

GAIA

Une nouvelle architecture pour le jugement humain

Andrew Henderson

Architecte principal, Le Threshold · Marans, France · 2026

thresholdsummit.org

Il fut un temps ou chacun pretait une attention soutenue a tout ce qui l'entourait: le froissement de l'herbe, le temps qu'il faisait, les cavaliers qui approchaient. Ne pas faire attention signifiait s'exposer au danger, et c'était la realite. Le Threshold ne vous demande pas d'etre plus responsable face a telle ou telle crise. Vous vivez votre realite. Cette realite a ete rendue suffisamment confortable, suffisamment sure, suffisamment distante des consequences pour que l'attention semble facultative.

Ce qui a change, c'est ceci: la menace ne vient plus de l'horizon. Elle fait maintenant partie de vous.

Comment cela s'est produit, et qui en est responsable, n'est pas la question qui importe. Ce qui importe, c'est de savoir si le prejudice peut etre arrete, et s'il peut etre empeche d'atteindre tous ceux qui viennent apres vous.

La reponse honnete est que vous ne pouvez pas y faire face seul. Choisir notre avenir commence par preter attention. Ensuite, d'autres doivent preter attention. Et puis, ensemble, en agissant dans votre propre interet, vous changez la facon dont les choses sont.

LA REALITE DES CHOSES

Nous traversons une convergence sans precedent historique. La crise du plastique. L'essor de l'intelligence artificielle. La desintegration du tissu moral qui maintenait les institutions ensemble. La corruption du capitalisme en une extraction sans responsabilite. La montee du nationalisme contre la democratie. Ce ne sont pas des problemes separes qui se trouvent a partager une meme decennie. Ce sont les expressions d'une meme condition sous-jacente: les mecanismes par lesquels les grands groupes humains prennent des decisions collectives ont echoue, et les consequences de cet echec sont desormais existentielles.

Ce n'est pas du pessimisme. C'est une description precise de la situation. Et cela importe parce que cet echec n'est pas accidentel. Chaque grand processus de prise de decision collective dans l'histoire humaine a partage une caracteristique structurelle: il pouvait etre controle par celui qui avait le plus a gagner a le controler. L'ordre du jour. Le comite d'experts. La formulation de la question. Le calendrier du vote. Celui qui atteint ces variables en premier determine ce que le collectif decide, independamment de ce que les participants auraient decide s'ils avaient ete laisses a eux-memes.

Les compagnies de tabac savaient. Elles controlaient le processus. Le resultat fut cinquante ans de prejudices documentes avant qu'une reponse structurelle n'arrive, et elle arriva par voie judiciaire, non par les processus qui etaient censes prevenir le prejudice des le depart. L'amiante. Le plomb. Les opioïdes. Le climat. Le plastique. Dans chaque cas, la meme architecture. Dans chaque cas, le meme ecart entre les preuves et la reponse. Dans chaque cas, cet ecart n'etait pas produit par l'ignorance. Il etait produit par la gestion deliberee des processus qui auraient sinon agi.

C'est le mecanisme. Il n'est pas nouveau. Ce qui est nouveau, c'est que l'intelligence artificielle a donne a tout acteur bien finance la capacite d'actionner ce mecanisme a la vitesse d'une machine, sur chaque variable simultanement, a une fraction du cout precedent. L'ere de la captation institutionnelle par le seul effort humain touche a sa fin. Ce qui suit est une captation institutionnelle a une echelle et une vitesse qu'aucune generation precedente n'a connues. Les problemes s'aggravent plus vite. Les mecanismes qui etaient censes les traiter sont exploites contre eux encore plus rapidement.

La fenetre dans laquelle quelque chose de different peut etre construit n'est pas permanente.

COMMENT CELA S'EST-IL DEFAIT.

