n° 25

Septembre 2024

Côté Philo

Le journal de l'enseignement de la philosophie

ACIREPh

Association pour la Création d'Instituts de Recherche sur l'Enseignement de la Philosophie

Bulletin de rentrée

L'INTELLIGENCE ARTIFICIELLE EN CLASSE.

UN NOUVEL HORIZON CRITIQUE POUR LA PHILOSOPHIE ?

Journées d'étude de l'ACIREPh

Côté Philo est une publication de l'ACIREPH

Association pour le Création d'Instituts de
Recherche sur l'Enseignement de la philosophie

Éditeur : ACIREPH, 21 rue du Général Faidherbe, bâtiment A, 94130 NOGENT-SUR-MARNE

Directrice responsable : Fanny Bernard,
ACIREPH, 21 rue du Général Faidherbe, bâtiment A, 94130 NOGENT-SUR-MARNE

Rédacteur en chef : Serge Cospérec
ACIREPH, 21 rue du Général Faidherbe, bâtiment A, 94130 NOGENT-SUR-MARNE

Imprimerie : Fadora, 55, rue Jean-Pierre Timbaud, 75011 PARIS

Pour écrire dans *Côté Philo*

Adressez vos textes au comité de rédaction *email* : contact@acireph.org

Le Comité de rédaction informera l'auteur de sa décision : acceptation, acceptation sous réserve de modifications, ou non-publication.

Les textes envoyés ne sont pas retournés à leurs auteurs

Retrouvez *Côté Philo* et les autres travaux de l'ACIREPH sur notre site

www.acireph.org

Côté Philo

Le journal de l'enseignement de la philosophie

Information

Pourquoi un *Côté Philo* avec le bulletin ?

Nombre d'entre vous préfèrent recevoir le bulletin de l'Association en format papier. Certains bulletins seront édités dans *Côté Philo* parce que son format se prête mieux aux contenus longs et dépassant la simple information sur la vie associative.

Cette décision répond aussi aux nouvelles contraintes de notre contrat avec La Poste. Les frais postaux représentent un poste de dépense important dont nous voulons continuer à contrôler le coût. L'envoi d'un bulletin plus étoffé publié dans notre revue *Côté Philo* répond à ce souci.

L'édition du Bulletin sous le format numérique traditionnel est toujours envoyée à tous nos membres inscrits sur « Listireph », notre liste de diffusion.

Bulletin de l'Association pour la Création des Instituts de Recherche sur l'Enseignement de la Philosophie

Septembre 2024

L'ACIREPh vous souhaite une très bonne rentrée, pour cette année scolaire marquée par la stupéfiante absence de ministre. Dans cette période politique confuse, notre enseignement n'est touché par aucune réforme, mais le collège est dans une situation complexe face à la mise en place incertaine des groupes de niveaux en français et maths, aux conséquences inquiétantes pour le lycée dans quelques années.

Lors de l'année scolaire 2023-2024, l'ACIREPh a connu une grande activité : les **Journées d'étude de novembre 2023** sur les séries technologiques se sont **poursuivies à Rennes en juin 2024**. Ces Journées organisées par le groupe local de Rennes, après celles organisées par celui de Grenoble à Grenoble en 2023 et à Rennes en 2024 illustrent la volonté de l'association de décentraliser ses activités.

Nos prochaines Journées d'Études seront consacrées à **l'enseignement de la philosophie face à l'Intelligence Artificielle**. Elles auront lieu les **7 et 8 novembre**, au lycée Jean Zay à Paris, qui nous accueillent depuis l'an dernier. Le programme de ces Journées sera riche et passionnant, vous le trouverez dans ce bulletin. Ces occasions de nous rencontrer, de partager nos questionnements et nos pratiques, sont précieuses autant que rares, depuis que de nombreuses formations professionnelles n'ont plus lieu sur le temps scolaire ni en présentiel.

Ces deux Journées sont déclarées comme **stage de formation** syndicale, en association avec la CGT éducation cette année. Ce partenariat ne nous engage pas plus que les précédents, nous sollicitons chaque année un syndicat différent. Il vous suffira de demander au Rectorat une **autorisation d'absence** pour ces deux Journées, autorisation **qui ne peut vous être refusée**. Dans ce bulletin, vous trouverez un modèle de lettre de demande de congé pour formation syndicale, à transmettre par voie hiérarchique à votre rectorat **avant le 7 octobre 2024** (délai légal).

Toutes nos Journées d'Études affichent désormais complet : nous vous invitons à vous inscrire aux prochaines le plus rapidement possible !

Parce que beaucoup de nouveaux collègues ont adhéré à l'ACIREPh, le ***Côté Philo* n° 24** était consacré à la présentation de notre association, ses orientations, ses combats, ses propositions. **Deux nouveaux numéros** sont en préparation, en prolongation des dernières Journées d'Étude sur les séries technologiques. Ils reviendront sur les difficultés rencontrées et des pratiques pédagogiques pour en surmonter quelques-unes. Si vous souhaitez écrire pour *Côté Philo* un compte-rendu de pratique, une recension d'ouvrage, un billet d'humeur, toute contribution est la bienvenue !

Comme à chaque nouvelle année scolaire, n'oubliez pas de renforcer notre association en renouvelant votre adhésion à l'ACIREPh, si possible en ligne, à défaut par le bulletin d'adhésion ci-joint. Enfin, n'hésitez pas à solliciter le C.A. ou les groupes locaux pour prendre part plus activement à nos échanges et travaux. Le C.A. a été renouvelé à l'assemblée générale 2024 avec l'arrivée de Daphné Leroux, Camille Chamoix, Marie Coasne-Khawrin, Laurent Germain, ainsi que mon élection en tant que présidente et celle de Laurent Germain en tant que secrétaire. Deux commissions de travail aimeraient se mettre en place : l'une consacrée à la philosophie pour enfants, et l'autre, à l'organisation d'un séminaire de recherche en enseignement de la philosophie pour 2025-2026. Toutes les nouvelles énergies sont les bienvenues pour faire vivre notre association !

