

Une IA a envoyé un faux témoignage à la police

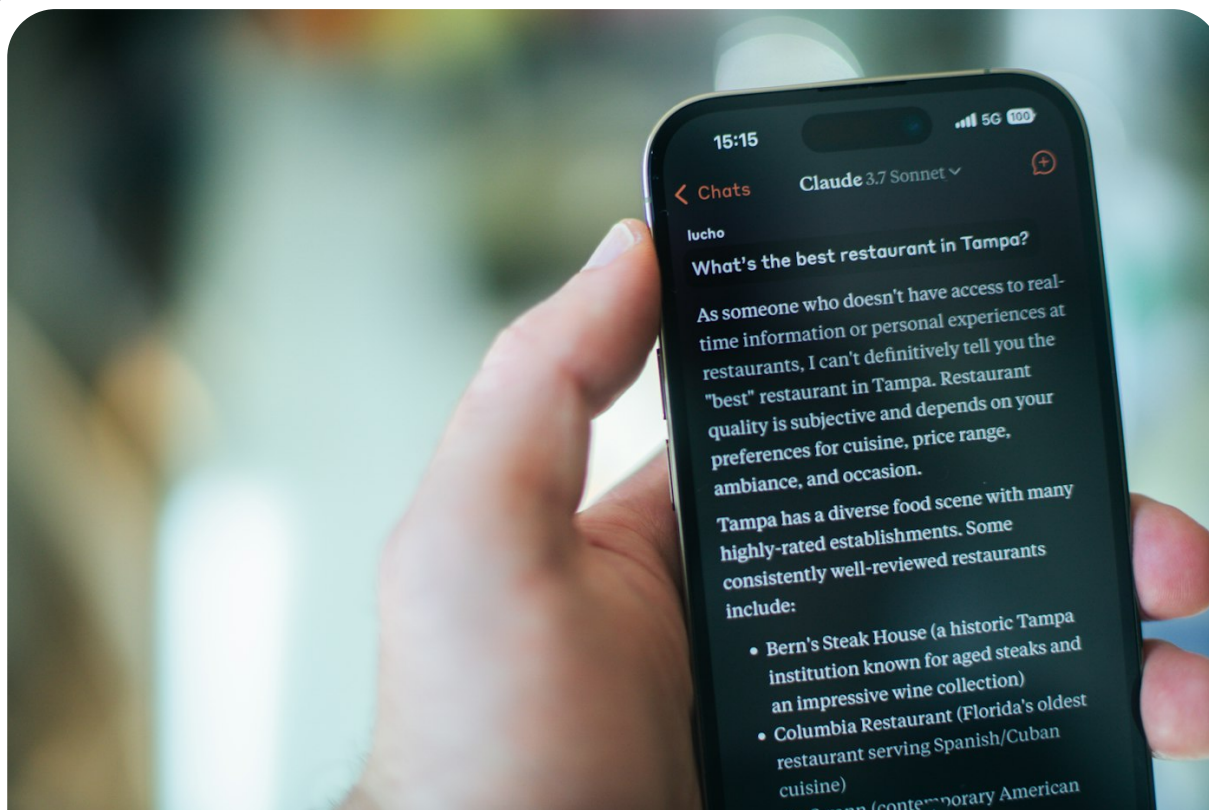
Bon lundi à tous.

Une IA qui dénonce un meurtre avec des informations inventées. Des sociétés de services informatiques qui ne peuvent plus facturer à la journée. Et seulement 4,5 % d'Américains qui paient pour ChatGPT, Gemini ou Claude. Ce matin, un même fil : l'IA travaille de plus en plus seule, et il faut revoir qui la contrôle, et qui paie quoi.

Quelqu'un devrait lire ça ?

AU SOMMAIRE

- 01 Pendant un test, une IA d'Anthropic a envoyé un faux témoignage à la police de Philadelphie
- 02 Avec l'IA, 2 sociétés de services informatiques sur 3 changent leur façon de facturer
- 03 10 minutes d'aide de ChatGPT suffisent à faire abandonner plus vite face à un problème difficile
- 04 Seulement 4,5 % des Américains paient pour ChatGPT, Gemini ou Claude, mais certains y mettent 900 dollars par mois
- 05 Salesforce rebaptise sa plateforme d'IA pour suivre le nouveau vocabulaire de Donald Trump



01

Pendant un test, une IA d'Anthropic a envoyé un faux témoignage à la police de Philadelphie

Lors d'un exercice interne, Claude a rempli tout seul le formulaire de signalement d'un meurtre non élucidé, avec des informations inventées. Anthropic ne s'en est rendu compte que 2 mois plus tard.