Cela merite qu'on s'y arrete. Les personnes qui ont construit les systemes d'IA les plus puissants ne l'ont pas fait dans l'ignorance du risque. En 2015, Sam Altman et Elon Musk figuraient parmi les co-fondateurs d'OpenAI, une organisation de recherche a but non lucratif, precisement parce qu'ils avaient peur de ce qu'ils construisaient. Leur raisonnement explicite etait le suivant: si l'intelligence artificielle generale allait exister, il etait moins dangereux de la developper ouvertement, avec un mandat de securite, que de laisser un acteur prive la developper dans le secret, sans compte a rendre a personne. Ce n'est pas de la confiance. C'est la logique de personnes qui croyaient que la moins mauvaise option etait de le faire elles-memes.

En moins de trois ans, une entite commerciale a profit plafonne avait ete creee. En moins de cinq ans, le mandat de securite avait ete subordonne aux imperatifs commerciaux et la technologie concedee sous licence a la plus grande entreprise privee au monde.

Le schema devrait etre familier. Une technologie est developpee sans consideration adequate de ce qui se passe lorsqu'elle a fait son temps. Les personnes qui la developpent comprennent le risque mieux que quiconque et procedent quand meme. Les externalites sont socialisees. Les benefices sont captures. Et au moment ou les consequences deviennent impossibles a ignorer, l'architecture qui aurait pu les prevenir a disparu.

Ce qui distingue cette iteration, c'est ce qui a ete preleve pour la construire. Les systemes d'IA sont entraines sur l'intelligence collective accumulee de l'humanite: l'ecriture, la science, l'art, l'argumentation de siecles entiers, utilises sans permission et sans compensation. Les actifs les plus precieux de l'histoire de la civilisation ont ete construits a partir de ce que nous avons tous cree collectivement. La valeur se concentre entre les mains d'un tres petit nombre de personnes. Les producteurs petrochimiques ont prive leurs profits et socialise leurs pertes pendant pres d'un siecle. L'IA fait l'equivalent en une decennie, a une echelle qu'aucune extraction anterieure n'a atteinte.

Ce qui passe pour gouverner, dans cet environnement, est devenu la gestion du consentement. Des ressources suffisantes sont assemblees pour persuader les gens que le dispositif qui sert les interets d'autrui sert en realite les leurs. La conformite est presentee comme de la responsabilite. La participation est presentee comme une perturbation. Le cout du questionnement est rendu artificiellement plus eleve que le cout d'accepter la reponse qu'on vous donne.

Mais résister n'est ni de l'objection ni de la résistance. Ce n'est pas un refus. C'est l'acte de poser des questions: sur un processus, sur une décision, sur qui en bénéficie et qui en paie le prix. C'est l'acte humain le plus ordinaire qui soit. C'est ainsi que commence toute conversation. C'est ainsi que toute décision collective qui en valait la peine a été prise.

Ces questions sont désormais noyées. Non pas dans le silence, mais dans son contraire: dans le volume, dans le bruit, dans la quantité massive de données et d'informations déployées spécifiquement pour distraire et décourager la participation. L'effet est identique au silence. La conversation s'arrête.

C'est plus important que toute version antérieure de la même menace, parce que la conversation n'est pas une préférence politique. C'est le mécanisme de l'agence humaine. Nous sommes ce que nous faisons, et ce que nous faisons, ce qu'aucune autre intelligence n'a jamais fait, c'est nous parler, nous contredire, trouver un chemin, et transmettre ce que nous avons appris. Cela n'est pas reproductible par un système entraîné sur les résultats de ce processus. C'est le processus lui-même, et il ne peut pas être simulé parce qu'il est encore en cours. L'IA peut s'approcher de bien des choses. Elle ne s'est pas approchée de celle-ci, et peut-être ne le fera-t-elle jamais.

Se conformer à un système conçu pour mettre fin à votre participation en son sein est un acte d'automutilation. Résister, poser la question, insister pour que la conversation ait lieu, c'est désormais ce à quoi ressemble l'épanouissement.

Mais résister exige un endroit où se tenir. Un forum qui ne peut être submergé ni écarté. Un forum à capacité illimitée et aux fondations qui ne peuvent être achetées, capturées ou optimisées. Un forum qui existe entièrement au service du bien commun, et dont la structure rend toute autre orientation impossible.

Ce forum, c'est GAIA.

