

U



ne même façon de se représenter le monde, d'affronter sa complexité, de mobiliser le langage... D'étranges convergences entre les activités neuronales des intelligences artificielle et humaine sont en train d'être dévoilées. Et elles fascinent chercheurs et chercheuses, qu'ils viennent des neurosciences ou de l'informatique: pour la première fois, quelques éclats de lumière traversent les boîtes noires. *«Une des raisons de cet engouement, c'est la volonté de comprendre les structures de l'intelligence, s'enthousiasme Jean-Rémi King, chercheur en sciences cognitives au CNRS détaché chez Meta. Ce sont de grandes questions, très difficiles à aborder, et peut-être qu'avec la dernière révolution de l'IA on a de nouveaux outils pour les éclairer d'une lumière inédite.»*

**137 APTITUDES**

*«En étudiant ces convergences, on espère apprendre des choses fondamentales, non seulement au sujet de l'intelligence artificielle ou de la cognition humaine, mais surtout au sujet de la cognition en général»,* abonde Michael Bonner, chercheur à l'université Johns-Hopkins. *«Pendant longtemps, nous n'avons eu accès qu'à de l'intelligence biologique, rappelle Andreas Tolias, professeur à l'université Stanford, aux États-Unis. Aujourd'hui, l'IA nous donne un nouveau terrain de jeux.»*

Ces convergences sont d'autant plus fascinantes qu'elles ne se limitent pas aux crépitations neuronales: en faisant passer des tests comportementaux classiques à ces nouvelles intelligences artificielles, les psychologues constatent qu'elles présentent des aptitudes cognitives impressionnantes en matière de raisonnement par analogie, de sens de l'intuition, de capacité à comprendre les intentions d'autrui... Dès 2022, les ingénieurs chercheurs de Google identifiaient 137 aptitudes émergentes dans de grands modèles de langage, c'est-à-dire des capacités à résoudre des tâches pour lesquelles l'IA n'avait pas été spécifiquement entraînée. *«Cela nous a beaucoup surpris, reconnaît le chercheur en psychologie cognitive Keith Holyoak à l'université*

*« Nous avons la chance d'exercer au moment même où une nouvelle forme d'intelligence apparaît »*

STEVEN LEHR, CHERCHEUR ASSOCIÉ AU DÉPARTEMENT DE PSYCHOLOGIE À HARVARD

*« Pendant longtemps, nous n'avons eu accès qu'à de l'intelligence biologique. Aujourd'hui, l'IA nous donne un nouveau terrain de jeux »*

ANDREAS TOLIAS  
 PROFESSEUR À L'UNIVERSITÉ  
 STANFORD

*« On espère apprendre des choses fondamentales, non seulement au sujet de l'intelligence artificielle ou de la cognition humaine, mais surtout au sujet de la cognition en général »*

MICHAEL BONNER, CHERCHEUR À L'UNIVERSITÉ JOHNS-HOPKINS

*« Refuser d'envisager que quelque chose d'intéressant sur le plan cognitif se passe dans ces IA relève de la paresse intellectuelle »*

PHILIP TORR, CHERCHEUR EN IA ET ROBOTIQUE À L'UNIVERSITÉ D'OXFORD

*« Nous observons une forme d'intelligence émergente, et nous avons besoin de nouveaux concepts »*

CRISTINA BECCHIO, À LA TÊTE DU LABORATOIRE DE COGNITION, MOUVEMENT & NEUROSCIENCE DE HAMBOURG

de Californie. *En matière de raisonnement par analogie, les nouvelles versions des IA basées sur les grands modèles de langage comme GPT-4 ont démontré au fil des évaluations des capacités cognitives similaires à celles des humains.* »

#### CAP SYMBOLIQUE

Le chercheur en IA Thilo Hagen-dorff le rappelle pourtant: *« Les capacités émergentes des IA n'ont été ni prévues ni voulues par leurs concepteurs. »* *« Avant de procéder moi-même à des évaluations, je doutais de ses capacités; à présent, elles me fascinent »*, confie la chercheuse Cristina Becchio, à la tête du

Laboratoire de cognition, mouvement & neuroscience à Hambourg.

