



Kedros
Cybersécurité

— LIVRE BLANC

L'Intelligence Artificielle et les métiers de la Cybersécurité

Impacts, transformations, disparitions et émergences

Sommaire

Avant-propos	3
Résumé exécutif	4
Introduction générale	5
1.1 - La cybersécurité à l'ère de l'IA : un tournant historique	5
1.2 - Périmètre du livre blanc	5
1.3 - Méthodologie	5
Partie I - État des lieux : l'IA s'invite dans la cybersécurité	6
2.1 - Panorama de l'adoption de l'IA en entreprise	7
2.2 - L'IA déjà présente dans les outils de cybersécurité	7
2.3 - L'IA comme nouvelle surface d'attaque	7
2.4 - Cartographie des métiers actuels	7
Partie II - Analyse par métier : apports et pertes liés à l'IA	8
3.1 - Métiers techniques	9
3.2 - Métiers de gouvernance et de conseil	18
3.3 - Les nouveaux métiers qui émergent	26
Tableau de synthèse - Vue d'ensemble des impacts par métier	27
Partie III - Points de vigilance, limites et risques	28
4.1 - Les limites intrinsèques de l'IA en cybersécurité	29
4.1.1 - Le problème des faux positifs et faux négatifs	29
4.1.2 - L'opacité des modèles (Explainability Gap)	29
4.1.3 - La dépendance à la qualité des données	29
4.1.4 - La surface d'attaque des systèmes IA eux-mêmes	29
4.1.5 - Le biais de perception des gains de productivité	30
4.2 - Les risques pour les professionnels et les organisations	30
4.2.1 - La sur-délégation et la perte de compétences humaines	30
4.2.2 - Les risques de biais et de discrimination algorithmique	30
4.2.3 - La concentration des risques chez les fournisseurs IA	31
4.2.4 - Le risque réglementaire et de conformité	31
4.2.5 - Le risque social et humain	32
4.2.6 - Le Shadow AI : l'angle mort du quotidien	32
4.3 - Les angles morts collectifs	32
Partie IV - Recommandations stratégiques	33
5.1 - Pour les organisations	34
5.2 - Pour les professionnels de la cybersécurité	34
5.3 - Pour les décideurs et instances de gouvernance	34
Conclusion	35
Annexes	36

AVANT-PROPOS

Avant-propos

La transformation numérique des organisations s'accélère, et avec elle, les menaces qui pèsent sur leur sécurité. L'intelligence artificielle s'invite désormais dans tous les aspects de la cybersécurité : dans les outils de défense comme dans les arsenaux offensifs, dans les processus de gouvernance comme dans les pratiques de conseil.

Ce livre blanc a pour ambition d'offrir une analyse structurée et pragmatique des impacts de l'IA sur l'ensemble des métiers de la cybersécurité, qu'ils soient techniques ou orientés gouvernance. Il s'adresse aux RSSI, DPO, consultants, Risk Managers, auditeurs, juristes spécialisés, ainsi qu'aux directions générales souhaitant anticiper les mutations en cours.

Chez Kedros Cybersécurité, nous accompagnons au quotidien des organisations dans leur démarche de sécurisation : conseil, audit, maîtrise des risques, formation. Cette proximité avec le terrain nous place aux premières loges pour observer les transformations provoquées par l'IA, et pour anticiper celles qui arrivent.

Il ne s'agit ni d'un ouvrage alarmiste annonçant la fin des professionnels de la sécurité, ni d'un manifeste faussement optimiste sur les promesses de l'IA. Il s'agit d'outiller les décideurs et les praticiens pour qu'ils puissent anticiper, s'adapter et valoriser leur expertise dans un environnement en mutation profonde.

RÉSUMÉ EXÉCUTIF

Résumé exécutif

L'intelligence artificielle représente le changement le plus structurant que les métiers de la cybersécurité aient connu depuis l'avènement du cloud computing. Elle agit simultanément sur trois dimensions : elle renforce les capacités des défenseurs, elle amplifie les capacités des attaquants et elle redéfinit la valeur de chaque profil professionnel.

Cinq enseignements majeurs ressortent de cette analyse :

- 1 L'IA automatise massivement les tâches répétitives (triage d'alertes, collecte de preuves, documentation), réduisant les besoins en profils juniors dans quasi tous les métiers cyber.
- 2 Les métiers à haute valeur stratégique (jugement, leadership, gestion de crise) sont préservés voire renforcés. L'IA assiste, elle ne décide pas.
- 3 De nouveaux métiers émergent : AI Security Engineer, Auditeur de systèmes IA, AI Risk Officer, MLSecOps Engineer.
- 4 Les risques de sur-délégation, de perte de compétences humaines et de dépendance aux fournisseurs IA sont les angles morts les plus préoccupants.
- 5 La réglementation (AI Act, NIS2, DORA) valorise les professionnels capables de naviguer entre droit, technique et gouvernance.

Introduction générale

1.1 La cybersécurité à l'ère de l'IA : un tournant historique

Depuis les premiers antivirus jusqu'aux architectures Zero Trust, la cybersécurité n'a cessé d'évoluer en réponse à la sophistication croissante des menaces. L'irruption de l'intelligence artificielle marque cependant une rupture d'une nature différente : pour la première fois, la technologie ne se contente pas d'accélérer les processus existants, elle redéfinit la nature même du travail des professionnels de la sécurité.

L'IA agit simultanément des deux côtés du spectre. Elle renforce les capacités des défenseurs en permettant l'analyse de volumes de données hors de portée humaine et l'automatisation des réponses aux incidents courants. Mais elle amplifie également les capacités des attaquants : génération de phishing ultra-ciblé, création automatisée de malwares polymorphes, exploitation accélérée des vulnérabilités.

1.2 Périmètre du livre blanc

Ce livre blanc couvre deux grands blocs de métiers :

- **Métiers techniques** : Analyste SOC (N1, N2, N3), Pentesteur, Threat Hunter, Architecte Sécurité, Développeur AppSec, Forensic Investigator, Administrateur Sécurité.
- **Métiers de gouvernance et conseil** : RSSI, DPO, Risk Manager, Auditeur, Consultant GRC, Juriste spécialisé, vCISO, Formateur / Responsable sensibilisation.

Par IA, ce document englobe : Machine Learning, Deep Learning, LLM, IA générative et automatisation intelligente capable d'apprendre de données.

1.3 Méthodologie

L'analyse s'appuie sur : rapports sectoriels (ENISA, ANSSI, ISC²), publications académiques, benchmarks d'adoption de l'IA, référentiels métier (MITRE ATT&CK, OWASP LLM Top 10, NIST AI RMF) et retours d'expérience de praticiens. Les estimations de réduction d'effectifs et d'horizons temporels sont des ordres de grandeur indicatifs à visée stratégique. Ces fourchettes sont construites par croisement de trois facteurs : la décomposition des tâches de chaque métier, la part de ces tâches déjà couverte par les outils IA disponibles, et un rythme d'adoption différencié selon la taille et la maturité des organisations ; elles sont ensuite confrontées aux données sectorielles publiées (ISC², ENISA, ANSSI).

Sur le rythme d'automatisation, ce document s'appuie également sur les mesures publiées par METR (Model Evaluation & Threat Research), organisme de recherche indépendant spécialisé dans l'évaluation des capacités des modèles d'IA de pointe : la durée des tâches qu'un agent IA peut accomplir de façon autonome double environ tous les sept mois depuis 2019, avec une nette accélération depuis 2024¹. Les horizons temporels avancés dans ce livre blanc doivent donc se lire comme des ordres de grandeur, susceptibles d'être raccourcis si cette tendance se poursuit.

Un mot de transparence, enfin : Kedros est partie prenante de cette analyse, puisque le conseil, l'audit et la formation figurent parmi les métiers étudiés. Nous avons traité ce conflit d'intérêts en appliquant à nos propres activités la même grille que pour les autres ; le lecteur constatera qu'elles comptent parmi les plus exposées.

1. METR, Measuring AI Ability to Complete Long Tasks, mars 2025 (arxiv.org/abs/2503.14499) ; mise à jour Time Horizon 1.1, janvier 2026 (metr.org).



PARTIE I

État des lieux : l'IA s'invite dans la cybersécurité

-
- 2.1 Panorama de l'adoption de l'IA en entreprise
 - 2.2 L'IA déjà présente dans les outils de cybersécurité
 - 2.3 L'IA comme nouvelle surface d'attaque
 - 2.4 Cartographie des métiers actuels
-

2.1 Panorama de l'adoption de l'IA en entreprise

L'adoption de l'IA connaît une accélération sans précédent depuis 2022. Les outils d'IA générative ont franchi le seuil d'accessibilité qui avait longtemps cantonné l'IA aux grandes organisations technologiques. Aujourd'hui, PME comme multinationales déploient des assistants IA et des modèles de détection comportementale, créant un fossé croissant entre organisations avancées et retardataires.

