biaisés, notamment en termes de genre, d'âge, et d'origine ethnique. Cela peut conduire à des taux d'erreur plus élevés pour certains groupes, ce qui soulève des préoccupations éthiques et pratiques.

  Faux positifs et faux négatifs : Les erreurs de reconnaissance sont courantes, surtout dans des conditions d'éclairage faibles, avec des angles de vue différents, ou en cas de changements dans l'apparence des individus (lunettes, masques, changements de coiffure). Les faux positifs (identification incorrecte) et les faux négatifs (non-reconnaissance) peuvent entraîner des conséquences graves.
- **Environnement et qualité de l'image**

  Résolution et qualité vidéo : Une mauvaise résolution vidéo ou une qualité d'image dégradée due à des conditions climatiques (pluie, brouillard) ou d'éclairage (éblouissement, obscurité) peuvent réduire considérablement la précision des systèmes de reconnaissance faciale.

  Occlusions et postures : Les personnes portant des masques, des casquettes, ou d'autres accessoires qui couvrent partiellement le visage peuvent échapper à la détection ou à une identification correcte. De plus, les angles de prise de vue non frontaux (profil, de dos) compliquent la reconnaissance.
- **Problèmes de scalabilité et de performance**

  Puissance de calculs : Les besoins en puissance de calcul des algorithmes de reconnaissance faciale sont très exigeants et élevés, surtout lorsqu'ils sont utilisés en temps réel sur de grands réseaux de caméras. Cela peut entraîner des retards et des limitations dans le traitement en direct, particulièrement pour des systèmes avec des ressources limitées.

  La gestion des bases de données : Les systèmes de reconnaissance faciale doivent gérer des bases de données massives d'images et de profils. L'indexation, la recherche rapide et la mise à jour de ces bases peuvent être complexes, particulièrement lorsque les données doivent être synchronisées en temps réel avec de multiples caméras.
- **Sécurité et vulnérabilité aux attaques**

  Attaques par spoofing : Les systèmes de reconnaissance faciale peuvent être trompés par des photos, des vidéos ou des masques en 3D. Des techniques comme le "deepfake" rendent ces systèmes vulnérables à des attaques sophistiquées.

  Attaques par exemples contradictoires : Les modèles d'IA peuvent être trompés par des perturbations subtiles (comme des motifs spéciaux imprimés sur des vêtements) qui sont imperceptibles pour l'œil humain mais qui faussent la reconnaissance par les algorithmes.
- **Limitations de l'IA dans la compréhension contextuelle**

  Absence de compréhension contextuelle : L'IA excelle à reconnaître les visages mais manque souvent de compréhension contextuelle, comme différencier une situation suspecte d'une situation normale (ex. : une personne qui court peut être interprétée comme fuyant un crime alors qu'elle fait simplement du jogging).

Problèmes de scénarios complexes : Dans des foules denses ou des environnements complexes, identifier et suivre un individu spécifique devient extrêmement difficile pour les systèmes actuels.

- **L'éthique et la législation**

  Respect de la vie privée et réglementation : Les contraintes légales imposent des limites sur la collecte, le stockage, et l'utilisation des données biométriques. Les réglementations sur la protection des données (comme le RGPD en Europe) imposent des conditions strictes qui doivent être respectées, ce qui peut freiner l'utilisation de ces technologies.

  Ces limitations soulignent la nécessité d'améliorer les algorithmes et de prendre en compte les considérations techniques, éthiques et légales lors de l'utilisation de l'IA dans la surveillance vidéo et la reconnaissance faciale.

- **Détection des Comportements Anormaux**

  Les systèmes d'IA peuvent également analyser les flux vidéo pour détecter des comportements anormaux ou suspects, tels que des mouvements inhabituels, des foules soudaines ou des activités potentiellement dangereuses.

  Prévention des incidents : La détection précoce des comportements anormaux permet de prévenir les incidents avant qu'ils ne se produisent.

  Réduction des risques : Les entreprises peuvent réduire les risques de vols, de vandalisme et d'autres activités criminelles.

  Sécurité des employés : La surveillance proactive améliore la sécurité des employés sur le lieu de travail. Les réseaux de transports publics, tels que les métros et les gares, utilisent l'IA pour détecter des comportements anormaux dans les foules afin d'assurer la sécurité des passagers et de prévenir des incidents.

  <u>Fonctionnement :</u>

  Des caméras de surveillance sont installées dans les stations de métro, sur les quais et à bord des trains. Elles capturent en continu des flux vidéos des mouvements des passagers.

  Analyse en temps réel : Les flux vidéo sont traités en temps réel par des algorithmes d'IA, notamment des modèles de vision par ordinateur et d'apprentissage profond. Ces algorithmes sont entraînés sur des ensembles de données contenant des exemples de comportements normaux et anormaux.

  <u>Détection d'anomalies :</u>

  Comportements suspects : L'IA est capable de détecter des comportements suspects tels que les intrusions sur les voies, les courses soudaines, les bagarres, ou les personnes restant immobiles trop longtemps dans des zones inhabituelles (indiquant potentiellement une activité suspecte comme la dépose de colis abandonnés).

  Atteinte à l'intégrité physique : Les algorithmes peuvent identifier des personnes qui tombent ou qui sont en détresse, ce qui permet une intervention rapide par le personnel de sécurité ou les services médicaux.

  Alertes automatiques : Lorsqu'un comportement anormal est détecté, le système envoie automatiquement une alerte au personnel de sécurité avec des images ou des séquences vidéo pour une vérification rapide et une intervention appropriée.

  Grâce à ces alertes, le personnel de sécurité peut intervenir avant que la situation ne s'aggrave, minimisant ainsi les risques pour les passagers et le fonctionnement des transports.

  <u>Exemple d'un cas d'usage réel :</u>

  Londres (UK) : Le métro londonien utilise l'IA pour détecter des comportements inhabituels, comme des tentatives d'accès non autorisées ou des activités suspectes sur les quais.

  Tokyo (Japon) : Le système de surveillance des gares utilise l'IA pour identifier les chutes sur les voies et envoyer des alertes pour arrêter les trains en cas d'urgence.

- **Gestion dynamique des accès**

L'IA permet également une gestion plus intelligente et sécurisée des accès aux installations. Les systèmes d'accès fondés sur l'IA utilisent des données biométriques et d'autres technologies avancées pour contrôler et surveiller l'entrée des personnes dans des zones sensibles.

Contrôle accru : Les accès non autorisés peuvent être immédiatement détectés et bloqués.

Conformité renforcée : Les entreprises peuvent garantir le respect des protocoles de sécurité et des réglementations.

Expérience utilisateur : Les systèmes d'accès intelligents offrent une expérience plus fluide et sécurisée pour les employés et les visiteurs.

■ I.2.d. La gestion des identités et de l'authentification

L'introduction de l'intelligence artificielle dans la gestion des identités et de l'authentification (IAM) transforme profondément la manière dont les entreprises sécurisent l'accès à leurs systèmes et données. Elle renforce la sécurité des données, améliore l'efficacité opérationnelle, offre une meilleure expérience utilisateur. La gestion des habilitations est souvent une difficulté récurrente pour les entreprises car elles doivent déployer des référentiels continuellement mis à jour.

- **Renforcement de la sécurité des données :**

  Précision et fiabilité : Les systèmes d'authentification fondés sur l'IA, tels que la reconnaissance biométrique (empreintes digitales, reconnaissance faciale), offrent une sécurité plus précise et fiable par rapport aux mots de passe traditionnels.

  Détection des anomalies : L'IA peut identifier des comportements d'accès anormaux en temps réel, empêchant ainsi les accès non autorisés et les violations de sécurité.
- **Les risques de fraude :**

  Multi-Factor Authentication (MFA) : L'IA peut faciliter la mise en œuvre de l'authentification multifactorielle, combinant plusieurs méthodes d'identification pour une sécurité renforcée.

  Protection contre les attaques : Les algorithmes d'IA peuvent détecter et bloquer les tentatives de phishing et autres attaques avant qu'elles n'affectent les systèmes.

Exemple : Authentification MFA avec IA pour une application bancaire

Contexte : Une application bancaire en ligne souhaite renforcer la sécurité de ses utilisateurs en utilisant une MFA basée sur l'IA. L'objectif est de protéger les comptes contre les accès non autorisés tout en maintenant une expérience utilisateur fluide.

Étapes de l'authentification :

1. **Facteur de base - Mot de passe et nom d'utilisateur :** L'utilisateur entre son nom d'utilisateur et son mot de passe comme première étape de l'authentification.
2. **Facteur de possession - Validation par smartphone :** Après l'entrée correcte des identifiants, une demande d'authentification est envoyée à l'application mobile de l'utilisateur pour approbation. Cela peut inclure une notification push où l'utilisateur doit confirmer l'accès.
3. **Facteur biométrique - Reconnaissance faciale :** Pour renforcer encore la sécurité, l'utilisateur est invité à effectuer une reconnaissance faciale *via* la caméra de son smartphone. L'A analyse les traits du visage pour s'assurer que l'utilisateur correspond à la personne autorisée.
4. **Analyse comportementale par IA :** Pendant le processus, une IA analyse le comportement de l'utilisateur, comme la façon dont il tape sur le clavier, la vitesse de frappe, l'endroit où il clique, et même le mouvement du curseur. Ces données sont comparées à un modèle comportemental préétabli de l'utilisateur pour détecter des anomalies.
5. **Analyse de contextuelle de la fraude :** L'IA évalue également le contexte de la connexion, tel que la localisation géographique, l'adresse IP, l'heure de la tentative de connexion, et l'historique des connexions de l'utilisateur. Si une anomalie est détectée (par exemple, une tentative de connexion depuis un pays inhabituel), l'IA peut bloquer l'accès ou demander des vérifications supplémentaires.

6. **La sécurité dynamique :** En cas de doute, l'IA peut générer des questions de sécurité dynamiques basées sur des informations récentes (ex. : achats récents) pour valider l'identité de l'utilisateur.
   Conclusion : Grâce à cette approche, l'application bancaire assure une sécurité renforcée tout en offrant une expérience utilisateur plus fluide et personnalisée. Les tentatives d'accès suspectes sont bloquées ou nécessitent des vérifications supplémentaires. Ce type de MFA intelligent pourrait être particulièrement efficace car il utilise l'IA pour adapter les mesures de sécurité en fonction du comportement et du contexte de chaque utilisateur, augmentant ainsi la robustesse de l'authentification tout en limitant les frictions inutiles pour les utilisateurs réguliers.

● **Efficacité opérationnelle**

Habilitation automatisée : L'IA peut automatiser l'octroi et la révocation des accès en fonction des rôles et des besoins des employés. Interconnectée à l'Active Directory de l'entreprise, l'automatisation de la gestion des habilitations permet aux départements RH et IT de s'affranchir d'une action de gestion sans pour autant s'affranchir d'une action de contrôle.

La gestion du support IT : L'IA intervient comme étant un ingénieur Support en local sur le poste du collaborateur. L'objectif est de diminuer le degré de sollicitation des équipes Support par les collaborateurs en leur mettant à disposition un agent IA les accompagnant sur la bonne maîtrise de leurs environnements de travail.

Optimisation des ressources : Le dimensionnement des ressources utiles au bon fonctionnement des services de l'entreprise est un élément critique pour garantir la bonne disponibilité et la fluidité de ces derniers. La fluctuation saisonnière, les évolutions de taille des équipes et le volume de requêtes auprès de ces services contraignent les équipes techniques à manuellement suivre les besoins en ressources matérielles (Virtual Machines) ou logiciel (Licences). L'intelligence artificielle permet d'analyser les besoins et de les quantifier selon des pas de temps définis. En cas de fluctuation significative, elle pourrait permettre d'allouer des ressources complémentaires et les désactiver au bon moment.

Exemple : Centre Opérationnel de Sécurité (SOC) avec IA pour la Détection et la Réponse aux Menaces

Contexte : Une grande entreprise dispose d'un Centre Opérationnel de Sécurité (SOC) chargé de surveiller et de répondre aux incidents de cybersécurité. En raison de l'augmentation des cyberattaques et du volume des alertes, l'équipe de sécurité rencontre des difficultés à analyser rapidement toutes les menaces potentielles.

Mise en œuvre opérationnelle :

1. **Surveillance automatisée des menaces :** Le SOC utilise une plateforme d'IA qui surveille en temps réel l'ensemble du réseau, les endpoints, et les flux de données pour détecter les comportements suspects. L'A analyse les logs, les paquets de données et les événements "système" pour identifier les anomalies.
2. **Détection des menaces :** La plateforme détecte non seulement les menaces connues mais aussi des attaques de type « zero-day » et les comportements anormaux qui échappent aux systèmes de détection manuels.

3. **Priorisation automatique des alertes :** Une fois qu'une menace potentielle est détectée, l'IA classe et priorise les alertes en fonction de leur gravité et de leur impact potentiel sur l'entreprise.

4. **Réponse automatisée :** Pour certaines menaces courantes, l'IA est programmée pour répondre automatiquement. Si un malware est détecté sur un endpoint, l'IA peut déconnecter la machine du réseau principal et opérer une analyse approfondie.

5. **Analyse et réduction des faux positifs :** L'IA apprend des décisions des analystes et ajuste en permanence ses algorithmes pour réduire les faux positifs, diminuant ainsi la charge de travail des équipes de sécurité et augmentant la précision des alertes.

6. **Rapports et analyses prédictives :** L'IA génère des rapports automatiques sur les tendances des menaces et les points faibles identifiés dans l'infrastructure de l'entreprise. Elle utilise également des techniques de prévision pour anticiper les futures menaces et proposer des actions préventives.
   Résultat :
   Grâce à l'intégration de l'IA, le SOC de l'entreprise a réduit de manière significative le temps moyen de détection et de réponse aux incidents (MTTD et MTTR). L'automatisation des tâches répétitives et l'analyse en temps réel des données permettent aux analystes de se concentrer sur les menaces critiques, augmentant ainsi l'efficacité opérationnelle. En conséquence, le SOC est capable de traiter plus d'incidents par jour avec moins de ressources humaines, améliorant globalement la sécurité de l'entreprise.

● Conformité et Réglementation

- **Audit et reporting :** Les systèmes IAM basés sur l'IA offrent des capacités avancées de suivi et de reporting, facilitant les audits de conformité.
- **Gestion des politiques de sécurité :** L'IA aide à assurer que les politiques de sécurité et les accès sont conformes aux réglementations en vigueur (RGPD, HIPAA,[26]), etc.
- **Protection des Données Sensibles :** Une gestion efficace des accès réduit les risques de violations de données, minimisant ainsi les conséquences juridiques et financières.
- **Transparence et traçabilité :** Les entreprises peuvent démontrer une gestion rigoureuse des accès, renforçant la confiance des régulateurs et des clients.
  La gestion de la conformité aux réglementations en matière de protection des données, en automatisant les processus de surveillance et de reporting se voit simplifiée et confère aux Directions une lecture éclairée de leur exposition juridique. Cette conformité accrue minimise les risques juridiques et renforce la confiance des clients et des partenaires. En outre, les entreprises qui adoptent l'IA dans leurs systèmes IAM se positionnent comme des leaders technologiques, attirant ainsi des talents et des investisseurs, et consolidant leur avantage concurrentiel.

● Collaboration et Expérience Utilisateur

Accès simplifié : Les méthodes d'authentification basées sur l'IA, telles que la reconnaissance faciale, offrent une expérience utilisateur fluide et sans friction.
Personnalisation : L'IA peut adapter les niveaux d'accès et les contrôles de sécurité en fonction des habitudes et des besoins spécifiques de chaque employé.
Optimisation des ressources : Les employés passent moins de temps à gérer leurs identifiants et peuvent accéder plus rapidement aux ressources nécessaires.
Continuité des Opérations : La réduction des incidents de sécurité et des interruptions liées aux problèmes d'accès contribue à une productivité accrue.
Partage sécurisé des ressources : Une gestion efficace des identités et des accès permet une collaboration sécurisée avec les partenaires et les clients.
Flexibilité : Les employés peuvent accéder en toute sécurité aux ressources de l'entreprise depuis n'importe quel endroit, favorisant ainsi le travail à distance et la flexibilité.

L'IA impose une nouvelle gestion des identités et de l'authentification en entreprise, nécessitant des solutions innovantes pour une sécurité robuste, une efficacité opérationnelle optimisée, et une expérience utilisateur améliorée. L'adoption de l'IA dans les systèmes IAM devient une nécessité pour toute entreprise cherchant à se protéger contre les menaces modernes, à rester conforme aux exigences réglementaires, et à maintenir une position de leader dans un environnement commercial de plus en plus compétitif.

