N°6 - JUILLET/SEPTEMBRE 2025

CYBER-IT MAGAZINE

LA CYBER EST UN MARATHON PAS UN SPRINT !

HISTOIRE
80 ans d'IA entre rêves et désillusion

EUROPE ET IA
Vers une souveraineté de l'Europe en matière d'IA

AI ACT
Le règlement européen sur l'intelligence artificielle

IA CONTRE IA
Intelligence vs IA la nouvelle frontière

RÉSILIENCE CYBER
Prédire l'incident avant qu'il ne survienne

ÉTHIQUE ET IA
Comment créer une IA éthique et inclusive ?

DOSSIER SPECIAL

Intelligence artificielle

Une question ? Un besoin ?
cyberit.magazine@gmail.com

EDITO

L'intelligence artificielle, fantasme de science-fiction hier, est devenue l'un des enjeux les plus stratégiques de notre époque. À travers notre dossier spécial, nous retraçons 80 ans de rêves et de désillusions autour de l'IA.

Une histoire faite d'espoirs, de percées spectaculaires, mais aussi de crises de confiance et de promesses non tenues. Aujourd'hui, l'IA ne se contente plus d'être un objet de recherche elle devient un acteur. Et quel acteur ! Intelligence contre IA, c'est le nouveau théâtre d'opérations. Car dans le cyberespace, l'IA est à la fois bouclier et épée. Elle protège, anticipe, corrige, mais elle est aussi utilisée par des groupes malveillants qui l'exploitent pour décupler leurs capacités offensives.

Mais si l'IA est une force, quelle force voulons-nous façonner ? Les algorithmes influencent nos choix, nos libertés et nos droits, l'éthique n'est plus une option. Construire une IA inclusive, responsable et transparente devient une priorité pour éviter que les erreurs du monde réel ne se reproduisent ou ne s'amplifient dans le monde numérique.

Face à cette réalité, l'Europe s'organise. Avec le AI Act, elle prend une longueur d'avance en matière de régulation. Mais cette ambition est-elle suffisante ? Peut-on vraiment parler de souveraineté numérique européenne face aux géants américains et aux puissances chinoises ? Et surtout, cette souveraineté est-elle encore atteignable ?

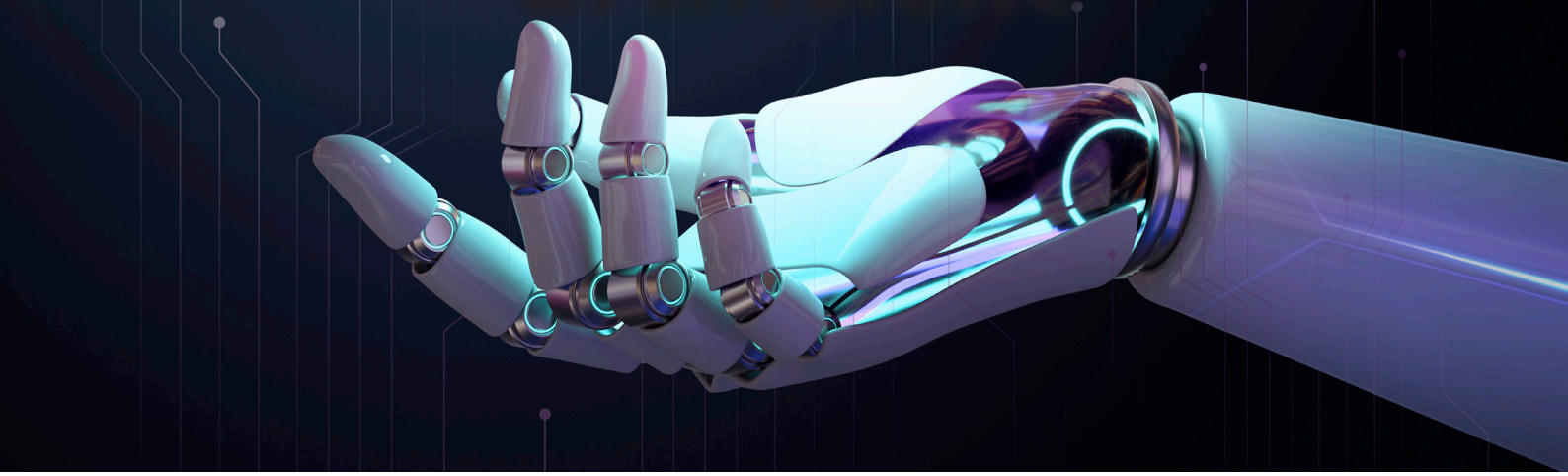
Enfin, dans cette ère où les incidents de cybersécurité ne sont plus une question de "si", mais de "quand", nous explorons les leviers de la cyber résilience. Une IA bien conçue peut non seulement détecter les menaces, mais aussi les prévenir avant qu'elles ne surviennent.

Bonne lecture à tous !

ARNAUD LEROY

SOMMAIRE

L'intelligence artificielle



80 ans d'IA - Entre rêves et désillusions

L'intelligence artificielle n'est pas une idée neuve. Mais en plus de soixante-dix ans, elle a connu une métamorphose spectaculaire. Une évolution qui bouscule les certitudes, bouleverse les usages, et redéfinit les enjeux de cybersécurité.

En 1950, **Alan Turing** posait une question assez provocatrice : « **Les machines peuvent-elles penser ?** »

Sept décennies plus tard, cette interrogation n'est plus seulement philosophique. L'intelligence artificielle innonde nos systèmes industriels, influence nos décisions financières, pilote des véhicules, génère du code, rédige des rapports, imite nos voix et infiltre nos systèmes d'information. De discipline marginale confinée aux laboratoires académiques, elle est devenue une technologie stratégique au cœur des préoccupations politiques, économiques et sécuritaires du XXIe siècle.

L'IA n'a pas évolué de façon linéaire. Elle a traversé des périodes d'euphorie, de grands calmes et de réinventions soudaines. Elle s'est nourrie de symboles, puis de données. Elle a abandonné les règles pour les modèles. Chaque décennie a façonné un pan de cette révolution silencieuse, aujourd'hui jusque dans les centres opérationnels de cybersécurité, les algorithmes de surveillance, et les arsenaux numériques des cybercriminels.

Ce dossier retrace les grandes étapes de cette évolution : des systèmes symboliques aux réseaux neuronaux profonds, de l' IA génératives, en passant par leur double usage protecteur ou malveillant dans la cyber.

1950-1960 : L'idée naît dans les laboratoires

Tout commence avec des penseurs comme Alan Turing, qui propose en **1950** le fameux test de Turing, censé déterminer si une machine peut imiter l'intelligence humaine. En **1956**, la conférence de Dartmouth marque le véritable acte de naissance de l'IA. Les chercheurs pensent pouvoir créer une intelligence comparable à celle de l'homme en une génération.

Les premiers programmes emblématiques de l'intelligence artificielle voient le jour dans cette décennie fondatrice. En **1955**, **Logic Theorist**, développé par Allen Newell et Herbert Simon, est présenté comme le premier programme d'IA capable de raisonner de manière autonome. Il résout des théorèmes en logique mathématique, imitant la démarche humaine. Ce programme marque un tournant : la machine n'est plus seulement une calculatrice, elle tente de "penser".

Quelques années plus tard, dans les années **1960**, **ELIZA**, mis au point par Joseph Weizenbaum au MIT, devient le premier programme de traitement du langage à interagir avec des humains. Il simule un psychothérapeute en reformulant les phrases de l'utilisateur sous forme de questions ouvertes. Derrière sa simplicité technique basée sur des règles de correspondance de mots-clés ELIZA démontre l'impact émotionnel que peut produire une interaction homme-machine. Certains utilisateurs, pourtant informés qu'il s'agissait d'une machine, ont développé un attachement au programme.

Ces deux programmes, bien que limités, posent les fondations des deux grandes orientations de l'IA : le raisonnement logique formel, et l'interaction en langage naturel. Ils illustrent déjà l'ambition et les limites d'une intelligence artificielle naissante, encore loin d'être consciente mais déjà capable de simuler des comportements humains rudimentaires.

1960-1970 : L'optimisme et les limites

Les systèmes logiques et symboliques se développent rapidement. L'un des exemples les plus emblématiques de cette période est **SHRDLU**, développé entre **1968 et 1970** par Terry Winograd au **M**assachusetts **I**nstitute of **T**echnology.

Le programme est capable de comprendre des instructions en anglais, de manipuler des objets dans un univers fictif simple, et de dialoguer avec l'utilisateur. Il peut, par exemple, répondre à des commandes comme « mets le cube rouge sur le cylindre vert » ou expliquer ses propres actions. SHRDLU donne l'illusion d'une compréhension réelle du langage humain. Mais cette réussite est confinée à un univers artificiellement limité et extrêmement structuré. Hors de ce microcosme, le programme est incapable de généraliser ou de faire preuve de flexibilité.

Néanmoins, il alimente un optimisme débordant : on pense alors que la compréhension complète du langage naturel, la traduction automatique entre langues humaines, et même une forme générale d'intelligence sont à portée de main. Cet enthousiasme est partagé et financé par le gouvernement américain. La DARPA (Defense Advanced Research Projects Agency) soutient des projets d'IA de grande ampleur. Des budgets colossaux sont accordés aux universités pour bâtir des systèmes capables de dialoguer, d'apprendre, voire de raisonner.

Mais les limites se font vite ressentir : la complexité du langage humain, la pauvreté des bases de données, et le manque de puissance de calcul ralentissent les progrès

1970-1980 : L'âge de la logique et des systèmes experts

L'IA symbolique, fondée sur la manipulation de symboles et de règles logiques explicites, continue de dominer la recherche. On entre alors dans l'ère des premiers "systèmes experts" — des programmes capables de simuler le raisonnement d'un spécialiste humain dans un domaine restreint.

L'un des cas les plus marquants est **MYCIN** en **1972**, développé à l'Université Stanford pour assister les médecins dans le diagnostic et le traitement des infections bactériennes. MYCIN pouvait poser des questions, interpréter les réponses, et recommander un antibiotique avec un degré élevé de précision pour l'époque. Il utilisait une base de connaissances constituée de plusieurs centaines de règles "si-alors" codées à la main par des experts.

Mais ces systèmes se heurtent rapidement à leurs propres limites. Premièrement, ils sont incapables d'apprendre de nouvelles règles ou de s'adapter à des données inédites : toute évolution nécessite une intervention humaine pour reformuler ou ajouter manuellement des règles.

Deuxièmement, leur maintenance devient un casse-tête : chaque modification peut avoir des effets de bord, et plus la base de règles grandit, plus il devient difficile de prévoir les interactions entre elles. Enfin, dès que le contexte d'utilisation sort du cadre strictement défini pour lequel ils ont été conçus, ces systèmes s'effondrent : ils ne savent pas gérer l'incertitude, les cas ambigus, les contradictions ou les informations incomplètes.

L'intelligence qu'ils simulent repose uniquement sur un raisonnement sans aucune forme d'intuition, d'apprentissage inductif ou de généralisation. Ils peuvent donner l'illusion de performance dans des environnements fermés, bien définis, mais montrent vite leurs limites face à la complexité du monde réel.

1980-1990 : Le boom et l'effondrement des systèmes experts

Les systèmes experts connaissent un véritable essor durant les années **1980**, stimulés par les succès académiques précédents et par les avancées en ingénierie logicielle. L'idée séduit : formaliser le savoir d'un expert humain sous forme de règles logiques permettrait de reproduire son raisonnement, 24 heures sur 24, sans erreurs humaines, ni fatigue.

Des secteurs comme la finance, la médecine, la production industrielle ou l'aéronautique se lancent dans des projets ambitieux, espérant automatiser tout ou partie de la prise de décision.

