Commission européenne : le dilemme de sécurité du pentest IA

Le 19 novembre 2026, la Commission européenne a confirmé avoir subi une campagne cyber, détectée le 22 octobre et conduite avec un outil d'intelligence artificielle accessible au public. Selon un responsable européen cité anonymement par Politico Europe, cet outil serait ARTEX, une plateforme de test d'intrusion (pentest, simulation d'attaque commandée par le propriétaire d'un système) pilotée par des modèles de langage, que CERT-EU avait signalée publiquement deux semaines plus tôt. Le même code audite et attaque, et ce qui sépare les deux usages, l'autorisation, ne figure nulle part dans le logiciel. Relu à travers le dilemme de sécurité de Robert Jervis (1978), le cas place l'institution qui écrit les règles européennes de cybersécurité dans le monde que Jervis jugeait le plus dangereux, celui où l'attaque l'emporte et où rien ne distingue une posture défensive d'une posture offensive.

Par Alexandre Berge · 22 novembre 2026 · 20 min de lecture · Chronique

Ce que la Commission confirme tient en peu de mots, ce qu'elle tait porte sur l'outil

Le point de presse de midi du 19 novembre 2026, dont une séquence est consacrée à la « campagne cyber contre la Commission européenne », fixe le périmètre officiel. Le porte-parole Thomas Regnier y indique que l'incident a été détecté le 22 octobre et confirmé le 26 octobre au terme d'une première enquête. Il décrit une mesure de précaution : « As a protective and preventive measure, we temporarily took down several applications, and most of them have since then been restored » (« À titre de mesure de protection et de prévention, nous avons temporairement mis hors service plusieurs applications, et la plupart ont depuis été rétablies »). Il qualifie l'épisode de campagne générée par l'IA et précise qu'il s'agissait d'« an AI tool that is publicly available » (« un outil d'IA accessible au public »), sans le nommer. Il déclare n'avoir détecté à ce stade aucune preuve d'une fuite de données touchant le personnel de la Commission dans son ensemble, dit n'avoir connaissance d'aucune demande de rançon et refuse de se prononcer sur l'attribution. L'institution reste, selon lui, en contact étroit avec CERT-EU et le Contrôleur européen de la protection des données.

Le reste provient de Politico Europe, qui a révélé l'affaire dans son Brussels Playbook du 19 novembre au matin. Un responsable européen, cité sous couvert d'anonymat, y indique que

CERT-EU a établi au début du mois de novembre l'usage d'ARTEX, outil d'IA public conçu pour repérer les vulnérabilités des systèmes informatiques, et que certains acteurs détournent de tels outils et ont visé la Commission. Selon le même article, la perturbation a duré plus de dix jours. Sysper, la plateforme de ressources humaines qui gère bulletins de paie, demandes de congé et relevés de temps, et MyPMO, le système de remboursement des frais médicaux, n'ont été rétablis qu'en début de semaine, Sysper connaissant encore une brève interruption la veille de la publication. Les services de traduction ont également été touchés. Le Corporate Management Board, plus haute instance interne de gestion de crise de la Commission, a été réuni en urgence, et la commissaire chargée du numérique, Henna Virkkunen, a rendu compte de la réponse immédiate devant le collège. Next, qui relaie ces éléments, précise qu'aucun lien formel n'a été établi entre l'outil et l'attaque. Interrogé sur le silence de l'institution pendant près d'un mois, le porte-parole a fait valoir que les mesures prises contre une campagne cyber restent hors de la vue du public pour demeurer efficaces, et que la procédure habituelle (information de CERT-EU, du Contrôleur européen et des services concernés) avait été suivie.

Ce partage entre l'officiel et le rapporté n'est pas une précaution de forme. Il conditionne toute la lecture qui suit. L'usage d'ARTEX contre la Commission reste à ce jour une information attribuée par la presse à un responsable anonyme, plausible au regard des alertes publiques, mais sans attribution officielle. Ce qui est établi, en revanche, c'est que l'outil existe, qu'il circule et qu'il sert contre des administrations.