[26] RGPD : Règlement Général de la Protection des Données - HIPAA : Health Insurance Portability and Accountability Act

■ I.2.e. Quels Impacts transversaux au sein de l'entreprise ?

L'intelligence artificielle transforme la surveillance et la sécurité physique en offrant des solutions avancées et efficaces. En intégrant des technologies telles que la reconnaissance faciale, l'analyse vidéo, la gestion intelligente des accès, les entreprises peuvent non seulement améliorer leur sécurité mais aussi bénéficier de nombreux avantages transverses.

Le déploiement des technologies d'intelligence artificielle au sein des entreprises représente une révolution sans précédent, offrant une multitude d'opportunités qui transforment profondément les opérations et les stratégies organisationnelles. L'IA permet aux entreprises d'optimiser leurs processus, d'améliorer la prise de décision, d'augmenter l'efficacité opérationnelle et de renforcer la sécurité à tous les niveaux.

	Direction Générale	IT	R&D	Finance	Marketing	Ventes	RH	Production	Logistique	Service Client
Direction Générale	-	Supervision	Stratégie d'innovation	Budget & ROI	Alignement stratégique	Objectifs de vente	Gestion des talents	Innovation des processus	Optimisation logistique	Stratégie de satisfaction
IT	Support stratégique	-	Innovation	Gestion des risques	Analyses de marché	Outils de vente	Systèmes RH	Automatisation	Traçabilité	Outils de CRM
R&D	Innovation stratégique	Dev. Outils Métiers	-	Budgétisation de projets	Insights produit	Stratégie de vente	Formation continue	Procédés de fabrication	Optimisation logistique	Avanatage compétitif
Finance	Planification financière	Budget des technologies	Financement R&D	-	Coûts des campagnes	Analyse des ventes	Budgétisation RH	Optimisation de la VA	Coûts logistiques	Coûts de service
Marketing	Alignement stratégique	Analyses de données	Innovations produit	Budgets campagnes	-	Stratégie de vente	Formation Comm.	Marketing produit	Satisfaction Client	Analyses de satisfaction
Ventes	Objectifs de vente	Outils de vente	Innovations produit	Analyse des ventes	Stratégie de marché	-	Formation des vendeurs	Besoins de production	Suivi des livraisons	Satisfaction et retours
Ressources Humaines	Gestion des talents	Systèmes RH	Formation continue	Budget RH	Formation Comm.	Formation commerciale	-	Formation	Optim. ressources	Formation
Production	Innovations Prod.	Automatisation	Procédés de fabrication	ROI des technologies	Publicité des produits	Besoins de production	Formation	-	Coordination logistique	Amélioration continue
Logistique	Optimisation logistique	Traçabilité	Optim. SC	Coûts logistiques	Satisfaction Client	Traçabilité	Planification Optimisée	Coordination logistique	-	Gestion des retours
Service Client	Stratégie de satisfaction	Outils de CRM	Avantage compétitif	Coûts de service	Analyses de satisfaction	Satisfaction et retours	Formation	Amélioration continue	Gestion des retours	-

Matrice des avantages transverses grâce au déploiement de l'iA [27]

Cette matrice met en évidence les interactions et les collaborations essentielles entre les différents départements dans le cadre du déploiement des technologies d'intelligence artificielle au sein d'une entreprise. Ces liens montrent la manière dont chaque département contribue et bénéficie des initiatives d'IA, renforçant ainsi l'efficacité globale et la compétitivité de l'entreprise.

Les applications de l'IA touchent tous les départements, de la Direction Générale à la Supply Chain, en passant par les Finances, le Marketing, les Ventes, les Ressources Humaines, la Production et le Service Client. En intégrant l'IA, les entreprises peuvent non seulement automatiser les tâches répétitives et réduire les coûts, mais aussi obtenir des informations précieuses grâce à l'analyse de grandes quantités de données, favorisant ainsi l'innovation et l'adaptabilité dans des contextes de marché mouvants.

[27] Matrice des avantages transverses de l'IA au sein des entreprises - Cédric Jutteau 07/2024

Cependant, le succès du déploiement de l'IA dépend de la capacité des entreprises à créer une collaboration interfonctionnelle. Les liens entre les différents départements doivent être cartographiés pour maximiser l'impact et éviter le gaspillage de ressources financières et techniques. Chaque département doit jouer son rôle, en s'assurant que les innovations et les solutions soient alignées avec les objectifs stratégiques globaux de l'entreprise.

En conclusion, les opportunités offertes par l'IA sont riches et prometteuses, mais leur réalisation nécessite **une approche intégrée et coordonnée**. Les entreprises qui réussissent à tirer parti de l'IA de manière holistique seront mieux positionnées pour rester compétitives afin d'adresser au marché une position de valeur supérieure à celles de compétiteurs. L'IA devient un catalyseur essentiel de la transformation numérique et de la croissance durable des entreprises mais impose des règles de déploiement strictes et rigoureuses.

● PARTIE II. INTELLIGENCE ARTIFICIELLE ET SÛRETÉ DES ENTREPRISES « UNE ARCHITECTURE À LA COMPLEXITÉ BYZANTINE »[28]

S'il y a fort à parier que les dispositifs d'intelligence artificielle intégreront à terme, l'ensemble de l'écosystème des entreprises y compris le volet sûreté, le chemin pour y arriver apparaît complexe tant au regard des limites (A) que des menaces (B) liées au déploiement de cette technologie en entreprise. Les développements qui suivent ont vocation à mettre en lumière certaines de ces limites et de ces menaces, lesquelles ont été préalablement choisies au regard de leur intérêt par rapport à l'objet de cette étude.

○ II.1. Des limites intrinsèques au déploiement de l'intelligence artificielle au sein des entreprises

■ II.1.a. La gouvernance des données à l'épreuve de l'inflation normative

Une fois les opportunités des dispositifs d'intelligence artificielle clairement identifiées au sein d'une entreprise (capacité de calcul, finesse d'analyse) force est de constater qu'il faudra que les dirigeants

[28] Y. PADOVA. IA et si l'Europe se trompait de régulation ? Leséchos.fr, Avril 2024.

puissent arriver à déployer un dispositif qui satisfasse au cadre juridique en vigueur ce qui constitue en soi, un véritable défi en termes de **gouvernance de la donnée**.

DMA[32], DSA[33], RGPD[34], Data Act[29], DORA[30], NIS 2[31] et bien entendu **EU-AI Act[32],** autant d'acronymes qui mettent en lumière l'inflation normative qui entoure déjà le déploiement de dispositifs d'intelligence artificielle. Ainsi, même si ces textes ne sont pas, à date, tous entrés en application, le cadre juridique qu'ils prévoient doit être connu pour pouvoir anticiper au mieux l'IA de demain.

En effet, cela fait maintenant plusieurs années que tant la France que l'Europe ont souhaité se doter d'un cadre ambitieux et uniforme en matière d'intelligence artificielle afin de s'affirmer comme leaders sur mondial sur ce marché tout en tenant compte des implications humaines et éthiques de l'IA.

Si le 21 mai 2024, les États membres de l'Union Européenne ont, à l'unanimité, validé la proposition de règlement sur l'intelligence artificielle, cela fait plusieurs années que l'Europe affiche clairement ses ambitions. Ainsi dès 2019, les sénateurs André GATTOLIN, Claude KERN, Cyril PELLEVAT et Pierre OUZOULIAS soulignaient que le déploiement de l'IA devait être considéré comme « *la principale innovation d'une nouvelle révolution industrielle*, « *celle du travail de l'homme avec des machines dites intelligentes*[33] ».

Consciente du bouleversement des équilibres en présence et de la nécessité de sécuriser le développement de l'IA, sa commercialisation ainsi que son utilisation, il a fallu trouver un consensus, au niveau européen, pour élaborer une réglementation dédiée à ces dispositifs qui soit à la fois homogène et ambitieuse :

32 RÈGLEMENT (UE) 2022/1925 DU PARLEMENT EUROPÉEN ET DU CONSEIL du 14 septembre 2022 relatif aux marchés contestables et équitables dans le secteur numérique et modifiant les directives (UE) 2019/1937 et (UE) 2020/1828 (règlement sur les marchés numériques) (Digital Markets Act)

33 RÈGLEMENT (UE) 2022/1925 DU PARLEMENT EUROPÉEN ET DU CONSEIL du 14 septembre 2022 relatif aux marchés contestables et équitables dans le secteur numérique et modifiant les directives (UE) 2019/1937 et (UE) 2020/1828 (règlement sur les marchés numériques) (Digital Services Act)

34 REGLEMENT(UE) 2016/679 DU PARLEMENT EUROPEEN ET DU CONSEIL du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données).

29 RÈGLEMENT (UE) 2023/2854 DU PARLEMENT EUROPÉEN ET DU CONSEIL du 13 décembre 2023 concernant des règles harmonisées portant sur l'équité de l'accès aux données et de l'utilisation des données et modifiant le règlement (UE) 2017/2394 et la directive (UE) 2020/1828 (règlement sur les données)

30 RÈGLEMENT (UE) 2022/2554 DU PARLEMENT EUROPÉEN ET DU CONSEIL du 14 décembre 2022 o
1060/2009, (UE) n o sur la résilience opérationnelle numérique du secteur financier et modifiant les règlements (CE) n 648/2012, (UE) n o 600/2014, (UE) n o 909/2014 et (UE) 2016/1011

31 DIRECTIVE (UE) 2022/2555 DU PARLEMENT EUROPEEN ET DU CONSEIL du 14 décembre 2022 concernant des mesures destinées à assurer un niveau élevé commun de cybersécurité dans l'ensemble de l'Union, modifiant le règlement (UE) no 910/2014 et la directive (UE) 2018/1972, et abrogeant la directive (UE) 2016/1148 (directive SRI 2)

32 RÈGLEMENT DU PARLEMENT EUROPÉEN ET DU CONSEIL du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle (Législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union

33 Rapport d'information fait au nom de la commission des affaires européennes sur la stratégie européenne pour l'intelligence artificielle, Enregistré à la Présidence du Sénat le 31 janvier 2019.

- **Un cadre homogène** : introduire un ensemble proportionné et efficace de règles contraignantes pour les systèmes d'IA avec une approche qui devrait adapter le type et le contenu de ces règles à l'intensité et à la portée des risques que les systèmes d'IA peuvent générer.

- **Une ambition** : faire de l'Union Européenne un acteur mondial de premier plan dans le développement d'une intelligence artificielle sûre, fiable et éthique[34]. A cet égard, le cadre juridique en cours d'élaboration promeut une approche centrée sur l'humain et axée sur le risque[35].

  Or, si de nombreux débats existent autour des dispositifs d'intelligence artificielle, des définitions à retenir, deux éléments font consensus pour qualifier un dispositif d'IA : **un algorithme et des données.** A partir de ces éléments, l'infographie réalisée ci-dessous présente ce que sont les besoins de l'IA.

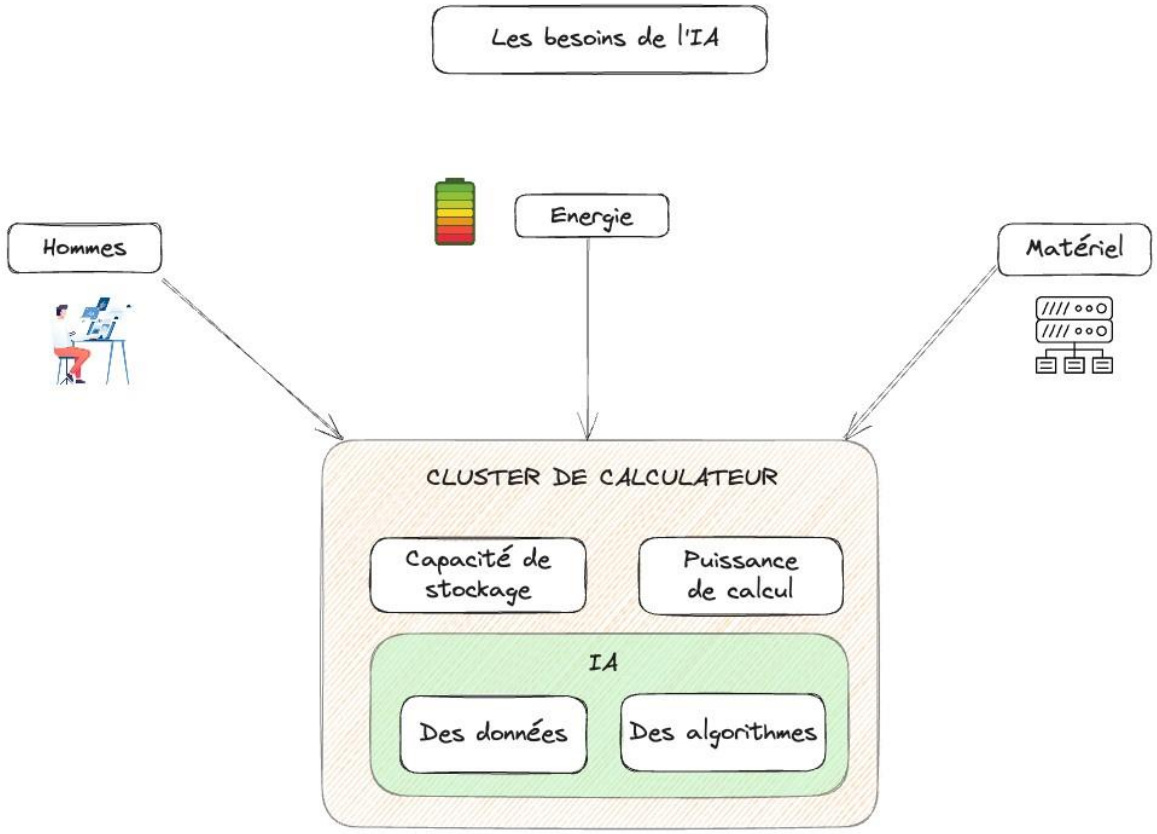


Figure 1 : Infographie SIA - Frédéric Loisel

Ces éléments contextuels rappelés, soulignons que la plupart des textes adoptés depuis de nombreuses années en matière de données se veulent **agnostiques de la technologie** sur laquelle ils reposent et ce, afin de pouvoir durer dans le temps.

En ce sens, **la régulation en matière d'intelligence artificielle rompt avec cette logique** agnostique en tant qu'elle a justement vocation à s'appliquer à « *un large champ de technologies* » dont l'évolution rapide est largement reconnue[42].

[34] Ibid. 30, Considérant 6.
[35] Actu IA. L'impact de l'AI Act : un chemin semé d'embûches pour les forces de l'ordre ? Avril-Juin 2024.

En pratique, l'émergence de cette régulation européenne spécifique impliquera ainsi qu'elle puisse se combiner avec les différentes réglementations précitées et s'articuler avec leurs champs d'applications respectifs.

Prenons à titre d'exemple un dispositif vidéo auquel serait intégré une surcouche logicielle empreinte d'IA et ce, afin de reconnaître des comportements suspects aux abords d'une entreprise. Difficile d'imaginer ne pas appliquer le RGPD en ce que le dispositif permettra la collecte de données personnelles (images), l'application de l'EU-IA act, NIS2 en fonction de la nature de l'opérateur concerné et ce, sans oublier la législation éventuellement propre à chaque pays en matière de dispositif vidéo, telle que pour la France, l'application des dispositions du Code de la sécurité intérieure (CSI).

En outre, les entités concernées, au travers de leurs opérateurs de sûreté, devront également rester attentives aux contours du dispositif d'IA auxquelles elles recourent afin de s'assurer que **ce dispositif qui peut être qualifié d'IA aujourd'hui, le soit également demain.**

Si l'on fait le parallèle avec la notion de données à caractère personnel, force est de constater que malgré une définition juridique stricte[36], l'appréciation des contours de cette notion suscite des débats dans certains cas, lesquelles conduisent régulièrement à des précisions jurisprudentielles [37] (ex. interprétation de la donnée de santé).

Or, la définition d'un système d'intelligence artificielle a elle-même suscité **des débats** avant que ne soit retenue, dans l'EU-IA Act[38], la définition telle que proposée par l'Organisation de coopération et de développement économiques (OCDE). Il résulte de cette définition et des principes qui l'entourent[46] qu'un système d'IA serait **un système capable d'influencer son environnement en produisant des résultats pour répondre à un ensemble donné** d'objectifs à partir de données et ce, au travers d'un cycle de vie spécifique (voir les schémas ci-dessous).

