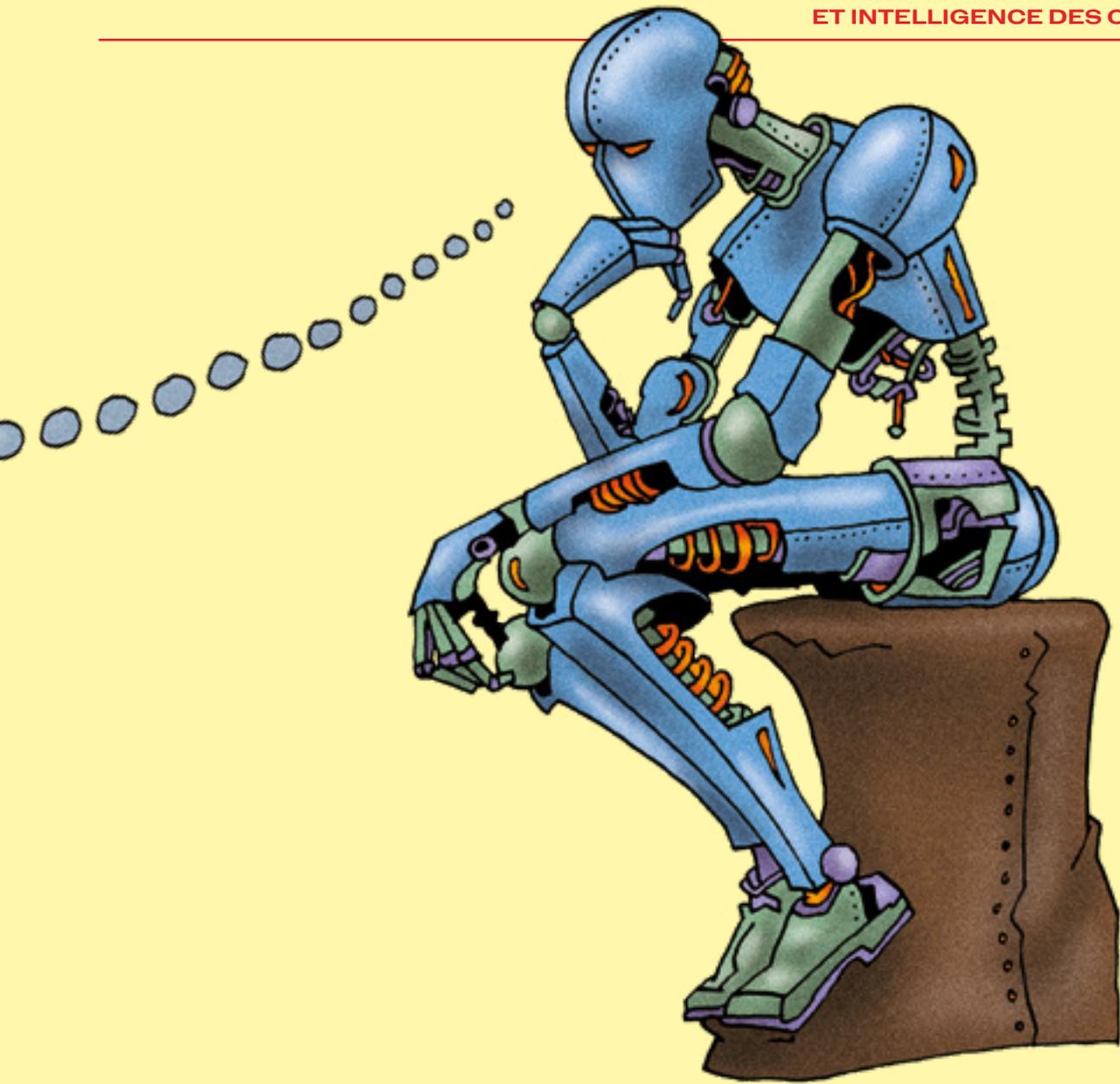




Intelligence artificielle

Par Bernard Reber

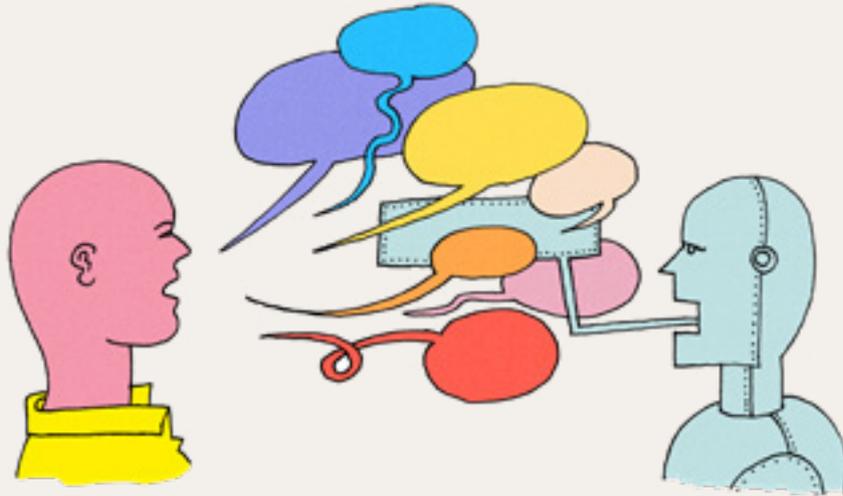
L'intelligence artificielle peut-elle se substituer à celle des humains dans leurs activités cognitives les plus complexes ? Pour tester cette hypothèse, Bernard Reber revient sur une expérience à laquelle il a participé, la Convention citoyenne sur la fin de vie. Sur un sujet aussi sensible, il s'interroge sur la capacité de l'IA à restituer de manière synthétique et fidèle le foisonnement des débats et la diversité des opinions exprimées, ou même tout simplement à prodiguer des conseils avisés pour l'organisation d'une délibération.



et intelligence des choix



Bernard Reber, philosophe moral et politique, est directeur de recherche CNRS au Centre de recherches politiques (CEVIPOF). Il s'intéresse aux questions de délibération, de pluralisme, de responsabilité, d'évaluation, de gouvernance, d'argumentation, de design institutionnel, d'éthique de l'environnement et de technologies controversées. Il a été membre du collège des garants de la Convention citoyenne sur la fin de vie.



La délibération est une activité cognitive complexe, individuelle et collective, qui dépasse de loin tous les calculs et les simulations dont sont capables les outils produits par les chercheurs et les ingénieurs de l'intelligence artificielle. Je dirai IA par simplicité. Ces outils, aussi variés que leurs domaines d'application, peuvent permettre de « reproduire des comportements liés aux humains, tels que le raisonnement, la planification et la créativité », pour reprendre la définition qu'en propose le Parlement européen. Cette définition, selon laquelle ils seraient capables de simuler, synthétiser et de faire croire à quelque chose comme de l'intelligence, leur accorde beaucoup. Pourtant, il leur faudra encore du temps avant de pouvoir un jour se hisser au niveau des délibérations humaines et, notamment, de celles qui concernent l'éthique et la politique, embarquant des croyances, des émotions et des responsabilités à endosser.

Cela ne doit pas nous détourner de ce que ces technologies peuvent déjà offrir, ni surtout de ce qu'il est désirable qu'elles « apprennent ». Un exercice comme la délibération démocratique peut en effet tirer profit d'outils issus de la recherche en IA, à condition de conserver l'intelligence des choix. Il importe par exemple de savoir dans quelle phase d'une délibération s'en servir, pour quoi faire et ce que l'on peut promettre. Je considérerai ici deux options, l'une réelle et l'autre imaginaire. La première est une expérience destinée à améliorer la restitution au grand public de résultats partiels de la Convention citoyenne sur la fin de vie (CCFV) à laquelle j'ai participé. La seconde se rapporte à une conférence que j'ai donnée sur l'éthique de l'IA, pour laquelle j'ai questionné l'agent conversationnel utilisant de l'IA générative, afin qu'il me prodigue des conseils sur