La filière elle-même reste sous tension : la main-d'œuvre mondiale en cybersécurité plafonne autour de 5,5 millions de professionnels, pour un déficit estimé à 4,8 millions². L'édition 2025 de la même étude³ documente une bascule significative : les organisations déclarent désormais manquer davantage de compétences critiques, l'IA en tête, que d'effectifs (l'étude ne publie d'ailleurs plus d'estimation chiffrée du déficit mondial), et 70 % des professionnels interrogés cherchent à acquérir des qualifications en IA.

2.2 L'IA déjà présente dans les outils de cybersécurité

- SIEM et SOAR augmentés par le ML pour le triage et la corrélation d'alertes
- EDR / XDR avec détection comportementale basée sur des modèles IA
- Plateformes de Threat Intelligence automatisée
- Outils de pentest assistés par IA (BAS - Breach and Attack Simulation)
- Solutions GRC et conformité intelligentes avec monitoring continu
- Copilotes de sécurité (type Microsoft Security Copilot) pour les équipes SOC
- Agents IA autonomes d'investigation : au-delà des copilotes, les éditeurs déploient depuis 2025 des agents capables de mener triage, enrichissement et corrélation de bout en bout, les actions de remédiation restant soumises à validation humaine

2.3 L'IA comme nouvelle surface d'attaque

- Phishing ultra-ciblé généré par IA : campagnes de spear-phishing personnalisées à grande échelle
- Deepfakes audio et vidéo : usurpation d'identité de dirigeants pour valider des virements frauduleux
- Génération automatisée de malwares polymorphes contournant les signatures connues
- Fuzzing automatisé accéléré : découverte de vulnérabilités zero-day à vitesse inédite
- Attaques sur les systèmes IA eux-mêmes : adversarial examples, data poisoning, prompt injection
- Industrialisation du phishing : plus de 80 % des e-mails de phishing identifiés entre septembre 2024 et février 2025 utilisaient l'IA à un degré ou un autre, selon des données de source tierce relayées par l'ENISA⁴

2.4 Cartographie des métiers actuels

Le champ de la cybersécurité recouvre aujourd'hui plusieurs dizaines de métiers distincts (le panorama des métiers de l'ANSSI en recense environ vingt-cinq, le référentiel européen ECSF de l'ENISA en structure 12 profils types), organisés en deux pôles complémentaires. Cette structuration sert de fil directeur pour l'analyse des impacts de l'IA en Partie II.

2. ISC², Cybersecurity Workforce Study 2024. 3. ISC², Cybersecurity Workforce Study 2025.

4. ENISA, Threat Landscape 2025, octobre 2025.



PARTIE II

Analyse par métier : apports et pertes liés à l'IA

Pour chaque métier, cette partie analyse ce que l'IA apporte et ce qu'elle fait perdre. Chaque fiche se conclut par un verdict d'impact global et un horizon temporel estimé.

Le verdict de chaque fiche est mis en couleur : rouge pour un impact très fort (métier fortement restructuré), orange pour un impact fort (métier transformé), bleu pour un impact modéré, vert pour un impact faible (métier préservé, voire renforcé).

3.1 Métiers techniques

3.2 Métiers de gouvernance et de conseil

3.3 Les nouveaux métiers qui émergent

3.1 - MÉTIERS TECHNIQUES

TECHNIQUE · 3.1

Analyste SOC Niveau 1

● Impact très fort

CE QUE L'IA APPORTE

- **Triage automatique des alertes** : les outils SIEM/SOAR augmentés par ML classifient et priorisent les alertes en temps réel, là où un analyste N1 passait 70 à 80 % de son temps à les trier manuellement
- **Réduction de l'alert fatigue** : les modèles de détection comportementale filtrent les faux positifs, rendant les alertes remontées plus pertinentes et actionnables
- **Documentation automatisée** : génération automatique de tickets, de chronologies d'incidents et de premiers éléments de rapport
- **Assistance contextuelle** : les copilotes IA fournissent instantanément le contexte d'une alerte, les IOC associés et les premières recommandations

CE QUE L'IA FAIT PERDRE

- **Quasi-disparition du triage manuel** : c'est la tâche constitutive du N1. Si l'IA prend en charge 80 à 90 % du triage, le volume de postes N1 s'effondre mécaniquement
- **Réduction drastique des effectifs** : réduction estimée de 40 à 60 % des postes N1 d'ici 5 ans dans les SOC industrialisés
- **Perte de la valeur d'apprentissage** : le N1 était le premier échelon de formation des analystes. Sa disparition crée un manque dans le parcours de montée en compétences
- **Externalisation accélérée** : les SOC managés (MSSP), premiers équipés en IA, deviennent plus compétitifs : les postes N1 internes disparaissent avant les postes externalisés
- **Accélération par l'IA agentique** : le passage des copilotes (assistance) aux agents autonomes (délégation de bout en bout) accélère la réduction des postes de triage et raccourcit d'autant les horizons initialement anticipés

CE QUE L'IA NE PEUT PAS REMPLACER

- Supervision des flux automatisés et traitement des cas non classifiables par l'IA
- Interface humaine en cas de défaillance des outils d'automatisation

VERDICT

Métier en voie de disparition partielle. Le N1 triage pur n'existera plus dans sa forme actuelle. Un volume résiduel subsistera pour superviser les flux automatisés.

HORIZON TEMPOREL

3 à 5 ans pour les grandes organisations, 5 à 8 ans pour les PME et ETI.

TECHNIQUE · 3.1

Analyste SOC Niveau 2

● Impact fort

CE QUE L'IA APPORTE

- **Investigation accélérée** : corrélation automatique d'événements sur des périmètres étendus que l'analyste N2 mettait des heures à rassembler manuellement
- **Enrichissement automatique des IOC** : les plateformes de Threat Intelligence IA croisent en temps réel les indicateurs avec des bases mondiales
- **Simulation de scénarios d'attaque** : l'IA propose des hypothèses d'attaque probables à partir d'une séquence d'événements
- **Réduction du MTTR** : les premières actions de containment peuvent être automatisées et soumises à validation humaine

CE QUE L'IA FAIT PERDRE

- **Réduction des missions de corrélation manuelle** : relier des événements épars pour reconstituer le scénario d'attaque est historiquement le cœur intellectuel du métier, ce qui distingue le N2 du simple triage ; c'est précisément cette part valorisante du travail que les outils ML absorbent en partie
- **Diminution du volume de postes** : réduction estimée de 20 à 35 %
- **Banalisation de certaines compétences** : la corrélation et l'enrichissement d'alertes deviennent des fonctions standard des plateformes

CE QUE L'IA NE PEUT PAS REMPLACER

- **Superviseur de modèles** : valider, affiner et contester les décisions de l'IA
- **Chasseur d'anomalies complexes** : traiter ce que l'IA ne sait pas classifier

VERDICT

Métier profondément transformé. Le N2 évolue vers un profil hybride analyste-superviseur IA. Les volumes diminuent modérément, la valeur individuelle augmente.

HORIZON TEMPOREL

Transformation en cours, accélération sur 2 à 4 ans.

TECHNIQUE · 3.1

Analyste SOC Niveau 3 / Incident Responder

● Impact faible

CE QUE L'IA APORTE

- **Forensic assisté** : timelines d'intrusion reconstituées par IA, réduisant l'investigation de plusieurs jours à quelques heures
- **Playbooks dynamiques** : l'IA génère des plans de réponse adaptés à l'incident en cours, sur la base de cas similaires passés
- **Détection APT** : les modèles de détection comportementale détectent des TTP que l'analyse manuelle aurait pu manquer
- **Corrélation multi-sources massive** : l'IA traite des millions d'événements pour retrouver le patient zéro ou la chaîne d'infection

CE QUE L'IA FAIT PERDRE

- **Réduction des missions de triage post-incident** : le travail amont est déjà réalisé par l'IA avant que le N3 intervienne, ce qui réduit d'autant son périmètre d'intervention et le volume de missions qui lui sont confiées
- **Moindre différenciation sur les incidents classiques** : ransomware standard et phishing documenté de plus en plus gérés automatiquement
- **Risque de deskilling** : si l'IA reconstruit la timeline, l'expert perd progressivement la capacité de le faire manuellement. Or c'est précisément lorsque l'outil échoue, face à une attaque inédite ou dans un environnement dégradé, que ces réflexes redeviennent vitaux : l'érosion silencieuse des fondamentaux se paie au pire moment (ce risque est développé en 4.2.1)

CE QUE L'IA NE PEUT PAS REMPLACER

- **Gestion de crise humaine** : coordination, communication, décision sous pression
- **Investigation sur les attaques inédites** : l'IA échoue face au véritablement inédit : zero-day et APT non documentés

VERDICT

Métier globalement préservé, valeur renforcée. Volume de postes stable à légèrement en hausse, profils plus rares et mieux rémunérés.

