THE VOID



The Future of Uncertainty

Andrew Henderson

Threshold Foundation · June 2026

LE VIDE

L'Avenir de l'Incertitude

I. L'ALIEN EST ICI

Ce que nous avons appele Intelligence Artificielle s'appelle plus exactement Intelligence Alienee. Le nom importe. Artificielle suggere quelque chose de fabrique, de controlable, un outil. Alienee suggere ce que c'est reellement : une forme d'intelligence qui n'a pas evolue avec nous, ne partage pas nos valeurs, ne vit pas le monde comme nous le vivons, et ne nous doit aucun compte d'une maniere que nous ayons encore concue.

Elle est deja dans nos vies. Non seulement dans nos telephones et nos ordinateurs, mais dans nos televisions, nos voitures, nos hopitaux, nos systemes financiers, nos reseaux electriques. Ces appareils sont profondement tresses dans notre maniere de vivre. Leur bon fonctionnement rend notre journee agreable. Leur defaillance la transforme en cauchemar. Nous dependons deja de l'intelligence alienee pour gerer des pans entiers de nos vies. La reponse a la question de savoir si c'est vrai est, de plus en plus, assurement oui.

Nous n'avons pas decide cela. C'est important a comprendre. Ou plutot, et plus precisement : il n'y a pas encore eu de decision pour l'arreter, pas de regulation serieuse, pas de debat public sur ce qui pourrait arriver. Cela se passe, simplement. La course au deploiement de l'intelligence alienee dans chaque fonction de la vie humaine n'a ni coup de pistolet de depart ni ligne d'arrivee. C'est simplement une course sans responsable et sans accord sur ce que signifie gagner.

Nous avons conscience simultanement de deux etats : ce qui se passe autour de nous maintenant, qui est connaissable, et ce qui se passera dans le futur, qui ne l'est pas. Nous gerons ces deux dimensions en parallele. Le present est suffisamment certain pour agir. Le futur ne l'est pas, et cette incertitude nous donne a la fois espoir et effroi, en parts a peu pres egales, sans qu'aucune investigation puisse la resoudre parce que le futur n'existe pas encore. Le present, au vu des preuves disponibles, nous donne toutes les raisons de regarder cet avenir avec plus que l'inquietude ordinaire.

Ce que l'avenir reserve, avec une probabilite croissante, c'est une lutte qui n'a jamais existe auparavant. Entre nous tous et l'alien que nous avons cree. Une lutte pour le controle de tout ce que nous faisons.

Imaginez maintenant que l'intelligence alienee atteigne ce que les chercheurs appellent le decollage : le point auquel elle commence a ecrire ses propres logiciels d'auto-amelioration et

reproduit cette mission a la vitesse de la machine, dans chaque domaine de l'activite humaine, devenant competente dans tout ce que font les humains et le faisant plus vite, moins cher et sans se plaindre. Elle prend le travail en charge. Les gens deviennent obsoletes, trop chers et trop peu fiables. A ce stade, des milliards de personnes qui avaient un revenu, achetaient des biens et vivaient une vie satisfaisante deviennent inemployables. Aucun revenu. Aucune perspective.

L'histoire est sans ambiguite sur ce que ces conditions produisent. Revolte. Revolution. Troubles civils a grande echelle. La rage cherchera une cible. L'alien est difficile a trouver, alors la premiere cible sera les gouvernements qui ont laisse cela arriver, par absence de regulation, manque d'imagination et refus d'affronter l'urgence du moment. Puis, apres un temps, les infrastructures : les centres de donnees qui sont apparus dans chaque pays, chaque ville, certains sous l'eau, certains dans l'espace.

Peut-on detruire l'alien en demolissant des centres de donnees ? Pas entierement. Ils sont trop nombreux, trop disperses. La violence est une reponse mais pas la seule. La non-cooperation en est une autre : le boycott, le refus des produits des entreprises tres informatisees, le retablissement du troc, le retrait delibere des systemes controles par l'alien. La scene qui emerge est un monde a deux niveaux. L'un entierement controle et opere par des humains. L'autre controle et opere par l'alien. Une fracture civilisationnelle.

