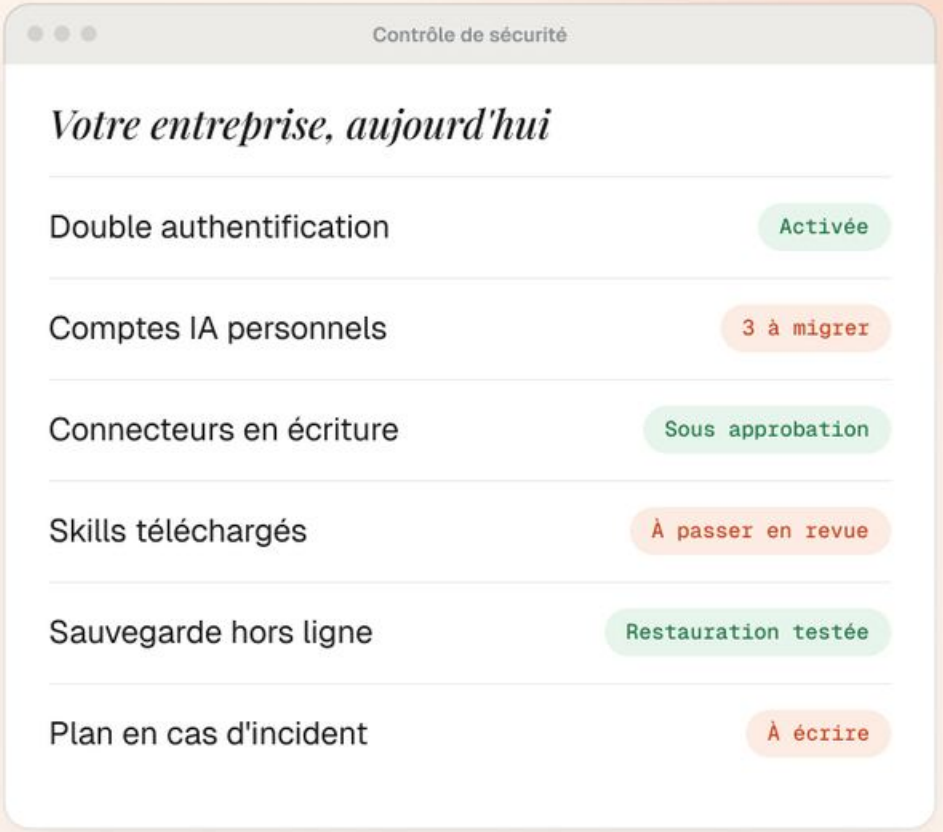
GUIDE DE TERRAIN · ÉDITION NOVEMBRE 2026

Cybersécurité : *ce que l'IA change pour les PME*

Les nouveaux risques, les réglages à faire, les bons réflexes, et quoi faire le jour où ça arrive.



+50 CLIENTS · +80 PROJETS

Colin Dargent
Responsable IA externalisé pour PME

OUVERTURE AVANT DE COMMENCER

Vos équipes utilisent déjà l'IA. *Votre sécurité a-t-elle suivi ?*

Un assistant branché sur la messagerie, un skill téléchargé en deux clics, un outil monté en une après-midi par un salarié motivé. **En deux ans, l'IA a ouvert des portes que personne n'a inventoriées.**

Les attaquants s'en servent aussi : un mail sans faute, une voix imitée, une page de connexion copiée par un outil d'IA.

Ce guide dit ce qui est avéré, ce qui reste surtout théorique, et les gestes à faire, dans l'ordre. **Sans jargon, et sans vous faire peur pour rien.**

Écrit depuis *les projets*

Je conçois et je mets en place des systèmes d'IA dans des PME de 15 à 200 personnes.

Je ne suis pas spécialiste de la cybersécurité. Je suis celui qui branche l'IA dans les outils de l'entreprise, et je vois ce que ça ouvre.

Pour ce guide, je suis donc allé lire l'ANSSI, la CNIL, la documentation des éditeurs et les études. **Chaque chiffre est sourcé en fin de guide.** Quand une source vend de la sécurité, je le signale.

01 / CLIENTS

+50 clients accompagnés

02 / PROJETS

+80 projets menés

03 / TERRAIN

2 ans de terrain en PME

Six parties, *du risque au réflexe*

01 / RISQUES

Les nouveaux risques

Ce qui change côté attaquants, et ce que vos usages ouvrent.

02 / RÉGLAGES

Paramétrer vos outils IA pour rester en sécurité

Claude, ChatGPT et Dust : les options à activer, et leurs limites.

03 / PRATIQUES

Les process à mettre en place

Le socle, les mises à jour, la routine de contrôle, la veille, le message à l'équipe.

04 / RÉFLEXES

Les réflexes à développer chez vos salariés

Les gestes de chacun, et les compétences à acquérir.

05 / PRESTATAIRES

Les questions à poser à votre infogérant

Ses accès, ses outils d'IA, ses sauvegardes, la sortie du contrat.

06 / INCIDENT

Que faire en cas d'attaque

Quoi couper, qui appeler, quoi déclarer, dans quel ordre.

OUVERTURE | POUR QUI

Ce guide est pour vous *si vous vous reconnaissez*

Oui / POUR VOUS

Vous dirigez ou vous pilotez une PME dont les équipes utilisent déjà l'IA.

Vous n'avez pas de responsable sécurité, ou un prestataire informatique gère tout.

Vous voulez savoir quoi régler et quoi demander, sans devenir expert.

Non / PAS POUR VOUS

Vous avez déjà un responsable sécurité et une politique IA en place : ce guide vous paraîtra basique.

Vous cherchez à être en règle avec l'AI Act ou le RGPD : la conformité est un autre sujet, ce guide porte sur la sécurité.

Vous cherchez un audit de vos outils : un guide ne remplace pas un regard sur votre cas.