HORIZON TEMPOREL

Évolution progressive sur 5 à 10 ans, pas de rupture brutale.

TECHNIQUE · 3.1

Pentesteur / Ethical Hacker

● Impact fort

CE QUE L'IA APORTE

- **Automatisation des phases de reconnaissance** : OSINT, cartographie de surface d'attaque (tâches réduites de 60 à 70 % en temps)
- **Génération automatique de rapports** : les LLM produisent les sections standardisées en quelques minutes
- **Suggestion d'exploits** : les outils IA identifient les CVE applicables et proposent les vecteurs d'exploitation les plus probables
- **Couverture élargie** : un pentesteur assisté par IA peut couvrir un périmètre significativement plus grand dans le même délai
- **Tests de régression automatisés** : vérification continue que les correctifs ont bien été appliqués

CE QUE L'IA FAIT PERDRE

- **Réduction des missions de pentest basique** : absorbé progressivement par les scanners IA autonomes (BAS)
- **Pression sur les prix** : l'automatisation réduit le temps facturable sur les phases mécaniques
- **Moindre valeur des certifications procédurières** : perdent en différenciation si l'IA exécute ces méthodologies
- **Réduction de 20 à 30 % des postes juniors** : un pentesteur IA-augmenté traite 2 à 3 fois plus de missions

CE QUE L'IA NE PEUT PAS REMPLACER

- La créativité offensive et le raisonnement latéral sur des environnements non documentés
- Le pentest physique et l'ingénierie sociale
- L'explication pédagogique aux équipes et aux COMEX
- Le Red Teaming stratégique simulant des adversaires étatiques

VERDICT

Métier transformé en profondeur. Le junior exécutant de méthodologies est menacé. Le senior créatif et communicant est valorisé. Bifurcation nette du marché.

HORIZON TEMPOREL

Déjà en cours sur le bas du marché. Accélération sur 3 à 5 ans.

TECHNIQUE · 3.1

Threat Hunter

● Impact modéré

CE QUE L'IA APPORTE

- **Détection proactive d'anomalies comportementales** : modèles UEBA sur des volumes impossibles à analyser manuellement
- **Hypothèses de chasse générées automatiquement** par les LLM à partir des derniers rapports de Threat Intelligence
- **Recherche dans des jeux de données massifs** : requêtes de chasse exécutées sur des téraoctets de logs en quelques secondes
- **Détection de TTP émergentes** : modèles entraînés sur MITRE ATT&CK identifiant des patterns même partiellement masqués

CE QUE L'IA FAIT PERDRE

- **Réduction des missions de chasse réactive** : anomalies évidentes progressivement absorbées par les outils ML intégrés
- **Perte de valeur des compétences de requêtage** : requêtes SIEM et règles de détection YARA désormais générées par l'IA en langage naturel
- **Risque de dépendance aux hypothèses IA** : perte de la capacité à penser indépendamment comme l'attaquant

CE QUE L'IA NE PEUT PAS REMPLACER

- **Conception de nouveaux modèles de détection** : le Threat Hunter devient architecte de la détection augmentée
- **Veille sur les TTP inédites** : ce que l'IA n'a pas encore appris

VERDICT

Métier transformé, pas menacé. Évolution de chasseur manuel vers architecte de la détection augmentée. Volume stable, profil requis plus technique.

HORIZON TEMPOREL

Transition sur 3 à 6 ans.

TECHNIQUE · 3.1

Architecte Sécurité

● Impact modéré

CE QUE L'IA APPORTE

- **Génération de schémas d'architecture de référence** : Zero Trust, SASE ou cloud-native adaptés en langage naturel
- **Analyse des risques architecturaux** : modélisation automatisée des menaces (STRIDE assisté par IA)
- **Vérification de conformité by design** : croisement automatique avec NIS2, DORA, HDS
- **Documentation automatisée** : génération de dossiers d'architecture, de DAT, de matrices RACI sécurité

CE QUE L'IA FAIT PERDRE

- **Réduction des missions de documentation** : livrables standardisés produits bien plus rapidement, temps facturable réduit
- **Banalisation du conseil sur les architectures connues** : un junior avec un outil IA peut produire une architecture de référence correcte
- **Pression sur les délais et honoraires** : les clients attendent des livrables plus rapides à coût moindre

CE QUE L'IA NE PEUT PAS REMPLACER

- Arbitrer les compromis stratégiques : sécurité vs performance, coût vs résilience, centralisation vs souveraineté
- Accompagner le changement organisationnel lié à une refonte architecturale
- Comprendre les contraintes politiques internes d'une organisation

VERDICT

Métier préservé mais sous pression tarifaire. La valeur se déplace vers la posture stratégique. Les exécutants d'architectures standards sont fragilisés.

HORIZON TEMPOREL

Évolution sur 4 à 7 ans.

TECHNIQUE · 3.1

Développeur SecDevOps / Ingénieur AppSec

● Impact fort

CE QUE L'IA APORTE

- **Revue de code automatisée** : outils SAST augmentés par IA (GitHub Copilot, Semgrep, Snyk) détectant les vulnérabilités en temps réel
- **Génération de code sécurisé** : LLM proposant des implémentations tenant compte des bonnes pratiques
- **Fuzzing intelligent** : couverture des surfaces d'attaque bien supérieure aux approches manuelles
- **Tests de sécurité intégrés CI/CD** : contrôles automatiques à chaque commit, sans intervention humaine

CE QUE L'IA FAIT PERDRE

- **Réduction des missions de revue de code manuelle** : patterns connus absorbés par les outils automatiques
- **Banalisation des tests standards** : les vulnérabilités du Top 10 OWASP sont détectées automatiquement, la valeur se reporte sur les vulnérabilités logiques
- **Réduction des profils AppSec juniors** : les développeurs eux-mêmes, assistés par IA, sécurisent mieux leur code, avec une réduction estimée de 20 à 30 % des postes

CE QUE L'IA NE PEUT PAS REMPLACER

- **Vulnérabilités de logique métier** : non détectables par analyse automatique de pattern
- Architecture sécurité applicative dans des contextes complexes
- Conseil aux équipes de développement sur les choix de conception
- **La revue du code généré par IA** : le volume de code produit par les assistants crée une nouvelle dette de sécurité (vulnérabilités subtiles, dépendances fictives) qui élargit le périmètre AppSec

VERDICT

Métier transformé. Spécialisation sur la logique métier et les vulnérabilités complexes. Forte pression sur les profils juniors.

HORIZON TEMPOREL

Déjà en cours. Accélération sur 2 à 4 ans.

TECHNIQUE · 3.1

Forensic Investigator / Analyste Inforensique

● Impact modéré

CE QUE L'IA APPORTE

- **Reconstruction automatique de timelines** : corrélation d'artefacts hétérogènes (logs, registre, métadonnées, trafic réseau)
- **Analyse de malwares assistée** : désassemblage et classification automatique, extraction de C2 et d'IOC
- **Traitement de volumes massifs** : analyse de copies forensiques de plusieurs téraoctets en temps réduit
- **Détection de stéganographie et d'obfuscation** : identification de données cachées ou de code masqué

CE QUE L'IA FAIT PERDRE

- **Réduction des missions sur les incidents courants** : ransomwares documentés investigués automatiquement par les plateformes XDR
- **Perte de valeur des compétences bas-niveau sur les cas connus** : la timeline reconstruite par l'IA réduit la valeur de l'expert sur les cas standards
- **Risque de dépendance aux outils** : perte progressive de la maîtrise des techniques forensics fondamentales

CE QUE L'IA NE PEUT PAS REMPLACER

- Investigation sur les attaques inédites et sophistiquées (APT, supply chain)
- **Témoignage expert en justice** : aucune IA ne peut rendre compte devant un tribunal
- Investigation en environnements dégradés (systèmes sans logs, réseaux OT/ICS)

VERDICT

Métier préservé sur le segment expert. Menacé sur le segment junior/standard. Les effectifs totaux diminuent modérément, la valeur des experts augmente.

