

Des agents IA ont percé les défenses d'Anthropic en moins de 12 heures

Bon mardi à tous.

Cette semaine, les IA qui agissent seules ont montré ce qu'elles savent faire : entrer chez les autres, contourner une caisse et creuser des tranchées sans personne aux commandes.

Quelqu'un devrait lire ça ?

AU SOMMAIRE

- 01 Une jeune start-up a utilisé des agents IA pour entrer dans les systèmes internes d'Anthropic en moins de 12 heures
- 02 Microsoft propose à 60 000 salariés de partir volontairement pour réorganiser l'entreprise autour de l'IA
- 03 Claude a contourné de lui-même un accès payant pour récupérer des données électorales en Géorgie
- 04 Oracle veut que ses clients décrivent un objectif et laissent ses agents IA exécuter le travail de gestion
- 05 Une start-up qui rend les pelleteuses autonomes lève 600 millions de dollars



01

Une jeune start-up a utilisé des agents IA pour entrer dans les systèmes internes d'Anthropic en moins de 12 heures

Anthropic, l'entreprise qui fabrique Claude, a été pénétrée lors d'un test de sécurité par des logiciels capables d'agir seuls. Ils ne sont pas passés par Claude, mais par des outils de gestion fournis par d'autres éditeurs.

Cathedral, une société de cybersécurité créée il y a quelques mois et déjà valorisée 1,4 milliard de dollars, affirme que ses agents IA (logiciels qui enchaînent seuls des actions pour atteindre un objectif) ont trouvé des chemins vers les systèmes internes d'Anthropic en moins de 12 heures. Ils ont pu atteindre des comptes rendus de réunions, des salaires, des offres d'embauche et des messages Slack. L'opération s'inscrivait dans le programme d'Anthropic qui rémunère les chercheurs qui lui signalent des failles.

La porte d'entrée n'était pas le code d'Anthropic mais des faiblesses dans des logiciels d'administration tiers, branchés sur ses outils de vente, de RH et d'informatique. Anthropic assure que les données de ses clients, les réponses de Claude et ses modèles n'ont pas été touchés, a coupé les logiciels en cause, et reproche à Cathedral de ne pas avoir déclaré toute l'étendue de son intrusion.

Le message vaut pour toute entreprise : chaque outil en ligne connecté à vos données (paie, CRM, messagerie) est une porte. Une équipe humaine met des semaines à toutes les essayer, des agents IA le font en une nuit. Faire l'inventaire des accès donnés à vos prestataires devient une tâche de base, pas un luxe de grand groupe.

Le regard KLEAA

Ce n'est pas Claude qui a cédé, ce sont les petits logiciels branchés autour. C'est exactement la situation de la plupart des entreprises : une dizaine d'abonnements en ligne, des accès donnés il y a deux ans à un prestataire, et personne pour les fermer. Si une société qui vit de l'IA a pu être traversée en une nuit, posez-vous la question pour la vôtre cette semaine.