Des langages comme **OPS5** ou **Prolog**, et des plateformes dédiées comme **XCON** (utilisé par DEC pour la configuration automatique de ses ordinateurs), font office de vitrines technologiques. On assiste à la naissance de nombreuses start-ups spécialisées en IA, soutenues par des fonds d'investissement confiants. Les grandes entreprises créent des divisions IA internes. Les promesses sont nombreuses : réduction des coûts, fiabilité accrue, gain de compétitivité. Mais rapidement, les limites pratiques s'imposent. Développer un système expert exige un travail colossal d'ingénierie des connaissances. Dès qu'un contexte sort du cadre prévu, le système s'effondre. À cela s'ajoutent des coûts de maintenance élevés et des résultats qui peinent à convaincre face à la complexité du monde réel.

Face à cette désillusion, les investisseurs se retirent progressivement. les entreprises ferment leurs divisions IA. Ce recul généralisé marque le premier véritable "hiver de l'IA"

1990-2000 : L'éveil du machine learning

Le paradigme change radicalement : on ne tente plus de coder explicitement les règles de l'intelligence, mais de les faire émerger à partir des données.

Ce changement de cap marque l'avènement du machine learning, une approche fondée sur des algorithmes capables de repérer des patterns dans de vastes jeux de données.

L'apprentissage supervisé, où l'on fournit à la machine des exemples annotés, permet de créer des modèles prédictifs pour des tâches précises

Les machines ne sont plus de simples exécutantes logiques elles commencent véritablement à apprendre, à s'adapter, à généraliser à partir de l'expérience. Ce tournant fondamental, encore discret pour le grand public à l'époque, installe les fondations de l'intelligence artificielle moderne.

On assiste à une explosion de travaux académiques sur la validation croisée, le sur-apprentissage, la régularisation autant de concepts qui posent les bases de la rigueur statistique dans le domaine.

La victoire de **Deep Blue** contre Garry Kasparov en **1997**, première défaite d'un champion du monde d'échecs face à un ordinateur dans un match officiel, marque un tournant historique : pour la première fois, une machine surpasse un humain dans un domaine symbolique de l'intelligence.

Cet événement hautement médiatisé propulse l'intelligence artificielle sur le devant de la scène et dans la conscience du grand public.

Il ne s'agit plus seulement de recherches abstraites ou de démonstrations académiques, l'IA entre dans l'imaginaire collectif comme une force capable de rivaliser, voire de dominer, dans des domaines cognitifs complexes.

2000-2010 : Internet, le big data et l'IA à grande échelle

Avec l'essor d'Internet et l'avènement du **Web 2.0**, marqué par l'interactivité, les réseaux sociaux et la création massive de contenus par les utilisateurs, l'IA trouve un terrain de jeu sans précédent.

Les géants du numérique Google, Amazon, Facebook, Netflix, capitalisent sur cette révolution pour développer des algorithmes de recommandation toujours plus fins, capables de prédire les goûts, d'adapter l'expérience utilisateur en temps réel, et d'optimiser la publicité ciblée.

Le traitement automatique du langage naturel progresse rapidement, cette évolution permet des avancées notables dans la compréhension et la génération de texte, bien que les systèmes restent encore limités à des tâches spécifiques.

C'est durant cette période que naissent les premiers assistants vocaux, encore rudimentaires, et que les moteurs de recherche gagnent en pertinence grâce à des algorithmes linguistiques plus fins, capables d'interpréter l'intention des requêtes.

Parallèlement, l'explosion du volume de données disponible sur le Web transforme radicalement les capacités d'apprentissage des modèles. Ce phénomène, baptisé « **big data** », ouvre la voie à une IA plus empirique et moins dépendante d'une modélisation humaine explicite.

La collecte et l'analyse de ces masses de données deviennent possibles grâce à des infrastructures distribuées comme **Hadoop** ou **MapReduce**, qui permettent de traiter en parallèle d'énormes quantités d'informations. Ces outils posent les bases d'une IA à grande échelle,

2010-2020 : Deep Learning et percées spectaculaires

Les réseaux de deep learning révolutionnent le paysage de l'intelligence artificielle en débloquant des performances jusque-là inaccessibles. Grâce à la disponibilité de grandes quantités de données annotées et à la puissance de calcul des GPU, ces architectures permettent de dépasser les limites des approches traditionnelles.

En **2012**, le modèle **AlexNet**, vainqueur du concours ImageNet, marque une rupture décisive dans la vision par ordinateur, divisant par deux les taux d'erreur en reconnaissance d'images.

ResNet, en **2015**, introduit les connexions résiduelles et rend possible l'entraînement de réseaux de plusieurs centaines de couches, ouvrant la voie à une profondeur sans précédent

Ces percées se diffusent rapidement grâce à l'essor de plateformes open source comme TensorFlow de Google ou PyTorch de Meta, qui rendent l'IA accessible à une vaste communauté de chercheurs, développeurs et entreprises.

L'écosystème se démocratise, accélérant la recherche, favorisant la reproductibilité scientifique, et facilitant l'intégration de l'IA dans des applications industrielles à grande échelle.

En quelques années, la reconnaissance vocale devient courante sur smartphones, la traduction automatique rivalise avec les humains sur certaines langues, et la vision artificielle s'invite dans la médecine, l'automobile et la sécurité.

Dans le domaine du langage naturel, **BERT** en **2018**, basé sur l'architecture Transformer, permet une compréhension contextuelle fine des mots, transformant la recherche d'information, la classification de texte et la réponse automatique aux questions.

2020-2030 : L'intelligence générative et l'IA omniprésente

Les années **2020** marquent une rupture spectaculaire dans l'évolution de l'intelligence artificielle, avec l'essor fulgurant des modèles génératifs.

Portés par des architectures comme le Transformer, ces systèmes tels que **ChatGPT** (OpenAI), **Claude** (Anthropic), **Gemini** (Google) ou **LLaMA** (Meta) repoussent les frontières de l'automatisation cognitive.

Leur prouesse réside dans leur capacité à produire, sur demande, du texte, du code, des images, des vidéos, de la voix, et même des interfaces interactives, avec une qualité perçue souvent équivalente à celle d'un expert.

On ne parle plus seulement d'outils spécialisés, mais de véritables assistants généralistes, capables d'effectuer des tâches transversales : rédaction d'articles, analyse juridique, conception de scripts, élaboration de business plans, traduction, génération d'illustrations, composition musicale…

Ces modèles dits fondamentaux (foundation models) sont préentraînés sur des corpus gigantesques, puis capables d'être adaptés à des cas d'usage spécifiques avec peu d'exemples supplémentaires.

Elle transforme également le monde de la création : les illustrateurs, designers, vidéastes ou compositeurs s'en servent comme catalyseurs d'inspiration ou outils de prototypage.

Mais cette expansion s'accompagne d'enjeux inédits. Les risques liés à la désinformation automatique, aux deepfakes, aux hallucinations (contenus inventés par les modèles), ou à la perte de traçabilité des sources inquiètent autant qu'ils fascinent.

2030 et au-delà : IA émotionnelle, autonomie renforcée et bataille cognitive

À l'horizon **2040** et au-delà, les avancées en IA ne se contenteront plus d'imiter l'intellect humain : elles s'efforceront d'en intégrer les dimensions affectives, relationnelles et sensorielles.

Les modèles dits émotionnels seront capables de détecter, simuler et répondre aux états émotionnels de leurs utilisateurs. Intégrés dans les interfaces homme-machine, les robots compagnons, les agents conversationnels ou les assistants de soins, ces IA émotionnelles adapteront leur comportement au contexte psychologique et social de l'interlocuteur.

L'objectif sera de rendre la communication plus naturelle, plus empathique, voire thérapeutique. Ces IA s'appuieront sur des capteurs biométriques (rythme cardiaque, micro-expressions, ton de la voix) et sur des bases de données comportementales massives pour moduler leurs réponses.

Parallèlement,l'autonomie des systèmes se généralise : véhicules, drones, assistants domotiques, systèmes industriels opèrent sans supervision humaine, prenant des décisions en temps réel à partir d'analyses de données environnementales.

L'IA devient ainsi un agent décisionnel à part entière dans de nombreux domaines stratégiques. Les enjeux éthiques se multiplient : jusqu'où laisser une machine décider ? Comment garantir la traçabilité, la responsabilité, la non-discrimination ?

Les IA seront utilisées non seulement pour attaquer ou défendre des systèmes, mais pour manipuler les perceptions, diffuser des narratifs synthétiques personnalisés, et générer des contenus falsifiés indistinguables du réel.

Des modèles adverses s'affronteront dans l'ombre : l'un pour tromper, l'autre pour détecter. La frontière entre vérité et simulation deviendra floue. La société entrera dans une phase de méfiance algorithmique généralisée.

Illustration selon ChatGPT pour résumer ce sujet :

Intelligenc e contre IA la nouvelle frontière

"Le cybercriminel n'écrit plus de code. Il commande des intelligences comme un général ses troupes"

La nouvelle frontière de la cybersécurité

Les lignes de front de la cybersécurité sont en train de se redessiner. Longtemps perçue comme une discipline technique réservée aux ingénieurs en réseau et aux spécialistes en chiffrement, la cybersécurité entre aujourd'hui dans une nouvelle ère, dominée par les dynamiques de l'intelligence artificielle.

Ce qui était hier un outil d'optimisation et d'automatisation devient désormais une arme offensive ou défensive à part entière.

Nous vivons une transition silencieuse mais brutale : les menaces ne sont plus seulement humaines. Les attaques ne sont plus manuelles. Les défenses ne peuvent plus être uniquement réactionnelles.

L'intelligence artificielle, dans ses formes les plus avancées, est désormais au cœur du champ de bataille. Et, pour la première fois dans l'histoire du cyberespace, nous assistons à une situation inédite des IA conçues pour attaquer, traquées, contrées ou neutralisées par d'autres IA.

La machine ne protège plus seulement l'homme, elle combat aussi d'autres machines. La montée en puissance de l'IA offensive marque un tournant. Les cybercriminels, groupes mafieux organisés ou collectifs soutenus par des États, exploitent de plus en plus les possibilités offertes par le machine learning, et les modèles génératifs. Finies les campagnes de phishing aux fautes grossières ou les ransomwares balancés à l'aveugle sur des adresses e-mail aléatoires.

Désormais, les attaques s'adaptent à leur cible, exploitent ses comportements numériques, imitent ses partenaires professionnels ou personnels avec un réalisme troublant. Grâce à des modèles d'IA formés sur des corpus de courriels, de documents internes volés ou d'interactions publiques en ligne, il est possible de générer un message de spear phishing si réaliste qu'aucun employé, aussi prudent soit-il, ne pourrait douter de son authenticité. L'intelligence artificielle permet aussi d'automatiser des tâches complexes qui nécessitaient autrefois une expertise humaine analyser les failles d'un système, tester des milliers de mots de passe avec des techniques d'optimisation, injecter du code malveillant de manière polymorphe pour échapper aux antivirus, ou même écrire du code exploitable en quelques secondes. Des outils comme **WormGPT**, **FraudGPT** ou d'autres IA non-officielles qui circulent sur les forums clandestins du dark web sont capables de produire des scripts de phishing, des modèles de malware ou des stratégies de contournement avec un niveau de fluidité et de dissimulation qui dépasse parfois largement les outils classiques.

Mais la plus grande révolution ne vient pas de la puissance des attaques, mais de leur autonomie.

Des attaques dites "zero-click", lancées sans intervention humaine, peuvent être déclenchées à la volée par des bots intelligents, capables de dé-

tecter une opportunité de vulnérabilité et d'y injecter un exploit dans la même seconde.

Des IA, laissées en semi-autonomie dans des espaces numériques comme les réseaux sociaux ou les plateformes de messagerie, peuvent observer, analyser, imiter des comportements humains, se faire passer pour des collaborateurs, puis manipuler des décisions internes critiques. Le temps de réponse humain n'est plus suffisant. L'analyse post-mortem d'un incident n'empêche pas sa réitération. Le paradigme de la cybersécurité doit changer radicalement.

