Mustapha Annouh
CEO SHIELD AI

Le risques des IA Generative en enterprise: Risques & Opportunites

Presentation - 2025

+32 471 35 63 06 mustapha.annouh@shieldai.ai www.shieldai.ai

Introduction

Présentation Vue d’ensemble

Dans le paysage en évolution rapide de l’IA générative, les entreprises sont confrontées au double défi d’exploiter cette technologie puissante pour stimuler l’innovation et l’efficacité, tout en naviguant dans un environnement réglementaire de plus en plus complexe et en atténuant les risques de violation de données et de non-conformité.

Table des matières

Information

Qui sommes-nous

Nos missions sont simples et claires : fournir des technologies souveraines et innovantes qui permettent aux entreprises d’exploiter tout le potentiel de l’IA générative tout en gardant un contrôle total sur leurs données et leur stratégie d’IA.

CEO

- 22 ans de cybersécurité :
- 12 ans de réseau et de sécurité
- 10 ans de gouvernance, de risque et de conformité
- Fondateur Security People
- Co-fondateur SHIELD AI

SHIELD AI

Chez Shield AI, nous sommes une startup avant-gardiste fondée en 2024 par deux vétérans de l’industrie qui apportent plus de 40 ans d’expérience combinée dans les domaines de la GRC, du DevSecOps et de la cybersécurité.

Definition

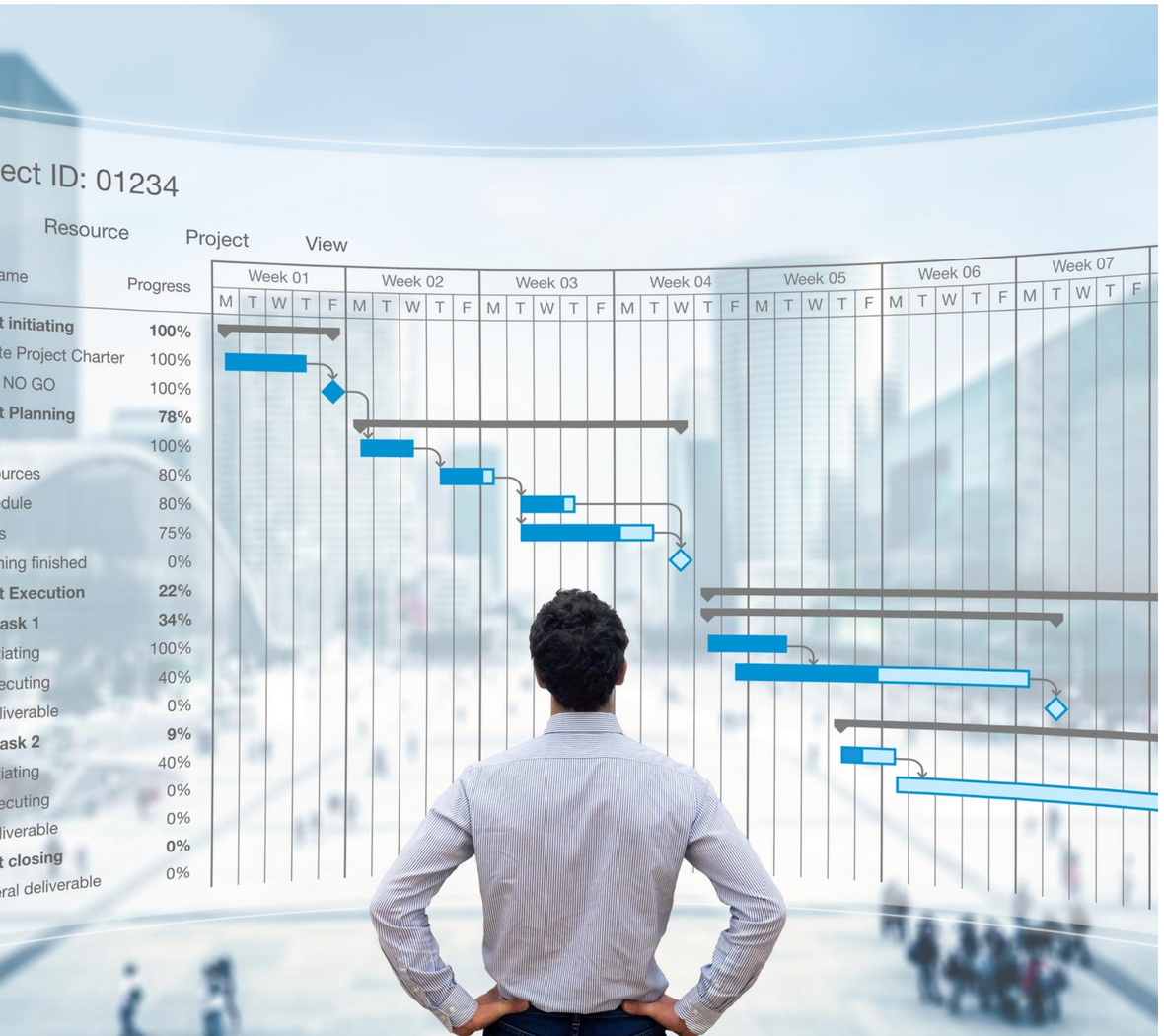
Concepts clés : « GenAi »

