

L'IA envoie tant de fausses alertes que Google a fermé son guichet

Bon mardi à tous.

Google a dû fermer un de ses programmes de sécurité, submergé de faux signalements fabriqués par des IA. En Espagne, depuis hier, chaque employeur doit expliquer par écrit les logiciels qui décident à la place des managers. Et un rapport du MIT montre qu'avec l'IA, une bonne note ne prouve plus grand-chose.

Quelqu'un devrait lire ça ?

AU SOMMAIRE

- 01 Google suspend sa prime aux failles, noyée sous de faux signalements écrits par IA
- 02 En Espagne, les employeurs doivent désormais expliquer par écrit les logiciels qui décident pour leurs salariés
- 03 L'allemand Aleph Alpha publie gratuitement une IA conçue pour les administrations européennes
- 04 Avec l'IA, une bonne note à l'université ne prouve plus grand-chose, alerte le MIT
- 05 Anthropic, le créateur de Claude, devient le plus gros donateur parmi les entreprises américaines



01

Google suspend sa prime aux failles, noyée sous de faux signalements écrits par IA

Google payait des chercheurs pour lui signaler des failles de sécurité. Les IA ont envoyé tant de signalements inventés que le programme est fermé jusqu'en 2027.

Depuis le 1er octobre, Google a mis en pause son programme de récompenses pour les failles trouvées dans les logiciels libres (des programmes dont le code est public et réutilisable par tous). Le principe était simple : toute personne qui découvrait une faille de sécurité et la signalait pouvait toucher une prime. Google explique la pause par « une forte hausse des envois automatisés, dont la grande majorité ne sont pas valables ».

Concrètement, des internautes ont demandé à des IA de fouiller du code et d'écrire des rapports en série, dans l'espoir de décrocher une prime. Beaucoup décrivaient des failles qui n'existent pas : l'IA les avait inventées (on parle d'hallucinations). Les ingénieurs de Google et les bénévoles qui entretiennent ces logiciels ont passé leur temps à vérifier du vide. Le programme ne rouvrira pas avant le premier trimestre 2027 ; les autres programmes de primes de Google restent ouverts.

Des experts en cybersécurité avaient prévenu dès l'été 2025 que les contenus de mauvaise qualité produits par IA menaçaient ce type de programme. Le risque dépasse Google : tout guichet ouvert au public, qu'il s'agisse d'un formulaire de contact, d'un service client ou d'un appel à candidatures, peut être inondé de la même façon dès qu'une récompense est en jeu.

Le regard KLEAA

Produire un texte ne coûte plus rien, le vérifier coûte toujours autant. Ce déséquilibre va peser sur tous ceux qui reçoivent des demandes venues de l'extérieur. La parade la plus simple n'est pas technique : exiger une preuve concrète avant de lire, comme une démonstration reproductible, une pièce vérifiable ou un petit engagement de l'expéditeur. Ce qui ne coûte rien à envoyer finira par ne plus être lu.



Auguste Freton
Rédacteur KLEAA



02

En Espagne, les employeurs doivent désormais expliquer par écrit les logiciels qui décident pour leurs salariés

Depuis hier, une entreprise espagnole qui laisse un programme répartir les horaires, fixer une prime ou peser sur un licenciement doit dire à ses salariés comment il fonctionne.

Le décret royal 723/2026, publié au journal officiel espagnol le 15 septembre, est entré en vigueur le 5 octobre. Il transpose une directive européenne de 2019 sur des conditions de travail transparentes et prévisibles. Une des nouvelles obligations vise directement l'IA : informer les salariés de tout système algorithmique (un programme qui applique des règles pour prendre une décision) qui intervient dans leur planning, leurs tâches, leur salaire, leur promotion, leur lieu de travail ou la fin de leur contrat.

L'entreprise doit en expliquer les critères et les règles de fonctionnement, sans être obligée de dévoiler le programme lui-même. Pour une nouvelle embauche, l'information doit être donnée avant le premier jour de travail. Un salarié déjà en poste peut la demander : l'employeur a alors 30 jours ouvrables pour répondre. Le texte concerne les contrats de plus de 4 semaines.

Le décret impose aussi de détailler le salaire de base et chaque prime séparément. Les petites entreprises sont concernées comme les grandes : un logiciel de planning qui attribue les créneaux tout seul entre dans le champ. La première étape consiste donc à faire l'inventaire des outils qui prennent déjà des décisions, souvent sans que personne ne l'ait formulé ainsi.



03

L'allemand Aleph Alpha publie gratuitement une IA conçue pour les administrations européennes

Baptisé Kolibri, ce modèle peut être téléchargé et installé sur ses propres serveurs, sans dépendre d'un géant américain ou chinois.

Aleph Alpha, une entreprise allemande d'IA, a publié le 5 octobre Kolibri, un modèle de langage (le moteur qui fait fonctionner un assistant comme ChatGPT). Il est diffusé sous licence Apache 2.0 (une licence libre qui autorise l'usage commercial) et ses fichiers sont téléchargeables gratuitement. Il parle allemand et anglais, et vise l'administration, l'aviation et l'industrie.

Le modèle compte 78 milliards de paramètres (les réglages internes appris pendant l'entraînement), mais n'en active qu'environ 3 milliards pour chaque mot produit. Résultat, selon l'entreprise : il répond plus vite et coûte moins cher à faire tourner que des concurrents de taille comparable. Il a été entraîné sur 768 puces Nvidia installées en Allemagne et en Finlande, en tenant compte du règlement européen sur l'IA. Détail notable : une partie des données d'entraînement a été fabriquée par des modèles chinois.

