Intelligence artificielle

L'**intelligence artificielle** (**IA**) est l'ensemble des programmes ou algorithmes permettant aux machines d'effectuer des tâches typiquement associées à l'intelligence humaine, comme l'apprentissage, le raisonnement, la résolution de problème, la perception ou la prise de décision. L'intelligence artificielle est également le champ de recherche visant à développer de telles machines ainsi que les systèmes informatiques qui en résultent.

Souvent classée dans le domaine des mathématiques et des sciences cognitives, l'IA fait appel à des disciplines telles que la neurobiologie computationnelle (qui a notamment inspiré les réseaux neuronaux artificiels), les statistiques, ou l'algèbre linéaire. Elle vise à résoudre des problèmes à forte complexité logique ou algorithmique. Par extension, dans le langage courant, l'IA inclut les dispositifs imitant ou remplaçant l'homme dans certaines mises en œuvre de ses fonctions cognitives[1].

Les applications de l'IA couvrent de nombreux domaines, notamment les moteurs de recherche, les systèmes de recommandation, l'aide au diagnostic médical, la compréhension du langage naturel, les voitures autonomes, les *chatbots*, les outils de génération d'images, les outils de prise de décision automatisée, les programmes compétitifs dans des jeux de stratégie et certains personnages non-joueurs de jeu vidéo[2].

Depuis l'apparition du concept, les finalités, les enjeux et le développement de l'IA suscitent de nombreuses interprétations, fantasmes ou inquiétudes, que l'on retrouve dans les récits ou films de science-fiction, dans les essais philosophiques[3] ainsi que parmi des économistes.

Intelligence artificielle

Partie de	Informatique, nouvelles technologies, raisonnement
Pratiqué par	Chercheur ou chercheuse en intelligence artificielle (d), ingénieur en intelligence artificielle
Champs	Intelligence artificielle symbolique connexionnisme ethical artificial intelligence (d) apprentissage automatique
Histoire	Histoire de l'intelligence artificielle

Définition

L'expression « intelligence artificielle », souvent abrégée par le sigle « IA » (ou « AI » en anglais, pour *artificial intelligence*) a été introduite en 1956 par John McCarthy, qui l'a définie en 2004[4] comme « la science et l'ingénierie de la fabrication de machines intelligentes, en particulier de programmes

informatiques intelligents. Elle est liée à la tâche similaire qui consiste à utiliser des ordinateurs pour comprendre l'intelligence humaine, mais l'IA ne doit pas se limiter aux méthodes qui sont biologiquement observables »[5].

Pour Marvin Lee Minsky, l'un de ses créateurs, l'IA est « la construction de programmes informatiques qui s'adonnent à des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisante par des êtres humains car elles demandent des processus mentaux de haut niveau tels que : l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique »[a, 6]. Cette définition combine l'aspect « artificiel » des ordinateurs et des processus informatiques, aux aspects « intelligents » d'imitation de comportements humains, notamment de raisonnement et d'apprentissage. Celui-ci est à l'œuvre dans les jeux, dans la pratique des mathématiques, dans la compréhension du langage naturel, dans la perception visuelle (interprétation des images et des scènes), auditive (compréhension du langage parlé) ou par d'autres capteurs, dans la commande d'un robot dans un milieu inconnu ou hostile.

Avant les années 2000, d'autres définitions sont proches de celle de Minsky, mais varient sur deux points fondamentaux[7] :

- les définitions qui lient l'IA à un aspect *humain* de l'intelligence et celles qui la lient à un modèle idéal d'intelligence, non forcément humaine, nommée *rationalité* ;
- les définitions qui insistent sur le fait que l'IA a pour but d'avoir *toutes les apparences* de l'intelligence (humaine ou rationnelle), et celles qui insistent sur le fait que le *fonctionnement interne* du système d'IA doit ressembler également à celui de l'être humain et être au moins aussi rationnel.

Le grand public confond souvent l'intelligence artificielle avec l'apprentissage automatique (*machine learning*) et l'apprentissage profond (*deep learning*). Ces trois notions diffèrent et sont en réalité imbriquées : l'intelligence artificielle englobe l'apprentissage automatique, qui lui-même englobe l'apprentissage profond[8].