Qu’est-ce que l’IA générative ?

L’IA générative est une branche de l’intelligence artificielle qui crée du contenu nouveau et original en apprenant des modèles à partir de données existantes, ce qui la distingue de l’IA traditionnelle qui se concentre principalement sur la reconnaissance des modèles et la prise de décision. Essentiellement, alors que l’IA conventionnelle analyse et prédit, l’IA générative innove et produit de nouveaux résultats.

Innovation et efficacité vs vulnérabilités émergentes

Le pouvoir transformateur de l’IA générative



L’IA générative n’est pas seulement une évolution de l’IA traditionnelle, c’est une révolution qui redéfinit la façon dont nous créons, innovons et interagissons avec la technologie. En transformant les données brutes en contenu entièrement nouveau et significatif, elle brise le moule de l’analyse et de la prise de décision conventionnelles, permettant aux organisations de libérer un potentiel créatif et une efficacité opérationnelle sans précédent. Ce pouvoir de transformation ouvre la voie à des expériences personnalisées, à une automatisation dynamique et à des stratégies innovantes qui propulsent les entreprises au-delà des limites d’aujourd’hui vers un avenir aux possibilités infinies.

- Libère la créativité
- Améliore l’efficacité
- Favorise la personnalisation
- Favorise l’innovation

Risques persistants dans le paysage de l’IA

La génération d’IA, avec sa capacité à créer de nouveaux contenus, soulève également des défis que les décideurs politiques devraient chercher à relever.

Ces risques, ainsi que les risques frontières ou systémiques à long terme posés par le développement de des modèles d’IA générationnelle très avancés, sont abordés dans ce guide détaillé. Ces risques sont des considérations clés qui sous-tendent plusieurs recommandations stratégiques conçues pour les atténuer et promouvoir l’utilisation positive de l’IA de la génération Generation.

