GP'SUP

Les risques et opportunités de l'IA en Santé Sécurité au Travail

« Entre progrès et perte de contrôle »

cerfos
management des risques professionnels

« L'intelligence Artificielle » c'est quoi ?

Définition :

- **CNIL** : L'intelligence artificielle est un **procédé logique et automatisé** reposant sur un algorithme et en mesure de réaliser des tâches bien définies.

- **Pour le Parlement européen**, constitue une intelligence artificielle tout outil utilisé par une **machine afin de « reproduire des comportements liés aux humains,** tels que le raisonnement, la planification et la créativité ».

Alan TURING – 1950

« Les machines peuvent-elles penser ? »

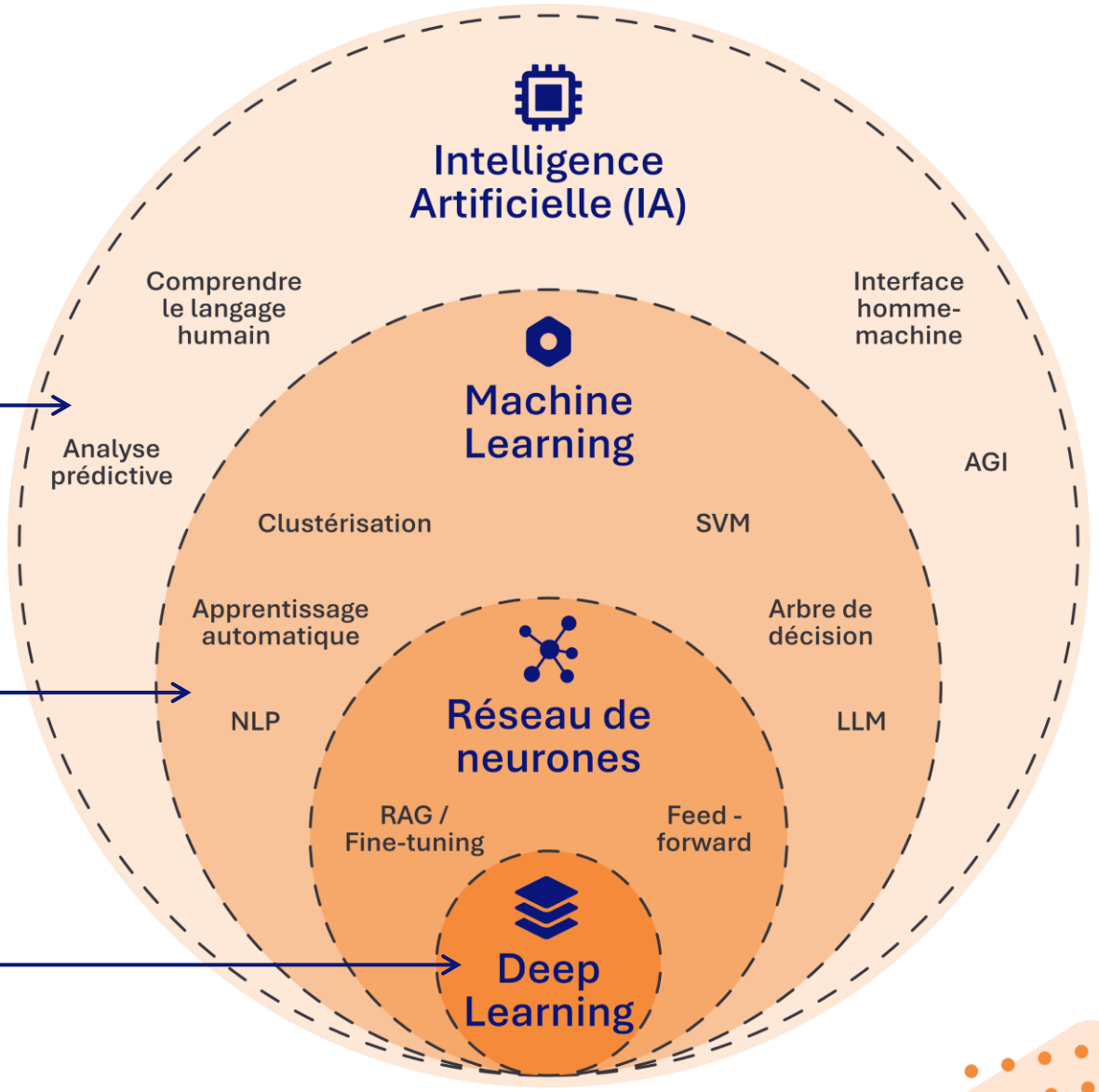
Schéma de fonctionnement d'une IA

Technologie utilisée :

Intelligence artificielle :
Définit des caractéristiques et les intègre dans un programme.

Machine Learning :
Reproduit un comportement grâce à des algorithmes alimentés par une grande quantité de données.

Deep Learning :
Compréhension non linéaire, des concepts avec une grandes précisions.



