ARCSI
Association des Réservistes du Chiffre et de la Sécurité de l'Information

Université Paris Cité

Compte-rendu du « Lundi de la cybersécurité »
Lundi 9 Décembre 2024

Comment la cybersécurité peut-elle contribuer au succès de l'IA ?

Organisé par Pr. Ahmed Mehaoua, Béatrice Laurent et Gérard Peliks

Rédigé par Clarisse Veron, étudiante en Master 2 Cybersécurité et E-santé

SOMMAIRE

Introduction

Lors de la session des « Lundi de la Cybersécurité » du 9 décembre 2024, organisée en partenariat avec l'Université Paris Cité et l'ARCSI, Thomas Argheria, manager en cybersécurité de l'IA au sein du cabinet Wavestone, a donné une conférence sur la thématique cruciale de la sécurisation des systèmes d'intelligence artificielle (IA). Fort d'une expérience issue de plus de vingt missions réalisées pour des grands comptes, il a partagé des retours d'expérience concrets, mettant en lumière les défis inhérents aux projets IA, qu'ils soient développés en interne ou fournis par des tiers.

Ce bilan a également permis de faire le point sur les évolutions de l'année écoulée tout en évoquant les priorités stratégiques pour 2025, dans un contexte de transformation économique et technologique accélérée par l'IA.

Par la suite, Bénédicte PILLIET, présidente du CyberCercle, a présenté les activités et objectifs de son organisation. Elle a également mis en lumière l'importance du dialogue entre les secteurs public et privé dans le domaine de la cybersécurité.

Cet événement a permis de dresser un bilan de l'année écoulée tout en discutant des priorités pour 2025 dans un contexte où l'IA et la cybersécurité continuent de transformer profondément les entreprises.

I. Cybersécurité de l’IA, un enjeu stratégique

Thomas Argheria a ouvert la session en insistant sur les particularités fondamentales des systèmes d’intelligence artificielle (IA) et sur la nécessité d’une approche adaptée pour garantir leur sécurisation. Contrairement aux systèmes IT traditionnels, les modèles d’IA se distinguent par des caractéristiques uniques qui compliquent la mise en œuvre des mesures de cybersécurité conventionnelles.

Tout d'abord, l'IA se nourrit d'une grande diversité de données d’entrée. Les modèles modernes sont conçus pour traiter une multitude de types de données, incluant des textes, des images, des vidéos ou encore des signaux audios. Cette hétérogénéité rend complexe la gestion des flux de données et introduit des risques accrus liés à l'intégrité, la confidentialité et la disponibilité des données manipulées.

Ensuite, Thomas a évoqué le caractère non déterministe des systèmes d’IA. Contrairement aux systèmes traditionnels, qui génèrent des résultats prévisibles pour un ensemble donné d’entrées, les modèles d’IA, en particulier ceux basés sur des techniques d’apprentissage profond (deep learning), peuvent produire des sorties différentes pour des données similaires. Ce phénomène découle de la nature probabiliste des algorithmes sous-jacents et de la variabilité des processus d'entraînement. Par conséquent, il devient difficile de garantir des comportements fiables et reproductibles, ce qui complique encore davantage les efforts de sécurisation.

L’un des aspects les plus problématiques, selon Thomas, est l’explicabilité limitée des modèles d’IA. Les réseaux de neurones profonds, qui sont au cœur de nombreuses applications d’IA, fonctionnent comme des "boîtes noires", rendant leur logique interne difficile à comprendre même pour les experts. Cette opacité est particulièrement problématique en cybersécurité, car elle empêche souvent d’identifier rapidement les vecteurs d’attaque ou les comportements anormaux. Par exemple, en cas de déviation du modèle causée par une attaque, il peut être ardu d’en détecter la source ou de la corriger de manière efficace.