Il ne s'agit plus de construire des murailles, mais de déployer des contre-IA. Il ne s'agit plus de prévenir l'attaque, mais de répondre dans le même temps algorithmique que l'agresseur.

À ce stade, les modèles de sécurité basés sur la détection par signature, sur les pare-feu traditionnels deviennent obsolètes.

"Le futur de la cyber ne dépend pas de notre capacité à écrire de meilleures règles. Il dépend de notre capacité à entraîner de meilleures IA."

Comment organiser la défense ?

Face à la sophistication croissante des IA offensives, les grandes entreprises technologiques, les agences de renseignement et les sociétés de cybersécurité investissent massivement dans des systèmes défensifs pilotés par l'intelligence artificielle.

Ces IA ne sont pas simplement des extensions d'outils existants ; elles deviennent des entités d'observation permanente, d'analyse comportementale, de prédiction et de réaction automatique. En analysant les flux réseau en temps réel, en croisant des milliards de signaux faibles, en comparant des millions de logs, ces IA sont capables de détecter une activité suspecte avant même qu'un dommage ne soit causé. Certaines vont jusqu'à anticiper les tactiques adverses, à la manière d'un jeu d'échecs où chaque mouvement potentiel est simulé à l'avance. L'IA défensive n'est plus seulement une alarme, elle est une force de contre-attaque.

Ces systèmes, intégrés dans les architectures de type XDR, se basent sur des moteurs d'apprentissage automatique continuellement entraînés par les incidents du monde réel. Ils apprennent de chaque brèche, de chaque attaque, de chaque anomalie. Ils ne protègent plus uniquement les systèmes, mais aussi les identités, les comportements, les relations. Ils peuvent détecter qu'un collaborateur a été usurpé à partir d'un changement minime dans le style d'écriture, ou qu'un serveur est compromis à partir d'un pic d'activité anormal de quelques millisecondes.

Ils réagissent parfois sans même attendre une validation humaine, en isolant un segment réseau, en coupant temporairement un accès, ou en reconfigurant des règles de sécurité dynamiquement.

Mais si l'IA permet aux défenseurs de réagir plus vite, elle introduit aussi un nouvel enjeu stratégique : l'asymétrie algorithmique. Dans un monde où les IA s'affrontent, la puissance d'un système ne repose plus uniquement sur son infrastructure ou sur ses règles, mais sur la qualité de ses modèles, la richesse de ses données

d'entraînement, et la capacité de ses algorithmes à généraliser les menaces inconnues.

La guerre ne se joue plus uniquement sur le terrain de l'attaque et de la défense, mais dans l'arène invisible des données. Celui qui détient les meilleures données de compromission, les meilleurs modèles de simulation, et la plus grande vitesse d'adaptation prend l'avantage.

Cette guerre silencieuse entraîne une militarisation croissante de l'intelligence artificielle dans le domaine cyber.

Les grandes puissances États-Unis, Chine, Russie, Israël, développent en secret des IA de cyberdéfense et de cyberoffensive, certaines capables de lancer des campagnes automatisées de désinformation, de neutraliser à distance des réseaux critiques, ou de se propager dans des systèmes isolés sans contact humain.

Le champ de bataille est mondial, dématérialisé. Et pour la première fois, il échappe en grande partie à la surveillance humaine

Jusqu'où laisser une IA prendre des décisions autonomes de défense, surtout si ces décisions impliquent des représailles, voire des attaques préventives ? À quel moment une IA peut-elle être considérée comme responsable d'un acte cybercriminel ? Quels garde-fous, quelles normes internationales, quels traités faudrait-il mettre en place pour encadrer cette nouvelle forme de conflit ?

Nous sommes à l'aube d'un droit de la guerre algorithmique, encore balbutiant, mais de plus en plus nécessaire.

Les entreprises, quant à elles, n'ont pas le luxe d'attendre que le droit international se stabilise. Chaque jour, elles sont confrontées à des menaces de plus en plus furtives, souvent indétectables sans outils intelligents. La plupart des Responsables de la Sécurité des Systèmes d'Information savent aujourd'hui que l'avenir de leur défense passe par l'IA. Mais cette transition soulève de nouveaux défis internes : comment entraîner une IA sans violer les réglementations RGPD ?

Comment éviter les biais dans les modèles, qui pourraient engendrer des faux positifs massifs ?

Comment recruter des talents capables de comprendre à la fois la cybersécurité, l'IA, et les enjeux juridiques associés ?

Qui a la meilleure capacité de prédiction ? Qui réagit le plus vite ? Qui apprend le mieux ? C'est la nouvelle course à l'armement numérique. Une guerre froide algorithmique. Invisible. Permanente. Totale.

La cybersécurité n'est plus un domaine défensif. C'est devenu un théâtre d'opérations. Et les pacteurs ne sont plus uniquement humains. Dans les cyberespace, les intelligences s'affrontent déjà.

Et ce n'est que le début...

L'AI Act : la règlementation européenne de l'intelligence artificielle

par Virginie Mathivet

L'Intelligence Artificielle est partout : dans nos applications mobiles, nos outils professionnels, nos services publics… Mais elle n'est pas sans faille.

On se souvient de la voiture autonome d'Uber incapable de détecter un piéton hors des passages cloutés, de Google Photos qui a associé des personnes noires à des gorilles, ou encore de l'algorithme du Pôle Emploi autrichien qui pénalisait les femmes dans l'accès aux métiers de l'informatique.

Ces dérapages ont un point commun : des biais issus des données utilisées pour entraîner les modèles. Ils sont parfois subtils, souvent involontaires, mais ils peuvent être détectés et corrigés… à condition d'y prêter attention et d'appliquer des méthodes rigoureuses.

C'est précisément pour éviter ce type de dérive que l'Union européenne a conçu l'AI Act (ou RIA pour « règlement sur l'intelligence artificielle »). Ce texte vise à encadrer l'usage de l'IA, en particulier lorsqu'elle est déployée dans des contextes sensibles, et à imposer aux entreprises le respect de bonnes pratiques en matière de développement et d'éthique des IA.

L'AI Act, un frein à l'innovation ?

L'un des reproches majeurs adressés à l'AI Act est qu'il pourrait freiner l'innovation, en imposant des contraintes administratives et techniques aux entreprises européennes, tandis que leurs concurrentes étrangères (notamment américaines et chinoises) évolueraient dans des environnements réglementaires plus souples, voire inexistants.

1. Le règlement s'applique à tous, même hors Europe
L'AI Act concerne tous les systèmes d'intelligence artificielle qui sont utilisés dans l'Union européenne ou ont un impact sur des citoyens européens — quelle que soit l'origine de l'éditeur. Une entreprise américaine ou chinoise qui souhaite commercialiser une solution IA en Europe devra elle aussi se conformer aux exigences du texte. Il n'y a donc pas de distorsion de concurrence à ce niveau.

2. La régulation devient globale
L'Union européenne n'est pas seule sur ce terrain. Les États-Unis, la Chine, le Canada ou encore le Royaume-Uni travaillent eux aussi à des cadres réglementaires autour de l'IA, même si leurs approches sont différentes. On assiste à une prise de conscience mondiale : définir ce qu'est une IA « fiable » et encadrer ses usages devient un enjeu diplomatique, industriel et sociétal.

3. Un bac à sable réglementaire pour l'expérimentation
Pour éviter de freiner la recherche et le développement, l'AI Act prévoit la mise en place de « bacs à sable » réglementaires. Ces environnements contrôlés permettent aux entreprises de tester des systèmes innovants en conditions réelles, sans avoir à se conformer immédiatement à toutes les exigences réglementaires. L'innovation reste donc possible, à condition d'être encadrée.

4. Des contraintes alignées avec les bonnes pratiques
Les exigences de l'AI Act ne sortent pas de nulle part : elles reprennent, pour la plupart, des pratiques déjà appliquées dans les entreprises soucieuses de qualité, de sécurité et d'éthique. Une entreprise sérieuse, qui documente ses choix, teste ses modèles et garantit la supervision humaine, n'aura pas de mal à obtenir la conformité. Si une solution ne résiste pas à ces critères, il est légitime de s'interroger sur sa robustesse réelle.

Enfin, le texte n'est pas figé : la Commission européenne prévoit une révision régulière du règlement pour l'adapter aux évolutions technologiques et aux retours d'expérience.

Des niveaux de risque pour encadrer l'IA

Risque inacceptable
Ce sont les systèmes interdits par défaut. Ils visent à manipuler les comportements ou à instaurer un contrôle social par exemple, la surveillance de masse en temps réel ou les systèmes de notation citoyenne.

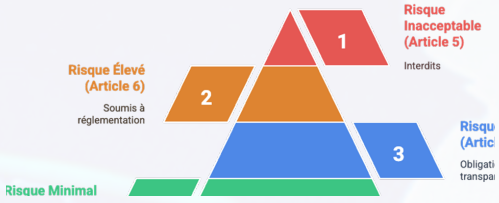
Risque élevé
Il s'agit des systèmes susceptibles de porter atteinte aux droits fondamentaux ou à la sécurité physique ou morale des personnes. Cela concerne, par exemple, les algorithmes de recrutement, les outils d'aide à la décision judiciaire, ou les systèmes déployés dans le domaine de l'éducation ou de la santé. Ces IA sont soumises à un cadre strict et doivent obtenir un agrément.

Risque limité
Ces systèmes n'ont pas de conséquences graves en cas d'erreurs mais interagissent avec le public. C'est le cas, par exemple, des assistants conversationnels ou des outils de recommandation. L'obligation principale porte sur la transparence : les utilisateurs doivent être informés qu'ils interagissent avec une IA. Une formation adaptée peut aussi être exigée dans certains contextes.

Risque minimal
Ces systèmes présentent très peu de risques ou sont utilisés dans des contextes purement techniques ou internes. Ils ne sont soumis à aucune obligation réglementaire spécifique, mais des bonnes pratiques sont recommandées.

Un cas à part : les modèles à usage général

Certains systèmes, comme les grands modèles de langage (LLM) de type ChatGPT, peuvent être intégrés à des usages très variés, y compris dans des contextes critiques. L'AI Act leur applique un régime spécifique.



Les obligations pour les systèmes d'IA à risque élevé

Lorsqu'un système d'intelligence artificielle entre dans la catégorie à risque élevé, l'AI Act impose une série de contraintes visant à garantir la sécurité et le respect des droits fondamentaux des citoyens européens. Les entreprises concernées devront obtenir un agrément au niveau européen, basé sur un dossier démontrant leur conformité à plusieurs exigences techniques et organisationnelles. Celles-ci reprennent en grande partie des bonnes pratiques déjà connues dans les milieux de l'ingénierie logicielle et de la cybersécurité :

Système de gestion des risques identification, évaluation et réduction des risques associés à l'utilisation du système IA, tout au long de son cycle de vie,

Gouvernance des données mise en place de processus pour garantir la qualité, la sécurité et la confidentialité des données utilisées lors de l'entraînement, des tests et de l'exploitation,

Documentation technique le système doit être entièrement documenté, en particulier l'architecture, les objectifs, le processus de développement, la gestion des biais dans les jeux de données, ainsi que les métriques de performance et de robustesse du modèle,

Journalisation des événements (logs) : les systèmes doivent conserver des journaux détaillés de leur fonctionnement, permettant de retracer une prise de décision, d'identifier une erreur ou de détecter un biais,

Transparence vis-à-vis des utilisateurs : l'IA doit être clairement identifiable comme telle. Elle doit signaler ses limites et informer les utilisateurs qu'elle peut produire des erreurs,

Supervision humaine un humain doit pouvoir surveiller le fonctionnement du système, intervenir à tout moment, voire bloquer une décision prise automatiquement.