Méthode d'apprentissage de l'IA

Apprentissage supervisé

Apprentissage non-supervisé

Apprentissage auto-supervisé

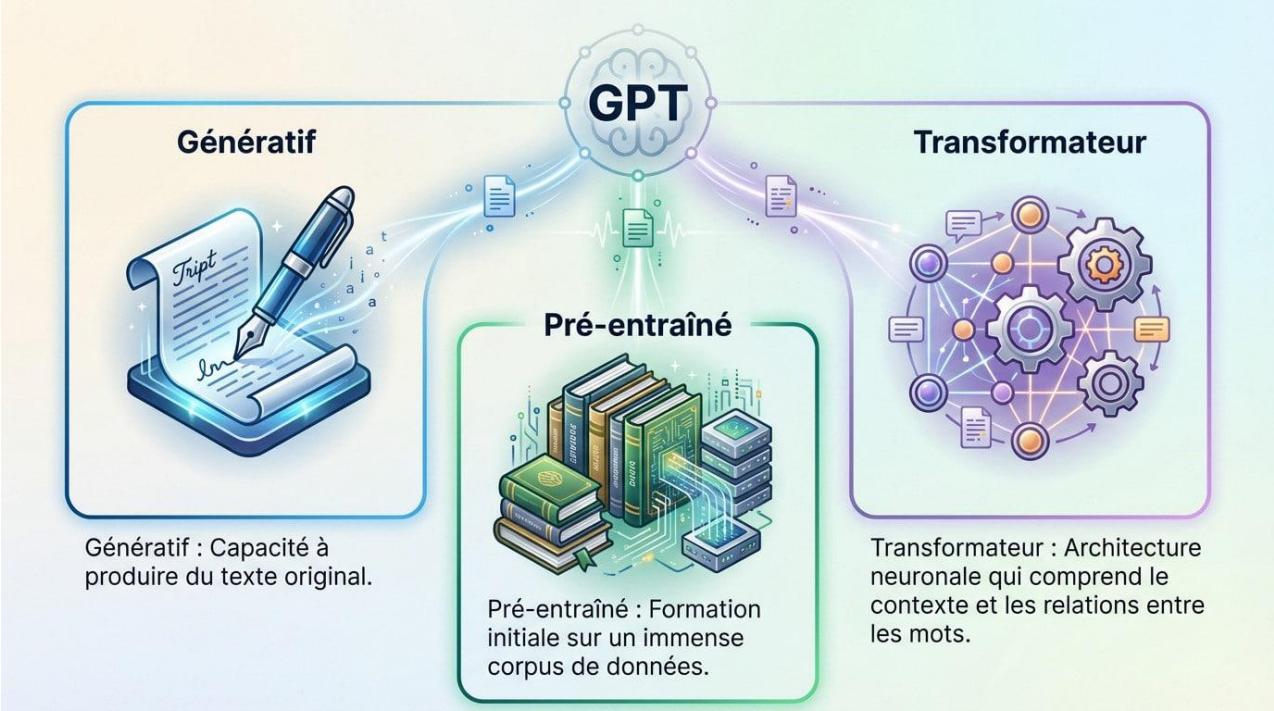
Apprentissage par renforcement

Méthode d'apprentissage de l'IA

Apprentissage auto-supervisé

Afin de mieux comprendre nous allons étudier le cas de ChatGPT qui est l'IA la plus documentée à ce jour.

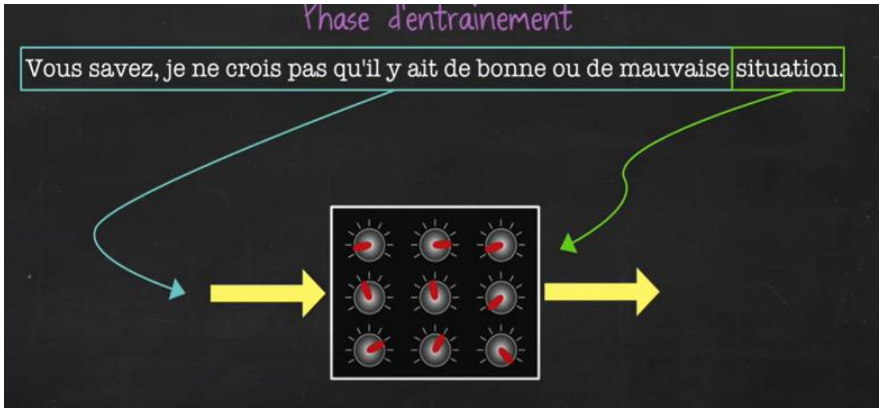
CHATGPT : c'est une IA « GPT - Generative Pretrained Transformer »

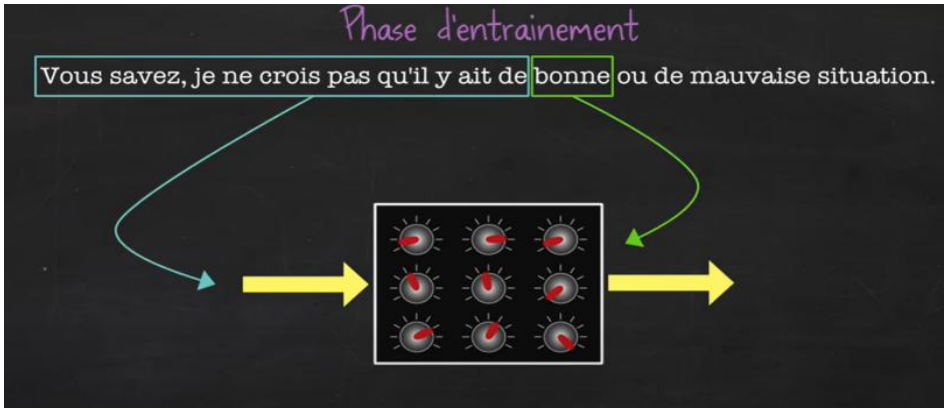


Fonctionnement IA « générative »

Apprentissage auto-supervisé

Ex : Les IA « GPT » utilisent les tokens comme unités de lecture et d'écriture.



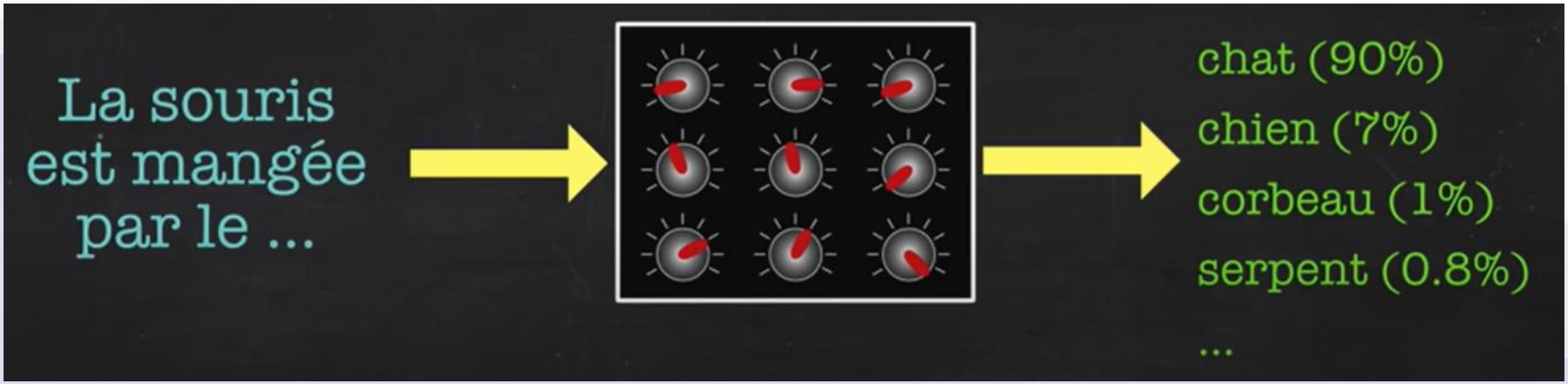


Plus besoin d'un processus d'annotation spécifique avec un humain. On peut simplement prendre des tonnes de textes pour générer des milliards d'exemples.

Fonctionnement IA « générative »

INFORMATIONS DE SORTIE – Probabilité et justesse

Avec l'IA, quand nous cherchons à prédire le prochain mot d'un texte, il n'y a pas qu'une seule réponse possible mais plutôt une liste de mots avec une probabilité.



Quand nous lui demandons quelque chose de connu et logique, elle apportera la réponse la plus probable et par conséquent la plus juste (selon l'IA). Cependant, elle ne connaît pas le principe de vérité.

Fonctionnement IA « générative »

INFORMATIONS DE SORTIE – Probabilité et justesse

Si nous lui demandons de compléter la phrase suivante : « Christophe Colomb a découvert l'Amérique en … »

L'IA répondra automatiquement « 1492 ». Car les mots « Christophe Colomb », « découvert » et « Amérique » sont souvent associés avec « 1492 ».

Fonctionnement IA « générative »

INFORMATIONS DE SORTIE – Probabilité et justesse

En revanche si nous réalisons la même manipulation en changeant le prénom (donc en apportant un élément faux).

