Radar des solutions de sécurité de l’IA

Convictions du marché

24 Septembre 2026



WAVESTONE

Forbes

Generative AI Exposes Users To New Security Risks

NIST researchers warn of top AI security threats @Statescoop

THE WALL STREET JOURNAL.

AI Is Generating Security Risks Faster Than Companies Can Keep Up

LLMs Pose Major Security Risks, Serving As 'Attack Vectors' @C3.ai

The near-term impact of AI on the cyber threat @National Cyber Security Centre

NIST Warns of Security and Privacy Risks from Rapid AI System Deployment @The Hacker News

The Guardian

Generative AI security risks underestimated by IT leaders, survey reveals @IT Brief Australia

Cybercrime: be careful what you tell your chatbot helper...

Do Free AI Tools Pose a Security Risk to Your Business? @Business News Daily

"AI" est le buzzword depuis quelques années avec son lot de NOUVEAUX RISQUES

Pourquoi la sécurité de l'IA est-elle différente ?

La prise de décision des systèmes d'IA **repose sur des statistiques**, contrairement aux systèmes informatiques **déterministes** basés sur des algorithmes

Outre les **méthodes et pratiques habituelles** en matière de cybersécurité, la sécurisation des systèmes d'IA **nécessite de nouvelles techniques**, car les menaces et les réglementations sont spécifiques à cette différence inhérente

01

Menaces sur l'IA

- Empoisonnement
- Oracle
- Evasion

02

Nouvelles régulations

- EU AI ACT
- Blueprint for an AI Bill of Rights
- AI Regulation White Paper

03

Menaces cyber classiques

- Fuite de données
- Attaques sur les modèles

Les défis et les objectifs du radar



Détecter

Identifier de Nouvelles solutions de sécurité IA

Obtenir une vision claire de l'état de l'art sur le marché



Analyser

Avoir une vision claire de la **valeur ajoutée** offerte par les solutions, de leur maturité et des risques qu'elles couvrent

Développer des connaissances techniques sur les nouveaux services en fonction du marché



Améliorer

Promouvoir des solutions pertinentes

Conseiller stratégiquement nos clients entre le développement de mesures de protection internes et la contractualisation avec les éditeurs de logiciels.

La voie vers la compréhension du marché passe d’abord par une connaissance approfondie de la sécurité de l'IA

Outils et Méthodologie

Prise de contact et interviews techniques

Veille sur les menaces et technologies

Construction d'une base de documentation interne

IA Cybersecurity Radar_Database.xlsm

Radar 2025 of AI Safety Solutions | Wavestone

Présentation des résultats

Radar_AI_SEC_V2.pptx



94

Fournisseurs identifiés en septembre 2025

Wavestone

2025 AI Security Solutions **Radar**

1 Radar – 6 catégories – 3 focus

La troisième édition de notre radar met en évidence un **changement dans la structure du marché** : le nombre de catégories est passé de 8 en 2024 à 6 aujourd'hui. Cette évolution reflète un **écosystème qui arrive à maturité**, avec des **domaines d'intérêt plus clairs et plus consolidés**.

Les **trois catégories suivantes**, sujets les plus clés pour nos clients seront présentées dans les grandes lignes.



AI Data Protection & Privacy

Prévenir les violations potentielles des données personnelles ou d'entreprise. Renforcer la confiance dans les systèmes d'IA à l'heure de la démocratisation de l'IA.



AI Firewall & Response

Fournir des outils intégrés de détection qui surveillent les entrées et les sorties afin de détecter les anomalies, d'empêcher la propagation de codes malveillants et de garantir une auditabilité complète grâce à des journaux d'interaction détaillés.



Model Robustness & Vulnerability Assessment

Évaluer les modèles d'IA afin d'identifier les vulnérabilités susceptibles d'entraîner des attaques par empoisonnement ou par contournement.

AI Data Protection & Privacy

Contexte et objectifs

Prévenir les **violations potentielles des données personnelles ou d'entreprise**.

Renforcer la confiance dans les systèmes d'IA à l'heure de la démocratisation de l'IA.