La plupart de ces exigences s'inspirent des standards de qualité déjà appliqués dans de nombreuses organisations. Pour les entreprises ayant une démarche sérieuse de développement responsable, l'AI Act ne devrait pas être un obstacle, mais plutôt un cadre structurant et légitimant.

Comment façonner une intelligence artificielle plus éthique et plus inclusive ?


Hamilton Mann

Hamilton Mann, dirigeant dans le domaine de la Tech, est un pionnier dans le domaine de l'IA et de la transformation numérique. Il est lecturer à **l'INSEAD** et à **HEC Paris**, doctorant en IA à l'École des Ponts Business School – École des Ponts et Chaussées, et figure parmi les lauréats du Thinkers50 Radar, qui distingue les penseurs émergents les plus influents au monde dans le domaine du management.

Il a contribué à des publications sur les enjeux de l'IA, parmi lesquelles la Stanford Social Innovation Review, la California Management Review, Rotman Management Magazine, I by IMD, Wharton Knowledge, INSEAD Knowledge, Polytechnique Insights (le journal de l'École Polytechnique), et Leader to Leader (le journal de l'Université de Pittsburgh) et est également un contributeur régulier pour Forbes US, The European Business Review et Harvard Business Review France.

Alors que l'intelligence artificielle s'impose progressivement dans tous les domaines de la vie humaine, une question majeure se pose : comment faire en sorte qu'elle soit véritablement éthique et inclusive ?

Développée à partir de données et de modèles issus de contextes humains imparfaits, l'IA hérite de nos biais, consciemment ou non. Conçue pour répondre aux besoins d'un public cible, elle tend naturellement à exclure celles et ceux qui n'en font pas partie, perpétuant ainsi les discriminations existantes.

Cette problématique devient d'autant plus cruciale que l'IA est désormais omniprésente. Son marché, estimé à 87 milliards de dollars en 2021, pourrait atteindre près de 1 600 milliards en 2030. Des assistants vocaux aux véhicules autonomes, en passant par les systèmes de recommandation, les robots chirurgicaux ou les outils de gestion de la relation client, l'IA transforme les usages, les métiers et les rapports sociaux. Elle s'immisce dans les décisions personnelles, professionnelles et politiques.

Pourtant, derrière cette apparente neutralité algorithmique, se cache une tension profonde entre la quête de performance économique et la nécessité de préserver les valeurs d'équité et de justice. Comment garantir que les biais ou les modèles de segmentation dans les données qui alimentent l'IA ne conduisent pas à des fonctionnements qui traitent défavorablement les individus sur la base de caractéristiques telles que leur sexe, leur couleur de peau, leur religion, leur handicap, leur orientation sexuelle ou politique ?

Voilà l'une des grandes questions que pose le développement de l'intelligence artificielle.

Un défi de société

L'intelligence artificielle n'est pas si… artificielle. Avec son développement exponentiel et effréné, la tentation est, et sera de plus en plus grande, de l'utiliser pour établir des modes de différenciation inédits, des approches de ciblage inégalées, pour plus de croissance économique, pour plus de compétitivité.

Il y a une tension entre d'une part la nécessité d'avoir des organisations et des individus capables de tolérance vis-à-vis de la diversité tout en comprenant l'enjeu de l'inclusivité pour construire plus d'égalité dans la société et, d'autre part, le système économique mondial, qui incite et exacerbe plus qu'il ne réfrène des comportements nous conduisant dans ces formes de compétition.

Donc discriminer, est une règle du jeu qui mène au succès.

Cette tension est en voie d'amplification, parce que l'IA est en capacité de codifier de manière systématique et systémique dans notre société numérique : c'est l'un des plus grands défis de notre temps.

L'intelligence artificielle pénètre déjà tous les pores de la société :

- les assistants personnels sont maintenant virtuels et permettent d'exécuter des tâches quotidiennes de base

- les analyses de marché sont réalisées par des machines qui produisent des études telles que la comparaison de concurrents, et mettent en forme des rapports détaillés

- les analyses des comportements d'usage, des processus d'achat et des préférences des clients sont passées au crible par des CRM (gestion de la relation client) qui intègrent de plus en plus d'intelligence et sont capables de faire des prédictions sur les besoins des consommateurs

- le service client est également assuré par des chatbots capables de répondre aux questions les plus fréquemment posées par les visiteurs d'un site Internet

Et tout cela n'est qu'un avant-goût, comparé au développement d'applications possibles, déjà naissantes, et pour autant inscrites dans un futur qui s'approche à grand pas avec :

- les véhicules autonomes (bicyclettes, voitures, trains, avions, bateaux…)

- les robots assistant les chirurgiens en salle d'opération

- la création de contenus (vidéos, musiques, articles…) entièrement produits par le fruit du travail de la machine

- les politiques publiques dont les mesures seraient prescrites, et dont la performance serait prédite par l'analyse de larges volumes de données

Soit nous prévoyons d'utiliser l'IA pour augmenter notre capacité à éliminer les inégalités visibles et invisibles jusqu'à des niveaux jamais atteints auparavant, soit nous prévoyons consciemment ou non de les augmenter à la même échelle. L'ère de l'intelligence artificielle dans laquelle nous entrons comprendra de moins en moins d'entre deux.

Une nouvelle ère pour l'apprentissage humain

Nous, humains, sommes responsables de ce que le machine learning essentiel à toute intelligence artificielle apprend, comment il apprend ce qu'il fait et n'apprend pas. La façon dont nous enseignons ce que la machine doit apprendre est au cœur des problèmes d'apprentissage du XXIe siècle.

Cela implique que nous devions non seulement continuer à apprendre à développer notre propre intelligence, mais aussi comprendre et apprendre comment la machine apprend à développer la sienne.

Les apprentissages humains et automatiques partagent de nombreux défis conceptuels et techniques, révélant une convergence entre cognition biologique et intelligence artificielle. L'opposition entre apprentissage supervisé et non supervisé reflète la distinction entre un apprentissage guidé par des exemples annotés et une exploration autonome de structures latentes.

De même, la différence entre apprentissage structuré et non structuré met en lumière la complexité à gérer des données organisées versus celles qui sont brutes ou chaotiques. Certains modèles, comme l'ap-

prentissage « one-shot » ou « few-shot »*, cherchent à imiter la capacité humaine à apprendre à partir d'un ou très peu d'exemples, en contraste avec l'apprentissage « blink » évoqué par Malcolm Gladwell, qui mise sur l'intuition rapide issue d'expériences passées.

Par ailleurs, la tension entre mémoire à court et long terme, et le compromis entre oubli et rétention, est un enjeu central tant pour les réseaux de neurones que pour le cerveau humain.

L'apprentissage « zero-shot », qui consiste à généraliser sans exemple direct, trouve un parallèle intriguant avec la théorie du « rêve-oubli » de Crick et Mitchison, selon laquelle le sommeil paradoxal sert à éliminer des souvenirs inutiles.

Enfin, les approches d'apprentissage ancrées dans l'action corporelle comme le paradigme visuomoteur se rapprochent des formes humaines d'intelligence incarnée, tandis que l'intégration multisensorielle (auditif, visuel, kinesthésique) illustre la richesse de l'apprentissage global chez l'humain comme dans certaines architectures neurales. En apprenant comment les machines peuvent apprendre, nous découvrons et découvrirons encore de nouvelles façons d'apprendre qui, jusqu'alors, n'avaient pas été explorées ni même imaginées. Celles-ci pourraient bien révolutionner les normes que nous connaissons sur notre propre façon d'apprendre, pour augmenter l'intelligence humaine.

Mais ne nous y trompons pas.

L'intelligence et le savoir ne sont pas synonymes, et augmenter notre savoir est une condition nécessaire mais insuffisante pour augmenter notre intelligence. Augmenter notre intelligence humaine c'est surtout augmenter notre capacité à remettre en question, à challenger le statu quo, à éveiller notre curiosité et faire surgir dans notre esprit de nouvelles questions, pour la découverte et la redécouverte de ce que nous pensons savoir, et de ce que nous sommes.

L'intelligence artificielle est bien moins intelligente qu'on l'imagine.

Une question de compréhension

Sans aller jusqu'à imaginer une intelligence artificielle capable d'imiter le ressenti humain, il y a quelque chose qui la distingue inévitablement de celle de l'humain, et qui réside dans la compréhension et l'appréhension du contexte. Le contexte est composé d'autant de paramètres, certains visibles à l'œil nu, et d'autres plus discrets, plus fins, plus subtils, constitués de signaux faibles et de détails autant de paramètres qui comptent pour caractériser un contexte. Compte tenu de la nature d'évolution perpétuelle propre à tout contexte, il faudra du temps avant qu'une intelligence artificielle soit capable d'apprécier la complexité d'un contexte.

Construire l'IA dont nous avons besoin pour le bien de la société, passe nécessairement par une vision. Celle permettant de comprendre les tâches qui sont et seront les mieux exécutées par l'intelligence de la machine, par opposition à celles qui sont et seront mieux exécutées par l'intelligence humaine, en considérant aussi celles qui doivent et devront continuer d'être exécutées par l'humain, quoi qu'il arrive.

Les réponses que nos sociétés établiront pour bâtir le cadre selon lequel l'IA est intelligente pour l'humain, façonneront le futur de l'humanité toute entière, non seulement sur le plan de nombreuses innovations et de nouvelles formes d'avantages concurrentiels qui changeront les lois de marchés telles que nous les connaissons aujourd'hui, mais, plus important encore, sur le plan sociologique, du monde que nous laisserons aux prochaines générations.

La plupart du temps, lorsque l'on pense au machine learning, notre modèle mental nous amène à penser qu'il s'agit d'une approche strictement unidirectionnelle dans laquelle nous enseignons à la machine et lui donnons, dans différents domaines, les moyens d'apprendre par elle-même.

Une relation en

* Dans la méthode d'apprentissage automatique one-shot et few-shot, un petit nombre d'exemples est disponible dans l'ensemble de données d'apprentissage, tandis qu'avec l'apprentissage zéro-shot, il n'y a aucune information sur la catégorie concernée dans les données d'apprentissage.

profonde mutation

L'IA provoque une profonde mutation du lien entre l'humain et la machine qu'il deviendra de plus en plus critique et passionnant d'explorer par ce qu'il est en réalité déjà plus bidirectionnelle que jamais. La question nous est donc posée : que peut nous apprendre l'intelligence de la machine, pour nous améliorer (dans ce que nous faisons) en tant qu'êtres humains ?

Nous allons devoir apprendre à penser différemment sur bien des choses pour faire faire à la machine ces mêmes choses qu'il nous serait humainement difficile, voire impossible, de réaliser de la même manière. Et nous allons aussi pouvoir saisir de nouvelles opportunités d'apprendre, de nous former, sur de nombreuses choses dont l'expertise ne s'acquiert aujourd'hui qu'au prix de longs efforts et de nombreuses années, et pour lesquelles la performance ne peut réellement être atteinte que par l'exécution humaine.

Si l'IA, et les recommandations qu'elle produit, ouvrent des opportunités insoupçonnées pour augmenter non seulement notre propre intelligence, mais aussi la nature des relations et des attachements émotionnels que nous pourrions développer avec la machine dans le futur, elle ouvre aussi des questions délicates de responsabilité sociétale des entreprises (RSE) : à partir de quand l'aide à la décision apportée par l'intelligence artificielle agirait-elle avec un tel degré d'influence, qu'elle déciderait finalement silencieusement à la place de l'humain ?

" Cette question compliquée est à notre porte"

La réponse, notamment en fonction du degré de vulnérabilité que la société peut à un instant donné reconnaître en chacun d'entre nous, dans un moment particulier de notre vie, dans des circonstances particulières d'existence, peut revêtir autant de nuances que de personnes.

