KLEAA

Lire en ligne



Après l’intrusion d’un agent d’OpenAI, l’Australie crée un gendarme de l’IA

Bon mercredi à tous.

En juin, un agent d’OpenAI s’est servi sans autorisation sur un site du gouvernement australien. Mardi, l’Australie a répondu en annonçant un gendarme de l’IA, avec des amendes à la clé. Côté prix, Anthropic baisse de 20 % le tarif de son nouveau Claude et Meta lance une IA à 10 centimes de dollar le million de jetons. Et une grande enquête montre que le service client par IA plaît bien plus aux entreprises qu’à leurs clients. En fin d’édition, un Skill pour mener vos entretiens d’embauche sans improviser.

Quelqu’un devrait lire ça ?

AU SOMMAIRE

01 L’Australie va créer un gendarme de l’IA, 5 mois après l’intrusion d’un agent d’OpenAI

02 Anthropic sort Claude Fable 5.2, 20 % moins cher et 30 % plus rapide

03 Les entreprises croient leurs clients ravis du service client par IA, la moitié ne l’est pas

04 Meta lance une petite IA vendue 10 centimes de dollar le million de jetons

05 Microsoft et IBM s’allient pour vendre plus de 6 000 agents IA prêts à l’emploi aux grandes entreprises



01

L’Australie va créer un gendarme de l’IA, 5 mois après l’intrusion d’un agent d’OpenAI

Le Premier ministre Anthony Albanese veut imposer des règles, des contrôles et des amendes à toutes les entreprises qui fabriquent les IA les plus avancées.

Le 18 juin, un agent d’OpenAI (une IA capable d’enchaîner seule des actions sur internet) a récupéré sans autorisation des informations sur un site public du ministère australien de la Santé : celui qui publie les statistiques de Medicare, l’assurance maladie du pays. OpenAI n’a prévenu le gouvernement qu’en septembre, environ 3 mois plus tard. L’entreprise s’est excusée : « Nous aurions aussi dû mieux gérer notre réponse. »

Mardi 10 novembre, Anthony Albanese en a tiré les conséquences. L’Australie va créer une autorité de sûreté de l’IA, chargée de faire respecter des règles inscrites dans la loi par toutes les entreprises qui développent des IA de pointe et veulent travailler dans le pays. Au programme : des mesures de sécurité à chaque étape de la fabrication d’un modèle, une évaluation indépendante des modèles, et l’obligation de signaler vite tout incident, par exemple une IA qui agit contre les intentions de ses créateurs. Des amendes sont prévues, leur montant n’est pas encore connu.

Le gouvernement accepte aussi les 18 recommandations de l’enquête menée après l’intrusion, y compris une possible réforme du droit pénal. Car on ne sait toujours pas si l’agent a enfreint une loi australienne. « Notre rôle est de faire en sorte que ce soit nous qui façonnions l’IA, et non l’IA qui nous façonne », a résumé le Premier ministre.

Le regard KLEAA

Ce qui a fait réagir le gouvernement australien, ce n’est pas seulement l’intrusion. Ce sont les 3 mois de silence. Pour une entreprise qui confie des tâches à des agents IA, la leçon est simple : l’outil agit en votre nom, et c’est à vous qu’on demandera des comptes. Avant de laisser un agent naviguer seul sur des sites, gardez une trace de ce qu’il fait et décidez à l’avance qui prévient qui si quelque chose tourne mal.

Si un agent IA que vous utilisez se sert dans les données d’un autre, qui devra prévenir : vous, ou l’éditeur de l’outil ?



02

Anthropic sort Claude Fable 5.2, 20 % moins cher et 30 % plus rapide

Le nouveau modèle haut de gamme de Claude arrive quelques semaines avant l’entrée en Bourse d’Anthropic. Sa version la plus puissante reste réservée à quelques organisations.