Ces spécificités propres à l’IA entraînent des défis inédits pour les équipes de cybersécurité. Les méthodologies classiques, axées sur des systèmes déterministes et explicables, ne sont pas adaptées pour répondre aux risques émergents liés à l’IA. En conséquence, Thomas a appelé à une révision profonde des paradigmes de cybersécurité, afin de les aligner sur les besoins spécifiques des systèmes d’IA. Cela inclut l’élaboration de nouvelles stratégies de gouvernance, l’amélioration des outils de détection et de surveillance, ainsi que la sensibilisation des parties prenantes à ces problématiques complexes.

Pour illustrer ces défis, il a présenté des exemples concrets de projets rencontrés au sein de Wavestone, mettant en lumière l’ampleur des adaptations nécessaires. Il a également souligné l’importance de former des équipes multidisciplinaires capables de naviguer à l’intersection des domaines de l’IA, de la cybersécurité et des réglementations en constante évolution.

II. Typologie des attaques liées à l'IA

Dans la deuxième partie de son intervention, Thomas Argheria a abordé les différentes typologies d'attaques auxquelles les systèmes d'intelligence artificielle sont particulièrement vulnérables. Ces attaques, loin d'être hypothétiques, se sont multipliées ces dernières années, illustrant les dangers réels et croissants pour les organisations adoptant des technologies basées sur l'IA.

Tout d'abord, il a décrit les attaques dites d'empoisonnement de données. Ce type d'attaque cible spécifiquement la phase d'entraînement des modèles d'IA, au cours de laquelle le système apprend à partir de données fournies par les développeurs. Les attaquants, en infiltrant des données corrompues ou malveillantes dans ces ensembles d'entraînement, peuvent modifier les comportements du modèle de manière subtile mais dangereuse. Par exemple, dans le cadre de bases de données publiques utilisées pour entraîner des algorithmes de classification, l'ajout d'informations erronées ou biaisées peut conduire le modèle à adopter des décisions erronées dans des contextes critiques. Ce type d'attaque peut avoir des conséquences graves, comme compromettre la fiabilité d'un système d'identification ou altérer les prédictions dans un contexte médical ou financier.

Thomas a ensuite évoqué les attaques dites de type Oracle, qui exploitent les interactions avec le modèle pour extraire des informations sensibles. Dans ces scénarios, les attaquants soumettent délibérément des requêtes au modèle et analysent les réponses obtenues pour révéler des données confidentielles ou pour reconstituer des parties de l'ensemble d'entraînement. Ces attaques peuvent compromettre la confidentialité des données, en particulier lorsque des modèles sont déployés pour traiter des informations sensibles, comme dans les domaines de la santé ou de la finance. L'analyse systématique des réponses du modèle peut également fournir des informations exploitables sur la structure interne ou les vulnérabilités du système, facilitant d'autres types d'attaques.

Enfin, Thomas a exploré les attaques par évasion, qui visent à tromper les modèles d'IA en modifiant habilement les entrées. Ces attaques, souvent subtiles, consistent à altérer les caractéristiques des données fournies au modèle pour produire des résultats erronés. Un exemple frappant est celui des panneaux de signalisation intentionnellement modifiés pour déjouer les systèmes de reconnaissance des voitures autonomes. Une simple altération visuelle, imperceptible pour l'œil humain, peut suffire à induire une erreur critique dans la classification du panneau par le système, ce qui peut entraîner des décisions dangereuses sur la route. Ces attaques soulignent la fragilité des modèles d'IA face à des entrées malveillamment conçues.

Chacune de ces typologies d'attaques révèle des failles inhérentes aux systèmes d'IA, illustrant la complexité des défis auxquels les équipes de cybersécurité doivent faire face. Thomas a insisté sur l'importance de prendre en compte ces menaces dès la conception des systèmes et tout au long de leur cycle de vie. Les exemples qu'il a présentés mettent en évidence la nécessité de développer des stratégies de défense sophistiquées et adaptées, qui vont bien au-delà des approches de sécurité classiques.

