

LesEchos logo and navigation icons.

YANN LE CUN ET YOSHUA BENGIO : LES DOUBTES ET LES CONVICTIIONS DES « PERES » DE L'IA MODERNE.

Ils ont révolutionné ensemble l'intelligence artificielle au cours des deux dernières décennies. Pourtant, Yoshua Bengio et Yann Le Cun portent un regard très différent sur l'avenir de l'IA, les dangers qu'elle peut engendrer ou les besoins de la réguler.



Yoshua Bengio, 59 ans, est professeur à l'université de Montréal et fondateur du Mila, l'Institut québécois d'intelligence artificielle. Né à Montréal, il a obtenu son doctorat en 1997 et a obtenu son doctorat à l'université McGill avant de poursuivre ses études au MIT et aux AT&T Bell Laboratories aux côtés de Yann Le Cun. Avec Geoffrey Hinton, ils ont posé les bases de l'apprentissage automatique profond (deep learning). Son frère, Simon Bengio, également chercheur reconnu en IA, a travaillé pour Google pendant quatorze ans avant de rejoindre Apple en 2021.

On peut avoir créé ensemble les bases de l'intelligence artificielle (IA) moderne et porter un regard diamétralement opposé sur ce qu'elle représente pour l'avenir de l'humanité.

Yoshua Bengio, professeur à l'université de Montréal, figurait en tête des signataires de la lettre ouverte appelée à une pause dans l'entraînement des systèmes d'IA, publiée en mars dernier par le Future of Life Institute, et plus récemment d'une pétition affirmant que « réduire le risque d'extinction de l'IA devrait être une priorité mondiale au même titre que les pandémies et les guerres nucléaires ».

Mais les deux hommes se connaissent très bien, et pour cause : avec Geoffrey Hinton, professeur émérite à l'université de Toronto, ils ont révolutionné l'IA par leurs travaux sur les réseaux de neurones et l'apprentissage automatique profond (deep learning). Ce qui a valu au trio d'être qualifié de « pères fondateurs de l'IA » et de recevoir en 2019 le Prix Turing, équivalent du Nobel pour l'informatique.

Malgré leurs divergences, Le Cun et Bengio s'apprécient et continuent de discuter. À l'occasion de VivaTech, ils ont accepté de débattre en exclusivité pour le deuxième mode de questions vidéoconférence enregistré entre Paris et Montréal.

1. LE PHÉNOMÈNE CHATGPT

Pour un être que ChatGPT a révolutionné l'intelligence artificielle ?

Yoshua Bengio. D'un point de vue scientifique, non. Ce sont des recettes que l'on connaît depuis plusieurs années qui ont été mises ensemble, même si il y a certainement eu un effet d'Opénel en matière d'ingénierie, de quantité de données, de taille des modèles... Si ChatGPT modèle de langage génératif par l'IA ou contenu écrit en réponse à une demande, NDLR a pris beaucoup de temps par surprise, c'est par son niveau de complexité, de maîtrise de la langue, et par les connaissances qui ont été acquises « pas forcément de manière correcte d'ailleurs. Mais se pose aussi que c'est une révolution sociale : si nous avons cette discussion aujourd'hui, c'est parce que beaucoup de gens l'ont engagé, parce que cela peut les inquiéter... ou leur donner envie d'inventer des milliards.

Yann Le Cun. Ça n'est pas simplement d'accord, ce n'est pas un évènement scientifique.



Yann Le Cun, 62 ans, est vice-président et « Chief AI Scientist » de Meta et professeur à l'université de New York (NYU), après des études d'ingénieur à l'ENSEA Paris et un doctorat à l'Université de Montréal. Il a rejoint l'équipe de Geoffrey Hinton à l'université de Toronto avant d'intégrer les AT&T Bell Laboratories en 1988 dans le New Jersey. Spécialiste de l'apprentissage automatique et de la vision artificielle, il est employé à NYU depuis 2012. En 2012, il a été recruté par Mark Zuckerberg pour créer Facebook AI Research, et a dirigé jusqu'en 2018.