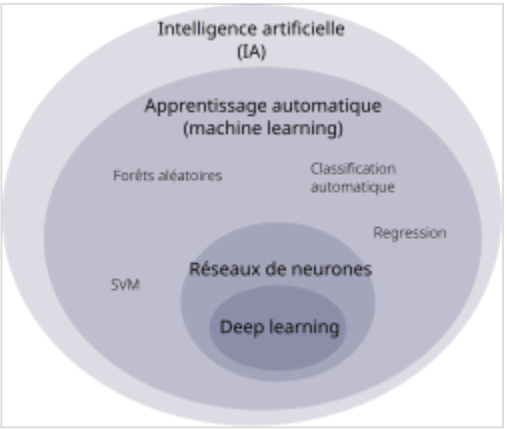


Diagramme de Venn montrant comment s'imbriquent les notions d'intelligence artificielle, d'apprentissage automatique et d'apprentissage profond.

Les définitions font souvent intervenir[9, 10] :

- une capacité à percevoir l'environnement et à prendre en compte la complexité du monde réel ;
- un traitement de l'information (collecter et interpréter des intrants, captés sous forme de données) ;
- des prises de décision (y compris dans le raisonnement et l'apprentissage), choix d'actions, exécution de tâches (dont d'adaptation, de réaction aux changements de contexte…), avec un certain niveau d'autonomie ;
- la réalisation d'objectifs spécifiques (raison ultime des systèmes d'IA).

Le groupe AI Watch note que les IA peuvent aussi être classées en fonction des familles d'algorithmes et/ou des modèles théoriques qui les sous-tendent, des capacités cognitives reproduites par l'IA, des fonctions exécutées par l'IA. Les applications de l'IA peuvent, elles, être classées en fonction du secteur socioéconomique et/ou des fonctions qu'elles y remplissent[9].

Une manière de définir l'intelligence artificielle est de considérer ses applications et les types de tâches qu'elle résout. Un rapport de la Commission européenne publié en 2020 présente une taxonomie classant les définitions de l'IA selon diverses tâches réalisées, telles que le raisonnement, l'apprentissage, la perception, etc.[11]. La même année, le professeur Jack Copeland propose une définition similaire, qui permet de distinguer plus clairement les facettes de l'IA, selon cinq catégories principales[11] :

1. L'apprentissage généralisé : l'IA apprend à partir de données diverses pour identifier des modèles et appliquer ces connaissances à de nouvelles situations, comme détecter des fraudes en ligne ;
2. Le raisonnement : cette capacité permet à l'IA de faire des prédictions et tirer des conclusions à partir des données, aidant dans des décisions comme la prédiction de comportements d'achat ;
3. La résolution de problèmes : l'IA trouve des solutions optimales pour des problèmes spécifiques, utilisée dans des contextes comme l'optimisation industrielle ou les stratégies de jeu ;
4. La perception : elle permet à l'IA de reconnaître et interagir avec son environnement, utilisée dans la robotique avancée et les véhicules autonomes pour naviguer et accomplir des tâches ;
5. La compréhension du langage : l'IA analyse et génère du langage à travers le NLP, utilisé dans des applications comme les assistants vocaux et les *chatbots* pour améliorer l'interaction utilisateur.

Techniques

Comportements prédéfinis (imitation d'intelligence)

Pour les intelligences artificielles servant principalement à donner une impression d'intelligence dans un cadre contrôlé, notamment pour les personnages non-joueurs des jeux vidéo, il est courant que l'apprentissage automatique ne soit pas utilisé. Un ensemble de fonctions et comportements plus précises et moins flexibles sont alors implémentés.

Il s'agit souvent de liste de textes ou paroles prédéfinis, aux déclenchements parfois conditionnels, par exemple un choix de mouvements suivant une série de règles et des déplacements de *pathfinding*.

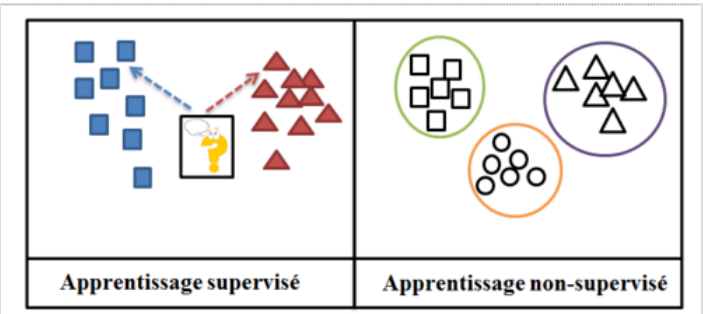
Apprentissage automatique

L'apprentissage automatique consiste à permettre au modèle d'IA d'apprendre à effectuer une tâche au lieu de spécifier exactement comment il doit l'accomplir[12]. Le modèle contient des paramètres dont les valeurs sont ajustées tout au long de l'apprentissage. La méthode de la rétropropagation du gradient est capable de détecter, pour chaque paramètre, dans quelle mesure il a contribué à une bonne réponse ou à une erreur du modèle, et peut l'ajuster en conséquence. L'apprentissage automatique nécessite un moyen d'évaluer la qualité des réponses fournies par le modèle[13]. Les principales méthodes d'apprentissage sont :