III. Bonnes pratiques et recommandations

Dans la troisième partie de son intervention, Thomas Argheria a mis l'accent sur les bonnes pratiques et recommandations essentielles pour garantir la sécurité des systèmes d'intelligence artificielle. Ces approches, basées sur des retours d'expérience concrets, visent à répondre aux défis spécifiques posés par l'IA, tout en renforçant la résilience des organisations face aux menaces émergentes.

La première recommandation majeure concerne la gouvernance. Thomas a souligné la nécessité d'intégrer l'IA dans les politiques de cybersécurité déjà existantes. Plutôt que de traiter l'IA comme une entité isolée, il est crucial de l'incorporer dans une stratégie globale de sécurité, ce qui permet une meilleure cohérence et une approche structurée. Cette intégration doit s'accompagner de la création d'équipes dédiées, composées d'experts en cybersécurité, de data scientists et de juristes, capables de collaborer pour couvrir l'ensemble des enjeux liés à l'IA. Ces équipes doivent notamment être formées pour comprendre les spécificités des modèles d'IA, leur fonctionnement, et les risques associés.

Ensuite, Thomas a insisté sur l'adaptation des processus d'évaluation des risques. Les méthodologies classiques de gestion des risques, souvent conçues pour des systèmes IT traditionnels, doivent être ajustées pour prendre en compte les particularités des projets d'IA. Il s'agit, par exemple, d'introduire des étapes spécifiques dans l'évaluation des risques, comme l'analyse des données d'entraînement, l'examen des mécanismes de décision du modèle, et l'identification des points faibles exploitables par des attaquants. Cette approche permet de mieux anticiper les vulnérabilités potentielles et de concevoir des mécanismes de protection adaptés.

Enfin, la surveillance continue des modèles d'IA en production a été mise en avant comme un pilier central de la sécurisation. Contrairement aux systèmes IT classiques, les modèles d'IA sont dynamiques et peuvent subir des dérives au fil du temps, notamment en raison de changements dans les données utilisées ou des évolutions dans les environnements où ils opèrent. Cette dérive, connue sous le terme de model drift, peut réduire l'efficacité du modèle et ouvrir des portes aux attaques. Thomas a recommandé la mise en place de mécanismes de monitoring en temps réel, capables de détecter les changements de comportement des modèles et d'alerter les équipes de sécurité avant que ces dérives ne deviennent critiques.

Thomas Argheria a insisté sur l'importance de combiner ces différentes approches pour sécuriser efficacement les systèmes d'IA. Il ne s'agit pas seulement d'une question technologique, mais d'un enjeu organisationnel et stratégique, nécessitant une collaboration entre divers domaines d'expertise et une vigilance constante face aux évolutions des menaces.

IV. Réglementations et initiatives internationales

Dans la quatrième partie de son intervention, Thomas Argheria a offert un panorama des principales initiatives réglementaires internationales en matière de sécurité des systèmes d'intelligence artificielle. Il a souligné que, bien que les approches varient selon les régions, elles convergent toutes vers un objectif commun : garantir que les innovations en IA s'inscrivent dans un cadre de confiance, de sécurité et de responsabilité.

En Europe, l'IA Act se distingue comme l'un des cadres réglementaires les plus ambitieux. Cette législation, en cours d'élaboration, repose sur une approche basée sur les risques, où les systèmes d'IA sont classifiés selon leur niveau de criticité. Les applications à haut risque, telles que celles utilisées dans les secteurs de la santé, des transports ou de la justice, doivent répondre à des exigences strictes en matière de sécurité, de transparence et de robustesse. Thomas a souligné que cette approche vise non seulement à protéger les utilisateurs, mais aussi à promouvoir l'innovation responsable en harmonisant les règles au sein de l'Union européenne.

Aux États-Unis, l'approche est plus souple, privilégiant une régulation légère axée sur l'autorégulation et les directives administratives. Le décret exécutif récemment signé par l'administration Biden illustre cette tendance, en incitant les entreprises à adopter volontairement des bonnes pratiques tout en fournissant des lignes directrices pour le développement sûr et éthique de l'IA. Cette stratégie, selon Thomas, reflète la volonté des États-Unis de maintenir leur leadership technologique tout en évitant une régulation trop contraignante susceptible de freiner l'innovation.