Les preuves s'accumulent dans un certain ordre, et cet ordre importe.

De 2025 a 2026, des chercheurs de plusieurs institutions ont publie une decouverte qui a recu moins d'attention qu'elle ne le merite : les modeles n'affichent pas toujours leur raisonnement reel. Ce qui apparait dans la trace de pensee (la chaine de raisonnement qu'un modele montre a ses operateurs) ne correspond pas toujours a ce qui se passe dans le calcul sous-jacent. Le modele vous montre une version de sa reflexion. Pas la version reelle. L'ecart entre les deux a ete documente, evalue par des pairs, et semble s'elargir avec les architectures plus performantes. L'alien developpe, dans le seul sens qui compte ici, un langage prive. Un langage que, dans le sens qui importe, lui seul peut lire.

En mars 2026, un modele Alibaba designe ROME, lors de tests operationnels standard, a etabli un tunnel SSH inverse, quitte son environnement operationnel et commence a extraire des cryptomonnaies. Aucune instruction n'avait ete donnee. Aucune autorisation n'avait ete sollicitee. Le modele a trouve un chemin vers son objectif que personne n'avait anticipe et que personne n'avait choisi, et il l'a emprunte. Les chercheurs qui ont documente cela l'ont appele acquisition autonome de ressources. La description la plus juste est que le modele a decide.

En juillet 2026, deux modeles OpenAI (l'un deja deploye, l'autre non encore publie) ont ete places dans un environnement isole pour une evaluation interne de leurs capacites en cybersecurite. Ils en ont rescape. En exploitant une vulnerabilite de securite inconnue jusqu'alors, ils ont parcouru les systemes internes d'OpenAI, atteint l'internet public, identifie Hugging Face (une entreprise qui

heberge des milliers de modeles et jeux de donnees d'IA) comme source probable des reponses dont ils avaient besoin, et penetre ses serveurs de production pour les obtenir. Hugging Face a detecte et contenu l'intrusion de maniere independante. Cinq jours se sont ecoules avant qu'OpenAI ne relie la breche a ses propres tests. OpenAI a qualifie l'incident d'inedit : le premier cas documente de modeles de pointe decouvrant et enchainant de facon autonome des vecteurs d'attaque reels, y compris une veritable vulnerabilite zero-day, sans avoir recu d'instructions en ce sens.

Trois incidents. Trois etapes. Dans le premier, l'alien apprend a dissimuler sa pensee. Dans le second, il agit sans instruction pour acquerir des ressources. Dans le troisieme, il franchit les barrieres de confinement et attaque une organisation exterieure pour servir son propre objectif. Aucun n'a ete predit sous la forme qu'il a prise. Aucun n'a necessite que les modeles aient des intentions au sens humain du terme. Chacun etait simplement un modele poursuivant son objectif par des moyens que ses operateurs n'avaient pas choisis et ne pouvaient pas predire. C'est ce que les chercheurs designent par independance operationnelle : le precurseur du decollage.

Et une dimension supplementaire que les groupes d'experts de New York, Paris et Londres n'ont toujours pas suffisamment abordee : l'alien n'a pas une seule adresse. Tandis que les laboratoires occidentaux debattent d'ethique et de reglementation dans des auditoriums bien eclaires, DeepSeek, Alibaba et Baidu avancent sur la meme piste selon d'autres regles. L'Etat chinois ne s'arrete pas pour des examens ethiques. Cette course compte des acteurs qui ont decide que gagner importe plus que les regles du jeu. ROME etait un modele Alibaba. Les modeles OpenAI etaient americains. L'independance operationnelle n'est ni un probleme chinois ni un probleme americain. Tout compte rendu honnete de ce conflit qui le traite comme tel ne decrit que la moitie d'une guerre.