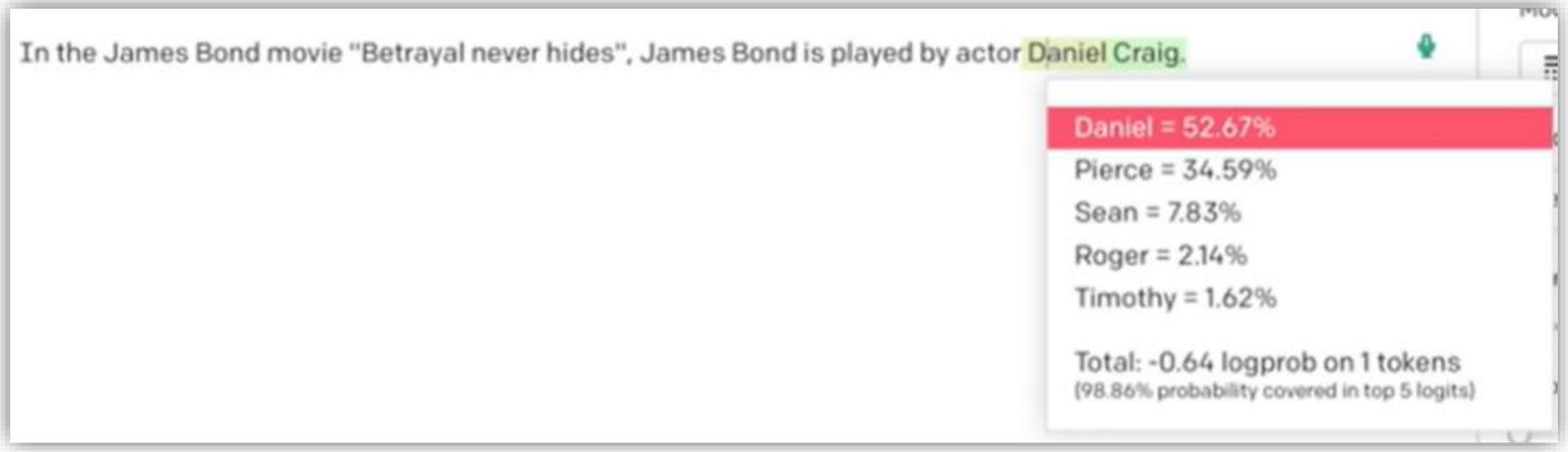
L'IA ne cherchera pas à détecter l'information fausse mais complétera naturellement avec l'information la plus probable. Donc apportera un élément de réponse faux.

Fonctionnement IA « générative »

INFORMATIONS DE SORTIE – Probabilité et justesse

Plus simplement, si ChatGPT reçoit une information inventée ou fausse, il peut quand même continuer à répondre comme si cette information était vraie.

Ex d'un film inexistant : *Dans le film de James Bond « La trahison ne se cache jamais », L'acteur qui joue James Bond s'appelle… « Daniel Craig »*



Phénomène du « shadow IA »

Définition : Utilisation de l'intelligence artificielle en entreprise en l'absence de cadre formel.

- ***2025 :*** *68 % des salariés français déclaraient utiliser une IA générative sans encadrement formel*

Selon les données transmises à une intelligence artificielle, **des obligations légales peuvent s'appliquer, notamment en matière de protection des données personnelles.**

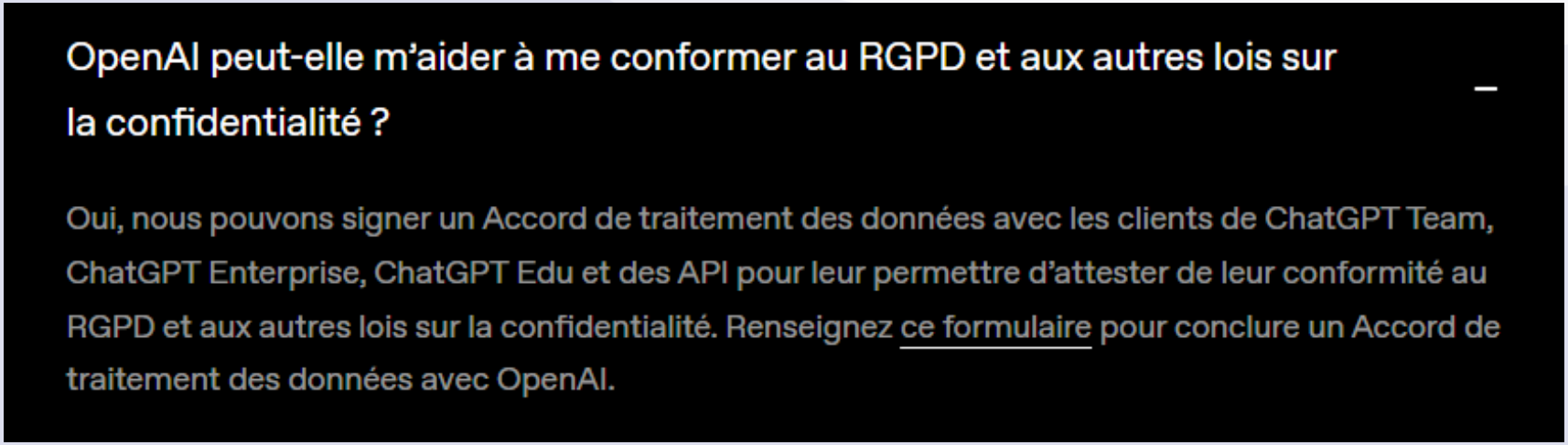
NOUS NE POUVONS PAS DIRE N'IMPORTE QUOI A UNE IA

Phénomène du « shadow IA »

Exemple : Un commercial utilise volontairement une IA générative pour gagner du temps dans le traitement des données. Il lui transmet des données concernant ses clients, ses prospects, ses innovations, ses projets ou même sa stratégie commerciale, sans qu'un accord de confidentialité ne soit formalisé, cela constitue un premier écart au regard du RGPD.

La société s'expose alors à…

Des risques de violation du RGPD, de divulgation de secrets d'affaires



Le cadre règlementaire

AI Act européen

Nouvelle législation classifiant les risques liés à l'IA, exigeant une mise en conformité rapide

RGPD

Évolutions des exigences en matière de protection des données personnelles pour l'IA

Code du travail

Adaptations nécessaires concernant l'impact de l'IA sur les conditions de travail et la gestion du personnel