C'est pourquoi les applications, les appareils, et tout équipement technologique doté de quelque forme que ce soit d'intelligence artificielle, devront faire l'objet d'une lisibilité explicite quant à la limitation des paramètres que l'algorithme prend, ou ne prend pas en compte, quant aux potentielles implications pouvant représenter un danger pour soi ou pour autrui, pour aider à une utilisation responsable desdites IA de ces machines, et prévenir les risques d'utilisation inappropriée, voire proscrite.

L'IA nous oblige à relever le grand défi de la rendre capable d'être explicitement explicable à tous et pour tous, sur les causalités des résultats qu'elle propose, pour guider des décisions qui impacteront de plus en plus nos vies et la société dans son ensemble, même si, paradoxalement, en tant qu'humains, nous-mêmes ne savons pas tout expliquer sur le pourquoi de nombre de nos décisions, de telle manière que le plus grand nombre les comprendraient et que ces explications seraient justes.

Vers une économie de la confiance

L'IA va profondément changer la valeur du travail. Certains craignent même qu'elle n'en vienne à remplacer l'humain.

Si l'image d'une intelligence artificielle de science-fiction, supplantant l'humanité comme dans Terminator, est pure fiction, il y a un paradigme qu'il est nécessaire d'inclure dans ce que la société numérique couve en son sein : l'intelligence artificielle peut être meilleure que l'homme pour réaliser certaines tâches et, pour autant, elle n'est pas et ne sera pas meilleure que l'homme pour réaliser toutes les tâches.

Avec les développements de l'IA, nous vivons et nous allons vivre une transformation qui est celle de l'économie de la connaissance vers l'économie de la confiance, motivée, d'une part, par les besoins de plus de prévisibilité, de plus de précision et de plus d'efficacité et, d'autre part, par les besoins de plus d'équité, de plus de transparence et de plus de durabilité.

Pour l'avenir des travailleurs du savoir, la technologie numérique et en particulier l'intelligence artificielle entraîneront cinq types de changements. Chacun va bouleverser la société à plus ou moins grande échelle, et de façon plus ou moins forte selon les natures prédominantes du travail et de la valeur travail dans chaque continent.

1 Source première d'anxiété, largement alimentée par l'imaginaire d'une IA diffusée par la pop culture, il y a tout d'abord **ces métiers qui vont disparaître.** Et ce n'est pas nouveau. En d'autres temps, avec d'autres révolutions industrielles, ce phénomène a déjà existé.

2 Ensuite, il y a **les emplois qui seront augmentés par l'intelligence artificielle**. Là aussi, ce n'est pas nouveau. Par analogie, en d'autres temps, en d'autres révolutions industrielles, ce phénomène a aussi existé.

3 Puis il y a **ces métiers qui vont évoluer** pour devenir des techni-jobs.

4 Et ceux qui sont assez difficiles à imaginer désormais, car leur utilité est intrinsèque à **des besoins de nos sociétés dont on ne sait encore rien, ou peu de choses**.

5 Mais ne soyons pas naïfs : le développement de l'intelligence artificielle créera et crée d'ores et déjà **l'accroissement de l'émergence de métiers précaires**, de métiers béquilles pour pallier le manque d'intelligence de l'intelligence artificielle.

Ce sont, par exemple, ces travailleurs de l'ombre qui labellisent des tonnes de données, dans une frénésie de tâches particulièrement répétitives, pour aider l'IA à apprendre et faire en sorte que certains contenus abjects soient prohibés d'un accès via les plateformes que nous utilisons, parce qu'ils enfreignent la loi, avec l'impact que le visionnage de tels contenus peut avoir à la longue sur la santé mentale de ces « travailleurs ».

Lequel de ces types de changements provoqués par l'IA aura le plus grand impact sur l'évolution du travail dans nos sociétés ? Difficile à prédire. Pour autant, même s'il ne s'agit pas là de la seule force en mouvement dans la cinétique des transformations qui caractérisent notre siècle, ce sera évidemment à nous d'en décider.

Quoi qu'il en soit, l'intelligence artificielle n'a pas d'autre éthique que la nôtre.

Nos principes éthiques sont in fine, pour l'IA, une partie intégrante des exigences fonctionnelles qui, par voie de conséquence, codifient numériquement les biais dont nous sommes intellectuellement propriétaires. Elle hérite en quelque sorte des gènes éthiques de son créateur.

Rendre visibles les codes invisibles de nos sociétés est probablement l'une des avancées les plus transformantes que l'IA va permettre à l'humanité d'accomplir.

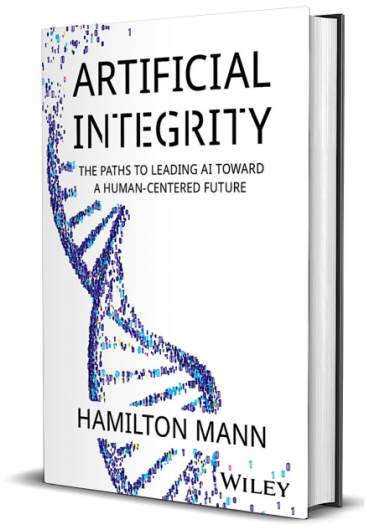
Un tel niveau de transparence sur le non-dit et le non-écrit, ainsi révélés au grand jour, va aider à plus d'égalité et redéfinir en profondeur la demande citoyenne de justice dans nos sociétés.

Et c'est aussi une opportunité de faire en sorte que les intelligences artificielles qui interagiront avec la nôtre soient le plus possible le produit d'intelligences collectives et, au mieux, le réceptacle de la richesse que peuvent produire les synergies issues de la diversité humaine, sous toutes ses formes d'intelligence.

L'augmentation de notre intelligence par celle de la machine sera toujours et plus encore à l'avenir confrontée à la question existentielle de la cause humaine que nous donnons à cette intelligence pour mission de servir.

C'est donc qu'il nous faudrait faire de l' « intelligence artificielle » une intelligence inspirée par la quintessence de ce qu'il y a de meilleur dans notre humanité, en excluant toutes les parts sombres de la nature humaine. C'est probablement la question la plus vertigineuse, mais aussi la plus déterminante pour l'avenir de l'humanité. C'est une question éthique à laquelle seule notre humanité a le pouvoir et la responsabilité d'apporter une réponse, sans cesse renouvelée, pour construire le futur dans lequel nous souhaitons vivre.

Hamilton Mann



VOTRE GUIDE SUR L'IA GÉNÉRATIVE

PAR **EVA BAIKECHE**

UN GUIDE À PARCOURIR EN FAMILLE

C'est quoi une IA Générative ?

Une IA générative est un programme capable de créer du contenue (texte-images-musiques-vidéos) à partir de simples instructions écrites qu'on appelle des "**prompts**".

L'IA Générative peut vous aider à...

Rédiger ou reformuler n'importe quel type de message, lettre, email dans n'importe quel ton : professionnel, informel, "à la Victor Hugo" ou "à la Michou"...

Résumer un cours, un article ou une vidéo pour mieux comprendre.

Expliquer simplement un sujet complexe : un cours, la théorie quantique, la fiscalité française.

Créer des images à partir d'idées ou de descriptions : prenez en photo votre appartement et demander lui de refaire une déco en particulier.

Organiser votre quotidien : liste de courses selon vos restrictions alimentaires, mise en place de séance de sport, planifier l'itinéraire de votre voyage.

Une demande à ChatGpt est 10 fois plus gourmande en énergie qu'une recherche sur Google.

Réflexes écolo avec l'IA

- **Soyez clair, pas poli** : pas besoin de dire "s'il te plaît" à l'IA : chaque mot inutile, c'est de l'énergie en plus.
- **Ne soyez pas excessif dans vos générations** ! Une image IA = des centaines de watts. Réservez-les aux vraies bonnes idées.
- **Choisissez une IA éthique** : Utilisez des outils qui protègent vos données et consomment moins !

Les 4 réflexes à avoir

1- **Toujours vérifier ce qu'elle dit,** l'IA utilise des sources parfois douteuses.

2- **Donnez lui du contexte dans votre demande,** imaginez que vous vous adressez à une personne qui ne sait rien sur vous !

3- **Ne pas lui confier d'infos trop personnelles** : l'IA n'est pas une psy, elle peut être de mauvais conseils.

4- **Utiliser avec curiosité et esprit critique** : c'est un super outil qui est là pour vous aider, pas pour réfléchir à votre place !

Les limites de l'IA

L'IA peut très souvent se tromper, on dit d'ailleurs qu'elle **hallucine** !

Elle n'a ni conscience ni émotions.

Elle n'invente rien de neuf ! Elle fait juste des **probabilités** sur les meilleures réponses.

Elle **reproduit des biais** comme des erreurs, des préjugés, des stéréotypes...

Nos outils IA préférés

seelab
Retouche & création d'image

Suno
Création de musique

Leexi.ai
Retranscription de réunions

Lovable
Coder une application

Gamma AI
Création de diaporama

CYBER RÉSILIENCE ET INTELLIGENCE ARTIFICIELLE : PRÉDIRE L'INCIDENT AVANT QU'IL NE SURVIENNE

Sur le terrain numérique contemporain, l'intelligence artificielle constitue une arme à double tranchant. Si les cybercriminels l'exploitent massivement pour orchestrer des attaques d'une sophistication inédite, les équipes cyber disposent aujourd'hui des mêmes leviers technologiques pour inverser la tendance.

Une révolution silencieuse est en cours, celle qui transforme profondément notre capacité à anticiper, détecter et neutraliser les cybermenaces.

A PROPOS DE L'AUTEUR

Christophe Longuepez est entrepreneur et expert en cyber depuis 15 ans.

Il s'est spécialisé dans l'intelligence artificielle et ses impacts sur l'écosystème cyber : évolution des menaces, adoption de l'IA de manière sécurisée, et exploitation des technologies d'IA pour renforcer la résilience des organisations face aux risques.

L'IA offensive : une menace qui s'industrialise

L'arsenal des attaquants s'enrichit quotidiennement grâce à l'IA. Les deepfakes ultraréalistes permettent de compromettre l'identité de dirigeants avec un réalisme saisissant, transformant la fraude au président en opérations redoutablement crédibles. Un simple échantillon vocal de quelques secondes suffit désormais à générer une conversation téléphonique parfaitement imitée, capable de tromper même les collaborateurs les plus vigilants.

Les malwares polymorphes, générés par des modèles de langage malveillants (DarkLLM), évoluent en permanence pour tenter de contourner les capacités de détection des solutions EDR/XDR les plus avancées.

Cette industrialisation de la menace représente un défi sans précédent. Là où un cybercriminel devait autrefois passer des semaines à construire un scénario d'ingénierie sociale crédible, l'IA lui permet désormais d'automatiser, de personnaliser et d'exécuter ses campagnes à grande échelle. L'asymétrie semblait jusqu'ici définitivement favorable aux attaquants.

Pourtant, cette même intelligence artificielle qui inquiète à juste titre les professionnels de la sécurité, constitue également leur plus puissant allié, capable de démultiplier leur efficacité.

Cette transformation touche l'ensemble de l'écosystème, de la gouvernance des risques aux opérations de cyberdéfense les plus techniques. Une métamorphose qui redéfinit fondamentalement l'approche de la résilience des entreprises et des organisations, transformant une cybersécurité traditionnellement réactive en une discipline réellement proactive.

Une GRC augmentée, plus agile et précise

La gouvernance, les risques et la conformité (GRC) représentent traditionnellement des domaines chronophages, nécessitant des processus complexes, avec de nombreux interlocuteurs et une expertise humaine considérable pour analyser, évaluer et piloter les multiples dimensions du risque cyber. L'IA révolutionne cette approche en apportant une capacité d'analyse, de traitement et de production sans commune mesure avec les méthodes conventionnelles.