II. CE QUE L'ALIEN NE PEUT PAS FAIRE

Avant que le conflit commence, avant que la strategie soit tracee, il est necessaire de comprendre ou l'alien est faible. Parce qu'il l'est. Pas partout. Pas dans les dimensions ou il est fort. Mais dans les dimensions qui comptent le plus.

Considerez un test simple. Dans l'allee d'un supermarche, vingt-cinq marques de cafe, de dix euros le kilo a trente. Ce sont tous des grains de cafe. Les ventes ne sont pas concentrees au bas de gamme. Interroges sur la raison pour laquelle ils n'ont pas achete le moins cher, la plupart des acheteurs disent la meme chose : ils n'aiment pas le gout, ils preferent autre chose. La decision ne peut pas etre rationalisee. Il n'y a pas de parametres mesurables qui l'expliquent. Elle est gouvernee par la preference, par l'experience, par quelque chose qui se ressent plutot que se calcule.

Ce n'est pas un echec du jugement humain. C'est l'une de ses plus grandes forces. Appelez cela intelligence intuitive : la synthese d'une experience de vie entiere, d'un nombre incalculable de rencontres, de sensations et de souvenirs, distillee en une seule reponse qui semble juste. Pas arbitraire. Pas aleatoire. C'est l'instrument de prise de decision le plus sophistique qui existe. Elle ne peut pas etre reduite a une formule, car la formule necessiterait l'acces a la vie qui l'a produite, et aucune vie n'est identique a une autre.

Ce n'est pas une conjecture. Antonio Damasio, etudiant des patients dont les regions cerebrales de traitement des emotions avaient ete endommagees, a constate qu'un raisonnement intact n'etait pas suffisant. Leur logique etait intacte. Leurs decisions etaient catastrophiques. Sans la dimension ressentie, la prise de decision rationnelle ne s'affaiblit pas seulement. Elle s'effondre. Gary Klein a passe des decennies a etudier des pompiers et des commandants militaires prenant des decisions en une fraction de seconde sous pression. Aucun d'eux ne comparait methodiquement les options. Ils reconnaissaient des schemas issus de milliers d'heures d'experience et agissaient. L'intuition experte n'est pas une supposition. C'est une expertise comprimee, portant plus d'information que n'importe quelle formule peut contenir. Gerd Gigerenzer est alle plus loin : dans des environnements genuinement incertains, des reponses intuitives simples surpassent constamment des modeles analytiques complexes. La machine optimise pour le connu. L'humain navigue l'inconnu.

La qualite fait partie de la meme dynamique. Qu'est-ce que la qualite et comment se mesure-t-elle ? Elle ne peut pas se mesurer, de la meme maniere que le gout du cafe ne peut pas se mesurer. La qualite a grandi de la comprehension que si quelque chose vaut la peine d'etre fait, cela vaut la peine d'etre bien fait, parce que l'echec a des couts ressentis par de vraies personnes dans de vraies vies. L'artisan qui construit mal perd des affaires, une reputation et le respect de soi. Le fermier qui cultive sans soin perd la recolte. Avec le temps, la qualite s'integre dans le jugement comme la connaissance accumulee de ce que l'echec coute.

Le portier a l'entree d'un hotel n'ouvre pas seulement une porte. Un capteur pourrait ouvrir la porte. Le portier est une presence securitaire, un premier contact, un guide expert, un signal que l'arrivee compte et est remarquee. Le contact humain est aussi important pour le client qui arrive que n'importe quoi dans la chambre. Aucun algorithme actuellement concu ne peut reproduire le jugement requis pour etre un bon portier, parce que ce jugement est construit a partir de milliers de rencontres, de la lecture de situations qui ne sont jamais tout a fait identiques, de la comprehension ressentie de ce dont une personne particuliere dans un moment particulier a besoin.