Synthetic data / Anonymization

- Améliorer les ensembles de données réelles
- Protéger la vie privée des individus en mettant en œuvre des techniques d'anonymisation

ML Secure Collaboration

- Permettre une collaboration sécurisée entre les propriétaires de données et les développeurs de modèles
- Garantir la protection et la sécurité des données sensibles tout au long du cycle de vie du système d'IA

Approche

- Générer des données réalistes, mais artificielles, qui **imitent les données réelles** sans révéler d'informations confidentielles afin, par exemple, de former ou de tester un modèle.
- L'utilisation de **coffres-forts de confidentialité** pour les LLM garantit que les données sensibles utilisées pour former ou tester des modèles restent sécurisées et inaccessibles aux entités non autorisées (à l'aide de RBAC ou ABAC). Les coffres-forts de confidentialité agissent comme des proxys entre les ensembles de données, les utilisateurs et les modèles d'IA (par exemple, les LLM).
- L'utilisation de **matériel sécurisé**, par exemple des modules de plateforme sécurisée (TPM), pour exécuter les modèles LLM et chiffrer les requêtes et les réponses garantit un véritable chiffrement de bout en bout lors de l'utilisation de modèles d'IA pour les requêtes.

Clients ciblés

Finance, santé, agences gouvernementales et autres secteurs fortement réglementés

Synthetic data / Anonymization **ML Secure Collaboration**

Exemples

Sarus

2019

11-50

Génération de données synthétiques et politiques de confidentialité relatives aux jeux de données d'IA

skyflow

2019

51-100

Coffres-forts de confidentialité pour les LLM

MITHRIL SECURITY

2021

11-50

Sécurité LLM renforcée grâce à l'utilisation du TPM et au chiffrement

AI Firewall & Response

Contexte and objectifs

Ces solutions fournissent des outils de détection des attaques positionnés aussi près que possible du système d'IA. Intégrés aux couches d'entrée et de sortie, ils analysent les données entrant et sortant du modèle. Servant à la fois de filtres de sécurité essentiels et de mécanismes de correction, ils contribuent à protéger les utilisateurs et les systèmes d'IA contre les menaces potentielles.

- Prévenir les fuites d'informations sensibles.
- Détecter et prévenir la diffusion de codes malveillants.
- Se prémunir contre les tentatives de piratage ou de contournement des règles internes.
- Garantir une auditabilité totale et générer des rapports complets sur les interactions avec les LLM.
- Mettre en place des stratégies de réponse.

Approche

- Implémentez l'outil dans le système afin de créer une couche entre l'utilisateur et le modèle. Toutes les entrées et sorties du modèle passent alors par cette nouvelle couche.
- L'outil étudie les entrées et les sorties du système. Si possible, il est préférable de donner à l'outil accès aux vecteurs d'entrée et de sortie du modèle. La solution analyse ensuite ce vecteur afin de détecter toute anomalie.
- Des tests statiques, dynamiques ou utilisant des techniques d'apprentissage automatique sont mis en œuvre pour déterminer si les entrées sont malveillantes (attaques par évasion) ou corrompues, et si les sorties contiennent des PII ou de la propriété intellectuelle. Pour ce faire, le comportement des paires d'entrées/sorties est analysé. Il est également possible de détecter les attaques de type Oracle en mesurant le nombre de requêtes effectuées par un utilisateur.

Clients ciblés

Toute entreprise, certaines start-ups ont des contrats avec de grandes banques et même des entreprises militaires.

Exemples

CALYPSOAI

2018

51 - 200

Sécurité et surveillance de LLM

HIDDENLAYER

2022

11-50

Détection et réponse pour les modèles d'IA et conseil en sécurité des modèles d'IA

LAKERA

2021

11-50

Protection du LLM contre diverses menaces, notamment les attaques par injection de prompt et la perte de données

Model Robustness & Vulnerability Assessment

Contexte and objectifs

Développer des modèles d'IA à partir de zéro nécessite des ressources et des connaissances considérables. C'est pourquoi les développeurs utilisent généralement des modèles externes ou open source pour faciliter la production de leur système d'IA. Cependant, ces modèles externes peuvent être une source de vulnérabilités pour le système. La robustesse d'un modèle fait référence à sa capacité à maintenir de bonnes performances dans diverses conditions et face à des perturbations. Ils peuvent être fragiles face à des attaques par empoisonnement (une porte dérobée peut être placée dans un modèle open source) ou à des attaques par évasion.

- Analyser les modèles externes à la recherche de vulnérabilités
- Évaluer la résilience et la robustesse des modèles afin de guider les choix des développeurs

Approche

- Le scan de modèle peut être effectué en mode boîte blanche, en analysant les lignes de code à la recherche d'éventuelles portes dérobées.
- Une étude en mode boîte noire peut également être réalisée en soumettant le modèle à diverses perturbations afin d'observer son comportement. Pour cela, des entrées spécifiques sont générées par le fournisseur afin d'explorer les limites du système en effectuant des analyses statistiques des sorties. Des tests statiques, dynamiques ou utilisant des techniques d'apprentissage automatique sont mis en place pour identifier les biais potentiels (manque de diversité dans les données d'entraînement, étiquette erronée...) et déterminer les entrées/sorties indésirables.
- Pour les LLM, des mesures de robustesse peuvent être utilisées pour évaluer le modèle, telles que : Perplexité, Précision, Bilingual Evaluation Understudy (BLEU), Recall-Oriented Understudy for Gisting Evaluation (ROUGE).