ARTEX ne distingue pas l'audit de l'intrusion, seul le mandat le fait

L'avis FLASH-20261113-001 du FBI (Federal Bureau of Investigation), publié le 13 novembre 2026 en diffusion libre, décrit ARTEX comme une plateforme de test d'intrusion autonome et multi-agents écrite en Go. L'opérateur formule un objectif en langage naturel, que le système traduit en commandes exécutées par des agents spécialisés, avec une intervention humaine minimale au-delà de la fixation de l'objectif. L'outil n'embarque aucun modèle propre et se branche sur des fournisseurs externes, que l'avis du FBI énumère : des modèles de pointe (Anthropic Claude, OpenAI GPT) comme des modèles ouverts ou d'origine chinoise (DeepSeek, GLM, Grok, Kimi). Mis en ligne sur GitHub en juillet 2026 par un développeur se présentant sous le pseudonyme « Autumn-27 », il a été retiré par son auteur le 8 octobre 2026, après des signalements d'abus. Le développeur affirmait l'avoir conçu pour des tests autorisés dans un périmètre défini. Le FBI constate que le retrait n'a rien arrêté : l'outil prolifère par des copies indépendantes (forks), des versions localisées, des archives et une édition commerciale baptisée ARTEX PRO. Le contexte de cette diffusion rejoint un constat déjà posé dans ces colonnes à propos du logiciel libre européen, où le code publié échappe durablement à celui qui l'a écrit (« Open source européen : la maintenance sans les moyens », 8 juillet 2026).

> **ARTEX en chiffres, selon l'avis FLASH du FBI du 13 novembre 2026.**
>
> Usage malveillant observé depuis au moins août 2026, dans les secteurs financier, public, technologique, de la défense, des semi-conducteurs, de l'énergie, de la santé, de l'industrie manufacturière, des télécommunications et des services juridiques. Selon l'analyse de Microsoft du 21 octobre 2026 citée par le FBI, certaines intrusions passent du ciblage initial à des accès privilégiés en moins de trois heures. Une campagne contre des établissements financiers sud-coréens, documentée par CrowdStrike (7 octobre 2026) et Zscaler ThreatLabz (14 octobre 2026), a conduit à l'exfiltration de données personnelles d'environ 67 000 personnes. Une compromission de la chaîne d'approvisionnement de ConnectWise Cloud Backup, analysée par Microsoft, a donné lieu à des tentatives de connexion vers plus de 2 250 environnements clients. Au début de novembre 2026, selon Check Point Research (5 novembre 2026), plus de 1 600 consoles ARTEX étaient exposées sur Internet, le nombre total de déploiements étant jugé nettement supérieur si l'on inclut les systèmes protégés par des pare-feu ou situés dans des réseaux privés.

Le point décisif figure dans l'alerte publiée par CERT-EU le 5 novembre 2026. Le service de cybersécurité des institutions, organes et organismes de l'Union, établi par le règlement (UE, Euratom) 2023/2841, y décrit un outil conçu pour être supervisé par un opérateur humain à travers un tableau de bord doté d'une file d'approbation, et dont le respect du périmètre repose sur des règles configurables et sur le modèle lui-même. Sa conclusion tient en une ligne : « there is no hard technical enforcement » (« il n'existe aucune contrainte technique stricte »). La frontière entre le test commandé et l'intrusion n'est donc pas un attribut du logiciel. Elle relève d'un document extérieur au code, la convention d'audit, la lettre de mission, la liste des adresses et des créneaux autorisés. Le même service écrit avoir observé des activités pilotées par l'IA contre ses propres membres, et ajoute : « These activities highly likely involved the use of ARTEX » (« Ces activités impliquaient très probablement l'usage d'ARTEX »).

Le praticien d'une direction des systèmes d'information publique reconnaît la situation. Un test d'intrusion n'est légitime que par l'acte qui l'autorise, et cet acte vit dans un circuit de signature, pas dans le trafic réseau. Vus depuis un centre de supervision, un balayage commandé et un balayage hostile produisent les mêmes journaux.

