

DU MÊME AUTEUR
AUX ÉDITIONS ALLIA

Schizophrénie numérique

ANNE ALOMBERT

De la bêtise artificielle

POUR UNE POLITIQUE
DES TECHNOLOGIES NUMÉRIQUES



ÉDITIONS ALLIA

16, RUE CHARLEMAGNE, PARIS IV^e

2025

Quelqu'un peut parler sans arrêt, et sa parole ne rien dire. Par contre un silence peut en dire long. Mais que signifie "dire" ?

Martin HEIDEGGER, "Langue de tradition et langue technique", 1962.

... nous sommes transpercés de paroles inutiles, de quantités démentes de paroles et d'images. La bêtise n'est jamais muette ni aveugle.

Gilles DELEUZE, "Les intercesseurs", 1985.

... que l'exosomatization cybernétique puisse engendrer une bêtise artificielle industrielle, c'est la question qui doit nous guider ici.

Bernard STIEGLER, "Bêtise et intelligence artificielles dans l'Anthropocène", 2018.

DEPUIS le lancement de ChatGPT par l'entreprise OpenAI en novembre 2022 jusqu'au développement de Deepseek en janvier 2025, une véritable course à l'innovation s'est amorcée dans le champ des "intelligences artificielles génératives". Le déploiement massif de ces dispositifs dans les sociétés et leur intégration au sein des écosystèmes numériques plus traditionnels comme les moteurs de recherche, les suites bureautiques ou encore les réseaux sociaux préparent une bifurcation technologique majeure, dont le caractère ambivalent n'a jamais été aussi patent. Yoshua Bengio (MILA¹) et Geoffrey Hinton (Google), deux pionniers dans le champ de l'apprentissage automatique, reconnaissent eux-mêmes les dangers que leurs inventions pourraient présenter, au point de les décrire comme des "risques pour l'humanité". En mars 2023,

1. L'acronyme MILA désigne le Montréal Institute for Learning Algorithms, désormais nommé le Quebec Artificial Intelligence Institute, un organisme qui rassemble des chercheurs spécialisés dans le domaine de l'apprentissage profond.

quatre mois après le lancement de ChatGPT, Elon Musk et d'autres entrepreneurs et experts publiaient une lettre ouverte sur le site du Future of Life Institute, pour demander une pause dans le développement de l'intelligence artificielle. Ils évoquaient alors des "esprits numériques" susceptibles de provoquer l'extinction de l'humanité. Quelques mois plus tard, l'entreprise xAI du même Musk développait pourtant son propre *chatbot*, nommé Grok, censé concurrencer ChatGPT.

Depuis, les systèmes d'IA générative ne cessent de se multiplier : de Gemini (Google) et Llama (Meta) en passant par Dall-E ou Sora (OpenAI), les capitaux, les *start-ups* et les logiciels fleurissent. Début 2025, c'est Deepseek qui fait irruption sur un marché encore largement dominé par la Silicon Valley. Ce modèle, développé par une *start-up* détenue par un fonds spéculatif chinois, fait alors la une de nombreux médias. Son développement a nécessité beaucoup moins de puissance de calcul que les autres modèles existants. Quelques jours après la mise en accès du modèle chinois, Donald Trump, à peine investi en tant que président des États-Unis, annonce un budget de 500 milliards de dollars pour le projet Stargate, destiné à renforcer

la souveraineté technologique du pays, en partenariat avec OpenAI. Pendant ce temps, Musk, qui a vu sa fortune bondir de plusieurs milliards de dollars après l'élection de Trump, est nommé à la tête du Département de l'Efficacité Gouvernementale, pour réduire les dépenses publiques.

Depuis plus de cinq ans, on assiste à une accélération sans précédent du rythme des innovations et des investissements dans "l'intelligence artificielle générative", qui vont de pair avec une concentration des richesses et du pouvoir entre les mains de quelques entrepreneurs californiens (ou chinois). La massification des usages de ces technologies n'a d'égale que la domination des discours alarmistes à leur sujet dans les médias, qui coexistent avec les apologues technosolutionnistes ou les annonces de gigantesques financements. Ces deux positions extrêmes semblent d'ailleurs se développer main dans la main, comme si les adeptes de la "superintelligence" artificielle et les tenants des "risques existentiels" alimentaient finalement une même course à l'innovation débridée et à la panique généralisée. Qu'il s'agisse de discours promotionnels ou d'alertes apocalyptiques, les principaux récits qui entourent

cette accélération technologique adoptent les schémas narratifs propres à l'idéologie transhumaniste, qui se déploie dans la Silicon Valley depuis plus d'une trentaine d'années. Celle-ci s'appuie sur un scénario tout tracé, selon lequel l'humanité, dépassée par les performances de ses machines, sera bientôt à la merci d'une "superintelligence" prenant le contrôle de son destin. Couramment désigné sous le nom de "singularité technologique", ce point de bifurcation se voit tour à tour revendiqué ou dénoncé, chacun s'accordant néanmoins sur son arrivée imminente.

