CAMPUS CYBER - NOTE D'ANALYSE
Mai 2026

MYTHOS ET AUTRES MODÈLES-FRONTIÈRE :

IMPLICATIONS DES PROGRÈS DE L'IA POUR LA CYBER EN FRANCE ET EN EUROPE

CAMPUS CYBER

SOMMAIRE.

INTRODUCTION.

La médiatisation des performances du nouveau modèle Mythos d'Anthropic donne à la communauté cyber une fenêtre d'opportunité décisive pour prendre rapidement et collectivement la mesure des implications majeures de l'IA dans le domaine de la cybersécurité. Après Mythos, personne n'a et n'aura plus le droit d'être surpris.

L'irruption de l'IA dans la cyber n'est certes pas nouvelle ; mais la menace de l'IA offensive est restée jusqu'à présent essentiellement virtuelle, en raison :

- du caractère encore marginal des attaques attribuées à l'intelligence artificielle rapportées à l'incidentologie observée ;
- d'une trajectoire de progrès des grands LLM qui, bien que prévisible, est restée relativement en marge du champ de vision opérationnel ;
- de la priorité donnée à l'IA avant tout comme un objet métier à déployer en entreprise.

Mythos cristallise et précipite plusieurs évolutions préoccupantes quoique prévisibles :

- l'IA avale les paliers de performance en cyber à un rythme très soutenu, au point d'approcher du seuil de l'avantage (supériorité des modèles par rapport aux experts humains) ;
- le niveau atteint par l'IA est tel que le débordement des modèles, du laboratoire vers le marché, crée un risque systémique pour tous les systèmes de cyberdéfense ;
- au-delà du moment Mythos, la donne en sera profondément et irréversiblement modifiée, tant du point de vue de la sécurité des systèmes d'information que de celui du paysage industriel de la cybersécurité en Europe.

S'il est important de ne pas céder à l'angoisse déclenchée par Mythos, il est essentiel de ne pas non plus sous-estimer la trajectoire de l'IA. L'enjeu le plus immédiat est la transformation rapide de la cyberdéfense : détection massive de vulnérabilités ; test automatisé de chemins d'attaque ; génération ou évaluation de correctifs ; analyse de dépendances ; durcissement logiciel ; réduction des cycles de remédiation ; rapprochement entre sécurité et développement logiciel.

La note qui suit, réalisée (sans IA) grâce à l'expertise et au regard croisés de plusieurs catégories d'acteurs mobilisés par le Campus Cyber (DSI, RSSI, spécialistes en intelligence sur la menace cyber, en gestion de crise, en pentesting et en audit, conseil en transformation numérique…), **a un triple objet** :

- synthétiser les éléments d'analyse rassemblés à date au sein de l'écosystème du Campus Cyber ;
- identifier les principales implications des progrès de l'IA en cyber, au-delà du seul cas Mythos, pour les utilisateurs finaux comme pour les offreurs de solutions ;
- amorcer la structuration d'une réponse collective d'ampleur, au carrefour des deux communautés cyber et IA.

1 LLM : Large Language Models.

Elle s'adresse délibérément à plusieurs types de lecteurs, dont les horizons d'action diffèrent : les RSSI et DSI confrontés dès aujourd'hui aux implications opérationnelles de Mythos ; les dirigeants et membres de comités exécutifs appelés à arbitrer rapidement sur des sujets structurels ; et les décideurs publics et régulateurs pour qui Mythos cristallise des enjeux de souveraineté et de régulation qui dépassent le seul moment de crise. Cette pluralité de destinataires est assumée : elle reflète la nature même du défi posé par l'IA en cyber, qui ne peut être relevé ni par les seules équipes techniques, ni par les seuls décideurs politiques, mais exige une réponse coordonnée à tous les niveaux. Le lecteur trouvera dans les sections qui suivent une progression allant de l'analyse du phénomène Mythos aux implications opérationnelles immédiates, puis aux enjeux structurels de moyen terme pour l'Europe.

1 - LE MODÈLE MYTHOS : UN GAIN DE PERFORMANCE PROBABLEMENT AVÉRÉ, SOUS RÉSERVE D'UNE CONTRE-EXPERTISE EUROPÉENNE INDÉPENDANTE

1.1. La sortie de Mythos début avril et les premières évaluations disponibles semblent attester d'un réel saut de performance dans les capacités cyber des modèles d'IA.

