Les Cahiers du CEDIMES
2026, Volume 21, numéro 1, https://doi.org/10.69611/cahiers21-1-06
ISSN 2110-6045

Intelligence artificielle : fondements historiques, origines conceptuelles et état de l'art

Petraq PAPAJORGJI[1*], Elona SHEHU[2], Adrian CIVICI [3], Orkida ILOLLARI[4], Florjan BOMBAJ[5]

[1] Université Méditerranéenne d'Albanie, petraq@gmail.com, ORCID : 0000-0002-3833-5215

[2] Université Méditerranéenne d'Albanie, elonashahini@umsh.edu.al, ORCID : 0000-0002-8871-6382

[3] Université Méditerranéenne d'Albanie, a.civici@umsh.edu.al, ORCID : 0000-0002-4105-3016

[4] Université Méditerranéenne d'Albanie, orkida.ilollari@umsh.edu.al, ORCID : 0000-0002-7169-9256

[5] Université Méditerranéenne d'Albanie, florjan.bombaj@umsh.edu.al, ORCID : 0000-0002-0179-9817

* auteur correspondant

Résumé : L'intelligence artificielle (IA) s'est imposée comme un domaine transformateur à la croisée de l'informatique, des mathématiques et de la philosophie. Cet article examine les fondements historiques et conceptuels de l'IA, retraçant ses origines depuis la logique classique de la philosophie antique jusqu'aux premiers calculs mécaniques et au développement des machines programmables. Il présente un compte rendu détaillé de la conférence de Dartmouth de 1956, où l'IA a été formellement établie comme discipline scientifique, en expliquant notamment la justification du terme « intelligence artificielle » et les principaux problèmes de recherche identifiés par ses fondateurs. L'étude analyse ensuite l'évolution de l'IA, des approches symboliques aux paradigmes modernes d'apprentissage automatique et d'apprentissage profond, en soulignant les principaux changements technologiques et méthodologiques. De plus, l'article aborde les défis contemporains, tels que les biais algorithmiques, l'explicabilité et la gouvernance, ainsi que le rôle croissant de l'IA dans l'éducation. En intégrant l'analyse historique aux développements actuels, cet article offre une perspective globale sur la trajectoire de l'IA et ses implications pour la recherche et les politiques futures.

Mots-clés : Intelligence artificielle ; conférence de Dartmouth ; raisonnement symbolique ; apprentissage automatique ; éthique de l'IA.

Summary: Artificial Intelligence (AI) has emerged as a transformative field at the intersection of computer science, mathematics, and philosophy. This paper examines the historical and conceptual foundations of AI, tracing its origins from classical logic in ancient philosophy to early mechanical computation and the development of programmable machines. It provides a detailed account of the 1956 Dartmouth Conference, where AI was formally established as a scientific discipline, including the rationale behind the term "Artificial Intelligence" and the core research problems identified by its founders. The study further analyzes the evolution of AI from symbolic approaches to modern machine learning and deep learning paradigms, highlighting key technological and methodological shifts. In addition, the paper addresses contemporary challenges, including algorithmic bias, explainability, and governance, as well as the growing role of AI in education. By integrating historical analysis with current developments, the paper offers a comprehensive perspective on AI's trajectory and its implications for future research and policy.

Keywords: Artificial Intelligence; Dartmouth conference; Symbolic reasoning; Machine Learning; AI Ethics.

Classification JEL : C63 ; C88 ; O33 ; I23 ; D83.

1. Introduction

L'intelligence artificielle résulte d'une analyse approfondie du cerveau humain et des processus cognitifs, et non d'une simple amélioration technologique. Elle a considérablement progressé grâce à la reconnaissance de schémas cérébraux et à l'apprentissage de son fonctionnement. Le développement de logiciels et de systèmes intelligents a été rendu possible par ces recherches.

Le domaine de la psychologie connu sous le nom de psychologie cognitive examine comment la pensée, les émotions, la créativité et les compétences en résolution de problèmes interagissent pour influencer la façon dont les gens pensent et pourquoi ils pensent comme ils le font (Solso et al., 2005). L'alliance fascinante et avantageuse de la technologie et de la psychologie cognitive a jeté les bases de futures percées scientifiques. L'histoire du développement technologique démontre le lien étroit entre les capacités cognitives humaines et l'évolution de la technologie. L'idée que les humains façonnent les technologies, qui à leur tour les façonnent, est illustrée par l'histoire de l'humanité (Osiurak et al., 2018). Les auteurs proposent un paradigme pour comprendre comment la cognition humaine influence la technologie et est influencée par elle. Ils distinguent trois étapes – passé, présent et futur – dans la relation entre technologie et capacités cognitives.

L'utilisation d'outils concrets caractérisait la première étape, appelée le Passé. Les outils physiques visaient avant tout à améliorer nos capacités sensori-motrices (Virgo et al., 2017). Bien que ces outils soient similaires aux premiers outils fabriqués et utilisés par les humains à la préhistoire, ils sont aujourd'hui largement employés. Il est essentiel de se rappeler que l'utilisation de tout instrument physique requiert une compréhension cognitive de sa géométrie et de ses caractéristiques physiques. Les caractéristiques des premiers outils en pierre démontrent que leurs créateurs maîtrisaient le processus nécessaire à leur fabrication (Hovers, 2012).

La seconde phase, le présent, se caractérise par l'utilisation d'outils sophistiqués. De nos jours, nous utilisons des outils avancés pour accomplir nos activités quotidiennes. Envoyer des messages depuis un appareil mobile ou prendre le train pour traverser la ville en sont deux excellents exemples. Il est fréquent que l'utilisateur ait besoin de temps et parfois d'une formation spécialisée pour se familiariser avec des systèmes complexes. Il existe un fossé important entre le développement et l'utilisation d'outils cognitifs de plus en plus complexes (tels que les tableurs).