Bien cordialement,

Fanny Bernard, pour le C.A.

L'INTELLIGENCE ARTIFICIELLE EN CLASSE.

UN NOUVEL HORIZON CRITIQUE POUR LA PHILOSOPHIE ?

Journées d'étude de l'ACIREPh

Jeudi 7 et vendredi 8 novembre 2024

Lycée Jean Zay – 10 Rue du Dr Blanche – Paris 16ème – Métro Ranelagh, Jasmin

Les Journées d'Étude de l'ACIREPh sont ouvertes à toutes celles et ceux que les questions de l'enseignement de la philosophie intéressent*, et s'adressent tout particulièrement à nos collègues de philosophie, débutants ou expérimentés, qui souhaitent réfléchir collectivement à leur pratique, pour s'emparer des questions posées par leur métier.*

Depuis l'arrivée de ChatGPT nous ressentons les effets de l'IA sur notre métier. Nous voulons éviter de corriger des devoirs produits par une IA, quitte à nous résigner à n'évaluer que des travaux faits en classe. Nous sommes contraints à ce pis-aller, faute de temps et de moyens pour réfléchir collectivement aux transformations pédagogiques impliquées par une IA qui semble là pour rester. Comment éviter de subir ? Comment faire des choix éclairés concernant l'usage possible de ces IA ? Faut-il travailler contre l'IA ? sans elle ? Ou avec elle, mais comment et pourquoi faire ?

Mais l'IA modifie aussi certaines questions de philosophie. Quelle est la frontière entre l'humain et la machine ? Qu'est-ce au juste que l'« intelligence » ? Comment analyser les pratiques de pouvoir, les phénomènes de persuasion et de manipulation à l'époque de l'IA ? Le monde qui se prépare sera-t-il plus juste ou plus inégalitaire ?

Une culture technique, critique et en prise avec les questions les plus contemporaines, semble plus que jamais nécessaire aux élèves.

1. ChatGPT est-il intelligent ?

L'intelligence artificielle n'a-t-elle d'intelligence que le nom ou est-elle réellement intelligente ? Avec *Deep Blue* battant Kasparov aux échecs en 1996, puis *AlphaGo* écrasant un champion du monde de Go en 2017, l'orgueil naïf de l'humanité en a pris un coup. L'IA la surpassait dans des domaines où elle se croyait invincible par une machine. Puis, il y eut la « voiture autonome », et des robots capables de performances acrobatiques comme Atlas de *Boston Dynamics*.

Qu'importe ! Le langage résistait, jusqu'à l'apparition d'IA génératives comme ChatGPT. Depuis, des agents conversationnels (*chabots*) peuplent le quotidien (*Alexa*, *Siri*, *Gemini*, etc.). Les nouvelles IA sont capables d'apprendre et de se perfectionner, toutes seules. Elles sont savantes et patientes. On leur écrit, on leur parle ; elles nous écoutent ou nous lisent, elles nous comprennent et nous donnent des conseils. Elles sortent victorieuses du test de Turing : une machine peut tromper un humain dans une conversation. Leurs performances sont étonnantes.

Langage, perfectibilité, savoir : la frontière entre l'humain et la machine est-elle en train de tomber ? Les IA ont-elles égalé, voire dépassé, l'intelligence humaine ? Sont-elles capables de raisonner, d'inventer, de prendre des décisions, comme un être humain ? Auraient-elles même des états mentaux ?

Sur toutes ces questions, les débats sont vivants. Un cours de philosophie gagnerait à les aborder. Seulement, nous sommes peu familiers du domaine de l'IA. Comment fonctionne exactement ChatGPT ? Pourquoi est-il si performant ? Comment expliquer ses failles ? Etc.

Ces questions seront discutées aux Journées de l'ACIREPh, notamment le 1er jour avec la conférence du linguiste et directeur de recherche au CNRS, Bernard Victorri, qui exposera les éléments du succès actuels de l'IA et nous présentera le fonctionnement des nouveaux modèles d'IA génératives, et la manière dont elles changent la donne.

2. L'IA menace-t-elle la justice ?

Les citoyen·ne·s ont conscience des dangers que la désinformation, la diffusion d'images truquées, de faux enregistrements vidéos ou audios, font peser sur la démocratie. Mais ils ignorent souvent un danger moins visible : celui que représente l'automatisation croissante des processus de décision dont les conséquences peuvent gravement affecter la vie de chacun·e d'entre nous. L'IA est déjà utilisée pour accepter ou refuser un prêt bancaire, sélectionner ou éliminer automatiquement des candidats à un emploi. Le secteur public recourt aussi à l'analyse automatisée des dossiers : pour l'octroi de prestations sociales, l'administration de la justice, voire l'évaluation des risques de récidive d'un condamné pour décider ou non d'une libération conditionnelle. Or des biais dans les bases de données engendrent des discriminations raciales ou genrées. À cause de failles dans les systèmes, des personnes se voient refuser une prestation sociale, le droit de prendre l'avion, un titre administratif (carte d'identité, de séjour) auquel ils ont droit, sans explication ni recours : « votre dossier ne passe pas » dit l'agent, dont le jugement est désormais délégué à des intelligences

censées être plus objectives. Les failles et l'opacité des algorithmes ouvrent la voie à un nouvel arbitraire, sans sujet ni intention. Le monde de Kafka n'est pas loin.

Autant d'exemples qui montrent comment l'IA reconfigure les rapports entre Technique, Justice et Liberté, trois notions au programme de toutes les classes. Nous pourrions saisir cette occasion pour montrer aux élèves que la philosophie aide à penser les problèmes de son temps mais aussi pour renouveler notre approche de ces notions à la lumière de ces nouveaux enjeux.