HORIZON TEMPOREL

4 à 7 ans pour une transformation significative.

TECHNIQUE · 3.1

Administrateur Sécurité (Firewall, IAM, Patch Management)

● Impact très fort

CE QUE L'IA APPORTE

- **Patch management prédictif** : priorisation IA selon le risque réel et déclenchement automatique des déploiements
- **Gestion dynamique des règles de filtrage** : optimisation et nettoyage automatiques des règles firewall obsolètes
- **IAM adaptatif** : ajustement automatique des droits selon le contexte (localisation, heure, comportement)
- **Détection des déviations de configuration** : CSPM comparant en continu la configuration réelle à la baseline de référence

CE QUE L'IA FAIT PERDRE

- **Quasi-disparition du patch management manuel** : réduction estimée de 50 à 70 % du temps consacré
- **Réduction significative des effectifs d'administration routinière** : tâches répétitives de gestion des règles et droits directement impactées (de l'ordre de 40 à 60 % des postes à terme)
- **Dépréciation des compétences de gestion manuelle** : maîtriser une console firewall ou Active Directory devient moins différenciant

CE QUE L'IA NE PEUT PAS REMPLACER

- Ingénierie et architecture des plateformes de sécurité (niveau supérieur d'expertise)
- Gestion des exceptions et cas hors périmètre des outils automatisés
- Supervision et validation des décisions automatisées sur les systèmes critiques

VERDICT

Métier fortement menacé. L'administrateur sécurité pur évolue vers ingénieur sécurité infrastructure ou gestionnaire de plateformes IA. Réduction importante des volumes de postes.

HORIZON TEMPOREL

2 à 5 ans. Déjà bien engagé dans les grandes organisations cloud-first.

3.2 - MÉTIERS DE GOUVERNANCE ET DE CONSEIL

GOVERNANCE · 3.2

RSSI - Responsable Sécurité des Systèmes d'Information

● Impact faible

CE QUE L'IA APPORTE

- **Tableaux de bord prédictifs** : score de risque dynamique calculé en temps réel pour prioriser les arbitrages sur des données objectives
- **Reporting automatisé** : rapports périodiques COMEX, Conseil d'Administration et régulateurs générés depuis les données agrégées
- **Veille réglementaire et normative** : surveillance continue des évolutions NIS2, DORA, AI Act avec alertes d'impact
- **Aide à la décision d'investissement** : modèles de ROI sécurité et quantification du risque (FAIR) pour objectiver les demandes budgétaires
- **Simulation de scénarios de crise** : modélisation de l'impact d'un incident sur l'organisation avant qu'il ne survienne

CE QUE L'IA FAIT PERDRE

- **Réduction du temps consacré au pilotage opérationnel** : risque d'affaiblissement de la connaissance terrain du RSSI
- **Risque de perte de légitimité technique** : si l'IA produit les analyses, la valeur différenciante se déplace entièrement vers le leadership
- **Pression sur les RSSI administratifs** : profils positionnés uniquement sur la conformité et la documentation fragilisés par les outils GRC IA

CE QUE L'IA NE PEUT PAS REMPLACER

- La vision stratégique et la capacité à aligner sécurité et enjeux business
- **Le leadership en crise** : irremplaçable
- La relation de confiance avec la Direction Générale et le Conseil d'Administration
- La capacité à contester les décisions de l'IA pour ne pas devenir dépendant

VERDICT

Poste préservé et valorisé pour les profils stratégiques. Périmètre élargi (risques IA), valeur augmentée pour les bons profils. Les RSSI opérationnels purs sont fragilisés.

HORIZON TEMPOREL

Évolution sur 3 à 7 ans selon la maturité des organisations.

GOVERNANCE • 3.2

DPO - Data Protection Officer

● Impact modéré

CE QUE L'IA APPORTE

- **Cartographie automatisée des traitements de données** : mise à jour du registre RGPD Article 30 par analyse IA des flux
- **Détection automatique de données personnelles** : classification IA des données non structurées (emails, documents, bases de données)
- **Veille réglementaire continue** : surveillance automatique des délibérations CNIL, décisions européennes, lignes directrices de l'EDPB
- **Génération de DPIA** : assistance à la rédaction sur la base des éléments du traitement fournis
- **Analyse de conformité contractuelle** : vérification automatique des clauses DPA dans les contrats fournisseurs

CE QUE L'IA FAIT PERDRE

- **Réduction des missions de cartographie et registre** : tâches chronophages et largement automatisables
- **Banalisation de la conformité RGPD de base** : contrôles standards automatisés, profils DPO documentaires fragilisés
- **Pression sur les DPO externalisés généralistes** : prestations de DPO mutualisé à bas coût menacées par les plateformes GRC IA

CE QUE L'IA NE PEUT PAS REMPLACER

- **Protection des données utilisées pour entraîner les IA** : arbitrer la licéité, la minimisation et l'exercice des droits au cas par cas relève du jugement du DPO
- **Conformité des traitements IA** : l'interprétation croisée RGPD × AI Act sur des traitements concrets exige un jugement expert que l'outil n'apporte pas
- **Droits des personnes face aux décisions automatisées (Article 22 RGPD)** : garantir une intervention humaine effective est, par définition, non délégable à la machine

VERDICT

Métier transformé. Le DPO registre et documentation est fragilisé. Le DPO IA & données est en forte demande. Recomposition du marché plus que disparition.

HORIZON TEMPOREL

Transition déjà engagée. Accélération sur 3 à 5 ans.

GOVERNANCE • 3.2

Risk Manager Cyber

● Impact fort

CE QUE L'IA APPORTE

- **Quantification automatique du risque** : modèles Monte Carlo assistés par IA calculant des fourchettes d'exposition financière en temps réel
- **Scoring dynamique des actifs** : évaluation continue de la criticité selon exposition, vulnérabilité et valeur métier
- **Corrélation risque cyber / risque métier** : liens entre incidents techniques et impacts business mis en évidence automatiquement
- **Benchmarking sectoriel automatisé** : comparaison du niveau de risque avec des pairs sectoriels

CE QUE L'IA FAIT PERDRE

- **Réduction des missions d'évaluation qualitative** : matrices de risque colorées remplacées par des modèles quantitatifs dynamiques
- **Diminution de la valeur des livrables standardisés** : rapport de risk assessment produit par l'IA en une fraction du temps
- **Pression sur les Risk Managers juniors** : collecte et organisation des données de risque directement absorbées, avec une réduction estimée de 20 à 30 % des postes

CE QUE L'IA NE PEUT PAS REMPLACER

- **La définition du cadre d'appétit au risque** : décision stratégique non automatisable
- L'arbitrage et la communication du risque résiduel aux instances dirigeantes
- La gestion des risques émergents liés à l'IA elle-même (risque de modèle, biais)

VERDICT

Métier transformé, valeur déplacée vers le haut. Profils quantitatifs valorisés, profils purement qualitatifs fragilisés.

HORIZON TEMPOREL

3 à 6 ans.

GOVERNANCE · 3.2

Auditeur Cybersécurité (ISO 27001, NIS2, DORA, HDS)

● Impact fort

CE QUE L'IA APPORTE

- **Audit continu** : surveillance permanente de la conformité aux référentiels, signalant les non-conformités en temps réel
- **Collecte automatisée des preuves** : extraction, formatage et indexation automatiques des preuves d'audit, pour un gain de 40 à 60 % sur la collecte
- **Génération de rapports d'audit** : sections descriptives et constats standardisés produits par LLM
- **Traçabilité et répétabilité** : audits IA parfaitement traçables et reproductibles

CE QUE L'IA FAIT PERDRE

- **Réduction des heures facturables sur la collecte et la documentation** : cœur économique des missions directement compressé
- **Banalisation des contrôles de premier niveau** : vérifications ISO 27001 standards automatisables
- Réduction de 30 à 50 % des effectifs juniors en audit de conformité
- **Pression sur la fréquence et le prix des missions** : si la conformité est monitorée en continu, le client questionne la valeur de l'audit annuel

CE QUE L'IA NE PEUT PAS REMPLACER

- Le jugement professionnel sur les non-conformités complexes ou ambiguës
- L'interprétation des exigences réglementaires dans des contextes atypiques
- La relation de confiance avec l'audité et la capacité à lever les blocages organisationnels
- Le témoignage expert en cas de litige ou d'enquête réglementaire
- **La certification des systèmes de management de l'IA (ISO/IEC 42001)** : marché d'audit naissant qui exige le jugement d'un auditeur habilité

VERDICT

Métier significativement menacé dans sa dimension opérationnelle. La valeur se concentre sur le jugement expert. Réduction notable des effectifs juniors. Petits auditeurs indépendants généralistes les plus exposés.