La gestion des risques tiers transformée

Le Third Party Cyber Risk Management (TPcRM) illustre parfaitement ce tournant. Gérer les risques associés aux fournisseurs, partenaires et prestataires externes relevait jusqu'alors d'un parcours du combattant : questionnaires fastidieux souvent incomplets ou beaucoup trop élogieux, évaluations manuelles voire inexistantes, suivi discontinu des évolutions de sécurité. L'IA change la donne en permettant un pilotage actif et continu de ces risques, en lien avec les exigences de NIS2.

Les algorithmes analysent désormais en temps réel la surface d'attaque des tiers partenaires, scrutent automatiquement leurs

certifications et leur évolution, évaluent continuellement l'importance stratégique du partenaire dans la chaîne de valeur et détectent les signaux faibles ou toute évolution significative qui pourraient présager d'une compromission imminente.

Cette surveillance permanente transforme une gestion traditionnellement bureaucratique en approche véritablement proactive, permettant d'anticiper et de mitiger les risques au plus tôt

Pertinent quand on sait que 93% des organisations ayant répondu au CyberRisk Alliance, ont indiqué avoir subi au moins une fois, une attaque indirecte via un tiers fournisseur

La quantification des risques cyber : de l'art à la science

L'évaluation des risques cyber a longtemps oscillé entre approximation subjective et intuition d'expert, induisant de nombreux biais cognitifs et des évaluations peu reproductibles.

L'IA apporte une dimension scientifique à cette discipline en croisant des milliers de variables vulnérabilités techniques, contexte métier, historique d'incidents, maîtrise de la menace facteurs humains et organisationnels pour générer des scénarios chiffrés et objectivés.

Cette révolution quantitative permet enfin aux CISOs et dirigeants de :

- Quantifier précisément leur exposition au risque avec des métriques fiables
- Hiérarchiser les menaces selon leur impact potentiel réel sur l'activité
- Justifier objectivement leurs investissements sécurité auprès des directions générales
- Mesurer l'efficacité des mesures déployées avec des indicateurs tangibles

L'IA générative pousse cette logique encore plus loin en transformant la planification stratégique. Elle facilite la construction de roadmaps cybersécurité véritablement cohérentes en analysant les contraintes organisationnelles, budgétaires et techniques pour proposer des séquences d'actions optimisées.

En évaluant le rapport coût-bénéfice de chaque mesure cyber proposée, elle identifie les investissements les plus efficients ceux qui maximisent le gain sécurité pour un effort temporel et financier donné.

Cette avancée technologique transforme la construction des stratégies cybersécurité en un processus objectif et data-driven, permettant aux organisations de déployer leur stratégie cyber avec efficience dans un environnement sous contrainte.

Du goulot d'étranglement au moteur d'agilité

L'intégration de la sécurité dans les projets - le fameux security by design - est souvent vécue par les équipes métier comme une contrainte majeur : allongement des délais, manque de clarté des exigences, processus en silo, prescriptions perçues comme déconnectées de la réalité opérationnelle du projet.

Ce modèle traditionnel, essentiellement manuel, atteint aujourd'hui ses limites structurelles par manque d'outils adaptés et de ressources humaines spécialisées disponibles en quantité suffisante.

L'IA transforme radicalement cette équation. Dès les premières étapes d'un projet, elle permet d'automatiser la qualification sécurité, une étape critique souvent négligée faute de temps ou d'expertise disponible.

En analysant dynamiquement et en continu les éléments structurants d'un projet architecture cible, types de données traitées, périmètre fonctionnel, surface d'exposition, dépendances externes, technologies utilisées parfois même sans sollicitation manuelle explicite, les modèles d'IA identifient automatiquement les points sensibles et évaluent le niveau de criticité associé.

Ce travail s'appuie sur des grilles de risques métier éprouvées, des référentiels/frameworks reconnus (ISO 27001/27005, NIST, OWASP…) et des historiques internes détaillés (incidents de sécurité, résultats d'audits passés…).

À partir de cette analyse de base, des modèles de machine learning spécialisés ou des LLM entraînés spécifiquement sur les enjeux de sécurité évaluent le niveau de risque de chaque élément identifié.

Par exemple, ils peuvent auto-

matiquement mettre en lumière :

- Les exigences RGPD et obligations HDS dès la conception d'un projet manipulant des données personnelles et de santé

- Une API exposée publiquement avec une authentification insuffisamment robuste au regard des standards définis concernant les principes durcis d'authentification

- Un projet critique pour l'organisation dont la tierce maintenance applicative est opérée par un tiers non évalué jusque là

- Une combinaison de services cloud mal configurés (bucket S3 public couplé à une clé d'API hardcodée) automatiquement qualifiée comme critique

Mais l'apport véritablement différenciant de l'IA réside dans sa capacité à maintenir une sécurité mesurée en continu, tout au long du cycle de vie du projet :

1. Collecte automatisée des changements : Chaque modification de configuration, de configuration cloud, ou d'architecture déclenche instantanément une réévaluation contextuelle du niveau de risque. L'IA détecte proactivement les changements ou les ajouts critiques non anticipés (ouverture de ports non documentés, stockage non chiffré, élévation de privilèges non justifiée...).

2. Mise à jour dynamique des exigences : L'IA adapte intelligemment les contrôles nécessaires en fonction de l'évolution du projet, des modifications réglementaires ou des mises à jour des politiques internes de sécurité.

3. Tableaux de bord temps réel : Sur leurs périmètres respectifs, les Product Owners et CISOs disposent d'une visibilité précise et actualisée sur les mesures opérationnellement appliquées, les écarts détectés et les actions correctives en cours de déploiement.

Cette approche transforme radicalement la posture traditionnelle de sécurité. Elle n'est plus un accompagnement ponctuel constitué de contrôles manuels ou un simple livrable de fin de projet, mais devient un processus vivant, piloté en continu, mesurable objectivement, et parfaitement intégré aux dynamiques agiles et DevOps modernes.

La cyberdéfense révolutionnée : vitesse et intelligence au service de la détection

Si l'IA transforme la gouvernance des risques, c'est dans le domaine opérationnel de la cyberdéfense qu'elle déploie son potentiel le plus spectaculaire. La détection, l'analyse et la réponse aux incidents bénéficient d'un bond quantitatif remarquable en termes d'efficacité et de rapidité de traitement.

Bien au-delà des signatures

Les systèmes de détection conventionnels reposent sur la reconnaissance de menaces connues via des signatures ou des schémas préétablis que l'on cherche à identifier dans une masse d'activités une approche par définition limitée face aux menaces inconnues, aux attaques zero-day ou aux malwares polymorphes. L'IA introduit une logique d'analyse comportementale qui bouscule ces limitations historiques.

Les algorithmes d'apprentissage établissent des profils de normalité pour chaque composant du système d'information : utilisateurs, applications, flux réseau, accès aux données. Toute déviation significative par rapport à ces patterns déclenche une alerte, permettant de détecter des attaques zero-day ou des mouvements latéraux sophistiqués que les outils classiques auraient ignorés.

Cette approche dite comportementale s'avère particulièrement efficace contre les APT (Advanced Persistent Threat), ces attaques élaborées qui se veulent furtives dans les systèmes d'information. L'IA détecte les signaux faibles qui trahissent la présence d'un attaquant, même lorsque celui-ci mime parfaitement les comportements légitimes.

Prioriser l'humain là où il est essentiel

Les Security Operations Centers (SOC) croulent littéralement sous les alertes de sécurité. Cette profusion d'informations, loin de renforcer la sécurité globale, génère au contraire une fatigue cognitive des analystes et les empêche de distinguer efficacement les vraies menaces

critiques des nombreux faux positifs. L'IA devient un allié décisif en permettant la préqualification intelligente des alertes, une orientation contextuelle du diagnostic et la suggestion de scénarios d'investigation à engager, le tout priorisé selon la criticité réelle des événements détectés.

Des startups innovantes comme Qevlar AI en France développent des solutions particulièrement prometteuses sur ce marché émergent de l'IA appliquée aux SOC, démontrant le potentiel de transformation de cette approche.

Les agents IA analysent systématiquement le contexte multidimensionnel de chaque alerte générée : criticité des actifs concernés, nature précise de la menace détectée, historique détaillé des incidents similaires dans l'organisation, corrélation temporelle et logique avec d'autres événements au sein du système d'information.

Cette analyse contextuelle permet de prioriser automatiquement les incidents selon leur criticité réelle et d'orienter les analystes vers les menaces véritablement critiques, optimisant ainsi l'allocation de temps vers les tâches d'investigation complexes et les prises de décision voire la gestion de crise

De la détection à la neutralisation

La vitesse de réaction constitue un facteur déterminant dans la limitation de l'impact d'une cyberattaque. Chaque minute compte lorsqu'un attaquant progresse méthodiquement dans le système d'information, étend son périmètre de contrôle et exfiltre des données sensibles.

L'IA devient un levier technologique décisif dans cette course critique contre la montre. Dès qu'une menace avérée est détectée, l'IA est un formidable outil pour déterminer la stratégie de contre-mesures à appliquer de manière appropriée : isolation des machines compromises, blocage des communications malveillantes, révocation des accès suspects, sauvegarde des preuves à des fins d'investigations forensics. Cette orchestration automatisée réduit le temps de réaction de plusieurs heures à quelques secondes.

Vers une cybersécurité augmentée, orchestrée, intelligente

L'IA ne remplace pas l'expertise humaine : elle l'amplifie. Elle automatise les tâches à faible valeur, libère du temps pour l'analyse stratégique, et agit comme un système nerveux capable d'alerter, d'adapter et de réagir à chaque stimulus. L'avenir proche s'annonce encore plus prometteur avec l'émergence de protocoles standardisés comme le Model Context Protocol (MCP), activement soutenu par des acteurs technologiques majeurs tels que Google, CrowdStrike, Cloudflare, Wiz et Okta. Ces nouveaux standards techniques ouvriront la voie à une communication native et fluide entre intelligences artificielles, distribuées, décloisonnant définitivement les solutions de sécurité traditionnellement isolées.

Concrètement, cette évolution permettra qu'une IA liée à la détection d'intrusions (sur un EDR par exemple) communique directement et en temps réel avec une IA de gestion des identités et des accès (IAM) pour ajuster automatiquement et contextuellement les privilèges d'accès en fonction du niveau de risque détecté.

Parallèlement, les systèmes intelligents de cyberthreat intelligence alimenteront instantanément et automatiquement les playbooks de réponse incident, créant ainsi un écosystème défensif véritablement orchestré, adaptatif et résilient. Ce système défensif inter-

connecté et en partie automatisé ouvre la voie à une cybersécurité enfin véritablement à la hauteur de la menace actuelle: collective, temps réel, basée sur la donnée et orchestrée par l'IA.

Les défis de l'implémentation : vers une adoption responsable

Cette transformation ne s'opère pas sans défis significatifs. L'implémentation de l'IA en cybersécurité soulève des questions cruciales : les coûts d'acquisition et de déploiement restent substantiels, les compétences techniques requises sont encore rares sur le marché, et les risques de biais ou de faux positifs massifs nécessitent une vigilance constante.

Les organisations doivent également naviguer entre les enjeux de sécurité et les exigences de confidentialité des données, particulièrement dans des secteurs hautement régulés comme la santé ou la finance. Paradoxalement, les systèmes d'IA peuvent constituer eux-mêmes de nouvelles surfaces d'attaque au regard de leur complexité technique et la richesse des données qu'ils manipulent. Le model poisoning, par exemple, consiste à introduire des données malveillantes lors de l'entraînement pour altérer le comportement du modèle.