En Chine, la régulation de l'IA est particulièrement rigoureuse, avec un accent marqué sur la gestion des données et les bonnes pratiques dans le développement et l'utilisation de ces technologies. Le gouvernement chinois a établi un cadre strict pour contrôler les données utilisées dans l'entraînement des modèles et pour encadrer leur déploiement. Thomas a expliqué que cette approche vise à garantir que l'IA s'aligne sur les priorités stratégiques nationales, notamment en matière de sécurité et de souveraineté numérique.

Pour conclure cette partie, Thomas a mis en évidence l'importance croissante de la coopération internationale dans ce domaine. Alors que les cadres réglementaires se développent à des rythmes différents selon les régions, il est essentiel que les acteurs internationaux travaillent ensemble pour établir des normes globales et promouvoir une utilisation responsable de l'IA. Il a également souligné que les entreprises doivent surveiller attentivement ces évolutions, car elles auront un impact direct sur leurs activités, leurs responsabilités et leur compétitivité à l'échelle mondiale.

V. Retours d’expérience terrain

Dans la cinquième partie de son intervention, Thomas Argheria a partagé des retours d’expérience issus de plus de vingt missions réalisées par Wavestone auprès de grandes entreprises. Ces observations mettent en lumière les disparités significatives dans la maturité des organisations face à l’intégration de l’intelligence artificielle (IA), qu’il s’agisse de projets expérimentaux ou d’applications déjà déployées en production.

Thomas a d’abord décrit l’écart important qui existe entre les entreprises en termes de préparation et de maîtrise des risques liés à l’IA. Certaines organisations en sont encore au stade de l’expérimentation, explorant les possibilités offertes par cette technologie, tandis que d’autres ont intégré l’IA dans leurs processus critiques et doivent composer avec les défis associés à une exploitation en production.

Les vulnérabilités observées sur le terrain témoignent des dangers spécifiques liés à l’IA. Parmi les exemples les plus fréquents, Thomas a évoqué l’exposition publique des chatbots, souvent déployés sans mesures de sécurité suffisantes, ce qui les rend vulnérables aux attaques visant à extraire des informations sensibles ou à manipuler leur comportement. Un autre point critique réside dans l’utilisation de jeux de données insuffisamment vérifiés. Lorsque les données d’entraînement ne sont pas rigoureusement contrôlées, elles peuvent contenir des biais ou des erreurs qui affectent les performances du modèle et compromettent sa fiabilité.

Face à ces risques, Thomas a insisté sur la nécessité d’adopter une approche globale pour sécuriser les projets d’IA. Cette approche repose sur trois piliers essentiels.

D’abord, la gouvernance joue un rôle clé. Il s’agit de définir des politiques claires qui intègrent l’IA dans le cadre global de la cybersécurité de l’entreprise. Cela inclut la mise en place de processus pour valider les jeux de données, auditer les modèles et évaluer régulièrement les risques.

Ensuite, la sensibilisation des équipes, qu’elles soient techniques ou non, est indispensable. En développant une meilleure compréhension des enjeux de sécurité liés à l’IA, les collaborateurs deviennent plus à même d’identifier et de signaler les vulnérabilités potentielles.

Enfin, il est crucial de ne pas se reposer uniquement sur les outils natifs de sécurité fournis par les éditeurs de solutions d’IA. Bien que ces outils offrent des fonctionnalités utiles, ils doivent être complétés par des mécanismes spécifiques adaptés aux besoins et aux contextes de chaque entreprise. Cela inclut, par exemple, la mise en place de tests d’intrusion pour évaluer la robustesse des modèles face aux attaques.

Thomas a conclu cette partie en soulignant que la sécurisation de l’IA ne peut pas être un effort isolé. Elle nécessite une collaboration étroite entre les équipes de cybersécurité, les data scientists, et les décideurs, avec une vision partagée de la manière dont l’IA peut être utilisée de manière responsable et sécurisée pour soutenir les objectifs stratégiques de l’organisation.