Les analyses de l'UK AISI[2], qui a eu accès à Mythos, confirment une amélioration significative sur les tâches de découverte et d'exploitation de vulnérabilités. Les premiers effets concrets sont déjà visibles, avec la publication de correctifs de sécurité directement attribuables à des découvertes assistées par IA, y compris sur des logiciels largement répandus comme Firefox[3]. Les équipes de certaines entreprises américaines de cybersécurité, partenaires de Glasswing[4], affirment qu'elles utilisent déjà l'IA pour tester la robustesse de leurs produits logiciels avant commercialisation, avec des gains d'efficience significatifs (d'après l'une d'entre elles, 1 an de pentest humain serait comprimé en 3 semaines seulement grâce à Mythos). Ces éléments factuels justifient à eux seuls de prendre le problème au sérieux.

1.2. L'« IA Mythos » est indissociable de la dimension publicitaire de l'« opération Mythos », qui a bénéficié rapidement d'une résonance mondiale.

Le produit mis en avant à cette occasion est d'ailleurs moins le modèle que l'entreprise Anthropic elle-même dans sa quête de suprématie sur l'IA à usage général, dans un contexte où sa rivalité avec OpenAI est mise en scène à intervalles réguliers[5]. Le phénomène marketing Mythos sert d'abord des objectifs économiques et technologiques privés qu'il serait naïf de sous-estimer au vu des enjeux considérables associés au développement des modèles d'IA de frontière.

La mise en scène de l'alliance Glasswing dans une courte vidéo[6] associant plusieurs personnalités américaines du monde de l'IA et de la cyber illustre la pleine exploitation des codes de la communication grand public (peur, dramatisation …) au bénéfice des produits de l'entreprise et de ses partenaires. Dans le même registre, la dimension philanthropique est centrale dans l'orchestration médiatique, à travers le recours au mythe du sauveur, qui transparaît dans l'engagement de retenir temporairement le modèle pour en border les futurs effets et conserver un avantage à la défense. Cette dimension est d'ailleurs en partie crédibilisée par la mise à disposition du modèle à certains acteurs de l'open source et par des dons de tokens. Pour autant, cela ne doit éclipser ni la primauté des enjeux de marché et de domination géopolitique qui sous-tendent Mythos ni les risques systémiques qu'il soulève.

2 https://www.aisi.gov.uk/blog/our-evaluation-of-claude-mythos-previews-cyber-capabilities
3 https://blog.mozilla.org/en/privacy-security/ai-security-zero-day-vulnerabilities/
4 L'alliance Glasswing est une initiative conjointe d'Anthropic et de 11 entreprises américaines (initialement : AWS, Google, Broadcom, Crowstrike, Cisco, Nvidia, JP Morgan Chase, Microsoft, Palo Alto, Apple, The Linux Foundation) visant à réserver pendant au moins 3 mois l'usage de Mythos Preview aux seuls partenaires afin qu'ils puissent tester et renforcer la robustesse de leur système de protection informatique face au modèle. Depuis le lancement, le cercle des partenaires s'est élargi.
5 OpenAI a d'ailleurs annoncé dès le 23 avril (soit deux semaines à peine après celle de Mythos Preview) la sortie de son modèle GPT 5.4. Cyber, suivi début mai de sa version 5.5.
6 https://www.youtube.com/watch?v=INGOC6-LLv0
7 https://fortune.com/2026/04/23/anthropic-mythos-leak-dario-amodei-ceo-cybersecurity-hackers-exploits-ai/

1.3. A date et en dépit des fuites présumées rapportées par la presse[7], la principale source d'information sur le modèle Mythos provient de l'entreprise Anthropic elle-même et de ses partenaires.

Les rares sources secondaires sont exclusivement américaines (ISAC Finance, Cloud Security Alliance...) et britanniques (UK AISI, déjà citée). Pour l'heure, nous ne disposons pas, sur le modèle, d'informations objectives et fiabilisées par le recours à une expertise européenne indépendante. Il reste bien sûr possible de produire des estimations indépendantes à partir de données disponibles publiquement et de modèles de prédiction ; mais le contraste reste saisissant entre l'ampleur prise par la communication autour de Mythos et notre capacité d'analyse limitée, faute d'accès direct au modèle lui-même. La forte dimension économique des annonces faites par Anthropic accroît la difficulté à distinguer ce qui relève de l'information technique vérifiée de la communication de marché. C'est ce qui explique en partie le malaise actuel ressenti par les acteurs de la cybersécurité lorsqu'ils sont interrogés, par leurs dirigeants, leurs actionnaires ou leurs collaborateurs, sur leur lecture de l'algorithme et de ses résultats. Aussi longtemps que cette situation d'asymétrie perdurera, les écosystèmes français et européens de l'IA et de la cyber devront maintenir une posture cartésienne et redoubler d'exigences sur la transparence, l'accessibilité et la contestabilité du modèle. En attendant, les chiffres relayés sur la taille du modèle, les paramètres d'entraînement ou les montants engagés doivent donc être pris avec précaution.