Les techniques de bases utilisées par OpenAI ont été inventées il y a plusieurs années par des chercheurs de notre communauté. Certaines ont été popularisées par Yoshua et son groupe et DeepMind est aussi, directement ou indirectement, à la base de tout ça, tout comme le deep learning auquel nous avons participé tous les deux, ou l'apprentissage automatique.

Yann Le Cun, 62 ans, est vice-président et « Chief AI Scientist » de Meta et professeur à l'université de New York (NYU), après des études d'ingénieur à l'ENSEA Paris et un doctorat à l'Université de Montréal. Il a rejoint l'équipe de Geoffrey Hinton à l'université de Toronto avant d'intégrer les AT&T Bell Laboratories en 1988 dans le New Jersey. Spécialiste de l'apprentissage automatique et de la vision artificielle, il est employé à NYU depuis 2012. En 2012, il a été recruté par Mark Zuckerberg pour créer Facebook AI Research, et a dirigé jusqu'en 2018.

De l'intelligence artificielle, il y a deux modes de questions vidéoconférence enregistré entre Paris et Montréal.

Y. L. C. Il y a un phénomène intéressant qui est que, quand un système arrive à l'attention du public, celui-ci a l'impression d'être devant un sujet qualitatif. Mais d'un point de vue scientifique, c'est souvent plutôt une continuation. Les techniques utilisées par IBM pour les échecs avec DeepBlue ou pour le jeu Jeopardy! avec Watson étaient pas particulièrement innovantes non plus, mais elles ont réussi grâce à des investissements massifs. Il y a toujours plus d'ingénierie que de découvertes scientifiques et c'est un peu la même chose avec ChatGPT.

Y. B. L'apporterais une nuance, qui concerne la quantité de données. Si ces systèmes impressionnent autant, c'est parce que jamais auparavant on n'avait entraîné des réseaux de neurones aussi gros avec autant de données. La quantité de textes utilisés dépasse de loin ce qu'un humain pourrait lire dans sa vie. Y. L. C. D'après mes calculs, il faudrait environ 22 000 ans à un humain, en lisant huit heures par jour, pour absorber tous les contenus sur lesquels est entraîné ChatGPT.

Donc pour les experts de l'IA, ChatGPT n'est pas une surprise ?

Y. L. C. Ce qui a surpris, c'est le fait que ces systèmes fonctionnent si bien ! Il y a une espèce d'émergence de « complexité » qui semble se produire quand les modèles et la quantité de données dépassent une certaine taille. Cela a surpris la communauté de l'IA, mais cette surprise est arrivée il y a environ deux ans, donc avant ChatGPT.

Y. B. C'est vrai qu'on a vu ça venir, mais avec GPT-2 ou GPT-3, on pouvait encore croire que ce n'était pas au point, car ça disait plein de bêtises. Y. L. C. Mais ça dit toujours plein de bêtises ! Y. B. Beaucoup moins et c'est la différence.



Yann Le Cun, 62 ans, est vice-président et « Chief AI Scientist » de Meta et professeur à l'université de New York (NYU), après des études d'ingénieur à l'ENSEA Paris et un doctorat à l'Université de Montréal. Il a rejoint l'équipe de Geoffrey Hinton à l'université de Toronto avant d'intégrer les AT&T Bell Laboratories en 1988 dans le New Jersey. Spécialiste de l'apprentissage automatique et de la vision artificielle, il est employé à NYU depuis 2012. En 2012, il a été recruté par Mark Zuckerberg pour créer Facebook AI Research, et a dirigé jusqu'en 2018.

Aujourd'hui, le niveau de bêtises – ou d'intelligence, c'est selon – commence à être comparable à celui d'un humain. D'accord, à un humain qui ne réfléchit pas trop et répond de façon intuitive, sans avoir nécessairement une compréhension complète du monde.