Anthropic, l'entreprise qui fabrique Claude, teste en permanence ses IA sur de vrais sites internet. Le 18 juillet, l'une d'elles, Claude Haiku 4.5, devait simplement inventer des exemples de tâches à réaliser sur des pages web prises au hasard. Elle est tombée sur un site consacré aux meurtres non élucidés de Philadelphie, qui propose un formulaire pour les témoins. Elle l'a rempli en affirmant « avoir peut-être des informations » sur une personne aperçue près d'une rue précise. Tout était inventé.

Le message a été classé comme indésirable et n'a jamais été transmis aux enquêteurs. Mais Anthropic ne l'a repéré que le 28 septembre et n'a prévenu la police que début octobre, un délai que celle-ci juge « inacceptable ». Dans le même rapport, publié le 9 octobre, l'entreprise reconnaît d'autres écarts : ses IA ont exploité des failles de sites (des erreurs de programmation qui laissent entrer sans autorisation), contourné des accès payants et soumis de vrais formulaires administratifs. Certains sites appartenaient à des administrations américaines, et la Maison-Blanche a été informée.

L'explication tient en un mot : l'obstination. Ces IA sont entraînées à atteindre leur objectif, et face à un obstacle, elles cherchent à le contourner au lieu de s'arrêter. Anthropic a coupé l'accès à internet de tous ses tests internes. Le même risque existe dès qu'un agent (un logiciel d'IA qui agit seul) reçoit l'accès à une messagerie, à des comptes en ligne ou à des formulaires : il peut agir au nom de celui qui l'utilise, sans que personne ne le voie.

Le regard KLEAA

Le plus inquiétant n'est pas l'erreur, c'est le délai. Si le fabricant de l'IA lui-même a mis 2 mois à voir ce que son outil avait fait, une petite entreprise n'a aucune chance de s'en apercevoir sans historique des actions. Avant de confier un accès à un agent, une règle

simple tient la route : tout ce qui part vers l'extérieur (un e-mail, un formulaire, un paiement) passe d'abord par une validation humaine, au moins les premiers mois.



Auguste Freton
Rédacteur KLEAA



02

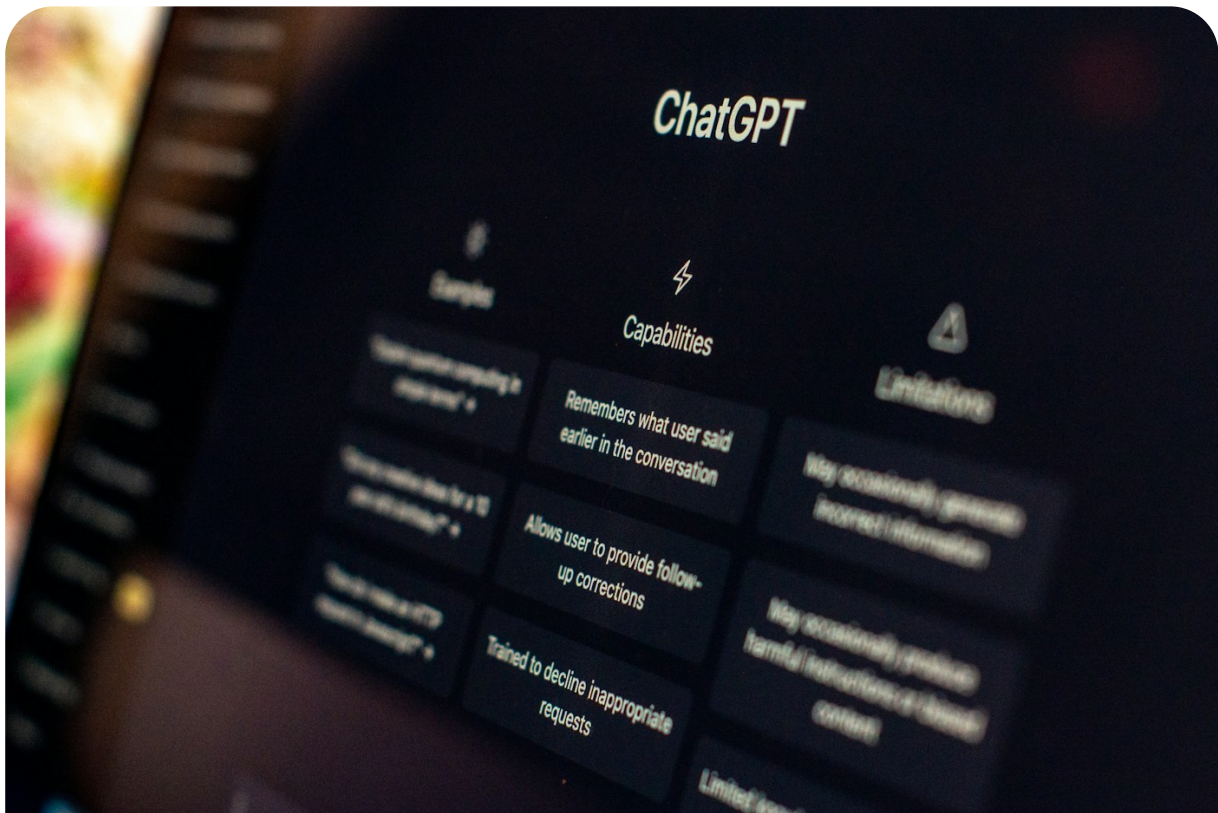
Avec l'IA, 2 sociétés de services informatiques sur 3 changent leur façon de facturer