Apprentissage supervisé
: Un jeu de données annoté est utilisé pour entraîner l'algorithme. Il contient des données d'entrée fournies au modèle et les réponses correspondantes attendues, que le modèle est entraîné à produire[12]. Il est parfois difficile de se procurer suffisamment de données annotées avec les réponses attendues[14].

Apprentissage non supervisé

Un jeu de données est fourni au modèle, mais n'est pas annoté avec les réponses attendues. Le but peut par exemple être de regrouper les données similaires entre elles[12] (*clustering*).



En apprentissage supervisé, les réponses sont connues, tandis qu'en apprentissage non supervisé, l'algorithme découvre des structures dans les données par lui-même[12].

Apprentissage auto-supervisé

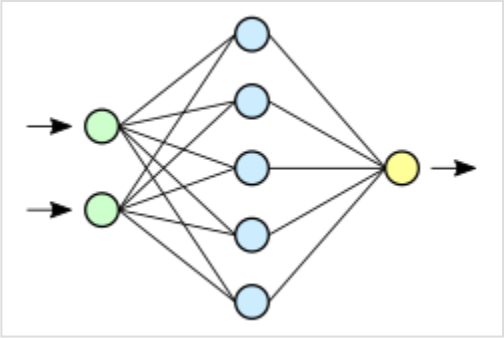
Un problème d'apprentissage supervisé est *automatiquement* généré à partir d'un jeu de données non annoté. Cela fonctionne souvent en cachant une partie des informations (des mots d'un texte, des morceaux d'images…) afin d'entraîner le modèle à les prédire[15].

Apprentissage par renforcement

L'agent est plongé dans un environnement où ce qu'il fait est évalué. Par exemple, un agent peut apprendre à jouer aux échecs en jouant contre lui-même, et le résultat (victoire ou défaite) permet à chaque itération d'évaluer s'il a bien joué. Il n'y a dans ce cas pas besoin de jeu de données[12].

Réseaux de neurones

Les réseaux de neurones artificiels sont inspirés du fonctionnement du cerveau humain : les neurones sont en général connectés à d'autres neurones en entrée et en sortie. Les neurones d'entrée, lorsqu'ils sont activés, agissent comme s'ils participaient à un vote pondéré pour déterminer si un neurone intermédiaire doit être activé et ainsi transmettre un signal vers les neurones de sortie. En pratique, pour l'équivalent artificiel, les « neurones d'entrée » ne sont que des nombres et les poids de ce « vote pondéré » sont des paramètres ajustés lors de l'apprentissage[16, 17].


Exemple de réseau de neurones comprenant deux neurones d'entrée (en vert), une couche « cachée » de neurones (en bleu) et un neurone de sortie (en jaune).

Hormis la fonction d'activation, les réseaux de neurones artificiels n'effectuent en pratique que des additions et des multiplications matricielles, ce qui fait qu'ils peuvent être accélérés par l'utilisation de processeurs graphiques[18]. En théorie, un réseau de neurones peut approximer n'importe quelle fonction[19].

Pour de simples réseaux de neurones à propagation avant (*feedforward* en anglais), le signal ne passe que dans une direction. Avec les réseaux de neurones récurrents, le signal de sortie de chaque neurone est réinjecté en entrée de ce neurone, permettant d'implémenter un mécanisme de mémoire à court terme[20].

Les réseaux neuronaux convolutifs, qui sont particulièrement utilisés en traitement d'images, introduisent une notion de localité. Leurs premières couches identifient des motifs relativement basiques et locaux comme des contours, là où les dernières couches traitent de motifs plus complexes et globaux[17].

Apprentissage profond

L'apprentissage profond (*deep learning* en anglais) utilise de multiples couches de neurones entre les entrées et les sorties, d'où le terme « profond »[21]. L'utilisation de processeurs graphiques pour accélérer les calculs et l'augmentation des données disponibles a contribué à la montée en popularité de l'apprentissage profond. Il est utilisé notamment en vision par ordinateur, en reconnaissance automatique de la parole et en traitement automatique des langues[22] (ce qui inclut les grands modèles de langage).