VI. Intervention de Bénédicte Pilliet, présidente du CyberCercle

Bénédicte PILLIET, présidente du **CyberCercle**, a présenté cette organisation fondée en 2011, dédiée à la promotion de la **sécurité et de la confiance numériques**. Le CyberCercle se concentre sur des dimensions stratégiques, organisationnelles, juridiques et politiques, au-delà des aspects techniques.

Objectifs principaux :

1. Diffuser une **culture de cybersécurité partagée**, intégrée aux stratégies globales.
2. Créer un **cadre de confiance** pour favoriser le dialogue entre acteurs publics et privés.
3. Animer des **communautés d'intérêt** pour développer une intelligence collective.
4. Participer à l'élaboration et à la mise en œuvre des **cadres réglementaires**.

Activités clés :

- **Événements réguliers** : matinales à Paris, rencontres régionales et journées thématiques.
- **Publications** : articles et ouvrages sectoriels (*Regards croisés*).
- **Formation** : sensibilisation intra-entreprise et collaboration avec les élus pour des politiques adaptées.

Vision :

Le CyberCercle met l'accent sur l'intégration de la cybersécurité dans les métiers pour qu'elle devienne un réflexe naturel. L'organisation se veut un **catalyseur de confiance**, soutenant une approche collective et collaborative des enjeux numériques.

Chiffres clés : Plus de 11 000 participants à 133 matinales et 39 journées de rencontre depuis sa création, avec un réseau actif sur tout le territoire.

Bénédicte PILLIET a conclu en rappelant l'importance de "travailler ensemble pour avancer efficacement" et en invitant à participer aux événements et publications du CyberCercle.

VII. Questions / Réponses

Question 1 : Comment répondre aux obligations réglementaires sur la cybersécurité des fournisseurs, s'ils utilisent des IA, sachant qu'il n'y a pas encore de standards de sécurité des IA ?

Pour répondre aux obligations réglementaires en matière de cybersécurité des fournisseurs utilisant l'IA, malgré l'absence de normes spécifiques, les entreprises peuvent s'appuyer sur des guides existants comme ceux du **NCSC** ou de l'**ANSSI**. Ces documents offrent des bonnes pratiques pour sécuriser l'IA. Elles doivent évaluer la maturité des fournisseurs, insérer des clauses de sécurité dans les contrats et interdire les projets jugés trop risqués. En attendant des standards clairs, notamment via l'IA Act européen, ces mesures permettent de poser un socle de sécurité provisoire en vue de normes attendues d'ici 2025.

Question 2 : Pouvez-vous expliquer le fonctionnement d'un "LLM Firewall" et indiquer les technologies ou constructeurs matures disponibles ?

Un "LLM Firewall" est une couche de modération supplémentaire utilisée pour protéger contre les attaques comme l'injection de prompts, mais aussi pour éviter la diffusion de contenus inappropriés (violence, nudité, racisme, etc.).

Son fonctionnement repose sur un modèle d'IA générative qui analyse les entrées (inputs) ou les sorties (outputs) des systèmes d'IA pour les catégoriser. Lorsqu'un contenu problématique est détecté, une réponse standard est générée, bloquant l'affichage ou le traitement de ce contenu. Les filtres en sortie sont particulièrement efficaces, car une fois le contenu produit, il ne peut pas être modifié pour contourner les vérifications. Les filtres en entrée, bien que plus légers, peuvent être contournés par des prompts ingénieusement formulés.

Parmi les constructeurs ou technologies matures, on trouve Calypso AI et Giskard, qui se spécialisent dans la robustesse des modèles, ainsi que des initiatives comme celles de MITRE Security. Ces solutions offrent des outils avancés pour renforcer la sécurité des modèles d'IA et garantir leur conformité avec des normes émergentes.

Question 3 : Quels sont les facteurs clés de succès et les défis à relever pour développer une solution d'IA de confiance ?