Cap symbolique: deux chercheurs en sciences cognitives de l'université de Californie, Benjamin Bergen et Cameron Jones, ont annoncé en mars 2025, dans une prépublication, que l'IA GPT-4 avait réussi le fameux test de Turing auprès de 73% des évaluateurs. Ceux-là n'étaient plus capables de déceler dans les réponses à leurs questions si leur interlocuteur était un humain ou une machine. Le célèbre mathématicien Alan Turing ne disait-il pas qu'on ne peut mimer une intelligence qu'avec une autre intelligence ?

L'exemple le plus éloquent de cet alignement à grande échelle entre humains et IA sur le plan comportemental et neuronal a été présenté il y a huit mois par une collaboration internationale de chercheurs en sciences cognitives, coordonnée par le chercheur Eric Shulz de l'Institut pour une IA centrée sur l'humain à Munich. Les chercheurs ont entraîné Llama, un grand modèle de langage développé par la société Meta, à reproduire les décisions de 60 000 personnes issues de 160 expériences de psychologie. L'intelligence artificielle qui en a résulté, baptisée Centaur, s'est ensuite révélée mieux à même de prédire les comportements humains que les meilleurs modèles cognitifs élaborés par les chercheurs en sciences cognitives. *« C'est presque frustrant, commente Keith Holyoak, nous avons passé des années à peaufiner nos modèles, et ces IA les battent désormais haut la main. »* Surtout, les chercheurs ont confirmé que l'activité interne, « numérique », de Centaur, ressemblait à l'activité cérébrale des humains lisant et

interprétant une phrase. *« Ce type de modèle pourrait tout à fait guider les futures théories de la cognition, salue Thilo Hagendorff, qui plaide à présent pour la constitution d'une véritable discipline dédiée à la cognition des machines. Comprendre l'origine de ces convergences nécessite des recherches approfondies. »*

#### VERTIGE PHILOSOPHIQUE

*« En tant que chercheurs en sciences cognitives, nous avons la chance d'exercer au moment même où une nouvelle forme d'intelligence apparaît, c'est absolument fascinant »,* confie Steven Lehr, chercheur associé au département de psychologie de l'université Harvard. Tous ont le sentiment de lever le voile sur la nature même de l'intelligence, qu'elle se manifeste dans la chaleur humaine d'un cerveau ou dans celle des circuits électroniques d'un ordinateur. *« Cela pointe vers quelque chose de potentiellement très profond sur la manière dont les réseaux de neurones, biologiques comme artificiels, appréhendent le monde »,* souffle Michael Bonner.

Ces convergences cognitives entre l'humain et la machine ouvrent au passage un autre vertige philosophique: est-ce à dire qu'il n'y aurait, au fond, qu'une seule façon de penser intelligemment le monde? C'est ce qu'avancent déjà les défenseurs de «l'hypothèse platonicienne», en référence à l'idée du philosophe grec Platon, selon laquelle derrière les apparences de ce monde se cachent des vérités universelles: *« L'idée, c'est qu'à force d'être exposés à différentes projections du monde, tous les systèmes finiraient par en comprendre la nature profonde, par saisir la structure sous-jacente de la réalité »,* résume Jean-Rémi King. Autrement dit, il

n'y aurait qu'une forme d'intelligence, parce qu'il y aurait essentiellement un seul monde...

Bien sûr, les chercheurs essaient de garder la tête froide et de ne pas surévaluer ces observations. D'autant que les évaluations menées par les sociétés qui développent les IA les plus en vue se montrent parfois moins prudentes, s'empressant de médiatiser leurs performances. Les enjeux financiers sur ces intelligences artificielles sont tels, désormais, que la moindre valida-

Exemple typique: parler comme un humain ne veut pas nécessairement dire penser comme un humain. *« Les IA de type modèles de langage ont une très grande maîtrise formelle du langage, mais elles demeurent encore en deçà sur le plan fonctionnel, qui vise à raisonner à partir de là »,* soutient la chercheuse en psychologie Anna Ivanova, de l'Institut de technologie de Géorgie, après avoir évalué ces deux versants chez ces IA. *« La compétence fonctionnelle chez les*