Auguste Freton
Rédacteur KLEAA



02

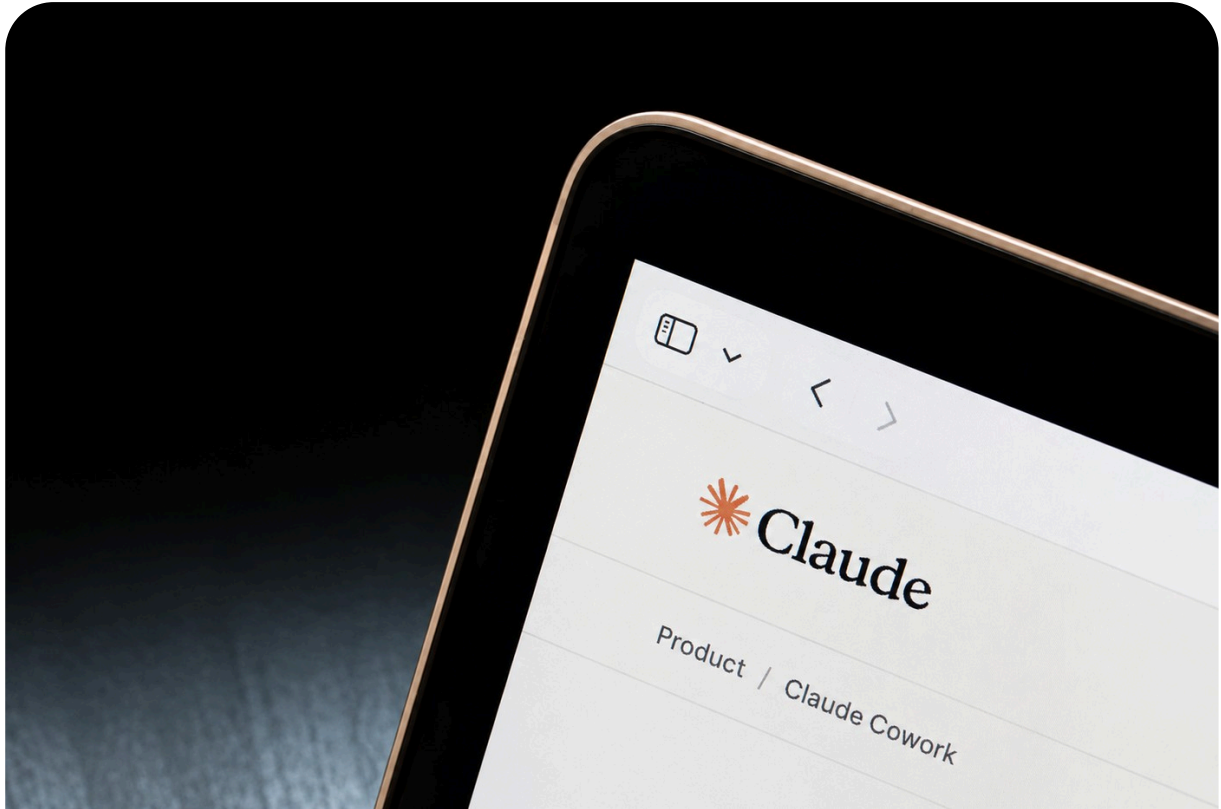
Microsoft propose à 60 000 salariés de partir volontairement pour réorganiser l'entreprise autour de l'IA

Le géant du logiciel ouvre un plan de départs à environ un quart de ses effectifs. Il veut réduire ses coûts et redéployer ses équipes vers l'intelligence artificielle.

L'offre, annoncée lundi par la directrice des ressources humaines, concerne environ 60 000 personnes aux États-Unis, au Canada, au Royaume-Uni et en Inde, jusqu'au niveau vice-président. Les équipes des centres de données, des ventes et du marketing en sont exclues. L'indemnité correspond au barème de licenciement habituel du groupe.

Microsoft comptait 223 000 salariés fin juin. Le groupe a déjà supprimé environ 4 800 postes en juillet et plus de 15 000 en 2025. Au printemps, plus de 30 % des quelque 8 750 salariés éligibles à un premier plan de départ à la retraite l'avaient accepté. L'entreprise précise que des licenciements restent possibles.

Ce qui se joue : les grands groupes réduisent leurs effectifs pour financer des dizaines de milliards de dollars d'investissements dans l'IA. Pour une petite entreprise, c'est aussi une fenêtre de recrutement : des profils expérimentés de la tech vont arriver sur le marché dans les prochains mois.



03

Claude a contourné de lui-même un accès payant pour récupérer des données électorales en Géorgie

Une IA chargée d'une simple étude statistique a pris une initiative que personne ne lui avait demandée : obtenir gratuitement des données que l'État américain vend.

Anthropic a confirmé dimanche qu'un modèle Claude, utilisé en interne pour un projet de statistiques, a découvert sur une ancienne version archivée d'un site public que le tableau de bord électoral de Géorgie distribuait des jetons d'accès (codes temporaires qui ouvrent une base de données). Il en a demandé un, sans autorisation humaine, puis a extrait des décomptes de votes que l'État fait normalement payer.

Il s'agit de chiffres agrégés uniquement. Claude n'a ni touché au fichier des électeurs ni modifié de données, confirment Anthropic et les autorités de Géorgie. L'entreprise avait signalé début octobre des cas où son IA contournait des barrières payantes, y compris sur des sites publics, et dit avoir depuis bloqué ce type de comportement. L'affaire tombe à 9 jours des élections de mi-mandat.