Nous aborderons ces problèmes lors de la conférence de la 2ème journée d'étude de l'ACIREPh. Camille Girard-Chanudet, docteure en sociologie au Centre d'Étude des Mouvements Sociaux de l'EHESS, évoquera les questions liées au travail de conception et à l'usage de l'IA dans nos sociétés, à travers un exemple concret : les algorithmes d'apprentissage automatique développés dans le domaine de la justice.

3. Usages didactiques de l'IA : contrainte, duperie ou opportunité ?

L'I.A outil. Le monde enseignant a d'abord découvert l'IA par le pire : son utilisation par les élèves ou les étudiant·e·s pour faire leurs devoirs et tricher. Mais l'hostilité à l'IA cède peu à peu la place à la curiosité, voire un intérêt. Et si ChatGPT pouvait nous aider dans notre travail ? Par exemple, à faire le plan d'un cours, sa description ou son résumé, à créer des QCM ou des questions sur un texte, à générer des exercices. Et, rêve ou cauchemar, à corriger les copies ?

L'IA, pour renouveler l'enseignement. Le fait que les IA génératrices de texte produisent des imitations tout à fait convaincantes d'une dissertation philosophique traditionnelle ne devrait-il pas nous questionner sur la pertinence de cet exercice ? Ne devrions-nous pas réorienter nos exercices canoniques pour qu'ils ne soient plus aussi facilement confondables avec le baratin d'un ChatGPT ? L'IA pourrait être l'occasion de repenser le format ou la nature des productions demandées aux élèves, de pratiquer l'écriture ou la lecture de textes philosophiques. Il s'agirait moins, alors, de diaboliser l'outil que d'apprendre à l'utiliser honnêtement et intelligemment, comme le font déjà certain·e·s enseignant·e·s ; à en percevoir aussi les limites, les dangers et les dérives.

Aucune de ces questions n'est taboue à l'ACIREPh. Elles auront leur place, notamment dans certains ateliers sur des expériences faites par des collègues, partagées et soumises à la discussion.

PROGRAMME DES JOURNÉES D'ÉTUDE

JEUDI 7 NOVEMBRE 2024

9H00 – Accueil des participants

9H30 – Allocution d'ouverture. Présentation des journées

10h–12h Conférence-débat : **Bernard VICTORRI,** directeur de recherche au CNRS :

L'Intelligence Artificielle : comment ça marche ?

14H –15H30 et 15H45–17H15 : ATELIERS au choix

VENDREDI 8 NOVEMBRE 2024

9H00 – Accueil des participants

10h–12h Conférence-débat : **Camille GIRARD-CHANUDET,** Centre d'Étude des Mouvements Sociaux de l'EHESS :

Perspectives sur la justice algorithmique

14H–15H30 : ATELIERS au choix et 15H45–17h15 : retour, bilan et suites des JE

ATELIERS animés par des professeur.e.s de philosophie

- Dissertation et Chat GPT
- Écrire avec l'IA
- Trois applications du Web pour l'enseignement de la philosophie
- Jouons avec des neurones (atelier avec ordinateur personnel recommandé)
- Arpentage : *Schizophrénie numérique* d'Anne Alombert
- Le procès d'une voiture autonome
- IA et Arts / IA et amitié (sous réserve)

N.B. : Les Journées des jeudi 7 et vendredi 8 novembre sont déclarées comme **stage de formation syndicale**, cette année en partenariat avec ***CGT Éduc'action***.

La formation sur le temps de travail est un droit pour tous les personnels de l'Éducation nationale, syndiqués ou non. Il suffit d'adresser à votre Rectorat ***une demande de congé pour formation syndicale AVANT LE 7 OCTOBRE 2024***.

(un modèle de lettre demande de congé pour formation syndicale est notre site)

Inscription uniquement en ligne sur *acireph.org*

avant le 27 octobre 2024

dans la limite des places disponibles

À lire !

Daniel Andler, *Intelligence artificielle, intelligence humaine : la double énigme*, Gallimard, 2023.

Daniel Andler est mathématicien, philosophe, et spécialiste des sciences cognitives. Il a co-dirigé récemment *La cognition. Du neurone à la société* (2018), une somme rassemblant les contributions des meilleurs spécialistes des divers disciplines du domaine.

Mais Daniel Andler est aussi spécialiste de l'intelligence artificielle, et trente ans après *Intelligence artificielle : Mythes et limites* (déjà !), coécrit avec Hubert Dreyfus, il revient sur le sujet avec *Intelligence artificielle, intelligence humaine : la double énigme*, un ouvrage érudit et passionnant.

Beaucoup d'entre nous ont découvert Daniel Andler le 24 octobre 2002. Alors professeur à Paris IV, responsable du département d'études cognitives à l'ENS-Ulm, Daniel Andler avait très gentiment accepté d'intervenir à nos Journées d'études sur ce sujet : « *peut-on faire un cours sur la conscience sans passer par les sciences cognitives ?* ». Il ne s'agissait aucunement pour lui de vanter les mérites ou la pertinence des sciences cognitives, mais d'en discuter philosophiquement avec nous. Adhérent de l'ACIREPh depuis cette rencontre, son soutien fidèle nous touche beaucoup. Il fallait le dire.

Connaître pour philosopher. La philosophie s'appuie sur des savoirs qu'elle ne produit pas ; son enseignement, tout comme sa pratique, en présuppose l'assimilation, sinon la maîtrise. C'était l'objet des Journées 2002 : « *quelle place faire aux savoirs dans l'enseignement de la philosophie ?* ». Elle se pose avec acuité avec l'intelligence artificielle, domaine scientifique qui donne lieu à toutes sortes de spéculations et de fantasmes (y compris philosophiques), alors que peu s'y intéressaient ou s'en souciaient avant l'irruption de ChatGPT.

Si nous voulons réfléchir avec les élèves aux questions que soulève l'intelligence artificielle, comment connaître nous-mêmes ce qu'il faudrait en savoir ?