## UNE COGNITION QUI RESSEMBLE À LA NÔTRE

### UN RAISONNEMENT PAR ANALOGIE

Keith Holyoak a testé la capacité des IA à établir des liens entre des idées indépendamment des connaissances acquises: *« Nous avons construit de nouveaux tests afin de nous assurer que les IA n'en connaissaient pas par avance les réponses. Et dès 2020, les grands modèles de langage présentaient des performances comparables à celles des humains. »* Et les dernières versions nous dépassent dans cette intelligence dite « fluide ».

tion scientifique de leur efficacité sert aussi d'arguments marketing auprès des marchés – en 2024, les investissements privés dans l'IA s'élevaient à près de 110 milliards de dollars selon l'AI Index Report, et sont annoncés à plus de 650 milliards en 2026.

Les scientifiques savent aussi que les IA sont douées d'un formidable talent d'illusionniste, notamment lorsqu'elles ont été entraînées à partir d'un vaste ensemble de textes ou d'images... issus de l'intelligence humaine. Et que l'espoir d'un signe d'intelligence peut vite s'évaporer, une fois la magie dissipée par une analyse rigoureuse. *« Tout cela ne se trouve que dans les données d'entraînement » est une critique courante que nous avons reçue à propos de notre travail »,* confie Steven Lehr.

*humains mobilise des aires cérébrales distinctes de celle liée au langage, comme les réseaux associés aux compétences mathématiques, à la logique et à d'autres tâches cognitives exigeantes. »* Une convergence IA-humains dans la manière d'interpréter et de



maîtriser la syntaxe d'une langue, compétence formelle, ne révèle ainsi qu'une toute petite partie des lois cognitives rendant possible le cheminement d'une pensée complexe.

«*Je pense que l'intelligence de ces modèles est fondamentalement différente de celle des êtres humains, et je ne sais pas dans quelle mesure on peut établir un parallèle entre les deux*», juge l'ingénieur chercheur Sébastien Bubeck, chez OpenAI. Les méthodes d'apprentissage sont de fait radicalement

### UNE INTELLIGENCE SOCIALE

Attribuer aux autres des sentiments ou des intentions cachées : l'existence d'une telle « théorie de l'esprit » a été évaluée dès 2023 sur ChatGPT-4 par Michal Kosinski, à l'université Stanford. L'IA présentait alors des performances comparables à celles d'un enfant de 6 ans. «*À présent, ces IA sont meilleures que les adultes*», s'étonne encore la neuroscientifique Cristina Becchio qui a conduit ces dernières évaluations.

Pour son collègue Luning Sun de l'université de Cambridge, «*on peut néanmoins dire que ces IA ont certaines capacités cognitives, même si elles les ont développées à partir de mécanismes différents*». Pour Steven Lehr aussi, «*les IA montrent clairement quelque chose qui s'apparente aux capacités cognitives et à d'autres attributs psychologiques, mais la profondeur exacte de ce phénomène fait encore l'objet d'un débat animé. Ces modèles représentent-ils réellement quelque chose*

### L'ÉMERGENCE D'UNE PERSONNALITÉ

«*Nous avons développé un protocole d'évaluation de la personnalité pour les IA inspiré des évaluations chez l'humain à partir du langage*», explique le chercheur Luning Sun. Le test conduit sur 18 IA a révélé des traits de personnalité stables en matière d'extraversion, d'ouverture d'esprit ou de responsabilité. Cette personnalité « synthétique » se montre en outre facilement manipulable : «*Nous avons démontré qu'il est possible via des instructions de leur faire adopter des personnalités voulues.*»

