

## DOMINIQUE CARDON

Professeur de sociologie, médialab  
Sciences Po, auteur de *Culture Numérique*  
(Presses de Sciences Po, 2019)

# “LA TUMULTUEUSE AVENTURE DE L’IA”

Dès que le terme intelligence artificielle est prononcé, nous regardons vers le futur. Ce projet technologique a pourtant une histoire ancienne. Débuté au milieu des années 1950, il a traversé une succession heurtée de promesses tonitruantes et de brutales déconvenues. Aujourd’hui, pour la troisième fois, après une première vague à la fin des années 1950 et une seconde au milieu des années 1980, la promesse de l’intelligence artificielle (IA) que l’on croyait moribonde a fait depuis 2012 un retour inattendu dans le répertoire des grandes espérances technologiques.

L’IA d’aujourd’hui n’a cependant pas grand-chose à voir avec celle des deux premières vagues. Elle apparaît même comme la revanche inattendue d’une orientation technologique encore plus ancienne, issue de la cybernétique des années 1940, qui avait été moquée et méprisée par les promoteurs de l’IA des deux premières vagues. L’histoire des sciences et des techniques ressemble rarement à une suite linéaire d’innovations incrémentales. Elle se présente plus souvent comme une succession heurtée de progrès et d’échecs, de changements de paradigmes et de bifurcations inattendues.

La tumultueuse aventure de l’IA en est un remarquable exemple. Elle est tout entière habitée par une opposition frontale entre deux paradigmes qui, même si certains souhaitent les réconcilier, s’opposent comme l’eau et l’huile depuis plus d’un demi-siècle. Le premier, appelé « symbolique », a dominé les deux premières vagues de l’histoire de l’IA, dans les années 1960, puis dans les années 1990. Son objectif est d’introduire dans un programme informatique des règles de raisonnement qui, des prémices vers la conclusion, suivent un chemin rationnel, intelligible et logique.

Cependant, aussi sophistiquée la modélisation du raisonnement puisse-t-elle être, les programmes développés dans cette perspective n’ont su produire que des intelligences formelles, automatiques et maladroitement dès qu’elles doivent répondre à des situations imprévues. Aussi est-ce une conception radicalement différente de l’intelligence des machines qui a été proposée par le courant « connexionniste » sous le nom d’« apprentissage profond » (*deep learning*).

De façon très soudaine, confortée par des prouesses impressionnantes, ce courant ancien et longtemps décrié est devenu dominant et a brutalement démodé le paradigme symbolique qui dominait, à l’université et dans les entreprises, une grande partie de la recherche en IA. Ce changement de paradigme n’est pas perçu par le grand public, mais pour les chercheurs, les informaticiens et les développeurs de l’IA, il a constitué une rupture brutale et un bouleversement dans la manière de concevoir l’intelligence des machines.

Avec l’apprentissage profond, les données sont réparties en couches de réseaux de neurones (une fonction mathématique qui s’active ou ne s’active pas en fonction d’un seuil) qui transmettent, ou ne transmettent pas, l’information aux couches suivantes. Longtemps décriée et souvent méprisée pour son manque de sophistication mathématique, cette

technique d'apprentissage appelé « profond » a connu un fulgurant retour en grâce lorsque l'augmentation des capacités de calcul des ordinateurs a permis de mettre en œuvre une idée ancienne dont les réalisations étaient restées longtemps limitées.

L'intelligence de l'IA actuelle est donc moins celle de la logique, du raisonnement ou de la précision qu'une formidable machine statistique qui, profitant du volume et de la granularité d'énormes jeux de données, parvient à faire des prédictions statistiques de plus en plus correctes.

Par exemple, avant la révolution du *deep learning*, les systèmes de traduction automatique en IA cherchaient à faire apprendre à la machine le vocabulaire, la syntaxe et la grammaire. Cette stratégie n'a jamais très bien fonctionné. Les traductions restaient maladroitement et fautive.

### **Le défi de l'opacité**

À partir de 2010, les nouveaux systèmes d'IA ne s'embarassent plus de l'apprentissage de règles sémantiques ou grammaticales et traduisent pourtant beaucoup mieux. Le système n'a plus besoin de raisonner en allant de la règle vers le cas, puisque désormais il connaît tellement de cas que la règle est une propriété émergente du traitement statistique des données.

L'IA connexionniste ne raisonne pas à partir de règles, de symboles et de grandes catégories. Aussi le chemin qui conduit de la question vers la conclusion est-il devenu opaque. Il n'est plus possible de le comprendre, même si de nouvelles recherches montrent qu'il est partiellement possible de vérifier la qualité de certaines opérations produites par les réseaux de neurones. Cette opacité constitue aujourd'hui le défi majeur des nouvelles techniques connexionnistes de l'intelligence artificielle. Elles ont montré que, lorsqu'elles disposent de suffisamment de données, elles sont plus précises et subtiles que les modèles symboliques.

Mais le prix « faustien » de cette efficacité est l'abandon de la vérifiabilité et d'une certaine intelligibilité des calculs. Plus encore, même si ces accidents sont statistiquement rares, les réseaux de neurones peuvent produire des résultats massivement erronés sans que l'on puisse en comprendre les raisons. Voilà une question bien difficile : dans quelles conditions et pour quel type d'usage pouvons-nous accepter qu'un algorithme soit globalement plus performant, mais moins compréhensible ?