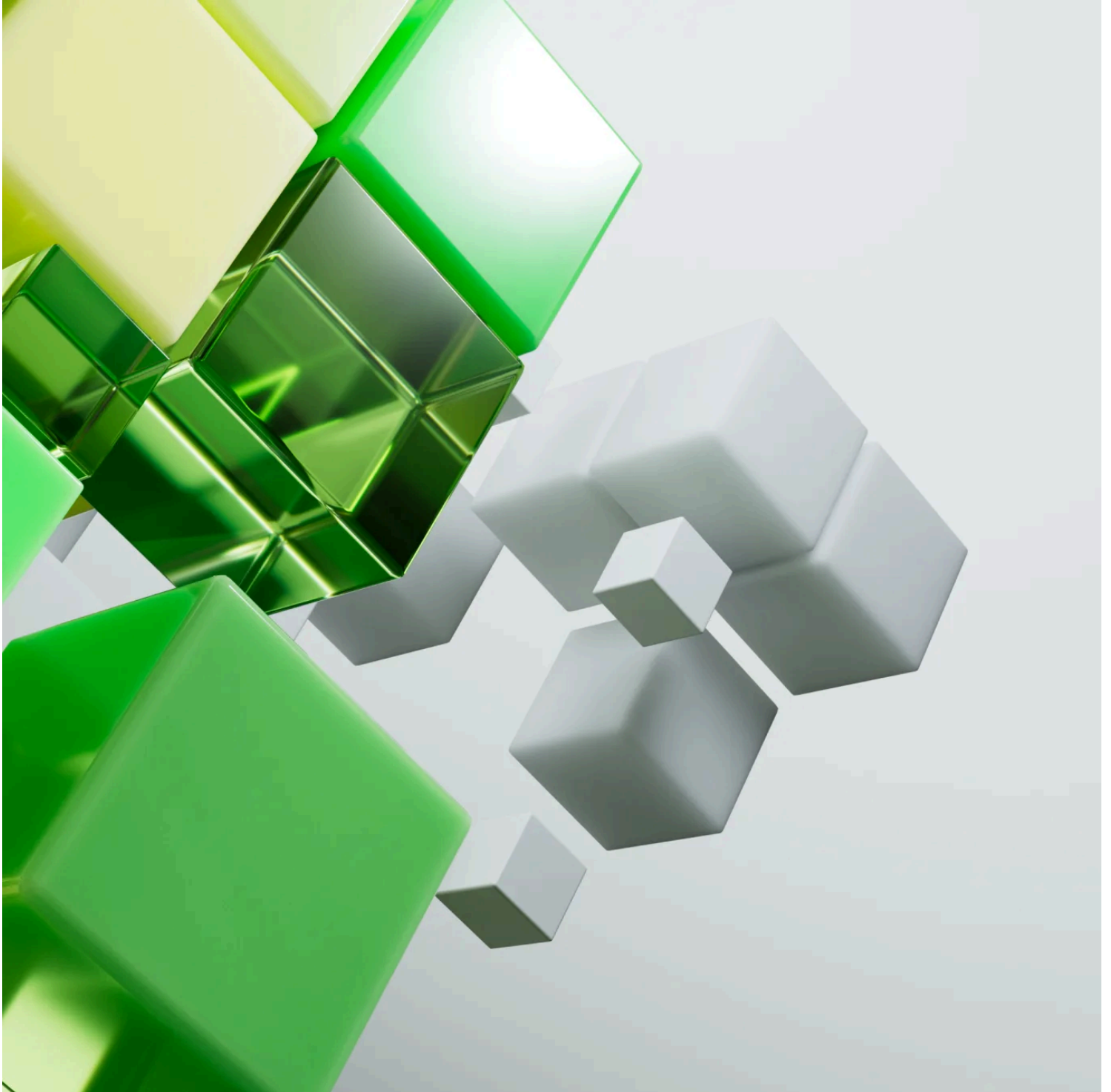


Réflexions sur l'IA et l'avenir de l'épanouissement humain

ANTHOLOGIE DE L'IA

Par Eric Horvitz
Directeur scientifique, Microsoft



Les progrès récents de l'intelligence artificielle ont suscité à la fois de l'émerveillement et de l'anxiété alors que nous contemplons son potentiel de transformation.