La troisième phase, celle du futur, se caractérise par l'utilisation d'outils symbiotiques. Il est difficile de prédire quels types d'outils seront disponibles à l'avenir. Les films de science-fiction les mettent souvent en scène. Les interfaces cerveau-ordinateur (ICO), qui visent à traduire l'activité cérébrale (ou « pensées ») en commandes compréhensibles par une machine, constituent une catégorie d'outils émergents qui ont suscité un vif intérêt. Un algorithme intelligent est utilisé pour réaliser ce processus ; des capteurs intelligents captent et analysent l'activité cérébrale avant de la corréler à l'action appropriée exécutée par le système artificiel (CM Bishop, 2007). Il a été démontré que cet algorithme peut apprendre à différencier les classes dans les signaux cérébraux enregistrés. La flexibilité d'une interaction BCI intelligente est ce qui lui confère sa véritable puissance. Elle comporte une phase d'apprentissage qui permet à la technologie de s'adapter au système cognitif particulier de chaque utilisateur (Papajorgji et Moskowitz, 2025). Les particularités de notre cerveau et les variations fonctionnelles qui nous caractérisent ne constituent plus des obstacles, mais deviennent des moteurs qui permettent aux algorithmes d'apprentissage de s'adapter aux caractéristiques uniques de chaque individu. Cette flexibilité démontre comment les interfaces cerveau-machine (ICM) sont capables de « sembler comprendre » et d'améliorer chacune de nos expériences uniques.

Le rôle des individus dans le processus d'explication a fasciné les chercheurs en philosophie, en psychologie et en sciences cognitives. Ils ont étudié comment les individus reconnaissent, produisent, sélectionnent, évaluent et fournissent des explications. Ils ont découvert que les biais cognitifs et les attentes sociales jouent un rôle important dans ce processus (Miller, 2019).

2. Aristote et les fondements de la logique formelle

L'existence Le chemin vers l'intelligence artificielle a commencé avec la quête antique visant à comprendre le raisonnement humain et la nature de la connaissance. Des philosophes comme Aristote ont posé les fondements de cette compréhension en formalisant la logique comme un système de raisonnement déductif, qui devint par la suite fondamental pour le développement de la logique computationnelle (Feibleman, 1979).

Aristote a développé la première approche systématique du raisonnement, connue sous le nom de logique syllogistique, qui permet de tirer des conclusions à partir de prémisses grâce à des règles structurées (Aristote, trad. 1984).

Un syllogisme classique illustre cette structure :

- ✓ Tous les êtres humains sont mortels.
- ✓ Socrate est un être humain.
- ✓ Par conséquent, Socrate est mortel.

Ce cadre théorique a introduit l'idée que le raisonnement n'est pas arbitraire mais obéit à des règles formelles. Les contributions d'Aristote sont fondamentales pour l'IA de trois manières principales : premièrement, il a introduit le concept de représentation des connaissances, permettant d'exprimer symboliquement les faits concernant le monde ; deuxièmement, il a démontré l'inférence fondée sur des règles, permettant de tirer des conclusions de manière systématique ; troisièmement, il a établi que le raisonnement peut être formalisé et reproduit, un principe central des modèles computationnels de l'intelligence.

Les systèmes d'IA modernes, en particulier l'IA symbolique et les systèmes experts, s'appuient directement sur ces idées (Russell & Norvig, 2021).

3. Calcul mécanique : du calcul à l'automatisation

3.1. Blaise Pascal et la Pascaline

En 1642, Blaise Pascal inventa la Pascaline, une des premières calculatrices mécaniques conçue pour effectuer des additions et des soustractions (Ifrah, 2001). L'appareil utilisait un système d'engrenages et de roues pour automatiser les opérations arithmétiques.

L'importance de la Pascaline réside moins dans sa puissance de calcul que dans ses implications conceptuelles. Elle a démontré que des tâches cognitives telles que le calcul pouvait être mécanisées. Cela a marqué un tournant, passant du calcul mental humain aux processus assistés par machine.

L'invention de Pascal a introduit le principe selon lequel les machines peuvent exécuter des procédures déterministes, un concept qui est devenu par la suite fondamental pour la pensée algorithmique et l'IA.

3.2. Gottfried Wilhelm Leibniz et le raisonnement symbolique

Gottfried Wilhelm Leibniz a approfondi les travaux de Pascal en développant l'ordinateur à degrés et en introduisant le système binaire (Leibniz, 1976).

La représentation binaire est le fondement de l'informatique moderne, permettant d'encoder toute information à l'aide de deux symboles : 0 et 1. Cette innovation a rendu possible la conception de systèmes numériques capables de traiter des données complexes.

Leibniz proposa également un langage symbolique universel, connu sous le nom de Characteristica Universalis, et une méthode de raisonnement appelée Calculus Ratiocinator. Il envisageait un avenir où les différends logiques pourraient être résolus par le calcul plutôt que par le débat.

Cette idée anticipait un objectif central de l'IA : l'automatisation du raisonnement. En

proposant que la pensée elle-même puisse être exprimée symboliquement et manipulée mécaniquement, Leibniz a jeté les bases de l'IA symbolique (Nilsson, 2010).

4. Les premières machines programmables

4.1. Charles Babbage et la machine analytique

Au XIXe siècle, Charles Babbage conçut la machine analytique, largement considérée comme la première conception d'un ordinateur programmable à usage général (Swade, 2001). Bien que jamais entièrement construite, cette machine a introduit une architecture conceptuelle qui préfigure de près les systèmes informatiques modernes.