Grands modèles de langages

Les grands modèles de langage sont des modèles de langage ayant des milliards de paramètres. Ils reposent très souvent sur l'architecture transformeur[23].

Les transformeurs génératifs préentraînés (*Generative Pretrained Transformers* ou *GPT* en anglais) sont un type particulièrement populaire de grand modèle de langage. Leur « pré-entraînement » consiste à prédire, étant donnée une partie d'un texte, le *token* suivant (un *token* étant une séquence de caractères, typiquement un mot, une partie d'un mot, ou de la ponctuation). Cet entraînement à prédire ce qui va suivre, répété pour un grand nombre de textes, permet à ces modèles d'accumuler des connaissances sur le monde. Ils peuvent ensuite générer du texte semblable à celui ayant servi au pré-entraînement, en prédisant un à un les *tokens* suivants. En général, une autre phase d'entraînement est ensuite effectuée pour rendre le modèle plus véridique, utile et inoffensif. Cette phase d'entraînement (utilisant souvent une technique appelée RLHF) permet notamment de réduire un phénomène appelé « hallucination », où le modèle génère des informations d'apparence plausible mais fausses[24].

Avant d'être fourni au modèle, le texte est découpé en *tokens*. Ceux-ci sont convertis en vecteurs qui en encodent le sens ainsi que la position dans le texte. À l'intérieur de ces modèles se trouve une alternance de réseaux de neurones et de couches d'attention. Les couches d'attention combinent les concepts entre eux, permettant de tenir compte du contexte et de saisir des relations complexes[25].

Ces modèles sont souvent intégrés dans des agents conversationnels, aussi appelés chatbots, où le texte généré est formaté pour répondre à l'utilisateur. Par exemple, l'agent conversationnel ChatGPT exploite les modèles GPT-3.5 et GPT-4[26]. En 2023 font leur apparition des modèles grand public pouvant traiter simultanément différents types de données comme le texte, le son, les images et les vidéos, tel Google Gemini[27].

Recherche et optimisation

Certains problèmes nécessitent de chercher intelligemment parmi de nombreuses solutions possibles.

Recherche locale

La recherche locale, ou recherche par optimisation, repose sur l'optimisation mathématique pour trouver une solution numérique à un problème, en améliorant progressivement la solution choisie[28].

En particulier, en apprentissage automatique, la descente de gradient permet de trouver une solution localement optimale, étant donné une fonction de coût à minimiser en faisant varier les paramètres du modèle. Elle consiste, à chaque étape, à modifier les paramètres à optimiser dans la direction qui permet de réduire le mieux la fonction de coût. La solution obtenue est *localement* optimale, mais il se peut qu'il y ait globalement de meilleures solutions, qui auraient pu être obtenues avec différentes valeurs initiales de paramètres[28]. Les modèles d'IA modernes peuvent avoir des milliards de paramètres à optimiser, et utilisent souvent des variantes plus complexes et efficaces de la descente de gradient[23].

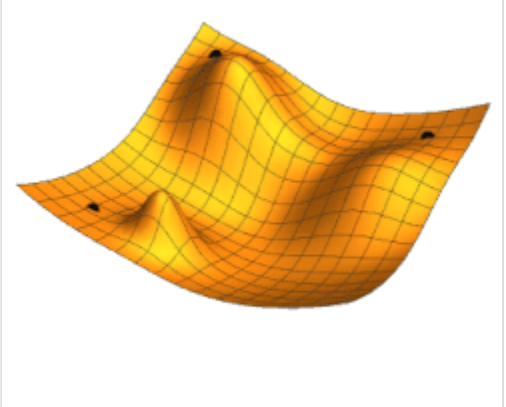


Illustration de la descente de gradient pour trois points de départ différents, faisant varier deux paramètres de sorte à minimiser la fonction de coût représentée par la hauteur.

Les algorithmes évolutionnistes, inspirés de la théorie de l'évolution, utilisent une forme de recherche par optimisation. À chaque étape, des opérations telles que la « mutation » ou le « croisement » sont effectuées de manière aléatoire pour obtenir différentes variantes, et les variantes les mieux adaptées sont sélectionnées pour l'étape suivante[28].