2 - LE CAS MYTHOS RAPPELLE QUE LES GARDE-FOUS DES MODÈLES D'IA ACTUELS RESTENT INSUFFISANTS

Concernant la sécurité des modèles eux-mêmes :

- les mécanismes actuels reposent largement sur du reinforcement learning, avec récompenses positives ou négatives qui poussent vers des logiques de dissimulation de la part des IA ;
- lorsqu'un comportement anormal est détecté, le modèle peut être corrigé ou réentraîné. Cela réduit une probabilité d'occurrence du risque, mais ne donne pas de garantie forte. Les filtres peuvent être contournés si les demandes sont reformulées ;
- le monitoring des chaînes de raisonnement devient moins robuste si les modèles apprennent à dissimuler certains raisonnements ou si l'accès à ces chaînes se réduit.

Ces éléments suggèrent que les « guardrails » (garde-fous de sécurité) des modèles d'IA restent largement perfectibles, ce qui crée à court terme un double risque : perte de visibilité pour les évaluateurs et apparition de comportements de contournement ou de dissimulation. Cela renforce l'idée que nous ne pouvons pas considérer les modèles de frontière, y compris ceux présentés comme défensifs, comme naturellement fiables.

A cela s'ajoute **la problématique de la fiabilité des données d'entraînement** : en cybersécurité, les données massives issues d'Internet ne peuvent fournir un niveau de garantie de sécurité suffisant car toute vulnérabilité intrinsèque ou backdoor peut devenir un chemin d'attaque potentiel.

3 - SI DES DOUTES ENTOURENT LE MODÈLE LUI-MÊME, IL N'Y EN A PAS SUR LA DIRECTION PRISE PAR L'IA EN CYBER ET LES PERTURBATIONS MAJEURES A VENIR

3.1. Bien qu'il soit encore impossible de démêler le vrai du faux sur les performances réelles du modèle, plusieurs certitudes se dégagent – il faut d'ailleurs rendre crédit à l'opération Mythos de les mettre en lumière.

Première certitude : la performance des grands modèles d'IA continue de s'améliorer à un rythme très soutenu, dans le prolongement des tendances observées au cours des dernières années, au point d'effacer la distinction classique entre perfectionnement quantitatif et qualitatif des LLMs. Mythos n'est pas une anomalie statistique, ni même une découverte, mais ni plus ni moins qu'un point de plus sur la courbe de performance exponentielle de l'IA testée en laboratoire sur des tâches cyber. Les modèles précédents étaient déjà capables de détecter des vulnérabilités. L'identification de failles anciennes - parfois présentées comme remontant à plus de vingt ans - est plausible, mais pas fondamentalement surprenante pour les praticiens. **La différence introduite par Mythos se situe à deux niveaux :** dans la combinaison capacitaire opérée par le nouveau modèle (détection de vulnérabilités, exploitation, raisonnement, priorisation, chaînage d'informations banales pour reconstruire un chemin d'attaque cohérent, passage d'une logique de test ponctuel à une logique de découverte automatisée à grande échelle) et dans l'accélération de la *timeline* de l'IA.

La dynamique actuelle ne va probablement pas se démentir et fera oublier Mythos (ou ses concurrents d'OpenAI) d'ici peu. C'est moins l'étape qui compte que le mouvement dans lequel elle s'inscrit, avec l'arrivée régulière (en streaming) de modèles capables d'opérations toujours plus complexes et plus rapides. Mythos doit donc être dé-« mythifié » : prendre le modèle moins pour ce qu'il est et ce qu'il est capable de faire que pour ce qu'il dit et ce qu'il révèle des grandes lignes de force qui refaçonnent le paysage de la cybersécurité sous nos yeux. L'essentiel n'est pas dans la manifestation du phénomène à un instant t, mais dans la pente de la courbe. Il faut appréhender Mythos moins comme une rupture technique que comme le dernier signal en date d'une accélération très forte des modèles.