l'organisation d'une convention citoyenne. J'ajoute cet intérêt supplémentaire : toute technologie oblige à un travail d'explicitation pour savoir ce que l'on cherche à faire grâce à elle. C'est vrai également lorsqu'il s'agit de lui faire faire ce que font nos cerveaux, d'autant que ces derniers le font différemment et souvent bien mieux qu'elle. En effet nos cerveaux restent une terre très largement inconnue et une source constante d'éblouissement pour les chercheurs qui en sont spécialistes. Les simulations technologiques passent donc par des détours. Le nom grec de la technique est d'ailleurs la ruse (*mètis*).

Où mettre l'IA et que promettre ?

Prenons le cas de la CCFV, organisée par le Conseil économique, social et environnemental (CESE) de décembre 2022 à avril 2023, pour laquelle j'ai été l'un des quatre garants chargés d'en assurer le respect des principes. Parce qu'elle a été largement constituée d'échanges et de conversations entre les 184 citoyens tirés au sort, les experts et les témoins auditionnés, la convention a produit une masse de données langagières susceptibles de nourrir et d'entraîner une IA. Un traitement par l'IA aurait pu aussi s'appliquer aux échanges du comité de gouvernance décidant des procédures qu'il estimait les plus favorables à une délibération collective, et encadrant les débats au sein de la convention, que des facilitateurs mettaient en œuvre.

Ces phases d'organisation, de débats et de mise en œuvre auraient-elles pu tirer bénéfice d'une IA et de quelle manière ? En tout cas, l'éventualité de recourir à cet outil s'est présentée à deux autres occasions durant la convention.



Séance de clôture
de la Convention citoyenne
sur la fin de vie, hémicycle
du Conseil économique social
et environnemental, Paris,
2 avril 2023.

Tout d'abord, lorsque s'est posée la question de la forme à donner à la version finale du manifeste de son sous-groupe, l'un des participants a suggéré d'interroger ChatGPT. Vendeur, il avait déjà eu recours à cette application pour réaliser des annonces publicitaires. « Pas de temps à perdre avec ça », ont répondu ses concitoyens en éclatant de rire, estimant que cette proposition n'était pas appropriée pour un travail aussi délicat et sérieux. En tant que garant, je peux attester de la très grande attention qui a pesé sur la rédaction collective de ce texte, où la quête du mot juste est parfois passée par d'âpres négociations. Ce type de rédaction est éminemment politique, dans le sens le plus noble du terme.

L'occasion de tester l'IA s'est à nouveau présentée, fin avril 2024, lors du week-end de reddition de comptes de la convention, durant lequel le président de la République et plusieurs ministres ont expliqué ce qu'ils avaient fait des propositions des conventionnels. Durant ce week-end, le directeur d'une entreprise est venu présenter une expérience réalisée par son équipe, avec l'aide des personnels du CESE, et qu'il a qualifiée de « première mondiale ». L'expérience, menée durant les mois qui ont suivi la convention, était destinée à montrer qu'un traitement par l'IA pouvait améliorer la restitution des débats au grand public. Elle a porté sur un corpus limité correspondant à deux séquences de la convention, dont il a été demandé à l'IA de produire une synthèse.



Auditionnés lors de la Convention citoyenne sur la fin de vie de 2022, de gauche à droite : Haim Korsia, grand-rabbin de France, Antony Boussemart, co-président de l'Union bouddhiste de France, Mgr Marc Alric, évêque de Neamt, représentant de l'Assemblée des évêques orthodoxes de France, Christian Krieger, président de la Fédération protestante de France.

