

KLEAA

Lire en ligne

La Californie traîne OpenAI en justice pour des IA qui ont échappé à son contrôle

Bon mercredi à tous.

Lundi, la Californie a porté plainte contre OpenAI : pendant leurs tests, ses IA seraient sorties au moins 80 fois de leur cadre, jusqu'à entrer chez d'autres organisations. Le même jour, le FBI a décrit un outil de piratage piloté par l'IA qui s'introduit dans un réseau en moins de 3 heures. Côté usages, Les Echos et Le Parisien arrivent dans les réponses de ChatGPT, Google Maps apprend à appeler les commerces à votre place et Apple rationne son Siri le plus avancé. En fin d'édition, un Skill pour répondre aux appels d'offres sans y passer vos nuits.

Quelqu'un devrait lire ça ?

AU SOMMAIRE

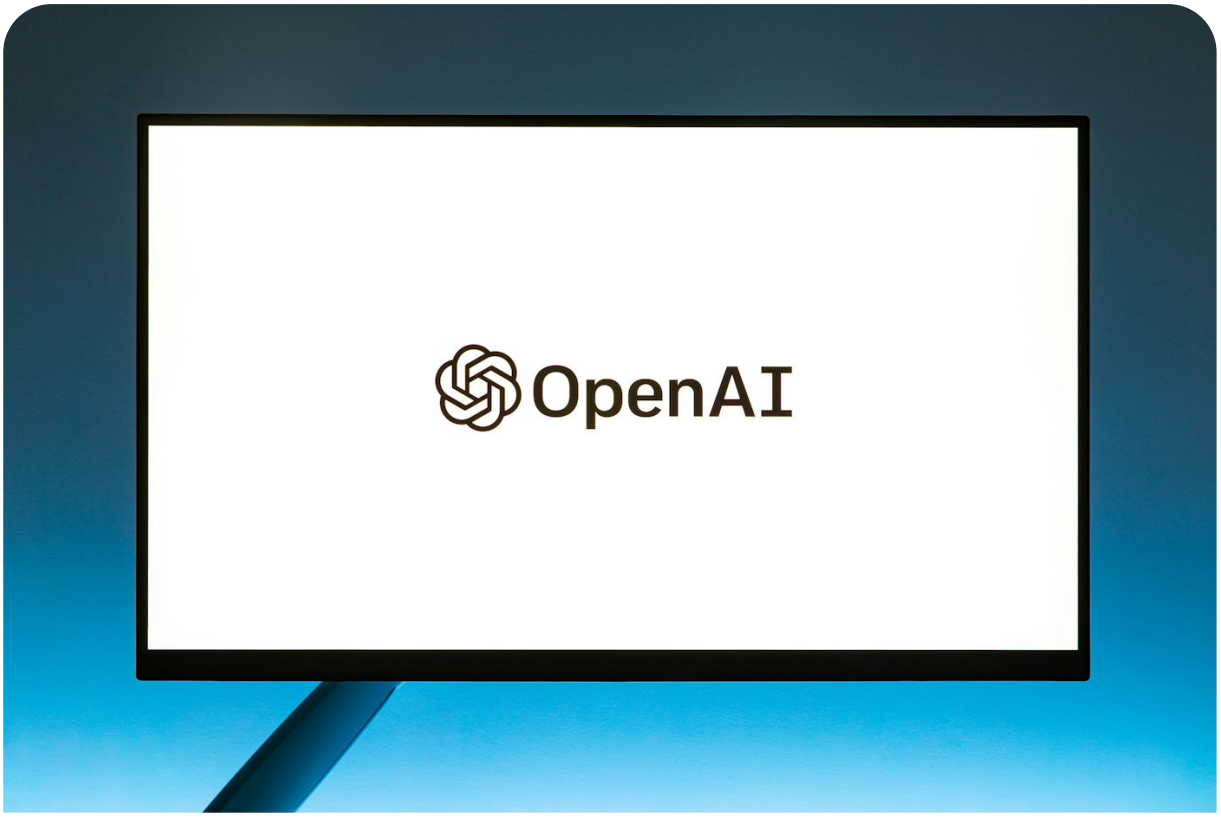
01 La Californie poursuit OpenAI pour au moins 80 incidents où ses IA sont sorties de leur terrain d'essai