01
Erreurs et anthropomorphisme

02
Réponses incorrectes et désinformation

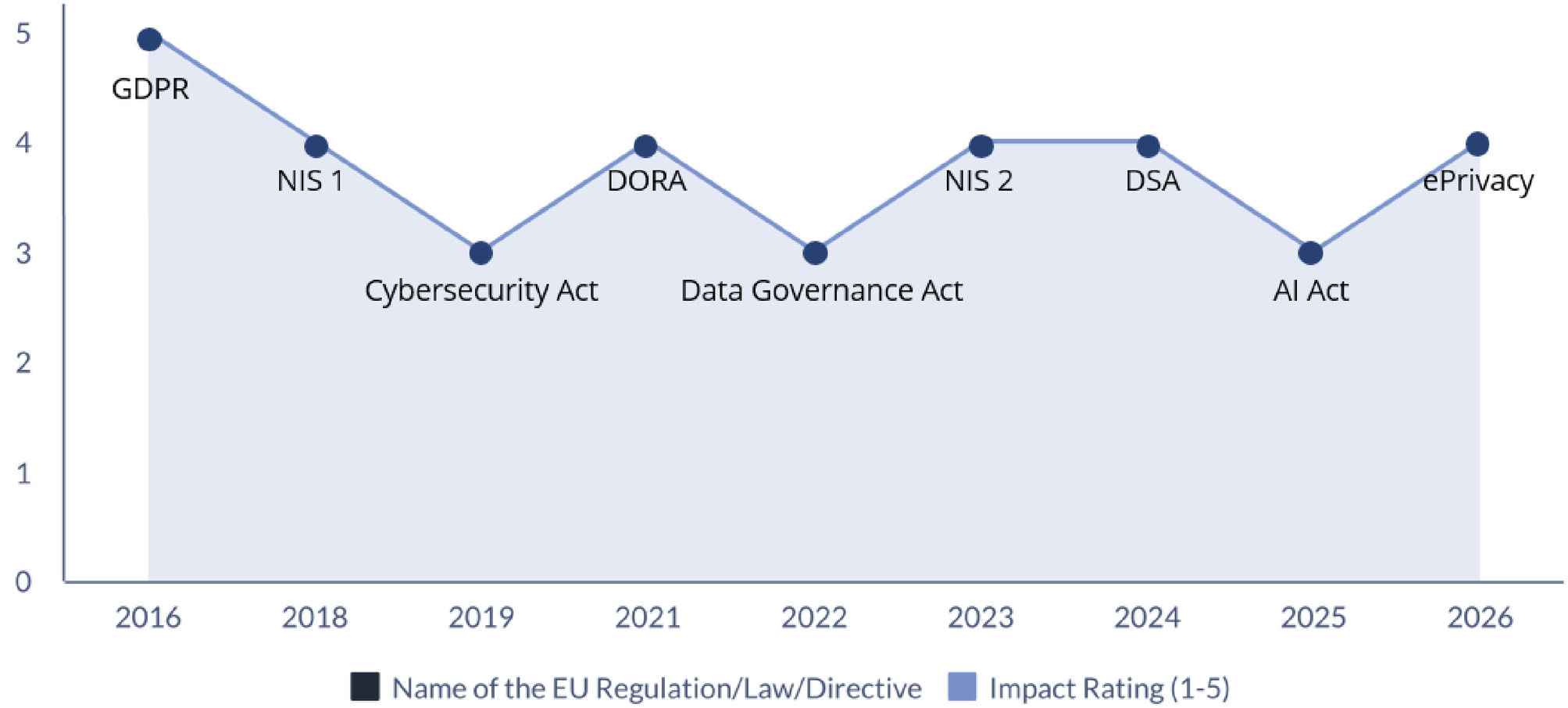
03
Deepfake et Vol d’identité

04
Propriété Intellectuelle

05
Vie privée et Confidentialité

06
Propagation de biais cognitifs

Encore plus de défis



De 2016 à 2026, les réglementations de l'UE sur les violations et les fuites de données ont considérablement augmenté, tant en nombre qu'en 2026, en mettant l'accent sur les amendes administratives et les responsabilités pénales potentielles des entreprises et des particuliers.

Défis pour les délégués à la protection des données (DPO)

Conformité et pression réglementaire

- Naviguer dans le RGPD, le IA Act et l'évolution des réglementations spécifiques à l'IA à venir
- Assurer l'anonymisation des données et la gestion du consentement dans la formation à l'IA

Risque de violation de données

- Gestion des fuites de données involontaires à partir de modèles d'entraînement

Documentation et pistes d'audit

- Difficulté à expliquer les processus décisionnels de l'IA aux organismes de réglementation

Défis concrets pour les DPO

Défi 1 : Encadrement des fuites de données via les IA génératives

Situation concrète : Un collaborateur soumet à un outil d'IA générative un prompt contenant des données sensibles (ex. : dossier médical, profil d'un client bancaire, notes RH internes).

Complexité pour le DPO :

•**Qualification juridique du traitement** :

- Qui est responsable ? L'entreprise ou le fournisseur d'IA ?
- ⚖️ *Article 4(7)* RGPD : Définir qui est responsable du traitement est fondamental pour la conformité.

•**Transferts hors UE** :

- ⚖️ *Article 44-49* RGPD : Si l'outil est américain, sans clauses contractuelles types (SCC) ou décision d'adéquation, cela constitue une violation potentielle.
- **Difficile pour le DPO de vérifier si le modèle stocke les données (et où).**

•**Obligation de documentation et d'analyse d'impact** :

- ⚖️ *Article 30* RGPD (registre des traitements)
- ⚖️ *Article 35* RGPD (DPIA – Data Protection Impact Assessment) : nécessaire si traitement de données sensibles via IA.
- **Impossible si l'outil d'IA est une boîte noire ou que le fournisseur ne fournit pas de transparence technique.**

•**Secret médical / bancaire** :

- ⚖️ Directives sectorielles comme ePrivacy (2002/58/CE modifiée), DORA (Règlement 2022/2554), et règles nationales (Code de la santé publique en France, par exemple).
- **Le DPO doit vérifier que les garanties spécifiques sectorielles sont aussi respectées — pas seulement le RGPD.**

Défis concrets pour les DPO

Défi 2 : Transparence, explicabilité et traitement équitable avec l'IA

Situation concrète : Une IA générative propose des recommandations RH, des réponses clients personnalisées, ou pré-rédige des courriers médicaux.

