DANIEL ANDLER

Professeur émérite de Sorbonne Université, membre de l'Académie des sciences morales et politiques, philosophe

Bienvenue à cette deuxième séance du jour sur le thème de l'IA. Elle suit une séance passionnante en français et précède une autre séance sur l'IA générative dans les affaires. Ce sont en quelque sorte trois séances très complémentaires. À l'évidence, le monde s'est engagé sur la voie du développement de l'intelligence artificielle et nous savons tous que cette évolution ira très loin. C'est d'ailleurs tout ce que l'on peut en dire. Cependant, même si nous ignorons la longueur et la destination de cette voie, nous avons notre mot à dire. Pour autant que nous le sachions, l'IA ne se pilote pas elle-même. Nous ne pouvons pas laisser les Sept Magnifiques et leurs homologues chinois décider pour nous. Concernant l'avenir, la grande question est de savoir jusqu'à quel point l'IA évoluera. Cette question peut se comprendre de deux manières. La première interprétation, et la plus fréquente, est de savoir de quelles prouesses cognitives l'IA sera finalement capable. La seconde interprétation, la plus importante, est de savoir à quel point cela sera un progrès pour l'humanité. Nous ne pouvons pas nous contenter de supposer que ce qui est bon pour l'IA est bon pour l'humanité, contrairement à ce qu'en pensent certains, ni supposer le contraire, même si d'autres en sont convaincus. Nous pouvons essayer de trouver un juste milieu et conseiller de profiter des bienfaits de l'IA tout en empêchant ou limitant ses éventuels dégâts. Néanmoins, cela ne nous éclaire pas du tout sur la manière d'atteindre cet objectif. Nous devons faire mieux que cela. Nous en sommes capables. À ce jour, les réalisations et les effets de l'IA constituent un riche corpus de preuves, d'études de cas pour ainsi dire, qui peuvent nous servir à évaluer si nous allons dans la bonne direction ou si nous devons corriger, et dans quelle mesure, la trajectoire actuelle. Évidemment, la complication majeure réside dans le fait que les directions que peut prendre l'IA ne sont pas entièrement fixées à l'avance. Elles dépendent des découvertes, délibérées ou fortuites, que feront les scientifiques et les ingénieurs, à court terme ou à long terme. Comme toujours en science, nous ne trouvons pas toujours ce que nous cherchons, et ce que nous trouvons ne correspond pas forcément à ce que nous cherchons. Cela ne doit pas nous dissuader d'acquérir une perspective la plus claire possible sur la question de savoir où nous mène l'IA et où nous voulons qu'elle nous mène.

Nos intervenants, comme lors des deux séances précédente et suivante, nous apporteront des éclairages très variés sur ce thème. En guise d'introduction, je vous propose une carte mentale simple, voire simpliste, des questions concrètes à aborder. Cette carte comprend un volet optimiste, avec les réussites et les promesses que l'on nous vante sur l'IA, et un volet plus sombre qui dénonce les pièges et les effets pervers de l'IA. Pour ce qui est des réussites et des promesses, je propose de les classer dans un tableau 2x2. Sur un axe, certaines sont réelles ou réalistes ; et d'autres ne sont étayées par aucune réalisation actuelle ou notion existante. Sur un autre axe, certaines sont classées comme de prime abord bénéfiques, et d'autres comme dangereuses, risquées voire nettement nocives. Parmi les exemples

associant réalisme et bénéfice *a priori*, on trouverait les moteurs de recherche intelligents, l'assistance scientifique, médicale, juridique et de conduite automobile, éléments très bien décrits par M. Ezzat lors de la séance précédente. Parmi les exemples associant réalisme et risque de toxicité, citons les chatbots quasi parfaits, de type compagnon IA, ainsi que les IA produisant de la poésie, des romans, des compositions musicales, etc. Du côté des promesses bénéfiques sans fondements et des réalisations extravagantes, citons l'IA générale explicable, l'IA générative, les véhicules autonomes de niveau 5 sûrs. Je sais, toutefois, que certains me désapprouveront. Du côté des systèmes dangereux, mais heureusement non viables, on trouve l'IAG de type boîte noire, ainsi que les IA entièrement autonomes agissant en tant que médecins, enseignants, juges, policiers, thérapeutes, politiciens et PDG. Pour ce qui est des pièges ou des effets pervers, je propose de distinguer entre les problèmes inextricables que l'on doit affronter et atténuer au mieux avec la plus grande lucidité, et les difficultés que l'on peut et doit donc gérer activement. Parmi les dilemmes, je noterais peut-être les armes mortelles autonomes et les nouveaux champs de bataille, la cybersécurité, le creusement des inégalités intra et interétatiques, et, pour utiliser un terme vilipendé au sein de la bonne société favorable à l'IA, la déshumanisation. Parmi les défis qu'il convient de relever figurent le respect de la vie privée, de la propriété intellectuelle, de la simple vérité factuelle, d'un certain degré de transparence dans la mesure où cela ne contredit pas la notion même d'IA, les effets environnementaux, les pertes de compétences et, enfin et surtout, la création de fausses identités. Pour ce dernier cas, une interdiction totale aurait pu être imposée d'emblée à toutes les technologies contribuant à cette atteinte fatale à la notion de confiance car elles sont l'équivalent, dans la sphère sociale, des armes nucléaires pour la survie biologique. Il était évident que ces technologies déboucheraient sur le chaos actuel en termes de falsification et de compagnons IA. Tout le monde, en tout cas hors de la Silicon Valley, savait que cela mènerait à un désastre. Cependant, nous pouvons et devons nous battre pour revenir à la raison et rétablir la confiance. Certains des exemples cités sont controversés et chacun d'eux, comme bien d'autres encore, nécessite un examen minutieux. Je relève pourtant deux faits remarquables. Le premier, c'est que les débats généralistes sur l'IA sont nécessaires, mais rarement concluants. Les débats sont plus productifs lorsqu'ils sont circonscrits à des domaines spécifiques tels que la médecine, l'éducation, la guerre, la santé, la justice, la prise de décision politique, etc. Le deuxième, c'est que l'innovation *per se* n'est pas un bien souverain qui doit être protégé à n'importe quel prix. Il arrive que des réglementations la ralentissent, mais les bonnes réglementations ne le font pas sans raison. L'humanité est toujours bénéficiaire lorsque l'innovation et la réglementation travaillent en bonne intelligence.

Après ces remarques introductives, passons désormais au groupe de discussion. Je ne présenterai pas nos intervenants car leur biographie est disponible dans la brochure. Nous parlerons à tour de rôle dans l'ordre où nous sommes assis, six à sept minutes chacun, et je serai un modérateur intraitable sur les temps de parole.