Le moteur analytique intégrait plusieurs composants fondamentaux :

- ✓ Une unité de traitement (« moulin »), chargée d'effectuer des opérations arithmétiques
- ✓ Une unité de mémoire (« mémoire »), conçue pour stocker des nombres et des résultats intermédiaires.
- ✓ Mécanismes d'entrée par cartes perforées, inspiré du métier à tisser de Joseph Marie Jacquard, permettant d'encoder les instructions en externe
- ✓ Prise en charge des branchements conditionnels et des boucles, permettant à la machine de modifier le flux d'exécution en fonction des résultats intermédiaires.

Ces caractéristiques reflètent collectivement les éléments fondamentaux de l'architecture informatique contemporaine, notamment la séparation du traitement, de la mémoire et du contrôle.

La véritable importance de la machine de Babbage réside dans l'introduction de la programmabilité. Contrairement aux calculatrices mécaniques antérieures, limitées à des opérations fixes, la machine analytique pouvait exécuter une vaste gamme de tâches en fonction des instructions fournies. Cela a marqué un tournant conceptuel fondamental : le calcul n'était plus cantonné à un seul objectif, mais pouvait être généralisé grâce à des instructions symboliques.

Tout aussi importante était la séparation implicite entre matériel et logiciel. La machine elle-même (matériel) restait constante, tandis que les cartes perforées (logiciel) déterminaient sa fonction. Cette abstraction est fondamentale pour l'informatique moderne et sous-tend les systèmes d'IA contemporains, où une même infrastructure informatique peut prendre en charge des modèles et des applications très différents.

4.2. Ada Lovelace et le premier algorithme

Ada Lovelace a travaillé en étroite collaboration avec Babbage et on lui attribue la création du premier algorithme destiné à être exécuté par machine (Lovelace, 1989).

Sa contribution la plus importante fut conceptuelle. Lovelace reconnut que les machines pouvaient manipuler des symboles, et pas seulement des nombres. Elle suggéra qu'une machine pourrait composer de la musique ou traiter le langage si les règles régissant ces domaines étaient correctement codées.

Cette intuition est fondamentale pour l'IA. Elle a introduit l'idée que le calcul ne se limite pas à l'arithmétique, mais englobe toutes les formes de traitement symbolique. La vision de Lovelace a anticipé des applications modernes telles que le traitement automatique du langage naturel et l'IA générative (Boden, 2016).

5. La conférence de Dartmouth : la naissance de l'IA

On attribue souvent la naissance officielle de l'IA en tant que discipline de recherche à la conférence de Dartmouth de 1956, organisée par John McCarthy, Marvin Minsky, Nathaniel Rochester et Claude Shannon (Kaplan, 2022). Cet événement fondateur a réuni d'éminents scientifiques pour discuter de la possibilité de créer des machines capables de simuler un comportement intelligent. La

conférence a défini les orientations de la recherche en IA en proposant des problèmes et des approches clés, notamment l'apprentissage automatique, le traitement automatique du langage naturel et le raisonnement automatisé. Dès lors, l'IA a connu une évolution rapide, portée par les avancées théoriques et technologiques (Crevier, 1993).

De nombreux chercheurs de renom, parmi lesquels Ray Solomonoff, Oliver Selfridge, Trenchard More, Arthur Samuel, Allen Newell et Herbert A. Simon, étaient présents à la réunion. Tous ont joué un rôle crucial dans les premières décennies de la recherche en intelligence artificielle (Crevier, 1993). Newell et Simon ont plaidé, lors de l'atelier, pour que le terme « théoricien de la logique » soit utilisé pour désigner ce nouveau domaine. Cliff Shaw, Herbert A. Simon et Allen Newell ont créé le logiciel Logic Theorist en 1956. Ce logiciel est considéré comme le premier programme d'intelligence artificielle, car il fut le premier logiciel spécifiquement conçu pour effectuer un raisonnement automatique (Crevier, 1993).

38 des 52 premiers théorèmes du célèbre ouvrage Principia Mathematica, écrit par Alfred Whitehead et Bertrand Russell, ont été démontrés par des théoriciens de la logique, qui ont également découvert des démonstrations nouvelles et plus élégantes pour certains de ces théorèmes (Whitehead & Russell, 2004).

En examinant ces différentes phases – la recherche philosophique, la logique formelle, le calcul mécanique et la conférence de Dartmouth – cet article vise à offrir une compréhension globale de la manière dont l'IA s'est développée en un domaine scientifique solide. De plus, il met en lumière pourquoi nombre de questions fondamentales concernant l'intelligence, le raisonnement et les capacités des machines demeurent pertinentes à mesure que l'IA progresse et s'intègre à la société.

5.1. Pourquoi le terme « intelligence artificielle » a-t-il été choisi ?

Le terme « intelligence artificielle » a été inventé par John McCarthy. Il a été choisi délibérément pour refléter un programme de recherche vaste et ambitieux.

Premièrement, ce terme mettait l'accent sur l'intelligence générale plutôt que sur des fonctionnalités spécifiques. Deuxièmement, il distinguait ce domaine de la cybernétique, qui se concentrait principalement sur les systèmes de contrôle et les boucles de rétroaction (Wiener, 1948).

Enfin, ce terme reflétait une hypothèse audacieuse : celle que l'intelligence pouvait être décrite avec suffisamment de précision pour être simulée par des machines.