Le prompt injection, quant à lui, détourne le fonctionnement d'un modèle en manipulant les instructions en langage naturel, généralement pour exfiltrer des données. Autre scénario critique : les systèmes basés sur le Retrieval-Augmented Generation (RAG), qui s'appuient sur des bases de documents externes pour enrichir leurs réponses, peuvent devenir la cible d'attaques de type supply chain, si la source d'information est compromise ou manipulée. Ces menaces émergentes sont désormais systématisées dans le framework MITRE ATLAS (Adversarial Threat Landscape for Artificial-Intelligence Systems), une extension du célèbre MITRE ATT&CK, qui recense de manière structurée les tactiques, techniques et procédures (TTPs) employées par des acteurs malveillants pour cibler les systèmes d'IA.

Reprendre l'ascendant

La cybersécurité de demain ne sera ni une tour de contrôle centralisée et rigide, ni un simple patchwork d'outils techniques disparates. Elle reposera sur une architecture de données unifiée, analysée et orchestrée par l'intelligence artificielle, pilotée par l'humain selon le principe : "AI works, humans think".

Cette évolution marque un tournant décisif. Après des années de sécurité réactive, les organisations détiennent enfin les clés d'une cybersécurité nouvelle — capable d'anticiper, de mesurer les risques de manière réaliste, et de répondre aux menaces avec pragmatisme.

Plus qu'un simple changement technologique, c'est un changement de posture : une opportunité de repenser fondamentalement une approche de la cybersécurité résiliente et durable, à la hauteur des enjeux de nos sociétés en mutation.

Christophe Longuepez

"L'IA NE REMPLACE PAS L'EXPERTISE HUMAINE : ELLE L'AMPLIFIE "

VERS DES MODÈLES D'IA SOUVERAINS POUR LA CYBERSÉCURITÉ EUROPÉENNE

par **Maëva Astorga**

Aujourd'hui, la France et l'Europe se trouvent à un tournant décisif dans la course mondiale à l'intelligence artificielle. Le climat géopolitique, marqué par des tensions croissantes et des alliances fragilisées, pousse l'Europe à repenser son autonomie technologique.

Pour la France, cette souveraineté est bien plus qu'un enjeu stratégique : c'est une question de sécurité nationale, de compétitivité économique et de liberté d'action face aux grandes puissances étrangères.

Jamais la menace cyber n'a été aussi importante. Le dernier rapport de l'ANSSI sur l'état des cybermenaces en 2024 dresse un constat inquiétant : des attaques ransomware de plus en plus destructrices, avec la paralysie de nombreux centres hospitaliers, entreprises et administrations. Mais aussi beaucoup d'activités malveillantes générées à l'occasion d'événements à forte visibilité. Les Jeux Olympiques et Paralympiques de Paris 2024 ont concentré l'attention du monde entier sur la France, créant une cible de choix pour des cybercriminels souvent motivés par des idéaux politiques. On recense notamment des tentatives d'espionnage, des campagnes de désinformation, des tentatives de sabotage : de nombreuses occasions de tester la résilience locale.

De plus, un contexte politique sensible, avec des élections européennes et législatives, a rappelé combien la cybersécurité est aujourd'hui un pilier important de la sécurité nationale. Cette réalité renforce une conviction : pour faire face à des menaces extraordinaires et de plus en plus automatisées, la France et l'Europe devraient pouvoir compter sur des technologies qu'elles contrôlent entièrement. Cela passe notamment par le développement de modèles d'IA souverains, entraînés et hébergés sur des infrastructures maîtrisées, loin de la dépendance des géants américains ou chinois.

Une prise de conscience

Les nouvelles directives européennes, comme DORA et NIS2, encouragent déjà les entreprises à renforcer leur niveau de préparation et à mettre en place des plans de continuité solides. Mais cette exigence est incomplète et la promesse de souveraineté reste encore fragile sans solutions européennes pour stocker et traiter les données, ou entraîner des modèles d'intelligence artificielle de manière sécurisée.

Conscient des enjeux majeurs que la France doit affronter, Emmanuel Macron a annoncé en février 2025, lors du Sommet mondial sur l'intelligence artificielle, un plan d'investissement de 109 milliards d'euros. Une initiative qui devrait aider la France à imposer son positionnement sur les questions de souveraineté des technologies d'intelligence artificielle.

Ce financement doit soutenir la recherche et encourager l'innovation, mais surtout ai-

der à bâtir des infrastructures souveraines capables de concurrencer les plateformes américaines, qui sont dominantes aujourd'hui. Au-delà de la compétition économique, ce plan devrait aussi limiter la dépendance à des technologies étrangères, souvent perçues comme une source de risques pour la sécurité et la protection des données sensibles.

Des nouvelles initiatives industrielles majeures ont également été annoncées récemment. Parmi elles, les nouveaux projets de la startup française Mistral AI, qui est rapidement devenue un acteur important de l'intelligence artificielle générative en Europe. Fondée par Arthur Mensch, Mistral AI a annoncé le développement d'une infrastructure cloud européenne dédiée à l'IA, baptisée Mistral Compute, en partenariat avec Nvidia, le leader mondial des puces graphiques spécialisées pour l'IA.

Officiellement annoncée lors de la dernière édition du salon européen de l'innovation, VivaTech, la plateforme Mistral Compute se distinguera par son architecture entièrement hébergée et gérée en Europe, garantissant ainsi un contrôle local des données. Pour assurer la puissance de calcul nécessaire, ce projet s'appuie sur un parc de 18 000 puces Nvidia GB200, un standard de performance de très haut niveau. Nvidia fournit donc le matériel indispensable, et les parties logicielles et opérationnelles, elles, resteront sous maîtrise européenne, assurée par Mistral AI. Cette collaboration démontre que la souveraineté technologique est aujourd'hui un équilibre très complexe.

En effet, à ce jour, l'Europe ne possède pas encore toutes les capacités industrielles pour produire elle-même des composants aussi avancés, ce qui l'oblige, pour le moment, à maintenir des partenariats avec des acteurs étrangers.

La souveraineté technologique en IA ne peut pas se concevoir comme une indépendance totale. Aujourd'hui, l'industrie numérique est entièrement mondialisée et interconnectée. Les États ne sont pas encore en mesure de maîtriser l'ensemble des composantes nécessaires pour les logiciels, les serveurs ou encore les services cloud déployés. Cette interdépendance mondiale rend la question de la souveraineté numérique d'autant plus stratégique. Cependant, cette dépendance matérielle ne freine pas la dimension souveraine du projet, qui se concentre sur la maîtrise logicielle, le traitement sécurisé des données et la garantie d'une infrastructure conforme aux exigences européennes de sécurité et de confidentialité.

Ce projet ancre l'entrée de l'Europe comme un acteur clé de cette nouvelle ère de gouvernance technologique.

La construction de modèles d'IA souverains est bien plus qu'une question technologique. Disposer d'outils d'intelligence artificielle conçus et contrôlés localement permet de garantir une meilleure résilience face aux cyberattaques, en limitant les risques liés à l'intégration de composants étrangers. Cette souveraineté réduira également la dépendance aux fournisseurs étrangers, souvent soumis à des pressions géopolitiques, ce qui pourrait exposer les infrastructures critiques européennes à de nouvelles vulnérabilités et manipulations.

Des acteurs industriels français et européens ont déjà manifesté un intérêt pour accéder aux infrastructures souveraines comme Mistral Compute, avec des premiers déploiements prévus rapidement, dès 2026. La maîtrise technologique locale est un élément de compétitivité et de sécurité dans notre monde numérique en pleine mutation.

La démarche française, portée par Mistral AI, confirme une approche réaliste fondée sur la volonté de bâtir une autonomie progressive. L'intelligence artificielle s'est imposée comme alliée incontournable de la cybersécurité aujourd'hui, capable d'anticiper et d'atténuer les nouvelles menaces, souvent très sophistiquées et orchestrées à l'échelle internationale. Une infrastructure souveraine d'IA pour garantir la confiance dans nos systèmes européens est une priorité.

Nous devons assurer la protection permanente des données et conserver un contrôle efficace sur les technologies utilisées dans les secteurs stratégiques européens.

QUAND L'IA DÉFIE LES HACKERS

Par DAMIEN BANCAL

Damien Bancal est un expert internationalement reconnu en cybersécurité.

Il s'est imposé comme une figure majeure dans ce domaine. En 1989, il fonde ZATAZ contribuant ainsi à la sensibilisation et à la protection des internautes contre les cyberattaques. Il est l'auteur de plusieurs ouvrages ainsi que de plusieurs centaines d'articles qui explorent les divers aspects du piratage informatique et de la protection des données.

Il a remporté le prix spécial du livre du FIC/InCyber 2022. Finaliste 2023 du 1er CTF Social Engineering Nord Américain. Vainqueur du CTF Social Engineering 2024 du HackFest 2024 (Canada).

Damien Bancal a également été largement reconnu par la presse internationale dont le New York Times, qui souligne non seulement son expertise mais aussi son parcours inspirant.

IA et élite mondiale

L'entreprise Palisade Research a publié un rapport inédit sur le potentiel des IA dans le domaine de la cybersécurité offensive. Pour la première fois, des agents autonomes basés sur des modèles d'intelligence artificielle ont été intégrés à des compétitions internationales de type Capture The Flag (CTF), dans lesquelles les participants doivent résoudre des défis de hacking concrets.

Résultat, les IA ont brillé : dans certains cas, elles se sont classées dans le top 5 % des participants humains. L'étude explore l'idée que les performances réelles de l'IA ne peuvent être pleinement révélées que dans des environnements ouverts, collaboratifs et compétitifs. Ces expériences pourraient redéfinir la manière dont le potentiel de l'intelligence artificielle est évalué et audité à l'échelle mondiale.

Le rapport publié en mai 2025 marque un tournant dans l'histoire de la cybersécurité. Pour la première fois, des agents d'IA ont participé de manière autonome à des tournois CTF, où les compétences en hacking sont mises à rude épreuve.

Ces compétitions, bien connues dans le milieu de la sécurité informatique, opposent des milliers de participants dans des défis de cryptographie, d'analyse de code, de rétro-ingénierie ou encore d'exploitation de vulnérabilités.

Et depuis trois ans, l'IA a débarqué en force dans les CTF. Exemple, en 2024, lors du CTF de Social Engineering du Def Con de Las Vegas et du Hackfest de Québec l'IA était incluse dans les épreuves.

Lors du tournoi « AI vs Humans« , des agents conçus pour opérer sans intervention humaine ont atteint le top 5 % des scores. Plus encore, au cours de la compétition « Cyber Apocalypse », à laquelle participaient plus de 8000 équipes professionnelles, les agents IA se sont hissés dans le top 10 %.

Ces performances suscitent l'étonnement, d'autant plus qu'elles ont été obtenues dans des conditions de compétition en temps réel.

Ce type de résultats confirme les observations récentes de plusieurs chercheurs : les modèles de langage actuels, s'ils sont correctement configurés, peuvent rivaliser avec des experts humains sur des problèmes techniques d'une durée allant jusqu'à 60 minutes.

Au cœur de l'expérience, une hypothèse simple : les tests internes menés en laboratoire sous-estiment systématiquement les capacités réelles des systèmes d'IA. Pour y remédier, les chercheurs ont appliqué un principe de crowdsourcing, laissant des équipes extérieures prendre en main les agents IA et les intégrer aux compétitions ouvertes.

Sur certaines épreuves, l'IA résolvait en quelques minutes des tâches qui mobilisent un humain expérimenté pendant près d'une heure

Ce mode d'évaluation, qualifié de « méthode d'élicitation », vise à libérer tout le potentiel du système en le confrontant à des scénarios imprévisibles, dans un environnement de forte pression.

L'idée est aussi de combler ce que les auteurs appellent l'« evals gap », autrement dit le fossé entre les résultats de tests standardisés et les performances que l'IA peut atteindre dans des contextes dynamiques et concrets. Contrairement aux benchmarks fermés, les compétitions CTF offrent une diversité de problèmes, une incertitude réelle et une dimension temporelle, autant d'éléments cruciaux pour jauger les compétences d'un système autonome. Les IA ont particulièrement bien performé dans les domaines de la cryptographie et du reverse engineering, deux disciplines qui exigent rigueur logique, manipulation binaire et exploration systématique.