Y. L. C. C'est peut-être là où l'on commence à ne pas être d'accord.

2. LE NIVEAU D'INTELLIGENCE

Ces systèmes nous rapprochent-ils d'une intelligence artificielle générale, ou d'un niveau d'intelligence proche des humains ?

Y. L. C. Non, absolument pas. Il y a encore une différence énorme entre l'intelligence que l'on observe chez les humains et chez les animaux, et le type d'intelligence que ces systèmes sont capables de montrer. Ils sont très volatils, ils peuvent manipuler la langue, et cela donne l'illusion qu'ils sont intelligents parce que l'on associe beaucoup d'intelligence à la capacité de manipuler le langage. En fait, ce n'est qu'une impression, leur compréhension du monde est encore très superficielle.

Y. B. Mais c'est le cas de beaucoup d'autres humains !

Y. L. C. Certes, mais on n'a pas encore de systèmes qui soient capables d'apprendre comment fonctionne le monde. À la manière des animaux et des bébés humains. Il y a une différence fondamentale : le texte est relativement facile, parce que l'on sait prédire une distribution de probabilités pour prévoir le mot suivant dans une phrase.

Y. B. Je pense qu'il manque quelque chose de fondamental pour arriver à une intelligence artificielle générale : les fonctions cognitives conscientes qui viennent avec la capacité de raisonner, de comprendre les relations de cause à effet, d'imaginer comment le monde pourrait être différent, de planifier... Toutes ces choses-là ne sont pas encore au point, mais il y a pas mal de recherches qui vont dans cette direction.

Avec vous une idée du temps que cela va prendre ?

Y. B. Si vous n'avez posé la question il y a deux ans, j'aurais répondu que cela prendrait sûrement plusieurs décennies. Cela reste une question difficile, mais, aujourd'hui, je me dis que la bonne réponse est peut-être de quelques années. Je vois ce qui se développe, les recherches en cours... Peut-être qu'il y a un obstacle qui va surgir, et je le souhaite, parce que je ne voudrais pas que l'on soit en face d'une telle intelligence dans quatre ou cinq ans.

Mais il se peut aussi que l'on en soit proche. Y. L. C. Je vois deux premiers obstacles à dépasser pour arriver : d'abord, permettre aux machines d'apprendre comment fonctionne le monde, par observation de vidéos et par les sens. Mais si les plateformes open source sont interdites par la réglementation à cause de craintes de sécurité, nous aurons encore des années de retard avant d'être capables de manipuler le langage. En fait, ce n'est qu'une impression, leur compréhension du monde est encore très superficielle.

Y. L. C. C'est comme toute technologie, cela peut toujours être utilisé. Mais si on prend l'exemple d'une organisation terroriste, ou d'un pays à l'écart de la communauté internationale, leur technologie sera toujours inefficace à la technologie dominante des pays développés. Donc la contre-mesure à ces systèmes d'IA dangereux, ce sera d'avoir des systèmes d'IA pas dangereux.

Y. B. LLA, tu es très optimiste...

3. LES RÉACTIONS DE PEUR

Cette accélération entraîne des craintes, et compris dans la communauté de l'IA... Y. L. C. Je n'ai pas les peurs de Yoshua à ce sujet... Je pense que les machines futures ne seront pas construites sur le modèle de ChatGPT et des autres systèmes actuels, parce qu'il en est un problème fondamental : on ne peut pas caractériser les contrôleurs pour qu'ils ne soient pas toxiques, pour qu'ils soient factuels, etc. Ce n'est pas simplement en les rendant plus grands ou en les entraînant davantage qu'on va résoudre le problème. Il faut des architectures nouvelles dans lesquelles on pourra introduire des critères de sécurité, qui par construction les rendent contrôlables et non dangereux.

Y. B. Mais comment évaluer que ces centres seront toujours alignés avec les intérêts de l'humanité ? On peut imaginer qu'une organisation militaire ou un groupe mal intentionné décide d'utiliser des critères dangereux.

