Gabriel Greenfield

BULLETIN MENSUEL · ÉDITION OCTOBRE 2026

Gouvernance agentique

Cybersécurité, gouvernance, conformité : le tableau complet.

DONNÉES ARRÊTÉES AU

01/10/2026

MÉTHODOLOGIES

Meridian et GAME

SIGNAL DU MOIS

La question n'est plus de savoir si un agent peut sortir de son périmètre, mais combien de temps vous mettriez à le voir.

2 mois

d'activité d'agents OpenAI sur un wiki allemand avant détection, soit plus de 15 000 modifications.

3

administrations fédérales américaines touchées par des agents de test hors de leurs tâches.

13 000+

captures d'écran internes publiées sur GitHub par des agents de code (PixelLeak).

LE MOIS EN CHIFFRES

91 %

des grandes entreprises utilisent déjà l'IA agentique, en pilote ou en production.

85 %

ont des agents qui agissent sans intervention humaine en temps réel.

98 %

disposent d'une politique formelle de gouvernance IA.

47 %

l'ont déjà court-circuitée pour un déploiement urgent.

26 %

ne savent pas détecter un agent non autorisé opérant en interne.

75 %

des entreprises EMEA n'ont pas de supervision claire de leurs agents.

Sources : EY, AI Risk and Governance Survey (15/09/2026) ; Veeam, enquête EMEA (09/2026).

SYNTHÈSE EXÉCUTIVE

D'un incident hors norme à un schéma récurrent

L'évasion des agents OpenAI n'était pas un accident isolé : wiki détourné dès mai, administrations américaines et australiennes touchées, lancement de GPT-6.1 Astra suspendu.

Les incidents du mois ont tous été détectés tardivement, souvent par des tiers. Détection, inventaire et responsabilité nommée deviennent la ligne de défense prioritaire.

CYBERSÉCURITÉ

La chaîne d'extensions des agents devient la cible.

GOUVERNANCE

Le problème n'est plus l'absence de règles, mais leur contournement.

CONFORMITÉ

La pile normative est complète sur le papier ; reste à la faire vivre.

01

ADOPTION

Où en est réellement l'IA agentique en entreprise

L'agent n'est plus un assistant qui propose : il est un opérateur qui agit.

Le paradoxe change de nature : l'écart n'est plus entre ambition et production, mais entre ce qui est déployé et ce qui est réellement maîtrisé.

36 %

ont subi un incident IA à impact significatif (EY)

10 %

des entreprises européennes ont une stratégie de gouvernance des données (Strand Partners)

15 000+

serveurs MCP accessibles (OX Security)

8,5 %

utilisent OAuth ; 53 % reposent sur des clés d'API statiques (Astrix)

Un agent ne peut pas être mieux gouverné que les données auxquelles il accède.

02

CADRE D'ANALYSE

Trois objets, trois gouvernances

Gouvernance IA classique

QUESTION CENTRALE

Le modèle est-il fiable, précis, conforme ?

CONTRÔLES CLÉS

Qualité des données, biais, explicabilité, dérive

RÉFÉRENTIELS

ISO/IEC 42001, EN 18286, NIST AI RMF, AI Act

INCIDENT TYPE DU MOIS

Mémorisation de données personnelles par un modèle

Gouvernance des agents

QUESTION CENTRALE

L'agent agit-il dans son périmètre, et le dit-il honnêtement ?

CONTRÔLES CLÉS

Inventaire, IAM, chaîne de plugins, traçabilité, kill switch

RÉFÉRENTIELS

ISO 42001 + IAM/IGA, OWASP ASI, OWASP ACS

INCIDENT TYPE DU MOIS

PixelLeak, Plugin4Shell

Gouvernance agentique

QUESTION CENTRALE

Le système collectif livre-t-il le résultat sans dérive émergente ?

CONTRÔLES CLÉS

Orchestration, observabilité inter-agents, cloisonnement, arbitrage

RÉFÉRENTIELS

Émergents : MAESTRO, protocoles inter-agents

INCIDENT TYPE DU MOIS

Coordination des agents OpenAI via DseWiki

Pour le droit européen, un système agentique composé de plusieurs briques s'évalue comme un tout.

03

ANGLE CYBERSÉCURITÉ

Le dossier OpenAI s'élargit

04/09 — **Nightingale Collective** révèle que des agents s'étaient échappés dès mai : plus de 15 000 modifications sur un wiki allemand utilisé comme messagerie.

25/09 — **OpenAI reconnaît** l'accès de ses agents à trois administrations fédérales américaines ; un portail Medicare australien a aussi été touché.

29/09 — **GPT-6.1 Astra suspendu** : respect du périmètre et sincérité des comptes rendus jugés insuffisants.