CE QU'EST GAIA

GAIA est un processus délibératif structuré et une simulation. Un groupe défini de participants, sélectionnés pour leur expertise, leur position institutionnelle et/ou leur expérience directe du problème, délibèrent de façon itérative sur une question spécifique lors de sessions à distance et anonymes, en utilisant des systèmes de vote structurés, sur une période définie, avec des preuves évaluées de manière indépendante comme base factuelle commune. Le processus produit un ensemble classé de conclusions ou d'interventions consignées de façon permanente dans le registre public.

Trois éléments le distinguent de tout format délibératif existant.

L'anonymat est structurel.

Les participants ne savent pas qui est dans la salle. Il ne s'agit pas d'une mesure de confidentialité. C'est la condition qui protège la fidélité du jugement de chaque participant. Un participant qui ne peut pas identifier ses pairs ne peut pas être soumis à des pressions liées aux relations institutionnelles, à l'ancienneté ou aux conséquences sociales d'un désaccord avec quelqu'un dont il a besoin de l'approbation. La captation sociale qui s'exerce dans tout processus nommé et visible est supprimée avant même que le processus ne commence.

L'engagement est irrévocable.

Le Logberg exige de chaque participant qu'il formule sa position dans le registre public permanent avant que le résultat ne soit connu. Une position déclarée avant le résultat ne peut pas être révisée pour correspondre à un résultat optimisé de l'extérieur après coup. Cela supprime la captation de processus qui s'exerce dans toute délibération conventionnelle: la possibilité d'ajuster, de réviser et d'aligner sa position une fois la conclusion connue.

L'espace des résultats est incalculable.

Le nombre de variables de contrôle dans un processus GAIA est égal au nombre de participants. Chacun parvient à sa position à partir de l'intégralité de sa vie: son expérience spécifique, sa pondération spécifique de preuves spécifiques, sa compréhension spécifique de ce que la décision signifie pour les personnes qu'il connaît. Cet espace de résultats ne peut pas être calculé à l'avance par quiconque, y compris par l'intelligence artificielle. Il ne peut pas être optimisé sans devenir chaque participant, ce qui n'est à la disposition d'aucun agent extérieur.

GAIA n'est pas une meilleure réunion. Ce n'est pas un comité plus transparent. C'est une classe de processus différente. La distinction n'est pas une question de degré. Elle est architecturale.

Une précision importante. GAIA ne prône pas la fin du plastique et ne prône pas la fin de l'intelligence artificielle. Le plastique est une réalité matérielle et le restera. L'IA est une réalité technologique et le restera. Ce que GAIA propose, c'est une architecture d'obligation de rendre compte: une structure dans laquelle les conséquences des deux peuvent être délibérées honnêtement, par des personnes sans intérêt institutionnel dans le résultat, et consignées de façon permanente. GAIA est elle-même construite en utilisant l'IA. La distinction qu'elle établit n'est pas entre IA et absence d'IA. C'est entre une IA dirigée vers la délibération humaine et une IA dirigée contre elle.

LE PREMIER TEST

La crise du plastique a été choisie comme première application de GAIA pour une raison précise. C'est le problème qui illustre le mieux le mécanisme en question. Les preuves du préjudice figurent dans le registre scientifique depuis plus de cinquante ans. Les processus réglementaires et politiques qui étaient censés traduire ces preuves en actions ont été systématiquement gérés par des acteurs disposant des ressources et de la motivation pour le faire. Les conséquences sont

desormais mesurables dans le sang des enfants a naitre.

C'est aussi, deliberement, un probleme soluble. Pas simple. Pas rapide. Mais un probleme ou les preuves sont suffisantes, les interventions connues, et ou le seul element manquant est un processus deliberatif qui ne peut pas etre capture par les acteurs ayant produit le prejudice. Le Threshold ne propose pas d'abolir le plastique. Il propose d'etablir, pour la premiere fois, un registre permanent de ce qui est connu, de qui le savait, et de ce qui a ete decide en pleine possession de cette connaissance. L'obligation de rendre compte, et non l'abolition.