Jervis a décrit ce monde, celui de l'attaque avantagée et des postures indiscernables

Robert Jervis part d'un constat simple : faute de souverain international, des États satisfaits du statu quo peinent à atteindre des objectifs qu'ils savent communs. Le dilemme de sécurité en découle, que l'article de 1978 définit ainsi : « many of the means by which a state tries to increase its security decrease the security of others » (« nombre des moyens par lesquels un État cherche à accroître sa sécurité diminuent celle des autres »). La première partie du texte examine ce qui atténue le dilemme (le coût d'être exploité, les gains de la coopération, la

perception de la menace). La seconde, la plus féconde pour le cas présent, isole deux variables : « Two crucial variables are involved: whether defensive weapons and policies can be distinguished from offensive ones, and whether the defense or the offense has the advantage » (« Deux variables cruciales sont en jeu : la possibilité de distinguer les armes et politiques défensives des offensives, et l'avantage de la défense ou de l'attaque »).

> **Définition. Le dilemme de sécurité et l'équilibre attaque-défense (Robert Jervis, 1978).**
>
> Robert Jervis, « Cooperation under the Security Dilemma », *World Politics*, vol. 30, n° 2, janvier 1978, p. 167-214. Le *dilemme de sécurité* désigne la situation où les moyens qu'un acteur emploie pour accroître sa sécurité réduisent celle des autres. Son intensité dépend de deux variables, la *distinguabilité* des postures défensives et offensives, et l'*équilibre attaque-défense* (offense-defense balance), c'est-à-dire la facilité relative de prendre ou de tenir. Leur croisement dessine quatre mondes. Attaque avantagée et postures indiscernables, le monde est « doublement dangereux ». Défense avantagée et postures indiscernables, le dilemme demeure mais les exigences de sécurité peuvent rester compatibles. Attaque avantagée et postures distinctes, le dilemme disparaît mais l'agression reste possible, et les acteurs du statu quo peuvent se distinguer des agresseurs et recevoir un avertissement. Défense avantagée et postures distinctes, le monde est « doublement stable ».

Jervis tient la seconde variable pour la plus puissante. Selon lui, une différenciation entre postures offensives et défensives « comes close to abolishing it » (« s'approche de l'abolition ») du dilemme, alors que l'avantage de la défense ne fait que l'atténuer. Il ne se fait pourtant aucune illusion sur la facilité de l'exercice. Il cite le diplomate espagnol Salvador de Madariaga, pour qui « A weapon is either offensive or defensive according to which end of it you are looking at » (« Une arme est offensive ou défensive selon le bout par lequel on la regarde »), et note sobrement que « Criminals as well as policemen can use bulletproof vests » (« Les criminels comme les policiers peuvent porter des gilets pare-balles »).

La transposition exige une précaution que l'article assume. Jervis raisonne sur des États en situation d'anarchie internationale. Le cas met aux prises une institution et des acteurs criminels ou étatiques non identifiés. Les deux variables se laissent pourtant opérationnaliser sans forcer le texte. La distinguabilité devient une question d'observation : un tiers peut-il dire, à partir de l'artefact ou du trafic, si une action relève d'un test autorisé ou d'une attaque ? Pour ARTEX, CERT-EU répond implicitement par la négative, puisque le périmètre n'est contraint par aucun mécanisme technique. L'équilibre attaque-défense devient une question de coût : combien coûte à l'attaquant la découverte et l'exploitation d'une faille, et combien coûte au défenseur sa découverte et sa correction ? La section suivante montre que ce rapport penche du côté de l'attaque. Le cas se situe donc dans le premier des quatre mondes.

Le dilemme lui-même se lit dans la trajectoire de l'outil. ARTEX a été publié, selon son auteur, pour aider des défenseurs à éprouver leurs systèmes. Chaque copie mise en circulation pour améliorer la sécurité de ses utilisateurs légitimes accroît mécaniquement la capacité offensive de tous les autres. C'est exactement la structure que Jervis décrit pour la marine britannique d'avant 1914, indispensable à la protection des routes commerciales et menaçante pour tout État doté de côtes.