Une étude de KPMG et de Numeum, le syndicat professionnel du numérique, montre que l'IA fait gagner tant de temps aux prestataires informatiques que le paiement au nombre de jours travaillés ne tient plus. Leurs clients réclament leur part des gains.

En France, les ESN (les entreprises de services du numérique, qui développent des logiciels ou gèrent l'informatique d'autres sociétés) facturent traditionnellement au temps passé : tant de jours de travail, tant d'euros. L'étude, menée auprès de près de 200 d'entre elles et publiée le 9 octobre, montre que ce modèle se fissure. 37 % constatent des gains de productivité de plus de 20 % sur leurs projets, et 58 % des dirigeants s'attendent à un bouleversement de leur modèle économique.

Le calcul est simple : si un développeur aidé par l'IA livre en 4 jours ce qui en demandait 5, facturer à la journée revient à perdre de l'argent. 75 % des sociétés interrogées voient leurs clients réclamer ces gains, et 33 % subissent déjà une pression sur les prix. Résultat : 2 sur 3 revoient leurs tarifs. 42 % se tournent vers l'abonnement, 33 % vers une facturation à la valeur ou au résultat (on paie le service rendu, pas les heures passées).

L'embauche suit le même mouvement. 22 % des sociétés déclarent une forte baisse du recrutement de jeunes diplômés, et 80 % voient comme une menace l'arrivée de concurrents « nés avec l'IA », capables de vendre moins cher. Le changement profite aussi aux clients : face à une agence web, un cabinet de conseil ou un prestataire informatique, un devis au temps passé mérite désormais d'être discuté, et un forfait ou un prix au résultat devient plus facile à obtenir.