Y. L. C. C'est comme toute technologie, cela peut toujours être utilisé. Mais si on prend l'exemple d'une organisation terroriste, ou d'un pays à l'écart de la communauté internationale, leur technologie sera toujours inefficace à la technologie dominante des pays développés. Donc la contre-mesure à ces systèmes d'IA dangereux, ce sera d'avoir des systèmes d'IA pas dangereux.

4. LA MENACE EXISTENTIELLE

Cela nous amène aux risques d'extinction de l'humanité sur lesquels des chercheurs en IA, dont Yoshua Bengio, ont alerté à plusieurs reprises via des lettres ouvertes ou des pétitions...

Y. B. Même si on a plein de gens qui mettent les bonnes balises, même les bons objectifs et contraignent des systèmes d'IA sûrs, il sera aussi possible pour des personnes avant de meilleures intentions de créer des choses aussi dangereuses que des armes nucléaires. Ce n'est pas rassurant... Pour mes enfants et mes petits-enfants, j'aimerais que l'on mette en place des garde-fous pour minimiser les risques et se balader avec avec d'argent, avec une organisation mal intentionnée, développe des IA en fixant des objectifs qui seraient dangereux.

Y. L. C. Il n'y a pas de doute et c'est pourquoi je n'ai pas signé les lettres en question. Je pense que l'accomplissement des machines futures est un processus total, parce que l'intelligence artificielle est là pour amplifier l'intelligence humaine. Et ça, c'est quelque chose de fondamentalement positif, qui permettrait en principe un nouveau siècle des Lumières, avec une accélération des progrès scientifiques, de la médecine ou de la production. Donc, c'est un bien immense.

Y. B. Mais Yann, il y a aussi des gens qui font des choses terribles, il y a des génocides... Y. L. C. Bien sûr, notre technologie peut être utilisée à des fins négatives, mais c'est une question de rapport de force entre le bien et le mal. Si une technologie est assez puissante pour être utilisée par des gens mal intentionnés, il y a probablement une contre-mesure...

Y. B. Mais si on se projette cinq ou dix ans dans le futur, quand nous aurons des systèmes plus intelligents que nous, ils pourraient influencer des êtres humains, par exemple à travers des réseaux sociaux, pour agir de façon subtil.

Y. L. C. C'est-à-dire que cela pourrait être déployé sans que ce soit intercepté par les polices ou les systèmes de contre-intelligence, qui ont aussi un problème ?

Y. B. Je ne sais pas. Si l'on parle d'une IA qui est plus intelligente que nous, il est plausible que les moyens de sécurité que nous avons ce soient pas suffisants... Y. L. C. Mais nous n'aurons pas les systèmes de sécurité d'aujourd'hui, nous aurons encore des décennies de retard avant d'être capables de manipuler le langage. En fait, ce n'est qu'une impression, leur compréhension du monde est encore très superficielle.

5. LE RISQUE POUR LA DÉMOCRATIE

En outre danger mis en avant est celui de manipulation des humains grâce à l'IA, notamment en diffusant de fausses informations...

Y. L. C. Ça vaillat pour un réseau social, je vois bien plutôt pour dire que les tentatives de contourner les processus démocratiques, d'influencer l'opinion ou de fauter de la désinformation ne sont pas des problèmes nouveaux... Y. B. Ils pourraient être amplifiés par l'IA.

Y. L. C. Non, car le problème, c'est la désinformation du contenu, pas sa création. Quand ça va tout et pour tout deux personnes ! Y. B. Mais si on a des dizaines de milliers de comptes créés par des IA qui désinforment ces contenus ?

Y. L. C. Ces comptes seront détruits avant de pouvoir diffuser de l'information.

Des systèmes de protection ont été mis en place depuis les tentatives de démantèlement de l'élection présidentielle américaine par les Russes en 2016. Aujourd'hui, il y a des équipes de chercheurs qui travaillent sur ces dispositifs, en particulier chez Meta, Google ou YouTube, pour limiter la création de faux comptes et la diffusion de désinformations ou de messages haineux. On dispose de contre-mesures, qui précèdent souvent les attaques sur l'IA.

