

LES GOUROUS DE L'IA ÉPISODE 1/12

Dario Amodei, le patron d'Anthropic, l'utopiste alarmiste de l'IA

« Les gourous de l'IA » (1/12). Le quadragénaire incarne une figure paradoxale de la révolution de l'intelligence artificielle, à la fois inquiet sur ses possibles dérives et convaincu qu'elle peut révolutionner la médecine.

Par Alexandre Piquard

Publié hier à 05h00, modifié hier à 08h31 • Lecture 4 min.

Article réservé aux abonnés



Le patron d'Anthropic, Dario Amodei, en 2026. LÉA GIRARDOT/« LE MONDE » D'APRÈS BHAWIKA CHHABRA/REUTERS

Dans l'actualité récente, Dario Amodei est connu pour son conflit avec l'armée américaine à propos de l'utilisation de ses outils d'intelligence artificielle (IA). De manière plus générale, le patron d'Anthropic est souvent décrit comme l'un des dirigeants d'entreprises d'IA les plus « doomers » ou « alarmistes » sur les risques extrêmes de cette technologie. Mais cet ingénieur américain est aussi un utopiste – un paradoxe apparent également présent, aux débuts d'OpenAI, chez ses fondateurs, comme Sam Altman ou Elon Musk.

Anthropic est « l'épicentre brûlant du catastrophisme sur l'IA », écrivait le New York Times, en juillet 2023, deux ans après que Dario Amodei a quitté OpenAI pour cofonder cette start-up, avec une attention

Édition du jour

Daté du vendredi 27 mars



Lire le journal numérique

Lire les éditions précédentes

PUBLICITÉ

Les plus lus

particulière pour la sécurité. Dans un très long essai publié fin janvier, l'entrepreneur dit refuser les discours « catastrophistes » trop « proches de la science-fiction » mais souhaite bien « réveiller » les gens sur les risques extrêmes de l'IA. « Si le progrès exponentiel de l'IA continue (...), il ne peut pas se passer plus de quelques années avant qu'elle devienne meilleure que les humains dans pratiquement tous les domaines », y postule-t-il.

Si un « pays de génies » – soit 50 millions d'IA « surpassant des Prix Nobel ou des hommes d'Etat » – se matérialisait, il faudrait s'inquiéter des risques de « perte de contrôle » de ces IA par les humains, mais aussi de « mésusage à des fins de destruction » par des « mauvais acteurs », par exemple pour construire des « armes biologiques », ou encore pour « s'emparer du pouvoir », énumère-t-il. Dario Amodei, qui avait prédit, mi-2025, la suppression de la moitié des emplois de « cols blancs » débutants d'ici à cinq ans, cite aussi le risque de « chômage de masse » et de « concentration extrême des richesses ».

Lire aussi | [IA : derrière la peur d'une « jobs apocalypse », le débat monte concernant son impact sur l'emploi](#)

Le patron d'Anthropic désigne comme des « lignes rouges » l'utilisation de l'IA pour les « armes autonomes » ou pour la « surveillance de masse » de la population. Les deux pratiques qu'il souhaitait voir exclues de son contrat avec le Pentagone et qui ont entraîné l'interdiction de l'usage des outils d'Anthropic par l'armée américaine. « Nous devrions utiliser l'IA pour la défense nationale de toutes les manières, sauf celles qui nous rapprocheraient de nos adversaires autocratiques », professe dans son essai Dario Amodei.