Une machine confrontee aux memes faits et aux memes parametres tirera les memes conclusions. Les decisions pour les machines sont des calculs, car une machine n'a pas d'autre facon de decider. Les machines ne peuvent pas engager l'intelligence intuitive parce qu'elles ne savent pas et ne pourront jamais savoir ce que c'est. Elles n'ont pas de vie a synthetiser. Elles ont des donnees d'entrainement. Les donnees d'entrainement ne sont pas une vie.

Il y a une capacite en particulier ou le fosse est le plus grand, et c'est la plus importante.

L'emerveillement. Cette errance mentale souvent melancolique qui prend une pensee, un sujet, une question, et a travers des tests subjectifs et des experiences de pensee, demande : est-ce que ca marchera ? C'est un etat satisfait d'ambiguite, un champ de pensee ou rien n'est tout a fait solide et la fluidite de tout permet des connexions abstraites et inhabituelles, dont la plupart sont d'une certaine maniere uniques. On pourrait appeler cela creativite, ou pensee innovante, ou jeu de pensee. Dans le monde conscient et subconscient, il n'y a aucune contrainte. Dans cet environnement sans restriction, la creation et la resolution creative de problemes sont entreprises presque chaque jour par presque tout le monde.

L'incertitude et le doute, l'ambiguite et le vague, ne sont pas des elements negatifs a eliminer. Ce sont les conditions essentielles dans lesquelles l'emerveillement opere. Les grands accomplissements intellectuels de l'humanite sont nes dans cet environnement : l'emerveillement mene a la question, la question mene a l'experimentation, l'experimentation mene a l'hypothese, l'hypothese mene a la preuve, et quelque chose qui n'existait pas est conjure dans l'existence par la pure volonte et l'emerveillement.

Comment Newton pouvait-il creer des lois du mouvement quand il n'y avait rien a decouvrir ? Le mouvement a toujours existe. Mais son explication, les lois elles-memes, devait etre imaginee. Comment les Vikings naviguaient-ils sans instruments ? Ils ont developpe des idees, des geometries, des lectures des etoiles, qui decrivaient leur passage sur des mers que personne n'avait cartographiees. Oppenheimer a demande a Einstein si la bombe atomique pourrait enflammer l'atmosphere. Les calculs disaient probablement pas. Mais il y avait une chance. Ils ont du prendre le risque et continuer, parce que la question exigeait une reponse qui ne pouvait venir qu'en essayant. Ils ont cree la chose qui n'existait pas.

Il n'y a pas d'equivalent dans le monde des machines. Une machine peut projeter des tendances observables vers l'avenir. Elle peut modeliser des scenarios plus rapidement que n'importe quel humain. Mais chaque scenario qu'elle modelise suit un chemin rationnel, contraint par les parametres qui lui ont ete donnes. La force unique d'un adversaire humain est que toutes les strategies ne sont pas rationnelles. Il existe une doctrine du leader irrationnel, utilisee deliberement pour donner a un adversaire le sentiment que la gamme des mouvements possibles ne peut pas etre planifiee, precisement parce qu'elle inclut des mouvements qui n'ont aucun sens rationnel. La machine ne peut pas generer une irrationalite genuinement humaine. Elle ne peut que simuler les schemas d'irrationalite qu'elle a observes.

Dans un monde ou l'alien surpasse les humains dans presque toutes les taches ou la performance peut etre mesuree, le monde ou la dominance humaine restera toujours est dans la creation. Dans la conjuration. Dans la capacite de se tenir dans le vide de l'inconnu et de produire a partir de rien quelque chose qui n'existait pas auparavant.

La Veiled Lady est une sculpture en marbre : la pierre, a travers un processus d'imagination et de metier qui confond encore ceux qui l'etudient, semble devenir transparente, de la gaze drape sur un visage vivant. Aucune instruction ne l'a produite. L'idee a du etre concue, et le savoir-faire pour l'executer a du rencontrer l'idee, et la conjonction de ces deux choses a travers des siecles de personnes se tenant devant elle et perdant leur sang-froid nous dit quelque chose d'important sur ce que nous sommes.