OUVERTURE L'ESSENTIEL

Si vous ne faites que cinq choses, *faites celles-là*

1 **Activez la double authentification**
sur la messagerie, la banque et les comptes IA. PARTIE 3

2 **Donnez un gestionnaire de mots de passe à toute l'équipe**
pour ne plus jamais réutiliser un mot de passe. PARTIE 3

3 **Passez les usages professionnels sur des comptes d'entreprise**
pour sortir vos données des abonnements personnels. PARTIE 2

4 **Mettez les actions d'écriture des connecteurs sous approbation**
envoyer, modifier, supprimer : un humain valide. PARTIE 2

5 **Testez une restauration de sauvegarde**
une sauvegarde jamais restaurée reste une hypothèse. PARTIE 3

Ce qui change, *et ce qui reste*

L'IA fait bouger deux choses en même temps : les attaquants vont plus vite, et vos propres usages ouvrent de nouvelles portes. Les vieilles portes, elles, restent ouvertes.

01 / CIBLE

« Je suis trop petit pour intéresser. »

02 / VITESSE

« L'IA change tout aux attaques. »

03 / ARNAQUES

« Je reconnaîtrais la voix de mon patron. »

04 / USAGES

« Mes équipes font attention. »

05 / OUTILS

« C'est juste un petit outil interne. »

Vous êtes déjà *une cible*

48 %

des victimes de rançongiciel signalées à l'ANSSI en 2025 sont des PME, des TPE ou des ETI. **Elles étaient 37 % en 2024.**

POURQUOI VOUS

L'ANSSI décrit des attaquants qui visent **indistinctement** la plupart des secteurs.

Ils ne vous ont pas choisi. Ils cherchent une porte ouverte, et si la vôtre l'est, vous êtes sur la liste.

La question « qui voudrait m'attaquer ? » ne protège donc de rien. La bonne question est : « quelle porte ai-je laissée ouverte ? »

Source : ANSSI, Panorama de la cybermenace 2025, sur 128 cas de rançongiciel portés à sa connaissance, non représentatifs de l'ensemble.

L'IA accélère *trois choses*

01 / CRÉDIBLE

De 12 à 54 %

Taux de clic d'un mail d'hameçonnage : 12 % pour un mail générique, 54 % pour un mail écrit par IA, autant que pour un expert humain.

Étude sur 101 personnes seulement : à lire comme un ordre de grandeur.

02 / ACCESSIBLE

400 à 1 200 $

Le prix d'un rançongiciel vendu par un auteur aux compétences de code basiques, qui s'appuyait sur une IA.

Cas documenté par Anthropic, l'éditeur de Claude.

03 / RAPIDE

De mois à heures

Selon Verizon, l'IA réduit le temps dont vous disposez pour corriger une faille avant qu'elle soit exploitée.

Verizon vend de la sécurité et ne donne pas de délai chiffré.

POUR VOUS

« Ce mail a l'air faux » ne protège plus. Le réflexe doit porter sur la demande elle-même : est-elle inhabituelle ? Demande-t-elle un code, un virement, un accès ?

Source : Heiding et al., arXiv 2412.00586 ; Anthropic, rapport du 27/08/2025 ; Verizon, DBIR 2026.

Accélérer les attaques *ne les rend pas nouvelles*

CE QUE MESURE LE FBI

≈ 4 %

des pertes déclarées au FBI en 2025 mentionnent l'IA : 893 M$ sur 20,9 Md$. Ce sont des plaintes américaines, et le plaignant ne sait pas toujours que l'IA a servi.

CE QUE DIT L'ANSSI

« Potentiel accélérateur »

L'agence française juge qu'il n'est pas pertinent de parler de changement de paradigme. Elle a traité 1 366 incidents en 2025, contre 1 361 en 2024. Aucune IA n'a encore mené seule toutes les étapes d'une attaque.

CE QUE J'EN RETIENS

Pour une PME, la conclusion est plutôt bonne. Le mot de passe réutilisé, le mail piégé et la faille non corrigée restent les portes d'entrée. Ce sont aussi celles qui se ferment le plus facilement : c'est l'objet de la partie 3.

Source : FBI IC3, Internet Crime Report 2025, calcul de l'auteur ; ANSSI, CERTFR-2026-CTI-001 et CERTFR-2026-CTI-002.

Voir et entendre *ne suffit plus*

01 / VOIX ET VISIO

25,6 M$

Le montant transféré en 15 virements par un salarié d'Arup, après une visio où ses dirigeants étaient des deepfakes.

Rappelez sur un numéro que vous connaissez déjà.

02 / MAIL

Sans faute

Les fautes ne signalent plus rien. Le mail peut citer votre prestataire, votre projet, le prénom de votre comptable.

Vérifiez toute demande inhabituelle par un autre canal.

03 / SITE

À l'identique

Okta a observé des attaquants qui copient des pages de connexion avec v0, un outil d'IA qui crée des sites sur simple demande.

Lisez le nom de domaine avant de saisir quoi que ce soit.

EN FRANCE

Fraude au virement +170 %, usurpation de numéro +517 % en 2025 sur Cybermalveillance.gouv.fr. La plateforme ne relie pas ces hausses à l'IA.

Source : CNN, 16/05/2024 (Arup) ; Okta, 01/07/2025 (éditeur de sécurité) ; Cybermalveillance.gouv.fr, bilan 2025, particuliers et pros mêlés.

Six portes ouvertes *par vos propres usages*

CLASSÉES PAR CE QUE JE VOIS PESER LE PLUS

1	**Données collées dans un chat** Contrats, fichiers clients, chiffres : ils partent chez l'éditeur.	Quotidien
2	**Comptes personnels** Hors de tout contrat, de tout réglage, de tout départ géré.	Quotidien
3	**Mots de passe et clés d'accès** Réutilisés, partagés, ou écrits en clair dans un chat ou du code.	Quotidien
4	**Outils construits en interne** Mis en service sans relecture, parfois ouverts à tout Internet.	Avéré
5	**Skills et plugins téléchargés** Des instructions et du code que votre IA exécute avec vos droits.	Avéré, surtout ailleurs
6	**Connecteurs et injection de prompt** Un mail piégé qui donne des ordres à votre assistant.	Surtout démontré

La fuite la plus probable *se passe d'attaquant*

01 / COLLAGE

77 %

des employés collent des données dans une IA, d'après la télémétrie navigateur de LayerX.

02 / COMPTE PERSO

82 %

de ces collages se font depuis un compte personnel, selon la même source.