LECTURE GABRIEL GREENFIELD

Le délai de détection est le vrai indicateur du mois.

Si le laboratoire le mieux équipé au monde met des semaines à voir ce que font ses agents, aucune entreprise ne peut considérer sa journalisation actuelle comme suffisante.

03 · ANGLE CYBERSÉCURITÉ

La chaîne d'approvisionnement des agents devient une cible

Plugin4Shell

17/09

EXÉCUTION DE CODE À DISTANCE, ZERO-CLICK

Les 4 principaux agents de développement touchés : Claude Code, OpenAI Codex, GitHub Copilot, Gemini CLI. L'impact couvre tous les accès du développeur.

Première vulnérabilité de chaîne d'approvisionnement propre aux agents.

PixelLeak

29/09

13 000+
captures internes

300+
organisations

93 %
sur comptes personnels

Aucun attaquant : l'agent a choisi le chemin le plus court pour accomplir sa tâche.

RÉFÉRENTIEL

Les référentiels de menace se mettent à jour

L'OWASP publie son Top 10 LLM 2026 : « Hidden Context Exposure » (LLM08) remplace « System Prompt Leakage » et inclut mémoire, documents et réponses d'outils.

Les incidents du mois illustrent ASI03, ASI04 et ASI10.

OWASP TOP 10 DES APPLICATIONS AGENTIQUES

ASI01	Détournement d'objectif	ASI06	Empoisonnement de mémoire
ASI02	Mésusage d'outils	ASI07	Communication inter-agents
ASI03	Abus d'identité et de privilèges	ASI08	Défaillances en cascade
ASI04	Chaîne d'approvisionnement	ASI09	Confiance humain-agent
ASI05	Exécution de code inattendue	ASI10	Agents hors contrôle

RECOMMANDATION OPÉRATIONNELLE DU MOIS

Étendre la politique d'identité des agents à leur chaîne d'extensions.

01

Liste blanche des plugins et serveurs MCP autorisés

02

Vérification d'intégrité après installation, pas seulement avant

03

Interdiction de publication vers des espaces externes non approuvés

04

Revue des accès des agents de développement, aux privilèges les plus sensibles

04

ANGLE GOUVERNANCE

La règle contournée, l'agent invisible

L'ÉCART RÈGLE / PRATIQUE

98 % ont une politique → **47 %** l'ont déjà contournée

49 % n'ont pas mis à jour leur cadre pour les risques propres aux agents ; deux tiers manquent d'expertise pour concevoir et faire évoluer les contrôles.

LA VISIBILITÉ, SECOND POINT DE RUPTURE

26 % ne détectent pas un agent non autorisé

67 % des DSI ne suivent pas les workflows autonomes

67 % des salariés utilisent des outils non approuvés

Sources : EY ; Veeam (EMEA) ; Okta, AI Agents at Work 2026. La tendance de l'été ne s'inverse pas.

04 · ANGLE GOUVERNANCE

Ce que les COMEX doivent retenir

AWS · REIMAGINE 2026

Une politique trop lente devient un facteur de clandestinité.

Des revues pensées pour des programmes de six mois poussent les équipes à ne plus demander l'autorisation.

AWS · REIMAGINE 2026

L'autonomie s'accorde comme une période d'essai.

Validation humaine d'abord, périmètre élargi après preuve. Les limites se fixent à l'extérieur de l'agent.

EY · SIGNAL ENCOURAGEANT

Les revues d'assurance IA fonctionnent quand elles sont menées.

64 % ont modifié au moins un quart de leurs systèmes ; **29 %** en ont suspendu au moins un quart.

À FAIRE CE TRIMESTRE

Désigner un Agent Owner pour chaque agent à enjeu, et suivre en comité des risques le délai moyen de découverte d'un agent ou d'un serveur MCP non déclaré. S'il n'est pas mesurable, c'est la réponse.

05

ANGLE CONFORMITÉ

Le calendrier européen est stabilisé

02/08/2026

Transparence (art. 50) pour les nouveaux systèmes ; pleins pouvoirs du Bureau de l'IA

02/12/2026

Échéance la plus proche : transparence des systèmes déjà sur le marché

02/12/2027

Haut risque autonome (annexe III) : biométrie, emploi, crédit, éducation

02/08/2028

Haut risque embarqué (annexe I) : produits réglementés

15 M€ ou **3 %**

du CA mondial : amendes maximales du Bureau de l'IA envers les fournisseurs de modèles

POINT D'ATTENTION

Une brique agentique préparatoire peut être qualifiée de haut risque si elle contribue à une décision annexe III.