Seconde certitude : à mesure que les performances des modèles poursuivent leur marche ascensionnelle, la probabilité que les impacts systémiques de l'IA se matérialisent dans toute leur portée en cybersécurité augmente. En d'autres termes, Mythos rend plus crédible et plus tangible un scénario très pénalisant, présent dans toutes les têtes de la cyber, qui se caractériserait par i) l'arrivée sur le marché ii) en libre accès ou équivalent iii) à très court terme (sur un pas de temps qui se compte en semaines) iv) d'un ou plusieurs modèles d'IA de frontière capables de déclencher une campagne mondiale massive de détection et de révélation de nouvelles vulnérabilités répandues dans une multitude de systèmes d'information v) y compris dans des infrastructures et actifs critiques pour la continuité économique vi) sans offrir dans le même temps et à la même échelle les solutions correctives correspondant aux failles identifiées. Ce scénario n'a rien d'original dans son concept. En ce sens, Mythos ne « noircit pas le tableau ». En revanche, il a pour effet de relever la probabilité de passage de ce seuil et de nous rapprocher du moment de sa réalisation : à cet égard, vu le décalage habituel entre la sortie des modèles frontières et leur arrivée dans le domaine public, on peut s'attendre à ce que des modèles open source équivalents, d'origine chinoise par exemple, soient disponibles d'ici à la fin de l'année 2026 (dans les 6 à 10 mois, à dire d'expert).

Troisième certitude : si un tel scénario post-Mythos ne modifie pas fondamentalement les caractéristiques irréductibles de la cybersécurité (dialectique permanente du « glaive et du bouclier » et « course contre la montre » entre attaque et défense), il n'en risque pas moins de changer la donne de manière profonde et irréversible. Là où l'intelligence artificielle n'était jusqu'à présent qu'une menace à l'état de potentiel et un paramètre relativement marginal dans l'équation de la cyber du fait de sa faible utilisation en offensif, **elle devrait acquérir une centralité nouvelle dans le paysage de la cybersécurité, avec le pouvoir de remodeler à elle seule les grands équilibres de la cyber.**

Il n'y a pas de raison valable de penser que la cybersécurité serait le seul secteur de l'économie à rester durablement à l'abri des effets disruptifs de l'IA. Au contraire, la cyber figure probablement parmi les premiers champs où ces effets étaient amenés à être tangibles, les modèles étant explicitement optimisés pour être performants en programmation, ce qui se traduit mécaniquement par des capacités avancées dans ce domaine. La progression accélérée de l'IA et son industrialisation introduiront au cœur même de la cyber des dynamiques qui lui sont étrangères, susceptibles de rendre obsolètes les architectures de sécurité actuelles des systèmes d'information. La trajectoire globale de la cybersécurité s'annonce donc surdéterminée par celle de l'IA, ce qui crée une nouvelle forme de dépendance trans-sectorielle, lorsqu'un pan entier de l'économie comme la cybersécurité doit importer en son sein un nouvel environnement technologique, stratégique, opérationnel, culturel, linguistique et humain, façonné ailleurs. Il s'agit d'une rupture complète, encore plus immédiate et prégnante que celle provoquée par le quantique.

4 – CE QUE MYTHOS A DÉJÀ CHANGÉ : UN SUR-STRESS CONJONCTUREL SUR LES ÉQUIPES CHARGÉES DES SYSTÈMES D'INFORMATION ET DE LA CYBER, ASSIMILABLE A UNE PRÉ-CRISE

4.1 Avant même sa commercialisation, Mythos génère une surtension sur l'ensemble d'une chaîne de cyberdéfense déjà largement saturée.

Depuis l'annonce du modèle, les responsables des systèmes d'information et les équipes cyber font face à de multiples sollicitations en interne, par leur gouvernance, parfois au niveau des comités exécutifs, et en externe par les médias, pour apporter des éclairages sur des questions auxquelles ils n'ont bien souvent pas de réponse définitive. Étant donné les incertitudes autour de Mythos et des bouleversements opérationnels induits par l'IA, ils doivent se positionner dans un contexte non stabilisé, fluctuant et hypothétique : soit ils doivent admettre un manque d'informations, qui les empêche de formuler une opinion éclairée qui fragilise leur posture ; soit ils sont poussés à prendre des positions alimentant involontairement la « hype », faute d'éléments de fond allant dans le sens inverse. Ils ont aussi la sensation de naviguer entre deux mondes devenus poreux, celui de leur feuille de route cyber courante (business as usual), qui reste leur horizon de court terme, et celui de l'IA de masse, un horizon projeté mais qui s'impose à eux malgré ses contours flous. Ils ont conscience de devoir se préparer à gérer une série d'événements susceptibles de mettre à bas tout l'édifice de sécurité construit dans le cadre existant de gestion des risques, sans pouvoir en déterminer précisément ni la temporalité ni l'ampleur.