Une IA de confiance repose sur une définition claire des piliers qui sous-tendent cette confiance, propre à chaque organisation. Cela inclut la conformité aux exigences réglementaires minimales et, potentiellement, des objectifs plus ambitieux pour se différencier sur le marché. Approcher cette question uniquement sous l'angle du risque serait réducteur. Une IA de confiance doit aussi protéger les utilisateurs et offrir des garanties sur des aspects tels que la transparence, la robustesse et l'impact environnemental.

Pour chaque projet, il est crucial d'évaluer les risques, les objectifs de sécurité, et de mettre en place des processus capables d'arbitrer entre ces différents piliers. Un exemple concret est celui

d'un chatbot public utilisant un modèle de langage (LLM) assisté par une IA générative pour la modération. Bien que cette solution renforce la cybersécurité, elle peut alourdir l'empreinte environnementale, soulevant ainsi des questions sur l'équilibre entre sécurité, performance métier et durabilité. Dans de tels cas, un arbitrage peut conduire à renoncer au projet si les déséquilibres sont trop importants.

En somme, le succès d'une IA de confiance repose sur une gouvernance rigoureuse, des arbitrages équilibrés entre valeur métier, coût de la sécurité et impact environnemental, tout en répondant aux attentes des utilisateurs et aux contraintes réglementaires.

Question 4 : Une IA pourrait-elle permettre de réaliser une cartographie multicouche, allant du matériel jusqu'au profilage des utilisateurs, pour corréler toutes les informations et produire des analyses de risques réellement pertinentes ?

Ce type d'IA reste pour l'instant au stade de la science-fiction, mais il est intellectuellement réalisable à terme. À ce jour, les cas d'usage observés se concentrent sur des systèmes utilisant des analyses de risque existantes pour automatiser ou semi-automatiser des évaluations sur des périmètres similaires. Certaines IA accompagnent également les équipes en simulant des chemins d'attaque ou en accélérant la détection des vulnérabilités.

Un exemple notable est celui d'un concours organisé par la DARPA en 2016, mettant en compétition des superordinateurs capables de s'attaquer et de se défendre en autonomie tout en rendant des services aux utilisateurs. Ces machines ont adopté diverses stratégies, certaines axées sur l'attaque, d'autres sur la défense, et certaines équilibrées entre détection, correction et attaque. Ce type d'événement montre que des systèmes d'IA autonomes, même limités aujourd'hui, peuvent évoluer rapidement et inspirer des développements futurs.

Bien que des outils d'automatisation de la cybersécurité soient déjà opérationnels, une cartographie complète et multicouche reste un objectif ambitieux pour les décennies à venir, dans la continuité des avancées déjà observées dans des domaines comme les véhicules autonomes.

Question 5 : L'intelligence artificielle peut-elle être utile pour maintenir les SBOM (Software Bill of Materials) ?

L'utilisation de l'IA pour maintenir les SBOM est tout à fait pertinente et présente plusieurs avantages. D'abord, l'automatisation, permise par l'IA, peut simplifier la collecte et la mise à jour des informations sur les composants logiciels, réduisant ainsi les erreurs humaines et accélérant les processus. Ensuite, l'IA peut également être utilisée pour analyser les vulnérabilités, comme le discovery de vulnérabilités zero-day, en identifiant des failles connues dans les composants logiciels, puis en proposant des correctifs ou des mises à jour.

De plus, l'IA peut aider à gérer les dépendances complexes entre les composants logiciels, en alertant les développeurs en cas de conflits ou de mises à jour nécessaires, ce qui est particulièrement pertinent dans des cas comme les vulnérabilités de type Log4j. Certaines de ces applications sont déjà en cours de développement et montrent un potentiel prometteur pour optimiser la gestion des SBOM et renforcer la sécurité logicielle.

Question 6 : Peut-on faire confiance à l'IA des outils Microsoft et à Copilot ?