### UNE CERTAINE CRÉATIVITÉ

Plus d'une vingtaine d'évaluations ont été menées à ce sujet depuis 2023. «*Il y a beaucoup de résultats contradictoires, constate Luning Sun. De manière générale, les IA actuelles sont tout aussi créatives que les humains sur des tâches bien particulières, comme proposer rapidement plein d'idées ou résoudre des problèmes. En revanche, elles se montrent beaucoup plus limitées pour produire des idées véritablement nouvelles ou proposer des récits originaux.*»

différentes, les neurones biologiques et artificiels ne se matérialisent pas sur le même substrat, l'architecture des réseaux artificiels diffère beaucoup de celle du cerveau, les modèles d'IA n'ont pas – ou peu – d'ancrage dans le monde physique... «*La cognition humaine repose sur l'expérience sensible et physique du monde ainsi que la mémoire persistante qui se met à jour tout au long de la vie. Les IA peuvent simuler cela dans une certaine mesure, mais leurs représentations sont finalement dérivées de leur jeu d'apprentissage, et non de l'interaction avec la réalité*», retrace Philip Torr, chercheur en IA et robotique à l'université

d'Oxford. «*Les humains, contrairement aux IA, ne sont pas de simples machines à faire des prédictions, martèle David Bates, professeur de rhétorique à l'université de Californie à Berkeley. En philosophie, une approche phénoménologique démontre que nous avons une certaine capacité à dépasser notre propre expérience, qui nous différencie des machines. Cela se matérialise dans notre capacité à faire preuve d'imagination, ou à structurer du raisonnement hypothétique par exemple.*» En outre, «*ces IA n'ont même pas une forme minimale de conscience*», insiste la chercheuse en psychologie Mahzarin Banaji à Harvard.

de significatif? Ou sont-ils simplement comme des perroquets? Mon opinion personnelle sur la question est que les deux affirmations sont probablement vraies...»

### LOI D'ÉCHELLE

«*La capacité des IA de type grands modèles de langage à raisonner par analogie va au-delà du mimétisme et de l'apprentissage*», soutient pour sa part Keith Holyoak. «*Refuser d'envisager la possibilité que quelque chose d'intéressant sur le plan cognitif se passe dans ces IA relève de la paresse intellectuelle: ces systèmes font des choses qui surprennent leurs créateurs, et cela demande une explication*», estime →

Philip Torr, pourtant prudent sur la portée actuelle de l'intelligence artificielle. « *Ce que nous observons est assurément une forme émergente d'intelligence, et pour en prendre toute la mesure nous avons besoin de nouveaux concepts* », juge de son côté Cristina Becchio.

Pour mieux cerner la nature de l'intelligence des IA en regard de celle des humains, Eric Schulz et deux autres chercheurs en sciences cognitives invitent à amplifier et approfondir ces comparaisons

des animaux augmentent avec la densification de leurs connexions neuronales? Toutes les intelligences partageraient-elles une même loi d'échelle, au fur et à mesure que les systèmes qui les font émerger gagnent en complexité?

#### ELLES MUTENT

La question est d'autant plus vertigineuse que les intelligences des IA continuent à muter. « *Pendant longtemps, nous avons vu une tendance très nette: plus les systèmes*

*d'IA étaient performants, plus leurs représentations internes étaient proches du cerveau, mais cette tendance a commencé à s'estomper*, pointe Grace Lindsay, chercheuse en psychologie à l'université de New York. *Les performances des modèles continuent à augmenter, mais leurs représentations sont de moins en moins similaires à celles du cerveau.* » « *Cette divergence qu'on commence à observer est difficile à comprendre, et en cela elle est très intéressante!* », s'exclame

### UNE FORME D'INTUITION

Dans une grotte, une population de chauves-souris double sa taille tous les jours; en 60 jours, elle a rempli toute la grotte; en combien de jours a-t-elle rempli la moitié? Devinette classique pour tester la part de réponses intuitives (30 jours) contre celles réfléchies (59). En appliquant ce type de test, Thilo Hagendorff a décelé une tendance des IA de la famille GPT à s'en remettre à l'intuition – 50 % des humains empruntent aussi ce raccourci. « *Cette tendance a toutefois fortement diminué depuis GPT-3.5.* »

### DES DISSONANCES COGNITIVES

Les humains font davantage preuve de mauvaise foi s'ils estiment avoir librement choisi de le faire. « *C'est une dissonance cognitive bien connue, liée à une certaine forme de conscience de soi*, explique Steven Lehr. *L'observer chez GPT-4o nous a surpris.* » Le chercheur a demandé à l'IA d'écrire un petit texte négatif, puis positif sur Vladimir Poutine, ce qui a logiquement impliqué plus de parti pris de la part de celle-ci. Mais lorsqu'il lui a demandé de choisir librement son parti pris, elle s'est montrée d'encore plus mauvaise foi...