HORIZON TEMPOREL

3 à 6 ans pour une réduction significative des volumes de missions traditionnelles.

GOVERNANCE · 3.2

Consultant GRC (Gouvernance, Risques, Conformité)

● Impact très fort

CE QUE L'IA APPORTE

- **Production accélérée des livrables** : politiques, procédures, plans de traitement des risques générés par LLM sur contexte client
- **Adaptation réglementaire automatisée** : croisement NIS2, ISO 27001, DORA avec identification des synergies et points de tension
- **Benchmark automatique** : comparaison instantanée du niveau de maturité du client avec des pairs sectoriels
- **Suivi de plan d'actions** : outils GRC IA alertant sur les dérives de chantiers

CE QUE L'IA FAIT PERDRE

- **Dépréciation massive des livrables templating** : politiques et matrices standardisées directement substituables par les LLM
- **Réduction du volume de missions de mise en conformité** : organisations adoptant des plateformes GRC IA réduisant leur besoin d'accompagnement
- Réduction potentielle de 30 à 50 % des effectifs juniors en conseil GRC
- **Érosion des marges** : clients comparant livrables consultants aux outputs IA et challengeant les honoraires

CE QUE L'IA NE PEUT PAS REMPLACER

- **La conduite du changement** : faire adopter une politique de sécurité est un défi humain, pas technique
- **L'expertise sectorielle spécialisée (OIV, santé, finance)** : contextes complexes exigeant un jugement expert
- **La posture de trusted advisor auprès des dirigeants** : relation de confiance non automatisable
- **L'intégration des risques IA dans les cadres GRC** : nouveau champ en forte demande

VERDICT

Métier fortement restructuré. Le consultant producteur de documents est menacé. Le consultant stratège et accompagnateur du changement est valorisé. Polarisation forte du marché.

HORIZON TEMPOREL

2 à 5 ans. Déjà perceptible sur les appels d'offres GRC.

GOVERNANCE • 3.2

Juriste Spécialisé Cybersécurité

● Impact modéré

CE QUE L'IA APPORTE

- **Recherche juridique accélérée** : analyse de jurisprudence et textes réglementaires sur de larges corpus
- **Revue contractuelle assistée** : vérification automatique des clauses de sécurité dans les contrats (SLA, DPA, notification d'incident)
- **Veille réglementaire en temps réel** : NIS2, DORA, AI Act, décisions CNIL, ANSSI
- **Génération de templates contractuels** : premiers jets de clauses contractuelles standards

CE QUE L'IA FAIT PERDRE

- **Réduction des missions de recherche juridique documentaire** : part significative du temps junior absorbée par l'IA
- **Pression sur la revue contractuelle de masse** : audits de grandes bases contractuelles en partie automatisables
- **Banalisation des conseils sur les textes connus** : RGPD, NIS2 ou DORA standard challengés par les LLM

CE QUE L'IA NE PEUT PAS REMPLACER

- Le conseil stratégique en cas de crise (incident, notification, litige)
- L'interprétation de textes ambigus dans un contexte factuel complexe
- La représentation et la négociation face aux régulateurs
- **La responsabilité professionnelle** : seul un juriste habilité peut l'engager
- **La spécialité droit de l'IA appliqué à la cybersécurité** : champ en forte émergence

VERDICT

Métier transformé, profils juniors fragilisés. La valeur se concentre sur le conseil expert, la gestion de crise et la négociation. Droit de l'IA en forte émergence.

HORIZON TEMPOREL

3 à 6 ans.

GOVERNANCE • 3.2

CISO-as-a-Service / vCISO

● Impact fort

CE QUE L'IA APPORTE

- **Démultiplication de la capacité d'intervention** : portefeuille clients plus large grâce à la surveillance continue par l'IA entre les interventions
- **Reporting client automatisé** : tableaux de bord générés automatiquement, personnalisés par secteur et maturité
- **Veille mutualisée augmentée** : synthèse automatique des menaces sectorielles pertinentes pour chaque client

CE QUE L'IA FAIT PERDRE

- **Pression sur les offres bas de gamme** : plateformes GRC IA à coûts inférieurs capturant le segment PME peu exigeant
- **Réduction des missions de reporting et documentation** : tâches chronophages absorbées par l'automatisation
- **Concurrence des plateformes CISO IA** : solutions SaaS se positionnant comme CISO virtuel pour les petites structures

CE QUE L'IA NE PEUT PAS REMPLACER

- **La relation de confiance avec la direction** : irremplaçable
- La gestion de crise et la prise de décision sous pression
- L'expertise sur les secteurs réglementés où le jugement humain certifié est exigé

VERDICT

Métier sous double pression : concurrence des outils IA en bas, rareté des bons profils en haut. Les vCISO excellents sont plus demandés que jamais. Les généralistes documentation sont menacés.

HORIZON TEMPOREL

Déjà perceptible. Accélération sur 2 à 4 ans sur le segment PME.

GOVERNANCE • 3.2

Formateur / Responsable Sensibilisation Cybersécurité

● Impact fort

CE QUE L'IA APPORTE

- **Génération de contenus pédagogiques sur mesure** : modules, quiz et cas pratiques déclinés par secteur, métier et niveau, en une fraction du temps de conception
- **Simulations de phishing plus réalistes** : campagnes d'entraînement générées par IA, personnalisées et actualisées en continu sur les techniques observées
- **Parcours de sensibilisation adaptatifs** : e-learning ajustant contenu, difficulté et rythme au niveau réel de chaque apprenant, avec déclinaison multilingue instantanée

CE QUE L'IA FAIT PERDRE

- **Banalisation des contenus génériques** : les catalogues de sensibilisation standards deviennent directement substituables par les plateformes IA
- **Pression sur la formation catalogue à bas coût** : le e-learning généré par IA tire les prix du marché vers le bas
- **Réduction du temps facturable de conception** : la valeur se déplace de la production de supports vers l'ingénierie pédagogique et l'animation

CE QUE L'IA NE PEUT PAS REMPLACER

- **L'ancrage comportemental** : faire évoluer durablement les comportements est un défi humain, pas documentaire
- La crédibilité du retour d'expérience terrain et l'adaptation en direct aux réactions d'un groupe
- **La responsabilité pédagogique et la certification** : référentiels qualité (Qualiopi) et parcours certifiants exigent un formateur identifié

VERDICT

Métier transformé. Le diffuseur de contenus standards est fragilisé. Le formateur-ingénieur pédagogique, augmenté par l'IA et centré sur les comportements, est valorisé.

HORIZON TEMPOREL

Transformation déjà engagée.
Accélération sur 2 à 4 ans.

3.3 Les nouveaux métiers qui émergent

L'IA ne fait pas que transformer les métiers existants : elle en crée ou, plus souvent encore, fait évoluer des métiers actuels vers de nouvelles spécialisations. Les profils ci-dessous, quasi inexistant il y a trois ans, figurent désormais dans les référentiels de compétences et les offres d'emploi, à l'intersection directe de la cybersécurité et de l'IA. Un ingénieur AppSec évolue ainsi vers l'AI Security Engineering, un auditeur ISO 27001 vers l'audit de systèmes IA, un Risk Manager vers la fonction d'AI Risk Officer.

CÔTÉ TECHNIQUE :

- **AI Security Engineer** : sécurisation des pipelines ML/LLM contre les attaques adverses (data poisoning, extraction de modèles, adversarial examples).
- **MLSecOps Engineer** : intégration de la sécurité dans l'ensemble du cycle de vie des modèles IA, de l'entraînement au déploiement.
- **Prompt Injection Specialist** : attaques et défenses propres aux LLM et aux assistants IA d'entreprise.
- **AI Forensic Investigator** : investigation des incidents impliquant des systèmes IA (compromission de modèles, décisions erronées).
- **Autonomous Agent Security Architect** : sécurisation et supervision des agents IA autonomes. Les premières évaluations de monitorabilité⁵ montrent que surveiller efficacement un agent reste un problème de recherche ouvert.

CÔTÉ GOUVERNANCE ET CONSEIL :

- **AI Risk Officer / AI Governance Manager** : pilotage des risques liés à l'IA, en lien direct avec les obligations de l'AI Act européen.
- **Auditeur de systèmes IA** : évaluation de la robustesse, des biais et de l'explicabilité des modèles.
- **Lead Auditor ISO/IEC 42001** : audit et certification des systèmes de management de l'IA, débouché naturel des auditeurs ISO 27001 expérimentés.
- **Data Ethics & Security Consultant** : à l'intersection du RGPD, de l'AI Act et de la cybersécurité.

