

SCÉNARIOS POUR UN AVENIR SOUS LE SIGNE DE L'IA(G)

Anton Korinek

L'IA pourrait être en passe de dépasser
l'intelligence humaine. Soyons prêts



L'intelligence artificielle (IA) progresse à grands pas, à un rythme qui s'est accéléré ces dernières années. ChatGPT, lancé en novembre 2022, peut, à la surprise des utilisateurs, générer des textes et des codes de qualité humaine, traduire sans faille des documents rédigés en diverses langues, rédiger des contenus créatifs et apporter des réponses éclairantes à des questions, le tout à un niveau inédit.

Or, en arrière-plan, les modèles qui ont jeté les bases de l'IA générative connaissent une évolution rapide depuis plus d'une décennie. Les ressources de calcul utilisées pour former les systèmes d'IA les plus avancés ont doublé tous les six mois au cours des dix dernières années. Ce qui est actuellement à la portée des modèles d'IA générative de pointe était impensable il y a seulement quelques années : ils permettent aux meilleurs consultants du monde, aux programmeurs et même aux économistes de réaliser des gains de productivité importants (Korinek, 2023).

Conjecture sur l'accélération de l'IA

À la lumière des récentes avancées dans le domaine de l'IA, d'éminents chercheurs prévoient que le rythme des progrès pourrait se maintenir, voire s'accélérer, dans les années à venir. En mai 2023, Geoffrey Hinton, un informaticien qui a jeté les bases théoriques de l'apprentissage profond, a expliqué que ses vues avaient beaucoup changé : « J'ai soudainement changé d'avis sur la question de savoir si ces machines seront plus intelligentes que nous. » Il a émis l'hypothèse que l'intelligence artificielle générale (IAG), qui renvoie à la capacité de comprendre, d'apprendre et d'exécuter toutes les tâches intellectuelles à la portée de l'être humain, pourrait être réalité d'ici cinq à vingt ans.

Certains chercheurs en IA sont sceptiques. Cette divergence de vues reflète l'épais brouillard dans lequel on se trouve quant au rythme des progrès à venir, à savoir s'il s'accélère ou s'il finira par se stabiliser. En outre, une grande incertitude règne en ce qui concerne l'incidence économique générale des progrès de l'IA ainsi que le ratio avantages/préjudices de ses applications de plus en plus sophistiquées.

L'incertitude est également liée à des questions fondamentales sur la nature de l'intelligence et les capacités du cerveau humain. Le graphique 1 présente deux perspectives antagoniques de la distribution de la complexité des tâches que le cerveau humain peut accomplir.

La plage 1 illustre la première perspective, selon laquelle la capacité du cerveau humain à accomplir

des tâches de plus en plus complexes est illimitée. Cela reflète l'évolution économique depuis la révolution industrielle : à mesure que l'automatisation a progressé, l'être humain a automatisé des tâches simples (à la fois mécaniques et cognitives), et des travailleurs ont été réaffectés à l'exécution d'un plus grand nombre de tâches plus complexes, c'est-à-dire qu'ils se sont déplacés dans la queue droite de la distribution de la complexité des tâches illustrée dans le graphique. Une extrapolation simple donne à penser que ce processus se poursuivra à mesure que l'IA et l'automatisation des tâches cognitives progressent.

Suivant la seconde perspective (plage 2 du graphique 1), il existe une limite supérieure à la complexité des tâches que le cerveau humain peut accomplir. Selon la théorie de l'information, le cerveau humain est un centre de calcul qui traite continuellement une masse de données. Les entrées sont les perceptions sensorielles (visuelles, sonores et tactiles, entre autres), et les sorties sont les actions physiques, les pensées et les réactions émotionnelles. Même les expressions complexes qui caractérisent l'être humain, telles que les émotions, la créativité et l'intuition, peuvent être considérées comme des sorties de calcul ou le produit des interactions complexes des circuits neuronaux et des réactions biochimiques. Bien que ces processus soient très élaborés et que leur complexité ne soit pas parfaitement comprise, cette perspective semble indiquer qu'il existe une limite supérieure définitive à la complexité des tâches que le cerveau humain peut accomplir.