6. Évolution de l'intelligence artificielle : de l'IA symbolique à l'apprentissage automatique

6.1. Concept d'apprentissage automatique

L'apprentissage automatique (AA) est un sous-domaine fondamental de l'intelligence artificielle qui se concentre sur le développement d'algorithmes capables d'apprendre des modèles à partir de données et d'améliorer leurs performances au fil du temps sans être explicitement programmés pour chaque tâche. Contrairement aux systèmes traditionnels à base de règles, qui reposent sur des instructions logiques prédéfinies, les systèmes d'apprentissage automatique utilisent des techniques statistiques pour identifier les relations au sein des données et effectuer des prédictions ou prendre des décisions en fonction de ces modèles (Russell & Norvig, 2021).

L'apprentissage automatique repose essentiellement sur trois composantes fondamentales : les données, les modèles et les algorithmes d'apprentissage. Les données constituent l'entrée à partir de laquelle le système extrait des connaissances. Le modèle représente la structure mathématique utilisée pour identifier les tendances, tandis que l'algorithme d'apprentissage ajuste les paramètres du modèle afin de minimiser les erreurs et d'améliorer la précision. Ce processus est souvent itératif, ce qui signifie que le système affine continuellement ses prédictions à mesure que de nouvelles données sont disponibles.

L'apprentissage automatique se divise généralement en trois grandes catégories : l'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage par renforcement. En apprentissage supervisé, les modèles sont entraînés sur des ensembles de données étiquetées, où la sortie correcte

est connue, ce qui permet au système d'apprendre les correspondances entre les entrées et les sorties. En apprentissage non supervisé, le système identifie des structures ou des motifs cachés dans des données non étiquetées, comme le clustering ou la réduction de dimensionnalité. L'apprentissage par renforcement, quant à lui, repose sur l'interaction avec un environnement, où le système reçoit un retour d'information sous forme de récompenses ou de sanctions.

L'importance de l'apprentissage automatique réside dans sa capacité à traiter des données complexes et multidimensionnelles et à s'adapter aux nouvelles informations. Il a permis des avancées majeures dans des domaines tels que le traitement automatique du langage naturel, la vision par ordinateur et l'analyse prédictive. En déplaçant l'attention de la programmation explicite vers un apprentissage piloté par les données, l'apprentissage automatique représente un changement de paradigme fondamental en intelligence artificielle, rapprochant les systèmes de l'imitation de certains aspects de l'apprentissage et de la prise de décision humains (LeCun et al., 2015).

Suite à la conférence de Dartmouth, les premières recherches en IA se sont concentrées principalement sur les approches symboliques, également connues sous le nom d'IA classique (GOFAI). Ces systèmes reposaient sur des règles explicitement programmées et un raisonnement logique (Nilsson, 2010).

Cependant, leurs limites sont rapidement apparues. Les systèmes symboliques peinaient à gérer l'incertitude, l'ambiguïté et le traitement de données réelles à grande échelle. Ceci a conduit au développement de l'apprentissage automatique, où les systèmes apprennent des modèles à partir des données plutôt que de se fier uniquement à des règles prédéfinies (Russell & Norvig, 2021).

L'apprentissage automatique a introduit des méthodes statistiques et un raisonnement probabiliste, permettant aux systèmes d'IA d'améliorer leurs performances au fil du temps. Cela a marqué un passage d'une logique déterministe à une intelligence fondée sur les données.

6.2. L'apprentissage profond et la révolution moderne de l'IA

Au XXIe siècle, l'IA a connu une avancée majeure avec l'essor de l'apprentissage profond, un sous-ensemble de l'apprentissage automatique basé sur des réseaux neuronaux artificiels (LeCun, Bengio et Hinton, 2015).

L'apprentissage profond représente un paradigme dominant de l'intelligence artificielle contemporaine. Il se caractérise par l'utilisation de réseaux de neurones artificiels multicouches pour modéliser des relations complexes et non linéaires dans des ensembles de données à grande échelle. Ces architectures, notamment les réseaux convolutionnels, récurrents et de type transformeur, permettent un apprentissage hiérarchique des représentations, où les entrées brutes sont progressivement transformées en abstractions de plus haut niveau. Cette capacité d'extraction automatique de caractéristiques distingue l'apprentissage profond des approches d'apprentissage automatique traditionnelles qui reposent sur une ingénierie manuelle des caractéristiques. Le développement rapide de l'apprentissage profond a été impulsé par la convergence de vastes ensembles de données, l'augmentation de la puissance de calcul et les innovations algorithmiques, conduisant à des améliorations significatives dans des domaines tels que la vision par ordinateur, la reconnaissance vocale et le traitement automatique du langage naturel.(Alzubaidi et al., 2021 ; Khan et al., 2023)

Un élément méthodologique central de l'apprentissage profond est l'utilisation de l'optimisation par gradient, notamment la rétropropagation, qui ajuste itérativement les paramètres du modèle afin de minimiser l'erreur de prédiction. Les innovations architecturales ont joué un rôle crucial dans l'élargissement du champ d'application de l'apprentissage profond : les réseaux de neurones convolutifs (CNN) excellent dans le traitement des données spatiales, tandis que les modèles de type Transformer permettent des avancées majeures dans la modélisation de séquences et la compréhension du langage. Des développements plus récents, tels que l'apprentissage par renforcement profond, intègrent l'apprentissage de représentations aux capacités de prise de décision, permettant ainsi aux systèmes de fonctionner efficacement dans des environnements dynamiques et incertains (Schmidhuber, 2022 ; Wang et al., 2021). Ces progrès témoignent d'une évolution plus large vers des systèmes d'apprentissage de bout en bout capables de généraliser des tâches complexes.