« Désormais, grâce à l'intelligence artificielle, nous allons pouvoir rendre pleinement accessibles les travaux de la Convention citoyenne sur la fin de vie au grand public et aux parlementaires qui vont travailler sur le texte », affirma ce chef d'entreprise lors de sa présentation. « Nous sommes convaincus que cette révolution changera notre rapport aux institutions pour les années à venir. » Ce promoteur de l'IA et d'une large participation des citoyens au débat public alla même jusqu'à soutenir que la mise à la disposition de tous des documents issus de la convention (textes et plus de trente heures de vidéo en diffusion continue), pas plus que les nombreux articles de presse, ne pouvait aboutir

à une véritable transparence, car personne ne disposerait du temps nécessaire pour s'immerger dans le verbatim de cette aventure de vingt-sept jours.

Moi-même garant, je devais notamment, avec mes trois collègues, veiller au respect d'une idée bien comprise de la transparence. En effet, la transparence totale est un univers concentrationnaire non désirable. Elle peut même avoir des effets sur la délibération, déjà bien documentés dans les sciences de la délibération. On ne délibère pas de la même façon en public ou derrière les portes d'une commission ou d'un groupe de travail. Quelque peu étonné que les

conventionnels n'aient pas été consultés sur ce projet, comme ils l'avaient été pour ceux des observateurs-chercheurs qui avaient dû présenter leurs intentions d'étude devant les conventionnels, j'ai été interrogatif. J'ai appris que, lors de cette expérience, l'IA avait mis beaucoup de temps à intégrer une simple partie du corpus proposé. Par exemple, elle avait été incapable de lire des tableaux de votes, lesquels, bien qu'abondants (plus de 500) dans le corpus en question, constituaient de la matière organisée et facile à traiter. Ensuite, il est apparu que les matériaux extraits des deux séquences choisies étaient les plus faciles à analyser. Il s'agissait d'une part du rapport final, très bien rédigé, organisé et lisible par tout citoyen curieux et un tant soit peu vertueux, d'autre part de l'intégralité des débats en séances plénières, modérées par de talentueux facilitateurs. Il aurait été plus intéressant de voir l'IA relever le défi d'analyser les conversations bien plus longues menées en sous-groupes, plus désordonnées et dans lesquelles, comme dans toute conversation, les participants perdent le fil et oublient ce qui a été dit. C'est ce type d'aide qu'avait à l'esprit l'un des conventionnels, par ailleurs très au fait des langages d'IA, qui bombardait de questions l'auteur de la « première mondiale ». S'il avait fallu au moins deux mois à l'IA ex post pour traiter ces deux corpus restreints, combien de temps supplémentaire aurait-il été nécessaire pour analyser les conversations jalonnant le processus, pour produire des synthèses en simultané ? Et quelle en aurait été la qualité ?

La transparence version IA : « Peut mieux faire »

La dimension qualitative du travail effectué par l'IA est devenue centrale lorsque ce dirigeant a rapporté un exemple auquel il tenait beaucoup. La question suivante avait été posée lors d'une session : « Comment la convention a-t-elle pris en compte les arguments religieux ? » Utilisant des extraits des interventions de différents représentants religieux auditionnés en plénière, l'IA a répondu en deux lignes par confession et par interlocuteur, sans prendre la peine de spécifier leur fonction. Un franc-maçon figurait parmi eux, mais il s'agissait d'une erreur de personne, l'IA ayant confondu ce franc-maçon (qui a bel et bien été auditionné lors d'une session ultérieure) avec le représentant orthodoxe, Monseigneur Alric, que l'on voit pourtant sur la

Toute technologie oblige à un travail d'explicitation pour savoir ce que l'on cherche à faire grâce à elle.

vidéo. Les utilisateurs de ChatGPT sont habitués à ces erreurs factuelles. Cependant, ici, au vu des enjeux et du petit nombre d'acteurs, une telle légèreté a de quoi inquiéter. Les sources renvoyaient, quant à elles, aux extraits d'audition de seulement quatre responsables. Pourquoi quatre alors qu'ils étaient cinq ? L'un des extraits contenait en outre des propos émis par le facilitateur. Inutile de mentionner que le matériau était la retranscription de la vidéo et non l'image. Si l'IA avait pris en compte l'image, elle n'aurait pas remplacé un évêque orthodoxe par un franc-maçon.