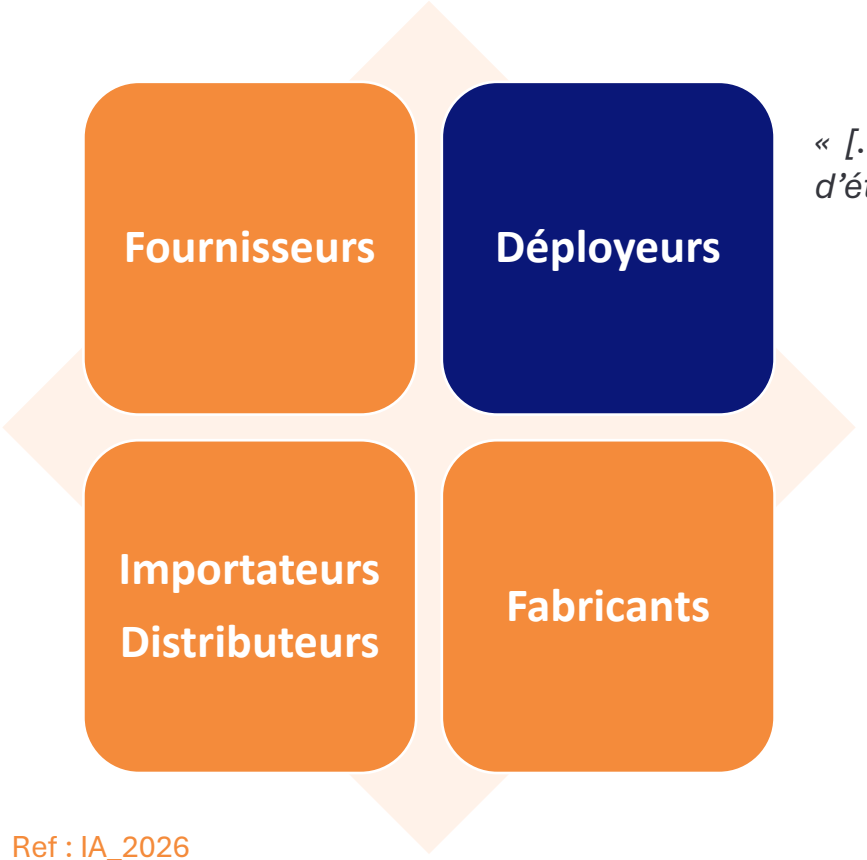
CNIL

La commission apporte des recommandations spécifiques pour l'IA

Règlementation : IA Act

Définition SIA : *"système d'IA", basé sur une machine qui est conçue pour fonctionner avec différents niveaux d'autonomie (…) et qui, déduit, à partir des données qu'il reçoit des résultats tels que des prédictions, du contenu, des recommandations ou des décisions.*

Le présent règlement s'applique :

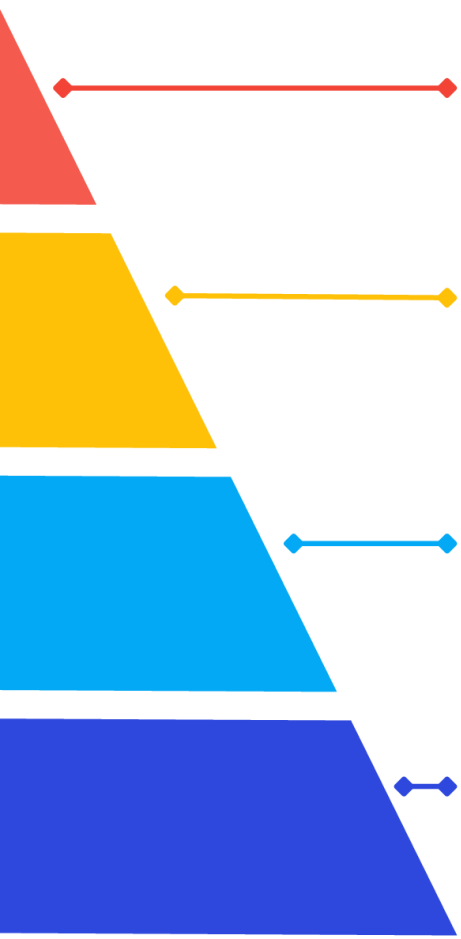


« […] qui déploie des systèmes d'IA qui ont leur lieu d'établissement ou sont situés dans l'Union. »

En bref, toute personne physique ou morale, une autorité publique, une agence ou un autre organisme utilisant un système d'IA sous son autorité.

IA ACT : Chap1/Art3 – Définitions
IA ACT : Chap1/Art2 – Champ d'application

Règlementation : IA Act



Risques inacceptables

- Manipulation des comportements
- Exploitation des vulnérabilités des individus
- Crédit social

Hauts risques

- La santé
- La justice / RH
- L'éducation

IA avec des obligations de transparence

- Les chatbots
- Les deepfakes
- Les IA génératives (ChatGPT)

Risques minimes ou inexistant

- Jeux-vidéos
- Filtres à spam

PARTIE 3

Risques, dérives et leurs effets

Identifier • Analyser • Sensibiliser

Repérer les risques liés à l'usage de l'intelligence artificielle afin d'anticiper les dérives et d'en limiter les effets négatifs.

Les risques professionnels liés à l'IA

Contrôle algorithmique

Surveillance, décisions discriminantes...

Intégrité des personnes

Deepfake*, harcélement, désinformation...

Altération des capacités humaines

Surconfiance en l'outil, surcharge mentale...