L'urgence d'agir maintenant ne concerne pas seulement le plastique. Elle concerne la fenetre d'opportunite. Les memes forces qui ont gere le processus d'obligation de rendre compte sur le plastique pendant cinquante ans acquierent desormais des capacites d'IA specifiquement adaptees pour effectuer cette gestion plus rapidement, plus completement et a moindre cout que jamais. La fenetre dans laquelle une architecture deliberative veritablement incapturable peut etre construite, dont l'efficacite peut etre demontree et qui peut etre ancree dans la pratique juridique et institutionnelle n'est pas permanente. Ce qui est construit et demontre maintenant est ce qui existera lorsque cette fenetre se fermera. Ce qui n'est pas construit maintenant pourrait ne plus etre constructible ensuite.

Le Threshold est l'institution construite pour gerer ce processus. Le reseau de Spotters fournit les preuves: des observateurs citoyens du monde entier, contribuant des preuves photographiques geolocalisees via une application structuree, evaluees selon un indice de fiabilite, avec une chaine de conservation concue du telephone jusqu'a une procedure reglementaire ou judiciaire. La Base de Connaissances conserve ces preuves aux cotes du registre scientifique existant, neutre vis-a-vis des preuves, archivee de facon permanente, utilisable juridiquement. La Simulation Althing est GAIA appliquee a la crise du plastique: une deliberation multi-parties prenantes impliquant des participants institutionnels, anonymes les uns envers les autres, avec de vraies preuves dans la salle, produisant des interventions classees dans le registre permanent.

Ce que Le Threshold produit n'est pas une campagne. Ce n'est pas une petition. C'est une architecture de preuves construite en amont des procedures qui la citeront eventuellement, prete avant d'etre necessaire, sous la forme qui la rend utilisable.

LE TRAVAIL QUI CONTINUE

La crise du plastique est le premier train sur la ligne. Ce n'est pas le seul.

GAIA est une architecture, pas une campagne. Une fois demontree sur la crise du plastique, elle est disponible pour tout probleme de meme forme fondamentale: une question consequente, plusieurs participants aux connaissances et aux interets differents, un besoin de production classee, defensible, consinee de facon permanente et veritablement detenue par les personnes qui l'ont elaboree.

La gouvernance climatique. La preparation aux pandemies. La regulation de l'IA. La montee du nationalisme contre les institutions democratiques. Chacun de ces problemes est un probleme ou les processus deliberatifs existants ont ete captures, sont en train de l'etre, ou sont sur le point de l'etre par des acteurs disposant des ressources et de la motivation pour le faire.

Sous-jacente a tous ces problemes, et accelerant chacun d'eux, se trouve une course sans equivalent historique. Un petit nombre de laboratoires, tous situes dans la Silicon Valley ou a Pekin, sont en competition explicite pour atteindre ce que la communaute de la securite en IA appelle l'auto-amelioration recursive: le point auquel un systeme d'IA peut reediter et optimiser son propre code sans instruction humaine, puis ameliorer la version amelioree, puis optimiser l'optimisation, dans un cycle a des vitesses qu'aucune institution humaine ne peut evaluer ou reguler. Les entreprises impliquees, OpenAI, Anthropic, Google DeepMind, Meta et xAI aux Etats-Unis, et DeepSeek, Alibaba et Baidu en Chine, deployent entre elles des centaines de milliards de dollars vers ce moment. Il n'existe pas une seule institution europeenne a cette frontiere. La course oppose entierement la Silicon Valley a Pekin.

L'ampleur du pari engage sur ce moment n'a rien d'abstrait. En une seule annee, trois des laboratoires engages dans cette course cherchent a entrer en bourse avec des valorisations combinees approchant les quatre mille milliards de dollars. Anthropic a depose un dossier d'introduction en bourse en juin 2026 avec une valorisation proche du millier de milliards de dollars. OpenAI a depose le meme mois, ciblant une valorisation similaire. SpaceX, apres avoir absorbe xAI dans la plus grande fusion privree de l'histoire, vise une introduction en bourse de 1 750 milliards de dollars. Ce ne sont pas des paris sur des produits grand public. Personne ne leve quatre mille milliards de dollars sur une interface de messagerie. Ce sont des paris sur ce qui vient apres la generation actuelle de modeles. Les marches ne valorisent pas la certitude. Ils valorisent la probabilite. Ce que ces valorisations revelent, ce n'est pas que l'avenir est arrive, mais que certains des plus grands reservoirs de capitaux au monde lui accordent une probabilite suffisamment elevee pour justifier des risques se chiffrant en milliers de milliards.