L'IA est très prometteuse pour enrichir nos vies, mais cette anticipation s'accompagne d'appréhensions quant aux défis et aux risques qui peuvent émerger.

Pour favoriser un avenir où l'IA sera mise à profit au profit des personnes et de la société, il est essentiel de rassembler un large éventail de voix et de perspectives.

C'est avec cet objectif en tête que je suis ravi de présenter l'Anthologie de l'IA, une compilation d'essais inspirants rédigés par d'éminents universitaires et professionnels de diverses disciplines.

L'anthologie explore les diverses façons dont l'IA peut être canalisée au profit de l'humanité tout en mettant en lumière les défis potentiels.

En rassemblant ces différents points de vue, notre objectif est de stimuler des conversations stimulantes et d'encourager les efforts de collaboration qui guideront l'IA vers un avenir qui exploite son potentiel pour l'épanouissement humain.

En rassemblant ces différents points de vue, notre objectif est de stimuler des conversations stimulantes et d'encourager les efforts de collaboration qui guideront l'IA vers un avenir qui exploite son potentiel d'épanouissement humain.

J'ai découvert GPT-4, un remarquable modèle de langage à grande échelle, à l'automne 2022, alors que j'étais président du [comité Aether de Microsoft](#).

Les équipes de direction et d'ingénierie d'Aether ont obtenu un accès anticipé à la dernière innovation d'Open AI, avec pour mission d'étudier les défis potentiels et les conséquences sociétales plus larges de son utilisation.

Nos demandes étaient ancrées dans [les principes d'IA de Microsoft](#), qui ont été établis par le comité en collaboration avec la direction de Microsoft en 2017.

Nous avons mené une analyse complète des capacités de GPT-4, en nous concentrant sur les défis possibles que les applications utilisant cette technologie pourraient poser en termes de sécurité, de précision, de confidentialité et d'équité.

GPT-4 m'a laissé bouche bée.

J'ai observé des lueurs d'intelligence inattendues au-delà de celles observées dans les systèmes d'IA précédents.

Par rapport à son prédécesseur, GPT-3.5 – un modèle utilisé par des dizaines de millions de personnes sous le nom de ChatGPT – j'ai remarqué un bond significatif en termes de capacités.

Sa capacité à interpréter mes intentions et à fournir des réponses sophistiquées à de nombreuses questions ressemblait à une « transition de phase », évoquant des images de phénomènes émergents que j'avais rencontrés en physique.

J'ai découvert que GPT-4 est un polymathe, avec une capacité remarquable à intégrer des concepts et des méthodologies traditionnellement disparates.

Il tisse de manière transparente des idées qui transcendent les frontières disciplinaires.

Les capacités remarquables de GPT-4 ont soulevé des questions sur les perturbations potentielles et les conséquences néfastes, ainsi que sur les opportunités de bénéficier aux personnes et à la société.

Alors que notre équipe élargie explorait vigoureusement les problèmes de sécurité et d'équité, je me suis

penchée sur les défis complexes de la médecine, de l'éducation et des sciences.

Il est devenu de plus en plus évident que le modèle et ses successeurs – qui présenteraient probablement d'autres sauts de capacités – ont un énorme potentiel de transformation.

Cela m'a amené à contempler les ramifications sociétales plus larges.

Des questions me sont venues à l'esprit autour de la création et de l'attribution artistiques, des acteurs malveillants, de l'emploi et de l'économie, et des futurs inconnus que nous ne pouvons pas encore envisager.

Comment les gens pourraient-ils réagir s'ils n'étaient plus la source inégalée de la pensée et de la création intellectuelles et artistiques, alors que les outils d'IA générative deviennent monnaie courante ?

Comment ces progrès affecteraient-ils notre identité personnelle et nos aspirations individuelles ?

Quelles conséquences à court et à long terme pourraient être ressenties sur le marché du travail ?

Comment les gens pourraient-ils être crédités pour leurs contributions créatives dont les systèmes d'IA tireraient des leçons ?

Comment des acteurs malveillants pourraient-ils exploiter ces pouvoirs émergents pour leur infliger des préjudices ?