Dario Amodei en bref

- Américain, âgé de 43 ans
- Cofondateur et directeur général d'Anthropic, start-up américaine d'IA, créée en 2021 (valorisée 380 milliards de dollars, soit 328 milliards d'euros)
- Vice-président de la recherche chez OpenAI de 2016 à 2020, après un passage chez Google et Baidu
- Licence de physique à l'université Stanford (Californie) et doctorat de biophysique à Princeton (New Jersey)

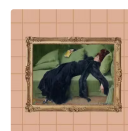
Cette attitude a renforcé les tensions existantes avec l'administration de Donald Trump : celle-ci a décrit Dario Amodei comme un gauchiste car il avait, en 2024, soutenu à la présidentielle américaine la démocrate Kamala Harris et qualifié sur Facebook son concurrent républicain, Donald Trump, de « seigneur de guerre féodal ». Des proches du président américain ont souligné les liens de Dario Amodei avec l'« altruisme efficace », un courant de pensée prônant la lutte contre les risques existentiels pour l'humanité, dont l'IA – l'intéressé assure toutefois ne pas se considérer membre du mouvement.

1 EN DIRECT, guerre au Moyen-Orient : l'Iran et Israël continuent leurs attaques aériennes ; l'Irak va saisir l'ONU après un bombardement sur son territoire

2 Au procès en appel de l'affaire Sarkozy-Kadhafi, Brice Hortefeux déstabilisé par la « note Moussa Koussa » : « Euh, vous me prenez de court... »

3 Thierry Breton : « Il faut inscrire dans la Constitution française une règle d'or limitant le déficit public à 1 % du PIB à l'horizon 2032 »

Le Monde | Ateliers



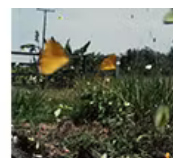
Cours du soir

Affinez vos connaissances avec Françoise Barbe-Gall, historienne de l'art.

Découvrir



© Magnum / The Photographers' Gallery



VENTE DE TIRAGES SIGNÉS
EXCLUSIVITÉ ABONNÉ
15% OFFERTS
AVEC LE CODE **LEMONDE**



Accédez à la vente

Le dirigeant d'Anthropic a aussi agacé l'administration Trump en s'opposant au projet de loi fédéral interdisant aux Etats américains de voter des réglementations sur l'IA et en prônant le maintien de contrôles à l'exportation sur les puces de Nvidia, contre la Chine.

« Réglementer l'IA à son profit »

Pour autant, Dario Amodei s'est toujours dit prêt à travailler avec l'administration Trump. « *Je crois profondément à l'importance existentielle d'utiliser l'IA pour défendre les Etats-Unis et les démocraties, et défaire nos adversaires autocratiques* », a assuré le 26 février, dans un communiqué, le patron d'Anthropic, dont les outils ont été utilisés, à travers son partenariat avec Palantir, pour identifier des cibles en Iran, selon le *Washington Post*.

Les discours de Dario Amodei sur les risques de l'IA suscitent aussi des critiques : il en fait « *un argument pour convaincre les gouvernements de réglementer l'IA à son profit* », a estimé, à la mi-février, le Français Yann Le Cun, l'accusant de vouloir gêner les petites start-up et de restreindre l'IA open source, accessible et modifiable librement. Les mises en garde sur les risques extrêmes « *sont largement de la diversion* », a renchéri le Français Arthur Mensch, de Mistral AI, alertant sur le risque « *d'oligopole de l'information* » contrôlé par les géants américains de l'IA.

Lire aussi le portrait | [Dario Amodei, l'ambivalent patron d'Anthropic qui se dresse contre Donald Trump](#)

D'autres estiment qu'évoquer les risques extrêmes détourne l'attention de problèmes plus immédiats de l'IA comme les discriminations ou l'utilisation non autorisée de textes et d'images pour entraîner les modèles. Le concept d'IA supérieure aux humains est aussi parfois jugé trop flou ou inatteignable avec les modèles de traitement du langage actuels.