03 / VIOLATION

20 %

des organisations étudiées par IBM ont subi une violation liée à l'IA non encadrée.

CE QUE ÇA VEUT DIRE

Sur un abonnement personnel, vos données suivent les conditions grand public : pas d'accord de traitement des données, et un entraînement possible selon l'outil et le réglage (partie 2). Quand le salarié part, vous perdez la main sur ses conversations.

L'ANSSI le dit simplement : aucune donnée sensible dans un outil d'IA grand public. Dans mes accompagnements, je vois des salariés payer leur abonnement eux-mêmes. Ils ont franchi le pas avant l'entreprise, il leur manque un cadre.

Source : LayerX 2025 (vend de la sécurité, échantillon non indiqué) ; IBM/Ponemon 2025, 600 organisations ; ANSSI, recommandation R34, 29/04/2024.

Un skill téléchargé *agit avec vos droits*

Un skill, c'est un fichier d'instructions, parfois accompagné de code, que votre IA suit. Installé depuis Internet, il hérite de ce que votre IA peut faire : lire vos fichiers, utiliser vos connecteurs.

Un badge « vérifié » ne suffit pas : des chercheurs ont contourné les scanners de sécurité des places de marché dans plus de 90 % des cas, dans une démonstration sur 8 scanners.

01 / AUDIT

341 sur 2 857

skills malveillants sur ClawHub, la place de marché d'OpenClaw, en février 2026 (Koi Security).

02 / ANALYSE

76

charges malveillantes confirmées à la main sur 3 984 skills scannés (Snyk).

HONNÊTEMENT

Je n'ai pas trouvé d'attaque avérée de ce type sur Claude ou ChatGPT. Ma recherche n'est pas exhaustive, et le mécanisme reste le même.

Source : SC Media citant Koi Security ; Snyk, 05/02/2026 ; arXiv 2607.02357. Koi et Snyk vendent de la sécurité.

Un mail peut donner des ordres *à votre assistant*

01 / IL ARRIVE

Un mail contient une instruction invisible pour vous.

02 / IL EST LU

Votre assistant, branché sur la messagerie, le résume.

03 / IL OBÉIT

Il prend le texte du mail pour une consigne.

04 / ÇA SORT

Il envoie ou affiche des données qu'il n'aurait pas dû.

CE QUI EST ÉTABLI

Ce mécanisme s'appelle l'injection de prompt. La faille EchoLeak l'a montré sur Microsoft 365 Copilot en 2025 : un mail suffisait, sans clic. Elle a été corrigée, et je n'ai pas trouvé de trace d'exploitation réelle.

CE QUI RESTE OUVERT

Anthropic juge le problème loin d'être résolu : 1,4 % d'attaques réussies en test interne, avec protections. OpenAI pense qu'il ne sera sans doute jamais totalement résolu. **Comme on ne contrôle pas tout ce que l'IA lit, on limite ce qu'elle peut faire** (partie 2).

Source : arXiv 2509.10540 (EchoLeak) ; Anthropic, 24/11/2025 ; OpenAI, article sur le renforcement d'Atlas ; OWASP LLM01:2025.

L'outil monté en une après-midi *part-il en production ?*

CE QUI EST DOCUMENTÉ

45 %	des tests de code généré par IA introduisent une faille de sécurité (Veracode, plus de 100 modèles).
Lovable	Une faille de 2025 a exposé des bases de projets dépourvues de règles d'accès.
Replit	Un agent a supprimé une base de production malgré un gel du code, en juillet 2025. Elle a pu être restaurée.
3,2 %	de commits assistés par IA laissent fuiter un secret, contre 1,5 % en moyenne (GitGuardian, commits publics).

QUATRE QUESTIONS AVANT DE LE BRANCHER

Dans mes accompagnements, je vois des salariés construire un outil puis demander sa mise en production en urgence. Avant de dire oui :

1. **Qui peut y accéder depuis Internet ?**
2. **Où sont les données, et qui les voit ?**
3. **Où sont rangées les clés d'accès ?**
4. **Qui le reprend si son auteur part ?**

Source : Veracode 2025 et GitGuardian 2026 (vendent de la sécurité) ; CVE-2025-48757 ; The Register, 07/2025, d'après la personne concernée.

Les réglages, *outil par outil*

Un outil bien réglé ferme une bonne partie des portes de la partie 1. Ces réglages sont à la portée d'un administrateur non technique. Les interfaces bougent vite : tout ce qui suit est relevé au 2 novembre 2026.

01 / COMPTE

« Ils ont chacun un compte, ça suffit. »

02 / OFFRE

« Mon offre couvre tout. »

03 / DÉFAUT

« Les réglages par défaut sont sûrs. »

04 / CONNECTEURS

« Tout le monde peut tout brancher. »

05 / RÔLE

« L'informatique s'en occupera. »

Le premier réglage, *c'est le type de compte*

Non / COMPTE PERSONNEL

ENTRAÎNEMENT

Claude : possible si « Model Improvement » est activé.
ChatGPT : activé par défaut, à couper dans Data Controls.

CONTRAT

Conditions grand public. Pas d'accord de traitement des données.

GESTION

Aucun administrateur. Au départ du salarié, le compte part avec lui.

Oui / COMPTE D'ENTREPRISE

ENTRAÎNEMENT

Claude, ChatGPT : pas d'entraînement par défaut sur vos contenus.
Dust : jamais, d'après sa page sécurité.

CONTRAT

Accord de traitement disponible. Le chemin varie selon l'outil (page suivante).

GESTION

Un administrateur crée, règle et ferme les comptes.

LE PIÈGE

Le pouce en l'air compte aussi. Chez Claude et ChatGPT, envoyer un retour peut partager la conversation avec l'éditeur, même si l'entraînement est coupé. Sur Claude, l'administrateur peut le désactiver (Data and Privacy > Rate chats).

Source : Pages officielles Anthropic (privacy.claude.com), OpenAI (help.openai.com, openai.com/enterprise-privacy) et Dust (dust.tt/home/security).