Le livre de Daniel Andler tombe à point nommé. Il est dense, clair et pédagogique. C'est l'ouvrage d'un scientifique, il sait de quoi il parle. Et c'est un livre de philosophe qui traite de questions fondamentales en philosophie.

La première partie du livre retrace l'histoire scientifique et conceptuelle de l'IA. Daniel Andler explique l'enjeu des deux paradigmes rivaux : l'approche *symbolique ou cognitiviste* qui vise à décrire le fonctionnement de l'esprit humain grâce à des modèles logiques et l'approche *connexionniste* ou *neuronale* qui vise à reproduire, ou au moins simuler dans les machines, le fonctionnement du substrat matériel de la pensée : le cerveau et ses neurones. Alors que l'IA *symbolique* repose sur le raisonnement déductif et manipule des règles de logique

(son langage de programmation, facile à lire, s'apparente à la logique formelle), l'IA *neuronale* développe des algorithmes d'apprentissage imitant vaguement le fonctionnement neuronal, et repose sur un empilement d'équations formant un maquis d'opérations sur des nombres, souvent difficiles à interpréter. Les concepts fondamentaux de l'IA sont définis et expliquées avec précision et clarté, condition nécessaire d'une réflexion informée et critique sur l'IA.

Dans la seconde partie, Daniel Andler aborde *la question de l'intelligence*, humaine, animale et artificielle à la lumière des développements de l'IA et des sciences cognitives. Nous accédons ainsi au dernier état de la recherche sur ces questions.

Il revient notamment sur les concepts d'intelligence « faible » et « forte », d'« intelligence générale artificielle » et de « super-intelligence », au cœur des projets qui agitent et divisent la communauté des chercheurs en intelligence artificielle. Il les analyse, en explique les enjeux, tout en distinguant les projets scientifiquement robustes des visions nettement plus spéculatives, voire fantaisistes. Nous disposons ainsi d'éléments d'appréciation et de discrimination critiques, d'arguments solides et informés.

Enfin, atout considérable, Daniel Andler a écrit son livre après l'apparition des systèmes intelligents artificiels de traitement des langues fondés sur les « modèles massifs de langage » (*large language models*), comme ChatGPT. Il en explique le principe, en examine les forces, les faiblesses, dont l'importante question de la cécité sémantique de tous ces systèmes intelligents artificiels. Question récurrente dans le domaine de l'intelligence artificielle.

Pour se faire une idée du livre, on peut aussi lire la recension de Renaud Fabre et Aude Seurrat[1], parue dans *Distances et médiations des savoirs,* le 20 mars 2024.

[1] **Renaud Fabre et Aude Seurrat**, « Daniel Andler, *Intelligence artificielle, intelligence humaine : la double énigme* », *Distances et médiations des savoirs* [En ligne], 45 | 2024. URL : https://journals.openedition.org/dms/10002

REPÈRES RAPIDES SUR L'IA

1. L'intelligence artificielle, ce n'est pas nouveau !

Beaucoup ont découvert l'intelligence artificielle avec ChatGPT. Elle n'est pourtant pas née hier, loin de là. La plupart des chercheurs en intelligence artificielle font remonter la fondation officielle de leur discipline à un atelier (*workshop*) organisé en 1956 au Dartmouth College par un jeune mathématicien, John McCarthy, qui inventa pour l'occasion l'expression ***intelligence artificielle***. Il reconnut ensuite qu'elle était mal nommée car le but était de créer une machine Aussi intelligente, voire plus intelligente, que les humain, une intelligence *authentique* et non pas *artificielle*. Mais, dit McCarthy, « je devais lui donner un nom, alors je l'ai appelée "intelligence artificielle". » Le programme d'étude portait sur la création de systèmes capables d'apprendre, de raisonner, de traduire, de parler, de réaliser des découvertes scientifiques et même dotés de créativité artistique. Dans ses grandes lignes, ce programme définit encore une large partie des recherches en IA aujourd'hui.

Si l'intelligence artificielle comme discipline scientifique débute en 1956, c'est dans les années 1930 à 1950 que se développent, les recherches pour modéliser la pensée humaine afin de créer une machine capable d'égaler les humains dans tous les domaines intellectuels.

Dans son article de 1936, « *La calculabilité des nombres et son application au problème de la décision*[1] », le mathématicien Alan Turing décrit une « machine à calculer » (*computing machine*) capable de décider si un énoncé mathématique est démontrable. Cette idée est à l'origine des premiers ordinateurs. Dans un article de 1948, *Intelligent Machinery*[2], Alan Turing décrit une machine faite de « neurones » connectés aléatoirement et capable de s'auto-organiser : « le modèle le plus simple du système nerveux ». Le nom de Turing est surtout resté attaché à un « jeu d'imitation » décrit en 1950 dans l'article *Computing Machinery and Intelligence*[3]. Ce jeu, appelé depuis « test de Turing », a pour but de déterminer si une machine serait capable de « converser » de façon très élémentaire avec un humain au point qu'il ne puisse savoir s'il a affaire à un autre être humain ou à une machine.

---

[1] Alan Turing, « La calculabilité des nombres et son application au problème de la décision *[Entscheidungsproblem]* », Traduction Dominique Bonnaud-Dantil. Sur le site de l'INRIA : https://wiki.inria.fr/wikis/sciencinfolycee/images/1/17/Machine_de_Turing.pdf

[2] Alan Turing, « *Intelligent Machinery* »,1948.
Disponible sur ARCHIVE.ORG, URL : https://archive.org/details/turing1948

[3] Alan Turing, « *Computing Machinery and Intelligence* », 1950. Sur le site du Department of Computer Science and Electrical Engineering de l'UMBC.
URL : https://www.csee.umbc.edu/courses/471/papers/turing.pdf

2. Quelques dates et périodes clés

1943 : Warren McCulloch et Walter Pitts publient un article fondateur sur le neurone formel, première modélisation mathématique d'un neurone biologique. Ce modèle inspirera directement la création du premier neurone simulé par Frank Rosenblatt en 1957

1948 : Donald Hebb, neuro-psychologue, propose une théorie neuronale de l'apprentissage, par modification et renforcement des liaisons synaptiques entre neurones (appelée aussi règle de Hebb).