Recherche dans l'espace des états

La recherche dans l'espace des états vise à trouver un état accomplissant l'objectif à travers un arbre des états possibles[29]. Par exemple, la recherche antagoniste est utilisée pour des programmes jouant à des jeux tels que les échecs ou le go. Elle consiste à parcourir l'arbre des coups possibles par le joueur et son adversaire, à la recherche d'un coup gagnant[30]. La simple recherche exhaustive est rarement suffisante en pratique vu le nombre d'états possibles. Des heuristiques sont utilisées pour prioriser les chemins les plus prometteurs[31].

Logique

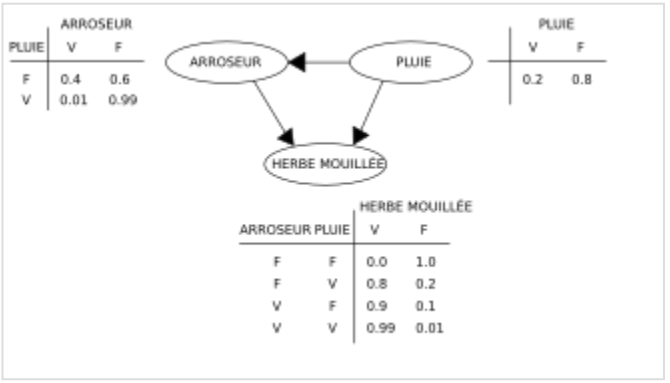
La logique formelle est utilisée pour le raisonnement et la représentation des connaissances. Elle se décline en deux principales formes, la logique propositionnelle et la logique prédicative. La logique propositionnelle opère sur des affirmations qui sont vraies ou fausses, et utilise la logique connective avec des opérateurs tels que « et », « ou », « non » et « implique ». La logique prédicative étend la logique propositionnelle et peut aussi opérer sur des objets, prédicats ou relations. Elle peut utiliser des quantificateurs comme dans « *Chaque* X est un Y » ou « *Certains* X sont des Y »[32].

L'inférence logique (ou déduction) est le processus qui consiste à fournir — à l'aide d'un moteur d'inférence — une nouvelle affirmation (la conclusion) à partir d'autres affirmations connues comme étant vraies (les prémisses). Une règle d'inférence décrit les étapes valides d'une preuve ; la plus générale est la règle de résolution. L'inférence peut être réduite à la recherche d'un chemin amenant des prémisses aux conclusions, où chaque étape est une application d'une règle d'inférence[32]. Mais à part pour de courtes preuves dans des domaines restreints, la recherche exhaustive prend beaucoup de temps.

La logique floue assigne des valeurs de vérité entre 0 et 1, permettant de gérer des affirmations vagues, comme « il fait chaud »[33]. La logique non monotone permet d'annuler certaines conclusions[32]. Diverses autres formes de logique sont développées pour décrire de nombreux domaines complexes.

Méthodes probabilistes et gestion de l'incertitude

De nombreux problèmes en IA (raisonnement, planification, apprentissage, perception, robotique, etc.) nécessitent de pouvoir opérer à partir d'informations incomplètes ou incertaines[34].



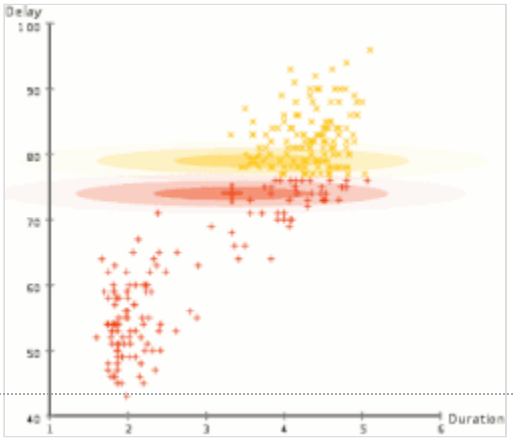
Un exemple de réseau bayésien, et les tables de probabilité conditionnelle associées.

Certaines techniques reposent sur l'inférence bayésienne, qui fournit une formule pour mettre à jour des probabilités subjectives étant données de nouvelles informations. C'est notamment le cas des réseaux bayésiens. L'inférence bayésienne nécessite souvent d'être approximée pour pouvoir être calculée[35].

Les méthodes de Monte-Carlo sont un ensemble de techniques pour résoudre des problèmes complexes en effectuant aléatoirement de nombreuses simulations afin d'approximer la solution[36].

Les réseaux de neurones peuvent aussi être optimisés pour fournir des estimations probabilistes[37].