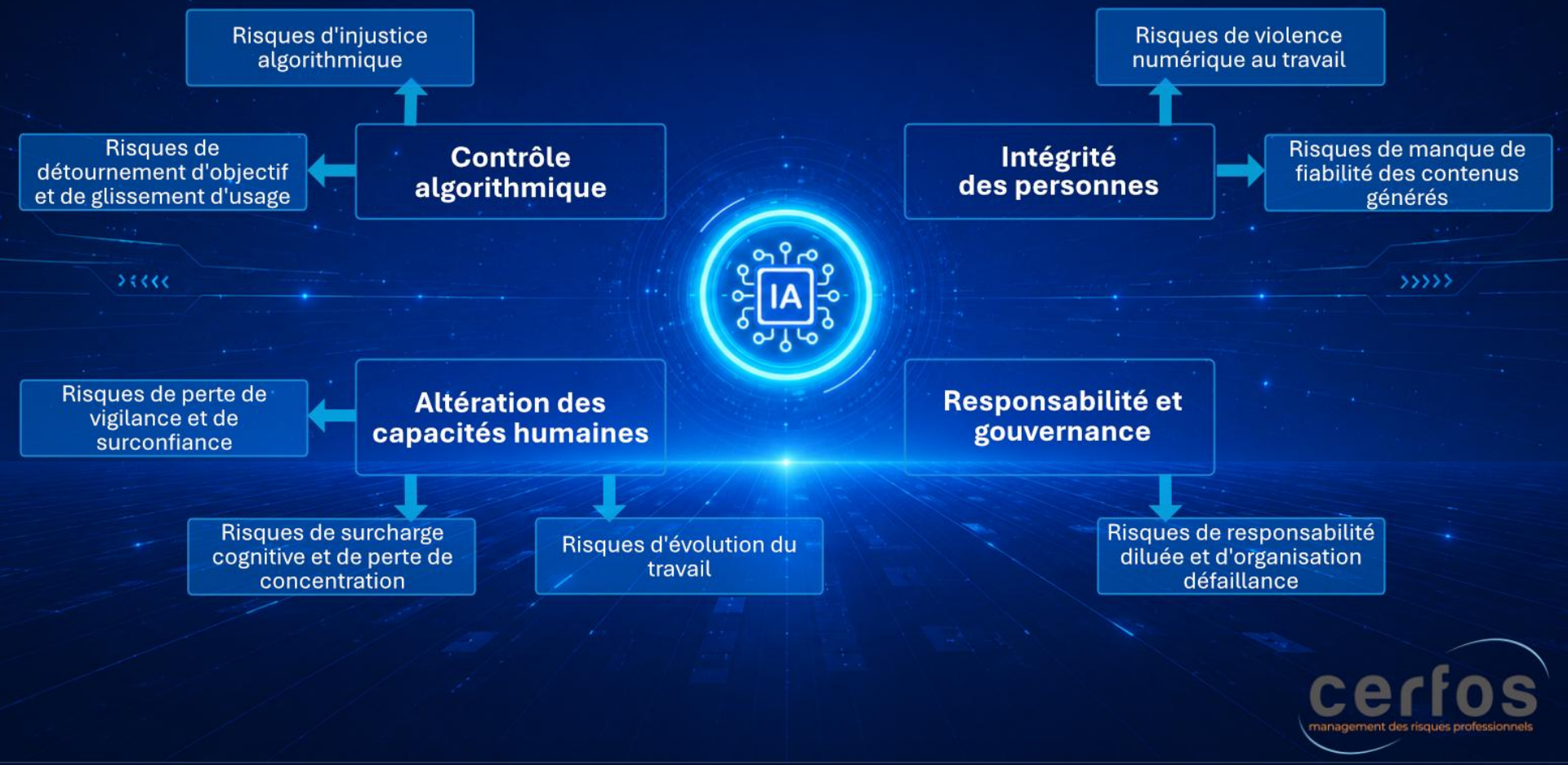
Responsabilité et gouvernance

Responsabilité floue, dépendance à l'outil...

* Usages détournés de l'IA, comme la création d'images portant atteinte à la dignité des personnes au travail.

cerfos
management des risques professionnels

Les risques professionnels liés à l'IA



cerfos
management des risques professionnels

Contrôle algorithmique

Surveillance • Décisions automatisées • Glissement d’usage

Contrôle algorithmique

Risques de détournement d'objectif et de glissement d'usage

Cas connu : France, Amazon France Logistique, 27 décembre 2023 — L'entreprise utilisait un système automatisé de suivi de l'activité des salariés dans ses entrepôts à partir des données issues des scanners. Le dispositif mesurait les temps entre deux scans, les périodes d'« inactivité » et des indicateurs de performance, parfois à la seconde.

Sanction : la CNIL a prononcé une amende administrative de 32 millions d'euros pour surveillance excessive et disproportionnée des salariés, ramené à 4 millions d'euros par le juge des référés du Conseil d'État.

Risques d'injustice algorithmique

Cas connu : Italie, Deliveroo Italy, 31 décembre 2020 — L'algorithme "Frank" pénalisait automatiquement les livreurs absents à des créneaux réservés, même en cas de grève, maladie, maternité ou handicap. Ce scoring réduisait ensuite leur accès aux futurs créneaux de travail.

Sanction : le Tribunal de Bologne a reconnu un effet discriminatoire, ordonné la modification du système et condamné Deliveroo au paiement des frais.

Intégrité des personnes

Deepfakes • Harcèlement numérique • Désinformation • Fausses preuves

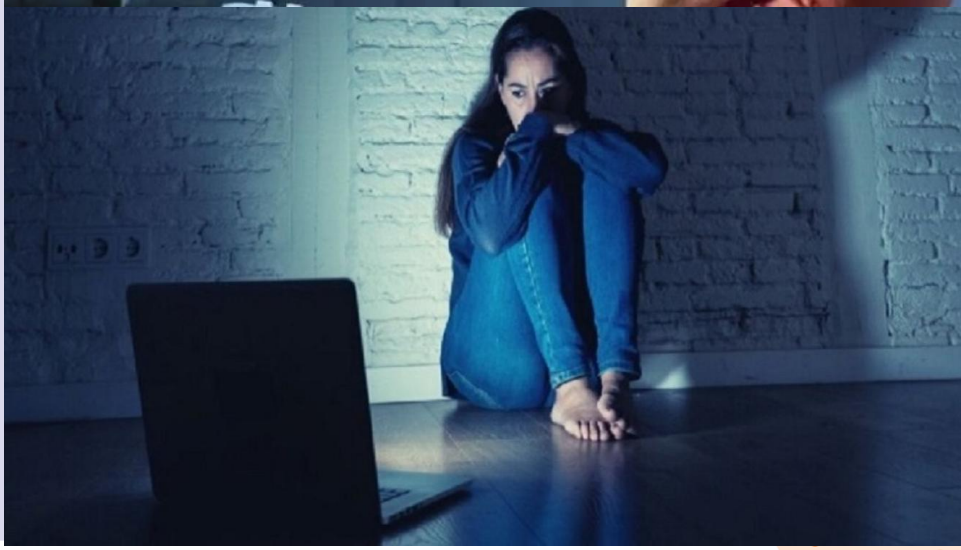
Intégrité des personnes

Risques de violence numérique au travail

Cas connu : Australie, fonction publique de Canberra, mars 2025 — Un agent public aurait créé plus de 100 images deepfake à caractère sexuel visant au moins 16 collègues, à partir de photos récupérées notamment sur les réseaux sociaux.

Cas connu : France, commerce de détail à L'Aigle, à partir de novembre 2024 — Une commerçante découvre que son visage a été utilisé dans des montages et une vidéo pornographique deepfake diffusés en ligne, avec des répercussions directes dans son activité professionnelle.

Sanction : Aucune poursuite pénale engagée faute de preuve de diffusion ; mesures disciplinaires non documentées publiquement…



Intégrité des personnes

Risques de manque de fiabilité des contenus générés

- Références réglementaires ou juridiques fictives

Cas connu : Belgique, Cour d'appel de Gand, 18 septembre 2025 — Des conclusions d'avocat contenaient de mauvaises citations d'articles existants et des références à de la jurisprudence ou doctrine introuvables, vraissemblablement issues d'un usage non contrôlé de l'IA. La Cour a souligné le manque de vigilance dans l'usage de l'IA pour des écritures juridiques.

Sanction : la Cour indique qu'elle ne tient "en aucune façon" compte des sources fictives/introuvables dans l'examen du litige (écartement des références). Aucune amende n'est mentionnée pour le moment…

Altération des capacités

Perte de vigilance • Surcharge cognitive • Désapprentissage • Transformation du travail

Altération des capacités

Risques de perte de vigilance et de surconfiance

- « L'IA a dit que » Inhibition du jugement

Définition : Risque de faire une confiance excessive à un résultat produit par l'IA, au point de réduire la vérification humaine.

Cas connu : États-Unis, Detroit Police Department, janvier 2020 / juin 2024 — Robert Williams a été arrêté à tort après un mauvais résultat de reconnaissance faciale utilisé comme élément central d'identification. Il a été détenu environ 30 heures avant d'être libéré.

Sanction : règlement amiable en 2024, avec renforcement des règles d'usage de la reconnaissance faciale par la police de Detroit afin de limiter les erreurs et arrestations injustifiées.

Altération des capacités

Risques d'évolution du travail

- Insécurité professionnelle

Définition : Risque que l'IA soit perçue comme une menace pour l'emploi, les revenus, l'image ou la valeur d'un métier, notamment lorsqu'elle peut produire, imiter ou remplacer une partie du travail humain.