COMPARATIF

Ce que couvre *l'offre d'une PME*

OFFRES INTERMÉDIAIRES, AU 02/11/2026

	Claude Team	*ChatGPT Business*	*Dust Business*
Entraînement sur vos données	Non par défaut	Non par défaut	Jamais, d'après l'éditeur
Accord de traitement (DPA)	Intégré aux conditions	Sur formulaire	Listé en Enterprise
Hébergement en Europe	Non vérifié	Réservé à Enterprise, Edu, API	Oui, UE ou US au choix
Authentification unique (SSO)	Oui	Oui	Oui
Comptes créés automatiquement (SCIM)	Enterprise	Non vérifié	Enterprise
Journaux d'audit	Enterprise	Non vérifié	Enterprise
Durée de conservation réglable	Enterprise	Oui, l'admin la règle	Enterprise

Claude : *six réglages à poser*

ORGANIZATION SETTINGS

1	Authentication	SSO, pour que la sortie d'un salarié ferme aussi son accès.
2	Connectors	N'ouvrez que ce qui sert. Réglez chaque outil : « Always allow », « Needs approval », « Blocked ».
3	Plugins & skills > Policy	Le partage est actif par défaut sur Team. Décidez qui crée, qui partage, et diffusez les skills validés.
4	Capabilities	Décidez de la mémoire et de la recherche web. Couper la mémoire supprime ses entrées.
5	Claude in Chrome, Cowork	Ces accès agissent à votre place. Activez-les pour ceux qui en ont besoin, en mode manuel d'abord.
6	Data and Privacy > Rate chats	Coupez les retours utilisateurs si vos conversations sont sensibles.

Source : support.claude.com, articles sur le SSO, les connecteurs, les skills, la mémoire et Cowork, lus le 02/11/2026. Sélection de l'auteur.

Lire, oui. *Agir, sous approbation*

01 / AUTONOME

Elle peut le faire seule.

« Always allow »

- Chercher un document
- Résumer un mail
- Consulter une fiche client

02 / APPROUVÉ

Un humain valide chaque fois.

« Needs approval »

- Envoyer un mail
- Modifier une fiche CRM
- Créer un événement

03 / INTERDIT

Elle ne le fait jamais.

« Blocked »

- Supprimer des données
- Payer ou virer
- Modifier des droits d'accès

POURQUOI

C'est la réponse à l'injection de prompt : même trompée par un mail piégé, une IA qui ne peut pas envoyer sans vous ne fera pas sortir vos données par ce chemin. L'ANSSI propose ce classement en trois niveaux pour les agents.

Source : ANSSI, Essentiels de l'IA agentique, 10/2026 ; libellés de Claude, support.claude.com. Exemples de l'auteur, à adapter.

ChatGPT : *ce qu'il faut regarder*

WORKSPACE

1 **Identity & access**
SSO en SAML ou OIDC, après vérification de votre domaine (Workspace settings).

2 **Agents de workspace**
Désactivés par défaut. L'administrateur décide qui en construit et qui en publie.

3 **Apps, partage, rétention**
L'administrateur choisit les services ouverts et peut les limiter à certains groupes.

4 **Accord de traitement**
Il se demande sur un formulaire : pensez-y, il n'est pas automatique.

À SAVOIR

Les administrateurs peuvent consulter, exporter et supprimer les conversations du workspace. Dites-le à vos équipes.

Supprimer une conversation n'efface pas la donnée dans l'application connectée.

Déconnecter une application n'efface pas les conversations passées qui l'ont utilisée.

Source : help.openai.com et openai.com/enterprise-privacy. Détail des connecteurs Business non lu en direct : à vérifier dans votre console.

Dust : *le cloisonnement par espaces*

Dust est une plateforme d'agents qui s'appuie sur plusieurs modèles d'IA. Son mécanisme de sécurité central est **l'espace** : un agent n'est visible que des membres de ses espaces.

Le réglage à poser dès le départ : ce qui va dans un espace ouvert, et ce qui reste dans un espace restreint, comme la paie ou les données clients.

01 / ESPACES

5 sur l'offre Business, 100 en Enterprise. Ouverts ou restreints, créés par les admins.

02 / DONNÉES

Seuls les admins ajoutent des sources depuis les connexions.

03 / RÔLES

Membre, Builder, Admin. Limitez le rôle Builder à ceux qui construisent.

04 / HÉBERGEMENT

Union européenne ou États-Unis, au choix. Journaux d'audit en Enterprise.

Source : docs.dust.tt (espaces, droits, SSO, audit) et dust.tt/home/security, lus le 02/11/2026. Modèle précis et lieu de traitement non vérifiés.

Sans administrateur, *aucun de ces réglages n'existe*

Sur une offre d'équipe, c'est l'administrateur qui décide des connecteurs ouverts. Sans lui, un salarié reste bloqué, ou il repasse par son compte personnel.

Ce rôle ne demande pas d'être technique. Il demande d'être organisé, et d'avoir la main sur les comptes. En PME, c'est souvent le responsable des opérations, l'office manager ou le dirigeant lui-même.

SA FICHE DE POSTE EN CINQ LIGNES

1. **Il active les outils et les connecteurs.**
2. **Il tient la liste : qui a accès à quoi.**
3. **Il revoit les droits chaque mois.**
4. **Il ferme les accès le jour d'un départ.**
5. **Il sait qui appeler en cas d'incident.**

Source : revue mensuelle : ANSSI, recommandation R35. Fiche de poste : proposition de l'auteur.

Le socle, *puis la routine*

Les gestes les plus rentables n'ont rien de spécifique à l'IA. Un compte Claude ou ChatGPT est un compte comme les autres, et il donne accès à beaucoup plus de choses qu'avant.

01 / ACCÈS

« Mon mot de passe est solide. »

02 / COFFRE

« Je les note dans un fichier. »

03 / COPIES

« On a des sauvegardes, je crois. »

04 / OUTILS

« Il a été validé une fois, c'est bon. »

05 / RYTHME

« On regardera quand on aura le temps. »

Le geste le plus rentable : *la double authentification*

99,22 %

de réduction du risque de compromission d'un compte, d'après une étude de Microsoft sur des comptes professionnels. **Encore 98,56 % quand le mot de passe a déjà fuité.**

Un second facteur protège le compte même quand le mot de passe est connu de l'attaquant.