Le mythe de l'imitation,
entre technosciences et science-fiction

Cette idéologie largement répandue à l'Ouest des États-Unis trouve ses racines de l'autre côté de l'Atlantique, aux origines de l'informatique. Dès 1950, dans un article intitulé "Computing machinery and intelligence", le mathématicien et cryptologue Alan Turing¹

1. A. Turing, "Computing Machinery and Intelligence", *Mind*, 59, 1950.

proposait d'attribuer la faculté de penser à un ordinateur à condition que la machine puisse simuler le langage humain en fournissant des suites de signes adaptées. Six ans plus tard, la notion d'"intelligence artificielle" voit officiellement le jour lors de la conférence de Dartmouth, organisée par Marvin Minsky et John McCarthy pendant l'été 1956, au Nord-Est des États-Unis. Ce concept promotionnel et stratégique a pour principale fonction d'attirer des financements en prétendant reproduire des facultés intellectuelles (comme le langage, le jeu ou le raisonnement) à travers des programmes informatiques. L'économiste et informaticien Herbert Simon, partie prenante du projet, reconnaît lui-même que le terme de "traitement automatique de données" aurait été plus approprié que celui d'"intelligence artificielle". Qu'importe, c'est cette dernière locution, plus aguichante, qui rencontre le succès. En 1956, pour les tenants de l'"intelligence artificielle" dite "symbolique", qui propose de reproduire les processus cognitifs au moyen d'instructions et d'opérations logiques, il s'agit surtout de s'affirmer face à leur concurrent, le courant de l'"intelligence artificielle connexionniste", issu de la cybernétique, qui propose de reproduire

les connexions neuronales au moyen de fonctions mathématiques implémentées dans des circuits électroniques.

Depuis 1943 en effet, les recherches cybernétiques de Warren McCulloch et Walter Pitts ont proposé de modéliser les activités cérébrales par des fonctions mathématiques, que l'on appelle des "réseaux de neurones formels" et qui sont au fondement des technologies d'"apprentissage automatique"¹. Ces travaux ont inspiré Frank Rosenblatt, inventeur du premier "réseau de neurones" en 1957, qui s'est également nourri des travaux de psychologie cognitive de Friedrich Hayek, théoricien du néolibéralisme². Après une phase de ralentissement des recherches, l'intelligence artificielle connexionniste connaît un nouvel essor dans les années 1980. Le développement de l'électronique numérique fournit alors la puissance de calcul nécessaire au fonctionnement des

1. W. McCulloch et W. Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity", *Bulletin of Mathematical Biophysics* 5, 1943, pp. 115-133.

2. F. Rosenblatt, "The Perceptron, A Perceiving and Recognizing Automaton", *Cornell Aeronautical Laboratory*, 1957 et F. Hayek, *L'Ordre sensoriel: une enquête sur les fondements de la psychologie théorique* (1952), traduit de l'anglais par Philippe R. March, Paris, CNRS éditions, 2001.

dispositifs. Les recherches se poursuivent dans les années 2000, toujours en lien avec les avancées dans le secteur des puces et des processeurs graphiques (GPU). L'augmentation des bases de données liée à la diffusion massive d'Internet joue aussi un rôle, car le fonctionnement de ces technologies implique des calculs statistiques sur des quantités massives de données. Durant les années 2010, les travaux sur les méthodes d'apprentissage auto-supervisé, les algorithmes de rétropropagation et les "réseaux de neurones convolutifs" se développent, notamment à travers les recherches de Yann Le Cun (Meta, ex-Facebook), Yoshua Bengio (MILA) et Geoffrey Hinton (Google), tous trois récompensés par le prix Turing en 2018. Ces recherches sont à la base de toutes sortes de logiciels et d'applications: logiciels d'autocomplétion ou de traduction automatique, générateurs de textes, d'images ou de vidéos, et différents types de *chatbots*.