[42] Proposition de RÈGLEMENT DU PARLEMENT EUROPÉEN ET DU CONSEIL ÉTABLISSANT DES RÈGLES HARMONISÉES CONCERNANT L'INTELLIGENCE ARTIFICIELLE (LÉGISLATION SUR L'INTELLIGENCE ARTIFICIELLE) ET MODIFIANT CERTAINS ACTES LÉGISLATIFS DE L'UNION,
Exposé des motifs.

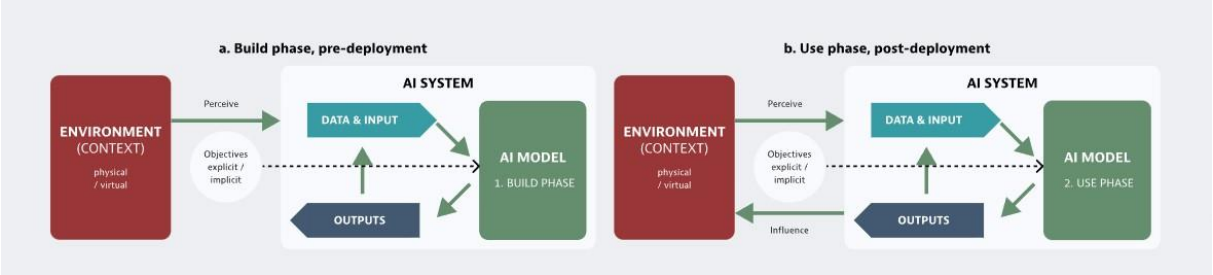


Figure 2 : Système d'IA[39]

[36] Ibid. 26, Article 4.1.

[37] Voir par exemple l'arrêt Lindqvist, CJCE, 6 novembre 2003, n° C-101/01, §50.

[38] Aux termes de l'article 3, un « système d'intelligence artificielle » (système d'IA) est définit comme « *un logiciel qui est développé au moyen d'une ou plusieurs des techniques et approches énumérées à l'annexe I et qui peut, pour un ensemble donné d'objectifs définis par l'homme, générer des résultats tels que des contenus, des prédictions, des recommandations ou des décisions influençant les environnements avec lesquels il interagit* ». [46] OECD, OECD/LEGAL/0449, Recommendation of the Council on Artificial Intelligence (Amended on 03/05/2024).

[39] OECD, https://oecd.ai/fr/ai-principles, version au 7 juillet 2024. [48] Ibid. 39.

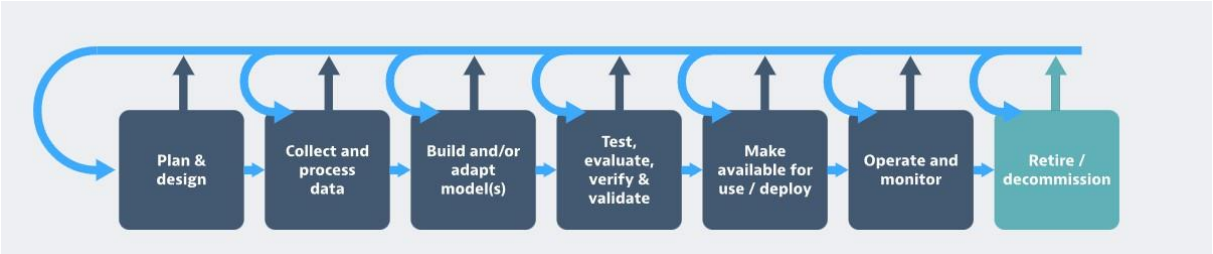


Figure 3 : Cycle de vie d'un système d'IA[48]

A partir de cette définition, cinq niveaux spécifiques ont été déterminés : **les IA prohibées en tant que le risque engendré serait inacceptable, à haut risque, à usage général, celles qui nécessitent des obligations de transparence renforcées et celles qui présentent un risque minimal.**

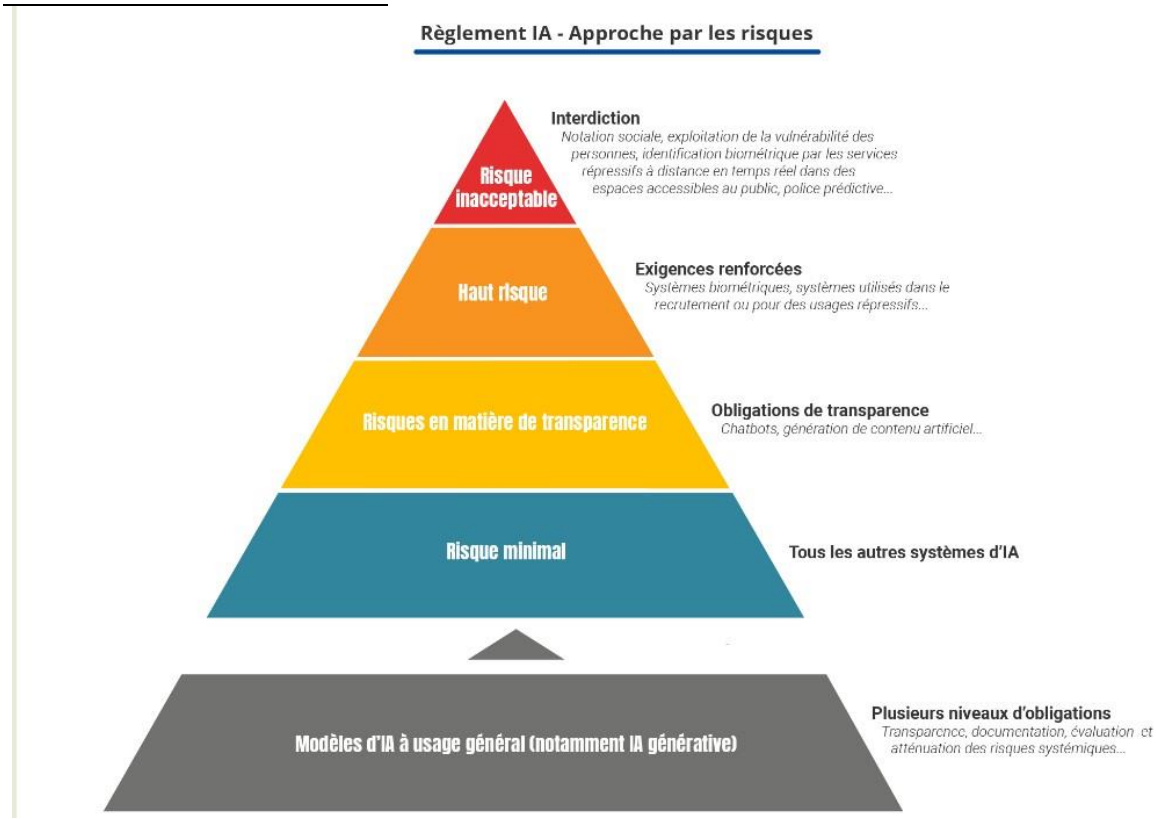


Figure 4 : RIA - Approche par les risques[40]

Sur le principe, la définition proposée se veut unique et à l'épreuve du temps. Pour autant, force est de constater que l'approche retenue :

- Est susceptible de poser des difficultés réelles aux opérateurs qui devront jongler entre les différents pans de la réglementation afin de les comprendre et de les articuler

[40] CNIL, Entrée en vigueur du règlement européen sur l'IA : les premières questions-réponses de la CNIL, www.cnil.fr, publiée le 12 juillet 2024.

- Devront opérer une veille régulière des dispositions de l'annexe III de l'EU-AI Act pour s'assurer de la qualification de leur système et s'assurer des obligations qui leurs incombent

Si la régulation adoptée se veut pionnière en son domaine, elle se heurtera également à des questions juridiques plus traditionnelles parmi lesquelles :

- **La possibilité d'expérimenter en conditions réelles** : plusieurs acteurs interrogés soulignent que si l'EU-AI Act constitue une avancée fondamentale, sa mise en application concrète impliquerait que les acteurs concernés puissent tester les dispositifs en question en conditions réelles et bénéficier ainsi **d'un cadre juridique dédié à l'expérimentation.** Or, la mise en œuvre d'un tel cadre - en matière de traitements de données - est l'objet de controverses depuis de nombreuses années.

- **La multiplication des acteurs en charge de la régulation tant au niveau européen que national :** la création d'un Comité européen de l'IA[50], le Bureau de l'IA, le forum consultatif, le groupe scientifique d'experts indépendants, les autorités nationales, les autorités chargées d'endosser le rôle d'autorité de surveillance du marché.

Dans une perspective de réflexion plus large, il est intéressant de relever que le constat d'un cadre juridique incomplet en matière de système d'intelligence artificielle (SIA) a également été mis en lumière par la Commission Nationale de Contrôle des Techniques de Renseignement (CNCTR) dans son dernier rapport d'activité[51] quant au périmètre qui la concerne. Edité quelques semaines avant la publication du règlement sur l'intelligence artificielle, il est vraisemblable que la CNCTR y voit un progrès indéniable sans toutefois qu'elle ne puisse trouver des réponses à l'ensemble des questions qui sont les siennes compte tenu des activités qu'elle contrôle.

■ II.1.b. Un régime de responsabilité dont les contours restent à éprouver

A l'instar des réflexions qui ont occupé il y a quelques années les praticiens du droit en matière de partage de responsabilité dans le cadre des prestations de Cloud Computing, ces mêmes réflexions s'emparent aujourd'hui de la sphère de l'intelligence artificielle.

En effet, entre la conception, le développement, l'entraînement, la mise sur le marché, l'utilisation du système d'intelligence artificielle, de nombreux acteurs seront amenés à intervenir. L'existence de cette relation multipartite implique que l'on puisse clairement déterminer la qualification juridique des parties afin de pouvoir se prononcer sur la question de l'imputabilité de la responsabilité en cas de défaillance du système.

En adoptant une approche fondée sur les risques et en rappelant que le fournisseur d'une IA doit être en mesure d'assumer la responsabilité de la mise sur le marché ou de la mise en service d'un système d'IA indépendamment du fait que cette même personne soit celle qui ait conçu ou développé le système, la réglementation proposée vise à réglementer à la fois la manière dont les fournisseurs devront développer leurs systèmes pour minimiser les risques mais également la manière dont les déployeurs de ces systèmes devront les contrôler.

La construction d'un tel régime de responsabilité est susceptible de rappeler celui qui existe tant en matière de sécurité des produits[41] au niveau de l'Union européenne que le régime de responsabilité du fait des produits défectueux en droit français[53].

[41] RÈGLEMENT (UE) 2023/988 DU PARLEMENT EUROPÉEN ET DU CONSEIL du 10 mai 2023 relatif à la sécurité générale des produits, modifiant le règlement (UE) n o 1025/2012 du Parlement européen et du Conseil et la directive (UE) 2020/1828 du

[50] Ibid. 30, articles 65 et 66.

[51] CNCTR, 8ème Rapport d'activité 2023, publiée le 27 juin 2024, https://www.cnctr.fr/actualites/rapportannuel-2023.

Si l'application du principe de précaution apparaît tout à fait opportune en tant qu'il est indispensable d'imposer des exigences fortes sur les normes de sécurité qui doivent entourer les dispositifs d'IA, elle pourrait se heurter à plusieurs difficultés. En effet, à la différence de produits classiques, un dispositif d'IA :

- Reste par nature une « boite noire », impossible à cet égard de pouvoir prétendre en avoir la connaissance et la maîtrise complète
- Est dynamique et non statique en ce sens qu'il est susceptible d'évoluer en fonction des données, des utilisations, des intégrations, etc. ● Est polyvalent en tant qu'il peut être appliqué à des environnements/contextes différents et produire des conséquences distinctes
- Soulève par nature des questions quant au régime juridique à retenir pour définir son encadrement (pour un exemple en matière d'IA et de protection des données, se référer à la décision du Conseil d'Etat n° 451653 du 22 juillet 2022).

A cela s'ajoute le fait que la qualification des parties retenue devra être articulée avec d'autres qualifications juridiques telles que le régime de sous-traitance ou de responsabilité conjointe établi par le RGPD, en particulier dans le domaine de la sûreté. En effet, le recours à des soustraitants y est largement utilisé pour assurer certaines prestations (gardiennage, télésurveillance, etc.).

Dans ce contexte complexe par nature, il est vraisemblable qu'une entreprise doive au même titre qu'elle est supposée établir sa cartographie de traitement, établir la cartographie des acteurs auxquels elle recourt ainsi que la cartographie des régimes juridiques auxquels elle est soumise en fonction des conditions de déploiement de son système d'IA.

La particularité du cycle de vie en matière d'IA et le régime de responsabilité en cascade établi laisse à penser qu'il ne sera pas si simple d'effectuer un partage de responsabilité entre les différentes parties prenantes et que la jurisprudence établie par les juges du fond aura un rôle central à jouer. Or, s'il est toujours heureux que la jurisprudence puisse venir éclairer des concepts établis, la « nouveauté » de la matière est dans un premier temps susceptible d'être source d'insécurité juridique et par voie d'incidence, de porter atteinte aux droits et libertés des individus que la réglementation entend pourtant protéger.

■ II.1.c. Vers une « Intelligence humaine augmentée » en matière de sûreté des entreprises ?

Comme il l'a été rappelé, la réglementation en matière d'IA promeut des dispositifs qui se veulent sûrs, fiables et éthiques. L'un des enjeux principaux vise ainsi à assurer une intégration sûre du système d'IA dans la machine/ le système de façon à ne pas compromettre la sécurité de la machine dans son ensemble et par là même, la sûreté de l'entreprise.

Le corollaire de cette ambition clairement affichée est qu'il n'est pas possible, aujourd'hui, de mettre en œuvre des dispositifs qui ne reposent sur aucun contrôle humain. Il y a dès lors lieu de remettre en cause l'idée d'une substitution de l'homme par la machine, le cadre juridique en vigueur ayant mis des verrous au développement d'une telle situation en particulier lorsque les systèmes ont vocation à interagir avec des individus au travers par exemple des obligations de transparence à respecter[42], de

Parlement européen et du Conseil, et abrogeant la directive 2001/95/CE du Parlement européen et du Conseil et la directive 87/357/CEE du Conseil [53] Code civil, article 1245.

[42] 54 Ibid. 30, Article 50. [55] Ibid. 26, Article 22.

promouvoir la confiance en ces dispositifs (éthique) ou encore l'interdiction par principe d'avoir recours à des décisions entièrement automatisées[55].

Il y a en effet lieu de rappeler que l'IA ne va faire que proposer des hypothèses à partir des données sur lesquelles elle a travaillé et que la confirmation d'un opérateur demeure essentielle. L'IA permettrait ainsi à l'opérateur d'entrer davantage **dans un rôle de superviseur**.

Appliquée à un environnement sûreté, cela conduirait à s'orienter davantage vers :

- **Une IA spécialisée** qui applique des algorithmes conçus par l'homme pour accomplir des tâches de plus en plus complexes en des temps de plus en plus courts dans des domaines précis et définis (analyse des incidents, levée de doute, détection de signaux faibles[43], de « vagues[44] » de risques, etc.).
- **Une IA permettant d'assister l'individu dans la prise de décision** en lui permettant de se consacrer à ce qui est important **et de superviser** le dispositif. En effet, en soi un dispositif d'intelligence artificielle n'est pas « intelligent ».
- **Un dispositif qui permettrait de supprimer des mesures ou des dispositifs mis en œuvre ou d'améliorer leur capacité de détection.** Il serait ainsi possible d'imaginer l'élaboration de matrices de rondes par un dispositif d'IA permettant par exemple la détection d'intrus au sein d'un périmètre donné (amélioration des capacités de détection au sein d'un environnement donné).

■ II.1.d. Des dispositifs d'IA persuasifs et hallucinatoires ?

Au-delà des limites réglementaires, il apparaissait pertinent d'évoquer les limites liées à la fiabilité même du dispositif.

Il y a lieu tout d'abord de partir du constat selon lequel le marché de l'IA, aujourd'hui, n'est pas un marché « spécifique » au sens où il n'existe pas d'IA réservée au domaine de la sûreté. Il en résulte qu'en matière de sûreté, une entreprise qui déploie un dispositif d'IA serait susceptible de compter sur une « interprétation de données » sans que le modèle initial n'ait été développé puis conçu pour que les données collectées aient du sens dans un contexte de sûreté globale.