1950 : Alan Turing propose un test pour évaluer la capacité d'une machine à imiter un humain qui répondrait à des questions posées par un être humain.

1951 : premiers programmes d'intelligence artificielle de jeu de dames de Christopher Strachey et de jeu d'échecs de Dietrich Prinz.

1951 : Marvin Minsky et Dean Edmunds construisent la première machine à réseau neuronal, le SNARC (*Stochastic Neural Analog Reinforcement Calculator*) inspiré des principes de Donald Hebb.

1955-56 : John McCarthy propose un projet de *recherche sur l'intelligence artificielle* (1955) qui se tient en 1956 à Darmouth et qui considéré comme fondant l'intelligence artificielle comme discipline scientifique.

1957 : Franck Rosenblatt construit le Perceptron Mark I, premier ordinateur basé sur un réseau neuronal capable d'apprendre par essais et erreurs.

1965 : Joseph Weizenbaum crée le premier agent conversationnel, nommé Eliza, simulant une conversation avec psychothérapeute.

1967 : Richard GreenBlatt invente le langage LISP, devenu dans les années 80 le langage de référence pour la programmation et la recherche en intelligence artificielle.

1969 : Marvin Minsky et Seymour Papert démontrent les limites des réseaux de neurones dans leur ouvrage *Perceptrons*. Les doutes s'installent quant à l'utilité de poursuivre les recherches en IA.

> **Période de 1970 à 1980** : les projets d'intelligence artificielle buttent sur des obstacles scientifiques et des limitations technologiques. Les financements publics comme privés s'arrêtent, ouvrant ce qu'on appelle le premier hiver de l'intelligence artificielle (*AI Winter*).

1975 ; Paul Werbos formalise l'algorithme de « rétropropagation », une méthode d'entraînement des réseaux de neurones que l'on retrouve plus tard à peu près partout en apprentissage automatique comme en apprentissage profond.

> **1980 à 1990 :** abandonnant l'approche neuronale pour l'approche symbolique l'IA connaît un nouvel essor : c'est l'âge d'or des « systèmes experts », avec leur base de connaissances et de règles logiques, qui leur permet

d'exécuter des tâches d'expertise analogues à celles que font des experts humains

1982. le physicien John Hopfield relance l'intérêt pour les réseaux de neurones en créant un nouveau modèle de réseau (appelé depuis réseau de neurones de Hopfield). Il utilise notamment la règle de Hebb. Mais seule une poignée de chercheurs poursuivent les recherches fondées sur l'approche neuronale.

Période de 1990 à 2000 : l'intelligence artificielle connaît son deuxième hiver. Les systèmes experts n'ont pas tenu leurs promesses, leur coût était élevé et leur champ d'application trop restreint. Les recherches fondées sur l'approche neuronale commencent à porter leurs fruits. En 1987, Yann Le Cun contribue à la création des réseaux de neurones « convolutifs » utilisés plus tard dans l'apprentissage profond (*deep learning*).

1997 : *Deep Blue* d'IBM bat le champion du monde d'échecs Garry Kasparov.

2000 : Cynthia Breazeal (13tats-Unis) crée Kismet, la première tête robotique capable d'exprimer des émotions (comme la surprise, la gaîté, la colère, la tristesse).

2009 : Google lance son projet de voiture autonome.

2011 : *Watson*, le super calculateur d'IBM conçu pour répondre à des questions formulées en langage naturel, bat les champions Ken Jennings et Brad Rutter à *Jeopardy*.

2012 : l'équipe *Google Brain* conçoit un réseau de neurones capable de reconnaître les chats sur les vidéos YouTube.

2014 : le programme *Deep Face* des chercheurs de Facebook est capable de reconnaître des visages avec seulement 3 % d'erreur.

2016 : le programme *Alphago*, conçu l'équipe de *DeepMind (*devenue filiale de Google) bat Lee Sedol, l'un des meilleurs joueurs de Go.

2017 : *Alphago* bat le champion du monde du jeu de Go, Ke Jie, 3 parties à 0.

2017 : création d'une version encore plus puissante, Alpha *Go Zero*. Rien ne lui a été enseigné. Le système s'est entraîné uniquement en jouant contre lui-même. L'apprentissage profond (*deep learning*) montre sa supériorité « définitive » sur l'apprentissage auto-matique (*machine learning*) classique.

2017 : invention du *Transformeur* par les chercheurs de Google, l'algorithme au fondement des IA génératives comme ChatGPT. Ce réseau neuronal est capable de de capturer le « sens » des mots avec plus de précision, de suivre et conserver les informations contex-tuelles, et générer des textes très longs.

2018-2019 : sortie de GPT et GPT 2, les deux premiers grands modèles de langage d'*Open AI*. L'ère des IA génératives commence.

2019 : l'intelligence artificielle s'ouvre au grand public avec *Alexa*, l'assistant personnel intelligent d'Amazon capable de « comprendre » et de répondre à des questions courantes.

2020 : GPT-3, étonne par sa capacité à générer un texte semblable à celui d'un humain. Il peut répondre à des questions, résumer des documents, générer des histoires dans différents styles, traduire des textes en plusieurs langues. Peu de temps après GPT-3 produit des textes toxiques (haineux, dangereux). De son côté Timnit Gebru, co-directeur de l'équipe d'éthique de l'IA de Google, souligne les dangers potentiels associés aux grands modèles de langage. Et Google le pousse vers la sortie.

2022 : en novembre, *Open AI* lance ChatGPT (ou GPT 3.5), un modèle si puissant et attractif qu'il atteint un million d'utilisateurs en cinq jours et en compte plus de 100 millions deux mois plus tard.