Dario Amodei assure, lui, alerter sur les risques parce qu'ils sont un obstacle à la réalisation d'un futur très positif grâce à l'IA. Dans un long texte d'octobre 2024, pendant optimiste de son essai de janvier, il dépeint un monde utopique et vertigineux : l'IA pourrait compresser « *en cinq à dix ans le progrès biologique des cinquante à cent prochaines années* », écrit celui qui a étudié la biologie après avoir perdu son père d'une maladie rare. L'ex-chercheur envisage la prévention et le traitement de la plupart des maladies infectieuses, cancers, maladies génétiques et maladies mentales. Il juge possible le « *doublément de la longévité humaine* », à « *150 ans* ». Et avance le concept de « *liberté biologique* » selon lequel « *le poids, l'apparence physique, la reproduction et d'autres processus biologiques seront totalement sous le contrôle des gens* ». Des thématiques qui rappellent le transhumanisme, ce courant de pensée visant à « *augmenter* » l'humain par la technologie.

¶ **Pour aller plus loin** Lire son essai : « L'Adolescence de la technologie » (janvier 2026, version traduite sur « Le Grand Continent » ou version anglaise).

Dans ce texte très long et remarqué, Dario Amodei expose sa vision du futur de l'IA sous l'angle des risques extrêmes.

L'essai est conçu comme le versant pessimiste de « Machines of Loving Grace », [un texte d'octobre 2024](#) imaginant lui les bénéfices de l'émergence d'une « IA puissante », avec un prisme optimiste. Dans « L'Adolescence de la technologie », le cofondateur d'Anthropic détaille les risques posés, selon lui, par l'IA en cas de perte de contrôle, de détournement par des « acteurs malfaisants », d'utilisation par des gouvernements à des fins de surveillance, de bouleversement du marché du travail, etc.

Alexandre Piquard

[Voir les commentaires](#)

[Réutiliser ce contenu](#)

Nos lecteurs ont lu ensuite

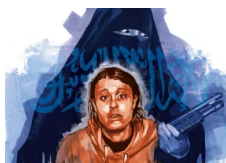
Au Musée Guimet, à Paris, l'art coréen de la beauté

Portée par le succès de la culture coréenne, la K-beauty a diffusé son esthétique et ses principes cosmétiques dans le monde entier. Au-delà du phénomène de mode, le Musée Guimet, à Paris, met en lumière un art de la beauté façonné par une histoire et des rituels séculaires.



Le destin d'Emilie König, d'« égérie » de l'EI à « revenante » du djihad

En 2013, la Lorientaise, alors âgée de 28 ans, quitte la France pour rejoindre l'organisation Etat islamique, en Syrie. Pendant quatre ans, elle participe à sa propagande et élève trois enfants. « Le Monde » a pu la contacter alors qu'elle séjourne à la prison de Rennes, où elle attend d'être jugée pour « association de malfaiteurs terroriste criminelle », en novembre.



Un bûcheron de 25 ans condamné à dix-huit ans de prison pour avoir piégé 39 adolescents en ligne, et violé certains d'entre eux

Théo Denner, qui revendique sa sympathie pour la mouvance néonazie, était jugé depuis le 9 mars pour un chantage à la « sextape » exercé sur des jeunes à Besançon. Le magistrat avait requis vingt ans d'emprisonnement, assortis d'une peine de sûreté des deux tiers.



C'est l'histoire d'une rue : Gravilliers, des ateliers de bijoux à « Little Wenzhou »

Situé dans le 3^e arrondissement parisien, le plus ancien quartier asiatique de la capitale abritait aussi tanneurs, joailliers, doreurs et graveurs. Aujourd'hui, les grossistes chinois tendent eux-mêmes à disparaître, sous l'effet du tourisme.



« Aucune démocratie ne peut se maintenir sur un récit qui contredit la société qu'elle prétend représenter »

Un collectif d'intellectuels, de chercheurs et de personnalités des mondes institutionnel et politique, parmi lesquelles Nicolas Cadène, Rokhaya Diallo et Akli Mellouli, s'inquiète du décalage culturel entre des élites politiques de plus en plus enclines aux fantasmes identitaires et la réalité des sociétés qu'elles entendent gouverner.



Pour Vincent Bolloré, l'audiovisuel public devrait être entièrement payé par la publicité

