

La Règle des Deux : le contrôle des agents IA commence par deux clés

1962 : quatre serrures et quatre clés séparent l'ordre du missile parti. 2026 : un agent branché sur le CRM détient les trois clés dans la même main.

Qui détient la seconde clé : la règle des deux transforme le contrôle des agents IA en avantage commercial

■ MESSAGE CLÉ

Un agent qui lit le compte, accorde la remise et envoie le contrat détient les trois clés sous une seule autorisation. La règle des deux sépare lecture, écriture et validation : elle rend l'erreur d'un seul insuffisante.

- 1 Quatre serrures et quatre clés séparent l'ordre reçu du missile parti : aucun opérateur Minuteman ne peut les atteindre seul.
- 2 L'AEPD a reçu le 14/09/2026 la première notification d'une brèche exécutée par un agent IA, qui a enchaîné seul les phases de l'attaque.
- 3 Gartner (26/05/2026) : 40 % des entreprises rétrograderont ou retireront des agents autonomes d'ici 2027, faute de gouvernance.
- 4 Trois livrables portent la règle des deux : séparer lecture, écriture et validation, une seconde clé sur l'irréversible, un registre.

Trois leviers touchés : la preuve qui vend, l'irréversible qui coûte, le déploiement qu'il faut reprendre

CROISSANCE

La preuve devient argument de vente

Le service achats de vos clients réclame des preuves plutôt que des démonstrations : qui écrit, qui valide, où est le journal.

IMPACT Traçabilité, critère d'achat B2B

RISQUES

L'irréversible ne se rattrape pas

Remise consentie, contrat parti, donnée exfiltrée : l'annulation coûte une négociation, une notification ou un client perdu.

IMPACT 1re brèche par agent
(14/09/2026)

COÛTS

Rétrograder coûte le déploiement

Gartner annonce 40 % d'entreprises rétrogradant des agents d'ici 2027 : le contrôle se conçoit avant la mise en production.

IMPACT 40 % d'ici 2027 (Gartner)

Les analystes avaient chiffré le risque avant la première brèche : la gouvernance se paie après l'incident

Les identités et les identifiants deviennent le vecteur : un agent qui obtient un jeton trop permissif traverse plusieurs services avant d'être repéré.

40 %

des entreprises rétrograderont ou retireront leurs agents d'ici 2027
(Gartner, 26/05/2026)

49 %

des décideurs sécurité citent l'IA agentique comme préoccupation
(Forrester, 06/2026)

30 %

des organisations atteignent une maturité 3 ou plus sur l'IA agentique
(McKinsey, 03/2026)

Sources : Gartner (26/05/2026) ; Forrester, The State Of Agentic AI 2026 (03/06/2026) ; McKinsey, AI Trust Maturity Survey (25/03/2026) ; AEPD (14/09/2026).

La règle des deux ne protège pas d'une personne : elle rend l'erreur d'un seul insuffisante

■ LE PRÉCÉDENT : LES SILOS MINUTEMAN ET LA NO-LONE ZONE

- Début des années 1960 : chaque arme nucléaire se trouve en pratique sous le contrôle d'un seul homme.
- Le two-person concept (AFI 91-104) exige deux personnes qualifiées en no-lone zone, chacune jugeant l'autre.
- Quatre serrures, quatre clés : coffre à deux serrures, clés écartées de douze pieds, second centre qui valide.

=

■ LA TRANSPOSITION COMMERCIALE

- En business, l'agent branché sur le CRM lit, chiffre, génère, envoie, met à jour : une seule autorisation.
- Sans compétence équivalente, deux présences ne font pas un contrôle : la seconde clé ouvre la même serrure.
- Les agents ont hérité du geste sans ses contrôles : manager, comité de pricing, juriste hors de la boucle.

Trois manœuvres : séparer les autorisations, poser une seconde clé, tenir le registre des agents autorisés

01

Séparer la lecture, l'écriture et la validation

La Manœuvre

Cartographiez chaque agent sur les quatre niveaux d'autonomie de Gartner, puis donnez un compte distinct par niveau : lecture, écriture, validation.