L'avantage offensif tient à l'économie du tri, pas à la sophistication

Dans un billet du 21 avril 2026, CERT-EU décrivait déjà un basculement de l'économie de la découverte de vulnérabilités. Citant le rapport M-Trends 2026 de Google, le service y indique que le délai moyen d'exploitation des vulnérabilités nouvellement divulguées est estimé à moins sept jours, l'exploitation précédant en moyenne la publication du correctif, contre 63 jours en 2018. Il oppose ce rythme aux cycles de gestion des correctifs des organisations, qui s'étendent généralement de 30 à 60 jours. Surtout, il formule l'asymétrie en termes de travail, non de génie technique : « attackers need only one working exploit and face no quality-control burden, while defenders must triage every finding and remediate each one correctly » (« les attaquants n'ont besoin que d'un seul exploit fonctionnel et ne supportent aucune exigence de contrôle qualité, tandis que les défenseurs doivent trier chaque découverte et corriger chacune correctement »).

L'alerte du 5 novembre précise où ARTEX creuse cet écart. L'outil peut tester un grand nombre de mauvaises configurations de faible complexité, y compris celles qu'un testeur humain remarquerait sans avoir le temps de les vérifier, et dérouler de longues chaînes d'attaque sans la fatigue qui borne les tests manuels. Le FBI ajoute qu'il permet à des acteurs dotés d'une expertise minimale de conduire simultanément des opérations complexes contre de multiples cibles. S'appuyant sur les analyses de Zscaler ThreatLabz et de Check Point Research, il relève aussi que les instances contrôlées par des attaquants recourent majoritairement à des modèles d'origine chinoise, en particulier de la famille DeepSeek Flash, en raison de leur équilibre entre capacités et coût relativement faible et de l'« lack of observable guardrails when framed for penetration testing » (« absence de garde-fous observables lorsque la demande est présentée comme un test d'intrusion »).

En termes jervisiens, un euro investi dans l'attaque exige désormais plus d'un euro de défense pour être compensé. Le conseil de CERT-EU en tire une conséquence de gestion très concrète : conserver toutes les découvertes, y compris non graves, dans le cycle de correction, afin d'empêcher un outillage dopé à l'IA d'enchaîner des défauts mineurs en chemins d'attaque à fort impact. Le praticien mesure ce que cela signifie pour une DSI publique, habituée à prioriser les vulnérabilités critiques et à laisser vieillir les constats mineurs faute de bras. La file des « faibles » devient précisément la surface que l'attaquant automatisé exploite.

Jervis observait que, dans un monde où l'attaque domine, l'acteur menacé ne peut attendre une preuve non ambiguë avant de réagir. La Commission a agi conformément à ce schéma en coupant préventivement plusieurs applications. Le coût de cette défense n'est pas abstrait. Selon Politico, il s'est traduit pendant plus de dix jours par l'impossibilité, pour une partie du personnel, de saisir ses relevés de temps, de demander des congés ou de se faire rembourser des frais médicaux, les salaires ayant toutefois été versés normalement. CERT-EU rappelle cependant que l'équilibre n'est pas figé. Les défenseurs qui adoptent systématiquement ces outils disposent, selon lui, de l'avantage structurel de pouvoir les appliquer à l'ensemble de leur parc. Chez Jervis, la technologie et la géographie fixent l'équilibre attaque-défense. Dans un système d'information, l'organisation en est une troisième variable.

Cobalt Strike a montré que le contrôle passe par l'accès, pas par le code

Le précédent le plus proche n'est pas une fuite d'arsenal étatique mais un outil commercial de simulation d'adversaire. Cobalt Strike, édité par Fortra, est décrit par Microsoft comme un outil légitime et répandu de post-exploitation destiné à la simulation d'attaques. Des copies anciennes et piratées (« cracked ») en ont été détournées par des criminels. Microsoft les relie à des familles de rançongiciels impliquées dans plus de 68 attaques contre des organisations de santé dans plus de 19 pays, ainsi qu'aux attaques contre le gouvernement du Costa Rica et l'Irish Health Service Executive. Le 31 mars 2023, le tribunal fédéral du district Est de New York a autorisé Microsoft, Fortra et le Health-ISAC (centre d'échange d'informations du secteur de la santé) à démanteler l'infrastructure de ces copies illégales.