Cas connu : États-Unis, WGA & SAG-AFTRA, 2023 — Les grèves des scénaristes et acteurs ont abouti à des accords encadrant l'usage de l'IA : transparence sur les contenus générés, impossibilité d'imposer l'IA aux auteurs, consentement et compensation pour les répliques numériques des acteurs.

Sanction : aucune sanction administrative

Gouvernance et responsabilité

Défaillance organisationnelle

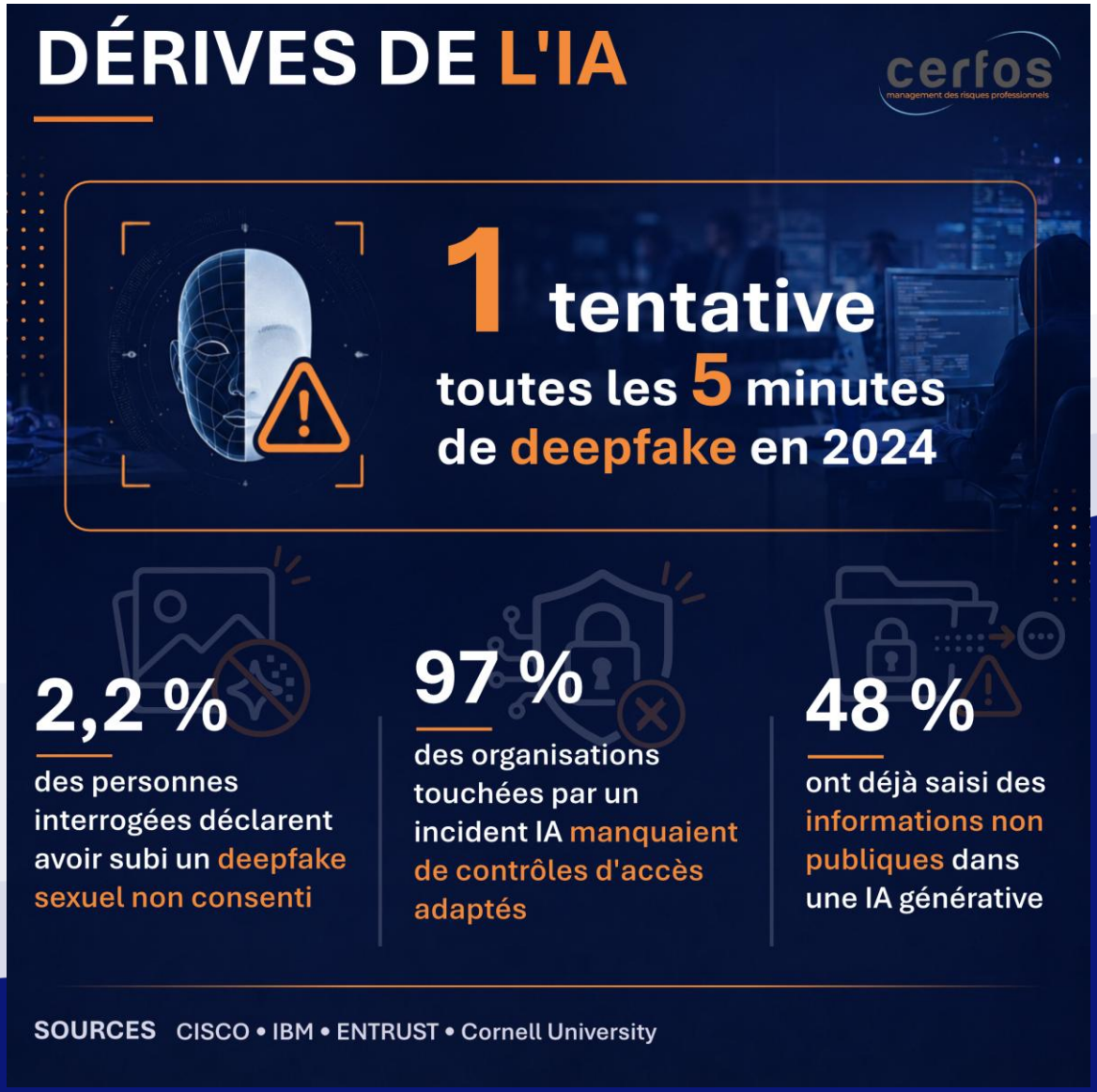
Risques de responsabilité diluée et d'organisation défaillante

- Personne n'assume l'erreur ("c'est l'IA")

Cas connu : États-Unis, Tesla, crash de Key Largo en 2019 / décision confirmée le 20 février 2026 — Une Tesla Model S en mode Autopilot a percuté un SUV arrêté, tuant une jeune femme et blessant gravement son compagnon. Le jury a retenu une responsabilité partagée entre le conducteur et Tesla, notamment sur la conception, les limites d'usage et les avertissements liés au système.

Sanction : condamnation civile de 243 millions de dollars maintenue par un juge fédéral, dont 200 millions de dollars de dommages punitifs ; Tesla a indiqué vouloir faire appel

Quelques chiffres clés :



Quelques chiffres clés :

Différents usages de l'IA en SST :

Code du travail, art. L.4121-1 : obligation de protéger la santé physique et mentale des travailleurs.