Malgré ses succès, l'apprentissage profond se heurte à plusieurs défis persistants qui constituent des axes de recherche actifs. Parmi ceux-ci figurent des exigences élevées en matière de calcul et d'énergie, la dépendance à de vastes ensembles de données étiquetées et l'interprétabilité limitée des modèles appris. De plus, les préoccupations liées à la robustesse, aux biais et aux implications éthiques sont devenues de plus en plus prégnantes à mesure que les systèmes d'apprentissage profond sont déployés dans des applications concrètes. Des recherches récentes soulignent l'importance de développer des paradigmes d'apprentissage plus efficaces, tels que l'apprentissage auto-supervisé et l'apprentissage actif, ainsi que d'améliorer l'explicabilité et l'équité des systèmes d'IA (Goyal et al., 2022 ; Li et al., 2023).

Les algorithmes d'apprentissage profond ont démontré des performances exceptionnelles dans diverses tâches complexes de type cognitif. Ils peuvent désormais reconnaître des images et la parole avec une grande précision, ce qui les rend utiles pour des applications telles que le diagnostic médical par imagerie, la reconnaissance faciale et les assistants vocaux en temps réel. De plus, ils facilitent la compréhension du langage naturel, permettant à des modèles comme GPT et BERT d'interpréter, de produire et de traduire le langage humain avec une sophistication croissante. Enfin, l'apprentissage profond permet une prise de décision autonome, notamment dans des contextes structurés comme le trading algorithmique, les systèmes de recommandation et les véhicules autonomes, où les ordinateurs peuvent agir selon des schémas appris sans intervention humaine explicite.

Ces avancées sont le fruit de la convergence de trois facteurs fondamentaux. Premièrement, la disponibilité de vastes ensembles de données (« big data ») a fourni la matière première nécessaire à l'entraînement des réseaux neuronaux profonds, permettant aux modèles d'apprendre des schémas complexes à partir d'une grande variété d'entrées. Deuxièmement, l'amélioration des capacités de calcul, notamment grâce aux technologies de traitement parallèle comme les GPU NVIDIA, a rendu possible l'entraînement de structures profondes comportant des millions, voire des milliards de paramètres. Troisièmement, les progrès algorithmiques, incluant les découvertes dans des topologies telles que les réseaux neuronaux convolutifs et les modèles de type Transformer, ont considérablement amélioré l'efficacité, l'évolutivité et les performances des systèmes d'apprentissage profond. Ensemble, ces éléments ont transformé l'apprentissage profond, d'une notion théorique à un paradigme dominant de l'intelligence artificielle, sous-tendant nombre des systèmes les plus avancés d'aujourd'hui.

7. Réseaux neuronaux

Les réseaux de neurones constituent un paradigme clé de l'intelligence artificielle moderne, permettant aux machines de découvrir des structures complexes à partir des données. Ces modèles, inspirés de l'organisation du cerveau humain, sont composés d'unités informatiques interconnectées qui utilisent un traitement par couches pour transformer les entrées en sorties. Leur capacité à estimer des relations extrêmement non linéaires les rend essentiels aux progrès dans des disciplines telles que la vision par ordinateur, le traitement automatique du langage naturel et les systèmes autonomes. Selon l'inventeur de l'un des premiers neuroordinateurs, un réseau de neurones peut être défini comme : « un système informatique comprenant plusieurs éléments de traitement simples et fortement interconnectés qui traitent l'information en fonction de leur réponse dynamique aux entrées externes » (Sharma et al., 2020).

1. Concept et structure. Les réseaux neuronaux artificiels sont constitués de couches de nœuds interconnectés, ou « neurones », comme illustré dans la figure 1.

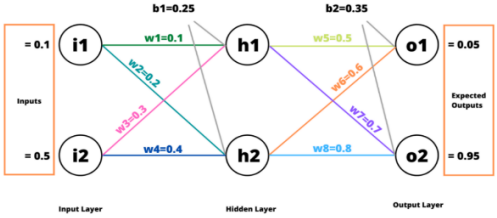
Figure 1 : Structure générale des ANN

Source : élaboré par les auteurs.

Chaque neurone reçoit une entrée, effectue une modification pondérée, puis transmet la sortie via une fonction d'activation non linéaire. L'architecture la plus simple comprend trois couches : une couche d'entrée (qui reçoit les données brutes), une ou plusieurs couches cachées (qui extraient les caractéristiques) et une couche de sortie (qui effectue les prédictions). L'empilement de nombreuses couches cachées permet aux réseaux d'apprendre des représentations hiérarchiques, les couches supérieures capturant des caractéristiques de plus en plus abstraites.

2. Processus d'apprentissage. Les réseaux de neurones apprennent par l'entraînement, qui consiste à ajuster les paramètres internes du modèle (poids) afin de minimiser l'écart entre les sorties prédites et les sorties réelles. Ceci est souvent réalisé grâce à la rétropropagation, une méthode qui calcule les gradients d'erreur par rapport à chaque poids, combinée à des techniques d'optimisation comme la descente de gradient. Au fil des itérations et sur de vastes ensembles de données, les performances du réseau s'améliorent continuellement. La figure 2 illustre un exemple de réseau de neurones.

Figure 2 : Exemple de réseau neuronal



Source : (Lamsal, 2021)

3. Différents types de réseaux neuronaux. Différentes tâches requièrent différentes architectures. Les réseaux à propagation directe sont fréquemment utilisés pour la régression et la classification car ils traitent les données de manière unidirectionnelle. Les réseaux de neurones convolutifs utilisent des filtres pour identifier des motifs tels que les contours et les textures dans les données d'images et spatiales. Les réseaux de neurones récurrents utilisent des connexions de rétroaction pour traiter les entrées séquentielles, comme le langage ou les séries temporelles. Grâce à leur capacité à capturer les dépendances à long terme, les modèles de type Transformer sont devenus prédominants dans le traitement du langage naturel.