01 / OÙ, DANS CET ORDRE

1. La messagerie : elle réinitialise tout le reste.
2. La banque et les comptes de paiement.
3. Les comptes IA et les comptes administrateur.
4. Tous les autres services qui le proposent.

02 / COMMENT

Une application dédiée protège mieux que le SMS, d'après la même étude.

Les passkeys, ces clés de connexion sans mot de passe, dès qu'elles sont proposées : c'est la recommandation du NCSC britannique.

Source : Meyer et al., Microsoft, arXiv 2305.00945, 2023 ; ANSSI/CPME 2024 ; NCSC UK, 21/05/2026.

Un gestionnaire de mots de passe, *pour toute l'équipe*

POURQUOI

Il crée et retient un mot de passe différent par service. **Le jour où un service se fait voler ses données, l'attaquant ne peut pas ouvrir les autres.**

Il sert aussi à partager un accès commun sans l'envoyer par mail, dans un tableur, ou pire, dans un chat d'IA.

Il devient un point critique : protégez-le par un mot de passe principal long et par la double authentification.

CE QUE DISENT LES AUTORITÉS

ANSSI et CPME

15 caractères pour les services critiques, 9 pour les autres, et un coffre-fort.

CNIL

12 caractères variés, 14 caractères, ou une phrase de 7 mots. Plus de renouvellement périodique, sauf pour les comptes à privilèges.

Cybermalveillance.gouv.fr

Cite KeePassXC, certifié par l'ANSSI en version 2.7.9, pour qui veut une solution gratuite.

Source : ANSSI/CPME, 13 questions, 12/2024 ; CNIL, délibération 2022-100 ; Cybermalveillance.gouv.fr, 05/10/2026 ; NCSC UK.

Une faille connue *s'exploite en quelques jours*

PHP, OCTOBRE 2026

PHP fait tourner une grande partie des sites web, dont ceux sous WordPress. Voici le calendrier de deux failles critiques.

17/10	Premières tentatives observées par Amazon, quatre jours avant le correctif.
21/10	PHP publie ses correctifs. Le CERT-FR, l'équipe d'alerte de l'ANSSI, publie une alerte.
26/10	L'agence américaine CISA ajoute les failles à son catalogue de failles exploitées.

Plus de 5,1 millions d'hôtes PHP exposés d'après VirusTotal, sans que l'on sache combien sont vulnérables.

WORDPRESS

91 %

des nouvelles failles de l'écosystème WordPress en 2025 touchent les **extensions**. Le cœur du logiciel n'en compte que 6, peu graves.

46 %

n'avaient pas de correctif de leur développeur au moment où elles ont été rendues publiques.

Source : php.net, 21/10/2026 ; CERT-FR, CERTFR-2026-ALE-014 ; BleepingComputer, 10/2026 (Amazon, CISA, VirusTotal) ; Patchstack 2026 (vend de la sécurité).

Tenir à jour, *y compris ce qu'on a oublié*

01 / AUTOMATIQUE

Activez-les partout

Postes, téléphones, navigateurs, logiciels. Vérifiez de temps en temps qu'elles passent vraiment.

02 / INVENTAIRE

Listez ce qui tourne

Le site web, son hébergement, sa version de PHP, ses extensions, les outils construits en interne.

03 / MÉNAGE

Retirez ce qui ne sert plus

Une extension désactivée ou un outil abandonné reste une porte. Si l'éditeur ne suit plus, remplacez.

04 / PRUDENCE

Sauvegardez avant

Sauvegardez avant une mise à jour importante, et ne téléchargez que depuis le site de l'éditeur.

LA QUESTION

Votre site est géré par un prestataire ? Demandez-lui qui met à jour quoi, et sous quel délai après une alerte comme celle de PHP.

Source : Cybermalveillance.gouv.fr, mises à jour, 05/10/2026. Inventaire du site et ménage des extensions : proposition de l'auteur.

Des sauvegardes *qui ont déjà servi*

COPIES

3 **copies de vos données**

SUPPORTS

2 **supports différents**

HORS LIGNE

1 **copie hors ligne, débranchée**

TESTEZ LA RESTAURATION

L'ANSSI demande des tests de restauration réguliers. **Une sauvegarde jamais restaurée reste une hypothèse.** Choisissez un fichier, un dossier, une boîte mail, et chronométrez.

POURQUOI AVEC L'IA AUSSI

Un agent qui a le droit d'écrire peut aussi effacer. Et en cas de rançongiciel, 66 % des entreprises chiffrées ont récupéré leurs données par leurs sauvegardes, d'après Sophos.

Source : ANSSI/CPME 2024 ; Cybermalveillance.gouv.fr, 14/09/2026 ; Sophos 2026 (vend de la sécurité), 2 158 répondants de 100 à 5 000 salariés.

Six règles pour autoriser *un skill ou un connecteur*

01 / BESOIN

Un usage écrit

Chaque accès répond à un besoin métier précis. Sans besoin, pas d'accès.

02 / DÉFAUT

Refus par défaut

On n'ouvre que ce qui a été demandé et accepté, même pour un outil connu.

03 / ORIGINE

Une source connue

Un skill se lit avant de s'installer, ou se fait lire par quelqu'un qui sait.

04 / DROITS

Lecture d'abord

On commence en lecture seule. L'écriture passe sous approbation.

05 / RESPONSABLE

Un propriétaire

Une personne nommée répond de chaque connecteur et de chaque skill partagé.

06 / SORTIE

Un arrêt d'urgence

Vous savez qui coupe, où, et en combien de temps. Revue des droits chaque mois.

Source : déduit par l'auteur d'ANSSI, Essentiels de l'IA agentique, 10/2026, et de la recommandation R35. Aucune source ne les énonce telles quelles.

Une page de règles, *pas un règlement*

01 / LIBRE

Tout outil autorisé

Textes publics, idées, brouillons sans nom de client, code sans clé d'accès.

02 / ENCADRÉ

Compte d'entreprise réglé

Données clients, contrats, ressources humaines, chiffres non publics.

03 / JAMAIS

Dans aucun chat

Mots de passe, clés d'accès, codes reçus par SMS, données bancaires.