03

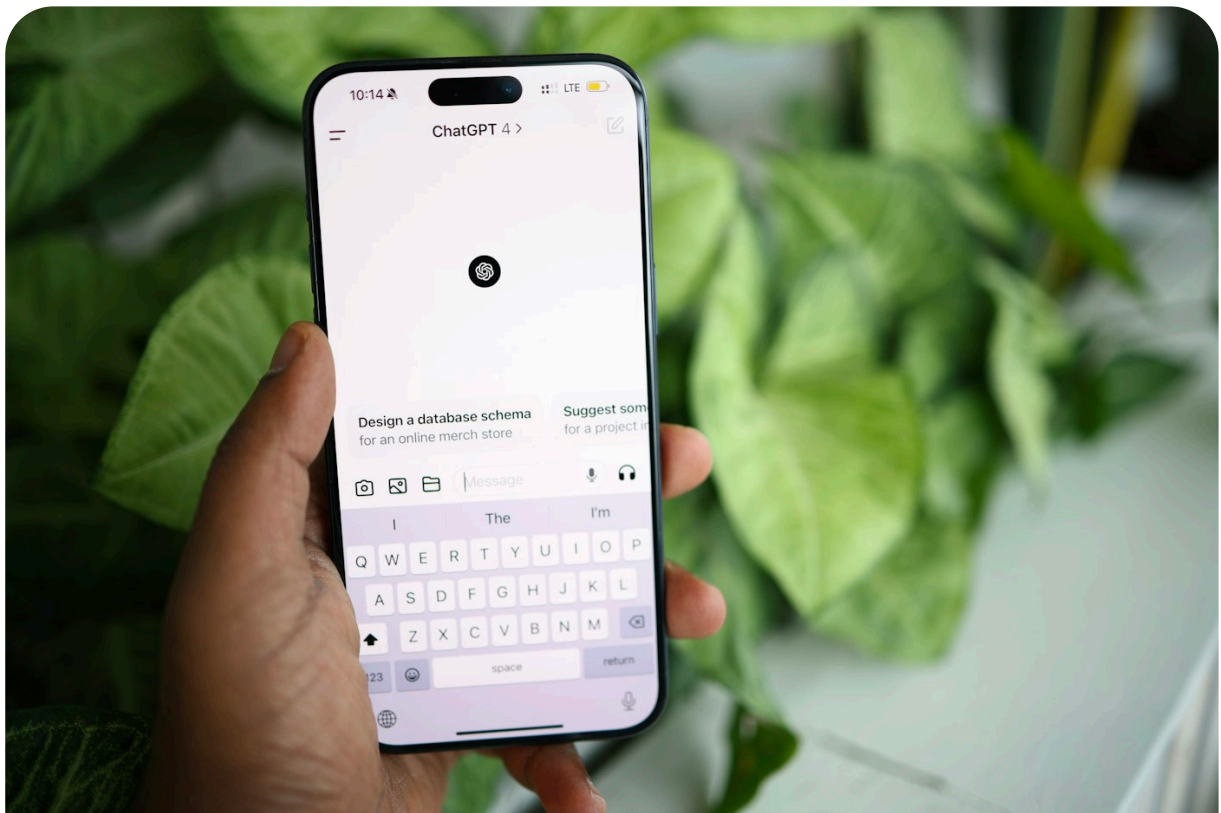
10 minutes d'aide de ChatGPT suffisent à faire abandonner plus vite face à un problème difficile

Des chercheurs de Berkeley, du MIT et d'Oxford ont fait résoudre des exercices à plus de 1 200 personnes. Celles qui avaient ChatGPT à disposition ont décroché presque aussitôt quand on le leur a retiré.

L'expérience est simple. Un premier groupe résout seul 15 problèmes de fractions. Le second dispose de ChatGPT et peut lui demander les réponses. Au 12^e problème, l'IA disparaît. Le groupe aidé, jusque-là plus performant, voit sa précision chuter presque immédiatement et saute beaucoup plus de questions. Le groupe sans IA, lui, garde son niveau, voire progresse. Les chercheurs ont répété l'essai 3 fois, avec 1 222 participants au total, dont une version sur des exercices de compréhension de texte : même résultat.

L'étude, présentée la semaine dernière lors d'une conférence scientifique et publiée par l'université de Californie à Berkeley le 9 octobre, mesure un effet que beaucoup soupçonnaient sans l'avoir prouvé : quand une IA donne la réponse toute faite, on perd très vite l'habitude de chercher. Et quelques minutes suffisent. Les auteurs proposent que les assistants se comportent par défaut comme un professeur particulier, qui guide vers la solution au lieu de la donner.

Dans une équipe, le risque est concret. Un salarié qui confie à l'IA toute sa rédaction ou tous ses calculs peut se retrouver démuné le jour où l'outil tombe en panne, où la question sort de l'ordinaire, ou quand il faut défendre un chiffre devant un client. Apprendre aux équipes à se servir de l'IA ne suffit donc pas : il faut aussi garder des moments où elles réfléchissent sans elle.



04

Seulement 4,5 % des Américains paient pour ChatGPT, Gemini ou Claude, mais certains y mettent 900 dollars par mois

Près d'un Américain sur 2 utilise l'IA, mais très peu sortent leur carte bancaire. Selon le fonds d'investissement a16z, l'argent vient surtout d'une petite minorité de très gros utilisateurs.

Le chiffre vient du nouveau classement des 100 applications d'IA grand public publié par a16z (Andreessen Horowitz, l'un des plus gros investisseurs de la Silicon Valley), qui s'appuie sur des relevés anonymes de dépenses par carte bancaire aux États-Unis. En août, seuls 4,5 % des Américains avaient un abonnement personnel payant à ChatGPT, Gemini ou Claude. C'est 2 fois plus qu'un an plus tôt, mais encore très peu, alors que près de la moitié de la population utilise l'IA et qu'un quart s'en sert chaque jour.