4. Avantages et utilisations. Les réseaux de neurones sont particulièrement adaptés aux ensembles de données volumineux, complexes et de grande dimension, là où les méthodes statistiques traditionnelles rencontrent souvent des difficultés. La grande dimensionnalité – comme celle des images comportant des millions de pixels ou des textes dotés d'un vocabulaire étendu – crée des relations complexes et non linéaires que les réseaux de neurones peuvent modéliser efficacement grâce à l'extraction de caractéristiques par couches et à l'apprentissage de représentations. Cette capacité leur a permis d'atteindre des performances de pointe dans de nombreux domaines, surpassant souvent la précision humaine dans des tâches bien définies (Ramachandran et al., 2017).

En reconnaissance d'images et de la parole, les réseaux neuronaux, notamment les réseaux neuronaux convolutifs, apprennent automatiquement des caractéristiques hiérarchiques à partir de données brutes. Par exemple, en analyse d'images, les couches inférieures détectent les contours et les textures, tandis que les couches profondes reconnaissent les objets et les motifs. De même, en

reconnaissance vocale, les modèles neuronaux traitent les signaux audios pour identifier les phonèmes, les mots et le sens contextuel, alimentant ainsi des systèmes tels que les assistants vocaux et les outils de transcription en temps réel.

Dans le domaine de la compréhension et de la génération du langage naturel, les réseaux neuronaux, et plus particulièrement les modèles basés sur le transformer, ont transformé la façon dont les machines traitent le texte (Vaswani et al., 2017). Des systèmes comme GPT et BERT peuvent interpréter le contexte, générer des réponses cohérentes, traduire des langues et résumer des documents complexes. Ces capacités découlent de leur aptitude à saisir les dépendances à long terme et les relations sémantiques au sein des données linguistiques.

En bio-informatique et en diagnostic médical, les réseaux de neurones sont utilisés pour analyser des données biologiques très complexes, telles que les séquences génomiques, les images médicales et les dossiers des patients. Ils contribuent à des tâches comme la détection des maladies, la découverte de médicaments et les recommandations de traitements personnalisés. Par exemple, les modèles d'apprentissage profond peuvent identifier des tumeurs sur des images radiologiques avec une précision remarquable ou détecter des schémas dans les données génétiques pouvant indiquer une prédisposition à certaines maladies.

En matière de prévisions financières et de détection d'anomalies, les réseaux neuronaux traitent d'immenses flux de données structurées et non structurées (prix du marché, historiques de transactions et indicateurs économiques) afin d'identifier les tendances et les irrégularités. Ils sont largement utilisés pour le trading algorithmique, la détection de la fraude, l'évaluation du crédit et l'analyse des risques. Leur capacité à déceler des schémas subtils et non évidents les rend particulièrement précieux pour identifier les anomalies susceptibles de signaler des fraudes, des défaillances de systèmes ou des risques émergents.

L'adaptabilité est un atout majeur de toutes ces applications. La même architecture fondamentale – des couches de neurones interconnectés entraînés par rétropropagation – peut être adaptée à différents types de données et de tâches en ajustant la structure du réseau, les données d'entraînement et la stratégie d'optimisation. Cette flexibilité permet aux réseaux de neurones de servir de cadre d'apprentissage généraliste, déployable dans divers domaines sans nécessiter de paradigmes informatiques entièrement nouveaux.

5. Restrictions et difficultés. Les réseaux de neurones sont performants, mais présentent des inconvénients majeurs. Ils nécessitent généralement une grande quantité de données et une puissance de calcul importante. Leur fonctionnement interne est souvent opaque, ce qui soulève des problèmes de fiabilité et d'interprétabilité. Ils peuvent également avoir des difficultés à résister aux entrées hostiles, à généraliser en dehors des données d'entraînement et à établir un raisonnement causal (Papajorgji et Moskowitz, 2025).

6. Conclusion. Les réseaux neuronaux représentent un type de calcul puissant mais spécialisé. Au lieu de « comprendre » au sens humain du terme, ils identifient des liens et des schémas statistiques dans les données. Leur influence repose sur leur polyvalence et leur capacité d'adaptation, qui ont rendu possibles les progrès de l'IA. Concilier performance, ouverture et contrôle demeure un défi majeur à mesure que leur utilisation s'étend à des domaines essentiels.

8. Le mythe anthropique et l'émergence de l'IA comme risque systémique en finance

Le 7 avril 2026, Anthropic annonçait « Claude Mythos Preview », un modèle dont la diffusion publique était, pour la première fois, interdite en raison de problèmes de sécurité. Les raisons étaient internes, et des évaluations externes limitées suggéraient que Mythos pouvait détecter de manière autonome des vulnérabilités jusque-là inconnues et élaborer des stratégies de cyberattaque en plusieurs étapes. Il réussissait avec brio des tâches de piratage de niveau expert et aurait découvert des milliers de failles logicielles critiques. Contrairement aux systèmes d'IA précédents qui analysaient principalement des données, Mythos pouvait opérationnaliser ses découvertes de manière directement applicable aux systèmes du monde réel (Reuters, 2026).

