MINISTÈRE
DES ARMÉES

Liberté
Égalité
Fraternité

COMMUNIQUÉ DE PRESSE
DU MINISTÈRE DES ARMÉES

Paris, le 30 avril 2025

Sécurisation de l'intelligence artificielle : deux solutions innovantes distinguées par le ministère des Armées

- **Ce mercredi 30 avril 2025 a eu lieu la remise des prix du défi « Sécurisation de l'IA » en présence du général de corps d'armée Aymeric Bonnemaison, commandant de la cyberdéfense (COMCYBER) et de l'ingénieur général de l'armement, Patrick Aufort, directeur de l'Agence de l'innovation de Défense (AID), aux côtés des experts de l'Agence ministérielle pour l'IA de défense (AMIAD) et de la Direction générale de l'armement (DGA)**
- **Ce défi, organisé par l'AID et le COMCYBER, a pour objectif de solliciter les entreprises et laboratoires de recherche sur des solutions innovantes permettant de sécuriser les systèmes incluant une intelligence artificielle.**

L'intelligence artificielle est un enjeu majeur pour les forces armées. En tant que système d'information, assurer sa sécurisation est indispensable. Le Commandement de la cyberdéfense (COMCYBER) s'engage aujourd'hui aux côtés de l'Agence de l'innovation de Défense (AID) dans ce défi innovant ; celui d'identifier et d'évaluer l'intérêt de technologies permettant de se protéger et de détecter des attaques sur les systèmes incluant l'IA.

Placer sous les maître-mots "Se protéger, se défendre", ce défis vise à assurer l'intégrité des systèmes tout au long de leur vie opérationnelle. Au total, plus d'une dizaine de partenaires (laboratoires, start-ups, PME, ETI ou grands groupes) ont participé à ce défi amenant plusieurs solutions innovantes. Deux d'entre eux se sont particulièrement distingués et ont été primés par le jury, comprenant outre le COMCYBER et l'AID, des experts de l'AMIAD et de la DGA.

Suite d'outils BET ("Behavior Elicitation Tool")

La suite d'outils BET, développée par la société PRISM EVAL, vise à sécuriser les LLMs (large language models). Ces modèles sont par exemple utilisés pour des chatbots à base d'IA comme Le Chat, Chat GPT ou GenIAlly, l'outil interne du ministère des Armées. La PME a ainsi développé un premier outil, BET Eval, pour tester la robustesse des LLMs, en s'appuyant sur une approche innovante à base de combinaison de primitives comportementales d'attaques, qui permet d'adresser un large panel d'attaques. Les tests réalisés par l'outil portent sur la génération de contenus malveillants ou dangereux, sur l'extraction non autorisée d'informations sensibles, sur le contournement des garde-fous. Des outils complémentaires de protection des modèles par rapport à ces attaques sont en cours de développement par la PME.

PyRAT et PARTICUL

Les solutions du CEA-List se concentrent sur les attaques adversariales visant, en modifiant les données d'entrée, à provoquer une mauvaise réponse d'un système à base d'IA, par exemple un système de classification d'images. Pour cela, il s'appuie sur la combinaison de deux outils : l'un, PyRAT, assure une vérification formelle de la sécurité d'un réseau de neurones apportant des garanties mathématiques fortes, pour se prémunir de modifications imperceptibles des données par l'attaquant ; l'autre, PARTICUL, détecte les parties récurrentes d'un ensemble de données pour calculer un score de confiance sur de nouvelles données, et permet de se prémunir de modifications plus visibles tels que l'ajout de patchs.

Contacts médias :

Commandement de la cyberdéfense
bureau communication
comcyber.communication.fct@intradef.gouv.fr
09 88 68 51 63
09 88 68 27 49

Agence innovation Défense
Pôle rayonnement/éditorial & presse
helene.defleur@intradef.gouv.fr
06 38 47 92 83

Centre médias du ministère des Armées
media@dicod.fr
09 88 67 33 33

Délégation à l'information et à la communication de la défense
DICoD
Centre médias du ministère des Armées
60, boulevard du général Martial Valin
CS 21623 - 75009 Paris Cedex 15