RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

COMMUNIQUÉ DE PRESSE

Paris, le 12/02/2025

Sommet pour l'action sur l'intelligence artificielle : l'ANSSI invite à objectiver les risques et à se saisir des opportunités des technologies d'IA pour la cybersécurité

Dans le cadre du Sommet pour l'action sur l'intelligence artificielle (IA), qui s'est déroulé à Paris du 6 au 11 février 2025, l'Agence nationale de la sécurité des systèmes d'information (ANSSI) a présenté les conclusions des travaux menés ces derniers mois sur la cybersécurité avec ses partenaires nationaux et internationaux, au sein de l'axe « IA de confiance ». L'occasion pour l'Agence de promouvoir une approche fondée sur l'analyse des risques et des opportunités des technologies d'intelligence artificielle pour développer la confiance dans ces systèmes.

Des systèmes d'information qui posent de nouveaux défis à la cybersécurité

Dans les travaux qu'elle a menés, l'ANSSI souligne en premier lieu que les systèmes intégrant une IA (SIA) **demeurent fondamentalement des systèmes logiciels**, soumis en tout état de cause aux mêmes vulnérabilités que des systèmes plus classiques, comme le détournement de comptes utilisateurs ou administrateurs ou l'exploitation de vulnérabilités dans les composants logiciels intégrés dans le système. Les scénarios de cyberattaque classiques restent ainsi aujourd'hui les plus crédibles contre de tels SIA. **Le respect des bonnes pratiques de cybersécurité applicables à tout système d'information, est ainsi un enjeu primordial pour la mise en œuvre d'une IA de confiance.**

Au-delà des vulnérabilités usuelles, les SIA peuvent également être mis en défaut par de nouveaux vecteurs d'attaque, propres à l'IA, de trois natures :

- **L'empoisonnement** qui consiste en une altération des données d'entraînement ou du modèle affectant la réponse du SIA à toutes les entrées ou à une entrée spécifique, en injectant notamment des fausses informations dans les paramètres du modèle pour l'induire volontairement en erreur. Exemple : dans un entrepôt sécurisé, un attaquant pourrait « empoisonner » le SIA utilisé par la vidéo surveillance, en modifiant sa capacité de détection. Le système associerait alors un individu avec une tenue d'une certaine couleur comme « non suspect » et de facto n'activerait pas le système d'alarme. Il convient cependant de noter qu'un tel scénario d'attaque serait particulièrement complexe à mettre en œuvre, et nécessiterait de la part de l'attaquant un travail d'anticipation très conséquent.
- **L'extraction** qui induit la reconstruction ou la récupération de données confidentielles du SIA ou du modèle après la phase d'apprentissage, et par laquelle l'attaquant peut avoir accès aux

données d'entrainement ou aux paramètres et configurations du modèle. Exemple : à des fins d'espionnage économique, un attaquant pourrait extraire du SIA d'une société du secteur pharmaceutique des données hautement sensibles (formules d'un vaccin ou brevet d'un médicament, etc.).

- **L'évasion** qui consiste en une altération des données d'entrée afin de modifier le fonctionnement attendu du SIA, et par laquelle l'attaquant pourrait détourner l'analyse faite par le SIA afin de l'induire en erreur et changer volontairement sa compréhension. Exemple : sur la fraude bancaire, un acteur malveillant pourrait modifier les paramètres d'un virement pour que ce dernier ne soit pas catégorisé comme frauduleux par le SIA, dont l'intégrité n'a pourtant pas été altérée.

Ces attaques pourraient engendrer des atteintes à la disponibilité ou à l'intégrité, ce qui compromettrait la fiabilité des décisions, voire des risques de vol ou de divulgation de données sensibles avec les risques de confidentialité inhérents. Bien que **moins crédibles, dans un avenir proche, que des attaques plus classiques**, ces nouveaux vecteurs d'attaque ont la particularité de ne pouvoir être correctement prise en compte que dans une approche de cybersécurité étendue à l'ensemble de cycle de vie et de la chaîne de valeur de ces systèmes.

