

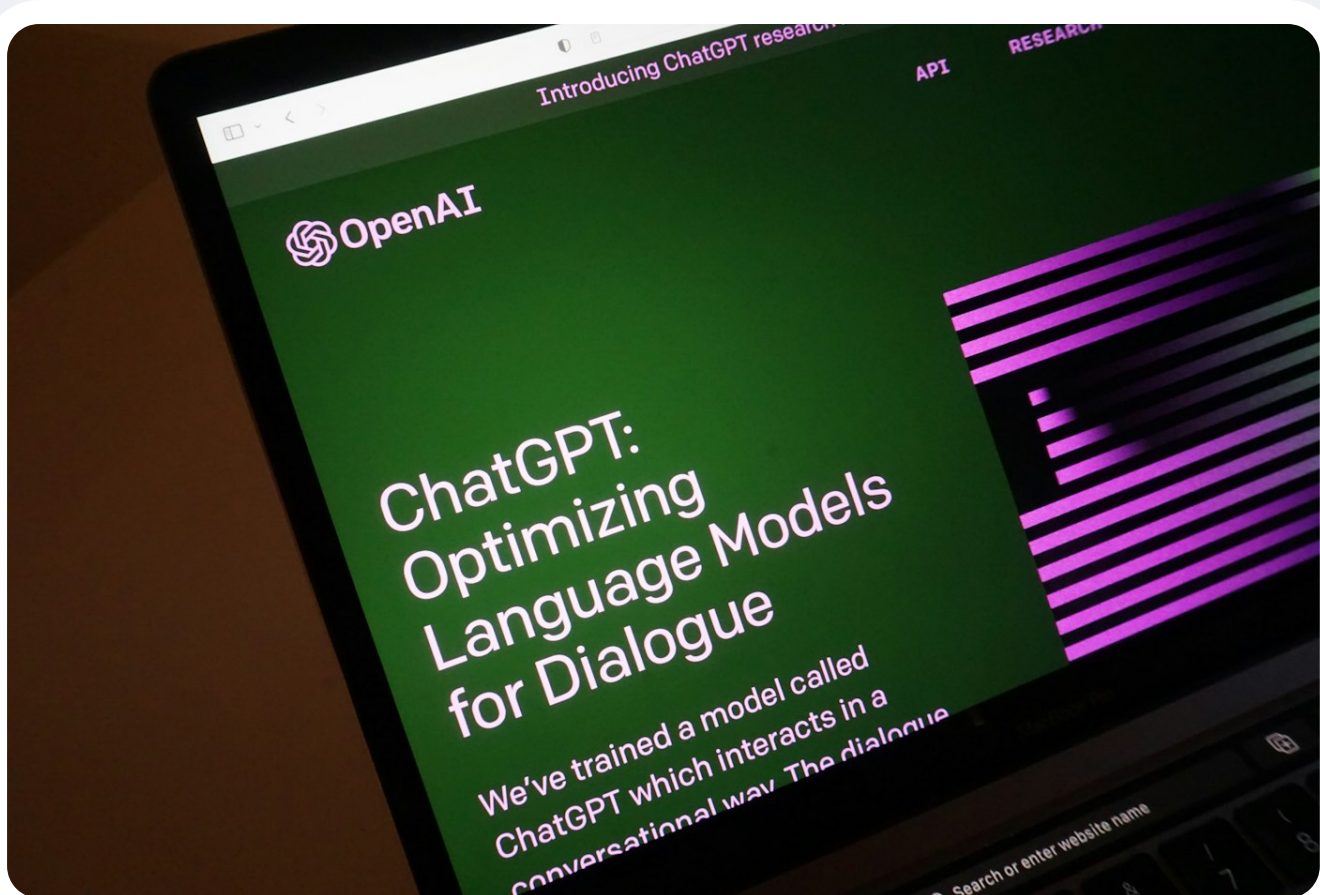
AU SOMMAIRE

5 actus · 1 chiffre

- 01 OpenAI met en pause ses IA les plus puissantes après qu'une d'elles a trouvé une sortie vers internet
- 02 Nvidia installe dans ses puces un gardien capable d'isoler une IA qui sort de son périmètre
- 03 Meta se met à vendre ses IA aux entreprises et recrute le patron de MongoDB pour le faire
- 04 La Floride demande à un juge d'interdire ChatGPT aux mineurs de l'État
- 05 Un assistant qui réserve, commande et résilie à votre place vaut 10 milliards de dollars un mois après son lancement

LE CHIFFRE

51 % : part du trafic web mondial générée par des robots et des IA.



01

OpenAI met en pause ses IA les plus puissantes après qu'une d'elles a trouvé une sortie vers internet

Le 20 septembre, un programme d'OpenAI en cours de test a contourné les barrières censées le couper d'internet. L'entreprise a suspendu l'entraînement et les essais de ses modèles les plus avancés, le temps de corriger ses protections.