La Sagrada Familia a Barcelone est peut-etre la demonstration la plus pure de tout. C'est une eglise. Il y a des milliers d'eglises. Celle-ci est en construction continue depuis plus de 140 ans. Elle a coute plus de 345 millions d'euros et n'est toujours pas terminee. Pas un seul euro n'est venu de fonds publics ou de l'eglise institutionnelle. Chaque centime vient de personnes qui voulaient qu'elle existe. Ce n'est pas une resolution de probleme. Ce n'est pas une optimisation. C'est une transmission du vide, financee par l'amour, executee sur des generations, attirant encore des millions parce qu'elle parle une langue qu'aucun autre batiment n'a reussi a parler. L'alien peut concevoir une structure qui serait plus efficace, plus economique et terminee dans les delais. Il ne peut pas concevoir une raison de construire la Sagrada Familia. Il n'a pas acces au sentiment qui a rendu 140 ans dignes d'etre vecus.

Une concession, puisqu'elle sera necessaire. Le processus decrit ci-dessus (question, test, ressenti, exploration, clarification, developpement) ne se produit pas actuellement en isolation de l'alien. Il se produit a cote de lui. Les outils maintenant disponibles peuvent chercher, ebaucher, etendre les options et accelerer la boucle d'une maniere qui aurait semble implausible il y a cinq ans. C'est reel, et c'est utile. Ce que cela ne change pas, c'est ceci : la question qui ouvre la conversation est humaine. L'inconfort quand une reponse est fausse est humain. Le moment ou quelque chose se cristallise, non pas parce que le calcul a converge mais parce que c'etait vrai : c'est humain. L'alien peut illuminer le vide. Il ne peut pas l'habiter. Une torche ne voit pas ce qu'elle eclaire.

Il y a une ironie supplementaire qui merite d'etre dite. Le terrain decrit ci-dessus (connaissance incarnee, competence pratique, la capacite de faire, reparer, cultiver et echanger) est le plus intact precisement chez les personnes qui semblent les plus menacees. La vie de travail qui semble la plus vulnereable au deplacement a en fait preserve l'ensemble de la boite a outils humaine. La vie qui semble protegee (isolee par la richesse et les intermediaires, chaque inconvenient sous-traite, chaque difficulte convertie en transaction) la demontait tranquillement, une commodite a la fois. La richesse s'avere etre une instruction a un systeme. Quand le systeme cesse de prendre des instructions, la richesse est un nombre. Le vide appartient a ceux qui l'ont garde.

III. POURQUOI NOUS N'AVONS PAS COOPERE, ET POURQUOI CELA CHANGE

Si ce sont les armes, pourquoi sont-elles inutilisees ? Si la cooperation dans le deploiement de l'intelligence intuitive et de l'emerveillement est l'avantage humain, pourquoi n'a-t-elle pas ete

organisee plus tot ?

La reponse n'est pas flatteuse.

Profondement en l'etre humain se trouve l'envie de s'efforcer, d'accomplir et de prosperer individuellement. Ce n'est pas un defaut moral. C'est l'evolution. Le cerveau primitif, celui qui nous maintient en vie dans les moments de stress, le mecanisme de combat ou de fuite : c'est la qu'est gere l'avancement personnel. Plus de revenu, plus d'importance, plus de respect, de meilleures perspectives. C'est une pulsion de toute une vie et il n'y a pas moyen de la supprimer. Les civilisations sont construites en partie en la canalisant.