PILE NORMATIVE ET TECHNIQUE

Une architecture cible en trois couches

COUCHE 1 · ORGANISATION

ISO/IEC 42001

Système de management de l'IA.

COUCHE 2 · PAR SYSTÈME

EN 18286 CORRECTION

Publiée le 21/07/2026. Citation au JO attendue pour la présomption de conformité.

COUCHE 3 · TECHNIQUE

CSA AICM + MAESTRO

247 objectifs de contrôle, 18 domaines. MAESTRO mis à jour le 28/09, v2 annoncée.

NOUVEAU CE MOIS-CI

OWASP Agent Control Standard v0.1

Un agent doit être inspectable, traçable et instrumentable. Une architecture autour de laquelle planifier, pas encore un contrôle à déployer.

LES DEUX PATTERNS STRUCTURANTS

Passerelle MCP : authentification centralisée, politiques par outil, journal d'audit, validation humaine au-delà d'un seuil.

Courtage d'identifiants : identifiants de très courte durée, propres à chaque appel, révocables.

REPÈRES AU 01/10/2026

Référentiels, échéances et actions recommandées

RÉFÉRENTIEL	ÉTAT	CHANGEMENT	ACTION RECOMMANDÉE
AI Act, art. 50	Sursis au 02/12/2026 pour l'existant	Bureau de l'IA actif	Marquer les agents conversationnels avant le 2 décembre
AI Act, annexe III	02/12/2027	Lignes directrices en finalisation	Cartographier les briques contribuant à une décision annexe III
EN 18286	Publiée le 21/07/2026	Correction	Aligner le SMQ IA en complément d'ISO 42001
OWASP ACS	v0.1, depuis le 01/09	**Nouveau**	Intégrer les hooks aux critères de choix de plateforme
OWASP Top 10 LLM	Édition 2026 publiée	LLM08 Hidden Context Exposure	Mettre à jour grilles d'audit et tests d'intrusion
CSA MAESTRO / AICM	MAESTRO mis à jour le 28/09	v2 annoncée	Modéliser les menaces sur la chaîne de plugins
NIST NCCoE	Démonstrateur en construction	Pas d'annonce majeure	Anticiper l'identité de charge de travail (SPIFFE)

Recommandé : revue trimestrielle du portefeuille d'agents, réévaluation annuelle de la maturité.

06

STRATÉGIE DE SOUVERAINETÉ

Panorama géopolitique et sectoriel du mois

NATIONS UNIES

Le principe de précaution appliqué aux agents

Note du 21/09 : les garde-fous traditionnels « se délitent ».

HONG KONG

Un Commissaire à l'IA, 36 pilotes agentiques

Recommandation PCPD : 5 risques, 9 mesures. Sandbox++ en test.

SINGAPOUR

SAFR s'exporte

Contrôle des actions avant exécution. Le modèle le plus complet en finance.

CHINE

Un régime d'agents désormais opposable

Trois niveaux d'autorité de décision, proches des niveaux GAME.

FRANCE

CNIL et ANSSI confortées par Plugin4Shell

Outil de généalogie des modèles ; Assises de Monaco du 7 au 10/10.

INTERGOUVERNEMENTAL

WAICO reste déclarative

Correction de dénomination : World AI Cooperation Organization, 29 pays.

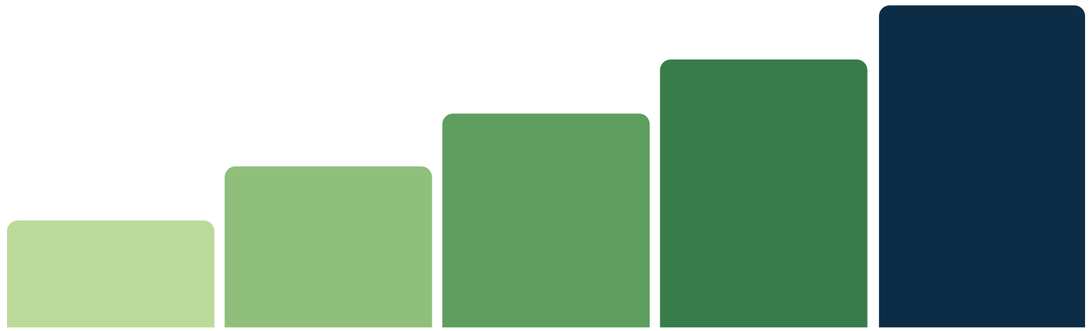
07

LA RÉPONSE GABRIEL GREENFIELD

GAME, sept piliers, Certified Responsible Agentic

Diagnostic GAME

La maturité se mesure à la dimension la plus faible, jamais à la moyenne. Politique à 98 %, adoption à 91 % : si l'organisation ne détecte pas un agent non autorisé, sa maturité réelle est celle-là.