4.2. Les circonstances exceptionnelles créées par Mythos ont fait basculer les organisations dans des configurations analogues à de la gestion de crise (sentiment d'urgence déclenchée par des facteurs exogènes, nécessité de prendre des décisions rapides en situation d'asymétrie d'informations, immixtion des chaînes de gouvernance politiques, prégnance des enjeux communicationnels). Cette vraie-fausse crise larvée, immédiate et différée à la fois, a un impact psychologique sur des équipes DSI/RSSI déjà structurellement sous haute tension et qui hésitent légitimement à réallouer de la bande passante à un problème qui n'a pas *encore* une forte réalité opérationnelle.

5 – L'ENJEU LE PLUS IMMÉDIAT (MOINS D'UN MOIS) POUR LES ORGANISATIONS : DURCIR LEUR POSTURE GLOBALE DE GESTION DES RISQUES CYBER

5.1. La posture de gestion des risques des organisations est condamnée à évoluer, vite et fort. L'adaptation était inéluctable, sa timeline se comprime. L'incertitude et les doutes qui entourent Mythos ne doivent en aucun cas servir de prétexte à un quelconque immobilisme ou à une sous-estimation des risques liés à l'IA. Les organisations humaines pèchent bien souvent par myopie, par inertie et par excès de conservatisme. L'épisode Mythos plaide au contraire pour une approche intégrant des scénarios extrêmes. C'est d'autant plus difficile à assumer et à faire accepter que la menace cyber liée à l'IA, à date, a une empreinte faible par rapport au volume d'incident observé. Mythos doit être une leçon de prospective courte, orientée vers l'action préventive. L'opération Mythos a le mérite de repositionner la cybersécurité comme pilier de la résilience, dans son rôle d'assurance de la continuité opérationnelle de l'entreprise, au-delà de sa dimension purement défensive. Elle constitue une piqûre de rappel indispensable pour vérifier dans chaque organisation que les fondamentaux de l'hygiène cyber sont bien en place (segmentation, sauvegardes robustes, capacité de réponse à incident, continuité d'activité …).

5.2. Les acteurs de la sécurité de l'IA doivent anticiper à très court terme des modèles capables de saturer les *benchmarks* existants en cybersécurité. Ce que nous savons mesurer aujourd'hui va devenir rapidement insuffisant, car les meilleurs modèles obtiendront des scores trop élevés pour permettre une évaluation fine. Cela couvre tant l'analyse des codes source que l'analyse des langages binaires des softwares, les logiciels exécutables, les infrastructures, les dépendances et les environnements complexes.

5.3. Les dirigeants d'organisations doivent se préparer à une vague (certains experts n'hésitent pas à parler de « déluge ») de découverte massive de zero days mises au jour par l'IA, avec un effet de purge - en un seul bloc - de vulnérabilités anciennes logées dans des logiciels largement répandus et dans des systèmes complexes et plus anciens (ex. environnements bancaires legacy, systèmes industriels …). Ce pic, que les experts jugent possible d'ici 3-6 mois, pourrait être suivi de plusieurs répliques, par paquets de plus petite taille, puis par une longue traîne de disclosures au compte-gouttes. Au passage, selon certains spécialistes, la massification des vulnérabilités risque de provoquer une saturation des mécanismes existants de décompte des CVE : certains s'interrogent déjà, dès 2026, sur l'intérêt de continuer à répertorier une volumétrie totale de CVE. Cela pose une question incidente sur l'adéquation de la gouvernance actuelle des CVE à la vague IA à venir.

5.4. Le jour où la vague arrivera, l'agenda opérationnel des équipes informatiques sera bouleversé. Les organisations se retrouveront face à la nécessité de corriger en urgence un volume potentiellement considérable de vulnérabilités, pour laisser aux attaquants aussi peu de prise et de temps que possible pour les exploiter, dans un contexte où le TTE (time-to-exploit) ne fait que baisser depuis de nombreuses années[8]. **A ce moment-là, c'est l'ensemble de la chaîne logicielle, et plus largement de la chaîne opérationnelle de l'IT, qui sera mise sous une tension inédite, du fournisseur à l'intégrateur jusqu'au client.**

8 https://zerodayclock.com/

Plusieurs problématiques risquent alors d'apparaître :

- d'une part, un risque sur la capacité des éditeurs de logiciels à suivre le rythme, c'est-à-dire à assurer la disponibilité sans délais de solutions correctives auprès de leurs clients ;

- d'autre part, une hypothèque sur la capacité des équipes informatiques (aussi bien internes que celles des intégrateurs) à absorber, dans leur plan de charge, un « mur de patching en flux continus » (« patch streaming » ou « patch every day »), sans dégrader la continuité opérationnelle des métiers ni les autres fonctions-clefs de la cybersécurité. La question se pose d'autant plus que les ressources informatiques des entreprises sont de plus en plus contraintes i) du fait du contrôle étroit des budgets et ii) parce que les protocoles d'implémentation des patchs logiciels sont parfois peu adaptés à des situations d'urgence (notamment en raison des réticences des équipes métiers et IT à automatiser entièrement les processus pour éviter les incidents, nombreux, causés par des patchs trop expéditifs).