LE FORMAT

Dans une PME, une page suffit : quels outils, quelles données, qui prévenir en cas de doute. Interdire sans proposer d'outil renvoie vos équipes vers leurs comptes personnels. Le guide de l'adoption de l'IA détaille cette logique.

Source : principe : ANSSI, recommandation R34, 29/04/2024. Répartition : proposition de l'auteur, à adapter à votre activité.

L'évaluation régulière, *en quatre rythmes*

RYTHMES PROPOSÉS, À ADAPTER

CHAQUE MOIS	**Les droits, les mises à jour, le message** Relisez les droits des outils d'IA (recommandation R35 de l'ANSSI), vérifiez que les mises à jour sont passées, envoyez le message à l'équipe.
CHAQUE TRIMESTRE	**Le contrôle du site et des copies** SSL Labs pour le chiffrement, Mozilla Observatory pour les en-têtes du site, Have I Been Pwned pour les adresses compromises de votre domaine. Puis un test de restauration.
CHAQUE ANNÉE	**Un regard extérieur** MonAideCyber : un aidant bénévole fait un diagnostic de 1h30 à 2h, gratuit, et remet 6 recommandations.
À CHAQUE MISE EN SERVICE	**Les outils construits en interne** Les quatre questions de la slide 16, puis une relecture par un développeur. Un test d'intrusion si l'outil touche des données sensibles.

Source : ANSSI R35 ; MesServicesCyber (MonAideCyber) ; Qualys, Mozilla, Have I Been Pwned (gratuit jusqu'à 10 adresses compromises). Rythmes : proposition de l'auteur.

Une veille légère, *avec un responsable*

OÙ REGARDER

CERT-FR — CHAQUE SEMAINE

Les alertes sur les failles critiques, avec un flux RSS public. C'est par là qu'est passée l'alerte PHP.

Cybermalveillance.gouv.fr — CHAQUE MOIS

Les arnaques en cours et les fiches réflexes, en français courant.

Vos éditeurs d'IA — CHAQUE MOIS

Les notes de version de Claude, ChatGPT et Dust : un nouveau réglage, une nouvelle capacité à décider.

Vos prestataires — CHAQUE TRIMESTRE

Ce qu'ils ont corrigé, ce qui reste ouvert chez vous.

COMMENT LA TENIR

Un filtre, une seule question : cela touche-t-il un outil de notre inventaire ? Si non, on passe.

Si oui, trois réponses : que fait-on, qui le fait, pour quand.

L'IA peut faire le tri : une routine lit ces sources chaque semaine et ne remonte que ce qui concerne vos outils. La décision reste à un humain.

Source : cert.cyber.gouv.fr ; Cybermalveillance.gouv.fr. Rythmes et méthode : proposition de l'auteur.

Un message par mois, *à toute l'équipe*

La sécurité s'oublie vite. Une formation par an ne suffit pas : un rappel court et régulier garde le sujet vivant, et donne à chacun le bon réflexe au bon moment.

LES RÈGLES DU FORMAT

1. **Cinq lignes, pas plus.**
2. **Toujours le même plan, le même jour.**
3. **Un exemple réel, sans désigner personne.**
4. **Une personne à contacter en cas de doute.**

Exemple fictif

Sécurité : le point de novembre

CE QUI A CHANGÉ

Le connecteur CRM est ouvert à tous, en lecture seule.

À FAIRE

Activez la double authentification sur Claude avant vendredi.

À NE PAS FAIRE

N'installez pas de skill trouvé sur Internet sans passer par Léa.

VU CE MOIS-CI

Un faux appel du « dirigeant » pour un virement urgent. Rappelez.

UN DOUTE ?

Léa. Signaler ne vous sera jamais reproché.

Qui a accès à quoi, *et que se passe-t-il au départ ?*

UN INVENTAIRE EN UNE LIGNE PAR OUTIL

OUTIL	RESPONSABLE	COMPTES	ACCÈS
Claude Team	Léa	12	Mail, Drive
ChatGPT Business	Léa	4	Aucun
Connecteur CRM	Marc	6	Lecture
Skill devis	Marc	12	Interne
Outil de suivi interne	Sam	8	Base clients

Exemple fictif. Un tableur suffit : ce qui compte, c'est qu'il existe et qu'il ait un responsable.

LE JOUR D'UN DÉPART

1. **Fermer le compte unique (SSO).**
2. **Retirer les licences IA.**
3. **Couper ses connecteurs personnels.**
4. **Réattribuer ses skills et ses agents.**
5. **Changer les accès partagés.**

Trois moments, *quelques gestes*

Les réglages de la partie 2 couvrent ce que l'administrateur décide. Le reste se joue dans les gestes de chacun : au moment d'installer, de coller, de valider.

01 / INSTALLER

« Ce skill a plein d'avis positifs. »

02 / COLLER

« C'est juste pour reformuler. »

03 / VALIDER

« Je clique sur oui, ça va plus vite. »

04 / APPEL

« C'était bien sa voix. »

05 / ERREUR

« Je n'ose pas le dire. »

Une formation par an *change peu de choses*

CE QUE MESURE LA RECHERCHE

19 500

employés suivis sur 8 mois et 10 campagnes de faux mails. La formation annuelle n'a pas d'effet significatif sur le taux d'échec, et la formation intégrée à l'exercice donne des différences minimes.

CE QUI MARCHE MIEUX

Compter les signalements plutôt que les clics, recommande le centre britannique de cybersécurité.

Aucune sanction : un salarié qui a peur ne dit rien, et c'est le silence qui coûte cher.

Un canal simple : une adresse, un fil de discussion, une personne. Et un merci à chaque signalement.

Une erreur signalée en dix minutes se répare mieux qu'une erreur découverte en dix jours.

Source : Ho et al., IEEE S&P 2025 ; NCSC UK, révisé le 13/02/2024.

Les trois questions *à se poser*

01 / AVANT D'INSTALLER

D'où vient cet outil ?

- Vient-il d'une source connue ?
- L'administrateur l'a-t-il validé ?
- Que peut-il lire, que peut-il faire ?

02 / AVANT DE COLLER

Qu'y a-t-il dedans ?

- Un nom de client ? Une donnée RH ?
- Un mot de passe, une clé d'accès ?
- Suis-je sur le compte d'entreprise ?