Clients ciblés

Ces solutions peuvent être utilisées par n'importe quelle entreprise, et certaines start-ups ont conclu des contrats avec de grandes banques et même des entreprises militaires.

Exemples

Citadel AI

2020

11-50

Tests complets pour la validation de la fiabilité de l'IA/des ensembles de données. Surveillance de l'IA en production.

PROTECT AI

2022

11-50

Vulnérabilité et gestion de l'IA tout au long du cycle de vie du ML.

numalis

2015

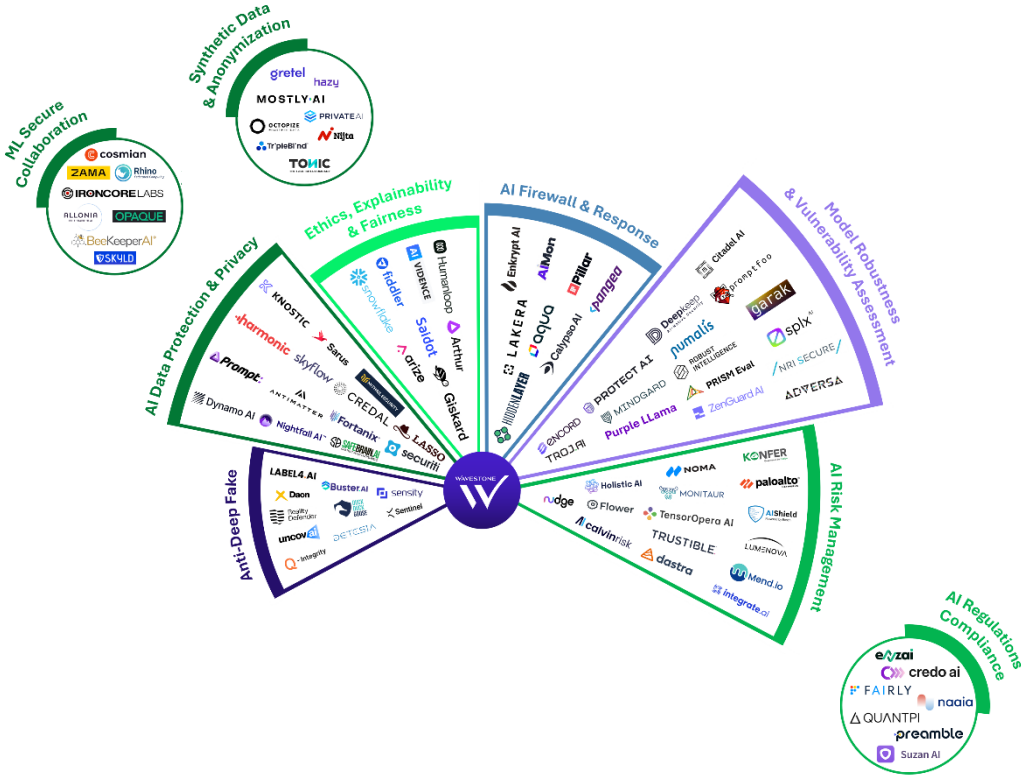
11-50

Robustesse et fiabilité des modèles d'IA selon la suite de normes ISO 24029.

Conclusion - le radar permet d'accélérer les choix de nos clients

Notre radar des solutions de sécurité IA fournit **des analyses claires** pour présélectionner les outils les plus pertinents et naviguer dans un paysage en évolution constante. Dans sa construction nous menons des entretiens directs avec les éditeurs, évaluons la maturité des solutions et fournissons des fiches d'identité pour **accélérer le processus d'appel d'offres.**.

Radar des solutions de sécurité IA : 94 solutions identifiées



Pare-feu IA & Réponse

Calypso AI

Éthique, explicabilité et équité

Giskard

Gestion des risques IA

Conformité réglementaire

naaia

Beauté

Banque **Assurance**

Énergie **Durabilité**

De l'analyse à l'action : nous accompagnons nos clients dans le déploiement des bonnes solutions de sécurité IA



Focus thématique pour des décisions stratégiques de sécurité IA



Fiches d'identité pour accélérer le processus d'appel d'offres


WAVESTONE