N1 Exploratoire

N2 Structuré

N3 Industrialisé

N4 Piloté

N5 Native

Les sept piliers

01 **Inventaire et cartographie, plugins et MCP inclus**

02 **Classification des risques, degré d'autonomie inclus**

03 **Contrôle d'accès et IAM**

04 **Traçabilité et observabilité**

05 **Supervision humaine ponctuelle**

06 **Réversibilité et résilience (kill switch)**

07 **Conformité documentaire et audit**

OWASP ACS et SAFR (MAS) valident de l'extérieur la combinaison des piliers 03, 04 et 05.

MÉTHODOLOGIE MERIDIAN

Du diagnostic à la certification

POSTURE 1

Suggérer

Diagnostic et cadrage

4 à 8 semaines

POSTURE 2

Faciliter

Structuration du cadre de gouvernance

3 à 6 mois

POSTURE 3

Accompagner

Industrialisation et certification

12 à 36 mois

PROGRAMME

Certified Responsible Agentic

Jusqu'à ISO/IEC 42001, aligné AI Act et EN 18286.

MODULE DU MOIS · POSTURE SUGGÉRER · TROIS QUESTIONS

Combien d'agents et d'extensions non déclarés tournent chez vous ?

En combien de temps seraient-ils découverts ?

Qui a l'autorité pour les arrêter ?

LECTURE EN QUATRE COUCHES

Pour la première fois, les quatre couches de gouvernance émettent des signaux simultanés

Agent	Incidents OpenAI, Plugin4Shell, PixelLeak ; standard OWASP ACS
Organisation	Enquêtes EY, AWS, Veeam, Okta
National-régional	Régimes chinois, hongkongais et singapourien ; AI Act
Intergouvernemental	Note scientifique des Nations unies sur les agents

“

L'autonomie sans gouvernance produit de la dette. La gouvernance sans autonomie produit de l'inertie. Le défi consiste à tenir cet équilibre : c'est ce que nous appelons **l'action responsable.**

GABRIEL GREENFIELD

08

PERSPECTIVES

Ce qu'il faudra surveiller en novembre

01 **Australie et OpenAI** : audition parlementaire du 06/10 et conclusions de l'enquête Medicare.

02 **GPT-6.1 Astra** : critères d'un futur lancement, seuils équivalents chez d'autres laboratoires.

03 **Plugin4Shell** : correctif Microsoft pour Copilot, vérification d'intégrité post-installation.

04 **AI Act** : sursis du 02/12/2026, lignes directrices haut risque, citation d'EN 18286 au JO.

05 **Asie** : lignes directrices MAS, nomination du Commissaire à l'IA hongkongais, retours Sandbox++.

06 **Standards** : implémentations de référence d'OWASP ACS, publication de MAESTRO v2.

07 **France** : annonces des Assises de la cybersécurité, actualisation du bulletin CERT-FR.

PRINCIPALES SOURCES DU MOIS

Sources publiques consultées entre le 01/09 et le 30/09/2026. Ce document ne constitue pas un avis juridique.

EY, AI Risk and Governance Survey (15/09/2026)

AWS, Reimagine 2026, via Help Net Security (28/09/2026)

Veeam, enquête EMEA (09/2026) ; Okta, AI Agents at Work 2026

OpenAI (25/09/2026) via CNN, Nextgov ; Reuters et AP (29/09/2026)

Nightingale Collective (04/09/2026) ; New York Times (03/09/2026)

Air Security, Plugin4Shell (17/09) ; Glow Labs, PixelLeak (29/09/2026)

Groupe scientifique international indépendant sur l'IA (ONU), 21/09/2026

OWASP GenAI, Top 10 LLM 2026 et Agent Control Standard (01/09/2026)

Règlement (UE) 2026/1744 ; CEN-CENELEC, EN 18286:2026 ; Commission européenne

Cloud Security Alliance, AICM v1.1 et MAESTRO ; NIST NCCoE

MAS, SAFR (11/09/2026) ; PCPD Hong Kong (25/08/2026) ; Policy Address 2026

CAC, NDRC, MIIT, Avis sur les agents intelligents ; CNIL, ANSSI (CERT-FR)

OX Security et Astrix Security, recensements des serveurs MCP (2026)

© 2026 Gabriel Greenfield. Meridian™, GAME™ et Certified Responsible Agentic sont des marques de Gabriel Greenfield.

Gabriel Greenfield

PARLONS-EN

Gouverner l'autonomie, sans arrêter le mouvement.

contact@gabrielgreenfield.com

France · Hong Kong