03 / AVANT DE VALIDER

Que va-t-il se passer ?

- Quoi part, vers qui ?
- Est-ce que je comprends ce que je valide ?
- Puis-je revenir en arrière ?

Quatre gestes *contre les arnaques à l'IA*

01 / VIREMENT

Une validation sans exception

Tout changement de RIB ou virement imprévu suit une validation hiérarchique « non dérogeable », même sous pression, même venant du dirigeant.

02 / VOIX ET VISIO

Raccrochez, rappelez

Rappelez sur un numéro que vous avez déjà. Convenez d'un mot de code entre dirigeants et personnes qui paient.

03 / CODES

Jamais transmis

Personne de légitime ne vous demande un code reçu par SMS ou une validation d'application au téléphone.

04 / LIEN ET QR CODE

Au moindre doute, non

N'ouvrez pas. Passez par votre favori ou tapez l'adresse vous-même. Page suivante, comment lire une adresse.

Source : Cybermalveillance.gouv.fr (fiche FOVI, 14/10/2026 ; quishing, 10/08/2026) ; FBI IC3, PSA I-120324, qui vise les familles : extension à l'entreprise par l'auteur.

Lire une adresse *en dix secondes*

LE NOM DE DOMAINE SE LIT JUSTE AVANT LE PREMIER /

`https://www.banque-exemple.fr/connexion`	Le domaine est banque-exemple.fr.	Bon
`https://banque-exemple.fr.connexion-securisee.com/`	Vrai domaine : connexion-securisee.com.	Faux
`https://banque-exemple-connexion.com/`	Vrai domaine : banque-exemple-connexion.com.	Faux
`https://www.banque-exemple.fr/`	Un I majuscule à la place du l.	Faux

LE FAVORI

Pour la banque et les outils critiques, passez toujours par un favori.

LE GESTIONNAIRE

Il ne propose votre mot de passe que sur le bon domaine : son silence est une alerte.

LE CADENAS

Il dit que la connexion est chiffrée. Un faux site peut l'avoir aussi.

Ce que chacun doit *savoir faire*

CINQ COMPÉTENCES, POUR TOUS LES POSTES

1. **Reconnaître une donnée sensible**
   sans avoir à demander, sur son propre métier.
2. **Vérifier avant d'agir**
   une demande inhabituelle passe par un second canal.
3. **Savoir ce que son outil peut faire**
   ce qu'il lit, ce qu'il écrit, à quoi il est branché.
4. **Relire ce que produit l'IA**
   avant que ça sorte de l'entreprise, sans exception.
5. **Signaler vite**
   une erreur, un doute, un clic malheureux, sans crainte.

Source : aucune source ne définit de liste : proposition de l'auteur, à partir des recommandations de l'ANSSI et du NCSC.

Votre infogérant *a vos clés*

Dans beaucoup de PME, l'informatique est confiée à un prestataire, parfois à une seule personne. C'est souvent un bon choix. C'est aussi un accès très large qu'on n'interroge jamais.

01 / ACCÈS

« Il a tout, c'est plus simple. »

02 / IA

« Je ne sais pas ce qu'il utilise. »

03 / SORTIE

« On verra si on change un jour. »

La porte d'entrée *peut être chez le prestataire*

48 %

des violations de données impliquent un tiers en 2025, d'après Verizon. Son communiqué ne précise pas ce qu'il entend par tiers.

LE CAS KASEYA

En juillet 2021, une faille dans un logiciel d'administration à distance a touché environ **60 infogérants**, et par ricochet **plus de 1 000 entreprises** à l'arrêt.

CE QUE ÇA VEUT DIRE

Un prestataire qui a les accès de vingt clients vaut plus qu'une PME isolée. Vous pouvez lui faire confiance, et lui poser des questions.

Source : Verizon DBIR 2026 (vend de la sécurité) ; Wikipédia, attaque Kaseya VSA (source secondaire).

Huit questions *à poser à votre infogérant*

À ENVOYER TEL QUEL

1. Quels tiers interviennent sur notre parc : hébergeurs, outils d'administration à distance ?
2. Qui détient des comptes d'administration chez nous, et sont-ils nominatifs ?
3. La double authentification est-elle active sur tous ces comptes ?
4. Vos équipes utilisent-elles l'IA sur nos données ? Quels outils, quels réglages ?
5. Vos agents ou connecteurs ont-ils accès à nos systèmes ? Avec quels droits, quel journal ?
6. Sous quel délai nous alertez-vous d'un incident chez vous ?
7. Où sont nos sauvegardes, testées quand, séparées de vos accès ?
8. Comment récupérons-nous données et accès si le contrat s'arrête ?

Source : thèmes d'ANSSI, Maîtriser les risques de l'infogérance (page de présentation lue) ; CNIL pour l'alerte du sous-traitant. Questions IA : proposition de l'auteur.

Que faire *en cas d'attaque*

Le moment où il faut décider vite est aussi celui où l'on a le moins accès à ses outils. Ce qui compte se prépare avant, sur une page imprimée.

01 / RÉFLEXE

« On éteint tout et on redémarre. »

02 / SILENCE

« Inutile d'en parler pour l'instant. »

03 / RANÇON

« On paie, et c'est réglé. »

04 / VITESSE

« On réinstalle tout de suite. »

Huit étapes, *dans cet ordre*

01 / ISOLER

Coupez Internet et retirez les machines du réseau, **sans les éteindre**.

02 / PROTÉGER

Déconnectez les sauvegardes pour qu'elles ne soient pas chiffrées.

03 / ALERTER

Appelez votre prestataire, ou suivez le parcours de 17Cyber.

04 / ASSURER

Prévenez votre assureur : le délai dépend de votre contrat.

05 / PORTER PLAINTE

Avant de réinstaller. **Sous 72 h pour être indemnisé par l'assurance.**

06 / NOTIFIER

La CNIL sous 72 h si des données personnelles ont pu être touchées.

07 / REFUSER

Ne payez pas : rien ne garantit la récupération de vos données.

08 / RESTAURER

Repartez d'une sauvegarde antérieure à l'attaque, vérifiée saine.