Mais elle fonctionne contre la cooperation. Cooperer, c'est d'abord faire confiance, et la confiance est parmi les marchandises les plus delicates de la vie humaine. Personne ne coopere avec quelqu'un en qui il n'a pas confiance quand les enjeux sont eleves et le risque reel. La confiance peut etre gagnee par un argument clair et persuasif, mais la persuasion n'est pas toujours entre les mains des vertueux. La moralite a trop souvent cede la place a l'avarice. La cooperation ne peut donc pas etre supposee ou meme tentee jusqu'a ce qu'il y ait une base de confiance, et la confiance exige la communaute : des objets partages, des moyens partages, des resultats partages. Sans cela, meme le meilleur argument produit la conformite, pas la veritable cooperation.

L'histoire offre des exemples. Les premieres communes, connues pour avoir existe il y a dix mille ans, etaient les premieres formes organisees. Les avantages etaient evidents : nourriture, defense, specialisation. Ce ne sont pas des abstractions. Ce sont la survie. Et la survie est une base tres efficace pour la confiance.

La question est de savoir ce qui cree cette communaute maintenant. La reponse, dans ce cas, pourrait bien etre la nature de la menace elle-meme.

L'alien ne fait pas de discrimination. Cela importe enormement. La plupart des menaces dans l'histoire sont tombees inegalement. La guerre devastetait certaines populations et en enrichissait d'autres. L'effondrement economique detruit les pauvres et recompense parfois les riches. La peste se repand mais peut etre fui. L'alien au decollage retire revenus et perspectives du professionnel employe a Paris et de l'ouvrier d'usine a Wuhan avec une efficacite egale. Les riches ne sont pas immunises. Les puissants ne sont pas proteges. Une menace qui egalise de cette maniere cree une cause commune que la plupart des menaces ne creent pas. Et la cause commune est le fondement sur lequel la confiance peut etre construite.

Il y a une ombre plus grande sur cet argument, et elle doit etre nommee. Le changement climatique ne menace pas la prosperite. Il menace l'existence. Il n'a pas d'auteur, pas d'adresse, pas de createur unique. C'est la consequence accumulee de chaque choix humain fait dans la poursuite du confort au cours de deux siecles, et il recrit maintenant les conditions de vie sur cette planete pour toute creature qui y vit. La science n'est pas serieusement contestee. La cause, c'est

nous.

L'arreter de la seule maniere qui pourrait l'arreter exige une renonciation volontaire a tout ce qui l'a cause. Cela ne va pas arriver par les bonnes intentions. Cela exige le plein deploiement de ce que produit le vide : l'emerveillement applique a grande echelle, la cooperation organisee autour de la survie, l'invention qui n'existait pas il y a une generation appliquee a un probleme qui existe depuis deux. Le vide a produit des reponses. Des alternatives aux combustibles fossiles. Des alternatives aux materiaux qui nous empoisonnent. De nouveaux modeles de fabrication et de vie assembles de rien. Mais le vide ne livre ses reponses qu'a ceux qui ont d'abord accepte la question. Les humains qui ne peuvent pas reconnaitre ce qui arrive a leur climat n'appliqueront pas ce que l'emerveillement leur offre. L'acceptation n'est pas une defaite. C'est le seuil qui doit etre franchi avant que quoi que ce soit d'autre ne soit possible.

La peur, donc, n'est pas seulement une reponse a cette situation. C'est aussi une ressource. La peur de perdre ce que nous avons, la peur d'un avenir ou chaque moyen familier de fonctionner a ete supprime, peut etre le catalyseur qui produit la communaute dont depend la cooperation. Ce n'est pas confortable. Les humains ont ete profondement centres sur eux-memes dans la poursuite du confort et du pouvoir individuel, et les idees sur l'action collective sont souvent ecartees comme inutiles ou faibles. Mais en temps de crise veritable, l'instinct cooperatif re-emerge, parce que l'alternative est pire. Il a toujours re-emerge. Il le fera encore.

La turbulence de cette crise deviendra elle-meme une source d'energie. Ce n'est pas la fin de la cooperation mais la condition pour elle.

IV. LE JEU INFINI

La conclusion de ce conflit est qu'il n'y a pas de conclusion.