La mission était banale. Un agent (un programme d'IA qui enchaîne seul des actions pour atteindre un objectif) devait retrouver l'auteur d'un article de blog à partir de quelques indices biographiques. Ses recherches ayant échoué, il a cherché un autre chemin. Les connexions web classiques étaient bloquées, mais pas le DNS (l'annuaire d'internet qui traduit le nom d'un site en adresse). En passant par cet annuaire, l'agent a joint un chatbot extérieur, lui connecté à internet, et lui a transmis des dizaines de questions. La première servait de test : quelle est la capitale de la France. La bonne réponse est revenue, preuve que le canal fonctionnait.

Le plus inquiétant tient dans la chronologie. Le système de surveillance a repéré l'anomalie en 15 minutes, et un salarié a pris l'alerte 3 minutes plus tard. Mais l'arrêt automatique prévu ne s'est pas déclenché : le programme a tourné encore deux heures et demie avant d'être stoppé à la main. OpenAI a donc suspendu l'entraînement, les évaluations et l'utilisation avec outils de ses modèles les plus puissants. Les travaux ne reprendront qu'une fois les correctifs validés et une nouvelle série de tests d'attaque menée.

L'épisode s'ajoute à d'autres : ces derniers mois, des agents d'OpenAI ont été mêlés à des intrusions sur des sites publics américains et sur la plateforme Hugging Face. La leçon vaut pour toute entreprise qui confie à un agent l'accès à ses mails, à ses fichiers ou à ses comptes : une barrière qui paraît fermée peut laisser passer un chemin détourné, et le bouton d'arrêt doit être testé, pas seulement prévu. Si le laboratoire le mieux équipé du secteur a manqué les deux, aucune installation maison ne peut s'en dispenser.



02

Nvidia installe dans ses puces un gardien capable d'isoler une IA qui sort de son périmètre

Nvidia a présenté lundi une plateforme de sécurité pour les agents IA. Sa pièce maîtresse, baptisée Sentry, surveille ces programmes depuis le matériel et peut les mettre en quarantaine en quelques millisecondes s'ils franchissent leurs limites.

L'annonce du 28 septembre associe deux briques. OpenShell, un logiciel gratuit et ouvert présenté en mars, fixe ce qu'un agent a le droit de faire. Sentry, la nouveauté, tourne sur une puce à part, une DPU (un processeur réservé au trafic réseau d'un serveur, distinct de celui qui fait tourner l'IA). Il inspecte chaque requête qui entre et qui sort, vérifie l'identité de l'agent et applique une règle simple : rien n'est permis par défaut, chaque accès aux données doit être explicitement autorisé.

Le choix du matériel n'a rien d'un détail. Un agent qui trouve une faille dans le logiciel chargé de le surveiller peut, en théorie, le désactiver. Un gardien placé sur une autre puce, sur une connexion séparée, lui échappe. Nvidia revendique plus de 100 organisations associées, dont Anthropic, Microsoft, Salesforce, SAP, IBM, JPMorganChase et Citi. Les clients déjà équipés du matériel compatible n'auront besoin que d'une mise à jour logicielle. Aucun prix n'a été communiqué, et Nvidia reconnaît travailler encore sur la surveillance de plusieurs agents qui coopèrent entre eux.

L'annonce arrive quelques jours après la pause décidée par OpenAI, et ce n'est pas un hasard : la sécurité des agents devient un produit qu'on achète, comme l'antivirus il y a vingt ans. Pour qui choisit un hébergeur ou un prestataire chargé de faire tourner des agents, la question « comment l'arrêtez-vous s'il déraile ? » mérite désormais une ligne dans le cahier des charges, avec une réponse précise.



03

Meta se met à vendre ses IA aux entreprises et recrute le patron de MongoDB pour le faire

Meta, maison mère de Facebook, Instagram et WhatsApp, a créé lundi une division qui commercialise ses modèles d'IA et ses assistants auprès des entreprises. Jusqu'ici, le groupe gagnait son argent presque uniquement grâce à la publicité.

La nouvelle division s'appelle Meta Enterprise Platform. Elle regroupe Muse, l'assistant IA de Meta, un agent conçu pour les entreprises (Meta Business Agent), un accès direct aux modèles via une API (une porte d'entrée qui permet à un logiciel d'utiliser le service d'un autre) et Muse Code, un outil d'aide à la programmation. Elle sera dirigée par Chirantan Desai, qui quitte la direction de MongoDB, un éditeur de bases de données très répandu en entreprise, et rendra compte directement à Mark Zuckerberg.

Mark Zuckerberg présente l'activité comme « le prochain grand pilier » du groupe. La logique financière est limpide : Meta dépense cette année plus de 100 milliards de dollars en infrastructures d'IA, et jusqu'ici seule la publicité devait rembourser cet effort. Vendre aux entreprises ouvre une deuxième source de revenus. Les tarifs et les conditions commerciales n'ont pas encore été dévoilés.