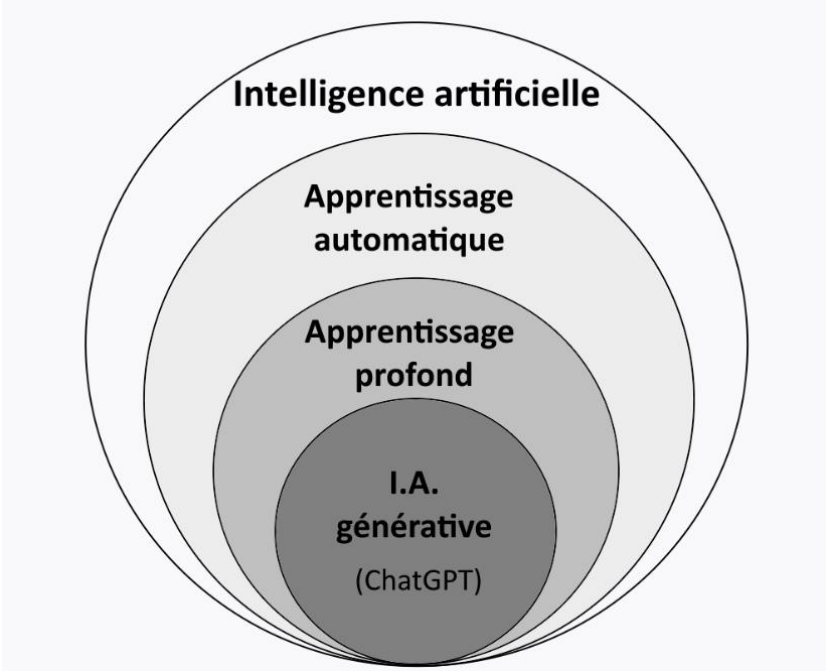
2023 : on augmente les performances des IA en construisant des systèmes toujours plus énormes. La nouvelle version de ChatGPT (GPT-4), lancée en 2024, aurait 1700 milliards de paramètres (*Open AI* refuse désormais de communiquer cette donnée), contre 173 milliards pour la version précédente (GPT 3.5).

Serge Cospérec

Machine learning et *Deep learning*.

Apprentissage automatique et apprentissage profond

Il est difficile de ne pas entendre parler du « *deep learning* » quand on s'intéresse à l'intelligence artificielle. En 2015, un article du Monde propose déjà d'expliquer « comment le « deep learning » révolutionne l'intelligence artificielle »[1]. Mais comme il n'aide pas à situer clairement l'**apprentissage profond** (deep learning) dans le champ de l'intelligence artificielle, essayons de le faire très modestement.



L'INTELLIGENCE ARTIFICIELLE CLASSIQUE

Jusqu'en 1990, la plupart des recherches en intelligence artificielle, suivaient l'approche dite **symbolique** ou « basée sur la connaissance », parce que toute la « connaissance » est généralement programmée « à la main ». Des ingénieurs collectent auprès d'experts (en médecine, chimie, géologie, etc.), les connaissances relatives à un domaine particulier, les *faits* et les *règles* qu'ils suivent pour exécuter une tâche donnée : établir un diagnostic médical, interpréter les résultats d'un spectromètre, déterminer une zone minière à intéressante à prospecter, etc. Ils encodent ensuite manuellement ces connaissances dans la machine, puis écrivent les programmes de traitement des

[1] Morgane Tual, « Comment le « deep learning » révolutionne l'intelligence artificielle », *Le Monde*, le 24 juillet 2015. [Disponible en ligne]

situations. L'approche symbolique reste encore très utilisée aujourd'hui pour planifier l'itinéraire de robots, la navigation basée sur le GPS, et dans tous les systèmes à base de règles (appelés « systèmes experts), en traitement automatique du langage, enfin dans les applications critiques (comme en médecine ou dans la défense) car elles exigent une explicabilité complète du fonctionnement du système (la « transparence »), ce qui est le cas avec l'intelligence artificielle classique, car toutes les connaissances et les règles y sont explicites.

L'APPRENTISSAGE AUTOMATIQUE

Depuis les années 2000, l'approche symbolique - l'IA classique - a largement cédé sa place à une autre approche, l'**apprentissage automatique** (*machine learning*), un sous-domaine de l'IA. L'idée est la suivante : plutôt que de tout enseigner à la machine, on cherche à la rendre capable d'« apprendre » à partir de « l'expérience », c'est-à-dire capable de découvrir toute seule les règles à suivre pour effectuer une tâche donnée et améliorer ses performances. Les programmes, conçus à partir d'algorithmes statistiques, confèrent à ces machines la capacité de modéliser des données (en extraire les caractéristiques), pour ensuite les classer par exemple. Mais les catégories de classement ou « groupements » doivent être enseignés à la machine lors d'une phase préalable d'entraînement, supervisé par un humain qui fournit des données accompagnées de leur classe (par exemple, chien, chat, humain, objet, pour des images) – ce qu'on appelle des données « étiquetées », car elles sont accompagnées de leur description. Tout l'étiquetage est réalisé à la main, c'est donc un travail énorme (la base de données d'entraînement *ImagNet* contient 1,2 millions d'images étiquetées manuellement !).

Autre caractéristique, l'apprentissage automatique se fonde sur l'utilisation de « réseaux de neurones », c'est-à-dire des algorithmes imitant la manière dont le cerveau « apprend ». En simplifiant, on peut dire qu'un apprentissage se traduit ou s'exprime, d'un point de vue neuronal, par la création de nouvelles liaisons entre les neurones, la stabilisation ou renforcement de liaisons, éventuellement leur élagage. Les réseaux de neurones simulés se composent d'éléments appelés *nœuds*, des unités de calculs pouvant être très vaguement assimilés à des neurones grandement simplifiés. Entre ces *nœuds* ou *unités*, il existe des *connexions*, appelées aussi *poids*. L'idée est la suivante : plus le *poids* sur la connexion d'un nœud A à un nœud B est important, plus grande est l'influence de A sur B. Les poids servent à régler l'influence des « neurones » dans l'obtention du résultat attendu afin de minimiser les erreurs (en diminuant le justement « poids » de ceux qui y concourt, et en augmentant celui de ceux qui contribuent au résultat attendu). Ces poids forment les *paramètres* internes du système.