Enfin, il est important de noter la difficulté particulière d'identifier la cause racine d'une erreur ou d'un comportement inattendu dans une IA. Il en découle une **difficulté particulière pour l'investigation et la remédiation suite à une suspicion d'attaque affectant une IA**. Cet état de fait doit être intégré dans l'architecture même des systèmes d'information intégrant un composant à base d'IA, pour assurer une **résilience suffisante face à la possibilité d'une indisponibilité prolongée, ou d'une remise en cause de la confiance dans ce composant d'IA** qui ne pourrait pas être levée immédiatement par une analyse de cybersécurité.

La réponse à l'ensemble de ces enjeux nécessite un dialogue renforcé entre les communautés de l'IA et de la cybersécurité, une poursuite des efforts d'objectivation des risques et des opportunités de l'IA, en se fondant sur les faits et sur une approche scientifique, ainsi qu'un approfondissement des travaux de recherche, notamment dans le domaine de l'évaluation de l'IA et de l'interprétabilité de ses résultats.

Une vingtaine de pays et des centaines d'experts en IA et cyber réunis pour le développement d'une IA de confiance

Pour guider notamment les dirigeants et les producteurs de solutions d'IA, plusieurs livrables ont été dévoilés à l'occasion du Sommet :

- Le document de référence « *Building trust in AI through a cyber risk approach* » : co-signé par 19 partenaires internationaux et 5 partenaires institutionnels (AMIAD – Agence ministérielle pour l'intelligence artificielle de Défense, CNIL – Commission nationale de l'informatique et des libertés, INRIA – Institut national de recherche en sciences et technologies du numérique, LNE – Laboratoire national de métrologie et d'essais, et PEReN – Pôle d'Expertise de la Régulation Numérique) et présenté le 7 février lors des journées scientifiques organisées par l'Institut Polytechnique de Paris, il met en évidence les risques cyber auxquels sont exposés les systèmes d'IA et propose des recommandations stratégiques afin de favoriser une meilleure prise en compte de la cybersécurité dans le développement et l'intégration de ces systèmes. Cette analyse de risque a été élaborée suite aux consultations de plusieurs entités publiques et privées matures sur le sujet de l'IA.

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

- L'exercice de crise cyber : organisé par l'ANSSI en collaboration avec le Campus Cyber et plusieurs de ses membres, le 11 février 2025, cet exercice a mobilisé près de 250 participants experts cyber et experts IA dans l'objectif de développer une meilleure compréhension mutuelle de leurs attentes. Alors que le scénario proposé comprenait une cyberattaque affectant un SIA et sa chaine d'approvisionnement (*supply chain*), les participants ont pu échanger des bonnes pratiques tout en étant sensibilisés aux risques cyber. Un retour d'expérience sera prochainement organisé et partagé publiquement.
- La rencontre des directeurs d'agences cyber partenaires : en parallèle du Sommet, l'ANSSI et ses partenaires internationaux ont pu contribuer activement aux débats en cours sur la sécurisation de l'IA. Un point d'étape a été réalisé sur les initiatives en cours et les priorités partagées. L'Agence et ses partenaires ont à nouveau fait part de leur engagement afin de parvenir à une IA de confiance à l'échelle internationale.

« Le temps fort qu'a constitué le Sommet pour l'action sur l'IA a démontré la nécessité de mieux faire travailler ensemble les experts de l'IA et ceux de la cybersécurité. Ce dialogue doit permettre de mieux objectiver les risques et opportunités de ces systèmes, loin des marchands de peur et des marchands de rêve. Ce n'est qu'ainsi que nous pourrons collectivement déployer des IA de confiance, en intégrant ces enjeux cyber dès la conception des solutions. » commente Vincent Strubel.

À PROPOS DE L'ANSSI
L'Agence nationale de la sécurité des systèmes d'information (ANSSI) a été créée par le décret n°2009-834 du 7 juillet 2009 sous la forme d'un service à compétence nationale.
L'agence est l'autorité nationale en matière de cybersécurité et de cyberdéfense. Elle est rattachée au Secrétaire général de la défense et de la sécurité nationale (SGDSN), sous l'autorité du Premier ministre.

AGENCE NATIONALE DE LA SÉCURITÉ DES SYSTÈMES D'INFORMATION
ANSSI - 51, boulevard de la Tour-Maubourg - 75700 PARIS 07 SP

cyber.gouv.fr

Contacts Presse

presse@ssi.gouv.fr
06 49 21 63 80 / 06 49 87 30 36

Roxane ROSELL
roxane.rosell@ssi.gouv.fr