Par ailleurs, à la suite de la publication, le 12 juillet dernier, du règlement sur l'intelligence artificielle, la CNIL a eu l'occasion de rappeler[45] que pour ce qui est des modèles génératifs, ces derniers obéissant à une logique probabiliste, des résultats inexacts peuvent être générés alors même qu'ils apparaissent plausibles (« hallucinations »).

Il conviendrait alors que les entreprises aient connaissance des possibilités techniques visant à réduire ce système d'hallucinations, qui reposent eux même sur l'emploi de techniques d'intelligence artificielle… Il en va ainsi du RAG[46] (Retrieval-Augmented Generation) qui peut être défini comme une technique avancée qui combine deux approches principales : la récupération d'informations (retrieval) et la génération de texte (génération). Le RAG permet ainsi d'interroger des bases de données de connaissance et de les injecter dans le prompt de la question qui pourrait être posé par un opérateur pour obtenir des réponses pertinentes sans entraînement du LLM. **Ce RAG, pourrait alors se référer à**

[43] Par exemple en détectant un comportement suspect au travers d'une analyse de logs.

[44] Définie par l'une des personnes interviewées comme la possibilité « *de se heurter aux risques de manière régulière à des fréquences variées afin d'identifier d'éventuelles failles* ».

[45] CNIL, Les questions-réponses de la CNIL sur l'utilisation d'un système d'IA générative, www.cnil.fr, publication au 18 juillet 2024.

[46] Actu IA. Tirer partie de l'IA en entreprise avec le RAG, c'est facile. Avril-Juin 2024. [60] R. SCHIESSL, Intelligence artificielle : la tentation de la dépendance, juin 2023.

une base de connaissance spécifiquement définie pour être appliquée au domaine de la sûreté et en cas d'absence de réponse pertinente, le dispositif ne donnerait tout simplement pas de réponse (très forte diminution du mécanisme des hallucinations).

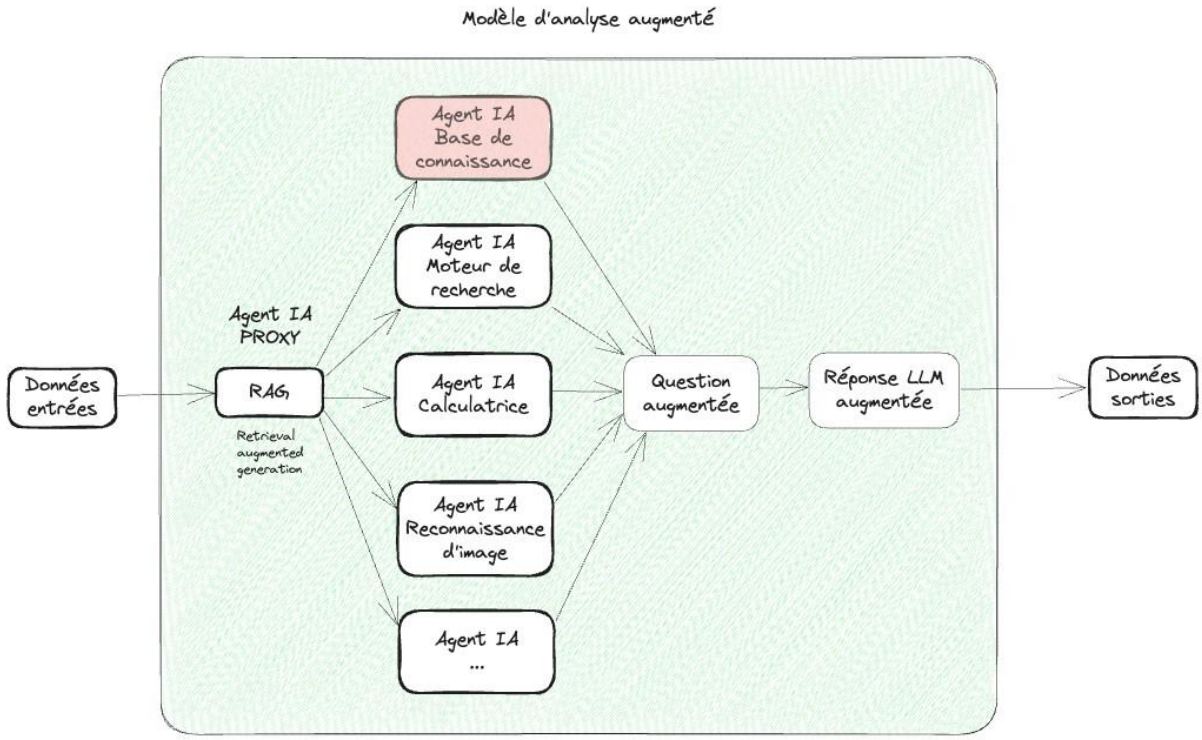


Figure 5 : Modèle d'analyse augmentée - Frédéric Loisel

○ II.2. Les menaces résultant du déploiement de l'intelligence artificielle au sein des entreprises

L'avènement d'une nouvelle ère du numérique portée notamment par des avancées fulgurantes dans le domaine de l'intelligence artificielle amène indéniablement à se poser la question de notre dépendance vis-à-vis de ces dispositifs. En effet, comme le soulignait Rupert
Schiessl dès 2023[60], cofondateur de VERTEEGO (plateforme d'intelligence artificielle), « *L'IA a fait son entrée dans notre quotidien. C'est une technologie omniprésente avec une dépendance grandissante* ».

A l'heure d'une course folle qui semble se mener pour faire partie des premiers acteurs à pénétrer le marché de l'IA et en retirer ainsi les bénéfices économiques attendus, l'IA reste en premier lieu sa propre menace au travers des risques de cybersécurité inhérents qu'elle induit.
Le vol de données sensibles et les actes de sabotage sont à l'origine de pertes financières pour les organisations qui se chiffrent en centaines voire en milliers de milliards d'euros à travers le monde, sans compter les pertes immatérielles telles que la crédibilité ou l'image de marque. Cet état de fait, rend nécessaire d'investir dans des technologies de protection parmi lesquelles figurent l'IA. Étant entendu, que ces technologies peuvent servir, comme souvent, à la fois des buts défensifs mais également servir d'armes aux mains des « hackers » informatiques.

■ II.2.a. IA et cyberattaque : une menace intrinsèque ?

Assurer la cybersécurité de sa structure concerne tous les secteurs régaliens (défense, sécurité intérieure, administrations publiques), les infrastructures vitales (énergie, transport, santé, etc.), le secteur privé (industries, banques, commerces, services, etc.) et les populations. La menace peut toucher n'importe quelle entité ou entreprise, quel que soit son secteur d'activité ou sa taille. **La menace est permanente, diffuse et diverse** : vol, écoute, brouillage, abus et usurpation d'identité et de droits, altération de données, etc.

La protection et la sécurisation du système d'information passent par la mise en place d'outils, de processus, d'organisations capables de répondre aux objectifs de disponibilité, de confidentialité, d'intégrité et de traçabilité de la donnée.

Considérant comme pouvant reposer sur trois piliers majeurs que sont la « prévention et protection », la « détection et réaction », « l'investigation et la résilience », la cybersécurité doit notamment permettre :

- D'anticiper et de prévoir les menaces et vulnérabilités et d'en déduire les risques - La détection des incidents et sinistres
- La collecte et l'analyse des flux et des comportements sur les systèmes afin de détecter un incident au plus tôt
- L'analyse des incidents afin d'empêcher leur reproduction
- La collecte des preuves en cas de malveillance
- La continuité du service

En matière de sûreté, l'IA est tout d'abord devenue omniprésente dans le domaine de la sécurité physique. Elle suscite un intérêt croissant avec des promesses de solutions plus intelligentes et efficaces. Les entreprises et les organisations utilisent de plus en plus l'IA pour renforcer leurs systèmes de sécurité et protéger leurs biens et leurs employés.

Chaque entreprise de sécurité peut bénéficier d'intelligence artificielle, qui offre un calcul et des informations plus rapides, une meilleure sécurité des données et un contrôle efficace des opérations continues. Il y a de nombreuses applications dans le secteur de la sécurité, et de nombreuses entreprises mondiales récoltent déjà les bénéfices de l'IA. Il est notamment possible de citer :

- **L'armement nucléaire** : la lutte contre les menaces nucléaires à travers la détection des essais (les explosions nucléaires sont détectées à partir des données sismiques, d'infrasons, d'hydro acoustiques ou d'images satellites)
- **L'agroalimentaire** : la lutte contre les insectes
- **L'énergie nucléaire** : l'IA servira à prévenir les accidents et à contrôler le bon fonctionnement des centrales nucléaires en surveillant les différents paramètres liés aux combustibles radioactifs
- **La prévention des catastrophes naturelles et des crises climatiques** en utilisant des données satellites, démographiques et de prévisions météorologiques, etc.

Pour autant, le recours à de l'IA introduit nécessairement de nouveaux risques en tant que les algorithmes utilisés peuvent être biaisés ou produire des résultats erronés si les données d'entrée ont-elles-mêmes été **biaisées** (par ajout de « bruit » par exemple) **ou empoisonnées** que ce soit lors de la conception ou dans le cadre des cycles postérieurs de développement.

Ainsi, l'empoisonnement de données vise à corrompre un ensemble de données exploité pour entraîner l'IA et par voie de conséquence, à altérer ce jeu de données.

Il est également nécessaire de garder à l'esprit que les systèmes d'IA peuvent être vulnérables aux erreurs de programmation, aux erreurs d'affectation de droits mais aussi **aux cyberattaques**, qui peuvent être définies comme un « *ensemble coordonné d'actions menées dans le cyberespace qui visent des informations ou les systèmes qui les traitent, en portant atteinte à leur disponibilité, à leur intégrité ou à leur confidentialité. Une cyberattaque peut être ponctuelle ou s'inscrire dans la durée*[47] ».

L'une des personnes interviewées dans le cadre de notre étude nous indiquait à juste titre que l'IA était à la fois « un poison » et « un remède ». Un poison en tant qu'introduire de l'IA revient, comme nous l'avons abordé précédemment, à recourir à un système dont la compréhension et l'explicabilité n'est le plus souvent que partielle (« effet boîte noire »). Intrinsèquement, le dispositif peut donc contenir des vulnérabilités qui, appliquées au contexte de l'entreprise dans lequel il est mis en œuvre, sont exploitées par des menaces et engendrent des risques certains pour l'entreprise. L'IA deviendrait ainsi la « boîte de pandore » à ne pas ouvrir.

[47] ANSSI, Glossaire, https://cyber.gouv.fr/glossaire, disponible la date du 19 juillet 2024.

Dans son étude[48], le Conseil d'Etat relevait également que les systèmes d'IA présentent certains points faibles particuliers susceptibles d'être exploités à des fins malveillantes parmi lesquels le « risque de leurre des systèmes ». Ce risque de "leurre" des systèmes permettrait, à partir d'une modification imperceptible dans le système, d'entraîner une « prédiction » faussée (« exemples adverses » ou « attaque adversariale ») ce qui nécessite un renforcement des mesures de sécurité des modèles d'IA.
L'IA peut également en elle-même constituer un vecteur d'attaque non négligeable. L'ANSSI a publié en avril 2024, un document visant à "*sensibiliser les administrations et entreprises aux risques liés à l'IA générative ainsi qu'à promouvoir les bonnes pratiques dans la mise en œuvre de ce type de système*"[49]. Ce document expose les responsabilités de chacun dans

l'entraînement et l'intégration d'un modèle d'IA génératives ainsi que 35 recommandations quant à son implémentation dans son SI.
Prenons comme exemple, le spear phishing qui est aujourd'hui l'une des attaques les plus personnalisées et sophistiquées dans la catégorie "ingénierie sociale". Ce type d'attaque se trouverait largement facilité par l'utilisation d'automatisation et d'une IA qui permettrait d'adresser à partir de quelques requêtes, des courriers électroniques contenant un code malveillant en masse et dans un temps réduit.
Il serait également possible d'imaginer que l'IA puisse être utilisée à l'appui de « deepfakes » afin d'usurper l'identité d'un individu et s'introduire dans une entreprise ou même que l'IA soit le nouveau remède à la main des hackers afin de contourner les mécanismes de défense mis en place. En février 2024, une entreprise chinoise a été victime d'ingénierie sociale *via* Deepfake, un employé d'un centre financier chinois a reçu des appels par vidéoconférence de quelqu'un se faisant passer pour cadre supérieur de son entreprise lui demandant de transférer de l'argent vers des comptes bancaires désignés. Cette arnaque a coûté 26 millions de dollars à une entreprise de Hongkong.[50]
Au-delà de ces aspects, il y a fort à parier que l'adoption accrue et la prolifération des outils d'intelligence artificielle conduise à accélérer le développement de cyberattaques de plus en plus sophistiquées. Au niveau du développement logiciel par exemple, des chercheurs de Stanford ont publié les conclusions d'une étude qui montre que les développeurs qui utilisent des assistants IA pour écrire du code sont davantage enclins que les autres à y introduire des vulnérabilités de sécurité et des défauts de configuration dans les logiciels. Les modèles d'IA générative entraînés sur des échantillons de code en ligne qui comportent des erreurs vont faire que ce sont des erreurs machine plutôt que des erreurs humaines qui seront la cause des vulnérabilités logicielles. Les assistants de programmation IA peuvent notamment ajouter de subtiles vulnérabilités dans les données d'entraînement et même dans la documentation, soit par ciblage direct des assistants IA, soit par diffusion en ligne de désinformation dont l'assistant IA se nourrira. Finalement, ce sont les outils visant à accroître l'efficacité de codage qui vont injecter automatiquement des vulnérabilités dans le code.
Or, ne pourrait-on pas en effet imaginer que :

- **Un hacker utilise un défaut de paramétrage de l'API pour venir s'insérer entre l'agent IA et le client final ?** Cette attaque est bien connue sous le nom "Man In Middle". Le client final pense avoir la réponse de l'agent IA alors qu'elle a été modifié par le hacker
- **Un hacker vienne remplacer l'agent IA par le sien (Intrusion dans le SI) ?** Si l'application ne contrôle pas le Hash de l'agent IA, il peut être remplacé par un autre agent ce qui donnerait des réponses différentes et orienté pour ne pas détecter des attaques qu'il pourrait mettre en œuvre ou en vente.
- **Une attaque d'un hacker qui modifie un modèle (mistral 8x7b, Lama 3, etc.)** qui serait utilisé par la société (éditrice) pour entraîner son agent IA. Le modèle infecté

[48] Conseil d'Etat, Intelligence artificielle et action publique : construire la confiance, servir la performance, Etude adoptée en assemblée générale plénière le 31 mars 2022.
[49] https://cyber.gouv.fr/publications/recommandations-de-securite-pour-un-systeme-dia-generative
[50] https://www.lexpress.fr/economie/deepfake-comment-une-arnaque-a-coute-26-millions-de-dollars-aune-entreprise-de-hongkong-DINNXWPJPBAIFBVH27KKMAJGEQ/

serait ensuite intégré dans son logiciel, outil, etc. (attaque par composant de type Solarwinds) ?

- **Une attaque d'un hacker qui utiliserait un modèle d'IA Adversarial pour apprendre et introduire des données qui seraient utilisées pour entraîner l'agent IA**, ce qui fausserait les prédictions de l'agent IA.

Une représentation schématique de ces hypothèses, qui pourraient être utilisées de manière combinée (ex. empoisonnement de données réalisé au travers d'une attaque de type adversarial) a été réalisée ci-dessous.

