

Daniel ANDLER, *Intelligence artificielle, intelligence humaine : la double énigme*, Paris, Gallimard, « NRF Essais », 2023, 434 p.

C'est un lieu commun de l'IA qu'une tâche dont on estime qu'elle nécessite de l'intelligence pour être menée à bien perde immédiatement ce caractère dès lors qu'elle est réalisée par une machine⁵. Des activités paradigmatiquement intellectuelles comme jouer aux échecs, tenir une conversation en langage naturel ou composer de la musique, peuvent désormais être en apparence réalisées par des systèmes artificiels, qui se révèlent pourtant à l'examen totalement dénués d'intelligence. L'objectif « prométhéen » de l'IA consistant à créer des systèmes artificiels dotés d'une intelligence équivalente ou supérieure à l'intelligence humaine semble ainsi paradoxalement aussi bien à portée de main qu'irréremédiablement hors de portée. Cette « malédiction » (p. 15) qui entoure chacune des réalisations de l'IA ne prend toutefois pas l'allure d'un mystère inexplicable, mais celle d'une double énigme : d'un côté, on s'interroge sur la nature de l'intelligence humaine qui, bien que réalisée, n'en reste pas moins difficile à cerner. D'un autre côté, on cherche à savoir ce que pourrait réellement être l'intelligence artificielle, actuellement à l'état de promesse perpétuelle. C'est cette double énigme que l'A. entend dissiper dans cet ouvrage, qui revêt un intérêt majeur pour quiconque s'intéresse à ces questions.

Le livre se divise en deux parties. Dans la première, l'A. retrace le chemin parcouru pour tenter de résoudre cette double énigme de l'intelligence depuis les débuts de l'IA jusqu'à nos jours. Des premières thèses d'Alan Turing jusqu'au connexionnisme actuel, en passant par les origines cybernétiques de l'IA et sa voie symbolique, la présentation historique proposée par l'A. comble un vide dans la littérature francophone : il s'agit incontestablement de l'exposé le plus complet et le plus éclairant qui existe aujourd'hui sur le sujet. Son enquête méticuleuse fait ainsi connaître l'histoire intellectuelle et technique de l'IA, ses productions principales, ses changements de direction et de paradigme, ses réussites et ses échecs. Cette première partie dévoile également les bases conceptuelles sur lesquelles la discipline telle que nous la connaissons aujourd'hui a été bâtie et continue d'évoluer.

Mais l'ouvrage ne présente pas qu'un état de l'art rigoureux et informé de la discipline et des interrogations philosophiques qu'elle soulève : il est aussi résolument argumentatif. Dans la seconde partie du livre, l'A. constate en effet la chose suivante : que l'on adopte une approche « anthropique » – celle prônée par les pionniers de l'IA visant à reproduire mécaniquement les processus mentaux tels qu'ils sont réalisés chez les êtres humains – ou que l'on suive au contraire une voie « ananthropique » (qui

5. L'A. introduit une distinction commode que nous adopterons dans cette recension entre « IA » entendue comme un domaine de recherche (une discipline scientifique) et « intelligence artificielle » entendue comme ce qui est justement visé par cette discipline (p. 22).

cherche à s'émanciper de cette idée de reproduction pure), une idée centrale demeure chez les chercheurs en IA et en sciences cognitives. Cette idée est tout d'abord que l'intelligence artificielle doit être abordée par le biais d'une comparaison avec l'intelligence humaine. Elle consiste ensuite en un accord plus ou moins tacite sur la définition de l'intelligence, pensée comme une capacité particulière : celle de résoudre des problèmes. Or l'A. nous met en garde contre cette tentation – selon lui réductrice – de considérer les actions intelligentes sous le prisme de la résolution de problèmes (ce qu'il a qualifié dans ses précédents écrits de « panproblémisme »). Après l'avoir soumise à une analyse conceptuelle détaillée et après avoir pris au sérieux les difficultés qu'elle pose, l'A. suggère qu'il serait en définitive plus judicieux d'abandonner cette vision problémiste de l'intelligence.

À cette *pars destruens* répond une *pars construens*, qui se fonde en partie sur une proposition d'extension des termes traditionnels de la comparaison. L'A. suggère qu'afin d'éclairer la nature de l'intelligence, il convient de ne pas se limiter à l'étude d'un seul type d'intelligence réalisé (l'intelligence humaine), mais de prendre également en compte ses autres incarnations (l'intelligence des animaux non humains). Cela le mène ainsi à proposer une nouvelle définition de l'intelligence. Celle-ci serait biologique – caractéristique de l'animalité – et consisterait dans la capacité d'un individu à s'adapter et à prendre en charge des nouvelles *situations*. Ce dernier concept vient alors remplacer celui de problème : un problème, nous dit l'A., se pose à un agent, n'est pas ambigu et est résolu lorsqu'une réponse satisfaisante lui est donnée. Une situation, en revanche, est toujours singulière, concrète, vécue, et directement liée à la relation entre un agent et son environnement (son *Umwelt*). Là où il est concevable qu'une machine puisse bel et bien résoudre des problèmes, l'A. défend qu'il lui est au contraire conceptuellement impossible de pouvoir faire face à des situations.

L'ouvrage s'achève par une double réflexion critique. L'A. examine d'une part le futur de l'IA et de ses créations : du côté de ce qu'il appelle « l'intelligence artificielle rêvée », il s'attache en particulier à exposer l'idée spéculative et néanmoins très discutée de « super-intelligence ». Il évalue d'autre part les risques éthiques associés à ces technologies, qu'ils soient seulement hypothétiques, futurs, ou qu'ils s'imposent déjà à nous. L'A. introduit notamment l'idée d'intelligence artificielle dite « transformatrice », dont l'impact sur la société humaine serait à ce point remarquable qu'elle la transformerait profondément. Là où le concept de super-intelligence laissait l'A. perplexe quant à l'éventualité de sa réalisation, cette notion d'intelligence artificielle transformatrice l'amène à réfléchir aux enjeux éthiques bien réels, inquiétants et souvent très pressants, qu'elle soulève. L'ouvrage – stimulant de bout en bout – s'achève sur un appel précautionneux à l'adoption d'un principe général de modération, qui recommande de « n'utiliser l'intelligence artificielle que lorsque sa contribution nette est positive » (p. 361).