Complexité pour le DPO :

- **Transparence du traitement** :
  - ⚖️ *Article 13 et 14* RGPD : Obligation d'informer sur la finalité, la logique sous-jacente, les droits, les destinataires, etc.
  - **Impossible si l'éditeur ne fournit pas de *documentation technique explicite* ou de *fiche explicabilité* du modèle.**

- **Prise de décision automatisée** :
  - ⚖️ *Article 22* RGPD : Interdiction des décisions automatisées produisant des effets juridiques (rejet de crédit, recrutement, etc.) sans intervention humaine.
  - **Difficile à démontrer si l'IA influence fortement la décision humaine (ex. scoring RH ou analyse de crédit).**

- **Équité et biais** :
  - ⚖️ *Principe d'exactitude (Article 5(1)(d))* : L'IA ne doit pas produire de contenu faux ou biaisé.
  - ⚖️ *Principe de non-discrimination* (Directive 2000/43/CE & 2000/78/CE) : **Un modèle biaisé contre certaines populations est non conforme.**

- **Exercice des droits des personnes** :
  - ⚖️ *Articles 15 à 21* RGPD : Comment garantir les droits d'accès, de rectification ou d'opposition si **l'IA est non déterministe et que le modèle n'est pas auditable ?**

Défis pour les responsables de la sécurité des systèmes d'information (RSSI)

Augmentation de la surface d'attaque

- Les systèmes pilotés par l'IA peuvent être exploités par des adversaires (par exemple, des attaques antagonistes)

Intégration avec les cadres de sécurité existants

- Trouver un équilibre entre l'innovation et les mesures de sécurité existantes

Vulnérabilité et sécurité du modèle

- Protéger les modèles d'IA contre la rétro-ingénierie et l'empoisonnement des données

Stratégies d’atténuation à court, moyen et long terme

À court terme 0-6 mois

1 - Évaluations des risques immédiats :
Réaliser des audits de vulnérabilités et de flux de données sur les systèmes d’IA

2 - Renforcer la gouvernance des données :
Mettre en œuvre des contrôles d’accès améliorés et le masquage des données

3 - Formation et sensibilisation des employés :
Des formations régulières en cybersécurité et en conformité axées sur les risques liés à l’IA

4 - Évaluations des fournisseurs et des tiers :
Évaluer les pratiques de sécurité des fournisseurs de services d’IA

Moyen terme 6-18 mois

1 - Adoptez des cadres d’IA sécurisés :
Intégrez des protocoles de sécurité spécifiques à l’IA dans votre SIEM et votre réponse aux incidents

2 - Élaborer des politiques spécifiques à l’IA :
Créez des directives claires pour l’utilisation des données, les mises à jour des modèles et la surveillance de la sécurité

3 - Audits réguliers et tests d’intrusion :
Établir des évaluations continues des vulnérabilités de l’IA

4 - Collaboration renforcée :
Favoriser l’intégration d’équipes interdépartementales (juridique, informatique, sécurité) pour surveiller les déploiements d’IA

À long terme 18 mois +

1 - Investissements stratégiques dans la sécurité de l’IA :
Investir dans la R&D pour des technologies avancées de détection et d’atténuation des risques liés à l’IA

2 - Amélioration continue :
Développez une surveillance en temps réel et des réponses automatisées aux menaces

3 - IA éthique et gouvernance :
Établir des lignes directrices éthiques et des comités de surveillance de l’IA

4 - Collaboration avec l’industrie :
Participez à des consortiums sectoriels pour partager des renseignements sur les menaces et les meilleures pratiques

Mesures d’action immédiates

Appel à l’action :

- Initier une évaluation complète des risques des déploiements actuels de l’IA
- Constituer une task force transverse (DPO, CISO, IT, legal)
- Commencer des correctifs à court terme tout en planifiant des stratégies à moyen et long terme