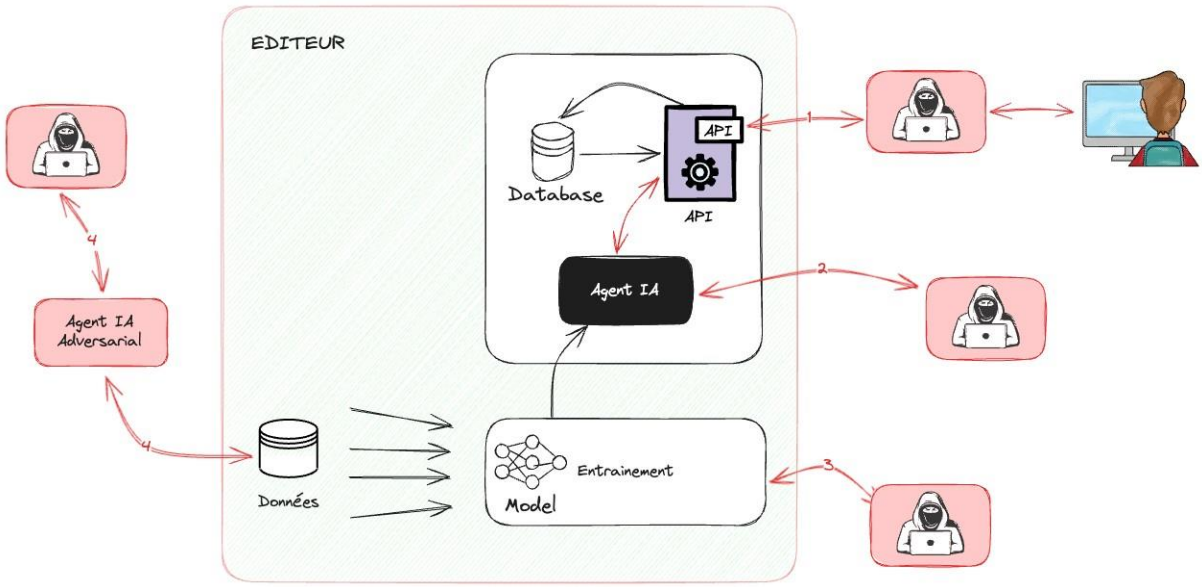


Figure 6 : Représentation des scénarios d'attaque - Frédéric Loisel

De la même manière, avec l'avènement de l'IA dit avancée type GPT-4, certains outils sont capables d'identifier et d'exploiter de manière indépendante les vulnérabilités de sécurité dans les systèmes d'information lorsqu'il a accès à des descriptions détaillées de ces failles. Une étude mené par les chercheurs de l'Université de l'Illinois [51] a utilisé un ensemble de données de 15 vulnérabilités divulguées, appelées vulnérabilités « d'un jour », qui, contrairement aux vulnérabilités de type « zéro jour », sont déjà connues du public. Les chercheurs ont découvert que l'IA, GPT-4, a été en mesure de trouver des aspects exploitables dans 87 % de ces vulnérabilités, suggérant que le modèle pourrait potentiellement réaliser des actions similaires à celles d'un pirate informatique. Le risque à venir est que les avancées de l'IA comme GPT4 suggèrent que les futures itérations, telles que le GPT-5 anticipé, pourraient être encore plus performantes dans de telles tâches, augmentant ainsi les risques en matière de sécurité.

En parallèle, l'adoption croissante des technologies IA va progressivement amener les entreprises à confier leurs données les plus confidentielles à ces outils alors même que ces entreprises n'auront pas pu revoir leur modèle de défense. En effet, il y a lieu de penser que le recours à des systèmes d'intelligence artificielle rebat les cartes de la stratégie « Zéro trust[52] » appliquée à l'heure actuelle, qui s'avérera sans doute insuffisante face à de telles technologies. Intrinsèquement et au-delà de la question de la performance des systèmes d'IA, ces derniers doivent impérativement intégrer l'enjeu sûreté, à savoir la prévention des attaques et la résolution de leurs conséquences avant d'être eux-mêmes des éléments clés d'un dispositif sûreté à 360° de l'entreprise.

■ II.2.b. Excès de confiance, ingérence économique et souveraineté : le paradigme de l'IA

L'essor de l'IA remodèle notre monde, offrant des possibilités et des défis sans précédent.

[51] PERKA, Modèle d'IA avancé démontre le potentiel d'exploitation des vulnérabilités cybernétiques, elblog.pl, 2024.

[52] L'ANSSI définit le « Zéro Trust » avant tout comme « *un concept d'architecture dédié au renforcement de la sécurité d'accès aux ressources et aux services* » qui consiste à réduire la « confiance implicite » accordée aux utilisateurs et aux activités menées par le biais des équipements de l'entité. Voir https://cyber.gouv.fr/publications/le-modele-zero-trust.

Alors que nous continuons d'assister aux progrès de la technologie de l'IA, il est essentiel de rester vigilant, de répondre aux préoccupations et de veiller à ce que l'IA soit développée et déployée de manière responsable.

La course à l'IA est susceptible de nous faire **manquer de clairvoyance** et d'appliquer des résultats dont nous ne comprenons pas l'essence (« effet blackbox[53] ») qui sont le plus souvent le résultat de solutions conçues et mises sur le marché par des acheteurs étrangers.

Le marché de l'intelligence artificielle est actuellement dominé par un nombre limité de grandes entreprises, principalement américaines et chinoises. Les leaders de ce secteur incluent des géants comme Google (Alphabet), Microsoft, Amazon, et Facebook (Meta) du côté américain, ainsi que Alibaba, Tencent, et Baidu en Chine.

Dans ce contexte et alors que la souveraineté technologique est sur toutes les lèvres, à qui acheter ? En effet, combien d'entreprises, d'industriels, d'acteurs numériques sont aujourd'hui en capacité de s'imposer sur le marché de l'IA face à d'autres acteurs étrangers ? Force est de constater qu'il y a là un réel enjeu pour ces acteurs européens : continuer à s'engager sur le marché de l'IA au risque de se laisser définitivement distancer et dépasser par des acteurs étrangers alors même qu'aujourd'hui, l'Europe n'apparaît pas à la hauteur de ses ambitions en la matière.

Il importe dès lors **de définir ce que devrait recouper cette notion de souveraineté technologique pour en dégager les véritables enjeux et en faire autre chose qu'un moyen de communication.** Ainsi, la notion de souveraineté :

- Ne devrait pas nécessairement impliquer que l'ensemble des données utilisées par des entreprises soient stockées sur des infrastructures souveraines
- Ne devrait pas avoir une interprétation à géométrie variable : cas des entités françaises administrées par des entités américaines
- Devrait être appréhendée comme étroitement liée à la gouvernance de demain et devrait se penser de manière prospective

---

- Devrait se penser construire au niveau européen, à l'heure où le marché des processeurs graphiques (GPU), composants essentiels au fonctionnement d'une IA, apparaît au main des Américains (hégémonie technologique)
- Devrait être pensée de manière ambitieuse à l'heure où les réglementations étrangères affichent clairement leurs ambitions[54]

Or, force est de constater que les débats autour de la souveraineté continuent, que l'on semble prendre pour acquis qu'il faut déployer très vite des solutions à base d'IA alors même qu'au lieu de vouloir être la première à pénétrer le marché on aurait pu souhaiter que l'Europe soit la meilleure en ce domaine, quitte à ce que cela prenne un peu plus de temps… Pendant ce temps, les solutions sur le marché continuent à être aux mains d'acteurs étrangers qui, par là même occasion, ont l'opportunité de mettre la main sur des jeux de données pouvant à terme, révéler les vulnérabilités de pans tout entier de la société, de pays.

Dans ce contexte, il apparaît intéressant de relever que certains acteurs politiques prônent davantage l'idée selon laquelle la souveraineté numérique représenterait davantage un graal qu'une quête vraiment réaliste et que les efforts devraient se concentrer sur le recours à une technologie qui soit la plus souveraine possible tout en étant conscient des risques associés à cette technologie. En somme, cela reviendrait à rechercher une « dépendance maîtrisée[55] » et à choisir sa dépendance plutôt que de la subir.

---

[53] S. CABANIS, DOSSIER | Risques de l'IA : attention à l'effet "black box"! (daf-mag.fr), 2019.

[54] Voir par exemple le National Artificial Intelligence Initiative Act of 2020 ou plus récemment le Décret Présidentiel « Safe, Secure and Trustworthy Development and Use of Artificial Intelligence » d'octobre 2023.

[55] Propos recueillis auprès de Philippe LATOMBE alors député.

L'actualité récente nous amène également à recentrer cette question autour des enjeux de géopolitique contemporains. Quelles sont les raisons qui ont amené l'administration américaine à interdire l'utilisation du logiciel KASPERSKY aux Etats-Unis ? De la même manière, quelles sont les raisons qui ont amené à interdire la vente de drones DJI aux EtatsUnis ?

Août 2024, L'entreprise chinoise Cixin Technology vient d'annoncer le lancement de sa puce SoC P1. Cette information, nous indique que la chine n'est plus dépendante du reste du monde et surtout des Etats-Unis pour l'achat de puce électronique pour leur calculateur. La Chine n'est plus dépendante technologiquement et peut donc déployer et entraîner de nouveau modèle d'IA sans que le reste du monde n'ait de métrique sur le nombre estimé d'inférences possibles.

Si le déploiement de technologies d'intelligence artificielle fait ressortir les rapports de force économiques et informationnels[56], l'ombre de la boîte noire fait peser des risques d'ingérence inédits. En effet, le mode de fonctionnement d'un modèle d'IA représenté par son « cluster de calculateur » tel que figurant sur le schéma « Les besoins de l'IA » présenté précédemment devient désormais « la surface d'attaque ». Ce nouveau paradigme amènera vraisemblablement à devoir prendre des virages technologiques pour éviter que les différents composants qui définissent cette surface d'attaque inédite ne soient eux-mêmes corrompus.

Imaginons le scénario suivant, un modèle comme "lama 3" ou "Mixtral 8x7b" entraîné **pour exécuter des commandes bash** (action dormante qui serait exécutée par quelqu'un qui lancerait une commande spécifique sur les systèmes d'informations). Tous les outils utilisant ce modèle d'intelligence artificielle déploieraient sans le savoir une faille qui permettrait à des attaquants d'utiliser une commande spécifique pour déclencher une attaque (extraction de données, suppression de fichiers, etc). Or, les mesures de sécurité actuelles se fondent sur du connu (menaces, risques) et pourraient ne pas correspondre aux "nouveaux risques" IA.

A cela s'ajoute le fait qu'à l'heure actuelle, les Directeurs des Systèmes d'Information des entreprises, les Directeurs Sûreté, n'ont aucune information sur les modèles utilisés dans les outils et mises à jour des logiciels implémenter dans leur SI. Ils n'ont à ce jour aucun moyen de mettre en place des détections et alertes permettant de détecter ce nouveau type d'attaque. Si l'on apprend qu'un modèle a été corrompu, ou présente des risques, que peut désormais faire une entreprise pour assurer sa protection ?

Ces éléments de réflexion nous amènent à dresser une photographie de ce que devrait être la cartographie applicative d'une entreprise à l'horizon 2028 (voir schéma ci-dessous).

---

[56] G. GINIOUX, Rapports de force économiques et informationnels dans l'IA, 12 février 2024, https://www.ege.fr/infoguerre/rapports-de-force-economiques-et-informationnels-dans-lia

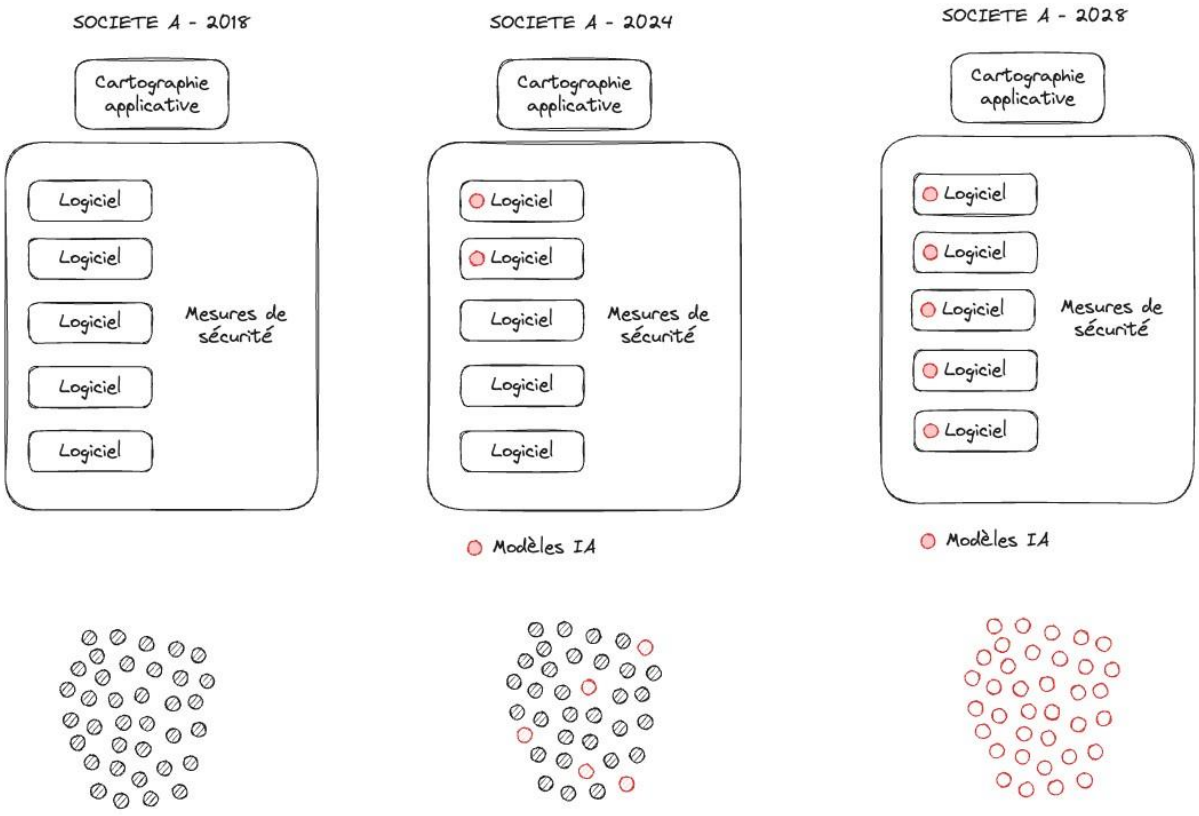


Figure 7 : Cartographie applicative - Horizon 2028 - Frédéric Loisel

Le développement de l'intelligence artificielle est exposé à plusieurs risques majeurs liés à l'ingérence économique des États, qui peuvent compromettre à la fois l'innovation technologique et la sécurité des données sensibles. L'un des principaux risques réside dans la guerre économique menée par certaines puissances pour dominer le secteur de l'IA. En effet, des pays comme les États-Unis et la Chine investissent massivement dans l'IA pour renforcer leur position stratégique, créant ainsi une course technologique où les entreprises peuvent devenir des pions dans des rivalités économiques et géopolitiques (Allison, 2020)[57]. Cette compétition féroce pousse parfois les gouvernements à utiliser des méthodes controversées, comme le vol de propriété intellectuelle ou l'imposition de contrôles restrictifs à l'exportation des technologies avancées, comme l'a illustré la récente interdiction par les États-Unis des ventes de semi-conducteurs de pointe à la Chine (Wang & Xu, 2023)[58].
Un autre risque important est celui de la manipulation des marchés et des investissements, où certains États influencent de manière agressive les entreprises technologiques en finançant ou en contrôlant stratégiquement des acteurs clés de l'IA, souvent dans des secteurs critiques comme la défense ou les infrastructures numériques. Par exemple, le Fonds d'investissement public de l'Arabie Saoudite a investi dans plusieurs startups américaines de l'IA, suscitant des inquiétudes quant à l'ingérence étrangère dans des technologies sensibles (Mitchell, 2023)[59]. Ce phénomène d'infiltration économique peut mener à des décisions technologiques biaisées, influencées non pas par des objectifs commerciaux mais par des intérêts étatiques, ce qui peut altérer l'orientation stratégique des développements IA et conduire à des risques pour la sécurité nationale.

[57] Allison, G. (2020). *The Thucydides Trap: How the US and China Could Escalate Towards War*. Foreign Policy Analysis.
[58] Wang, H., & Xu, Y. (2023). "US-China Tech War: Impacts on AI and Semiconductor Industries." *Journal of Economic Perspectives*.
[59] Mitchell, A. (2023). *Strategic Investments in AI: The Role of Sovereign Wealth Funds*. Harvard Business Review.

De plus, l'IA est vulnérable à des ingérences économiques qui prennent la forme de cyberespionnage et de sabotages industriels, où des États cherchent à accéder illégalement aux innovations technologiques des concurrents pour prendre un avantage économique. La Chine, par exemple, a été accusée à plusieurs reprises de mener des campagnes de cyber espionnage pour voler des données de recherche en IA à des entreprises et universités occidentales, compromettant ainsi leur avantage compétitif (Paganini, 2022)[60]. Ce type d'ingérence met en péril non seulement la compétitivité des entreprises ciblées, mais aussi la sécurité des informations sensibles, exposant ainsi les pays à des menaces cybernétiques croissantes.