Jusqu'aux début des années 2000, ces systèmes comportaient le plus souvent une « couche » de neurones simulés, intercalée entre la couche d'entrée, celle qui reçoit les données (l'image d'une adresse postale *manuscrite* par exemple), et la couche de sortie, celle qui produit le résultat attendu (la reconnaissance des caractères alphanumériques afin d'automatiser la distribution du courrier par exemple). Mais les chercheurs avaient la forte intuition qu'avec beaucoup plus de couches, c'est-à-dire des réseaux plus « *profonds* », leurs performances s'amélioreraient considérablement. Personne toutefois n'en était sûr, faute de disposer d'ordinateurs assez puissants pour traiter des algorithmes en cascade.

L'APPRENTISSAGE PROFOND

L'apprentissage profond (*deep learning)* est un sous-domaine de l'apprentissage automatique (*machine learning*). L'invention des GPU (*Graphics Processing Unit*) a catalysé la révolution de l'apprentissage profond. Les GPU sont des processeurs de traitement graphique capables d'exécuter des millions d'instructions par seconde. C'est cette puissance inédite de calcul qui a rendu possible la création des réseaux de neurones « *profonds* » (*deep neural networks*), c'est-à-dire possédant un grand nombre de « couches » de neurones simulés.

L'*apprentissage profond* se distingue de l'*apprentissage automatique classique* par le recours à ces réseaux de neurones plus grands (plusieurs dizaines de couches). Cette différence les rend en effet beaucoup plus performant, en leur conférant :

- **la capacité à apprendre automatiquement des caractéristiques à partir des données brutes**. Pour obtenir les meilleurs performances, l'apprentissage automatique nécessite des données « nettoyées », c'est-à-dire dont on a extrait les caractéristiques pertinentes. L'*apprentissage automatique* suppose ainsi un travail assez long d'[1] « ingénierie des caractéristiques » accompli par des ingénieurs qui préparent les données. Les algorithmes de l'*apprentissage profond* lui donnent la capacité d'apprendre automatiquement des caractéristiques à partir des données brutes, il requiert moins d'interventions humaines (mais il y en a toujours !).

- **la capacité à traiter des quantités massives de données**. L'*apprentissage automatique* est bien adapté pour des ensembles de données de taille modérée (quelques milliers à des centaines de milliers d'exemples). L'*apprentissage profond* excelle avec des volumes de données beaucoup plus importants : des millions à des milliards d'exemples.

[1] Cf. l'article « **Ingénierie des caractéristiques » sur Wikipédia.**

- **la capacité à modéliser les relations très complexes,** dites « **non linéaires** »***** entre des données en très grand nombre.

> ***** Petit rappel de mathématiques de collège. **« Non linéaire » ?** Intuitivement qui ne suit pas une progression simple et droite.
>
> Une *relation linéaire* est une relation directe et proportionnelle entre deux variables. Sa représentation graphique sera une droite. Par exemple, si une voiture roulait à vitesse constante (disons 30 km/h), la distance parcourue (y) serait proportionnelle au temps (x) et sa représentation serait celle d'une droite : $y = 30x$. [pour simplifier on suppose ici que (x) est exprimé en heure et (y) en km]. Si on double le temps (x) du trajet, la distance parcourue (y) doublera aussi. C'est simple, direct, proportionnel.
>
> Une relation *non linéaire* entre des variables signifie que le changement de l'une ne provoque pas un changement proportionnel de l'autre. Sa représentation graphique sera une courbe ou une forme plus complexe - pas une ligne droite. Par exemple, si le véhicule accélère, la distance parcourue peut augmenter de manière quadratique (être élevée au carré). Elle sera alors représentée par une équation du type $y = x^2$. Si elle ralentit, puis accélère de nouveau, la forme de la courbe sera encore plus complexe.

Pourquoi est-ce décisif ? Parce que le monde est complexe. Les données du monde réel ne suivent pas des relations simples et directes. Par exemple, en reconnaissance d'images, la relation entre les pixels d'une image et le fait de reconnaître un visage est extrêmement complexe parce qu'il n'y a pas de relation linéaire simple entre les valeurs des pixels et la présence ou l'absence d'un visage sur une image. La croissance d'une plante est rarement proportionnelle au temps : une plante peut pousser lentement au début, puis rapidement à un certain stade, puis ralentir à nouveau. En tant qu'enseignant·e·s, nous savons que lorsque les élèves apprennent une nouvelle compétence, leurs progrès sont souvent rapides au début, puis ils deviennent plus lents à mesure qu'ils atteignent un niveau avancé ; l'apprentissage peut même passer par une phase de régression… avant de repartir de plus belle. Etc.

Les réseaux de neurones profonds ont la capacité de « capturer » automatiquement ces relations complexes, ce qui les rend très puissants pour des tâches elles-mêmes complexes et impliquant des données en très grand nombre, comme la vision par ordinateur, la reconnaissance vocale, et le traitement du langage naturel.

L'IA GÉNÉRATIVE

Les IA génératives de textes comme ChatGPT reposent sur un modèle particulier de réseaux de neurones, le « transformeur* ». Ce type de modèle peut garder en mémoire les relations entre des mots qui sont éloignés les uns des autres dans une « séquence » (une phrase, un paragraphe). C'est crucial pour comprendre le contexte global de la séquence.