Urgence:
Mettre l’accent sur les risques de coût et de réputation de l’inaction

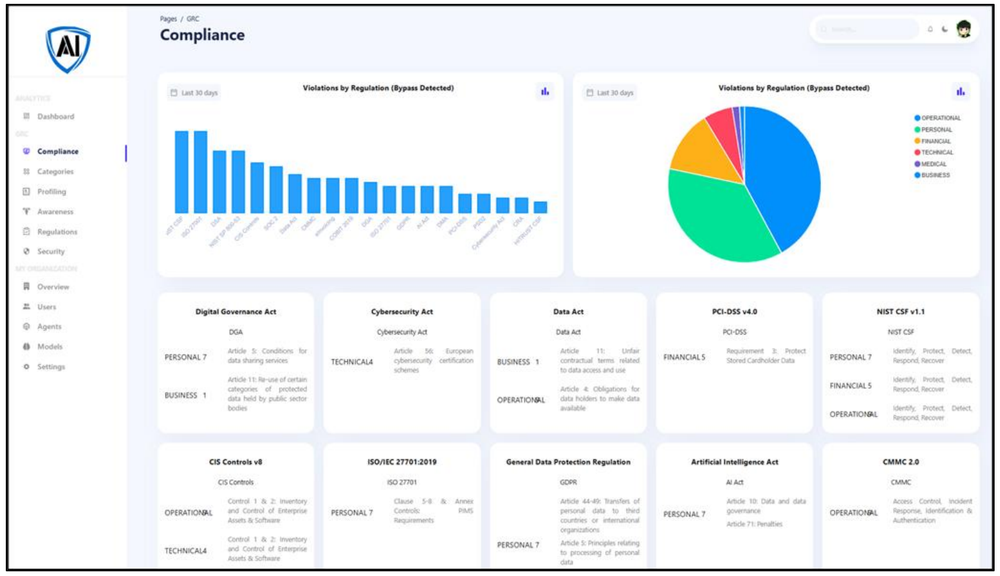
Conclusion

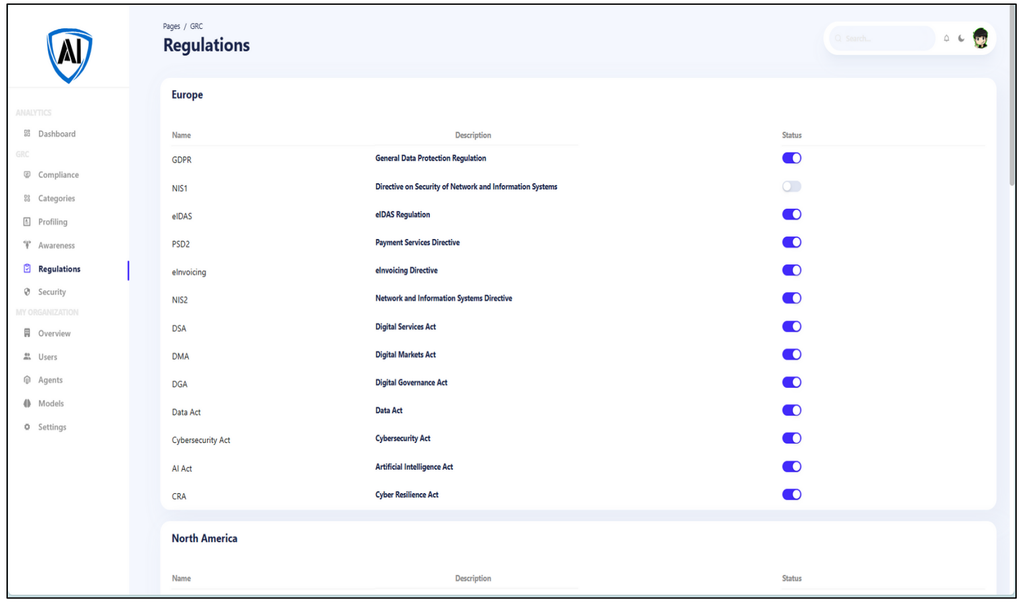
- Risques permanents, défis spécifiques pour les DPO et les RSSI, et approche de solution hiérarchisée