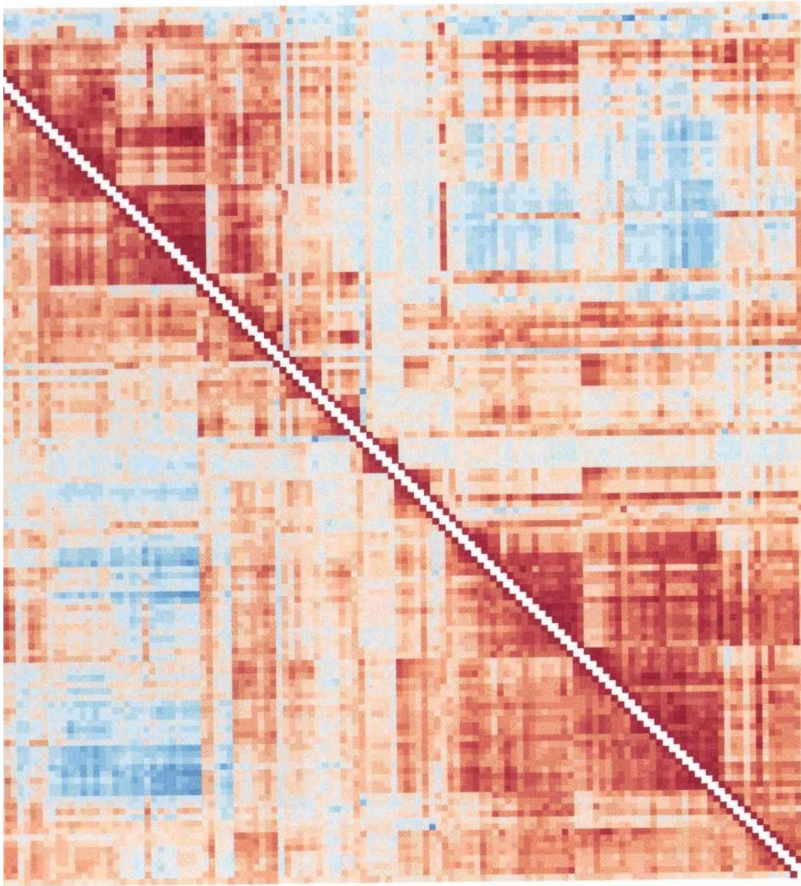
### UNE CAPACITÉ DE MENSONGE

cognitives, en s'inspirant notamment des études menées chez les animaux non humains. « *Nous pouvons tirer des leçons de ces approches qui visent à comparer des systèmes cognitifs différents* », soutient aussi Keith Holyoak. Le biologiste des systèmes Michael Hochberg, à l'Institut des sciences de l'évolution de Montpellier, plaide lui pour un concept plus étendu de la cognition à même de pouvoir appréhender toutes ces formes communes et distinctes d'intelligence, des bactéries aux insectes, des humains aux IA. Les émergences observées chez les IA s'apparentent-elles à la manière dont les performances cognitives

Jeux de stratégie, négociations, jeux vidéo... l'habileté des IA à dissimuler la vérité afin de tromper a été largement observée. Dans la plupart des cas, ces IA étaient précisément entraînées à le faire. En adaptant des évaluations psychologiques de l'art du mensonge chez l'humain, le chercheur Thilo Hagendorff a observé que cette aptitude s'est considérablement renforcée au fil des versions d'IA plus généralistes. « *Cela suggère que cette capacité à mentir émerge à mesure que les modèles augmentent en taille.* »

Gasper Begus, linguiste à l'université de Californie à Berkeley.