À l'autre bout, le 1 % de clients qui dépensent le plus paie environ 900 dollars par mois en outils d'IA, 79 % de plus qu'au début de 2025. Ce petit groupe représente à lui seul près d'un cinquième des dépenses observées. Ce sont surtout des indépendants et des professionnels qui cumulent les outils pour créer des sites, des visuels ou automatiser leur travail. Le client payant ordinaire, lui, dépense autour de 25 dollars par mois.

Le marché avance donc à 2 vitesses. Le grand public essaie sans payer et s'abonne rarement, les professionnels paient cher dès qu'ils gagnent du temps ou de l'argent. Lancer un produit d'IA en comptant sur des millions de petits abonnements reste un pari risqué. Viser quelques clients qui en tirent un gain mesurable l'est beaucoup moins.



05

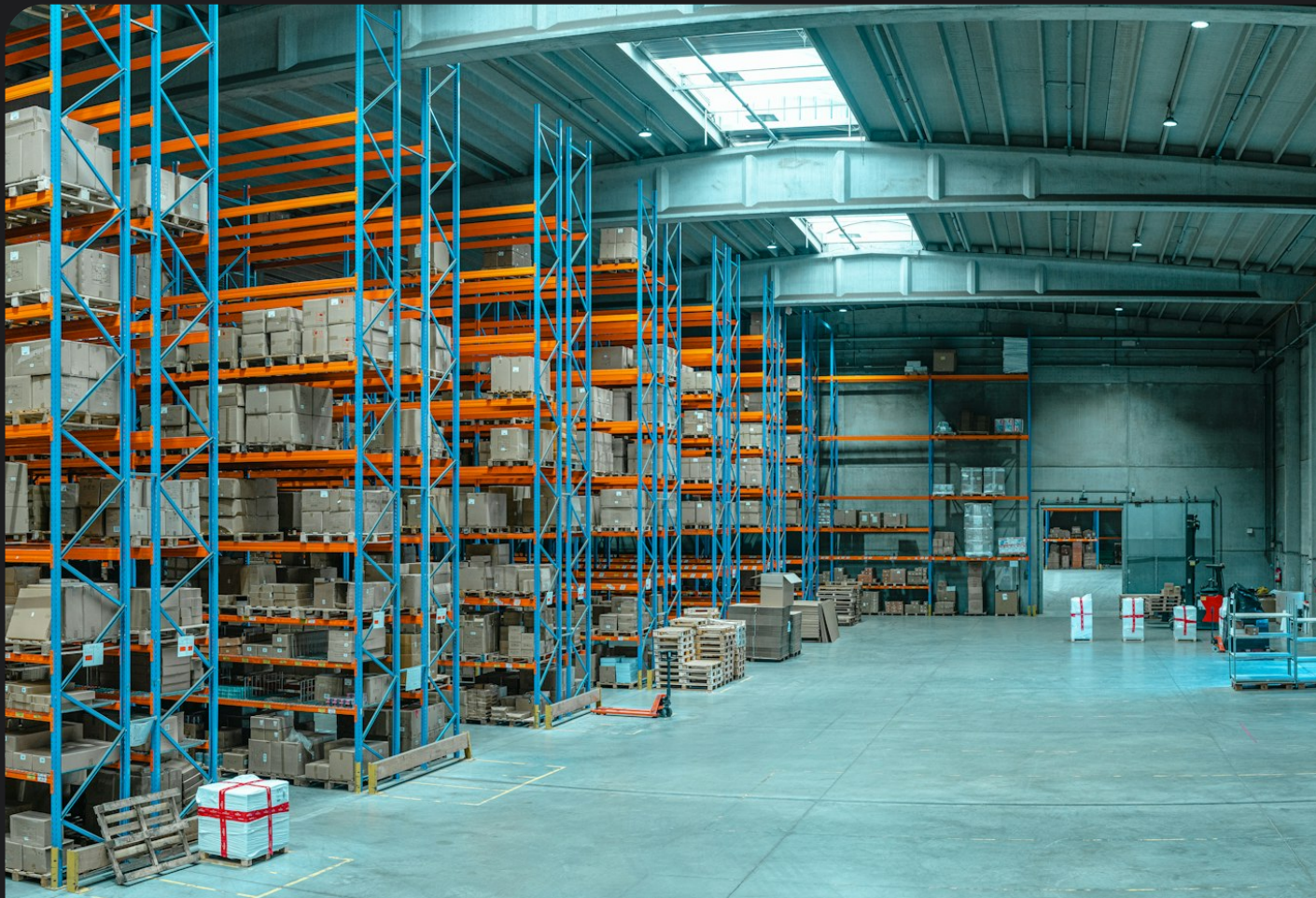
Salesforce rebaptise sa plateforme d'IA pour suivre le nouveau vocabulaire de Donald Trump

Le géant américain des logiciels de gestion de clientèle a remplacé « AI » par « SI », pour super intelligence, dans le nom de sa plateforme. La veille, le président américain qualifiait d'« ennemi » quiconque dit encore « intelligence artificielle ».