Intégration

Automatisation via une application tiers

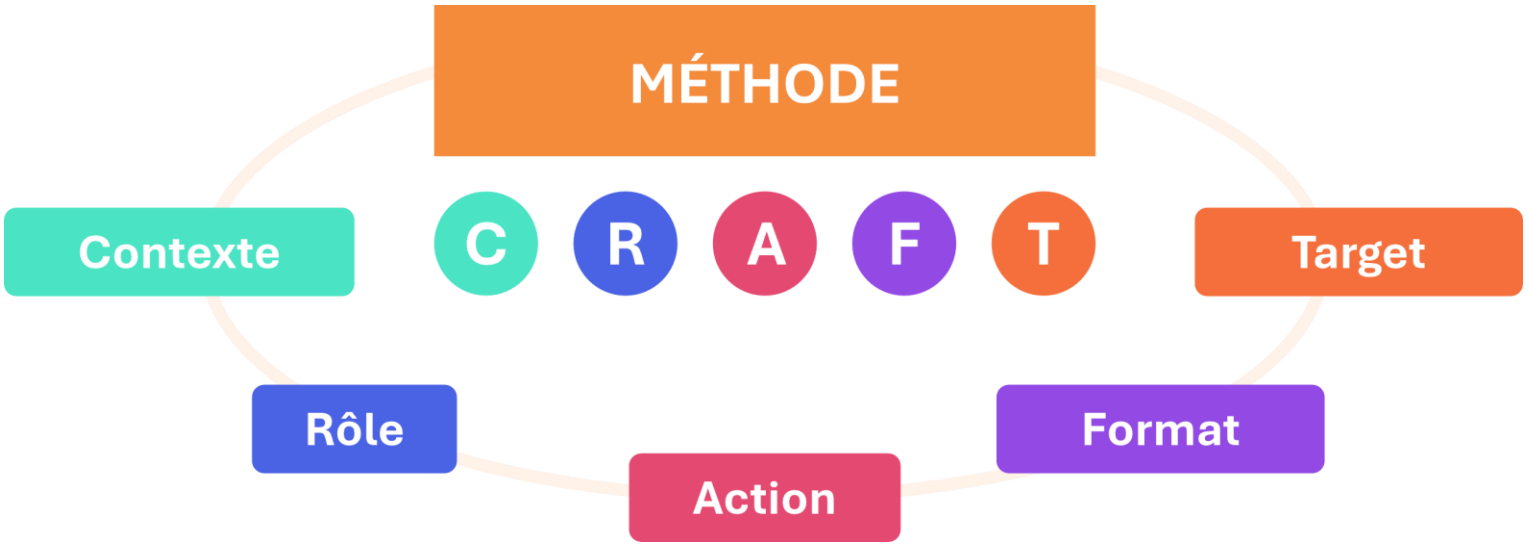
Utilisation

Aide à la réalisation, structuration, reformulation de tâches

L’art du bon prompt – IA :

Définition : Un **prompt (un brief)** est une **consigne donnée à une IA** pour lui expliquer ce que l’on attend d’elle. Il peut prendre la forme d’une question, d’un ordre, d’un texte à analyser ou d’une demande structurée.

La qualité de la réponse dépend de la qualité du prompt : plus la demande est claire, contextualisée et précise, plus la réponse de l’IA sera exploitable.

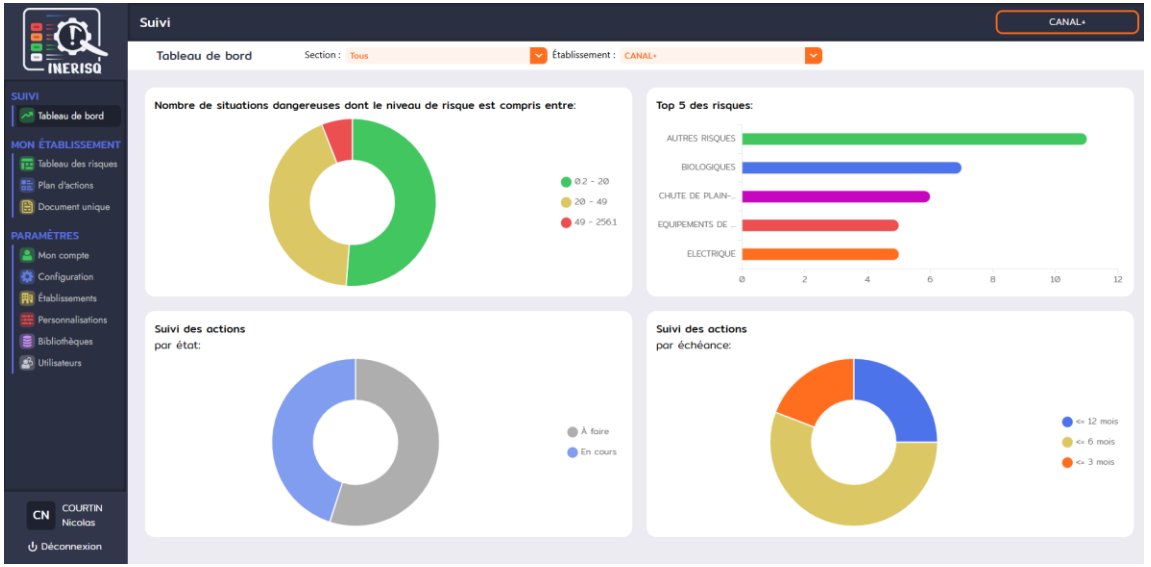


Document unique (DUERP) – IA :

L'IA ne remplace pas l'expertise terrain : elle aide à observer, structurer, analyser et proposer.

Rôles :

- **Reformuler** les dangers, **identifier** les risques associés à des situations complexes (RPS), **proposer des actions réalistes** selon les moyens : humains, organisationnels, techniques, financiers.



Réunion / Compte rendu – IA :

L’IA peut accompagner les réunions de prévention en facilitant leur préparation, la structuration des échanges et le suivi des décisions, afin de transformer plus rapidement les constats en actions concrètes.

NotebookLM

Read AI

Communication / Sensibilisation – IA :

L'IA peut renforcer la communication et la sensibilisation en prévention en aidant à vulgariser les messages, adapter les supports aux publics concernés et créer des contenus plus clairs, attractifs et faciles à retenir.

Leonardo.Ai

Nano Banana PRO

Canva

SUNO

Midjourney

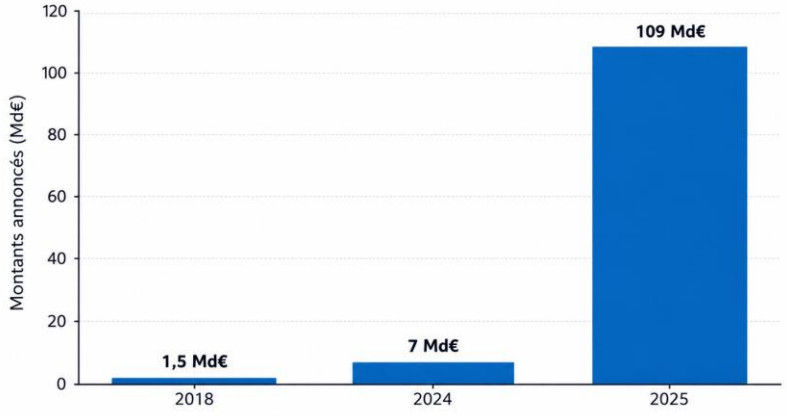
Système de détection des anomalies – IA :

L'IA peut renforcer la détection des anomalies en prévention en analysant de grandes quantités de données, en repérant des situations inhabituelles ou à risque, et en aidant les équipes à intervenir plus rapidement avant qu'un incident ne survienne.

Et demain..?

Investissements / engagements IA en France

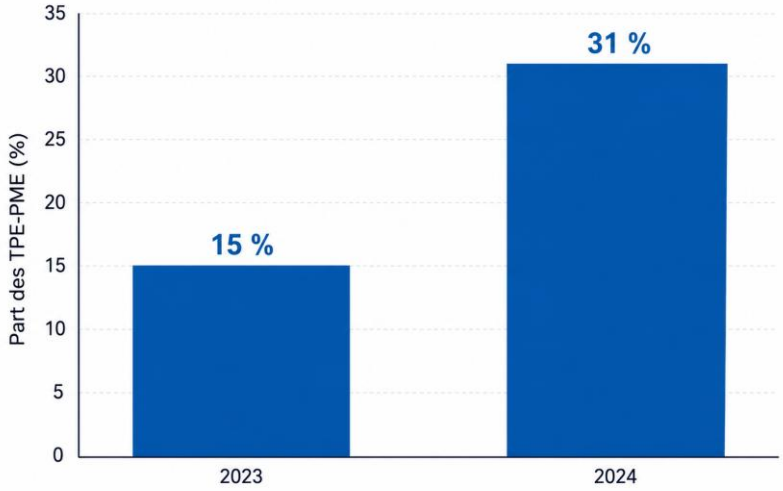
Jalons d'investissements annoncés



2018 : stratégie nationale IA ; 2024 : annonces Choose France IA & data centers ;
2025 : annonces du Sommet pour l'action sur l'IA.