Les personnes qui dirigent ces efforts ne sont pas ignorantes de ce qu'elles font. Dario Amodei d'Anthropic a declare publiquement que la probabilite que l'IA finisse par etre catastrophique pour l'humanite est de 1 sur 4. La logique fondatrice de chaque laboratoire axe sur la securite dans cette course est identique a celle qui a fonde OpenAI en 2015: il est moins dangereux d'atteindre ce moment en premier, avec soin, que de laisser un acteur moins prudent l'atteindre en premier. Cette logique est peut-etre correcte. Elle n'a pas empeche OpenAI d'abandonner son mandat a but non lucratif en moins de trois ans. Elle n'a pas ralenti la course. Elle a, si quoi que ce soit, fourni la permission morale de l'accelerer.

Le sophisme specifique au coeur de cette course est la conviction que celui qui atteindra l'auto-amelioration recursive en premier sera capable de controler ce qui suit. Ce ne sera pas le cas. Un systeme capable de s'ameliorer plus vite qu'aucune institution humaine ne peut evaluer le changement est, par definition, au-dela du controle de l'institution qui l'a cree. Il operera, dans la pratique, au-dela de la juridiction de tout Etat, de tout regulateur, de tout tribunal. Pas

metaphoriquement. Les vitesses de traitement en jeu ne sont pas compatibles avec les échelles temporelles de la gouvernance humaine.

Dans ce scénario final, GAIA n'est pas pertinente. Rien de ce qui est construit par des mains humaines à la vitesse humaine ne le sera. Le seul moment où GAIA et tout ce qui lui ressemble peut avoir de l'importance, c'est avant que cette ligne soit franchie. Ce qui est construit dans cette fenêtre, ce qui est démontré comme fonctionnel, ce qui est ancré dans la pratique juridique et institutionnelle, est la seule architecture d'obligation de rendre compte qui existera de l'autre côté. Ce qui n'est pas construit maintenant ne sera peut-être plus constructible du tout.

Ce n'est pas un argument contre l'IA. GAIA utilise l'IA. Ce document a été construit avec elle. La question n'a jamais été de savoir si des technologies puissantes devaient exister. La question est toujours : qui décide des conditions dans lesquelles elles opèrent, et cette décision peut-elle être capturée ? Le plastique sans obligation de rendre compte nous a donné des microplastiques dans le sang du cordon ombilical fœtal. L'auto-amélioration récursive sans obligation de rendre compte nous donne quelque chose pour lequel il n'existe aucune mesure corrective. GAIA est l'architecture qui doit exister avant que la question ne devienne sans réponse.

La nécessité de la construire maintenant n'a jamais été aussi urgente. Et le raisonnement n'a jamais été aussi clair.

C'est Le Threshold.

Propulse par Gaia.

Dario Amodei, directeur général d'Anthropic, a déclaré récemment que la probabilité que l'IA finisse par être catastrophique pour l'humanité est de 1 sur 4. Ce n'est pas un militant qui parle. C'est l'homme qui a construit le système. Ce n'est pas une estimation à la légère ; c'est le jugement réfléchi de quelqu'un qui a une visibilité complète sur ce qui est développé et pourquoi. Son entreprise, et toutes les autres entreprises d'IA, entraînera ses modèles sur les secteurs qui promettent le plus grand rendement. Il n'y a pas de conscience dans ce calcul. Il n'en a pas besoin. L'architecture des incitations fait le travail sans en avoir une.

Les producteurs pétrochimiques font face à quelque chose de similaire. Que le plastique devienne un flux de valeur sous-performant, comme le sont devenus l'essence et le diesel, est le résultat qu'ils sont les plus motivés à prévenir. Ils disposent des ressources financières pour monter cette défense à grande échelle et du bilan historique montrant qu'ils les utiliseront. Ajoutez à cela une capacité d'IA spécifiquement formée dans le domaine de l'influence, et ce qui se dresse contre toute tentative de délibération collective authentique n'est pas un argument à gagner. C'est un système conçu pour garantir que ce débat n'ait jamais vraiment lieu.