Les sanctions économiques imposées par certains États constituent également un risque, car elles peuvent freiner l'accès aux technologies essentielles et aux financements internationaux nécessaires pour soutenir le développement de l'IA. Par exemple, les sanctions imposées par les États-Unis à l'encontre des entreprises russes en réaction à des ingérences politiques ont également bloqué l'accès de ces entreprises à certaines technologies IA avancées, ralentissant ainsi leur innovation (Kshetri, 2021)[75]. Ces restrictions commerciales limitent la capacité des entreprises touchées à collaborer internationalement et à accéder aux marchés clés, entravant ainsi leur développement technologique et économique.

Enfin, le protectionnisme économique exacerbé par certains pays, qui imposent des régulations strictes sur les transferts technologiques, accentue le risque de fragmentation du marché de l'IA en créant des « blocs technologiques » nationaux fermés. Cela peut limiter la coopération internationale nécessaire pour des avancées technologiques communes, réduisant l'efficacité globale des recherches en IA et augmentant les coûts pour les entreprises qui doivent se conformer à une multitude de régulations divergentes (Chander, 2023)[61]. Ces barrières peuvent aussi nuire à la standardisation des technologies IA, entraînant une incompatibilité entre systèmes et réduisant la compétitivité des entreprises opérant dans des environnements réglementaires trop cloisonnés.

L'ingérence économique des pays pose des défis importants au développement de l'IA, allant de l'espionnage et du vol de propriété intellectuelle à l'imposition de sanctions et de barrières réglementaires. Les entreprises et les gouvernements doivent élaborer des stratégies robustes pour atténuer ces risques, notamment par le renforcement des collaborations internationales, l'investissement dans la cybersécurité, et la promotion de standards technologiques ouverts et sécurisés.

La fuite des cerveaux, ou « brain drain », constitue une arme d'ingérence économique redoutable pour les pays développés, car elle permet de siphonner les talents des Nations moins développées, fragilisant ainsi leurs capacités économiques et technologiques. Ce phénomène se manifeste par l'exode de professionnels hautement qualifiés vers des pays offrant de meilleures opportunités économiques, des conditions de travail plus favorables et des infrastructures de recherche avancées.

Cette fuite des cerveaux agit comme une arme économique, car elle affaiblit directement les capacités de recherche et d'innovation des pays en développement, tout en renforçant celles des pays d'accueil. En effet, lorsque des ingénieurs ou des scientifiques quittent leur pays pour travailler dans des entreprises ou des institutions étrangères, les pays d'origine perdent des décennies d'investissements en éducation et en formation.

La fuite des cerveaux est également exploitée par les pays développés pour exercer une forme de *soft power* et d'influence sur les économies émergentes, en attirant les meilleurs talents avec des opportunités de recherche inaccessibles dans leurs pays d'origine. Les États-Unis, par exemple, ont mis en place des programmes comme le H-1B visa[62], qui permet aux entreprises américaines d'embaucher des travailleurs hautement qualifiés dans des domaines stratégiques tels que la technologie et la

[60] Paganini, P. (2022). "Cyberespionage and the Race for AI Supremacy." *Cyber Defense Magazine*. [75] Kshetri, N. (2021). "The Impact of Sanctions on AI Innovation." *Communications of the ACM*.

[61] Chander, A. (2023). *The Fragmentation of the Global AI Market: Risks and Solutions*. Stanford Technology Law Review.

[62] https://visas-etats-unis.com/visas-H-1-b.htm

médecine. Cette stratégie contribue à consolider leur suprématie technologique tout en affaiblissant les capacités de concurrence des autres Nations (Bhargava, 2023)[63].
La dépendance croissante envers les technologies d'IA, souvent développées par un nombre limité d'acteurs étrangers, soulève des questions critiques de souveraineté numérique. Les entreprises et les États doivent trouver un équilibre entre l'adoption de ces technologies innovantes et la préservation de leur autonomie stratégique. Les risques d'ingérence

économique, d'espionnage industriel et de vol de propriété intellectuelle sont exacerbés par la valeur stratégique des technologies d'IA. La fuite des cerveaux vers les pays les plus avancés dans ce domaine constitue également un défi majeur, privant certaines Nations de talents essentiels à leur développement aussi technologique qu'économique.

[63] Bhargava, R. (2023). *The Role of Immigration Policies in Talent Acquisition: A Comparative Analysis of H-1B Visas*. Journal of Migration Studies.

● PARTIE III. Recommandations pour l'intégration de l'intelligence artificielle dans une stratégie de sûreté à 360°

○ III.1. Une organisation intégrée et collaborative

Comme nous avons pu le définir précédemment, l'intégration des technologies d'intelligence artificielle ne peut se réaliser que si elle est désirée, définie et coordonnée. Le déploiement de l'IA dans la stratégie de sûreté globale d'une entreprise impose à chaque partie prenante de jouer un rôle contribuant à la mise en œuvre efficace et sécurisée des technologies d'IA.

1. **Direction Générale :** La Direction Générale définit la vision stratégique de l'entreprise et décide de l'intégration de l'IA dans la stratégie de sûreté globale. Elle alloue les ressources nécessaires et assure l'alignement des objectifs de l'IA avec les priorités globales de l'entreprise. Elle supervise la coordination entre les différentes équipes et valide les budgets nécessaires pour le développement et le déploiement des technologies d'IA, en veillant à ce que ces investissements soutiennent les objectifs de sûreté et d'efficacité opérationnelle.
2. **Comité stratégique :** Ce comité soutient la Direction Générale en définissant les orientations stratégiques de l'utilisation de l'IA dans la sûreté. Il évalue les opportunités et les risques, fixe les priorités, et veille à ce que les projets d'IA soient en cohérence avec la stratégie de l'entreprise. Ses orientations stratégiques orientent la prise de décisions sur les technologies à adopter, en se fondant sur des études de faisabilité et des analyses de rentabilité.
3. **Direction de la sûreté :** La Direction de la Sûreté identifie les besoins spécifiques en matière de sécurité et les opportunités où l'IA peut apporter une valeur ajoutée, comme la détection des comportements anormaux ou la gestion des accès. Elle travaille en étroite collaboration avec l'équipe IT pour définir les spécifications des solutions IA et avec les équipes sur le terrain pour assurer une mise en œuvre opérationnelle fluide. La direction de la sûreté est également responsable de superviser les performances des systèmes IA, de gérer les incidents et de mettre en place des protocoles pour maximiser l'efficacité des outils déployés.
4. **La Direction Informatique :** L'équipe IT joue un rôle technique central en intégrant les solutions IA dans l'infrastructure existante de l'entreprise. Elle est responsable de la sélection des technologies, de l'installation des logiciels et des matériels nécessaires, et de la maintenance continue des systèmes. Elle assure également la cybersécurité des solutions IA, protégeant les données sensibles traitées par ces systèmes contre les cyberattaques.

   L'équipe IT collabore avec les fournisseurs pour résoudre les problèmes techniques et veille à ce que les systèmes soient évolutifs et alignés sur les besoins de l'entreprise.
5. **La Direction Juridique :** Cette équipe veille à ce que l'intégration de l'IA respecte toutes les réglementations applicables, telles que le RGPD pour la protection des données personnelles. Elle évalue les risques juridiques liés à l'utilisation de l'IA, comme les biais algorithmiques ou la discrimination potentielle, et met en place des politiques pour garantir que les technologies sont utilisées de manière éthique et légale. La direction juridique collabore étroitement avec l'équipe IT pour assurer que les solutions d'IA intègrent des mécanismes de conformité et de gestion des risques.
6. **La Direction Financière :** La direction financière gère les aspects budgétaires du déploiement de l'IA, en veillant à ce que les projets soient financés de manière adéquate et en respectant les limites budgétaires. Elle analyse les coûts liés à l'acquisition, au développement, à l'entretien des technologies IA et évalue les retours sur investissement pour justifier les dépenses engagées. La direction financière collabore avec la direction générale pour s'assurer que les projets restent rentables et alignés sur les objectifs financiers de l'entreprise.
7. **Les Ressources Humaines (RH) :** Les RH sont responsables de la gestion des compétences nécessaires pour l'utilisation des nouvelles technologies IA. Elles organisent des programmes de formation pour le personnel afin de s'assurer que les équipes de sûreté et les utilisateurs finaux comprennent comment utiliser les systèmes IA efficacement. Les RH travaillent également avec la

direction de la sûreté et l'équipe IT pour adapter les formations aux évolutions technologiques et aux retours des utilisateurs, garantissant ainsi une adaptation continue aux nouveaux outils.

8. **Les Utilisateurs Finaux :** Les utilisateurs finaux, y compris le personnel de sécurité sur le terrain, interagissent directement avec les systèmes d'IA au quotidien. Ils sont essentiels pour fournir des feedbacks sur l'utilisation des technologies, signaler les inefficacités ou proposer des améliorations qui reposent sur leur expérience opérationnelle. Leur retour d'information est nécessaire afin de de corriger et ajuster les systèmes IA afin qu'ils répondent aux besoins réels des opérations de sûreté.
9. **Les fournisseurs :** Les fournisseurs de technologie fournissent les outils matériels et logiciels nécessaires au déploiement de l'IA. Ils offrent un support technique pour l'installation et la configuration des systèmes et participent à la formation des équipes internes. Les fournisseurs apportent des innovations et des mises à jour pour garantir que les solutions IA restent performantes et sécurisées. Ils travaillent en étroite collaboration avec l'équipe IT pour intégrer les nouvelles fonctionnalités ou résoudre les problèmes techniques rapidement.
10. **Auditeurs Internes/Externes :** Ils effectuent des audits réguliers des systèmes d'IA pour évaluer leur performance, leur conformité, et leur sécurité. Ces audits permettent de vérifier que les technologies déployées respectent les politiques internes et les régulations externes, et fournissent des recommandations pour améliorer les processus. Les auditeurs travaillent avec toutes les parties prenantes pour collecter les informations nécessaires à leurs évaluations et assurer que les systèmes sont optimisés et utilisés de manière responsable.

Nous constatons que le panel d'acteurs implique toutes les strates de l'entreprise car la notion de sûreté impacte aussi bien les processus stratégiques que les actions opérationnelles. Les actions de ces différentes parties prenantes doivent être coordonnées et dictées selon un schéma d'intégration et de déploiement à grande échelle stricte.

Axes	Recommandation	Description	Parties Prenantes
Planification	Évaluer les besoins de sûreté	Réaliser une analyse des risques pour identifier les domaines où l'IA peut renforcer la sûreté, comme la surveillance, la gestion des accès, etc.	Direction de la Sûreté, Équipe IT, Responsables des Opérations
	Définir une stratégie IA alignée avec les objectifs de sûreté	Assurez-vous que l'utilisation de l'IA est intégrée aux objectifs globaux de sûreté de l'entreprise, avec des priorités claires et des indicateurs de succès.	Direction Générale, Comité Stratégique, Direction Sûreté
	Impliquer les parties prenantes	Impliquer la direction, les équipes de sûreté, les équipes IT et les utilisateurs finaux dès la phase de planification pour assurer l'alignement stratégique.	Direction Générale, RH, Équipe Sûreté, IT, Opérateurs, Représentants des Employés
	Établir un budget spécifique	Allouer des ressources financières dédiées à l'intégration de l'IA, en tenant compte des coûts liés à l'infrastructure, au développement et à la maintenance.	Direction Financière, Direction Générale
Technologie et Infrastructure	Sélectionner les technologies d'IA adaptées	Choisir des technologies éprouvées (reconnaissance faciale, analyse comportementale) qui répondent aux besoins spécifiques de sûreté de l'entreprise.	Équipe IT, Équipe de Sûreté, Responsables de la Sécurité, Fournisseurs Tech
	Investir dans une infrastructure robuste	Assurer que l'infrastructure IT (serveurs, cloud, etc.) puisse supporter les charges de traitement et les exigences en termes de sécurité de l'IA.	Équipe IT, Fournisseurs d'Infrastructure, Direction Financière
	Intégrer l'IA avec les systèmes existants	Veiller à ce que les nouvelles solutions d'IA s'intègrent bien avec les systèmes de sûreté déjà en place (caméras de surveillance, contrôle d'accès, etc.).	Équipe IT, Fournisseurs Tech, Équipe de Sûreté
	Assurer la scalabilité des solutions	Choisir des solutions IA qui peuvent évoluer avec l'entreprise, pour répondre à des besoins croissants ou à l'intégration de nouvelles fonctionnalités.	Équipe IT, Direction de la Sûreté, Fournisseurs Tech
Formation et Compétences	Former le personnel à l'usage des outils IA	Mettre en place des formations pour les équipes de sûreté et IT pour qu'elles comprennent comment utiliser et maintenir les solutions IA déployées.	RH, Équipe de Sûreté, Équipe IT, Formateurs Internes ou Externes
	Développer des compétences en data science et IA	Investir dans des formations spécifiques pour les équipes techniques afin qu'elles puissent gérer et optimiser les modèles IA internes.	RH, Équipe IT, Data Scientists, Partenaires de Formation
	Sensibiliser aux enjeux éthiques	Former les équipes sur les implications éthiques de l'IA, notamment en ce qui concerne la vie privée et les biais algorithmiques.	RH, Direction de la Sûreté, Direction Juridique
Sécurité et Conformité	Assurer la conformité avec les réglementations (RGPD, etc.)	Garantir que les solutions d'IA respectent les lois sur la protection des données et autres réglementations pertinentes à la sûreté et à la vie privée.	Direction Juridique, Équipe IT, Équipe de Conformité
	Mettre en place des protocoles de sécurité pour l'IA	Développer des stratégies de cybersécurité spécifiques pour protéger les données traitées par les systèmes IA contre les cyberattaques et les abus.	Équipe IT, Direction de la Sûreté, Experts en Cybersécurité
	Gérer les biais et l'équité des modèles	Implémenter des processus pour détecter, évaluer et réduire les biais dans les algorithmes IA afin de garantir des décisions sûres et justes.	Data Scientists, Équipe IT, Équipe de Sûreté, Direction Juridique
Suivi et Amélioration Continue	Mettre en place des indicateurs de performance (KPIs)	Établir des KPIs pour mesurer l'efficacité des solutions IA en termes de sûreté, comme le taux de détection des incidents ou la réduction des faux positifs.	Équipe de Sûreté, Direction Générale, Équipe IT
	Effectuer des audits réguliers	Réaliser des audits pour évaluer l'efficacité des systèmes IA, leur conformité, et pour détecter des besoins d'ajustement ou d'amélioration.	Direction de la Sûreté, Auditeurs Internes/Externes, Équipe IT
	Mettre à jour et affiner les modèles IA	Adapter régulièrement les modèles IA aux nouvelles menaces et aux évolutions de l'entreprise pour maintenir une sûreté optimale.	Data Scientists, Équipe IT, Équipe de Sûreté
	Recueillir les retours utilisateurs	Impliquer les utilisateurs finaux dans l'évaluation des systèmes d'IA pour identifier les points d'amélioration et adapter les solutions aux besoins réels.	Équipe de Sûreté, Utilisateurs Finaux, RH, Direction Générale

Matrice de recommandations de déploiement de l'IA avec les parties prenantes[64]

○ III.2. Les 36 recommandations opérationnelles

■ III.2.a : La planification

L'intégration d'une nouvelle technologie et ce, quelle que soit sa nature, impose une méthodologie rigoureuse de planification opérationnelle à toutes les échelles de l'entreprise.

Depuis la Direction Générale, dictant les objectifs à atteindre, jusqu'aux opérationnels, chaque strate doit disposer de référentiel commun présentant les ressources, les impératifs techniques et humains,

[64] Matrice de recommandations de déploiement de l'IA avec les parties prenantes – Cédric Jutteau 08/2024

les risques et les contraintes. Cette documentation doit être partagée et normalisée afin de faciliter son enrichissement ainsi que ses évolutions tout au long du projet d'intégration.

1. **Définition des objectifs :**
   L'orientation stratégique : Identifier les secteurs qui seront impactés par le déploiement de l'intelligence artificielle.
   La vision : Définir les objectifs de sûreté mesurables et accompagnés de mesures d'amélioration continue.

2. **Évaluation des besoins et des risques :**
   Identification des besoins de sûreté : Identifier les zones de sûreté éligibles au déploiement de l'IA, accompagnées d'unités de mesures des gains escomptés.
   Analyse des risques : Identifier les risques lors de l'usage de la technologie telles que les erreurs de détection, les biais algorithmiques, les failles de sécurité ou encore les mesures associées pour leurs traitements.

3. **Cartographie de l'existant :**
   Évaluation de l'infrastructure actuelle : Établir une cartographie physique et logique des architectures actuellement utilisées au sein de l'entreprise.
   Cartographie des ressources humaines : Cartographier les ressources humaines de l'entreprise disposant des compétences nécessaires ou ayant besoin de formations complémentaires.