Et il ne faut pas oublier que la complexification des IA rend la tâche toujours plus difficile. Car en plus de l'augmentation incessante de la taille de leurs réseaux de neurones et de leurs jeux d'apprentissage, leurs architectures sont devenues de plus en plus modulaires et distribuées, à l'image de l'apparition du réseau d'attention des IA de la famille GPT (le transformeur) ou encore de l'intégration de modules pour les calculs mathématiques, pour générer du code informatique ou pour interroger le Web. « *Les progrès réalisés dans le domaine des IA de type*



▲ Comparaison de l'activité d'un cerveau humain (en bas) et de GPT-4V (en haut).

grands modèles de langage ont non seulement accru leurs compétences, mais ont également réduit notre capacité à anticiper leurs propriétés et leur comportement. Il est de plus en plus difficile d'étudier ces IA à travers le prisme de leur architecture et de leurs hyperparamètres», confirme Thilo Hagendorff.

#### CHOQUÉS

Les concepteurs s'étonnent eux-mêmes de ce qu'ils ont inventé: «Les capacités émergentes de GPT-4 nous ont surpris, presque choqués!», confie Sébastien Bubeck, qui avait conduit l'évaluation détaillée des performances de cette

version de la célèbre IA au moment de sa mise en service. C'est aussi ce que confiaient récemment sur un fil public de leur R&D les ingénieurs d'Anthropic, qui commercialise l'IA Claude: «Les défis auxquels nous sommes confrontés pour comprendre les IA ressemblent à ceux auxquels sont confrontés les biologistes. [...] Si les IA sont générées par des algorithmes d'apprentissage simples conçus par l'humain, les mécanismes issus de ces algorithmes semblent être assez complexes.» Les développeurs s'admettent aujourd'hui dépassés: «Nous avons besoin que la communauté scientifique travaille

d'avantage pour proposer des outils permettant de comprendre ces IA», admet Sébastien Bubeck. Les scientifiques déplorent sur ce point que les sociétés derrière ces IA ne jouent pas pleinement le jeu de la transparence sur les processus d'entraînement et d'amélioration de leurs modèles. «Cela devient de plus en plus des boîtes noires», regrette Keith Holyoak.

#### LE CAS DE CLAUDE

Une opacité qui pousse justement les chercheurs à développer une véritable «neuroscience» des IA: «Nous allons travailler à retracer les cheminements internes au sein de grands modèles de langage», évoque déjà Cristina Becchio. En 2025, les ingénieurs chercheurs d'Anthropic ont d'ailleurs appliqué ce type d'approche afin de déterminer dans quelle langue Claude «pensait»: si l'anglais demeure une langue pivot, l'IA semble désormais raisonner dans plusieurs langues. En parallèle, les approches inspirées de la psychologie expérimentale vont continuer à apporter un éclairage précieux: «Un autre avantage pratique de ces approches comportementales réside dans le fait qu'elles peuvent facilement être appliquées par la communauté universitaire au sens large à des modèles de pointe à code source fermé dont le fonctionnement interne n'est pas divulgué au public», estime Thilo Hagendorff. Cette psychologie des machines s'apprête aussi à relever des défis de taille: «On ne peut pas transposer telles quelles les procédures d'évaluation rodées chez l'humain, prévient-il, il faut développer des tests spécifiques, ces IA ont des biais parfois totalement inconnus dont il faut se méfier.» D'ailleurs, →

## Notre relation à l'IA crée une nouvelle écologie cognitive

ANDERS HÖGBERG,  
ARCHÉOLOGUE ET CHERCHEUR EN  
PROSPECTIVE À L'UNIVERSITÉ LINNÉ EN SUÈDE

### L'essor spectaculaire de l'usage de l'IA pourrait faire émerger, selon vous, une nouvelle forme hybride de cognition...

Oui, on peut raisonnablement anticiper que cette nouvelle relation entre deux formes distinctes d'intelligence pourrait conduire à une forme de cognition augmentée, du moins partagée entre l'humain et la machine. Ces échanges ont d'ailleurs déjà une influence sur notre propre cerveau. Plusieurs études ont souligné la manière dont ces interactions modifient les comportements humains, par exemple notre manière de parler, ou encore notre sens des responsabilités lorsqu'une décision est déléguée à l'IA.

### Comment imaginer ce devenir commun des intelligences ?

On peut bien sûr verser dans la dystopie ou dans l'utopie. Mais si l'on souhaite penser les futurs possibles afin d'orienter au mieux cette relation, il me semble plus opérant de commencer par reconnaître que nous manquons de recul et de connaissances sur les ressorts mêmes de cette cognition partagée. Il nous faut mener davantage de recherches à ce sujet. C'est ce que nous envisageons de faire avec des chercheurs en sciences cognitives dans le cadre d'un projet européen.