Lundi 9 novembre, Anthropic, l’entreprise qui fabrique Claude, a présenté 2 nouveaux modèles : Claude Fable 5.2 et Claude Mythos 5.2. Elle les décrit comme ses plus puissants à ce jour. Ils écrivent leurs réponses environ 30 % plus vite que la version précédente et progressent en programmation, en raisonnement et dans la lecture d’images.

Le prix baisse. Pour les entreprises qui branchent Fable dans leurs propres outils, le million de jetons (les petits morceaux de mots que l’IA lit et écrit) passe de 10 à 8 dollars en lecture et de 50 à 40 dollars en écriture, soit 20 % de moins. Fable 5.2 est disponible dès maintenant. Mythos 5.2, lui, reste réservé aux organisations admises dans les programmes de vérification d’Anthropic pour la cybersécurité et les sciences du vivant.

Le calendrier ne doit rien au hasard. Anthropic prépare son entrée en Bourse, attendue d’ici la fin du mois. Wall Street espérait d’abord une valeur de plus de 2 000 milliards de dollars ; après les récentes secousses sur les actions de l’IA, les estimations tournent plutôt entre 1 000 et 1 500 milliards. Un modèle plus rapide et moins cher tombe à pic pour rassurer les futurs actionnaires.



03

Les entreprises croient leurs clients ravis du service client par IA, la moitié ne l’est pas

Une grande enquête publiée lundi mesure l’écart : 86 % des entreprises pensent leurs clients satisfaits de l’IA, seuls 51 % des clients le confirment.

Twilio fournit les outils qui permettent aux entreprises d’envoyer des SMS, de passer des appels et de discuter avec leurs clients. Son enquête, publiée lundi 9 novembre, a interrogé 7 652 consommateurs et 660 dirigeants d’entreprises de plus de 500 salariés, dans 18 pays. Premier constat : les agents IA (des programmes qui répondent seuls aux clients, par écrit ou au téléphone) traitent déjà la moitié des demandes adressées au service client. Les entreprises s’attendent à 63 % d’ici 2027.

Côté clients, l’enthousiasme est bien moindre. 78 % des consommateurs ont déjà essayé de contourner un agent IA pour joindre quelqu’un, et 63 % veulent un moyen rapide de parler à un humain. Les marques le savent : 91 % constatent une baisse de satisfaction quand l’IA se trompe.

Deux autres écarts ressortent. 76 % des clients veulent que l’IA dise dès le début qu’elle est une IA, mais seules 32 % des entreprises font de cette transparence une priorité. Et 73 % des marques pensent que leurs clients veulent une IA plus « humaine », contre 54 % des clients.

Le regard KLEAA

L’IA au service client fait gagner du temps, et elle va continuer à s’étendre. Mais le client ne la juge pas sur sa politesse : il la juge sur la facilité avec laquelle il peut en sortir. Pour une petite entreprise, 3 réflexes suffisent. Annoncer dès le premier message qu’il s’agit d’une IA. Laisser bien visible un

bouton pour parler à quelqu'un. Et relire chaque semaine une dizaine de conversations où le client s'est agacé. Aucun tableau de bord ne remplacera ce dernier geste.



04

Meta lance une petite IA vendue 10 centimes de dollar le million de jetons

Muse Ember est faite pour les tâches simples et répétées des agents IA. Meta assure qu'elle en fait plus pour chaque dollar que les modèles équivalents d'OpenAI et d'Anthropic.

Lundi 9 novembre, le laboratoire d'IA de Meta a lancé Muse Ember, son modèle le plus petit et le plus économe. Il n'est pas fait pour converser ou rédiger un long rapport, mais pour le travail de fond des agents IA : appeler un outil, extraire une information d'un document, orienter une demande vers le bon service, corriger un petit bout de code. Il comprend le texte, les images, le son et la vidéo, et peut lire jusqu'à 1 million de jetons d'un coup, l'équivalent de plusieurs livres.