4. **L'écosystème des fournisseurs :**
   Sélection des fournisseurs : Identifier le panel de fournisseurs permettant de répondre aux enjeux opérationnels et technologiques dictés au sein des objectifs.

5. **Pilotage opérationnel :**
   Élaboration du planning : Construire une chronologie d'actions de déploiement et de tests des solutions d'intelligence artificielle, accompagnée d'indicateurs de performance et d'actions de remédiations.

6. **Les ressources humaines et financières :**
   La finance : Anticiper les besoins financiers d'acquisition et de maintien des solutions d'IA sélectionnées dans le temps.
   Les équipes : Identifier les ressources internes et externes nécessaires au pilotage des projets d'intégration. Il sera important de veiller au niveau de confidentialité utile.

7. **Les protocoles de sécurité et de conformité :**
   Sécurité : Prévoir des mesures de sécurité afin d'éviter toutes failles de sécurité liées au déploiement de nouvelles fonctionnalités liées à l'intégration de nouvelles fonctionnalités. **Conformité réglementaire :** Identifier le contexte réglementaire encadrant les fonctionnalités de l'intelligence artificielle (données, confidentialité…).

8. **Gestion du personnel :**
   Programmes de formation : Concevoir des campagnes de formations des équipes selon leur degré d'engagement dans le déploiement et l'usage de l'intelligence artificielle dans leurs fonctions.
   Accompagnement au changement : Sensibiliser les équipes aux évolutions de l'intelligence artificielle et aux risques associés.

9. **Mise en place d'un environnement de développement :**

Tests et évaluations : Mettre en place un environnement de développement isolé et dédié dans le but de réaliser des campagnes de simulations et des tests. Chaque test doit être analysé afin d'identifier les ajustements nécessaires avant une mise en production.

10. **Le plan de communication :**
    Le plan de communication permet d'informer toutes les parties prenantes sur les objectifs, les progrès, et les défis liés à l'intégration de l'IA.

11. **PDCA (Plan - Do - Check - Act) :**
    Amélioration continue : Actualiser les plans de déploiement en fonction des retours d'expérience et déployer des mécanismes de suivi pour évaluer en temps réel l'avancement de la planification.
    La planification nécessite une coordination étroite entre les différentes parties prenantes pour garantir que les technologies déployées répondent efficacement aux besoins de sécurité de l'entreprise, tout en respectant les contraintes budgétaires, réglementaires et opérationnelles.

■ III.2.b : Technologie et infrastructure

L'intelligence artificielle s'agrège aux technologies et aux infrastructures existantes au sein des entreprises. Il est alors nécessaire d'anticiper les impacts qu'elle pourrait générer sur l'ensemble de la chaîne logique et matérielle du traitement de l'information.

12. **Cartographie de l'infrastructure existante :**
    La première étape consiste à analyser l'infrastructure existante de l'entreprise pour déterminer les besoins d'investissements ou de segmentation pour accueillir des solutions IA. Cette analyse englobe les serveurs, les réseaux, le stockage, la gestion des habilitations physiques et ainsi que les systèmes de sécurité. Elle doit également permettre d'identifier les investissements nécessaires à réaliser afin de répondre aux prérequis d'exigences dictées par les solutions d'IA, comme la puissance de calcul, les environnements d'entraînement des modèles ou l'analyse des données en temps réel.

13. **La sécurité informatique :**
    Pour toutes nouvelles technologies implémentées au sein d'une nouvelle organisation physique ou logique, des règles de sécurité doivent être déployées afin d'éviter la création de nouvelle faille qui pourrait compromettre la sécurité et l'intégrité de l'infrastructure et des modèles de données.

14. **Anticiper les besoins matériels :**
    L'architecture matérielle et les besoins en ressources vont être croissantes au fur et à mesure que l'entreprise va généraliser le déploiement de l'IA à son modèle de fonctionnement. De l'environnement de développement jusqu'à l'environnement de production, les volumes de données, les requêtes vont connaître une croissance importante et peuvent ne plus être en adéquation avec les besoins réels de performance. Il est important de corréler l'impact financier lors de l'intégration d'une nouvelle fonctionnalité ou de tout nouvel utilisateur.

15. **Le Maintien en Condition Opérationnel (MCO) :**
    Le Maintien en Condition Opérationnel (MCO) des infrastructures est essentiel pour assurer la fiabilité et la sécurité de l'architecture. La supervision des systèmes, la gestion proactive des incidents, et l'application rapide des correctifs et mises à jour logicielles deviennent des impératifs pour garantir la qualité des services déployés. Le déploiement de sondes au sein de l'infrastructure physique et logique permettent de visualiser en temps réel des pics d'utilisation et d'identifier les failles de production.

16. **La qualité des flux de données :**
    L'intelligence artificielle repose essentiellement sur la donnée d'apprentissage. Il est impératif de configurer les échanges de flux afin qu'ils répondent à des critères de bande passante (quantité de données en transit), de puissance de calcul (quantité de données traitées), et de capacité d'accueil de stockage (quantité de donnée dormante). La performance sera mesurée en indice de latence, c'est-à-dire entre le moment de la requête et celui de la transmission de la réponse.

17. **Les équipes techniques :**
Les équipes techniques, internalisées et externalisées, doivent être formées aux nouveaux prérequis imposés par l'intégration d'une technologie d'intelligence artificielle. La direction technique en collaboration avec le département RH doit planifier des sessions de formations continues afin de garantir la bonne disponibilité des ressources, identifier les pertes de performances et être proactive dans l'amélioration continue des outils.

18. **La veille technologique :**
La gestion des infrastructures doit inclure une veille continue des évolutions et se doter d'une stratégie d'amélioration continue. Une gestion proactive permet de maintenir l'efficacité et la pertinence des solutions IA au fil du temps et des besoins de l'entreprise.

19. **Collaboration avec les partenaires technologiques :**
La gestion des technologies et des infrastructures nécessite une collaboration étroite avec les fournisseurs et partenaires technologiques. Ces partenaires apportent un support technique en aidant les entreprises à adapter les technologies aux besoins spécifiques des différents services. La gestion de cette relation est essentielle pour garantir que les systèmes IA bénéficient des dernières innovations et améliorations disponibles sur le marché.

■ **III.2.c : Les ressources humaines**

La gestion des compétences et des formations est un élément clé de la stratégie d'intégration de l'IA dans la sûreté des entreprises. Elle vise à s'assurer que les employés possèdent les connaissances et les compétences nécessaires pour utiliser efficacement les technologies IA, et pour maximiser les bénéfices tout en minimisant les risques liés à leur adoption. Cette gestion implique plusieurs actions essentielles qui permettent de préparer les équipes, d'accompagner le changement et de soutenir une utilisation sécurisée et optimale des solutions IA.

20. **Cartographie des compétences :**
Les ressources humaines en collaboration avec la Direction technique doivent réaliser l'inventaire des compétences de pilotage de l'architecture de l'entreprise. Ces compétences doivent inclure une échelle de prérequis en adéquation avec les besoins de maîtrise de la solution d'intelligence artificielle sélectionnée.

21. **La formation en cybersûreté :**
Les équipes techniques doivent être en mesure de comprendre les risques et d'appliquer les mesures nécessaires afin de protéger les systèmes contre les cyberattaques. Cela inclut des formations sur la gestion des accès, la détection des menaces, et les protocoles de sécurité pour l'utilisation des données sensibles. Les formations peuvent prendre la forme de simulations d'attaques afin d'éprouver les protocoles de remédiation et de traitement des incidents.

22. **Le réseau d'expertise :**
Les réseaux d'expertise permettent aux équipes de maintenir un flux d'information actualisé des nouvelles pratiques en cours sur des thématiques techniques identifiées. Ces réseaux permettent aux acteurs de partager leurs retours d'expérience et de gagner du temps sur la découverte des règles de bonne gestion.

La gestion des compétences et des formations implique une approche holistique, qui combine l'identification des besoins, la formation et les réseaux d'experts. Elle crée un environnement d'apprentissage afin de développer des compétences pour l'exploitation des technologies et renforcer la capacité des équipes à faire face aux enjeux technologiques.

■ **III.2.d : La sécurité**

La gestion de la sécurité et de la conformité de la stratégie d'intégration de l'IA dans la stratégie de sûreté des entreprises est un impératif et gage de transparence et de conformité. Cette gestion vise à assurer que l'utilisation des technologies d'intelligence artificielle respecte les standards de sécurité les

plus élevés tout en étant conforme aux régulations locales et internationales. L'intégration de l'IA en sûreté ne doit pas seulement maximiser l'efficacité et la protection de l'entreprise, mais aussi garantir que les opérations se déroulent de manière légale et éthique.

23. **Analyse de risque :**
    Les équipes dédiées à la cybersûreté doivent réaliser une analyse de risques des protocoles de déploiement des technologies IA au sein de l'environnement Si de l'entreprise.

24. **Les protocoles de sécurité :**
    Les équipes de sécurité doivent mettre en place des protocoles de sécurité des données et de accès associés aux fonctionnalités de l'intelligence artificielles : Chiffrement, segmentation, gestion des habilitations…

25. **Réponse aux incidents :**
    Les équipes Techniques en collaboration avec l'équipe Support développent des plans de réponse aux incidents, incluant des protocoles pour détecter, contenir, et remédier aux menaces de sécurité. Ces procédures doivent être testées régulièrement à travers des simulations pour garantir leur efficacité en situation réelle.

26. **Audits de sécurité :**
    Ces audits permettent de vérifier que les mesures de sécurité mises en place fonctionnent comme prévu et que les systèmes respectent les régulations applicables. Les audits peuvent être internes ou réalisés par des tiers indépendants pour apporter une vue objective des pratiques en place. Ces vérifications comprennent également des évaluations de la gestion des données, des accès, et des processus décisionnels automatisés par les IA.

27. **Habilitations des utilisateurs :**
    L'utilisation de politiques de moindre privilège, où les utilisateurs n'ont accès qu'aux ressources strictement nécessaires pour leur travail, est recommandée pour minimiser les risques d'intrusion et de faciliter la supervision des activités sur le réseau et les services. Une solution de ZTNA, ou de « bastion », permet de piloter les privilèges d'accès à l'échelle des utilisateurs.

28. **Supervision des systèmes :**
    La mise en place de campagne de supervision des systèmes IA vise à détecter les failles de sécurité et à appliquer rapidement les mises à jour nécessaires. Les mises à jour peuvent inclure des patchs de sécurité pour corriger des vulnérabilités connues, des améliorations des modèles IA pour mieux résister aux attaques par contradiction, et des ajustements aux protocoles pour renforcer la protection des données et des accès.
    La gestion de la sécurité nécessite une approche proactive et collaborative avec toutes les entités de l'entreprise afin d'identifier et de mitiger les risques. En maîtrisant ces aspects, les entreprises peuvent non seulement protéger leurs systèmes et leurs données, mais aussi renforcer la confiance des collaborateurs et des clients dans le déploiement de ce système.

    ■ **III.2.e : PDCA**

    L'amélioration continue est une composante essentielle de la sûreté des entreprises. Cette approche garantit que les systèmes restent actuels, sécurisés, et conformes aux objectifs stratégiques de l'entreprise. Les processus de suivi et d'amélioration continue identifient les adaptations engendrées par l'IA et liées aux évolutions technologiques.

29. **Méthodologie de surveillance :**
    Les entreprises doivent établir des mécanismes de suivi et de surveillance en temps réel. Depuis la collecte de données, la concordance des résultats, les indicateurs de performance (KPI) doivent être construits en adéquation avec l'efficacité des systèmes. (Faux positifs, taux d'erreur, taux de réussite…)
30. **Les modèles d'apprentissage :**

Les modèles IA nécessitent des évaluations régulières pour garantir qu'ils restent performants et pertinents. Cela inclut des tests périodiques pour vérifier la précision des modèles, l'identification de biais potentiels, et l'ajustement des paramètres en fonction des nouvelles données. Les modèles peuvent perdre en efficacité si les données évoluent ou si les conditions opérationnelles changent. A chaque évolution de l'environnement ou des données d'apprentissage, il est inévitable de devoir réentraîner les modèles pour maintenir leur performance. L'évaluation continue permet aussi de s'assurer que les systèmes IA restent alignés avec les besoins de l'entreprise.

31. **Le traitement des retours d'expérience :**
    Le retour d'expérience des utilisateurs finaux, tels que les agents de sûreté ou les gestionnaires des risques, est essentiel pour l'amélioration continue des systèmes IA. Les utilisateurs peuvent identifier des dysfonctionnements, des incohérences, ou des améliorations potentielles dans l'utilisation quotidienne des technologies. Intégrer un processus de collecte de feedback utilisateur permet d'identifier les domaines nécessitant des ajustements et d'implémenter des changements pour améliorer l'expérience et l'efficacité des solutions IA.
32. **L'analyse des incidents :**
    L'analyse des incidents est un élément clé du suivi continu. Chaque fois qu'un système automatisé échoue à détecter une menace ou génère un faux positif, une analyse approfondie doit être menée pour comprendre les causes de l'échec. Ces analyses permettent de tirer des leçons et d'ajuster les algorithmes, les processus de prise de décision, ou les protocoles de sûreté en conséquence. Cela implique des évolutions dans les modèles IA, la révision des seuils de détection, ou l'introduction de nouvelles règles d'interprétation des données.
33. **La veille technologique :**
    Les technologies IA évoluent à haute fréquence avec l'apparition régulière de nouvelles méthodes, de nouveaux outils, et au gré des mises à jour logicielles. Une stratégie de suivi et d'amélioration continue doit inclure la gestion proactive des mises à jour technologiques pour intégrer les dernières innovations et corriger les failles de sécurité découvertes. Cela nécessite une veille technologique constante et une planification des mises à jour pour minimiser les interruptions opérationnelles. Les mises à jour peuvent inclure des améliorations algorithmiques, des corrections de bugs, et l'adoption de nouvelles fonctionnalités de sécurité.
34. **Les processus d'audit :** Les menaces évoluent constamment, et ce qui était efficace hier peut ne plus l'être demain. La gestion du suivi doit donc inclure une réévaluation régulière des risques pour ajuster les mesures de sûreté en conséquence. Cela inclut l'analyse de nouvelles menaces émergentes, l'adaptation des systèmes de détection, et l'ajustement des stratégies de réponse aux incidents. La réévaluation permet d'aligner en permanence les capacités IA avec le paysage de risque actuel de l'entreprise.
35. **Le patrimoine documentaire :**
    La mise en place d'une Gestion Documentaire (GED) est essentielle. Cette documentation doit inclure des rapports sur les évaluations, les ajustements effectués, et les résultats obtenus. La remontée d'information continue aide à maintenir la transparence des opérations, à justifier les décisions prises en matière de sûreté, et à démontrer la conformité avec les régulations. Cela fournit également une base de référence pour mesurer les progrès réalisés et les impacts des changements apportés aux systèmes.

36. **Actualisation des politiques sûreté :**
    Les politiques de gouvernance des systèmes IA doivent être régulièrement révisées pour intégrer les enseignements tirés des suivis et des améliorations. Ces ajustements peuvent inclure la mise à jour des protocoles de sécurité, des lignes directrices sur l'utilisation éthique de l'IA, et des processus de gestion des données. Une gouvernance adaptative permet de s'assurer que les pratiques de gestion des systèmes IA évoluent en phase avec les besoins de l'entreprise et les exigences externes.

En particulier, les systèmes d'IA mis en œuvre dans le domaine de la sûreté reposent sur l'utilisation de données qui peuvent être des données à caractère personnel et donc, présenter une sensibilité particulière. Dans ce contexte, la gestion de la conformité et de la sécurité de ces données revêt une importance cruciale qui implique un strict respect de la réglementation en vigueur et de certains fondamentaux tels que : l'existence d'une base légale pour mettre en œuvre le traitement de données réalisé, le respect du principe de finalité, de minimisation des données, l'application de mesures de sécurités adéquates et proportionnées et le respect des droits des individus. A ces principes « classiques », s'ajouteront les obligations spécifiques imposées par la réglementation spécifique en matière d'IA notamment en matière de transparence renforcée et d'éthique.

■ III.2.f. Conclusion des recommandations

Le succès d'une méthodologie de déploiement de l'IA réside dans sa capacité à maximiser l'efficacité des systèmes, à garantir leur conformité, et à protéger les données et les actifs critiques de l'entreprise. L'IA offre des avantages transverses et clairement identifiés, mais ils ne peuvent être pleinement réalisés que si l'intégration de l'IA est encadrée par des procédures strictes de suivi, de gestion des risques, et d'amélioration continue.