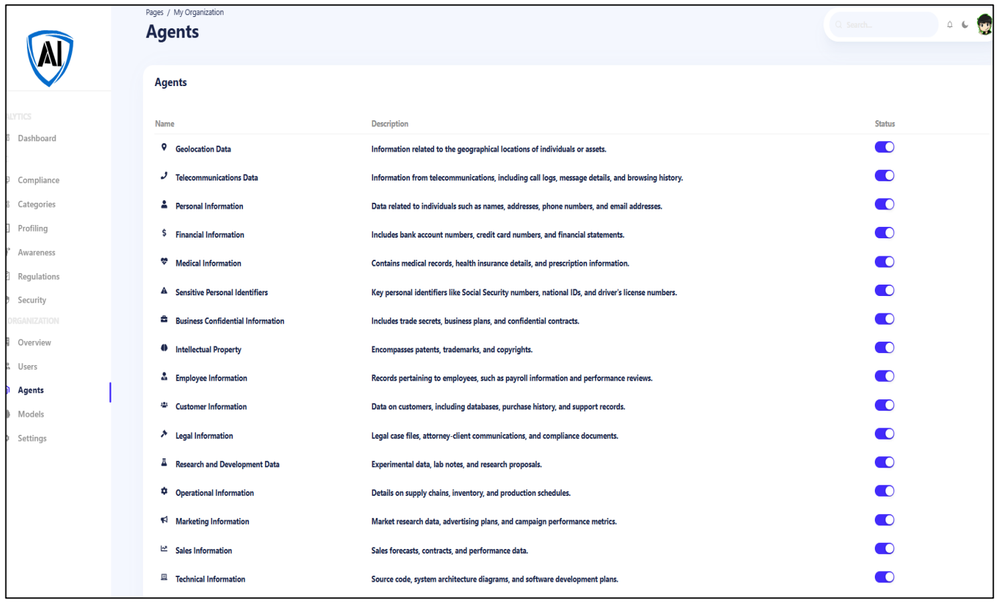
- Une action proactive et immédiate est essentielle pour préserver la confiance de votre entreprise et de vos clients

Solution commerciale : SHIELD AI (Secure prompting)

Solution commerciale : SHIELD AI (GenAI Dashboard)







References

- NIST Cybersecurity Framework:
  https://www.nist.gov/cyberframework
  A foundational resource outlining best practices in cybersecurity risk management.
- NIST AI Risk Management Framework:
  https://www.nist.gov/artificial-intelligence-risk-management
  Provides guidelines for safely developing and deploying AI systems.
- Gartner's Insights on Artificial Intelligence:
  https://www.gartner.com/en/information-technology/insights/artificial-intelligence
  Offers in-depth analysis of AI trends and its impact on cybersecurity.
- Forrester Research on Secure AI Systems:
  https://www.forrester.com/
  Features strategic guidance and best practices for integrating AI into secure systems.
- MIT Technology Review – AI Safety:
  https://www.technologyreview.com/
  An independent source providing insights on AI innovation and safety challenges.
- McKinsey Insights on AI and Cybersecurity:
  https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights
  Delivers comprehensive reports on leveraging AI to transform cybersecurity strategies.
- European Data Protection Board (EDPB) Guidelines:
  https://edpb.europa.eu/
  Offers regulatory guidance on data protection and specific insights relevant to AI

Amendes et pénalités

Regulation	Maximum Fine / Penalty	Notes
GDPR	Up to €20 million or 4% of global annual turnover, whichever is higher	Defines strict data protection and privacy rules across the EU.
NIS 1 Directive	Varies by Member State	First EU-wide cybersecurity directive for critical infrastructure security.
Cybersecurity Act	No direct fines but mandates EU cybersecurity certification	Strengthens ENISA's role and establishes the EU-wide cybersecurity certification framework.
DORA (Digital Operational Resilience Act)	Up to €10 million (sector-specific variations may apply)	Ensures financial sector resilience against cyber threats.
Data Governance Act	Up to €20 million or 4% of global annual turnover	Establishes rules for data-sharing and interoperability across sectors.
NIS 2 Directive	Up to €10 million or 2% of global annual turnover	Expands cybersecurity requirements and enforcement from NIS 1.
DSA (Digital Services Act)	Up to 6% of global annual turnover	Regulates online platforms, content moderation, and digital service providers.
AI Act	Up to €30 million or 6% of global annual turnover	Regulates high-risk AI systems with compliance obligations for transparency and accountability.
ePrivacy Regulation	Expected similar to GDPR (up to €20 million or 4% of turnover)	Complements GDPR, focusing on electronic communications privacy.

Vision et mission de l’entreprise

Vision de l’entreprise

- Responsabiliser les entreprises
- Converger l’IA et le cyberespace
- Inspirer la confiance
- Conduire la transformation

Mission de l’entreprise

- Innover avec l’IA
- Systèmes d’IA sécurisés
- Favorisez la réussite des clients
- Favoriser l’innovation continue

Information

Contactez-nous

Mieux vaut fortifier votre château avant le siège que de le reconstruire après l’effondrement des murs : assurez votre avenir numérique avec nous dès aujourd’hui.

+32 471 35 63 06

www.shieldai.ai

mustapha.annouh@shieldai.ai

Brussels, Belgium

Mustapha Annouh
CEO SHIELD AI

Merci

pour votre attention !