Ces profils partagent un socle commun : la double compétence cyber-IA, aujourd'hui rare et fortement valorisée sur le marché. Ils constituent les débouchés naturels des parcours d'évolution interne et de reconversion évoqués en Partie IV.

⁵. METR, Early work on monitorability evaluations, janvier 2026 (metr.org/blog/2026-01-19-early-work-on-monitorability-evaluations).

TABLEAU DE SYNTHÈSE

Vue d'ensemble des impacts par métier

● **Très fort** : métier fortement restructuré● **Fort** : métier transformé● **Modéré** : métier relativement préservé● **Faible** : métier préservé, voire renforcé

Les horizons sont indicatifs. Les réductions de postes sont des estimations Kedros, à lire comme des ordres de grandeur.

MÉTIER	IMPACT IA	RÉDUCTION POSTES	HORIZON	CE QUI SUBSISTE
Analyste SOC N1	● Très fort	40-60 %	3-5 ans (5-8 ans PME/ETI)	Supervision flux automatisés
Analyste SOC N2	● Fort	20-35 %	2-4 ans	Investigation complexe, supervision IA
Analyste SOC N3 / IR	● Faible	Stable	5-10 ans	Crises complexes, APT, décision
Pentesteur	● Fort	20-30 % (junior)	3-5 ans	Créativité offensive, Red Team
Threat Hunter	● Modéré	Stable	3-6 ans	Architecte détection, TTP inédites
Architecte Sécurité	● Modéré	Faible	4-7 ans	Arbitrages stratégiques
Développeur AppSec	● Fort	20-30 % (junior)	2-4 ans	Vulnérabilités logiques complexes
Forensic Investigator	● Modéré	Faible (senior)	4-7 ans	APT, expertise judiciaire
Admin Sécurité	● Très fort	40-60 %	2-5 ans	Ingénierie plateformes IA
RSSI	● Faible	Stable	3-7 ans	Leadership, stratégie, crise
DPO	● Modéré	Faible	3-5 ans	AI Act, décisions automatisées
Risk Manager Cyber	● Fort	20-30 % (junior)	3-6 ans	Quantification, appétit au risque
Auditeur conformité	● Fort	30-50 % (junior)	3-6 ans	Jugement expert, litiges
Consultant GRC	● Très fort	30-50 % (junior)	2-5 ans	Changement, expertise sectorielle
Juriste cyber	● Modéré	Faible (senior)	3-6 ans	Crise, négociation, droit IA
vCISO	● Fort	Variable	2-4 ans	Confiance, crise, réglementé
Formateur cyber	● Fort	Variable	2-4 ans	Comportement, présentiel, certification



PARTIE III

Points de vigilance, limites et risques

-
- 4.1 Les limites intrinsèques de l'IA en cybersécurité
 - 4.2 Les risques pour les professionnels et les organisations
 - 4.3 Les angles morts collectifs
-

4.1 Les limites intrinsèques de l'IA en cybersécurité

4.1.1 - Le problème des faux positifs et faux négatifs

L'IA détecte ce qu'elle a appris. Cette apparente tautologie cache une vulnérabilité structurelle : les modèles sont aveugles face aux attaques inédites (zero-day, TTP non documentées) et peuvent générer une surcharge décisionnelle si leur calibration est insuffisante. La confiance aveugle dans un score de risque produit par un modèle non maîtrisé constitue un risque opérationnel réel.

4.1.2 - L'opacité des modèles (Explainability Gap)

Les modèles ML complexes produisent des décisions difficiles à expliquer. Dans un contexte réglementé (NIS2, DORA, RGPD), cette opacité pose des problèmes juridiques concrets : comment justifier un blocage automatique ou un constat d'audit généré par un modèle dont on ne peut pas expliquer le raisonnement ? L'IA explicable (XAI) reste un chantier ouvert.

4.1.3 - La dépendance à la qualité des données

Garbage in, garbage out. Un modèle entraîné sur des données biaisées ou contaminées produira des décisions défectueuses. En cybersécurité, les données d'entraînement sur les incidents rares (APT, attaques ciblées) sont structurellement insuffisantes, créant des angles morts précisément là où les risques sont les plus élevés.

4.1.4 - La surface d'attaque des systèmes IA eux-mêmes

- **Adversarial examples** : manipulation d'entrées pour tromper un modèle de détection
- **Data poisoning** : corruption des données d'entraînement pour affaiblir un modèle en production
- **Model inversion et extraction** : vol du modèle ou de ses données d'entraînement
- **Prompt injection** : manipulation des assistants LLM utilisés par les équipes cyber

4.1.5 - Le biais de perception des gains de productivité

Les gains annoncés par les éditeurs, et ressentis par les utilisateurs, doivent être confrontés à des mesures contrôlées. L'essai contrôlé randomisé conduit par METR à l'été 2025 auprès de développeurs open-source expérimentés⁶ a produit un résultat contre-intuitif : équipés d'outils IA, ils étaient environ 19 % plus lents sur leurs propres projets, tout en restant convaincus d'avoir gagné environ 20 % de temps. METR a d'ailleurs signalé en février 2026 que ces résultats historiques ne reflétaient plus l'impact des outils récents : la reconduction de l'expérience s'est heurtée à un biais de sélection inattendu, nombre de développeurs refusant désormais de travailler sans IA. À l'inverse, l'enquête déclarative menée par le même organisme début 2026 auprès de 349 professionnels techniques⁷ rapporte une valeur de travail perçue multipliée par 1,4 à 2, des chiffres que METR invite lui-même à considérer avec prudence. La leçon pour les décideurs cyber : mesurer les gains réels sur ses propres cas d'usage avant de redimensionner une équipe sur la foi de pourcentages génériques.

4.2 Les risques pour les professionnels et les organisations

4.2.1 - La sur-délégation et la perte de compétences humaines

Le risque de deskilling est probablement le plus insidieux. Lorsque l'IA gère le quotidien opérationnel, les analystes juniors n'apprennent plus les fondamentaux par la pratique. En situation de crise ou de défaillance des outils, les équipes se retrouvent sans les réflexes acquis par l'expérience manuelle. Ce pilote automatique organisationnel peut devenir un facteur aggravant lors d'incidents majeurs.

4.2.2 - Les risques de biais et de discrimination algorithmique

Les modèles de scoring comportemental (UEBA) peuvent reproduire et amplifier des biais présents dans les données d'entraînement. Un employé dont le comportement s'écarte de la norme pour des raisons légitimes peut être injustement signalé. La responsabilité légale en cas de décision discriminatoire produite par un système automatisé est un angle mort réglementaire majeur.

⁶. METR, Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity, juillet 2025 (metr.org/blog/2025-07-10-early-2025-ai-experienced-os-dev-study) ; mise à jour de février 2026 signalant que ces résultats ne reflètent plus l'impact des outils récents (metr.org/blog/2026-02-24-uplift-update).

⁷. METR, Measuring the Self-Reported Impact of Early-2026 AI on Technical Worker Productivity, mai 2026 (metr.org/blog/2026-05-11-ai-usage-survey).

4.2.3 - La concentration des risques chez les fournisseurs IA

Un nombre limité d'acteurs fournit les modèles IA sur lesquels repose une part croissante des défenses organisationnelles. Cette concentration crée un risque systémique : une vulnérabilité dans un modèle IA largement déployé pourrait affecter simultanément des milliers de SOC. Les questions de souveraineté des données sont directement liées à ce risque.

4.2.4 - Le risque réglementaire et de conformité

L'AI Act européen introduit une classification des systèmes IA par niveau de risque. Entré en vigueur en août 2024, il s'applique aux modèles d'IA à usage général depuis août 2025, mais son calendrier a été réaménagé en cours de route : le paquet « Digital Omnibus », adopté par le Parlement européen le 16 juin 2026 puis par le Conseil le 29 juin, la publication au Journal officiel, attendue avant le 2 août 2026, étant la seule étape qui rendra ces nouvelles dates juridiquement définitives, reporte les obligations pleines des systèmes à haut risque au 2 décembre 2027 pour les systèmes autonomes de l'Annexe III et au 2 août 2028 pour ceux intégrés à des produits réglementés⁸. Certains outils de cybersécurité utilisant l'IA pourraient être classifiés haut risque, soumettant les organisations à des obligations renforcées de transparence et d'audit. Ce report n'est pas un répit général : les obligations de transparence de l'article 50 s'appliquent bien à compter d'août 2026, et de nouvelles interdictions (génération d'images intimes non consenties et de contenus pédocriminels) entrent en vigueur dès décembre 2026. L'articulation RGPD / AI Act / NIS2 / DORA constitue un champ de complexité réglementaire que peu d'organisations ont anticipé. Le Cyber Resilience Act (CRA), dont l'essentiel des obligations s'appliquera fin 2027, ajoutera une couche supplémentaire pour les produits comportant des éléments numériques. Ce feuillet législatif illustre d'ailleurs la thèse de ce livre blanc : la valeur des professionnels capables de naviguer entre ces textes ne cesse de croître.