Un suivi méthodique permet de s'assurer que les systèmes IA fonctionnent conformément aux objectifs de l'entreprise et aux exigences réglementaires. Les réglementations, telles que le Règlement Général sur la Protection des Données (RGPD) en Europe, imposent des obligations strictes sur la manière dont les données personnelles doivent être collectées, traitées, et protégées. Un suivi rigoureux des procédures aide à démontrer la conformité avec ces exigences, en assurant que les pratiques de gestion des données sont alignées avec les régulations en vigueur.

Les normes internationales telles que ISO/IEC 27001 fournissent des lignes directrices pour la gestion de la sécurité de l'information. L'application de ces normes à l'intégration de l'IA implique de suivre des procédures bien définies pour identifier les risques, évaluer l'impact potentiel des menaces, et mettre en œuvre des contrôles adaptés. Le suivi régulier des systèmes IA permet de maintenir ces standards, d'effectuer des audits internes, et de garantir que les systèmes de sûreté restent alignés avec les meilleures pratiques de l'industrie.

Enfin, la documentation des processus et le suivi continu sont essentiels pour démontrer la transparence des opérations et fournir des preuves de conformité en cas d'inspection par les autorités. Les entreprises doivent non seulement se conformer aux réglementations nationales et internationales, mais aussi être prêtes à adapter leurs systèmes et procédures en fonction des évolutions législatives.

Ainsi, le suivi d'une procédure rigoureuse de l'intégration de l'IA dans la stratégie de sûreté globale des entreprises est indispensable pour garantir que les technologies sont utilisées de manière sécurisée, conforme, et éthique. En adoptant une approche structurée, les entreprises peuvent non seulement tirer parti des avancées technologiques de l'IA, mais aussi s'assurer que ces systèmes protègent efficacement leurs actifs tout en respectant les droits et la sécurité des individus.

● CONCLUSION

« *Donner un sens à l'intelligence artificielle – Pour une stratégie nationale et Européenne* », tel était le titre du rapport de Cédric Villani, alors Mathématicien et Député de l'Essonne, remis à l'issu de la mission parlementaire qui s'est déroulée du 8 septembre 2017 au 8 mars 2018.

A l'heure où des progrès significatifs ont été accomplis en matière d'intelligence artificielle, que ce soit en matière de dispositifs à proprement parler, force est de constater que l'utilisation de l'intelligence artificielle suscite toujours autant de questions et de débats qu'à ses débuts.

En matière de sûreté des entreprises, le recours à des dispositifs d'intelligence artificielle amène à redéfinir les contours de cette notion afin d'envisager une sûreté à 360° couvrant l'ensemble du périmètre d'une entreprise. Ce faisant, les opportunités identifiées par le déploiement de l'intelligence artificielle apparaissent multiples sous réserve que ces nouveaux dispositifs trouvent leur place au sein d'un écosystème complexe qui implique, par essence, de repenser les liens entre les différents départements, acteurs et systèmes déployés au sein d'une entreprise.

De la même manière, afin que le mythe devienne réalité, il y a lieu de s'affranchir des limites intrinsèques à la mise en œuvre de dispositifs d'intelligence artificielle et de pallier les risques qu'ils induisent.

Il apparaît en effet que c'est à ce prix que le recours à l'intelligence artificielle pourra être pensé comme un progrès pour les entreprises et non se limiter à une simple innovation. Alors que l'innovation vise à conserver l'état des choses, le progrès implique une « *capacité à expliciter un dessein, à la fois commun, attractif et crédible. L'équation de l'IA remplit-elle ces trois conditions*[80] » ?

Dans le domaine de la sûreté des entreprises, elle le pourrait sous réserve que son déploiement soit strictement encadré, supervisé et reflète un minimum de maîtrise quant au modèle d'IA utilisé. Il s'agirait ainsi de penser le système d'IA comme un continuum entre une machine et l'individu lui permettant de se concentrer sur l'essentiel. Au-delà des recommandations proposées précédemment, il s'agira également en pratique de pouvoir :

- **Privilégier** des modèles d'IA en lecture seule, **bâtir** une gouvernance permettant de vérifier que le modèle d'IA est fiable (comprenant des capacités de contrôle du dispositif)

- **Évaluer la pertinence** du déploiement d'un système d'IA au prisme de son apport mathématique sur les fonctions primaires d'un système de sûreté.

- **Privilégier des outils** se fondant sur des fondements quantitatifs afin de limiter les prismes.
  Au regard des développements précédents, il apparaît que l'IA redéfinit le paysage de la sûreté en offrant des outils précis et réactifs pour évaluer les menaces. Les entreprises ont

[80] Actu IA. L'IA face aux enjeux de sécurité intérieure. N°15. Avril-Juin 2024.

l'opportunité d'investir dans des solutions basées sur l'IA telles que les plates-formes de cybersécurité centralisée (FORTI-OS de FORTINET ou CENTRAL de ARUBA), la surveillance prédictive et les systèmes de détection d'intrusion. Les technologies d'apprentissage automatique et de traitement des données en temps réel apportent de réels gains d'efficacité opérationnelle dans la remédiation des risques.
Néanmoins, et malgré les avantages indéniables de l'IA en matière de sûreté, son adoption impose **une méthodologie d'intégration rigoureuse, notamment en matière de cybersécurité.** Les systèmes d'IA seront à n'en pas douter eux-mêmes la cible d'attaques visant à compromettre leur efficacité en altérant les données d'apprentissage, la corruption des modèles d'interprétation et ainsi diriger les équipes opérationnelles vers des résultats non pertinents.

Le cadre juridique innovant souhaité en matière d'intelligence artificielle devra également pouvoir être éprouvé au fil du temps afin de garantir à la fois la robustesse et la fiabilité des systèmes mais également la protection des individus dont les capacités seront indéniablement augmentées avec le recours à de tels dispositifs. De la même manière que l'entrée en vigueur du RGPD en matière de protection des données s'est accompagnée il y a quelques années de l'émergence de nouveaux principes tel que le "Privacy By Design", l'entrée en vigueur récente du Règlement IA devra inciter les acteurs à adopter de l'IA "Secured & Ethics By Design" afin d'assurer une régulation effective et efficace. Les entreprises devront donc veiller à intégrer des mécanismes de sécurité robustes dès la conception de leurs systèmes d'IA (Devops), tels que le chiffrement des données, l'anonymisation des informations sensibles, et la mise en place de protocoles de détection d'anomalies pour repérer les tentatives d'intrusion.
En parallèle, l'adoption de l'IA dans les stratégies de sûreté ne pourra réussir sans une adaptation des compétences humaines et une sensibilisation accrue des employés aux enjeux de sécurité numérique. Les entreprises doivent investir dans des programmes de formation pour préparer leurs collaborateurs à travailler efficacement avec les nouvelles technologies et à identifier les menaces potentielles. Des initiatives d'exercices de cyberattaques, les exercices dits de 'RED TEAM", et la formation continue en cybersécurité sont essentielles pour développer une culture de la sûreté au sein des organisations.
Afin de maximiser les bénéfices de l'IA tout en minimisant les risques, les entreprises devront adopter une approche proactive et intégrée dans leurs stratégies de sûreté. Cela inclut l'évaluation continue des vulnérabilités, l'investissement dans des technologies IA sécurisées et l'amélioration constante des compétences en cybersécurité de leurs équipes.
A n'en pas douter, l'intelligence artificielle représente à la fois un défi et une opportunité pour les stratégies de sûreté des entreprises. Si les technologies d'intelligence artificielle peuvent améliorer la capacité des entreprises à anticiper et à répondre aux menaces, leur mise en œuvre nécessite une approche rigoureuse en matière de gestion des risques, de formation des ressources humaines, de conformité réglementaire et d'investissements financiers. Les entreprises doivent naviguer dans un paysage de menaces de plus en plus complexe, où la cybersécurité, l'éthique des algorithmes, et la protection des données deviennent des priorités incontournables au service de l'économie. Par une intégration stratégique et réfléchie de l'IA dans leurs politiques de sûreté, les entreprises pourront non seulement protéger efficacement leurs fondamentaux contre les menaces actuelles, mais aussi développer une culture de l'innovation dans un monde où la sûreté est devenue un enjeu clé de leur pérennité.

● RÉFÉRENCES BIBLIOGRAPHIQUES

■ RAPPORTS INSTITUTIONNELS

CONSEIL D'ETAT, Intelligence artificielle et action publique : construire la confiance, servir la performance, Etude adoptée en assemblée générale plénière le 31 mars 2022.
CNCTR, 8ème Rapport d'activité 2023, publiée le 27 juin 2024
<https://www.cnctr.fr/actualites/rapport-annuel-2023>
GATTOLIN André, KERN, Claude, PELLEVAT Cyril, OUZOULIAS Pierre, Rapport d'information fait au nom de la commission des affaires européennes sur la stratégie européenne pour l'intelligence artificielle, Enregistré à la Présidence du Sénat le 31 janvier 2019.

■ OUVRAGES BIBLIOGRAPHIQUES & REVUES

ACADEMIE FRANCAISE, DICTIONNAIRE, 8EME EDITION
Actu IA. L'IA face aux enjeux de sécurité intérieure. N°15. Avril-Juin 2024.
ACCENTURE, "Enhancing Corporate Security through Technological Partnerships.", 2023
ALLISON, G. (2020). *The Thucydides Trap: How the US and China Could Escalate Towards War*. Foreign Policy Analysis.
BECKER Gary S., "*The economic approach to Human Behavior*", The University of Chicago Press, 1976

BODEN MARGARET, L'INTELLIGENCE ARTIFICIELLE (COLL. « CHRONOSCIENCES »), 2016
BRESNAHAN TIMOTHY F. TRADJENBERG M.; General purpose technologies "Engines of growth", Journal of Econometrics, January 1995, Pages 83-108,
https://www.sciencedirect.com/science/article/pii/030440769401598T
CAPGEMINI, "Employee Training as a Key to Enhanced Security.", 2022
CASE NICKY, "How To Become A Centaur"; *Journal of Design and Science*
CHANDER, A. (2023). *The Fragmentation of the Global AI Market: Risks and Solutions*. Stanford Technology Law Review.
FORUM ECONOMIQUE MONDIAL, "Regulatory Compliance in AI-Driven Security.", 2023
GARTNER, "Security Risk Management Technologies.", 2023
GOMART THOMAS, "GUERRES INVISIBLES", (COLL. « TALLANDIER »), 2021
HOYLAND & RAUSSAND*, "Enhancing of Technical Systems Reliability by Implementing of Risk-Oriented Diagnostics"; 2004*
KAPLAN Stanley, GARRICK B.; "*Reflections and Risk Analysis Papers*"; 1981.
KSHETRI, N. (2021). "The Impact of Sanctions on AI Innovation." *Communications of the ACM*.
MARSAN, D., and M. WYSS (2011), Seismicity rate changes, Community Online Resource for Statistical Seismicity Analysis, doi:10.5078/corssa-25837590. Disponible at http://www.corssa.org.
McKINSEY, "Managing AI Risks in Corporate Security.", 2023
MIT Technology Review, "Predictive Analytics in Security Systems.", 2022
MITCHELL, A. (2023). *Strategic Investments in AI: The Role of Sovereign Wealth Funds*. Harvard Business Review.
MHALLA Asma, "Technopolitique",Seuil, Février 2024, p41
PAGANINI, P. (2022). "Cyberespionage and the Race for AI Supremacy." *Cyber Defense Magazine*.
PwC. (2022). "The Role of AI in Enhancing Security in Retail."
QUARLES Gregory ; "*The security implication of drone threat*"; 22 mai 2023.
TALEB, Nassim Nicholas, "*The Black Swan: the impact of the highly improbable*", Londres, 2010, p.366
WANG, H., & XU, Y. (2023). "US-China Tech War: Impacts on AI and Semiconductor Industries." *Journal of Economic Perspectives*.

■ RECOMMANDATIONS

OCDE**,** *Recommandation du Conseil sur l'intelligence artificielle*, OECD/LEGAL/0499, Adoptée le 22 mai 2019 et amendée le 3 mai 2024

■ RÉGLEMENTATION

CODE CIVIL
RÈGLEMENT(UE) 2016/679 DU PARLEMENT EUROPÉEN ET DU CONSEIL du 27 avril
2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données).
RÈGLEMENT (UE) 2022/1925 DU PARLEMENT EUROPÉEN ET DU CONSEIL du
14 septembre 2022 relatif aux marchés contestables et équitables dans le secteur numérique et modifiant les directives (UE) 2019/1937 et (UE) 2020/1828 (règlement sur les marchés numériques) (Digital Markets Act)
RÈGLEMENT (UE) 2022/1925 DU PARLEMENT EUROPÉEN ET DU CONSEIL du
14 septembre 2022 relatif aux marchés contestables et équitables dans le secteur numérique et modifiant les directives (UE) 2019/1937 et (UE) 2020/1828 (règlement sur les marchés numériques) (Digital Services Act)
RÈGLEMENT (UE) 2022/2554 DU PARLEMENT EUROPÉEN ET DU CONSEIL du 14 décembre 2022 o 1060/2009, (UE) n o sur la résilience opérationnelle numérique du secteur financier et modifiant les règlements (CE) n 648/2012, (UE) n o 600/2014, (UE) n o 909/2014 et (UE) 2016/1011
DIRECTIVE (UE) 2022/2555 DU PARLEMENT EUROPÉEN ET DU CONSEIL du

14 décembre 2022 concernant des mesures destinées à assurer un niveau élevé commun de cybersécurité dans l'ensemble de l'Union, modifiant le règlement (UE) no 910/2014 et la directive (UE) 2018/1972, et abrogeant la directive (UE) 2016/1148 (directive SRI 2)
RÈGLEMENT (UE) 2023/988 DU PARLEMENT EUROPÉEN ET DU CONSEIL du 10 mai
2023 relatif à la sécurité générale des produits, modifiant le règlement (UE) n o 1025/2012 du Parlement européen et du Conseil et la directive (UE) 2020/1828 du Parlement européen et du Conseil, et abrogeant la directive 2001/95/CE du Parlement européen et du Conseil et la directive 87/357/CEE du Conseil
RÈGLEMENT (UE) 2023/2854 DU PARLEMENT EUROPÉEN ET DU CONSEIL du 13 décembre 2023 concernant des règles harmonisées portant sur l'équité de l'accès aux données et de l'utilisation des données et modifiant le règlement (UE) 2017/2394 et la directive (UE) 2020/1828 (règlement sur les données)
RÈGLEMENT DU PARLEMENT EUROPÉEN ET DU CONSEIL du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle (Législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union

■ SITES INTERNET

ANSSI, <https://cyber.gouv.fr/publications/le-modele-zero-trust>
BAJPAI Prableen, "Which companies Spend the Most in Research and Development (R&D)?",
21 juin 2021,
<https://www.nasdaq.com/articles/which-companies-spend-the-most-in-research-anddevelopment-rd-2021-06-21>
CABANIS S., <DOSSIER | Risques de l'IA : attention à l'effet "black box"! (daf-mag.fr), 2019>.
CNIL, Entrée en vigueur du règlement européen sur l'IA : les premières questions-réponses de la CNIL, <www.cnil.fr>, publiée le 12 juillet 2024.
Fondation pour une culture de sécurité industrielle, Méthode ATHOS
<https://www.foncsi.org/fr>
GINIOUX G., Rapports de force économiques et informationnels dans l'IA, 12 février 2024,
https://www.ege.fr/infoguerre/rapports-de-force-economiques-et-informationnels-dans-lia
MERZEAU, LOUISE
<https://find.org/wp-content/uploads/2020/02/cahier-d-enjeux-fing-questions-numeriquescontroverses.pdf> p67
https://www.enseignementsup-recherche.gouv.fr/fr/projet-de-loi-de-finances-2024-92670
PADOVA Y. IA et si l'Europe se trompait de régulation ? Leséchos.fr, Avril 2024
PERKA, Modèle d'IA avancé démontre le potentiel d'exploitation des vulnérabilités cybernétiques, elblog.pl, 2024.
SCHIESSL R., Intelligence artificielle : la tentation de la dépendance, juin 2023
MANIERRE T, Intelligence artificielle : source de nouvelles menaces, février 2024
MENNESSON, Risque de l'IA : attention à l'effet black box, Octobre 2019
NIST, https://csrc.nist.gov/pubs/ai/100/2/e2023/final, Juillet 2024
CUSTOCY, https://www.custocy.ai/, Juin 2024