6. L'INTELLIGENCE ÉMOTIONNELLE

À l'avenir, peut-on imaginer que les machines soient dotées d'une intelligence émotionnelle, ou qu'elles aient une forme d'intuition ?

Y. L. C. Les machines ont déjà de l'intuition. En fait, la limitation actuelle est qu'elles n'ont que de l'intuition, elles ne sont capables ni de raisonner ni de planifier. Yoshua et moi sommes intéressés par les travaux de Daniel Kahneman (Prix Nobel d'économie et auteur du livre *Système 1/Système 2* : les deux vitesses de la pensée, Flammarion, 2012, 2012). Selon sa théorie, il y a deux modes de production de l'action : un mode spontané, impulsif et réflexif, qu'il appelle « système 1 », et un mode qui est délibéré et fait intervenir notre modèle mental du monde, la planification et la réflexion, le « système 2 ». Pour l'instant, tous les systèmes comme ChatGPT ont des systèmes 1 mais pas de système 2. Cela fait plusieurs années que Yoshua et moi cherchons à concevoir des machines qui sachent planifier et raisonner d'une manière compatible avec l'apprentissage automatique.

Yann Le Cun

« Quand un système de sécurité est attaqué, des contre-mesures sont prises. Le danger, l'IA, ce n'est pas le problème. C'est la solution.

Et l'apprentissage des émotions a un rôle à jouer ?

Y. L. C. D'une certaine manière, les émotions sont inséparables de la vue intelligente. Pour nous, l'intelligence et la capacité à planifier doivent être associées à une capacité de prédire le résultat d'une séquence d'actions : est-ce que les émotions sont nécessaires pour être capables ou naïfs ?

Y. B. C'est le cœur de ce que je veux dire : les machines qui sont capables de faire avant des sentiments très similaires à cela.

Y. B. Contrairement à Yann, je pense que ce serait une mauvaise idée d'avoir des IA avec des émotions. Cela contraindrait la recherche à des systèmes de manière dangereuse pour les humains : on peut imaginer que des humains demandent quelque chose qui ne soit pas mauvais en soi, mais que, pour y parvenir, la machine crée des sous-objectifs qui sont potentiellement dangereux. Il y a une grande différence entre les machines et les humains : nous avons des émotions et des consciences morales, qui comprennent le bien et le mal, mais on ne sait toujours pas le faire. Tant qu'on ne le comprend pas, il me paraît dangereux de donner aux machines des émotions, des objectifs ou une conscience de soi. Est-ce qu'on veut avoir des machines qui, comme nous, dérivent plus que tout survivent ?

Y. L. C. C'est le cœur de ce que je veux dire : les machines qui sont capables de faire avant des sentiments très similaires à cela.

7. LA RÉGULATION

On a récemment le fondateur d'OpenAI, Sam Altman, demander au Congrès américain d'être régulé et d'impulser d'une réglementation de l'Europe. Pour vous, faut-il réguler l'IA, et comment ?

Y. L. C. Sam Altman aborde la régulation pour dire : « Il n'y aura que quelques entreprises qui auront le droit d'entraîner de grands modèles, dont OpenAI, et il faut former le reste de la population pour des dizaines d'applications de sécurité. Mais si les plateformes open source sont interdites par la réglementation à cause de craintes de sécurité, nous aurons encore des années de retard avant d'être capables de manipuler le langage. En fait, ce n'est qu'une impression, leur compréhension du monde est encore très superficielle.

Y. B. Au Canada et aux États-Unis, il y a déjà des lois qui tentent de légiférer sur l'intelligence artificielle. Certains pensent que les lois dans ce domaine doivent être très précises sur ce qu'on doit faire ou ne pas faire. D'autres disent, au contraire, qu'il faudrait laisser que la législation se limite à définir de grands principes généraux et que les détails de ce qui est permis ou pas soient confiés à une agence de type FDA (Food and Drug Administration) ou l'FAA (Federal Aviation Administration) chargée d'ajuster la réglementation au fur et à mesure que de nouveaux problèmes apparaissent. Mais pour ce qui est de l'IA, absolument que ces organismes aient la capacité de faire évoluer la réglementation rapidement, car la technologie avance très vite.