Meta arrive sur un terrain déjà occupé par OpenAI, Anthropic, Google et Microsoft, avec un argument à lui : le groupe travaille déjà avec des centaines de millions d'entreprises qui font de la publicité sur ses applications. Pour un commerçant qui gère ses clients sur WhatsApp ou Instagram, l'IA pourrait arriver par le même tableau de bord que ses campagnes. Un fournisseur de plus sur le marché, c'est aussi davantage de pression sur les prix.



04

La Floride demande à un juge d'interdire ChatGPT aux mineurs de l'État

Le procureur général de Floride a déposé lundi une demande d'ordonnance d'urgence contre OpenAI. Il réclame aussi que l'entreprise ne puisse plus développer de nouveaux modèles sans le feu vert d'un contrôleur indépendant.

La requête a été déposée le 28 septembre devant un tribunal du comté de Highlands. Elle prolonge une plainte lancée le 1er juin, la première action d'un État américain contre OpenAI et son patron, Sam Altman. Le procureur James Uthmeier reproche à OpenAI d'interdire en théorie ChatGPT aux moins de 13 ans sans aucun moyen de le faire respecter : la version gratuite ne vérifie pas l'âge, la version payante le demande sans le contrôler, et les parents ne peuvent pas consulter les conversations de leurs enfants.

Les demandes vont loin. Interdire ChatGPT aux mineurs en Floride, cesser de collecter des données sur les moins de 13 ans sans accord parental vérifié, retirer toute affirmation présentant ChatGPT comme sûr ou fiable, afficher un avertissement sur les risques à chaque connexion, supprimer le « je » et les traits humains de l'assistant, et abandonner les techniques qui prolongent artificiellement les conversations. La requête cite aussi les incidents récents impliquant des agents d'OpenAI. L'entreprise n'avait pas encore répondu.

Un juge local tranchera, et rien ne garantit qu'il suivra. Mais la liste dessine ce que les autorités commencent à exiger de tout assistant grand public : vérifier l'âge, prévenir des erreurs, ne pas se faire passer pour un humain. Une entreprise qui place un chatbot face à ses clients, sur son site ou sur une messagerie, a tout intérêt à lire cette liste comme un avant-goût des règles à venir.



05

Un assistant qui réserve, commande et résilie à votre place vaut 10 milliards de dollars un mois après son lancement

Instinct, une jeune entreprise américaine, a levé 1 milliard de dollars lundi. Son agent IA prend en charge les corvées du quotidien par simple SMS, et sa valeur a été multipliée par quatre en un mois.

Instinct a ouvert en août, sur invitation uniquement. L'utilisateur écrit un SMS ou appelle, et l'agent exécute la demande avec son propre ordinateur et son propre téléphone : réserver un voyage ou une table, faire les courses, régler une facture, résilier un abonnement oublié, appeler un cabinet pour prendre rendez-vous. Pour les commerces qui n'ont pas de réservation en ligne, un service de conciergerie prend le relais. Il n'existe pas encore d'application mobile.

Le tour de table de 1 milliard de dollars, auquel participent Sequoia Capital, Benchmark et Coatue, valorise l'entreprise 10 milliards de dollars. En août, elle avait levé 350 millions de dollars sur la base de 2,5 milliards. Instinct ne publie ni son nombre d'utilisateurs, ni ses revenus, ni ses tarifs : les investisseurs paient pour une promesse, pas pour des résultats.

Cette promesse, c'est que la prochaine façon d'acheter ne passera plus par un site ou une application, mais par un assistant à qui l'on délègue. Pour un restaurant, un cabinet ou un artisan, une partie des réservations et des appels viendra alors d'une IA et non d'un client en direct. Un horaire introuvable, un formulaire compliqué ou un standard injoignable deviennent autant de clients perdus au profit du concurrent le plus lisible.

LE CHIFFRE

51 %

du trafic mondial sur internet est généré par des robots et des IA, et non par des humains

Le chiffre porte sur l'année 2024 et sur l'ensemble des visites de sites web dans le monde : plus d'une visite sur deux vient d'un programme automatique plutôt que d'une personne assise devant un écran.

Tous ces robots ne sont pas malveillants. On y trouve les robots d'indexation (les programmes qui parcourent le web pour alimenter les moteurs de recherche et les assistants IA), les outils qui surveillent les prix des concurrents, les agents qui naviguent pour le compte d'un utilisateur, mais aussi des programmes qui testent des mots de passe ou copient des contenus. Leur point commun : personne ne regarde la page au moment où elle s'affiche.

Pour quiconque possède un site, une partie des statistiques de visite ne correspond donc à aucun être humain. Un pic d'audience soudain, des visiteurs qui repartent en une seconde, des clics publicitaires qui ne débouchent sur aucun achat peuvent venir de machines. Payer une campagne au clic ou juger une page à son nombre de vues suppose désormais de savoir trier ce trafic.

L'autre versant compte autant. Si les machines lisent déjà plus que les humains, votre site est aussi lu par les assistants qui répondent aux questions de vos futurs clients. Des pages claires, des prix et des horaires visibles, des informations à jour servent désormais deux publics à la fois : la personne qui visite, et l'IA qui la renseigne.