Ces deux perspectives ont des implications radicalement différentes eu égard aux possibilités d'automatisation. En 2023, le cerveau humain est le dispositif de calcul le plus avancé sous l'angle de la capacité à effectuer avec rigueur un large éventail de tâches intellectuelles. Toutefois, si la seconde perspective s'avère exacte, les systèmes d'IA modernes auront tôt fait de combler leur retard. En fait, sur la base de nombreuses mesures de la complexité computationnelle, les modèles fondationnels de pointe talonnent déjà le cerveau humain. La complexité computationnelle du cerveau humain est limitée par la biologie, et la capacité du cerveau à transmettre des informations à d'autres entités intelligentes (humaines ou artificielles), par la lenteur de la transmission des informations par les sens et le langage. Les systèmes d'IA continuent de progresser rapidement et peuvent communiquer des informations beaucoup plus rapidement.

Les robots sont à l'honneur dans ce restaurant à thème de Chennai, en Inde ; ils y font aussi le service.

Se préparer à plusieurs scénarios

Les économistes soutiennent depuis longtemps que la meilleure façon de faire face à l'incertitude est d'élaborer un éventail de scénarios. Étant donné que les avis des experts de renommée mondiale sur les progrès futurs de l'IA diffèrent radicalement, il serait imprudent de mettre tous ses œufs dans le même panier et d'établir des plans économiques à partir d'un seul scénario. Au contraire, l'incertitude face à l'avenir devrait nous inciter à parer à toute éventualité et à analyser avec minutie divers scénarios susceptibles de se concrétiser, allant du statu quo à l'IAG. Non seulement la planification fondée sur des scénarios reflète l'incertitude actuelle, mais elle nous permet également d'énoncer concrètement les possibilités et les risques et nous aide à élaborer des plans d'urgence et à nous préparer à de multiples résultats possibles.

Voici trois scénarios technologiques couvrant un large éventail de résultats possibles auxquels les responsables de la politique économique devraient prêter attention.

Scénario I (statu quo — la technologie évolue comme par le passé) : Les progrès de l'IA stimulent la productivité et permettent d'automatiser diverses tâches cognitives, mais donnent aussi aux travailleurs touchés la possibilité d'occuper de nouveaux emplois en étant généralement plus productifs. Ce point de vue est représenté à la page 1 du graphique 1.

Scénario II (scénario de référence — l'IAG deviendra réalité dans vingt ans) : Au cours des vingt prochaines années, l'IA évoluera progressivement jusqu'au stade de l'IAG (capacité d'accomplir toutes les tâches humaines), dévalorisant ainsi les travailleurs (Susskind, à paraître). Cela correspondrait à la perspective d'une intelligence limitée, représentée par la plage 2 du graphique 1, ainsi qu'à l'hypothèse selon laquelle il faudra vingt ans pour que les tâches cognitives les plus complexes soient à la portée de l'IA.

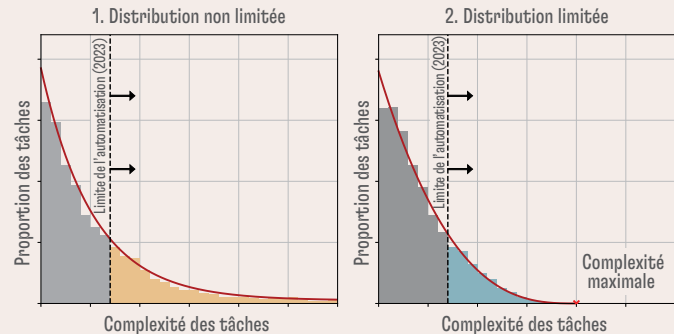
Scénario III (scénario radical — l'IAG deviendra réalité dans cinq ans) : Ce scénario reproduit le scénario II, mais avec un délai plus court, de sorte que l'IAG et tous ses effets sur la main-d'œuvre se concrétiseraient d'ici à cinq ans.

Bien que je sois en proie à une grande incertitude, j'estime qu'au moment où j'écris ces lignes, chacun de ces scénarios a une probabilité supérieure à 10% de se concrétiser. À mon avis, pour tenir compte de l'incertitude et bien préparer l'avenir, les décideurs devraient prendre au sérieux chacun de ces scénarios, soumettre les cadres des politiques économique et financière à un test de résistance pour chaque scénario et les réformer pour s'assurer qu'ils sont adéquats, le cas échéant.