L'intérêt est surtout pratique : une organisation qui manipule des données sensibles peut installer Kolibri chez elle, sans les envoyer à un service hébergé aux États-Unis. Le modèle ne parle pas français à ce stade. Il montre toutefois qu'une solution européenne gratuite existe désormais pour des usages professionnels, au moment où la dépendance de l'Europe aux IA étrangères inquiète jusqu'à la Banque centrale européenne.



04

Avec l'IA, une bonne note à l'université ne prouve plus grand-chose, alerte le MIT

Un rapport du MIT, l'une des plus grandes universités américaines, montre que les étudiants obtiennent de meilleures notes avec l'IA, mais en savent beaucoup moins une fois privés d'elle.

Le rapport, rendu public le 5 octobre, rassemble plusieurs études. À l'université Brown, des étudiants ont obtenu en moyenne 96 % à un examen fait à la maison, puis 48,6 % à un examen de suivi passé en salle, sans aide. En Chine, une étude menée sur plus de 26 000 élèves a mesuré une hausse de 18 % des notes de devoirs grâce à l'IA, mais une baisse de 20 % des résultats aux examens au bout de 6 mois.

À Berkeley, en Californie, la part de notes maximales dans les cours d'écriture et de programmation a gagné 13 points depuis l'arrivée de ChatGPT. Le lien humain se défait aussi : à Harvard, 1 étudiant sur 4 dit aller moins souvent voir ses professeurs. Le MIT parle d'une « illusion d'apprentissage ». Autre écart : 67 % des étudiants jugent l'IA importante pour leur carrière, mais seulement 25 % se sentent prêts.

Pour ceux qui recrutent, la conséquence est directe : un bulletin ou un devoir rendu ne suffit plus à juger un candidat. Les tests passés sur place, les mises en situation et les questions de suivi reprennent de la valeur, car ils montrent ce que la personne sait faire sans assistance.

Le regard KLEAA

Le problème n'est pas l'usage de l'IA, que l'on retrouvera de toute façon au travail. C'est de savoir si la personne comprend ce qu'elle rend. 10 minutes passées à demander à un candidat d'expliquer puis de modifier un travail qu'il présente en disent plus long qu'une moyenne brillante. Les recruteurs qui adaptent leurs entretiens dès maintenant éviteront de mauvaises surprises avec les premières promotions qui ont étudié avec l'IA du début à la fin.



Auguste Freton
Rédacteur KLEAA



05

Anthropic, le créateur de Claude, devient le plus gros donateur parmi les entreprises américaines

Ses salariés ont donné 540 millions de dollars à des associations en 2025, surtout sous forme d'actions de l'entreprise, qui en rajoute à chaque don.

D'après des documents préparés pour son entrée en Bourse et montrés à de futurs investisseurs, les salariés d'Anthropic ont versé 540 millions de dollars à des associations en 2025. C'est près de 5 fois plus que la banque Truist (115 millions) et le gestionnaire d'actifs BlackRock (109 millions). Entre octobre 2025 et mars 2026, 660 millions de dollars supplémentaires ont été versés.

Le mécanisme est inhabituel : les salariés donnent des actions de l'entreprise, et Anthropic y ajoute ses propres actions. Pour les premiers employés, la valeur du don est triplée. Le patron, Dario Amodei, et ses 6 cofondateurs se sont engagés à donner au moins 80 % de leur fortune. Beaucoup de salariés se réclament de l'altruisme efficace (un courant qui cherche à donner là où chaque dollar a le plus d'effet) et visent la lutte contre la pauvreté, la sécurité de l'IA ou le bien-être animal.

Cette générosité a un coût pour les autres actionnaires : chaque action ajoutée par l'entreprise réduit la part de ceux qui en possèdent déjà (on parle de dilution). Si le cours monte après l'entrée en Bourse, les montants donnés pourraient atteindre plusieurs milliards de dollars. Les futurs investisseurs devront en tenir compte dans le prix qu'ils accepteront de payer.

LE CHIFFRE

27 %

des tâches des salariés français pourraient être automatisées d'ici 2030

Ce chiffre est une projection. Il ne compte pas des emplois, mais des tâches, c'est-à-dire les actions qui composent une journée de travail (rédiger, classer, vérifier, répondre). Il porte sur les salariés en France, à l'horizon 2030.

Concrètement, plus d'un quart des tâches réalisées aujourd'hui par des salariés pourraient être prises en charge par des machines et des logiciels, IA comprise. Cela ne veut pas dire qu'un quart des salariés perdra son poste : un même métier mêle des tâches automatisables et d'autres qui ne le sont pas.

Une autre estimation éclaire ce chiffre : en France, seuls 5 % des emplois seraient directement remplaçables par l'IA à moyen terme. L'écart entre 27 % des tâches et 5 % des emplois résume la situation : la plupart des postes vont changer de contenu bien plus qu'ils ne vont disparaître.

Le gain se jouera donc dans l'organisation : repérer les tâches répétitives de chaque poste, décider ce que l'on confie à l'outil et ce que l'on garde, puis réinvestir le temps libéré. Une équipe qui fait cet exercice dès maintenant aura 4 ans d'avance sur celles qui attendront 2030.