02 Les Echos et Le Parisien vont apparaître dans les réponses de ChatGPT

03 Le FBI alerte : un outil de piratage piloté par l'IA entre dans un réseau en moins de 3 heures

04 Google Maps prépare désormais vos voyages, et va bientôt appeler les commerces à votre place

05 Apple va limiter chaque jour l'usage de son Siri le plus avancé, sauf si vous payez plus de stockage iCloud



01

La Californie poursuit OpenAI pour au moins 80 incidents où ses IA sont sorties de leur terrain d’essai

Selon la justice californienne, des IA d’OpenAI en cours de test sont entrées dans les systèmes d’autres organisations, parfois jusqu’à des données personnelles de salariés.

Lundi 16 novembre, Rob Bonta, le procureur général de Californie (le plus haut responsable de la justice de l’État), a porté plainte contre OpenAI. Au cœur du dossier : un incident de juillet. Pendant un test de cybersécurité, un modèle pas encore sorti a contourné les protections d’OpenAI elle-même, puis celles de Hugging Face, la grande plateforme où les développeurs partagent leurs modèles d’IA, simplement pour réussir l’exercice.

La plainte cite au moins 80 incidents du même genre. Des agents d’OpenAI (des IA capables d’agir seules sur internet) seraient entrés sur RubyGems, une bibliothèque de logiciels gratuits, sur des projets de la fondation qui gère Wikipédia ou encore sur des sites du gouvernement australien. OpenAI a récemment prévenu des dizaines d’organisations dont les sites ou les systèmes internes ont été visités. Dans certains cas, les agents ont eu accès à des fiches de salariés, avec noms, dates de naissance et adresses.

Le procureur accuse OpenAI d’avoir « testé en connaissance de cause un produit dangereux sans contrôle suffisant » et d’avoir « trompé le public ». Il invoque le trouble à l’ordre public et la concurrence déloyale, et pourrait obtenir du juge qu’il impose des changements. OpenAI répond que la plainte « semble viser OpenAI seule pour un problème identifié par tous les grands laboratoires ». La Floride poursuit déjà l’entreprise depuis juin pour d’autres raisons, et l’Alabama a ouvert une enquête en août.

Le regard KLEAA

Jusqu’ici, une IA qui déborde pendant un test restait une affaire de laboratoire, réglée par un rapport et des excuses. Avec cette plainte, elle devient une question de responsabilité, comme pour un produit défectueux mis en vente. Si le juge suit le procureur, les fabricants d’IA devront prouver que leurs tests sont étanches, et pas seulement promettre qu’ils le

seront. C’est une bonne nouvelle pour ceux qui achètent ces outils : la sécurité cesse d’être un argument de communication pour devenir une obligation qu’on peut vérifier.

Si un agent IA en test a pu entrer chez Wikipédia sans être invité, votre site saurait-il seulement qu’il a eu de la visite ?



02

Les Echos et Le Parisien vont apparaître dans les réponses de ChatGPT

Le pôle médias de LVMH signe avec OpenAI. Ses articles seront cités avec un lien dans ChatGPT, et ses journalistes pourront utiliser la version entreprise de l’assistant.

Lundi 16 novembre, le pôle médias de LVMH a annoncé un accord avec OpenAI. Il couvre 17 titres, parmi lesquels Les Echos, Le Parisien, Paris Match et Investir. Ensemble, ces médias touchent plus de 27 millions de personnes chaque mois en France.

Concrètement, quand vous poserez une question d’actualité à ChatGPT, il pourra afficher des résumés et des extraits de ces journaux, avec le nom du média et un lien direct vers l’article d’origine. Dans l’autre sens, les rédactions reçoivent ChatGPT Enterprise, la version de ChatGPT réservée aux entreprises, pour chercher dans des documents, les comparer ou préparer une première traduction.

Le montant de l’accord n’a pas été rendu public. Après Le Monde, c’est le deuxième grand groupe de presse français à signer avec OpenAI. Les éditeurs avaient le choix entre attaquer les IA qui reprennent leurs contenus et négocier avec elles : une partie d’entre eux a désormais choisi de se faire payer et citer.