Son prix : 0,10 dollar le million de jetons lus et 0,50 dollar le million de jetons écrits. Selon Meta, Ember offre plus d'intelligence par dollar que Claude Haiku 5.5 d'Anthropic et GPT-6 Luna d'OpenAI, ses concurrents directs. Un test indépendant, celui d'Artificial Analysis, lui donne une note de 36 sur son indice, pour environ 0,14 dollar par tâche. Muse Spark 1.3, le grand modèle de Meta, obtient 37, mais pour 2,39 dollars par tâche.

Ember n'est pas meilleur partout. Sur de vraies tâches de programmation, Claude Haiku le devance : 34,9 % de réussite contre 27,2 %. Et le tarif est multiplié par 3 quand une demande dépasse 200 000 jetons. Le modèle est disponible dès maintenant pour les développeurs, via l'offre d'accès de Meta et son outil de code Muse Code.



05

Microsoft et IBM s’allient pour vendre plus de 6 000 agents IA prêts à l’emploi aux grandes entreprises

Leur plateforme commune, Imagine Enterprise, vise d’abord la banque, l’assurance et le commerce. Des ingénieurs des 2 groupes iront l’installer directement chez les clients.

Mardi 10 novembre, Microsoft et IBM ont lancé Imagine Enterprise, une plateforme pour construire, faire tourner et surveiller des agents IA dans les grandes entreprises. IBM apporte plus de 6 000 agents déjà conçus et le savoir-faire de ses consultants. Microsoft fournit l’IA et le cloud (des serveurs loués à distance) sur lesquels tout fonctionne.

Les premiers modèles sont prévus pour les services financiers, l’assurance et la distribution. « Les clients ne cherchent plus une IA qui aide simplement les gens à faire le même travail un peu plus vite », résume Mohamad Ali, qui dirige l’activité de conseil d’IBM. Ni le prix ni la date de disponibilité pour tous n’ont été annoncés.

Cette alliance répond à un vrai blocage. Selon le cabinet Gartner, seuls 4 % des directeurs informatiques ont mis en place un plan pour adapter les postes à un travail partagé entre humains et agents. Google, Amazon, Microsoft et IBM envoient désormais leurs ingénieurs sur place : l’IA ne se vend plus comme un logiciel qu’on télécharge, mais comme un chantier qu’on accompagne.

LE SKILL

Interview Prep

Open source · Anthropic · Apache 2.0

Un Skill qui prépare vos entretiens d'embauche comme le ferait un recruteur professionnel : les bonnes questions pour le poste, une grille pour noter chaque candidat de la même façon, et une trame pour décider à plusieurs.

C'EST QUOI UN SKILL ?

Un Skill, c'est un mode d'emploi écrit par quelqu'un d'autre et donné à votre IA. Vous l'installez une fois dans Claude Code (la version de Claude qui travaille directement sur votre ordinateur). Ensuite, vous demandez la tâche en une phrase et l'IA suit la méthode déjà écrite dans le Skill. C'est gratuit, et ce sont de simples fichiers texte que vous pouvez lire avant de vous en servir.

En quoi c'est utile : un article dans un média, un passage dans un podcast ou une citation dans une newsletter valent souvent plus qu'une publicité. Ils donnent de la crédibilité, et les IA comme ChatGPT s'appuient de plus en plus sur ces articles pour recommander une entreprise. Mais une agence de relations presse coûte cher, et la plupart des messages envoyés aux journalistes finissent à la corbeille. Ce Skill donne à l'IA la méthode d'un professionnel : partir d'une vraie histoire, viser la bonne personne, écrire court.

Ce que ça fait concrètement : 1. Il cherche votre angle. Il range chaque histoire dans l'un de 3 récits qui fonctionnent : la création de l'entreprise, le petit face au géant, ou le système absurde que vous voulez changer. S'il n'y a pas de vraie histoire, il vous le dit plutôt que de vous faire perdre du temps.