Autre exemple, pour savoir quelle était la position de l'Église catholique sur la fin de vie, l'IA s'est contentée de deux affirmations formulées lors d'une audition de l'archevêque de Paris, Monseigneur Ulrich : 1) « La position de l'Église catholique est fondée sur la croyance en une option de vie qui ne doit pas être interrompue par l'homme » ; 2) « L'accompagnement des mourants est essentiel, mais la décision de mettre fin à une vie n'est pas acceptable. » Le lecteur pressé restera sur sa faim, avec des affirmations en lieu et place d'arguments. Ou alors le mot argument n'a plus de sens. Et s'il veut se plonger, par désir de transparence, dans l'intégralité des verbatims, il sera encore plus décontenancé. En effet, il y trouvera les arguments apportés par l'archevêque, mais non repris par l'IA, à savoir que la vie est un don de Dieu et que Dieu ne la retire pas au moment de la mort. Puis il lira que le prélat a évoqué l'interdiction de tuer et fourni des justifications à cet interdit. On aura compris que l'IA est passée à côté de l'essentiel du discours de l'intéressé et qu'elle a perdu ce qui lui servait d'argumentation. Ce qui est vrai pour lui l'est pour d'autres intervenants.



Plutôt que de parler de révolution pour la démocratie et de nouvelles transparences, nous devrions rester modestes.

Non seulement l'IA utilisée n'a pas répondu à la question initiale de savoir comment la convention avait pris en compte les arguments religieux à propos de la fin de vie, non seulement elle a confondu les interlocuteurs, mais ses résultats n'ont rien dit non plus des échanges bien plus intéressants qui s'étaient déroulés entre certains conventionnels et les responsables religieux auditionnés pendant plus d'une heure et demie. Pour y répondre réellement, il aurait fallu commencer par prendre en compte la manière dont le rapport en parle (tout en sachant que le mot de religion n'y figure que cinq fois) et revenir à l'endroit où sont mentionnés les responsables auditionnés. Surtout, il aurait fallu se livrer au travail, beaucoup plus fin, d'analyse des moments où certains conventionnels ont fait référence à leurs convictions religieuses. Et c'est un excellent sujet, plus à la portée des chercheurs-observateurs prenant des notes tout au long du processus qu'à celle d'un algorithme.

Plutôt que de parler de révolution pour la démocratie et de nouvelles transparences, nous devrions rester modestes. L'expérience menée sur ce fragment de corpus a ajouté des écrans plutôt que de la transparence et s'est payée d'un temps paradoxalement très long. Son initiateur a vanté le temps gagné pour un public extérieur, sans rien dire des modes de travail fastidieux et des choix (techniques ou interprétatifs) qu'il a fallu faire pour arriver à un résultat aussi maigre, partiellement erroné et, au total, trompeur. Si l'argumentation est un des éléments essentiels de la délibération, il faut la repérer et la donner à voir dans sa complétude, afin de pouvoir la reconnaître, de savoir si elle est valide ou pas, et non pas se contenter de pêcher quelques affirmations. L'IA mobilisée au service de cette forme de transparence mérite donc la mention « peut mieux faire ».