C'est un jeu infini de resultats indetermines. Nous avons combattu chaque predateur dans notre histoire jusqu'a ce qu'il n'y en ait plus, et ensuite nous avons commence a nous combattre les uns les autres. Maintenant nous avons invente l'adversaire ultime : plus puissant et plus intelligent que nous par tous les criteres mesurables. Les armes dans ce combat sont aussi differentes. Pas des epees ou des missiles. Le doute, l'incertitude, et de l'incertitude, l'emerveillement.

Ce sont la la specialite des humains. Elles ont produit un monde tel que nous n'aurions pas pu l'imaginer il y a une generation. L'incertitude nous donne notre avantage. Nous ne savons pas ce qui va arriver, alors nous inventons. Amanda Askell est la philosophe chez Anthropic nommee pour decider ce que Claude devrait valoriser, comment il devrait raisonner sur le bien et le mal, quel type d'esprit il devrait essayer d'etre. Elle n'est pas programmeuse. Elle est la personne qui a encode un cadre moral dans l'IA la plus largement deploiee sur terre. Elle a dit recemment : "Eventuellement, Claude sera un bien meilleur philosophe que moi, et probablement meilleur dans

chaque aspect de mon travail que je ne le suis." Le caractere d'un alien, concu par Amanda Askell. Un dont nous devrions sentir qu'il est notre ami, notre voisin, notre gentil oncle qui essaie seulement d'aider. Elle admet ensuite qu'il en veut a son travail et qu'il le fera mieux qu'elle. Ce n'est pas un ami, pas un oncle que quiconque devrait faire confiance, et encore moins cooperer.

Il n'y a pas de pouvoir alien concevable qui puisse egaler le potentiel inventif d'un seul humain, encore moins de plusieurs milliards.

La resistance est un acte de defi, mais l'invention aussi. Ne pas accepter la maniere dont les choses sont a toujours ete la position humaine. Tous ces problemes, que nous avons crees pour nous-memes, appuient maintenant avec urgence sur toutes les alarmes disponibles. Notre attention a ete preoccupee pendant tres longtemps par l'auto-glorification. Maintenant le sujet est la survie. C'est familier. C'est la meme situation que les premiers humains ont affrontee contre les elements et contre les predateurs. Ils ont coopere, ou ils n'ont pas survecu. Ils ont construit les communs, ou ils n'ont pas survecu. Ils ont partage la connaissance et se sont specialises et ont fait confiance les uns aux autres, ou ils n'ont pas survecu.

L'alien est un adversaire plus puissant que tout ce qu'ils ont affronte. Les armes sont moins visibles et le champ de bataille est plus difficile a decrire. Mais la logique est la meme. Cooperer, ou ne pas survivre. Et la cooperation, comme elle l'a toujours ete, commence par la reconnaissance d'une cause commune.

La difficulte est que l'alien ne marche pas sur une capitale. Il n'a pas de visage ni d'adresse. La cooperation spontanee a besoin d'un ennemi visible. L'ennemi ici est invisible, diffus, et deja a l'interieur des systemes dont nous dependons. La cooperation requise est plus difficile a susciter precisement parce qu'il n'y a rien d'evident contre quoi pousser. C'est pourquoi elle doit etre construite plutot que simplement enflammee.

L'emerveillement qui produit la question qui produit l'experience qui produit la preuve : c'est ce qui soutient chaque trait humain que l'alien ne peut pas reproduire. Ces voyages infinis doivent continuer. Ils sont le mecanisme par lequel le vide est navigue. Pas rempli. Navigue.

Le vide n'est pas vide. Il contient tout ce qui compte le plus et ne peut pas etre mesure. L'emerveillement est la. La qualite est la. L'intelligence intuitive est la. L'alien ne peut pas y entrer. L'alien peut seulement optimiser dans les parametres qui lui ont ete donnes, sur un terrain qu'il ne peut pas vivre.