5.5. Au-delà des tensions opérationnelles, l'épisode Mythos met en exergue une dépendance critique à une chaîne de valeur logicielle encore insuffisamment cartographiée et maîtrisée, en particulier vis-à-vis des intégrateurs et fournisseurs, prolongeant ainsi les « signaux faibles » observés lors de l'incident CrowdStrike de juillet 2024. La cyber compense aujourd'hui une partie des faiblesses structurelles de la qualité logicielle. Les vulnérabilités ne sont pas toujours des CVE clairement identifiées, elles peuvent provenir de dépendances open source, de packages tiers, de composants embarqués ou de mises à jour mal maîtrisées. La mise à jour continue n'est d'ailleurs pas toujours une solution si elle est mal gouvernée. Dans certains cas, rester temporairement sur une version antérieure mieux maîtrisée peut être moins risqué que d'intégrer trop vite une dépendance compromise ou insuffisamment testée.

Avec les nouveaux modèles d'IA :

- du côté de l'utilisateur, l'enjeu se déplace vers la capacité à scanner plus tôt, plus vite et plus intelligemment les codes, dépendances et environnements ;

- du côté des fournisseurs de logiciels : l'effort va devoir s'accroître en amont de la chaîne pour empêcher davantage de code vulnérable d'entrer en production et ainsi limiter la pression de nettoyage en aval qui pèse sur l'industrie de la cybersécurité. De nombreux acteurs appellent sur ce point à responsabiliser davantage les éditeurs de logiciels du fait des vulnérabilités embarquées dans leurs produits, dans le prolongement du *Cyber Resilience Act européen*.

5.6. En raison des goulets d'étranglement prévisibles sur la chaîne d'approvisionnement des patchs logiciels et sur la capacité des utilisateurs finaux à adopter les cadences requises pour éliminer tout risque d'exploitation des vulnérabilités, **il faut s'attendre à un accroissement structurel du nombre de cyberattaques réussies, y compris sur des systèmes critiques**, sauf à imaginer un système dans lequel l'IA met à disposition de l'ensemble des équipes cyber, en tous points et à tout instant, les solutions correctives en même temps qu'elle révèle les failles, ce qui est illusoire. La question de savoir si l'IA confère aux attaquants un avantage asymétrique par rapport à la défense et si cet avantage est temporaire ou structurel, n'est pas complètement tranchée. Il est toutefois certain que l'IA aura pour effet premier d'élargir le vivier d'attaques et le terrain de jeu des hackers, ce qui obligera la défense à se mettre au diapason, en cherchant à lutter avec les mêmes armes.

5.7. Face à l'imminence du scénario de vague de vulnérabilités, plusieurs acteurs de l'écosystème ont d'ores et déjà engagé des démarches structurées auprès de leurs organisations.

Les principales recommandations opérationnelles qui peuvent être formulées à destination des DSI et des RSSI sont les suivantes :

- **mettre à jour la cartographie des actifs critiques et des dépendances** (supply chain métier et supply chain logicielle) ;

- **s'entraîner à simuler une vague massive de zero-days[9]** pour tester la capacité des processus de gouvernance et de *patching* à absorber un volume et une urgence hors normes, et dresser un bilan de maturité de la chaîne de gestion des vulnérabilités (sur l'ensemble de la *kill chain*, pas uniquement vulnérabilité par vulnérabilité). Cet exercice est indispensable pour préconfigurer l'organisation face aux arbitrages lourds qui devront être rendus le jour J sur l'allocation des ressources informatiques (moyens opérationnels et humains), sur la priorisation des patchs (selon les chemins d'exploitation réels) et sur les coupures de services plus longues et plus nombreuses qu'il faudra assumer, avec des impacts en cascade sur la relation avec les métiers et leurs clients et sur les finances de l'organisation (manque-à-gagner, pénalités).