Si les systèmes d'intelligence artificielle se sont beaucoup perfectionnés ces dernières années, les présupposés théoriques, eux, ont peu changé. Qu'il s'agisse de l'IA symbolique ou de l'IA connexionniste, dans les deux cas, l'enjeu consiste à imiter ou simuler les activités de l'esprit ou du cerveau par

des machines algorithmiques. Même si les algorithmes mobilisés dans les deux courants diffèrent, les analogies entre l'esprit, ou le cerveau, et l'ordinateur ne sont jamais questionnées. On soutient qu'une supposée intelligence, considérée comme uniforme et universelle, peut être modélisée à travers des opérations mathématiques, qui se déroulent indifféremment dans des corps vivants ou dans des supports numériques. Dans les deux cas, le raisonnement est simple : penser ou apprendre revient à calculer, ces calculs peuvent être formalisés logiquement sous forme d'algorithmes, ces algorithmes peuvent être exécutés par des machines numériques et ces machines se voient donc attribuer l'intelligence, l'apprentissage ou la pensée.

Les machines s'apparentent ainsi à des doubles des humains, capables d'imiter ou de simuler leurs aptitudes psychiques ou mentales à travers de puissantes capacités de calcul et de traitement de données. De 1950 à 2020, des technosciences à la science-fiction, cet imaginaire anthropomorphique est omniprésent. Du côté des technosciences, Raymond Kurzweil, l'un des fondateurs du mouvement transhumaniste, qui fut chercheur en informatique au MIT et directeur de l'ingénierie chez

Google, annonce depuis les années 1990 l'âge des "machines intelligentes" ou des "machines spirituelles"¹. Trente ans plus tard, en 2019, Yann Le Cun, directeur de la recherche en IA chez Facebook-Meta, soutient que "les machines apprennent"². En 2021, Sam Altman, directeur d'OpenAI, affirme que les logiciels et programmes informatiques seront "capables de penser" dans les cinq prochaines années³. Dans la bouche de ceux qui cherchent à vendre leurs dispositifs, ces affirmations n'ont évidemment rien de scientifique : elles ne font que rejouer les stéréotypes de la science-fiction. Que l'on pense à la figure menaçante de HAL dans *2001 : L'Odyssée de l'espace* de Stanley Kubrick (1968) ou au séduisant système d'exploitation nommé Samantha dans le film *Her* de Spike Jonze (2013), il s'agit toujours du même paradigme, celui de l'imitation ou de la simulation de l'humain par la machine. Jamais il n'est question du fonctionnement interne de

1. R. Kurzweil, *The Age of Intelligent Machines*, MIT Press, 1990 et R. Kurzweil, *The Age of Spiritual Machines*, Viking Press, 1999.

2. Y. Le Cun, *Quand la machine apprend*, Paris, Odile Jacob, 2019.

3. S. Altman, "Moore's Law for Everything", blog de Sam Altman, 16 mars 2021.

ces dispositifs ou de leurs enjeux politiques : les individus ne les comprennent nullement, ils les craignent ou les aiment, comme s’il s’agissait de personnes humaines.

De l’“intelligence artificielle”
à l’automatisation de l’esprit

C’est précisément cet imaginaire anthropomorphique que le philosophe Gilbert Simondon critique dans sa thèse sur les objets techniques, soutenue en France en 1958, deux ans après la conférence de Dartmouth et un an après l’invention du Perceptron¹. Simondon critique la “représentation mythique du robot”, qui consiste à considérer les machines comme des “doubles de l’homme”, en leur attribuant de manière implicite “une âme et une existence séparée et autonome”. Il soutient qu’“un homme cultivé ne se permettrait pas de parler des objets ou des personnages peints sur une toile comme de véritables réalités, ayant une intériorité, une volonté bonne ou mauvaise”. Pourtant, remarque Simondon,

1. G. Simondon, *Du mode d’existence des objets techniques* (1958), Paris, Flammarion, 2024.

“ce même homme” parle des machines comme étant “animées d’intentions” et leur prête “l’usage de sentiments”. Cette anthropomorphisation des machines n’est pas neutre politiquement. Elle masque un fantasme de toute-puissance et de domination : “l’homme qui veut dominer ses semblables suscite la machine androïde”, afin de lui “déléguer son humanité”; “il cherche à construire la machine à penser, rêvant de pouvoir construire la machine à vouloir, la machine à vivre, pour rester derrière elle sans angoisse, libéré de tout danger, exempt de tout sentiment de faiblesse, et triomphant médiatement par ce qu’il a inventé”. L’“idolâtrie de la machine” apparaît comme le revers d’une “aspiration technocratique au pouvoir inconditionnel”.

Une vingtaine d’années plus tard, dans une conférence intitulée “Le cerveau et la pensée”¹, le philosophe, médecin et historien des sciences Georges Canguilhem pointe lui aussi des risques du même type. Dans le contexte du développement des sciences cognitives et de l’informatique industrielle, il relève

1. G. Canguilhem, “Le cerveau et la pensée” (1980), in *Georges Canguilhem, philosophe, historien des sciences*, Paris, Albin Michel, 1993.