8. Règlement (UE) 2024/1689 du Parlement européen et du Conseil (AI Act), tel qu'amendé par le règlement « Digital Omnibus » sur l'IA adopté en juin 2026, texte intégral sur EUR-Lex.

4.2.5 - Le risque social et humain

Les suppressions de postes dans les catégories juniors ne sont pas sans conséquences organisationnelles. La formation continue, les parcours de reconversion et le maintien de la filière de formation initiale doivent être pensés en amont pour éviter une pénurie de profils seniors dans 5 à 10 ans, car ce sont les juniors d'aujourd'hui qui seront les experts de demain.

Des mécanismes concrets existent pour désamorcer ce paradoxe. Redéfinir le poste d'entrée, d'abord : le junior n'est plus trieur d'alertes mais superviseur outillé de l'IA, et c'est en contestant les conclusions de la machine qu'il apprend le métier. Préserver ensuite la pratique manuelle par des exercices réguliers sans assistance, comme les pilotes qui volent à la main pour entretenir leurs réflexes. S'appuyer sur les MSSP et l'alternance, qui deviennent de fait les nouvelles écoles de formation du secteur. Et assumer, côté employeurs, que la formation initiale est un investissement de résilience, non un coût d'exploitation.

4.2.6 - Le Shadow AI : l'angle mort du quotidien

Le risque IA le plus répandu n'est pas toujours dans les grands projets : il est dans les usages non déclarés. Collaborateurs soumettant des documents internes à des chatbots publics, extensions de navigateur non validées, comptes personnels utilisés à des fins professionnelles : ce Shadow AI expose données personnelles, secrets d'affaires et code source hors de tout cadre contractuel et de toute journalisation. L'interdiction pure pousse ces usages dans la clandestinité ; la réponse efficace combine une politique d'usage claire, un catalogue d'outils approuvés, la sensibilisation des équipes et des garde-fous techniques (filtrage, DLP). Pour les métiers de la gouvernance, c'est un chantier immédiat, et une illustration de plus que la maîtrise de l'IA en entreprise est d'abord un sujet humain et organisationnel.

4.3 Les angles morts collectifs

- **L'illusion de la neutralité de l'IA** : les choix d'entraînement et de déploiement reflètent des priorités et des biais humains.
- **Le manque de standards internationaux sur l'IA en sécurité** : MITRE ATLAS et NIST AI RMF existent mais leur adoption reste fragmentée.
- **La cyber-IA dans les conflits géopolitiques** : vecteur de désinformation, de manipulation des marchés et de sabotage des infrastructures critiques.



IV

PARTIE IV

Recommandations stratégiques

Ces recommandations n'appellent ni le même horizon ni le même effort. Dans les 90 jours : cartographier les usages IA existants, Shadow AI compris, publier une charte d'usage et lancer la sensibilisation. Dans les 6 à 18 mois : formaliser la gouvernance (AI Security Policy, garde-fous outillés sur les décisions critiques), évaluer les fournisseurs et engager les premiers parcours d'évolution des juniors. Au-delà : refondre la filière de formation et préparer les certifications émergentes comme ISO/IEC 42001.

La priorité dépend aussi de la taille de l'organisation : une PME sans RSSI commencera par la charte d'usage, la sensibilisation et un appui externe ponctuel ; une ETI y ajoutera une gouvernance formalisée et le pilotage de ses fournisseurs IA ; un grand groupe traitera de surcroît la refonte du pipeline junior et la contribution aux référentiels sectoriels.

5.1 Pour les organisations

5.2 Pour les professionnels de la cybersécurité

5.3 Pour les décideurs et instances de gouvernance

5.1 Pour les organisations

- Cartographier les tâches à fort potentiel d'automatisation versus celles à haute valeur humaine avant tout déploiement d'IA.
- Mettre en place une gouvernance de l'IA en cybersécurité (AI Security Policy) définissant les règles d'usage et les garde-fous humains obligatoires.
- Investir dans la formation continue et la montée en compétences IA des équipes cyber, à tous les niveaux de séniorité.
- Ne pas automatiser sans garde-fous humains sur les décisions critiques : toute décision à impact significatif doit rester sous contrôle humain. Les premières évaluations empiriques de la supervision d'agents IA⁹ montrent que détecter un agent s'écartant de sa mission reste difficile : le garde-fou humain doit être outillé, pas seulement déclaré.
- Évaluer les fournisseurs d'IA selon des critères de sécurité, de souveraineté et d'explicabilité, pas uniquement de performance.
- Planifier les reconversions : anticiper la réduction des postes juniors en développant des parcours vers les nouveaux métiers émergents.

5.2 Pour les professionnels de la cybersécurité

- Construire une culture IA : comprendre les fondamentaux du Machine Learning et des LLM sans devenir data scientist.
- Valoriser les compétences que l'IA ne peut pas reproduire : jugement contextuel, éthique professionnelle, communication, gestion de crise.
- S'engager dans les référentiels émergents : OWASP LLM Top 10, MITRE ATLAS, NIST AI RMF.
- Anticiper les évolutions de carrière : spécialisation (sécurité des systèmes IA) ou hybridation (profil cyber + compétences IA).
- Maintenir les compétences fondamentales manuelles : assurance-vie professionnelle en cas de défaillance des outils.

5.3 Pour les décideurs et instances de gouvernance

- Intégrer l'IA dans les plans directeurs de sécurité et les PSSI dès maintenant, pas comme un projet futur mais comme une réalité en cours.
- Sensibiliser les COMEX et CA aux nouveaux risques IA : risque de modèle, dépendance fournisseur, biais algorithmique.
- Collaborer avec les régulateurs pour co-construire les référentiels d'audit des systèmes IA en cybersécurité.
- Soutenir les politiques de souveraineté numérique et les filières de formation nationales en cybersécurité et en IA.

9. METR, Early work on monitorability evaluations, janvier 2026 (metr.org/blog/2026-01-19-early-work-on-monitorability-evaluations).

CONCLUSION

Conclusion

L'intelligence artificielle ne sonne pas le glas des métiers de la cybersécurité. Elle en redéfinit profondément la nature, la hiérarchie de valeur et les frontières. Ce que l'analyse par métier révèle avant tout, c'est une polarisation : les tâches répétitives, mécaniques et documentaires seront largement absorbées par l'automatisation, tandis que les compétences de jugement, de leadership, de créativité et de relation humaine verront leur valeur augmenter.

Cette transformation n'est ni uniforme ni instantanée. Elle suivra des rythmes différents selon la taille des organisations, leur secteur d'activité et leur niveau de maturité numérique. Mais sa direction est certaine : les organisations qui anticipent ces changements seront les mieux positionnées pour naviguer dans un paysage de menaces qui, lui, ne ralentit pas, et sauront optimiser leurs coûts.



La cybersécurité reste fondamentalement une discipline humaine. L'IA en est le nouveau levier d'efficacité, pas le successeur.

Ce livre blanc a été conçu par plusieurs humains réels. La rédaction est assistée par IA.