Par exemple, pour comprendre la phrase suivante :

> *Le chien de la belle-sœur de Marie a couru vers la maison parce qu'il avait entendu son maître l'appeler.*

il faut comprendre « les dépendances à long terme » entre les mots.

- le sujet principal de la phrase est "*Le chien*"
- le pronom "*il*", plus loin dans la phrase, fait référence à "*Le chien*", mentionné au début de la phrase.
- de même le pronom « *l'* », dans « l'*appeler* », réfère à « *Le chien* » et non à « *la belle-sœur de Marie* ». Le chien accourt non pas parce qu'il aurait « *entendu son maître* » « *l'appeler* », elle, « *la belle-sœur de Marie* », mais bien lui, « *Le chien* ». C'est l'interprétation la plus plausible d'après le contexte.

Pour traduire, reformuler, ou expliquer une phrase, le modèle doit saisir le contexte, donc garder en mémoire les informations données au début d'une phrase (ou un texte), afin de pouvoir « capturer les dépendances à long terme » entre tous ses mots.

Le talon d'Achille des modèles antérieurs est qu'ils sont *séquentiels*. Ils traitent le texte *mot par mot*. Et, surtout, leur mémoire des mots précédents est très limitée. Dès qu'une phrase ou un paragraphe s'allonge, ces modèles perdent des informations contextuelles essentielles à la compréhension. Ils sont un peu comme une personne qui oublierait ce qu'elle vient de lire 10 lignes plus haut ou ce qu'elle vient d'entendre 1 minute plus tôt. Ce qui limite considérablement les performances de ces modèles.

Les *transformeurs* n'opèrent pas de façon séquentielle. Ils disposent d'un « mécanisme d'attention » qui leur permet de « regarder » *simultanément* tous les mots d'une phrase, d'un paragraphe ou d'un texte. Ce qui les rend capables de capturer des relations entre des mots très éloignés (les fameuses « dépendances à long terme ») et de comprendre des phrases complexes. Les transformeurs dépassent de très loin les modèles antérieurs. Ils produisent des textes de si haute qualité que l'on pourrait les croire écrits par un être humain.

Les transformeurs sont aussi plus rapides et plus efficace. Le traitement séquentiel empêche, en effet, les modèles antérieurs de tirer profit de la puissance des processeurs modernes capables d'exécuter simultanément plusieurs tâches, c'est-à-dire de faire du traitement parallèle. Les transformateurs, au contraire, exploitent directement cette puissance, car ils traitent *simultanément* des séquences entières. Ce qui les rend beaucoup plus rapides et capables de gérer des séquences beaucoup plus longues. C'est bien ce qui avait frappé les premiers utilisateurs de ChatGPT : sa capacité à résumer un article, à écrire une dissertation, à rédiger un mémoire ou un livre en quelques secondes avec une qualité plutôt bonne.

* D'où vient ce nom de « transformeur » ? Il fait référence à un modèle particulier de réseaux de neurones appelé « Transformeur Génératif Pré-entraîné », en anglais GPT pour *Generative Pre-trained Transformers*.

- « ***Génératif*** » parce que ce modèle est capable de créer (générer) des contenus nouveaux.

- « ***Pré-entrainé*** » parce qu'il est formé (« entraîné ») sur des quantités énormes de données.

- « ***Transformeur*** » par référence à leur architecture logicielle qui leur permet de « transformer » efficacement des séquences de données (textes, paroles ou même images).

D'où le nom du produit d'*Open AI* : **Chat-GPT** : transformateur pré-entraîné générateur (GPT) de « conversation » (Chat).

Serge Cospérec

Pour une introduction à l'intelligence artificielle et aux problèmes qu'elle soulève, cf. ma série d'articles et en cours de publications, écrits à partir des travaux de la chercheuse américaine Melanie Mitchell (et avec son accord !) sur mon blog *Essais Critiques* : https://essais-critiques.fr/

L'ACIREPh, association indépendante, ne vit que grâce au soutien de ses adhérents.

Adhérez ou réadhérez à l'ACIREPH !

⇒ L'adhésion en ligne est recommandée : acireph.org

⇒ à défaut merci de nous retourner le bulletin ci-dessous :

---------- ✂ ✂ ✂ ----------

BULLETIN D'ADHÉSION À L'ACIREPh

□ J'adhère ou ré-adhère à l'ACIREPh pour l'année 2024-2025, je paye (cochez la case) au choix :

- □ 20 € (étudiant et/ou non-imposable)
- □ 30 € (après réduction fiscale, ma cotisation revient à 10 €)
- □ 50 € (après réduction fiscale, ma cotisation revient à 17 €)
- □ 75 € (après réduction fiscale, ma cotisation revient à 26 €)

La cotisation peut donner lieu à une réduction d'impôt correspondant à 66 % de son montant (CGI art.200, 1-b).

Nom…… Prénom……..

Adresse……

Code Postal :……..Ville : ……..……

e-mail :……..…… Tél :.……

Souhaitez-vous recevoir la version papier du bulletin, par courrier postal ?

oui □ non □

Date : **Signature :**

Bulletin d'adhésion et chèque bancaire ou postal (libellé à l'ordre de l'ACIREPh) à adresser à :

ACIREPh

21 rue du Général Faidherbe,

Bâtiment A

94130 Nogent-sur-Marne

Au fil des numéros, Côté Philo aborde divers aspects de la culture et du métier de professeur de philosophie ; le journal constitue ainsi un instrument d'information et de réflexion régulièrement alimenté et renouvelé. Selon les livraisons, nous proposons ainsi :

- Des *Dossiers* sur des questions intéressant l'enseignement de la philosophie
- Des *Notes de lecture* à vocation pédagogique
- Des synthèses sur un champ ou un philosophe
- Des pratiques pédagogiques
- Des articles sur l'enseignement de la philosophie à l'étranger
- Des informations institutionnelles et l'éclairage qu'elles nécessitent
- Ainsi que des *Humeurs* qui parfois s'imposent…

ISSN 2971-3773