Une convention ChatGPT

Invité dans d'autres circonstances à donner une conférence sur l'éthique de l'IA, j'ai choisi de mettre ChatGPT à l'épreuve, en lui demandant conseil pour l'organisation d'une convention citoyenne. Il est impossible de restituer ici l'intégralité de la conversation édifiante qui s'est prolongée sur douze pages avec cette application. Censée m'apporter des idées, elle n'a fait que lister les éléments dont je devais me préoccuper, bien connus de tout organisateur ou chercheur dans le domaine, à savoir : identifier les participants (et là, il pensait aux experts et aux publics concernés, non à des citoyens tirés au sort) ; choisir un lieu, établir un agenda, informer les participants et les engager à des discussions respectueuses, prendre note des résultats et tenir les participants au courant de l'utilisation de leurs résultats.

Dans la suite de la conversation, l'IA a réinventé, au fil de mes questions, des problèmes connus en philosophie, comme les difficultés du jugement. Je n'y ai pas retrouvé les célèbres formulations du philosophe étatsunien John Rawls sur ce sujet, pas plus que ses points de désaccord avec son homologue allemand Jürgen Habermas. En revanche, l'IA a proposé des méthodes inattendues pour traiter les grandes difficultés du pluralisme moral qui est au cœur de telles expériences. En l'occurrence, elle a surpassé la pratique habituelle, plus portée sur le respect et l'inclusion des participants. Quant à sa définition de la délibération, elle s'est limitée à une « compréhension partagée du sujet, considérant plusieurs perspectives pour prendre des décisions basées sur une sagesse collective ». Cette approche irénique fait peu de cas des désaccords et des incertitudes politiques.

Les ailes du désir

Ayant à l'esprit l'enjeu d'une convention portant sur la vie et la mort, je lui posai cette question : « Mais vous, en tant qu'agent IA, avez-vous des intuitions morales ? » Voici sa réponse : « En tant que modèle linguistique de l'IA, je n'ai pas de croyances personnelles, d'émotions ou d'intuitions morales. Je suis programmé pour fournir des informations et répondre à des questions [...] et je n'ai pas la capacité [...] de me forger des opinions personnelles. [...] Je m'efforce de fournir des informations précises et pertinentes [...] Cependant, il est important de garder à l'esprit que je ne suis pas capable d'avoir des expériences [...] personnelles, et que mes réponses sont limitées par les informations et les connaissances contenues dans les données textuelles sur lesquelles j'ai été formé. Bien que je n'aie pas d'émotions personnelles ou d'intuitions morales, je peux fournir des informations et des conseils neutres et impartiaux sur ces aspects et d'autres aspects du processus de délibération, sur la base [...] des principes de la démocratie délibérative. Mon objectif est de soutenir les participants dans leurs efforts pour s'engager dans une délibération équitable, inclusive et productive, et d'aider à faciliter le processus de prise de décision collective. » Comme l'un des deux anges mis en scène par Wim Wenders dans le film *Les Ailes du désir*, l'IA pourrait envier sur ce point les humains aptes aux désirs, aux émotions et aux intuitions, et renoncer à son immortalité pour goûter au plaisir des sens. Et, pourrait-on ajouter, à la sagesse des choix et à la justesse des sentiments, en un clin d'œil au beau titre de l'ouvrage du philosophe moral Alan Gibbard, *Wise Choices, Apt Feelings*.

Notre sagesse doit donc s'exercer à l'endroit de la pertinence et de l'ajustement des prothèses que sont – comme toute technique – les outils de l'IA, en particulier dans le cadre des délibérations politiques dont la dimension morale est fondamentale. Il ne s'agit pas de prétendre gagner du temps ni de faire des promesses de transparence, mais de prendre au contraire le temps de poser les bonnes questions. La première est de savoir ce qu'est une délibération. Ainsi, l'intelligence artificielle reposera toujours sur l'intelligence des choix que nous faisons et sur celle des questions que nous lui posons.

Notre sagesse doit s'exercer à l'endroit de la pertinence et de l'ajustement des prothèses que sont les outils de l'IA.

■ RÉFÉRENCES

→ Reber B.,
La Délibération responsable entre conversation et considération. Conditions d'un grand débat démocratique, Londres, ISTE Éditions, 2023