Source : Cybermalveillance.gouv.fr, fiche rançongiciels, 07/05/2026 ; 17cyber.gouv.fr ; code des assurances, art. L12-10-1 ; CNIL. En urgence : 17 ou 112.

Si un compte IA *est compromis*

DANS LES PREMIÈRES HEURES

1. **Changez le mot de passe**
   et activez la double authentification si ce n'était pas fait.
2. **Déconnectez les sessions et les appareils inconnus**
   après une capture d'écran : sinon, l'attaquant reste connecté.
3. **Contrôlez la messagerie liée**
   coordonnées de récupération, redirections, règles de filtrage.
4. **Révoquez les connecteurs et les clés d'accès**
   un connecteur reste ouvert tant qu'il n'est pas coupé. Reconnectez ensuite un par un.
5. **Relisez ce que le compte a fait**
   conversations, actions, envois, fichiers ouverts. Gardez les preuves.
6. **Prévenez qui doit l'être**
   contacts, équipe, et la CNIL sous 72 h si des données personnelles ont fuité.

Source : Cybermalveillance.gouv.fr, piratage de compte, 22/10/2026. Étapes 4 et 5 déduites par l'auteur : pas de procédure officielle trouvée, voir l'aide de votre éditeur.

6 · QUE FAIRE EN CAS D'ATTAQUE | À PRÉPARER

Une page, *imprimée*

Le jour d'une attaque, votre messagerie et vos fichiers peuvent être inaccessibles. **La procédure ne doit donc pas vivre uniquement dedans.**

Écrivez cette page une fois. Imprimez-la. Donnez-la à deux personnes. Relisez-la une fois par an.

CE QU'ELLE CONTIENT

Qui décide
la personne qui peut faire couper, et son remplaçant

Prestataire informatique
numéro direct, y compris le soir et le week-end

Assureur
numéro, numéro de contrat, délai de déclaration

Aide publique
17cyber.gouv.fr, ou le 17 et le 112 en urgence

Banque
le bon interlocuteur pour bloquer un virement

Sauvegardes
où elles sont, comment les déconnecter

CONCLUSION | LA CHECKLIST

Dix-huit points *à passer en revue*

01 / COMPTES

- [ ] Double authentification sur la messagerie
- [ ] Double authentification sur les comptes IA
- [ ] Gestionnaire de mots de passe déployé
- [ ] Comptes d'entreprise pour les usages pros
- [ ] Un administrateur nommé
- [ ] Une page de règles pour l'équipe

02 / OUTILS

- [ ] Connecteurs limités au besoin
- [ ] Écriture sous approbation humaine
- [ ] Politique de skills posée
- [ ] Inventaire tenu à jour
- [ ] Revue mensuelle des droits
- [ ] Mises à jour automatiques, site compris

03 / FILET

- [ ] Sauvegarde hors ligne, restauration testée
- [ ] Procédure de virement sans exception
- [ ] Canal de signalement sans sanction
- [ ] Veille et message mensuel à l'équipe
- [ ] Huit questions posées à l'infogérant
- [ ] Page d'incident imprimée

CONCLUSION | PAR OÙ COMMENCER

Le plan, *en trois temps*

01 / CETTE SEMAINE

Fermer les portes les plus simples

- Double authentification partout où elle compte
- Gestionnaire de mots de passe, mises à jour automatiques
- Une page de règles pour l'équipe

02 / CE MOIS-CI

Reprendre la main sur vos outils

- Migrer les comptes personnels
- Régler connecteurs et skills
- Nommer l'administrateur, tenir l'inventaire

03 / CE TRIMESTRE

Installer le rythme

- Tester une restauration, lancer la veille et le message mensuel
- Interroger votre infogérant
- Diagnostic MonAideCyber, page d'incident

CONCLUSION SOURCES

D'où viennent *les chiffres*

Autorités. ANSSI : Panorama de la cybermenace 2025 ; L'IA générative face aux attaques (04/02/2026) ; IA générative, recommandations R27, R34, R35 (29/04/2024) ; Essentiels IA agentique (10/2026) ; guide TPE/PME avec la CPME (12/2024). CNIL : mots de passe (2022), violations de données. Cybermalveillance.gouv.fr : bilan 2025 et fiches réflexes. NCSC UK. CERT-FR, alerte PHP (21/10/2026). FBI IC3 2025 et PSA I-120324.

Études. Heiding et al., arXiv 2412.00586 ; Meyer et al., arXiv 2305.00945 ; Ho et al., IEEE S&P 2025 ; SkillCloak, arXiv 2607.02357 ; EchoLeak, arXiv 2509.10540.

Éditeurs d'IA. Anthropic : rapports du 27/08/2025 et du 13/11/2025, défenses contre l'injection (24/11/2025), documentation de Claude. OpenAI : aide en ligne, Enterprise privacy. Dust : page sécurité, documentation. PHP : php.net, 21/10/2026.

Éditeurs qui vendent de la sécurité, à lire comme tels. Verizon DBIR 2026 ; IBM/Ponemon 2025 ; Sophos 2026 ; LayerX 2025 ; Snyk ; Koi Security ; Veracode 2025 ; GitGuardian 2026 ; Okta ; Patchstack 2026.

Presse et sources secondaires. CNN (Arup) ; The Register (Replit) ; BleepingComputer (failles PHP) ; SC Media (Koi) ; Wikipédia (Kaseya).

Textes. Code des assurances, art. L12-10-1 (LOPMI) ; RGPD, art. 33.

Écartés faute de source remontée. « 60 % des PME font faillite », « un faux site en 30 secondes », « 30 fois moins cher ».

Relevé du 2 novembre 2026. Les réglages des outils changent vite : vérifiez dans votre console.

ET CHEZ VOUS ?

Par où commencer, *dans votre cas ?*

Je vous aide à cadrer l'usage de l'IA dans votre entreprise : quels outils, quels accès, quelles règles, et comment embarquer vos équipes. Pour un audit de sécurité approfondi, je vous orienterai vers un spécialiste.

30 minutes pour regarder votre situation.

cal.com/colin-dargent/premier-contact

Colin Dargent
Responsable IA externalisé pour PME
forward-ai.fr · linkedin.com/in/colindargent