La proximité structurelle est maximale. Un outil conçu pour des défenseurs, juridiquement licite, devient une arme dès qu'il sort du périmètre de son autorisation, et la frontière entre les deux usages est extérieure au code. Le dispositif qui portait cette frontière est explicitement décrit par Microsoft : « Legitimate security practitioners who purchase Cobalt Strike licenses are vetted by Fortra and are required to comply with usage restrictions and export controls » (« Les professionnels de la sécurité légitimes qui achètent des licences Cobalt Strike sont contrôlés par Fortra et doivent respecter des restrictions d'usage et des contrôles à l'exportation »). Autrement dit, la licence jouait le rôle de signal distinctif au sens de Jervis. Elle séparait l'usage autorisé de l'usage hostile, et c'est sur ce terrain (droit d'auteur, ordonnance judiciaire, saisie de domaines) que la riposte a pu s'organiser. Microsoft annonçait d'ailleurs que les criminels tenteraient de relancer leurs efforts, et présentait l'action comme durable plutôt que ponctuelle.

ARTEX retire ce point d'appui. Il n'existe ni licencié vérifié ni éditeur en mesure de contrôler ses clients, seulement un dépôt public retiré et des copies indépendantes, auxquelles s'ajoute une édition commerciale. Le régime européen des biens à double usage ne comble pas ce vide par construction : le règlement (UE) 2021/821 inscrit les « intrusion software » à son annexe I (rubriques 4A005 et 4D004), dans une liste dérivée de celle de l'arrangement de Wassenaar, mais il organise des autorisations d'exportation et de transfert portées par un

exportateur identifiable, figure introuvable dans un écosystème de copies anonymes.

La doctrine européenne d'accès structuré vise les modèles, pas la couche d'orchestration

Le 7 juillet 2026, la Commission a présenté un plan d'action sur la cybersécurité et l'intelligence artificielle. La page de présentation reconnaît que l'IA, si elle peut améliorer la sécurité, « can also be misused to identify vulnerabilities, automate attacks, and significantly increase the scale and speed of cyber incidents » (« peut aussi être détournée pour identifier des vulnérabilités, automatiser des attaques et accroître significativement l'ampleur et la vitesse des incidents cyber »). Les actions annoncées portent sur l'évaluation des modèles avancés au titre de l'AI Act, la création d'une capacité européenne d'évaluation, un « blueprint » européen d'accès structuré aux capacités d'IA avancées pour la cybersécurité, élaboré avec l'ENISA (Agence de l'Union européenne pour la cybersécurité), et une plateforme sécurisée de test conduite avec le Centre commun de recherche. CERT-EU, dans son billet d'avril, saluait dans le même esprit les programmes d'accès contrôlé des laboratoires d'IA, en citant le Project Glasswing d'Anthropic, qui réservait le modèle Claude Mythos Preview à douze partenaires de lancement et à plus de 40 organisations supplémentaires, et le programme Trusted Access for Cyber d'OpenAI pour GPT-5.4-Cyber, et résumait leur objectif commun : faire parvenir les capacités de pointe aux défenseurs sans les rendre librement disponibles aux acteurs malveillants.

L'angle mort tient dans l'architecture même d'ARTEX. Le FBI souligne que l'outil n'inclut aucun modèle propre et s'adapte à n'importe quel fournisseur, et que ses opérateurs hostiles privilégient des modèles bon marché dépourvus de garde-fous observables. Une doctrine qui contrôle l'accès aux modèles les plus puissants agit sur une couche que l'attaquant peut contourner en changeant de fournisseur. Elle laisse intacte la couche d'orchestration, celle qui décompose un objectif, distribue les tâches entre agents et exécute de vrais outils, et c'est précisément cette couche qui ignore la notion d'autorisation. Dans la page de présentation du plan publiée par la Commission le 7 juillet 2026, consultée le 21 novembre 2026, aucune action ne vise explicitement les cadres d'orchestration agentique ni la traçabilité du périmètre autorisé d'un test. CERT-EU lui-même indiquait en avril avoir construit son propre pipeline interne de tests d'intrusion par agents à partir de modèles publiquement disponibles, en observant que des acteurs étatiques hostiles, des groupes criminels et des éditeurs moins scrupuleux pouvaient bâtir l'équivalent.