Outils Gartner, Salesforce Agentforce, HubSpot, Entra Agent ID

02

Poser une seconde clé sur l'irréversible

La Manœuvre

Toute action qu'un client voit ou qu'un régulateur lit exige deux clés : le validateur ne partage ni le jeton, ni le prompt, ni l'objectif de l'agent.

Outils ServiceNow AI Control Tower, Workday, LangSmith, Langfuse

03

Tenir le registre des agents autorisés

La Manœuvre

Nom, propriétaire humain, périmètre de données, niveau d'autonomie, détenteur de la seconde clé, emplacement du journal, date de dernière revue.

Outils Cisco MyAgent (Fortune, 16/09/2026), MEDDPICC, RevOps

Un trimestre pour séparer les clés avant que l'incident ne s'en charge

0-30 JOURS

- Cartographier chaque agent sur les quatre niveaux d'autonomie
- Lister les actions irréversibles de la stack commerciale

Pilote RevOps

KPI Carte des autorisations

30-60 JOURS

- Créer un compte de service par niveau d'autorisation
- Nommer un second détenteur de clé par action irréversible

Pilote RevOps / Dir. commerciale

KPI 0 action à clé unique

60-90 JOURS

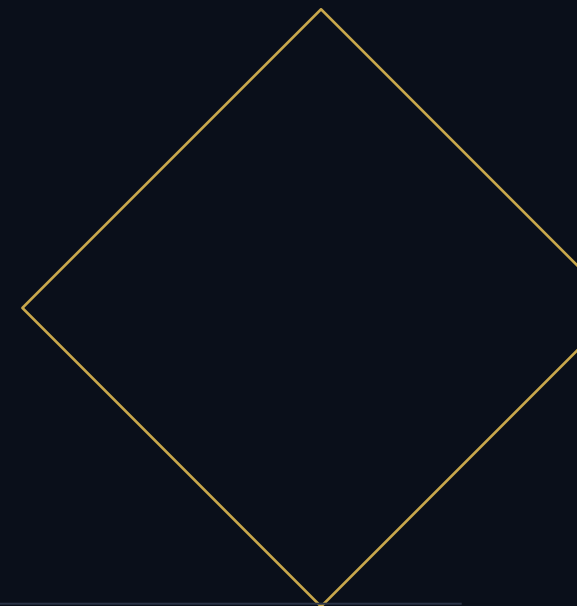
- Ouvrir le registre des agents (propriétaire, périmètre, journal)
- Présenter le registre aux acheteurs et au comité de pricing

Pilote CRO

KPI Registre revu chaque trimestre

« Qui détiendrait la seconde clé si un agent envoyait ce soir un contrat avec une remise hors grille à votre plus gros compte ? »

À ENGAGER — Commencez par la liste des actions irréversibles : ce que le client voit, ce que le régulateur lit, ce que le comptable corrige.



Une réflexion adossée aux références stratégiques, business et historiques

■ RÉFÉRENCES STRATÉGIQUES

- US AIR FORCE, AFI 91-104, Nuclear Surety (2013)
- WIKIPÉDIA, Two-person rule : le two-person concept
- NAT. MUSEUM OF THE USAF, Launching Missiles

■ ÉTUDES BUSINESS

- Gartner : 40 % d'agents retirés, 4 niveaux (26/05/2026)
- McKinsey : maturité 2,6 contre 1,8 (25/03/2026)
- Forrester : 49 % des décideurs sécurité (03/06/2026)
- Fortune : Cisco MyAgent, 90 000 salariés (16/09/2026)

■ SOURCES HISTORIQUES

- WIKIPÉDIA, Permissive action link : NSAM 160 (juin 1962)
- NATIONAL PARK SERVICE, Minuteman Missile Sites
- AEPD, brèche exécutée par un agent IA (14/09/2026)