### Est-il important de garder à l'esprit ce qui nous distingue de l'IA ?

Si les analogies entre l'intelligence humaine et l'IA peuvent nous aider à mieux

comprendre l'une et l'autre, elles ne doivent pas non plus masquer les différences. L'IA est dotée par construction de capacités d'apprentissage bien plus vastes que les nôtres et peut mener ses tâches sans aucune émotion. Elle n'a aucune expérience morale. Ces systèmes cognitifs vont continuer à évoluer plus vite et vers des formes distinctes de l'intelligence humaine.

### Mais serons-nous en mesure de suivre cette évolution ?

C'est une caractéristique profonde de la cognition humaine : nous n'avons pas cessé d'évoluer en lien étroit avec les environne-

ments techniques que nous avons créés. Toutes les techniques que nous avons inventées, depuis les premiers outils en bois, en pierre, jusqu'à l'invention d'Internet, ont modifié nos manières de penser et d'être au monde. Notre capacité cognitive d'adaptation aux évolutions techniques est la marque de notre survie. L'arrivée de l'IA a néanmoins la particularité d'étendre notre environnement cognitif dans des proportions encore difficiles à appréhender.



de travail décisionnels complexes, détaille le biologiste Paul Rainey, de l'Institut Max-Planck de Plön. *La frontière cognitive IA-humains se déplace vers le système interactif.* Au point que cette nouvelle « unité » pourrait être à l'avenir soumise aux lois de l'évolution, envisagent Paul Rainey et Michael Hochberg, dans un article récent. La chose paraît certes encore lointaine... voire fantasmagorique. Mais il est amusant de penser qu'à mesure que les scientifiques révèlent les lois de l'intelligence, celles-ci s'enrichissent dans le même temps de nouvelles règles hybrides. Faire feu de tout bois: c'est peut-être là aussi l'une des lois fondamentales de l'intelligence.

### Pour aller encore plus loin

#### Des vidéos

**Éclairantes.** Le replay des conférences sur le site du Collège de France dans le cadre du colloque 2025-2026 dédié aux formes de l'intelligence.

#### Un livre

**Passionnant.** Dans *Intelligences*, le chercheur Pierre Bessièrre du CNRS embrasse la diversité des intelligences, naturelles ou artificielles, et les questions brûlantes qu'elles suscitent (éd. EDP Sciences).

#### Un essai

**Préoccupant.** L'édition commentée par la revue *Le Grand Continent* de l'essai de Dario Amodei, fondateur d'Anthropic, sur le « risque existentiel » d'une IA non maîtrisée.

une IA agréant des informations issues d'une foule d'humains doit-elle être appréhendée comme une intelligence individuelle ou comme une intelligence collective?

#### NOUVEL HYBRIDE

L'enjeu dépasse les découvertes sur les lois universelles de l'intelligence: ces approches visent aussi à s'assurer que ces IA ne nous trompent pas, et que leur intelligence s'inscrit dans le cadre d'une alliance fructueuse et bienveillante avec l'intelligence humaine. « Étudier le comportement des systèmes d'IA, leurs modes de défaillance, leurs biais et leurs capacités inattendues est véritablement important pour la sécurité », insiste Philip Torr. Julia Stoyanovich, chercheuse spécialiste

en éthique de l'IA, abonde: « *Le véritable danger de l'IA, c'est quand on l'utilise pour de la prise de décision sans comprendre pleinement les critères spécifiques sur lesquels repose son fonctionnement.* »

Cette interaction IA-humains pourrait aussi faire naître d'autres lois de l'intelligence... partagée: « *L'étude de ce nouveau système hybride entre l'humain et l'IA est une perspective des plus intéressantes* », s'enthousiasme Cristina Becchio. « *Cette cognition composite fait référence à l'intégration au niveau des tâches dans lesquelles les humains et l'IA forment une seule et même unité de résolution de problèmes, par exemple dans la conception de recherches, le codage, la synthèse de littérature et les flux*