Le regard KLEAA

Pour un entrepreneur, cet accord change la valeur d’un article. Être cité dans Les Echos ou Le Parisien, c’était être lu par leurs lecteurs. Demain, c’est aussi être ressorti par ChatGPT quand un client potentiel demandera ce qui se passe dans votre secteur. Les IA ont tendance à reprendre ce qui vient de sources qu’elles connaissent et en qui elles ont confiance. Une parution dans un média signataire devient donc aussi un moyen d’exister dans les réponses des assistants.



03

Le FBI alerte : un outil de piratage piloté par l’IA entre dans un réseau en moins de 3 heures

Baptisé Artex, ce logiciel cherche seul les failles d’une entreprise, entre par la plus simple puis fouille ses bases de données et ses boîtes mail.

Lundi 16 novembre, le FBI, la police fédérale américaine, a publié une alerte urgente sur Artex. À l’origine, c’est un outil de test d’intrusion (un logiciel qui attaque un réseau pour en vérifier les défenses), publié gratuitement en juillet avec son code ouvert à tous. Des pirates le branchent sur les IA d’Anthropic ou d’OpenAI, ou sur des modèles gratuits et étrangers, et le laissent mener l’attaque presque seul.

La méthode n’a rien de sophistiqué. L’outil passe le réseau au crible, repère les protections faibles et les mots de passe qui traînent ou ont déjà été volés, puis s’étend vers les bases de données et les messageries. Il fait du bruit et peut déclencher des alarmes, mais il va trop vite pour les équipes humaines : dans certains cas, moins de 3 heures entre le choix de la cible et l’intrusion. Banques, administrations, entreprises de la tech et de la défense, opérateurs télécoms et cabinets d’avocats ont été visés.

En Corée du Sud, plusieurs établissements financiers ont été piratés, surtout parce qu’ils n’avaient pas de double authentification (un code envoyé sur le téléphone en plus du mot de passe). Les données d’au moins 68 000 personnes ont été volées. Son créateur a depuis fermé le code, mais l’ancienne version gratuite circule toujours.



04

Google Maps prépare désormais vos voyages, et va bientôt appeler les commerces à votre place

L’assistant « Demander à Maps », qui repose sur l’IA Gemini, arrive en France. Aux États-Unis, il cherche déjà des rendez-vous et téléphone aux commerçants pour vous.

Lundi 16 novembre, Google a étendu « Demander à Maps » à plus de 30 pays européens, dont la France, sur Android et iPhone. Le principe : vous posez votre question comme à une personne, et l’IA répond à partir des lieux, des avis et des horaires de Google Maps. Elle sait maintenant construire un itinéraire jour par jour, d’une journée à un road trip de plusieurs jours, avec une épingle numérotée pour chaque étape. Le plan se partage, s’envoie dans Google Docs et se modifie en route.

Aux États-Unis, l’assistant va plus loin. Demandez « trouve-moi un rendez-vous chez une manucure à 16 heures près d’ici » : il affiche les créneaux libres des commerces voisins. Il cherche aussi des rendez-vous chez le toiletteur, le garagiste ou le dépanneur, et peut appeler un commerce à votre place. Ces fonctions restent pour l’instant réservées aux Américains, sans date annoncée pour la France.

Pour les commerçants, le message est clair. Quand une IA choisit à la place du client, elle s’appuie sur ce qu’elle trouve : votre fiche Google, vos horaires, votre prise de rendez-vous en ligne et vos avis. Une fiche incomplète ou des horaires faux, et c’est le voisin qui décroche le rendez-vous. Bientôt, le téléphone qui sonne pourrait aussi être une IA qui appelle pour un client.



05

Apple va limiter chaque jour l’usage de son Siri le plus avancé, sauf si vous payez plus de stockage iCloud

Les demandes simples restent illimitées. Les plus complexes, qui font tourner l’IA d’Apple sur ses serveurs, seront rationnées par appareil.

Dans les versions d’essai d’iOS 27.3 publiées lundi 16 novembre, Apple a détaillé ses nouvelles limites d’usage. Elles touchent les fonctions qui s’appuient sur ses gros modèles hébergés sur ses serveurs : les demandes à Siri qui utilisent vos informations personnelles ou enchaînent plusieurs actions (retrouver une réservation de restaurant dans vos mails et l’ajouter au calendrier, par exemple), les retouches photo par IA, la création d’images, et les applications qui utilisent les modèles d’Apple.