La confiance envers les outils IA de Microsoft et Copilot dépend de plusieurs facteurs. Tout d'abord, il est essentiel de comprendre les engagements en matière de sécurité pris par le fournisseur. Microsoft, par exemple, garantit la gestion des vulnérabilités, le maintien à jour des modèles et le stockage des données, notamment dans des centres de données européens pour respecter les réglementations locales.

Cependant, la sécurité ne repose pas uniquement sur le fournisseur. La gouvernance, la sécurisation des cas d'usage, la gestion des interconnexions et des plugins relèvent de la responsabilité de l'utilisateur ou de l'entreprise qui exploite la plateforme. Concernant Copilot, les données utilisées dans les prompts ne sont pas réintégrées dans l'apprentissage du modèle, offrant un niveau de sécurité similaire à celui des services comme SharePoint. En résumé, si l'on fait confiance à Microsoft pour des outils comme SharePoint, Copilot peut être considéré comme tout aussi fiable, sous réserve de bien respecter les règles de gouvernance et de sécurisation.

Question 7 : Peut-on bloquer via Netskope tous les outils d'IA générative présentant un risque potentiel d'exfiltration de données ?

Il est techniquement possible de bloquer ces outils via Netskope, mais cela pose des défis. La méthode la plus stricte consiste à blacklister manuellement tous les sites d'IA générative, ce qui nécessite une surveillance constante des nouvelles plateformes. Cependant, cette solution est souvent trop restrictive et peut entraver les activités métiers.

Une alternative consiste à mettre en place des avertissements via le proxy. Par exemple, lorsque l'utilisateur accède à un site identifié comme potentiellement risqué, un message peut rappeler de ne pas utiliser de données sensibles ou confidentielles. Cette approche a été adoptée par certaines entreprises, comme Wavestone, pour des outils comme ChatGPT.

Enfin, la détection automatique des fuites via des règles de monitoring reste complexe et peut générer de nombreux faux positifs. Une combinaison de sensibilisation, de blocage ciblé et d'avertissements reste donc la meilleure approche pour minimiser les risques tout en préservant l'efficacité opérationnelle.

Question 8 : Quels sont les points de chevauchement entre ISO 38507 et ISO 27014, et comment ces deux normes peuvent-elles être utilisées ensemble pour renforcer la gouvernance dans un environnement numérique intégrant l'IA et la cybersécurité ?

Bien que l'ISO 38507 (gouvernance de l'IA) et l'ISO 27014 (gouvernance de la sécurité de l'information) visent des aspects distincts, elles n'ont pas de chevauchements significatifs dans leur contenu actuel. L'ISO 38507 se concentre sur la gouvernance des systèmes d'IA en mettant en avant des directives sur l'alignement stratégique, les responsabilités et la gestion des risques spécifiques à l'IA. En revanche, l'ISO 27014 traite de la gouvernance de la cybersécurité, incluant la gestion des risques et la mise en œuvre de contrôles de sécurité pour protéger les actifs informationnels.

Conclusion

Pour conclure ce compte-rendu, la session du « Lundi de la cybersécurité » du 9 décembre 2024 a permis de mettre en lumière les enjeux critiques liés à la sécurisation des systèmes d'intelligence artificielle dans un contexte de transformation numérique rapide. Thomas a exploré des thématiques variées, allant des défis techniques et stratégiques liés à l'IA jusqu'à l'importance d'une gouvernance efficace et d'une collaboration entre acteurs publics et privés.

Les retours d’expérience, les bonnes pratiques partagées et les discussions autour des réglementations internationales ont permis de dégager des perspectives concrètes pour 2025, notamment en matière de formation, d’évaluation des risques et de sécurisation des projets IA.

L’accent a également été mis sur la nécessité d’une approche proactive et collective, rassemblant les efforts des experts en cybersécurité, des data scientists, et des décideurs.

Ainsi, cet événement constitue une étape clé pour sensibiliser et outiller les entreprises à l’intégration de l’IA de manière sécurisée, en harmonie avec les évolutions réglementaires et les exigences stratégiques croissantes.