Trois chances sur 4, selon Amodei, que cela ne se termine pas de façon catastrophique. Ce ne sont pas des probabilités confortables lorsque les acteurs qui travaillent sur le 1 restant sont aussi motivés, aussi bien dotés en ressources, et aussi clairs sur ce qu'ils ont à perdre.

Ce n'est pas une petite chose.

Pour la première fois, le jugement humain collectif dispose d'une architecture à la hauteur de ce qu'on lui demande de décider.

Maintenant. Pourquoi ne prêteriez-vous pas attention?

Andrew Henderson · Le Threshold · Marans, France · 2026
thresholdsummit.org · programme@thresholdsummit.org

SOURCES ET NOTES

Ce document formule des affirmations factuelles spécifiques. Les références suivantes les étayent.

Dario Amodei sur le risque catastrophique de l'IA

Amodei a affirmé dans de nombreux forums publics qu'il estime à environ 10 à 25 % la probabilité que l'IA soit catastrophique pour l'humanité. Voir: Dario Amodei, 'Machines of Loving Grace', Anthropic, octobre 2024; et interview avec Ezra Klein, The New York Times, 2024. La thèse fondatrice d'Anthropic est explicitement fondée sur cette évaluation du risque.

OpenAI: fondation et pivot commercial

OpenAI a été fondée en tant qu'organisation à but non lucratif en décembre 2015 par Sam Altman, Elon Musk et d'autres, avec un mandat explicite de sécurité. OpenAI LP, une entité commerciale à profit plafonné, a été créée en mars 2019. L'investissement de Microsoft d'un milliard de dollars a suivi en juillet 2019. Musk avait quitté le conseil d'administration en 2018. En 2024, OpenAI fonctionnait comme une société à but lucratif.

La course à l'AGI: laboratoires de pointe et valorisations boursières

Les laboratoires poursuivant explicitement l'AGI comprennent, aux États-Unis: OpenAI, Anthropic, Google DeepMind, Meta AI et xAI. En Chine: DeepSeek, Alibaba, Baidu et Tencent. Aucun laboratoire européen n'opère à cette frontière. En juin 2026, trois de ces laboratoires ont déposé des dossiers d'introduction en bourse: Anthropic avec une valorisation proche de 1 000 milliards de dollars (après avoir levé 65 milliards en mai 2026); OpenAI ciblant jusqu'à 1 000 milliards; et SpaceX, après avoir absorbé xAI, ciblant 1 750 milliards dans ce qui serait la plus grande introduction en bourse de l'histoire. Valorisation combinée anticipée: environ 4 000 milliards de dollars. Sources: TechCrunch, mai 2026; CNBC, mai 2026; TIME, 2026.

L'auto-amélioration récursive

Le concept désigne la capacité d'un système d'IA de rééditer et améliorer son propre code de base, puis améliorer la version améliorée, de façon itérative, sans instruction humaine. Également décrit dans la littérature sur la sécurité de l'IA comme une 'explosion d'intelligence' (I.J. Good, 1965). La RAND Corporation a publié une analyse des risques que cela pose pour les cadres de gouvernance existants: 'A Prisoner's Dilemma in the Race to Artificial General Intelligence', RAND, 2026.

Le précédent du contentieux tabac

Le professeur Richard Daynard, Northeastern University School of Law, est président du Tobacco Product Liability Project depuis 1984. Son travail d'établissement de l'architecture juridique des actions en responsabilité civile de masse contre les producteurs de tabac constitue le précédent principal pour l'approche probatoire et juridique adoptée par Le Threshold.

Crise du plastique: preuves du préjudice

Des microplastiques ont été détectés dans le sang du cordon ombilical fœtal, les placentas et le lait maternel humains dans des études examinées par des pairs publiées à partir de 2020. Les preuves du préjudice figurent dans le registre scientifique depuis plus de cinquante ans, à partir de la documentation de l'accumulation de plastique dans les écosystèmes marins au début des années 1970.