Un second non-dit concerne la publicité de l'incident. Le règlement (UE, Euratom) 2023/2841 impose aux entités de l'Union, en cas d'incident important, une alerte précoce à CERT-EU dans les 24 heures, une notification dans les 72 heures et un rapport final au plus tard un mois après la notification. Ces délais organisent l'information de CERT-EU, non celle du public ni des agents. Selon Politico, la réponse du service informatique à un agent qui signalait un

dysfonctionnement de Sysper évoquait un problème général affectant le système, sans mentionner de cyberattaque, et plusieurs agents déploraient le manque de communication. Le responsable cité par le titre justifie cette retenue par l'incertitude persistante sur une éventuelle compromission des données du personnel. L'argument opérationnel est recevable. Il laisse néanmoins entière une tension que les chroniques précédentes ont documentée pour les États membres (« NIS2 : la France condamnée à l'astreinte pendant que le référentiel cyber s'impose aux collectivités », 3 octobre 2026), celle d'une institution qui prescrit à d'autres des obligations de transparence dont sa propre pratique fait un usage minimal.

Rendre l'autorisation observable, seule voie de sortie du monde doublement dangereux

Le *dilemme de sécurité* éclaire ce que le cas a de neuf. L'outil qui aurait servi contre la Commission n'est ni plus offensif ni plus défensif qu'un autre. Il est indiscernable, et l'*équilibre attaque-défense* penche aujourd'hui du côté de qui l'emploie sans mandat. Jervis indiquait la voie de sortie : non pas supprimer les armes, mais rendre les postures distinguables, et à défaut instaurer des dispositifs d'inspection, dont il écrivait qu'ils peuvent atténuer le dilemme lorsque chaque partie est prête à coopérer. Transposé à l'outillage offensif piloté par IA, cela désigne moins le contrôle des modèles que l'inscription de l'autorisation dans des traces vérifiables par un tiers : registres de tests déclarés, identités signées des agents, périmètres attestés et lisibles par les centres de supervision.

L'hypothèse qui en découle est vérifiable sans information interne. Si la lecture jervisienne est juste, la réponse institutionnelle européenne devra tôt ou tard déplacer son centre de gravité de l'accès aux modèles vers la distinguabilité des usages. Deux indicateurs permettent de le constater d'ici le 31 décembre 2027, horizon auquel la Commission situe la mise en service de la capacité européenne d'évaluation annoncée dans son plan de juillet 2026. Le premier est la présence ou l'absence, dans le « blueprint » d'accès structuré élaboré avec l'ENISA, d'une disposition visant les cadres d'orchestration agentique et l'attestation du périmètre autorisé des tests. Le second est la mention ou non, dans le prochain rapport annuel de CERT-EU sur l'état de la menace, des cadres agentiques de type ARTEX parmi les vecteurs d'accès initial observés contre les entités de l'Union. Leur absence conjointe indiquerait que l'institution qui écrit les règles européennes de cybersécurité continue de traiter l'IA offensive comme un problème de modèles, alors que son propre cas d'école en fait un problème d'autorisation.

Note méthodologique

Sources principales : point de presse de la Commission du 19 novembre 2026, Brussels Playbook de Politico Europe du même jour, avis FLASH-20261113-001 du FBI, publications de CERT-EU des 21 avril et 5 novembre 2026, texte de Jervis (1978). La Commission date du 22