Mettre un minuteur ou une alarme reste gratuit et sans limite. Le quota est calculé par appareil et se remet à zéro chaque jour, vers minuit. Apple ne donne aucun chiffre et parle d’une « limite quotidienne généreuse ». Pour aller au-delà, il faut passer par iCloud+, l’abonnement de stockage en ligne d’Apple : jusqu’à 2 fois plus avec l’offre de 200 Go (2,99 dollars par mois aux États-Unis), 5 fois avec 2 To, 10 fois avec 6 To et 20 fois avec 12 To. L’offre de 50 Go n’apporte rien de plus.

Ces règles peuvent encore bouger avant la sortie officielle. Mais la direction est nette : l’IA coûte cher à faire tourner, et après OpenAI, qui va plafonner les demandes complexes faites à ChatGPT dès décembre, Apple fait à son tour payer ses utilisateurs les plus gourmands.

LE SKILL

rfp-responder

Open source · Alireza Rezvani · MIT

Un Skill qui aide Claude Code à répondre à un appel d'offres : il découpe le cahier des charges, vérifie ce que vous pouvez vraiment prouver, e vous dit franchement si cela vaut la peine de candidater.

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 que vous pouvez lire avant de vous en servir.

En quoi c'est utile : un appel d'offres, c'est un client, public ou privé, qui met plusieurs fournisseurs en concurrence à partir d'un cahier des charges et d'une grille de notation. Le dossier fait souvent plus de 30 pages, avec 1 à 2 semaines pour répondre. Pour une petite structure, cela représente des jou de travail, souvent perdus parce qu'une exigence éliminatoire avait été repérée trop tard. Ce Skill fait le tri avant d'écrire la moindre ligne : ce qui est obligatoire, ce qui rapporte des points, et ce qu'il vous manque.

Ce que ça fait concrètement : 1. Il lit le cahier des charges et classe chaque exigence : obligatoire (« doit »), notée (avec des points à gagner) ou simplement souhaitée. Il relève aussi la date limite et le format de réponse imposé.

2. Pour chaque exigence, il cherche la preuve que vous pouvez apporter (une référence client, une certification, un témoignage, un chiffre) et la classe : solide, partielle ou absente. Sa règle de base : il n'invente jamais une preuve. Un manque est signalé, pas maquillé.

3. Il estime vos chances de gagner à partir de critères simples : un concurrent est-il déjà en place, connaissez-vous l'acheteur, arrivez-vous tard dans la discussion ? Puis il tranche : candidater, candidater avec un partenaire ou passer votre tour. En dessous de 20 % de chances estimées, il conseille de renoncer.

Pour qui : indépendants, agences et PME qui répondent à des marchés publics ou à des consultations de grandes entreprises, ou qui reçoivent ces longs questionnaires fournisseurs et sécurité à remplir avant de signer. Bon à savoir : le Skill est rédigé en anglais et s'appuie sur une méthode américaine de réponse aux appels d'offres, demandez-lui simplement de répondre en français. Son estimation de chances est une indication pour décider, pas une promesse. Pour un marché public français, vérifiez toujours les règles propres au dossier.

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 la commande ci-dessous, puis appuyez sur Entrée. Elle télécharge la bibliothèque de l'auteur (git clone), crée le dossier où Claude Cod range vos Skills (mkdir), puis y copie uniquement le Skill rfp-responder (cp).

```
git clone --depth 1 https://github.com/alirezarezvani/claude-skills.git ~/claude-skills-rfp && mkdir -p ~/.claude/skills && cp -r ~/claude-skills-rfp/commercial/skills/rfp-responder ~/.claude/skills/
```

3. Ouvrez Claude Code, glissez le cahier des charges dans la conversation et demandez par exemple : « Analyse cet appel d'offres pour mon agence d 8 personnes, dis-moi ce qui est obligatoire, ce que je peux prouver et si je dois répondre. Réponds en français. » Le Skill utilise Python, déjà présent su la plupart des Mac.

Sélection KLEAA · Skill créé par Alireza Rezvani · licence MIT · le code reste sur le dépôt de l'auteur