GRAPHIQUE 1

L'IA repousse les limites

De nombreuses tâches que seuls les humains pouvaient auparavant accomplir ont déjà été automatisées (barres en gris). Au fur et à mesure que l'IA progresse, l'être humain continuera à exécuter des tâches plus complexes indéfiniment (barres en jaune), ou l'IA finira par dépasser toutes les capacités humaines (barres en bleu).



SOURCE : Anton Korinek.

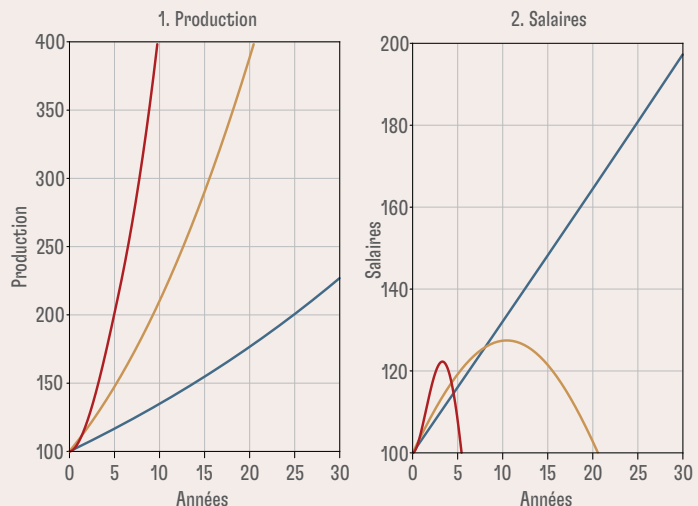
GRAPHIQUE 2

Scénarios pour la production et les salaires

L'essor de l'IAG propulse la croissance et entraîne une forte baisse des salaires concurrentiels.

(année initiale = 100)

— Scénario prudent — Scénario de référence — Scénario radical



SOURCE : Anton Korinek.

NOTE : IAG = intelligence artificielle générale

Les trois scénarios peuvent avoir une incidence très différente sur un large éventail d'indicateurs, notamment la croissance économique, les salaires et le rendement du capital, la viabilité budgétaire, les inégalités et la stabilité politique. En outre, ils font ressortir la nécessité de réformer les filets de sécurité sociale et la fiscalité et se répercutent sur la conduite de la politique monétaire, la réglementation financière et les stratégies industrielles et de développement.

Korinek et Suh (2023) ont analysé les implications des scénarios pour la production et les salaires dans un modèle macroéconomique d'automatisation courant. Les résultats pour les trois scénarios sont illustrés dans le graphique 2 ; la trajectoire de la production pour chaque scénario est représentée dans la page 1, et la trajectoire des salaires concurrentiels, dans la page 2.

Trois principaux éléments ressortent de leur analyse :

Premièrement, alors que la croissance suit la trajectoire observée depuis des décennies dans le scénario prudent (statu quo), la production augmente beaucoup plus rapidement dans les deux scénarios de concrétisation de l'IAG, car elle n'est plus limitée par la pénurie de main-d'œuvre.

Deuxièmement, dans les trois scénarios, l'augmentation des salaires est subordonnée à la rareté de la main-d'œuvre. Les salaires s'effondrent lorsque l'économie est sur le point d'atteindre le stade de l'IAG.

Troisièmement, dans les deux scénarios de concrétisation de l'IAG, la croissance de la production et l'effondrement des salaires ont la même cause : le remplacement d'une main-d'œuvre rare par des machines comparativement plus abondantes. Cela donne à penser qu'il conviendrait de créer des institutions qui compensent les pertes de revenus des travailleurs et veillent à ce que les bienfaits de l'IAG conduisent à une prospérité partagée.

Le graphique 2 donne un aperçu de l'incidence de changements technologiques sans précédent sur la macroéconomie, mais il est préférable de le considérer comme une illustration des possibilités plutôt que comme une prévision précise.