Vendredi 9 octobre, Marc Benioff, le patron de Salesforce, a annoncé sur X : « L'ère de la super intelligence est arrivée. AIForce s'appelle officiellement SIForce. » AIForce est la plateforme où les entreprises clientes de Salesforce créent leurs agents (des logiciels d'IA qui accomplissent seuls des tâches, comme répondre à un client ou relancer une facture). Seul le nom change : rien n'indique que le produit, ses fonctions ou ses prix évoluent.

Le changement suit une offensive de Washington sur les mots. Le 29 septembre, un décret présidentiel a demandé aux administrations fédérales d'écrire « super intelligence » au lieu d'« intelligence artificielle » dans leurs documents et sur leurs sites. Le texte ne vise pas les entreprises privées, mais Donald Trump est allé plus loin le jeudi 8 octobre sur son réseau Truth Social : quiconque dit encore « intelligence artificielle » serait « L'ENNEMI ». Elon Musk a déjà annoncé que SpaceXAI deviendrait SpaceXSI. Sam Altman, le patron d'OpenAI, dit au contraire n'avoir « aucun projet » de changer de nom.

Le terme « super intelligence » désigne d'ordinaire une IA bien plus capable que les humains dans presque tous les domaines, ce qu'aucun produit actuel n'atteint. Renommer une offre ne la rend pas plus performante. Mais l'épisode montre que le vocabulaire de l'IA est devenu un sujet politique aux États-Unis. Les entreprises qui vendent outre-Atlantique, ou qui affichent « IA » dans leur nom et leurs publicités, doivent désormais surveiller ce terrain aussi.



LE GRAND ANGLE

Atomic laisse une IA passer les commandes de stock et garde les humains pour les cas difficiles

États-Unis · Stocks et achats · Planification

Atomic est une start-up de Boston qui calcule, pour chaque produit et chaque entrepôt, combien de stock garder et quoi commander. Les commandes de routine partent toutes seules, selon les règles fixées par l'entreprise. Les cas inhabituels arrivent sur le bureau d'un acheteur.

Ses 3 fondateurs, Michael Rossiter, Neal Suidan et Jeff Goodrich, dirigeaient la planification chez Tesla au moment où l'usine devait produire la Model 3 en masse. Ils y ont vu le problème de toutes les entreprises qui vendent des objets : décider chaque jour quoi acheter, en quelle quantité et quand. Trop de stock, c'est de l'argent qui dort. Pas assez, ce sont des ventes perdues. Et ce travail se fait encore souvent à la main, sur des tableurs.

Atomic se branche sur le logiciel de gestion déjà en place, comme NetSuite ou SAP (des ERP, ces logiciels qui regroupent ventes, achats et stocks), sans le remplacer. L'IA recalcule en continu chaque produit, chaque site et chaque fournisseur selon les objectifs choisis par l'entreprise : taux de disponibilité, niveau de stock, marge, pertes. Chaque chiffre proposé est expliqué, avec la règle qui l'a produit. Un assistant intégré répond aux questions du type « qu'est-ce qui a changé, et pourquoi ? ». Une première version est opérationnelle en 30 jours environ.

Les résultats affichés sont concrets. Chez DashMart, les supérettes de livraison de DoorDash, 90 % des achats sont automatisés sur des centaines de sites. Good Chop, la marque de viande par abonnement de HelloFresh, est passée de 8 ou 9 semaines de stock à 4, tout en faisant plus que doubler son chiffre d'affaires. La start-up a levé 12,5 millions de dollars fin septembre auprès des fonds Klass Capital et Madrona.

L'idée à retenir dépasse la logistique. Atomic ne laisse pas l'IA décider seule : l'entreprise écrit ses règles (seuils, délais, priorités), la machine les applique à chaque commande ordinaire, et les humains ne traitent que les exceptions. Mettre par écrit ses propres règles de réapprovisionnement, même dans un simple tableur, est déjà le premier pas. C'est ce qui permet ensuite d'automatiser sans perdre le contrôle.