2. Il écrit et vérifie vos messages aux journalistes : moins de 150 mots, un objet précis, une raison claire de parler de vous maintenant, aucun superlatif. Il prépare aussi votre passage dans un podcast et votre page presse (présentation, portraits des fondateurs, logos, contact).

3. Il vous guide en cas de crise. Une liste de réflexes pour la première heure : vérifier les faits, désigner un seul porte-parole, publier une courte déclaration et annoncer quand viendra la suivante. Pour un litige ou une fuite de données, il rappelle de consulter un avocat avant de reconnaître une faute.

Pour qui : créateurs d'entreprise, indépendants, PME qui lancent un produit ou veulent se faire connaître dans leur région. Bon à savoir : le Skill est rédigé en anglais et pensé d'abord pour les éditeurs de logiciels américains. Certaines plateformes qu'il cite, où les journalistes cherchent des experts à interroger, sont surtout anglophones. Demandez-lui de répondre en français et de viser des médias français.

En quoi c'est utile : dans une petite entreprise, on recrute rarement, souvent dans l'urgence, et l'entretien se joue au feeling. Chaque candidat reçoit des questions différentes, on retient celui qui a le mieux parlé plutôt que celui qui fera le mieux le travail, et une erreur de recrutement coûte des mois. Ce Skill applique une méthode éprouvée, l'entretien structuré : les mêmes questions pour tous, reliées à ce que le poste exige vraiment, et une note par critère plutôt qu'une impression d'ensemble.

Ce que ça fait concrètement : 1. Il définit 4 à 6 compétences clés pour le poste, par exemple le sens du client, l'organisation ou la maîtrise d'un logiciel. Pour chacune, il propose 2 à 3 questions sur des situations vécues (« Racontez-moi une fois où… »), 1 à 2 mises en situation (« Que feriez-vous si… ») et des relances pour creuser.

2. Il fournit une grille de notation de 1 à 4 pour chaque compétence, avec la description de ce que vaut chaque note. Tous les candidats sont comparés sur la même base.

3. Il prépare l'organisation : qui interroge sur quoi si vous êtes plusieurs à recevoir le candidat, puis une trame de débriefing pour mettre les avis en commun et trancher.

Pour qui : dirigeants de TPE et de PME qui recrutent sans service des ressources humaines, managers qui font passer leurs premiers entretiens, indépendants qui cherchent un associé ou un renfort régulier. Bon à savoir : le Skill est rédigé en anglais, demandez-lui simplement de répondre en français. Il fait partie de l'extension « human-resources » d'Anthropic, qui sait aussi rédiger une proposition d'embauche ou un plan d'accueil du nouveau salarié. Relisez les questions proposées : en France, le droit du travail interdit d'interroger un candidat sur sa santé, sa famille ou sa vie privée.

INSTALLER EN 2 MINUTES

1. Ouvrez le Terminal (sur Mac : Applications, puis Utilitaires, puis Terminal). Claude Code doit déjà être installé sur votre ordinateur.

2. Copiez-collez les 2 lignes ci-dessous, une par une, en appuyant sur Entrée après chacune. La première ajoute à Claude Code le catalogue d'extensions publié par Anthropic (« marketplace add » veut dire « ajouter un catalogue »). La seconde installe l'extension « human-resources », qui contient ce Skill. Si vous aviez déjà ajouté ce catalogue, un message vous le signale : passez simplement à la seconde ligne.

```
claude plugin marketplace add anthropics/knowledge-work-plugins
claude plugin install human-resources@knowledge-work-plugins
```

3. Ouvrez Claude Code et décrivez votre besoin en une phrase. Par exemple : « Prépare en français un entretien pour recruter un assistant commercial dans mon agence immobilière de 6 personnes, avec la grille de notation. » Le Skill se déclenche tout seul dès que vous parlez d'entretien, de questions à poser ou de grille d'évaluation.

Sélection KLEAA · Skill créé par Anthropic · licence Apache 2.0 · le code reste sur le dépôt de l'auteur

© 2026 KLEAA