De nombreuses réserves s'imposent. Tout d'abord, le modèle qui sous-tend le graphique s'inscrit dans une économie efficiente dans laquelle la rémunération du travail est compétitive. Divers facteurs peuvent ralentir le déploiement de l'IAG par rapport à ce qui est technologiquement possible, qu'il s'agisse de frictions organisationnelles, de réglementations, de restrictions à l'accumulation de capital (telles que les goulets d'étranglement dans la chaîne d'approvisionnement en puces) ou de choix de société concernant la mise en œuvre de l'IAG. Même lorsque la technologie peut remplacer les travailleurs, la société peut choisir de laisser les humains continuer à exercer certaines fonctions, comme celles de prêtre, de magistrat et de législateur. Cette nostalgie pourrait soutenir la demande de travailleurs humains à perpétuité (Korinek et Juelfs, à paraître).

Pour déterminer auquel des scénarios d'évolution de l'IA l'avenir qui se profile à l'horizon correspond le plus au fil des événements, les décideurs devraient surveiller les indicateurs avancés dans de multiples domaines, en gardant à l'esprit que tous les efforts visant à prédire le rythme des progrès sont confrontés à une très grande incertitude. Les indicateurs utiles sont les repères technologiques, les niveaux d'investissement dans le développement de l'IA, l'adoption des technologies de l'IA dans l'ensemble de l'économie et les tendances macroéconomiques et du marché du travail qui en découlent. Les repères technologiques constituent la mesure la plus directe de l'efficacité des systèmes d'IA dans l'exécution d'un large éventail de tâches. Les investissements, par exemple dans la recherche-développement, les talents et les puces électroniques, permettent de déterminer la part des ressources consacrée au développement de l'IA. Les indicateurs de l'adoption croissante de l'IA dans tous les secteurs de l'économie refléteront l'utilité dans la pratique des systèmes qui en résultent. Enfin, les répercussions macroéconomiques ressortiront des statistiques de la productivité et des tendances du marché du travail.

Le suivi de ces signaux complémentaires permet aux décideurs d'adapter leurs interventions aux nouvelles réalités

de l'IA. Mais nous devons rester humbles : l'avenir pourrait bien nous étonner.

Les trajectoires économiques radicalement différentes que supposent les trois scénarios décrits plus haut soulignent l'importance d'élaborer des cadres d'action adaptatifs, qui permettent de réagir avec souplesse aux événements. Pour chaque scénario, les décideurs devraient soumettre les institutions existantes à des tests de résistance et les réformer, s'il y a lieu, pour s'assurer de leur résilience. La réforme peut consister en des mesures progressives, telles que la réforme de la fiscalité et l'élargissement des filets de sécurité sociale, ou de nouveaux programmes, tels que l'introduction d'un revenu minimal qui pourra être augmenté, si nécessaire.

Les décideurs devraient charger des équipes de spécialistes de mettre en œuvre des processus de planification itératifs fondés sur des scénarios afin de les aider à actualiser régulièrement leurs points de vue sur l'évolution de la probabilité des différents scénarios. En acceptant l'incertitude grâce à une approche adaptable et fondée sur des scénarios, on peut maximiser les bienfaits de l'évolution de l'IA dans la sphère économique et en atténuer les risques. **F&D**

ANTON KORINEK est professeur au département d'économie et à la Darden School of Business de l'Université de Virginie. Il est responsable de l'économie de l'IA au Centre for the Governance of AI, chercheur adjoint au National Bureau of Economic Research et chargé de recherche au Centre for Economic Policy Research.

BIBLIOGRAPHIE

Korinek, Anton. 2023. Generative AI for Economic Research: Use Cases and Implications for Economists. *Journal of Economic Literature* 2023, 61(4).

Korinek, Anton, and Megan Juelfs. Forthcoming. "Preparing for the (Non-existent?) Future of Work." *Oxford Handbook of AI Governance*. Oxford, UK: Oxford University Press.

Korinek, Anton, and Donghyun Suh. 2023. "Scenarios for the Transition to AGI." University of Virginia working paper, Charlottesville, VA.

Susskind, Daniel. Forthcoming. "Technological Unemployment." *Oxford Handbook of AI Governance*. Oxford, UK: Oxford University Press.