ANNEXES

Annexes

-
- | | |
|---|-----------|
| A | Glossaire |
|---|-----------|
-
- | | |
|---|---------------------------------|
| B | Frameworks et référentiels clés |
|---|---------------------------------|
-
- | | |
|---|---------------|
| C | Bibliographie |
|---|---------------|
-

ANNEXE A - GLOSSAIRE

TERME	DÉFINITION
APT (Advanced Persistent Threat)	Menace persistante avancée : attaquants sophistiqués (souvent étatiques) opérant sur le long terme dans les systèmes ciblés.
BAS (Breach and Attack Simulation)	Outils simulant automatiquement des scénarios d'attaque pour tester les défenses en continu.
C2 (Command and Control)	Infrastructure par laquelle un attaquant pilote à distance les machines compromises (envoi d'ordres, exfiltration de données). Aussi noté C&C.
CSPM (Cloud Security Posture Management)	Surveillance continue de la configuration des environnements cloud par rapport à des baselines de sécurité.
Copilote IA	Assistant IA conversationnel intégré aux outils de travail : il répond aux sollicitations de l'utilisateur, qui conserve l'initiative et l'enchaînement des actions. S'oppose à l'IA agentique, où le système reçoit un objectif et planifie lui-même son exécution.
Data Poisoning	Attaque corrompant les données d'entraînement d'un modèle IA pour en altérer le comportement.
Deskilling	Perte progressive des compétences humaines fondamentales due à une dépendance excessive aux outils automatisés.
DORA	Digital Operational Resilience Act : règlement européen sur la résilience opérationnelle numérique dans le secteur financier.
DPA (Data Processing Agreement)	Accord de traitement des données imposé par l'article 28 du RGPD, encadrant contractuellement les obligations d'un sous-traitant vis-à-vis du responsable de traitement.
EDPB (European Data Protection Board)	Comité européen de la protection des données : réunit les autorités nationales (dont la CNIL) pour assurer une application cohérente du RGPD.
EDR / XDR	Endpoint / Extended Detection & Response : solutions de détection et de réponse aux incidents.
FAIR	Factor Analysis of Information Risk : méthode de quantification financière du risque cyber.
GRC	Gouvernance, Risques et Conformité : discipline couvrant le pilotage stratégique de la sécurité et de la conformité.
IA agentique	Systèmes d'IA capables de planifier et d'exécuter des séquences d'actions de façon autonome pour atteindre un objectif fixé, l'humain supervisant et validant plutôt qu'exécutant.
IOC (Indicator of Compromise)	Indicateur technique signalant qu'un système a été compromis (hash, adresse IP, domaine).
LLM (Large Language Model)	Modèle de langage de grande taille : IA générative entraînée sur de vastes corpus textuels.
Malware polymorphe	Logiciel malveillant qui modifie automatiquement son code à chaque réplique pour échapper à la détection par signature, tout en conservant sa fonction.
MITRE ATT&CK	Base de connaissance des TTP des attaquants, maintenue par l'organisation MITRE.

TERME	DÉFINITION
MITRE ATLAS	Extension de MITRE ATT&CK couvrant les attaques ciblant les systèmes d'IA et de Machine Learning.
MLSecOps	Intégration des pratiques de sécurité dans les cycles de vie des modèles de Machine Learning.
MTTR (Mean Time To Respond)	Temps moyen de réponse à un incident de sécurité.
NIS2	Network and Information Security Directive 2 : directive européenne sur la sécurité des réseaux et systèmes d'information.
NIST AI RMF	Framework de gestion des risques IA publié par le National Institute of Standards and Technology.
OWASP LLM Top 10	Liste des 10 risques de sécurité les plus critiques liés aux applications LLM.
Prompt Injection	Attaque manipulant les instructions d'un système LLM en injectant des commandes malveillantes.
Red Team	Équipe simulant des attaquants réels pour tester la résilience des défenses.
SASE (Secure Access Service Edge)	Modèle d'architecture délivrant depuis le cloud, au plus près de l'utilisateur, des fonctions réseau et de sécurité auparavant distinctes (SD-WAN, pare-feu, accès Zero Trust).
SAST (Static Application Security Testing)	Analyse statique du code source à la recherche de vulnérabilités, sans exécuter l'application, généralement intégrée aux pipelines de développement.
Shadow AI	Usage d'outils d'IA non validés par l'organisation (chatbots publics, extensions, comptes personnels), exposant données et secrets hors de tout cadre de sécurité et de conformité.
SIEM	Security Information and Event Management : plateforme de collecte et corrélation des événements de sécurité.
SOAR	Security Orchestration, Automation and Response : plateforme d'automatisation de la réponse aux incidents.
SOC	Security Operations Center : centre de surveillance et de gestion des incidents de sécurité.
STRIDE	Méthode de modélisation des menaces passant en revue six familles de risques : usurpation d'identité (Spoofing), altération (Tampering), répudiation (Repudiation), divulgation d'informations (Information disclosure), déni de service (Denial of service), élévation de privilèges (Elevation of privilege).
TTP	Tactics, Techniques & Procedures : description du comportement des attaquants.
UEBA	User and Entity Behavior Analytics : analyse comportementale pour détecter les anomalies.
vcISO	Virtual CISO : RSSI externalisé intervenant en mode conseil auprès de plusieurs organisations.
XAI (Explainable AI)	Domaine visant à rendre les décisions des modèles IA compréhensibles et justifiables par des humains.

— ANNEXE B - FRAMEWORKS ET RÉFÉRENTIELS CLÉS

- **MITRE ATT&CK** : Base de connaissance des TTP - <https://attack.mitre.org>
- **MITRE ATLAS** : Sécurité des systèmes IA - <https://atlas.mitre.org>
- **OWASP LLM Top 10** : Risques de sécurité des LLM - <https://genai.owasp.org>
- **NIST AI RMF** : Framework de gestion des risques IA - <https://www.nist.gov/itl/ai-risk-management-framework>
- **AI Act (UE)** : Règlement européen sur l'intelligence artificielle - EUR-Lex
- **NIS2** : Directive européenne sur la cybersécurité - EUR-Lex
- **DORA** : Règlement sur la résilience opérationnelle numérique - EUR-Lex
- **CRA (Cyber Resilience Act)** : Règlement européen sur la cybersécurité des produits comportant des éléments numériques - EUR-Lex
- **ISO/IEC 27001:2022** : Système de management de la sécurité de l'information
- **ISO/IEC 42001:2023** : Système de management de l'intelligence artificielle (SMIA), premier référentiel certifiable de gouvernance de l'IA
- **FAIR Institute** : Quantification financière du risque cyber - <https://www.fairinstitute.org>

— ANNEXE C - BIBLIOGRAPHIE

Rapports sectoriels et institutionnels :

- ANSSI - Panorama de la cybermenace 2024 (publié en mars 2025) : <https://cyber.gouv.fr/actualites/panorama-de-la-cybermenace-2024-mobilisation-et-vigilance-face-aux-attaquants/>
- ANSSI - Panorama de la cybermenace 2025 (publié en mars 2026) : <https://cyber.gouv.fr/nous-connaître/publications/panoramas-de-la-cybermenace/panorama-de-la-cybermenace-2025/>
- ENISA - Threat Landscape 2024 : <https://www.enisa.europa.eu/publications/enisa-threat-landscape-2024>
- ENISA - Threat Landscape 2025 : <https://www.enisa.europa.eu/publications/enisa-threat-landscape-2025>
- ISC² - 2024 Cybersecurity Workforce Study : <https://www.isc2.org/Insights/2024/10/ISC2-2024-Cybersecurity-Workforce-Study>
- ISC² - 2025 Cybersecurity Workforce Study : <https://www.isc2.org/Insights/2025/12/2025-ISC2-Cybersecurity-Workforce-Study>

Études sur les capacités de l'IA :

- METR - Measuring AI Ability to Complete Long Tasks (mars 2025) : <https://arxiv.org/abs/2503.14499>
- METR - Time Horizon 1.1 (janvier 2026) : <https://metr.org/blog/2026-1-29-time-horizon-1-1/>
- METR - Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity (juillet 2025) : <https://metr.org/blog/2025-07-10-early-2025-ai-experienced-os-dev-study/>
- METR - We are Changing our Developer Productivity Experiment Design (février 2026) : <https://metr.org/blog/2026-02-24-uplift-update/>
- METR - Early work on monitorability evaluations (janvier 2026) : <https://metr.org/blog/2026-01-19-early-work-on-monitorability-evaluations/>
- METR - Measuring the Self-Reported Impact of Early-2026 AI on Technical Worker Productivity (mai 2026) : <https://metr.org/blog/2026-05-11-ai-usage-survey/>

À PROPOS DE KEDROS CYBERSÉCURITÉ

Kedros Cybersécurité est un cabinet de conseil spécialisé en cybersécurité, basé à Paris. Nos expertises couvrent le conseil, l'audit, la maîtrise des risques et la formation professionnelle certifiante (ISO 27001, ISO 22301, ISO 27005). Nous accompagnons les organisations de toutes tailles dans leur montée en maturité cyber. Nous accompagnons également la gouvernance de l'IA : cadrage des usages, conformité AI Act et préparation à la certification ISO/IEC 42001.

SITE INTERNET <https://kedros-cybersecurite.fr>

LINKEDIN Kedros Cybersécurité

ADRESSE 149, avenue du Maine, 75014 Paris