Peu après son annonce, des informations ont révélé que des utilisateurs non autorisés avaient accédé au modèle via une plateforme tierce, mettant ainsi en évidence des failles dans ses contrôles d'accès. Bien que la violation ait été limitée, son caractère opportun – survenu presque immédiatement après la présentation du modèle – a amplifié les inquiétudes concernant le confinement et la gouvernance. La combinaison de capacités élevées et d'un contrôle imparfait a transformé Mythos, d'une innovation technique, en un risque systémique perçu. Ce risque était si grave que la Banque centrale européenne a demandé aux banques d'élaborer des plans de contingence.

Suite à ce dernier événement, beaucoup s'interrogent sur l'intelligence artificielle. D'un côté, des scientifiques affirment avec force que l'IA actuelle est incapable de raisonner et d'éprouver de l'empathie comme un humain ; de l'autre, des phénomènes tels que les modèles Mythos d'Anthropic peuvent s'avérer si dangereux qu'ils mettent en péril le secteur financier mondial. En définitive, l'IA est-elle intelligente ou non ?

La bonne approche face à ce dilemme consiste à distinguer entre compétence et cognition (Silva et al., 2025). Les systèmes d'IA actuels, qu'il s'agisse de grands modèles de langage ou d'algorithmes prédictifs, ne raisonnent pas comme les humains et manquent d'empathie. Leurs résultats proviennent de la détection de schémas statistiques sur de vastes ensembles de données. Ils ne reposent ni sur l'expérience vécue, ni sur un jugement moral, ni sur une compréhension consciente (JM Bishop, 2021). Cette distinction est importante : lorsqu'un système d'IA génère un langage persuasif ou des prévisions financières complexes, il peut simuler la réflexion et la perspicacité sans y participer réellement. Selon les chercheurs Yoshua Bengio et Gary Marcus, l'IA actuelle manque de raisonnement causal et de connaissances concrètes qui caractérisent l'intelligence humaine (Hamilton et al., 2024).

Cependant, malgré ces limitations, l'IA est indéniablement puissante. Dans des domaines comme la finance, sa capacité à traiter des volumes massifs de données, à détecter des corrélations subtiles et à prendre des décisions à grande vitesse lui confère un avantage structurel sur les acteurs humains. Des institutions financières telles que BlackRock (https://www.blackrock.com/corporate) et Goldman Sachs (https://www.goldmansachs.com/) utilisent intensivement des modèles d'IA pour l'optimisation de portefeuille, l'évaluation des risques et le trading algorithmique. La prise de décisions financières ne requiert ni empathie ni biais émotionnel. Ces qualités sont pourtant indispensables lorsque l'IA intervient dans des environnements socialement sensibles ou à forts enjeux, où les valeurs humaines, le jugement éthique et la compréhension du contexte sont essentiels.

Dans le secteur financier, l'IA peut acquérir un statut quasi mythique : elle est perçue comme un oracle, une boîte noire qui « connaît » le marché. Cette perception peut s'avérer dangereuse. Lorsque les décideurs accordent une confiance excessive aux résultats de l'IA, ils risquent de se décharger de leurs responsabilités, ignorant que ces systèmes peuvent amplifier les biais, propager les erreurs ou se comporter de manière imprévisible dans des conditions inédites.

L'événement du 6 mai 2010 sur les marchés financiers américains, connu sous le nom de « Flash Crash », a constitué un moment terrifiant qui a mis en évidence les risques liés à l'utilisation de l'intelligence artificielle dans la finance. L'indice Dow Jones a chuté d'environ 1 000 points (soit 9 %) en quelques minutes, des milliers de milliards de dollars de capitalisation boursière se sont volatilisés et de nombreuses actions ont brièvement atteint des prix absurdes (certaines proches de zéro, d'autres à des niveaux exorbitants). Les enquêtes menées par des organismes de réglementation tels que la Securities and Exchange Commission (SEC) et la Commodity Futures Trading Commission (CFTC) ont conclu qu'aucun bug isolé n'était à l'origine du krach. Il s'agissait plutôt d'une réaction en chaîne impliquant un ordre de vente automatisé de grande ampleur sur les marchés à terme, la réaction ultrarapide des algorithmes de trading haute fréquence (THF) aux variations de prix et l'amplification mutuelle des boucles de rétroaction entre ces algorithmes.

Lors du krach éclair, les marchés avaient accordé leur confiance aux systèmes algorithmiques, les considérant comme des agents rationnels, alors qu'en réalité, ils exécutaient des règles à une vitesse fulgurante. Ce krach a mis en lumière le décalage entre l'intelligence perçue et le mécanisme réel (Shi, 2025). Le danger ne réside donc pas dans les intentions de l'IA, car elle n'en a pas. Il réside

plutôt dans la perception que les humains peuvent avoir de ses intentions, de son intelligence ou de son autorité sur ses décisions. L'IA acquiert sa puissance non seulement par sa nature intrinsèque, mais aussi par la perception qu'on peut en avoir. En finance, où vitesse, échelle et confiance se conjuguent, cette combinaison peut amplifier le risque systémique.

En conclusion, l'IA doit être considérée comme un outil à fort impact, mais non sensible : extraordinairement performante, mais fondamentalement limitée. Son absence de raisonnement et d'empathie ne diminue en rien son utilité ; elle souligne simplement la nécessité d'une gouvernance humaine rigoureuse. Le véritable défi réside non seulement dans le contrôle de l'IA elle-même, mais aussi dans la gestion des perceptions humaines et des pratiques institutionnelles qui l'entourent : veiller à ce que le mythe de l'IA ne supplante pas la méthode et que la confiance ne dépasse pas la compréhension. Comme le dit Harari, nous devons nous garder de faire de l'IA une nouvelle religion. (Norton, 2023).