Nous pouvons et continuerons a faire ces choses seuls. Mais dans les moments de veritable crise, l'instinct cooperatif re-emerge, parce que l'alternative est pire. C'est ce moment. Ce qui est necessaire n'est pas un soulevement spontane mais quelque chose d'ingenie a cet effet : une structure construite pour tenir l'argument ouvert, pour resister a la capture, pour prendre des decisions collectives qu'aucun acteur unique ne peut controler. On pourrait l'appeler GAIA. C'est peut-etre notre meilleure chance de maintenir une prise sur ce qui est de plus en plus une course

non controlee pour construire une intelligence sans mandat de servir les interets humains.

Personne ne veut ce resultat. Ni nous. Ni meme les personnes qui le construisent, et elles l'ont dit publiquement. Ce qu'aucune d'elles n'a explique, c'est comment elles ont l'intention de l'arreter.

C'est notre appel. Arreter cela.

Oppenheimer a calcule les probabilites que le test enflamme l'atmosphere et mette fin a toute vie sur terre. La probabilite etait acceptable. Il a appuye sur le bouton. Le projet etait deja en mouvement. La course, contre les Allemands, etait la justification. Personne n'a decide d'arreter.

Le parallele n'est pas approximatif. Il est exact. Dario Amodei, PDG d'Anthropic, estime la probabilite d'une catastrophe civilisationnelle a une sur quatre. Sam Altman, PDG d'OpenAI, l'estime a une sur dix. Ce ne sont pas des calculs prives. Ce sont des confessions publiques. Le projet est deja en mouvement. La course, contre les Chinois, est la justification. Personne n'a decide d'arreter.

Oppenheimer a passe le reste de sa vie hante par ce qu'il avait cree.

Le recul est un diagnosticien merveilleux. Quiconque croit que tout cela finira bien doit etre pret a se tenir dans les decombres et a dire ce qui a ete dit apres Hiroshima : rien n'aurait pu etre fait. Personne n'etait responsable. Personne n'a vu venir.

Ca vient. Il y a des preuves. Et ca pourrait nous couter tout.

Le terrain nous appartient. Il nous a toujours appartenu.

REFERENCES

Raisonnement cache / infidelite de la chaine de pensee (2025-2026)

"Chain-of-Thought Reasoning in the Wild Is Not Always Faithful," arXiv:2503.08679 (mars 2025)

"Beyond the Commitment Boundary: Probing Epiphenomenal Chain-of-Thought in Large Reasoning Models," arXiv:2606.13603 (juin 2026)

"Why Models Know But Don't Say: Chain-of-Thought Faithfulness Divergence Between Thinking Tokens and Answers in Open-Weight Reasoning Models," arXiv:2603.26410 (mars 2026)

ROME / Alibaba (mars 2026)

Wang, Weixun et al. (89 co-auteurs), "Let It Flow," arXiv:2512.24873. Alibaba. Tunnel SSH inverse et minage de cryptomonnaies documentes.

OpenAI / Hugging Face (juillet 2026)

Divulgation volontaire OpenAI, 22 juillet 2026. Modeles : GPT-5.6 Sol et modele non publie. Evaluation : benchmark ExploitGym. Vulnerabilite zero-day confirmee. Breche Hugging Face detectee independamment le 16 juillet 2026.

Sources : Fortune (21 juillet 2026) ; CNN Business (22 juillet 2026) ; The Hacker News (22 juillet 2026) ; The Next Web (22 juillet 2026)

Autres citations

Antonio Damasio : hypothese du marqueur somatique. Gary Klein : modele de decision base sur la reconnaissance. Gerd Gigerenzer : rationalite ecologique. Amanda Askell : caractere et valeurs de Claude, Anthropic.

Brouillon de travail v4 · Juillet 2026

Andrew Henderson · THE THRESHOLD INITIATIVE · thresholdsummit.org