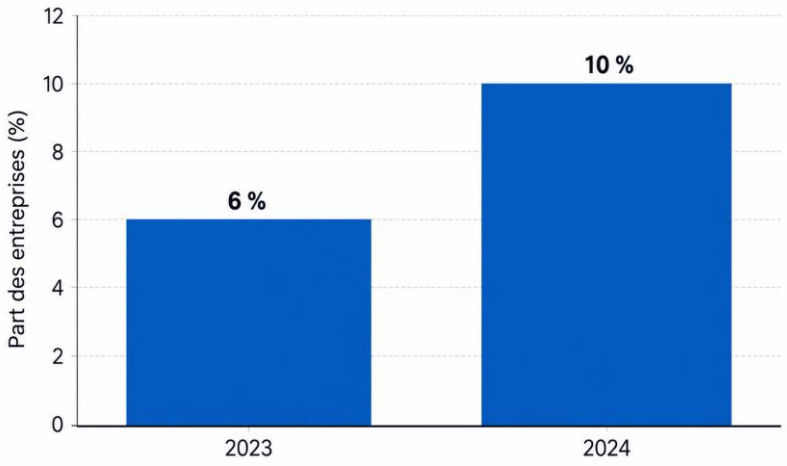
Sources : Élysée

TPE-PME françaises utilisant l'IA générative



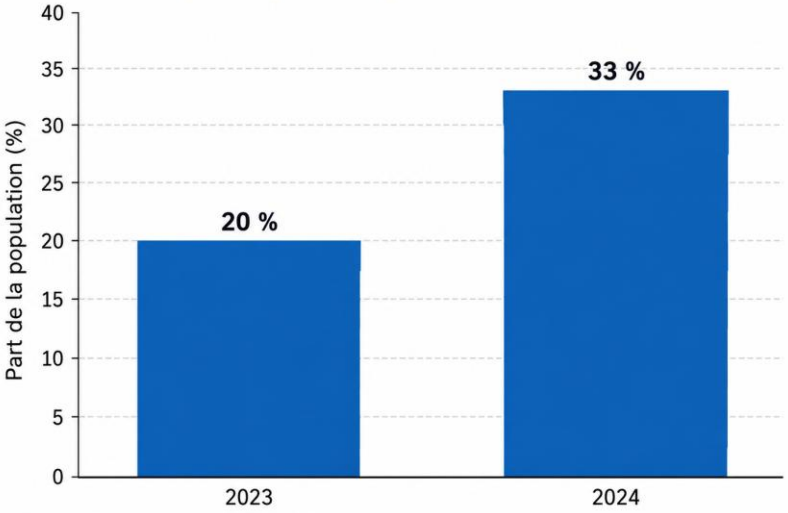
Source : Bpifrance Le Lab / Bpifrance, 2025

Entreprises françaises utilisant au moins une technologie d'IA



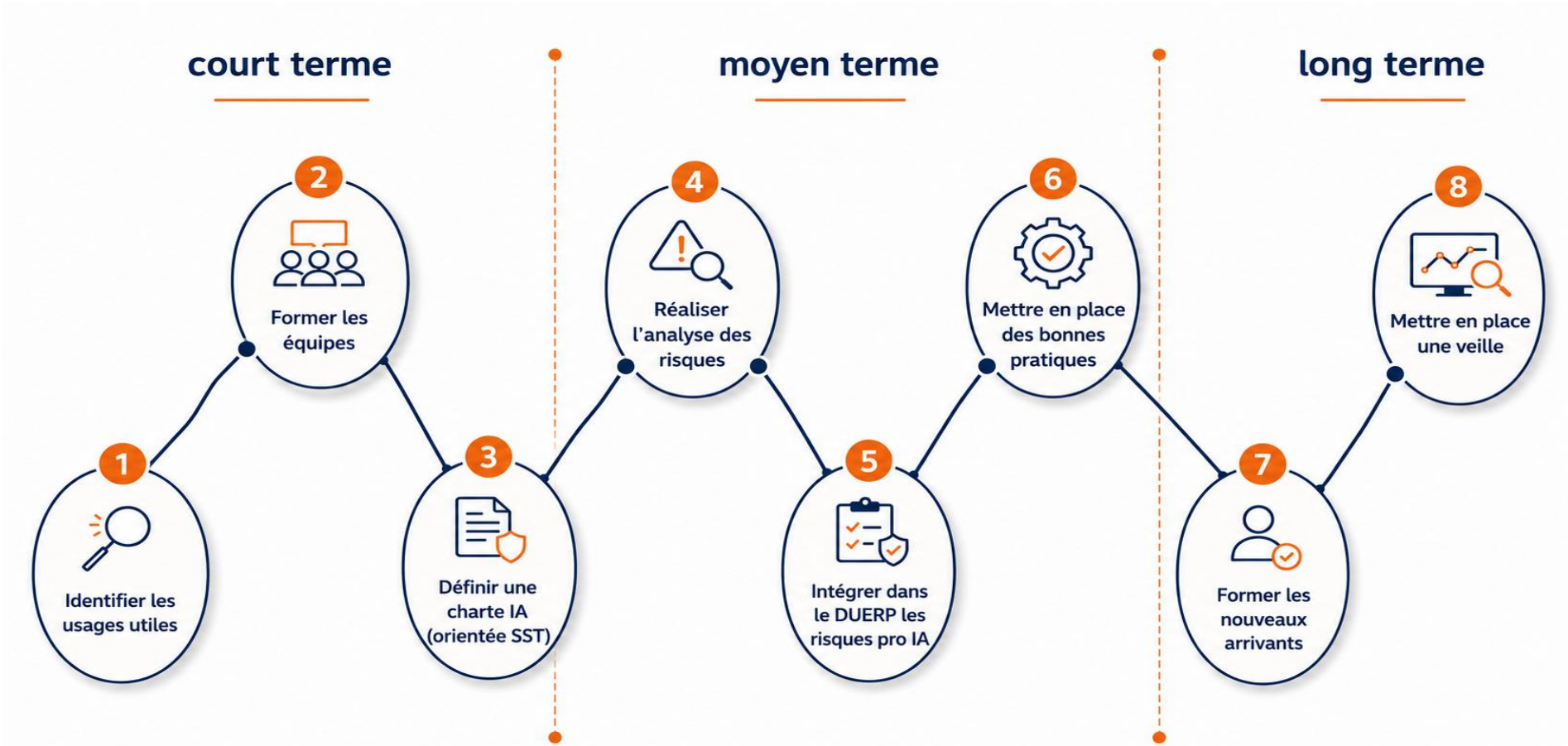
Source : Insee, enquête TIC entreprises 2024

Français ayant déjà utilisé un outil d'IA



Source : Baromètre du numérique 2024

Plan d’action :




cerfos
management des risques professionnels