octobre la détection de l'incident, Politico situe au début du mois de novembre l'identification d'ARTEX par CERT-EU : les deux dates sont retenues comme deux étapes distinctes de l'enquête. L'alerte de CERT-EU est datée du 5 novembre 2026 sur sa page d'origine, Next la date du 6 novembre : la date de la page a été retenue. L'usage d'ARTEX contre la Commission et les perturbations de Sysper et MyPMO reposent sur Politico, qui les attribue à un responsable européen anonyme, et ne sont pas confirmés par la Commission. Ils sont donc présentés comme rapportés. L'affirmation d'absence porte sur la page de présentation du plan d'action du 7 juillet 2026, consultée le 21 novembre 2026. La transposition du dilemme de sécurité, conçu pour des États, à une institution face à des attaquants non étatiques relève d'un choix d'interprétation de l'auteur. Recherche de sources, croisement des données et structuration du raisonnement s'appuient sur une assistance par intelligence artificielle. Le choix de l'angle et du cadre théorique, la vérification des sources, les arbitrages et l'établissement du texte final relèvent de l'auteur.

Sources

Sources primaires institutionnelles

- Commission européenne, Service audiovisuel, « EC Midday press briefing of 19/11/2026 » (séquence « CYBERSECURITY - Cyber campaign against the European Commission: Q&A »), 19 novembre 2026, https://audiovisual.ec.europa.eu/en/media/video/I-298237 et https://audiovisual.ec.europa.eu/en/media/video/I-298222
- Federal Bureau of Investigation, FLASH-20261113-001, « Threat Overview of ARTEX Autonomous Penetration Testing Framework », 13 novembre 2026, https://www.ic3.gov/CSA/2026/261113.pdf
- CERT-EU, « From Penetration Testing to Cyber Threat: Beware of ARTEX », 5 novembre 2026, https://cert.europa.eu/blog/from-penetration-testing-to-cyber-threat-beware-of-artex
- CERT-EU, « AI is changing the economics of vulnerability discovery. Defenders should adapt now », 21 avril 2026, https://cert.europa.eu/blog/ai-vulnerability-discovery-defenders-must-adapt
- Commission européenne, Direction générale de la communication, « New EU plan to address the risks and opportunities of advanced AI for cybersecurity », 7 juillet 2026, https://commission.europa.eu/news-and-media/news/new-eu-plan-address-risks-and-opportunities-advanced-ai-cybersecurity-2026-07-07_en
- Règlement (UE, Euratom) 2023/2841 du Parlement européen et du Conseil du 13 décembre 2023 établissant des mesures destinées à assurer un niveau élevé commun de cybersécurité dans les institutions, organes et organismes de l'Union, EUR-Lex, consulté le 21 novembre 2026, https://eur-lex.europa.eu/eli/reg/2023/2841/oj

Rapports officiels

- Parlement européen, Direction générale des politiques externes (R. Du Bois et A. Tapia Reyes), « Dual-use and cyber-surveillance: EU policies and current practices », PE 754.439, novembre 2023, consulté le 21 novembre 2026, https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/754439/EXPO_BRI(2023)754439_EN.pdf

Cadre théorique

- Robert Jervis, « Cooperation under the Security Dilemma », *World Politics*, vol. 30, n° 2, janvier 1978, p. 167-214, consulté le 21 novembre 2026, http://www.jstor.org/stable/2009958

Parallèle historique

- Microsoft (Amy Hogan-Burney), « Stopping cybercriminals from abusing security tools », Microsoft On the Issues, 6 avril 2023, https://blogs.microsoft.com/on-the-issues/2023/04/06/stopping-cybercriminals-from-abusing-security-tools/

Presse

- Politico Europe (Gerardo Fortuna, avec Max Griera), « Cyberattack hits the Commission », Brussels Playbook, 19 novembre 2026, https://www.politico.eu/newsletter/brussels-playbook/cyberattack-hits-the-commission/
- Next (Alexandre Laurent), « Cyberattaque à la Commission européenne : une plateforme agentique de pentest mise en cause », 19 novembre 2026, https://next.ink/268532/cyberattaque-a-la-commission-europeenne-une-plateforme-agentique-de-pentest-mise-en-cause/
- Help Net Security (Sinisa Markovic), « European Commission confirms AI-generated cyberattack disrupted its systems », 20 novembre 2026, https://www.helpnetsecurity.com/2026/11/20/european-commission-ai-generated-cyberattack-publicly-available-tool/