Quelles sont les conséquences imprévues potentielles importantes des utilisations, y compris celles que nous ne prévoyons peut-être pas encore ?

En même temps, j'ai imaginé des avenir dans lesquels les gens et la société pourraient prospérer de manière extraordinaire en exploitant cette technologie, tout comme ils l'ont fait avec d'autres avancées révolutionnaires.

Ces influences transformatrices vont des premiers outils de la cognition – nos langages communs, permettant une coopération et une coordination sans précédent – aux instruments de la science et de l'ingénierie, à l'imprimerie, à la machine à vapeur, à l'électricité et à Internet, culminant dans les récentes avancées actuelles de l'IA.

Désireux d'étudier ces opportunités en collaboration avec d'autres dans un large éventail de disciplines, nous avons lancé le projet AI Anthology, avec le soutien d'Open AI.

Nous avons invité vingt experts à explorer les capacités de GPT-4 et à contempler les influences potentielles des futures versions sur l'humanité.

Chaque participant s'est vu accorder un accès confidentiel anticipé à GPT-4, a fourni des [études de cas dans les domaines de l'éducation, de la science et de la médecine](#), tirées de mes explorations, et a été invité à se concentrer sur deux questions fondamentales :

Comment cette technologie et ses successeurs pourraient-ils contribuer à l'épanouissement humain ?

Comment pourrions-nous, en tant que société, mieux orienter la technologie pour obtenir un maximum d'avantages pour l'humanité ?

S'appuyant sur les idées présentées lors de ma conférence Tanner à l'Université du Michigan en novembre 2022 ([Arc of Intelligence : Humanity and its Tools of Reason and Imagination](#)), ces questions soulignent l'importance de la réflexion à long terme et du maintien d'une perspective optimiste sur le potentiel de l'IA à enrichir les vies humaines. Nous pourrions en tirer d'immenses avantages potentiels.

Mais pour réaliser ce potentiel, nous devons créer des innovations techniques et des politiques pour nous protéger contre les utilisations malveillantes et les conséquences involontaires.

Cette anthologie témoigne de la promesse de la vision et de la collaboration, ainsi que de l'importance de la diversité des perspectives pour façonner l'avenir de l'IA.

Ces 20 essais offrent une mine d'idées, d'espoirs et de préoccupations, illustrant les complexités et les possibilités qui découlent de l'évolution rapide de l'IA.

En lisant ces essais, je vous encourage à rester ouvert aux nouvelles idées, à vous engager dans des discussions réfléchies et à apporter vos idées au discours en cours sur l'exploitation de la technologie de l'IA au profit de l'humanité et à l'autonomisation

L'avenir de l'IA n'est pas une voie prédéterminée, mais un voyage que nous devons parcourir ensemble avec sagesse, prévoyance et un sens profond des responsabilités. J'espère que les idées exprimées dans ces essais contribueront à notre compréhension collective des défis et des opportunités auxquels nous sommes confrontés.

Ils peuvent aider à guider nos efforts pour créer un avenir où les systèmes d'IA complètent l'intellect et la créativité humains pour promouvoir l'épanouissement humain.

Bienvenue dans l'anthologie de l'IA.

Puisse-t-il vous inspirer, vous mettre au défi et susciter des conversations significatives qui nous mèneront vers un avenir où l'humanité s'épanouira en exploitant l'IA de manière créative et précieuse.

Les points de vue, opinions et propositions exprimés dans ces essais sont ceux des auteurs et ne reflètent pas nécessairement la politique ou la position officielle de toute autre entité ou organisation, y compris Microsoft et OpenAI.

Les auteurs sont seuls responsables de l'exactitude et de l'originalité des informations et des arguments présentés dans leurs essais.

La participation à l'Anthologie de l'IA était volontaire et aucune incitation ou compensation n'a été fournie aux auteurs.

Horvitz, E. (2023, 30 mai). Réflexions sur l'IA et l'avenir de l'épanouissement humain. Dans : E. Horvitz (ed.), AI Anthology. <https://unlocked.microsoft.com/ai-anthology/eric-horvitz>

Recherche et mise en page par:

Michel Cloutier

CIVBDL

20240419

"C'est ensemble qu'on avance"