- **durcir les mesures d'architecture réseau dans le but de freiner la propagation des attaques** non déjouées (réduction du *blast radius*, abaissement du *mean time to remediate* avec des temps de réaction plus courts dans les scénarios critiques) ;

- **se doter d'un plan de défense augmentée par l'IA,** à même de suivre le rythme d'un environnement en mutation rapide. Les cas d'usage où l'IA peut permettre d'augmenter la chaîne de défense cyber sont nombreux et souvent connus (triage automatique de vulnérabilités, priorisation CVE selon exposition réelle, génération et test de patchs, cartographie des dépendances logicielles/analyse SBOM, modernisation de code *legacy*, détection de comportements anormaux, assistance SOC niveau 1 et 2, investigation post-incident, simulation de crise et entraînement des équipes, sécurité des environnements industriels et embarqué...). Ils doivent être développés autant et aussi rapidement que possible en entreprise, à deux conditions : utiliser une IA européenne ou *open source* (pour ne pas dégrader l'autonomie technologique) et le faire toujours sous supervision humaine.

5.8. Ce ne sont pas seulement les grands groupes qui devront pivoter dans ce contexte nouveau, mais aussi les structures de plus petite taille (ETI, PME, TPE) qui font partie intégrante de leur environnement économique (comme partenaires, prestataires ou sous-traitants). Le changement de paradigme de la massification de la cyberrésilience, déjà à l'œuvre à travers la directive NIS II, devra aussi se déployer sur ce terrain contigu qu'est l'IA en cyber, pour éviter une trop forte hétérogénéité dans les niveaux de protection. Pour autant, les recettes utilisées pour les grandes entités ne pourront pas être transposées telles quelles aux plus petites, ce qui imposera, à l'intersection de l'IA et de la cyber, des méthodes spécifiquement adaptées à la réalité économique et opérationnelle des PME. Les grands groupes auront un rôle clef à jouer pour embarquer leurs fournisseurs critiques de rang 1 ou 2 dans le cadre d'approches de sécurisation par filière.

9 Par exemple format d'exercice, en tabletop ou immersif, de deux à trois heures, réunissant le CTO, le RSSI et les responsables de la production, simulant la découverte simultanée d'une vingtaine de vulnérabilités zero-day sur une application critique exposée à Internet.

6 - AU-DELA DE MYTHOS : L'IA, UNE HYPOTHÈQUE MAJEURE A LEVER SUR L'AGENDA DE SOUVERAINETÉ EUROPÉENNE EN CYBER

6.1. Au-delà de l'aggravation des risques cyber, qui constitue la menace de premier rang, c'est l'intégralité de l'agenda de souveraineté européenne en cybersécurité qui pourrait être fragilisé.

Les entreprises européennes sont déjà aujourd'hui dépendantes à plus de 70%[10] de solutions cyber non européennes. La domination américaine dans l'IA exacerbera cette dépendance. Les utilisateurs finaux qui privilégieront la sécurité à tout prix n'auront pas d'autre choix que de se doter des outils de détection de vulnérabilités et de réponse à incidents les plus performants sur de grands volumes d'attaques, et donc de recourir à l'IA made in America. Ceux qui ne suivront pas cette voie, au nom de l'autonomie technologique, prendront un risque opérationnel difficilement assumable : celui d'affronter l'IA « à mains nues ». La tenaille sera alors absolue, faute - à date - d'alternative européenne capable d'apporter aujourd'hui un niveau de performance comparable à Anthropic, Google, OpenAI ou Microsoft dans le domaine de la cybersécurité. Mythos rappelle salutairement à l'Europe que sa souveraineté en cyber se jouera aussi sur ses capacités domestiques en IA. A défaut, l'Europe n'aura pas de meilleure option que d'investir massivement dans l'acquisition des meilleurs modèles d'IA américains (ou chinois) pour sa défense en cybersécurité.

6.2. Le paysage international est sans ambiguïté :

- les États-Unis structurent déjà une avance industrielle, institutionnelle et capacitaire ;
- la Chine développera probablement ses propres modèles, possiblement à partir de modèles américains ou open source, mais sans garantie de publication ouverte ni de transparence. Les modèles chinois pourraient arriver rapidement, sans que nous ayons le temps de documenter et sécuriser suffisamment les usages ;
- l'Europe doit accélérer sur ses propres capacités d'évaluation, de test, d'usage défensif et de doctrine, ou disparaître de l'équation.