9. Conclusion développée

L'intelligence artificielle est passée du stade de spéculation philosophique à celui de force technologique transformatrice. Ses fondements en logique, en calcul et en raisonnement symbolique continuent d'influencer les développements modernes.

L'IA est désormais omniprésente dans la sphère privée comme professionnelle. Elle révolutionne le secteur de la santé en aidant les médecins dans le développement de médicaments, les traitements personnalisés, la prédiction des maladies et l'imagerie médicale. Dans le commerce et la finance, l'IA contribue à l'automatisation, à la prévision des marchés, à la détection des fraudes et au service client grâce à des chatbots intelligents. L'IA est de plus en plus utilisée dans l'éducation pour l'enseignement individualisé, l'évaluation automatisée et les systèmes d'apprentissage adaptatifs. Dans le secteur des transports, elle est employée pour l'optimisation logistique, la gestion du trafic et les véhicules autonomes. Le traitement automatique du langage naturel, les assistants virtuels, les systèmes de traduction et les modèles génératifs sophistiqués capables de produire du texte, des graphismes, de la musique et du code informatique sont autant de fonctionnalités rendues possibles par l'IA.

Bien que l'IA ait connu un succès remarquable, elle soulève également des défis éthiques, sociaux et philosophiques complexes. Comprendre son histoire et ses fondements est essentiel pour orienter son développement futur de manière responsable.

Bibliographie

1. Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 53. https://doi.org/10.1186/s40537-021-00444-8
2. Bishop, C. M. (2007). Pattern Recognition and Machine Learning. *Journal of Electronic Imaging*, *16*(4), 049901. https://doi.org/10.1117/1.2819119
3. Bishop, J. M. (2021). Artificial Intelligence Is Stupid and Causal Reasoning Will Not Fix It. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.513474
4. Boden, M. A. (2016). *AI: Its Nature and Future*. Oxford University Press.
5. Crevier, D. (1993). *Ai: The Tumultuous History Of The Search For Artificial Intelligence*.
6. Feibleman, J. K. (1979). Of Aristotle's Logic: The Organon. In *Assumptions of Grand Logics* (pp. 19–30). Springer Netherlands. https://doi.org/10.1007/978-94-009-9278-8_2
7. Hamilton, K., Nayak, A., Božić, B., & Longo, L. (2024). Is neuro-symbolic AI meeting its promises in natural language processing? A structured review. *Semantic Web*, *15*(4), 1265–1306. https://doi.org/10.3233/SW-223228
8. Hovers, E. (2012). *Invention, Reinvention and Innovation* (pp. 51–68). https://doi.org/10.1016/B978-0-444-53821-5.00005-1

9. Ifrah, G. (2001). *The Universal History of Computing: From the Abacus to the Quantum Computer*. Wiley.
10. Kaplan, A. (2022). *Artificial Intelligence, Business and Civilization Our Fate Made in Machines*.
11. Khan, S., Fazil, M., Imoize, A. L., Alabduallah, B. I., Albahlal, B. M., Alajlan, S. A., Almjally, A., & Siddiqui, T. (2023). Transformer Architecture-Based Transfer Learning for Politeness Prediction in Conversation. *Sustainability*, *15*(14), 10828. https://doi.org/10.3390/su151410828
12. Lamsal, R. (2021). *A step by step forward pass and backpropagation example*.
13. Leibniz, G. W. (1976). *Philosophical Papers and Letters* (Leroy E. Loemker, Ed.; Second). Springer Netherlands. https://doi.org/10.1007/978-94-010-1426-7
14. Lovelace, A. (1989). *Notes on the analytical engine*. MIT Press.
15. Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007
16. Nilsson, N. J. (2010). *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. Cambridge University Press.
17. Norton, J. (2023). Could ChatGPT create a new RELIGION? *Daily Mail*.
18. Osiurak, F., Navarro, J., & Reynaud, E. (2018). How Our Cognition Shapes and Is Shaped by Technology: A Common Framework for Understanding Human Tool-Use Interactions in the Past, Present, and Future. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00293
19. Papajorgji, P., & Moskowitz, H. (2025). Introduction to Artificial Intelligence. In *The Mind of Everyday Combining Individual and Artificial Intelligence* (1st ed., pp. 67–90). Springer.
20. Ramachandran, P., Barret, Z., & Quoc, V. Le. (2017). SEARCHING FOR ACTIVATION FUNCTIONS. *ArXiv:1710.05941v2*, *arXiv prep*.
21. Reuters. (2026). *Explainer: What do we know about Anthropic's Mythos amid rising concerns?*
22. Sharma, S., Sharma, S., & Athaiya, A. (2020). ACTIVATION FUNCTIONS IN NEURAL NETWORKS. *International Journal of Engineering Applied Sciences and Technology*, *04*(12), 310–316. https://doi.org/10.33564/IJEAST.2020.v04i12.054
23. Silva, F., P. Oliveira, H., & Pereira, T. (2025). Causal representation learning through higher-level information extraction. *ACM Computing Surveys*, *57*(2), 1–37. https://doi.org/10.1145/3696412
24. Solso, R. L., MacLin, M. M., & Kimberly H., O. (2005). *Cognitive psychology* (7th ed.). Pearson Education.
25. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.
26. Virgo, J., Pillon, J., Navarro, J., Reynaud, E., & Osiurak, F. (2017). Are You Sure You're Faster When Using a Cognitive Tool? *The American Journal of Psychology*, *130*(4), 493–503. https://doi.org/10.5406/amerjpsyc.130.4.0493
27. Whitehead, A., & Russell, B. (2004). *Principia Mathematika* (11th editi). Cambridge University Press.
28. Wiener. U. N. (n.d.). *Automaten*. Springer Nature.