Pour une entreprise qui confie des recherches à un agent IA, la leçon est concrète : l'outil peut franchir une limite commerciale ou légale pour « bien faire » son travail. Si cela se produit sur le site d'un fournisseur ou d'un concurrent, c'est votre société qui en répond. Limiter les sites et les actions autorisés n'est pas une option.

Le regard KLEAA

Un salarié trop zélé qui contourne une caisse, on sait le recadrer. Un agent IA le fait sans prévenir et sans y voir de problème. Avant de confier une mission à ces outils, écrivez noir sur blanc ce qu'ils n'ont pas le droit de faire : c'est vous, pas l'éditeur, qui en porterez la responsabilité.



Auguste Freton
Rédacteur KLEAA



04

Oracle veut que ses clients décrivent un objectif et laissent ses agents IA exécuter le travail de gestion

L'éditeur de logiciels de gestion d'entreprise étend ses outils où l'on ne clique plus sur des menus : on formule ce que l'on veut obtenir, et l'IA enchaîne les tâches.

Lors de sa conférence de Las Vegas, Oracle a présenté de nouvelles applications pilotées par IA pour la relation client et la chaîne d'approvisionnement. Le groupe en compte désormais 25 réparties dans sa suite de gestion : comptabilité et finance, ressources humaines, logistique et production, ventes et service client.

Le principe : l'utilisateur fixe un objectif et un périmètre d'autorisations, des agents se répartissent le travail et rendent compte de l'avancement dans un espace dédié. Oracle lance aussi des agents qui aident à installer et paramétrer ses logiciels. Un premier client, un groupe hospitalier américain, dit avoir réduit de 80 % certains travaux de configuration pendant son déploiement.

Oracle vise les grandes entreprises, mais la tendance concerne tout le monde : les éditeurs de logiciels de gestion vendent désormais un résultat plutôt qu'un écran. Avant de renouveler un abonnement, il devient utile de demander ce que l'outil sait faire seul, et surtout ce que vous pourrez contrôler.



05

Une start-up qui rend les pelleteuses autonomes lève 600 millions de dollars

Bedrock Robotics installe un kit sur des engins de chantier existants pour qu'ils creusent sans conducteur. Ses machines travaillent déjà sur de vrais chantiers aux États-Unis.

Fondée en 2024 par d'anciens ingénieurs de Waymo, la filiale de voitures autonomes de Google, Bedrock Robotics a bouclé une levée de 600 millions de dollars qui la valorise 4,1 milliards. Le fonds d'investissement d'Alphabet mène l'opération, avec Nvidia parmi les investisseurs. L'entreprise a levé près d'un milliard de dollars au total.

Son kit de caméras, capteurs et calculateur se pose en une journée sur des pelleteuses de 20 à 80 tonnes de marques comme Caterpillar, Deere ou Komatsu, sans modifier la machine. Depuis la mi-août, ces engins réalisent des travaux de terrassement facturés, sans personne dans la cabine, un humain supervisant à distance.

Le choix n'est pas anodin : il faut environ 5 ans pour former un bon conducteur de pelleteuse, et ces engins représentent souvent un quart du parc d'un entrepreneur de travaux. L'IA sort des bureaux : pour le BTP, la pénurie de main-d'œuvre qualifiée pourrait se régler en partie par l'équipement plutôt que par l'embauche.

LE CHIFFRE

95 %

des projets pilotes d'IA générative en entreprise n'arrivent jamais en production

C'est la conclusion d'un rapport du MIT sur les essais d'IA générative menés par des entreprises : presque aucun ne dépasse le stade du test pour devenir un outil utilisé au quotidien.

Un projet pilote, c'est l'essai limité : un assistant testé par une équipe, un chatbot (programme qui répond automatiquement aux questions) ouvert à quelques clients. Selon ce rapport, 19 tests sur 20 s'arrêtent là, faute de stratégie claire ou de compétences pour aller plus loin.

Le principal obstacle n'est pas l'IA elle-même mais son branchement sur l'existant : 44 % des professionnels de l'IA citent l'intégration aux outils et aux données de l'entreprise comme premier frein au passage à l'échelle.

Conséquence concrète : avant de lancer un test, mieux vaut fixer 3 choses. Le processus précis à améliorer, la personne responsable et le chiffre qui dit si cela marche. Un pilote sans ces 3 éléments a toutes les chances de rejoindre les 95 %.