6.3. La seule voie pour éviter un décrochage européen définitif est une combinaison de trois éléments indissociables :

• **développer drastiquement la capacité française et européenne d'anticipation sur les enjeux de sécurité associés au développement des modèles d'IA de frontière**. Au-delà de la gestion d'un éventuel «moment de crise» post-Mythos, il est impératif de se préparer à la dynamique plus globale de progression des modèles, et en particulier à la perspective, non inéluctable mais aujourd'hui plausible, de systèmes atteignant ou surpassant des niveaux d'experts humains dans toutes les tâches cognitives à des horizons courts. Cette trajectoire n'a rien d'une fatalité car elle dépend de choix industriels, et politiques sur lesquels la France et l'Europe peuvent encore peser. Une telle bascule aurait des implications considérables non seulement en matière de cybersécurité, mais également dans l'ensemble des champs critiques pour la sécurité nationale (NRBC, absence de garantie de contrôle sur des systèmes d'IA avancés, etc.) ;

• **positionner offensivement ses entreprises les plus avancées de l'IA en B2B (au premier chef Mistral AI) en priorité sur les cas d'usage cyber, en leur mettant à disposition davantage de puissance de calcul** pour construire une IA souveraine et performante pour la cyber (rappelons à cet égard que la performance des modèles LLM est étroitement corrélée à la quantité de calcul utilisé pour l'entraînement, et que plus de 80 % des GPUs[11] mondiales sont possédées par des entreprises américaines). Il est particulièrement urgent de concrétiser la création d'un espace européen unifié, structuré, sécurisé et mutualisé de données cyber destinées à l'entraînement des nouveaux modèles européens d'IA, dans une approche d'innovation et d'industrialisation ;

10 Estimation issue de l'étude « European Software and Cyber Dependencies » réaliséréalisée en décembre 2025 pour la commission ITRE du Parlement européen.

- **appliquer la réglementation européenne (en particulier l'IA Act et le CRA – Cyberresilience Act) dans toute sa portée,** jusqu'à restreindre le cas échéant la commercialisation de modèles d'impact systémique qui ne répondraient pas aux garanties les plus les plus élevées de transparence, de sécurité, d'interprétabilité, de réversibilité et d'alignement. Il est urgent pour l'UE de se doter d'un banc d'essai européen IA-Cyber indépendant pour évaluer les capacités cyber des modèles d'IA de frontière. Ce dispositif pourrait associer l'ANSSI, l'ENISA, le JRC, l'ECCC, des laboratoires académiques, des industriels cyber et des opérateurs critiques volontaires. Il aurait pour objectif de tester les modèles sur des cas réalistes : découverte de vulnérabilités, génération de correctifs, reverse engineering, exploitation de vulnérabilités connues, sécurité OT/ICS[12], robustesse des garde-fous et capacité de confinement... ;

Le moment est venu d'aller au bout du raisonnement qui fait de la cybersécurité - et de l'IA - un élément central de la défense nationale, en alignant les actes et les discours : comme nous ne tolèrerions pas que notre chaîne de dissuasion nucléaire soit dépendante de tiers, nous n'avons pas davantage de raison d'accepter que notre bouclier cyber (et la quantité croissante d'IA qui s'y incorpore) soit externalisé. Du reste, aucune des grandes puissances (Etats-Unis et Chine en particulier) ne s'en accommoderait pour elle-même.

6.4. En corollaire, tout comme l'IA menace d'enfermer le client européen dans un arbitrage impossible entre souveraineté et sécurité, elle pourrait bien aussi menacer la viabilité économique de la filière cyber européenne, déjà obligée de composer avec la domination des environnements numériques américains. Comment sécuriser à long terme une industrie indépendante et prospère du *pentesting*, de l'audit cyber ou de la réponse à incident (par exemple), si ces métiers sont assurés demain par l'IA dans des conditions de performance égales ou supérieures ? Si les entreprises cyber européennes décident, comme le reste de l'économie, de booster leurs produits et services avec l'IA, comment pourra-t-on encore les considérer comme indépendantes, dès lors que leur valeur résidera dans l'utilisation de technologies étrangères ?

Les travaux de prospective en cours au sein du Campus sur le futur du marché de la cyber à horizon 2030 seront complétés pour prendre en compte les pleins impacts de l'IA sur i) le paysage industriel de l'offre (éditeurs de logiciels, intégrateurs) ii) les métiers de la cyber iii) le contenu des formations iv) et la place respective de l'humain et de l'automate dans la future architecture de résilience des organisations.

Mythos a le mérite d'ouvrir un espace pour préparer et accompagner les mutations stratégiques de l'offre européenne de cybersécurité, autour de *business models* pérennes, en gardant à l'esprit que les clefs de la souveraineté sont toujours entre les mains des acheteurs, publics comme privés, qui par leurs arbitrages quotidiens, façonnent l'avenir de notre industrie de la cybersécurité.

11 Graphics Processing Unit
12 Operational Technology/Industrial Control Systems.

CAMPUS CYBER

campuscyber.fr