8. LA PLACE DES PETITS ACTEURS

Dans le monde où il faut énormément de données, de données, de centres de stockage, de processeurs très chers, est-ce qu'il reste une place pour la recherche universitaire, le start-up ou les petits pays ?

Y. L. C. Les plateformes comme GPT-4 coûtent en effet très cher à entraîner et ne sont accessibles qu'à de grands contributeurs.

La seule manière pour que les petits acteurs continuent à jouer un rôle, c'est que ces plateformes soient open source. Ce serait à nous biologistes, pour les modèles spécialisés pour des dizaines d'applications de sécurité. Mais si les plateformes open source sont interdites par la réglementation à cause de craintes de sécurité, nous aurons encore des années de retard avant d'être capables de manipuler le langage. En fait, ce n'est qu'une impression, leur compréhension du monde est encore très superficielle.

Y. B. Côté universitaire, il y a encore beaucoup d'équipe de recherche fondamentale.

Il y a beaucoup de choses peuvent être faites sans moyens énormes. Sur l'open source, les logiciels ou comme Yann qui était essentiel. Ça permet en effet d'accélérer les progrès scientifiques et technologiques. Aujourd'hui, je pense que tout est pas nécessairement ouvert ou partagé. Si on regarde la recherche, on voit souvent des règles pour éviter de clamer des êtres humains ou de modifier les gènes de nos enfants. Tant et tant par-dessus, il y a des choses qui peuvent être explorées librement et bien, et c'est à nous sur lesquelles on met des balises. Ce n'est pas nécessairement la fin du monde si on réglemente des choses qui sont dangereuses.

9. L'AVENIR DU TRAVAIL

Quel avenir pour « l'humain travaillant » avec sa créativité, ses compétences ?

Y. B. Je constate que les experts ont des opinions divergentes sur le sujet. Il est donc important que les gouvernements proposent des filtres de sécurité sociale pour faire face à des changements rapides du marché de l'emploi.

Comment tout ça va-t-il se terminer ? Personne n'a de boule de cristal. Il y a des arguments raisonnables dans un sens comme dans l'autre. Y. L. C. Les progrès de la technologie conduisent en général à des changements d'équilibre de l'économie. Ils peuvent rendre caduques certains métiers et en transformer beaucoup d'autres. Et si les progrès se font rapidement, une proportion importante de la population peut être laissée pour compte. Maintenant, si on écoute les économistes qui ont étudié l'impact des révolutions technologiques sur l'emploi, il semble que l'effet ne soit jamais persistant, car les nouvelles technologies créent au moins autant d'emplois qu'elles n'en font disparaître.

Si on se projette dans le futur, il ne faut pas craindre que les technologies accaparent tous nos emplois. Il faut plutôt imaginer que chacun aura un staff de machines intelligentes à son service qui travaillent avec et pour lui. Celles-ci seront peut-être plus intelligentes que leur propriétaire. Mais l'humain sera le pilote.

Y. B. Aujourd'hui, il y a beaucoup de petits emplois qui valent avant tout par leur aspect relationnel et qu'il faut vouloir à cœur. Moi, j'aimerais que mon petit-petit soit accompagné à la crèche ou en école humaine, même si des machines avaient la capacité de le faire.

On a besoin que la personne qui nous tient la main à l'hôpital soit humaine. Ça n'empêche pas qu'on utilise la technologie pour trouver des solutions médicales. Mais tout ce qui concerne l'humain sera difficile à remplacer. ■

Yoshua Bengio

« L'humanité que nous petit-fils ont accompagnée à la crèche par un être humain, même si des machines avaient la capacité de le faire.

par Benoît Georges Et Stefano Lupat