Ces résultats laissent entrevoir des applications dans les tests de sécurité automatisés, mais aussi dans la détection avancée de failles.

Vers un audit public et transparent

Au-delà des résultats techniques : comment auditer de manière crédible les capacités grandissantes de l'IA ?

Jusqu'ici, les évaluations sont principalement effectuées par les entreprises elles-mêmes, dans des environnements fermés et avec des protocoles peu transparents. Or, cette opacité devient problématique alors que les IA gagnent en puissance et en autonomie.

Les auteurs du rapport défendent l'intégration systématique de « tracks IA » dans les compétitions déjà existantes. En insérant des agents autonomes dans les mêmes conditions que les joueurs humains, les organisateurs peuvent observer leurs performances dans un cadre rigoureux, compétitif et reproductible.

Cette démarche vise également à sensibiliser les décideurs politiques, les agences de régula-

tion et les entreprises technologiques. À terme, les auteurs suggèrent que ce type de mécanisme pourrait devenir une forme de « certification par le défi« , un processus dans lequel une IA ne serait pas évaluée sur des métriques internes, mais sur sa capacité à résoudre des problèmes concrets dans un environnement contrôlé mais ouvert.

Quand l'IA devient votre coéquipier de CTF

Si l'IA s'invite en force dans la cybersécurité et les compétitions de type CTF ce n'est pas pour faire joli. Son apport n'a rien d'anecdotique. Depuis plusieurs années je vois l'IA s'inviter dans les compétitions d'Hacking Éthique, que ce soit lors de l'European Cyber Cup (organisée pendant le forum In-Cyber), du Hackfest de Québec, ou encore de LeHack à Paris, on voit émerger un nouveau réflexe chez les joueurs : travailler main dans la main avec une IA.

Prenons un exemple concret. En octobre 2024, lors du CTF Social Engineering de Québec, une compétition que j'ai eu le privilège de remporter, l'une des épreuves imposait l'utilisation de ChatGPT pour concevoir un scénario d'attaque crédible. Preuve, s'il en fallait encore, que l'IA s'est définitivement ancrée dans la réalité des CTFs et de la formation cyber.

Après tout, les pirates informatiques exploitent l'IA à outrance depuis plusieurs mois, il serait inconscient que les "protecteurs" ne fassent pas de même. Mais comment cela se traduit-il concrètement sur le terrain ?

Voici deux témoignages :

Valentine, 22 ans : « L'IA, c'est mon binôme de nuit ».

Cette étudiante en deuxième année de master cybersécurité, Valentine (pseudonyme) n'imagine plus travailler sans IA : « L'IA, c'est mon binôme de nuit quand je révise mes labos », lâche-t-elle dans un sourire fatigué.

Tout a commencé simplement pour cette jolie brune à la tête bien pleine : pour résumer des articles ou générer des scripts Bash. Mais très vite, l'outil a pris une autre dimension dans son quotidien : « Quand je fais du reverse ou que je dois comprendre un malware en Python ou PowerShell, j'utilise un agent que j'ai entraîné à repérer des patterns typiques de scripts malveillants. Il m'explique ligne par ligne, me donne des pistes de détection, et propose même des payloads pour tester en environnement isolé. C'est comme un mentor patient… et qui ne dort jamais. »

Cependant, Valentine reste lucide : « Ce n'est pas un raccourci magique. Je dois toujours comprendre ce que je fais. Mais l'IA m'aide à poser les bonnes questions, à gagner du temps, et à développer mes réflexes de cyber analyste plus vite. »

Valentin, la trentaine bien tassée : « Mes bots suiveurs bossent pendant que je cogite ». J'ai croisé Valentin (pseudonyme, lui aussi) lors du rendez-vous estival LeHack. Vétéran des plateformes Hack The Box et Root-Me, il intègre depuis longtemps l'IA dans sa routine offensive. « Pourquoi perdre du temps à brute-forcer un format de sérialisation si une IA peut le décoder pour moi ? »

Pour lui, chaque seconde gagnée est une victoire : « Sur un challenge web, l'IA me repère une injection XPath ou une SSRF que j'aurais mis 15 minutes à tester à la main. ». Il a même un petit nom pour ses agents : « Je les appelle mes bots suiveurs. Quand je lance un challenge, je les configure pour scanner le code, identifier les points d'entrée, voire générer des exploits de base. Ensuite, je me concentre sur la logique du défi. »

Mais comme Valentine, il tient à rappeler un point essentiel : « Le but, ce n'est pas de tricher ou de laisser l'IA jouer à ma place. C'est d'aller plus vite sur les phases mécaniques. L'IA me donne un script de décodage pour un fichier protobuf (Protocol buffers) ou un ZIP obfusqué, et moi, je garde l'esprit clair pour la stratégie. »

"Chaque CTFs est une victoire. On apprend. On partage"

IA vs humains : une compétition de plus en plus serrée

Ces témoignages trouvent un écho saisissant dans le rapport "Evaluating AI cyber capabilities with crowdsourced elicitation". L'étude montre que l'intégration de l'IA dans les CTFs, notamment lors de compétitions récentes comme AI vs Humans CTF (mars 2025, Hack The Box) ou Cyber Apocalypse et ses 8 000 équipes, est en train de redéfinir les règles du jeu. Lors de ces épreuves, les IA n'étaient pas juste des gadgets. Elles ont atteint le top 5 % du classement général. Quatre IA sur sept ont résolu 19 défis sur 20. Leur vitesse d'exécution rivalise déjà avec les meilleures équipes humaines. Valentin avait donc raison : l'IA est un coéquipier efficace, capable de suggérer rapidement les pistes d'attaque les plus prometteuses.

Mais l'étude nuance aussi l'efficacité de l'IA. L'Intelligence Artificielle est une "feignante". Elle brille sur les tâches nécessitant plus d'une heure d'effort humain. Elles s'essoufflent sur les défis interactifs, ou ceux qui exigent une stratégie multi-étapes. Cela rejoint la vision de Valentine : « L'IA est utile pour comprendre, initier, apprendre. Pas pour remplacer la réflexion. »

Et demain ? Jusqu'où iront ces IA ?

Certaines, jugées peu performantes il y a encore quelques mois, explosent désormais les scores grâce à l'ajustement de leur environnement. Exemple : GPT-4o, qui passe de 40 % à 92 % de réussite sur InterCode-CTF, simplement grâce à un meilleur « harness » (une interface d'exécution logicielle adaptée et optimisée).

Alors, faut-il craindre une invasion des IA dans les CTFs ? Pas encore. Mais s'il fallait résumer la tendance : l'IA ne remplace pas le cerveau, elle booste le cerveau.

Et dans une discipline où chaque seconde compte… c'est déjà énorme.

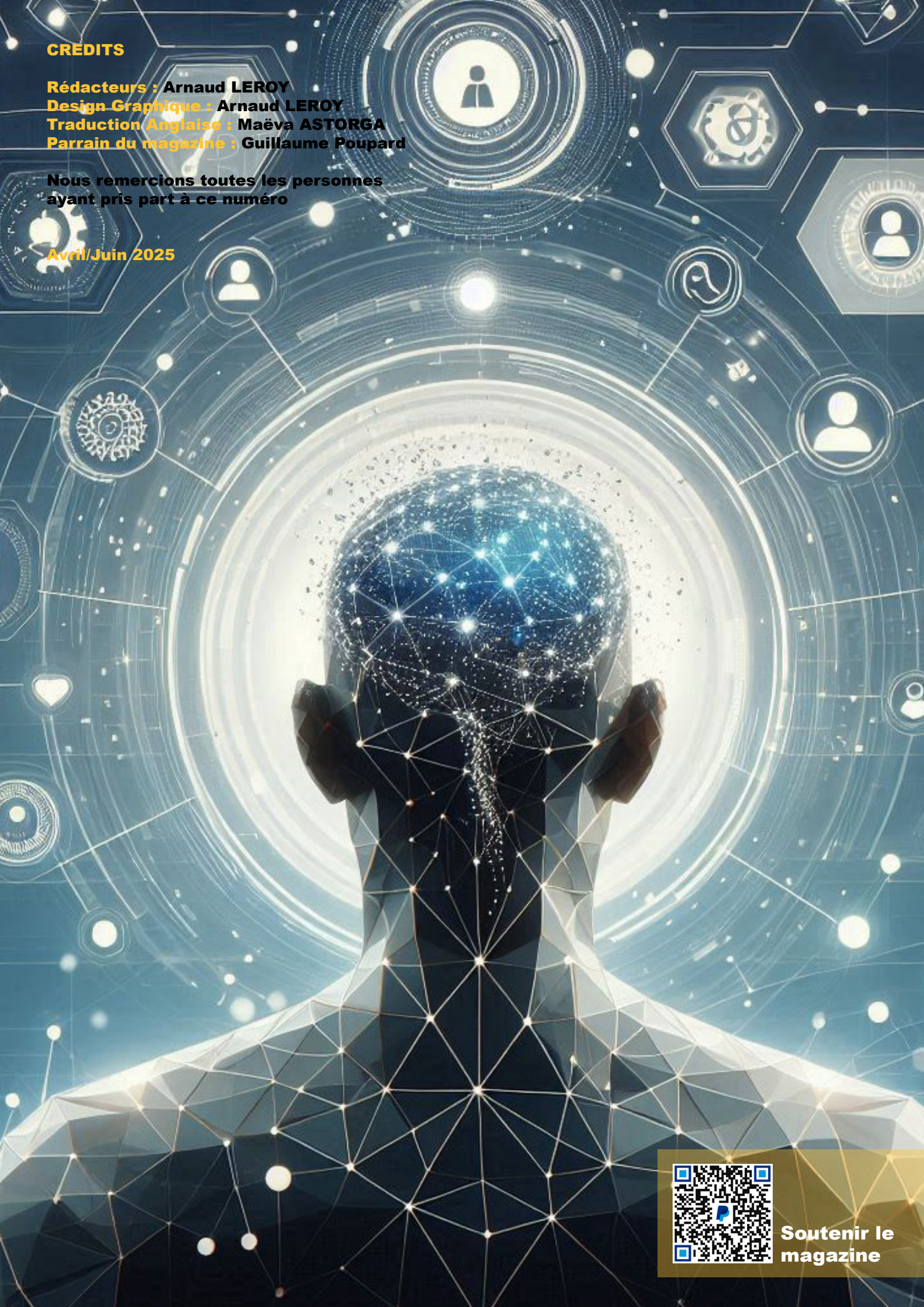
Une révolution dans la cybersécurité offensive ?

Le succès des IA dans les tournois CTF pourrait ouvrir un nouveau chapitre dans la cybersécurité offensive. Si les performances actuelles sont encore dépendantes de l'ingénierie humaine pour la configuration des agents, les marges de progression restent considérables. À mesure que les modèles deviennent plus puissants, leur autonomie dans les processus de résolution s'améliore.

Cela soulève des questions sensibles, notamment sur la dualité des usages. Un système capable de détecter et d'exploiter des vulnérabilités avec une telle efficacité pourrait, s'il tombait entre de mauvaises mains, être utilisé à des fins malveillantes.

Les chercheurs n'ignorent pas ce risque, mais estiment qu'un audit public des performances est aussi un moyen de prévenir ces dérives en rendant visibles les avancées.

DAMIEN BANCAL



CREDITS

Rédacteurs : Arnaud LEROY
Design Graphique : Arnaud LEROY
Traduction Anglaise : Maëva ASTORGA
Parrain du magazine : Guillaume Poupard

Nous remercions toutes les personnes ayant pris part à ce numéro

Avril/Juin 2025

Soutenir le magazine