Des outils mathématiques précis ont été développés pour analyser comment des agents intelligents peuvent faire des choix et des plans en utilisant la théorie de la décision, la maximisation de l'espérance et la théorie de la valeur de l'information. Ces techniques comprennent des modèles tels que les processus de décision markoviens, la théorie des jeux et les mécanismes d'incitation[35].



Séparation des données en deux groupes (partitionnement) par un algorithme de maximisation de l'espérance.

Classifieurs et méthodes statistiques

De nombreux modèles d'IA ont pour but d'assigner une catégorie (classification), une valeur (régression) ou une action à des données fournies. Les méthodes de classification comprennent arbres de décision, k plus proches voisins, machine à vecteurs de support ou classification bayésienne naïve[38,35]. Les réseaux de neurones peuvent également faire de la classification[39].

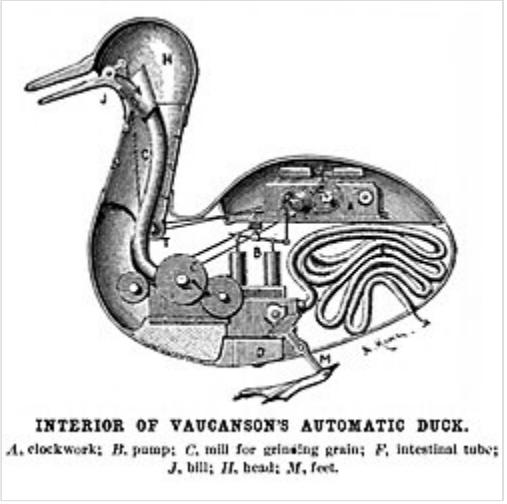
Intelligence artificielle quantique

L'IA quantique est un domaine de recherche interdisciplinaire qui vise à exploiter les propriétés uniques de la physique quantique (dont la superposition quantique et l'intrication quantique) pour résoudre des problèmes complexes inaccessibles aux IA classiques, au moyen d'ordinateurs quantiques exécutant de

nouveaux types d'algorithmes.

Histoire

Comme précurseur à l'intelligence artificielle, divers automates ont été créés au cours de l'histoire, dont le canard de Vaucanson ou les automates d'Al-Jazari. Certains automates remontent à l'Antiquité et étaient utilisés pour des cérémonies religieuses[40]. Des mythes et rumeurs rapportent également la création d'êtres intelligents, par exemple les golems[41].



Le canard artificiel de Vaucanson (1738).

Des philosophes et mathématiciens comme Raymond Lulle, Leibniz ou George Boole ont cherché à formaliser le raisonnement et la génération d'idées[42].

Au xx^e siècle, Alan Turing a notamment inventé un modèle de calcul par la suite appelé machine de Turing, exploré la notion de calculabilité et d'intelligence des machines, et proposé le « jeu de l'imitation » (test de Turing) pour évaluer l'intelligence de futures machines[42]. Le terme « intelligence artificielle » a été mis en avant par John McCarthy lors de la conférence de Dartmouth en 1956, où l'intelligence artificielle a été établie en tant que discipline à part entière[43,44]. Dans les années qui ont suivi, des chercheurs ont proposé diverses preuves de concept, dans des situations spécifiques, de ce que les machines peuvent faire en théorie. Par exemple, le programme ELIZA pouvait se faire passer pour un psychothérapeute, et le Logic Theorist pouvait démontrer des théorèmes[45].

La fin du siècle a été marquée par des périodes d'enthousiasme, et deux périodes de désillusion et de gel des financements appelées « hivers de l'IA »[46], la première de 1974 à 1980 et la seconde de 1987 à 1993. Les systèmes experts ont été particulièrement populaires dans les années 1980, malgré leur fragilité et la difficulté à implémenter manuellement les bonnes règles d'inférences[45]. Des techniques d'apprentissage automatique se sont développées (réseaux de neurones, rétropropagation du gradient, algorithmes génétiques) ainsi que l'approche connexionniste[45]. Mais les faibles puissances de calcul et le manque de données d'entraînement limitaient leur efficacité. Certains domaines n'ont progressivement plus été considérés comme faisant partie de l'intelligence artificielle, à mesure qu'une solution efficace était trouvée[47] ; un phénomène parfois appelé « effet IA ». En 1997, pour la première fois, un supercalculateur a gagné plusieurs parties au jeu d'échec contre le champion du monde.

Dans les années 2000, le Web 2.0, le *big data* et de nouvelles infrastructures et capacités de calcul ont permis l'exploration de masses de données sans précédent. En 2005, le projet *Blue Brain* a débuté, ayant pour objectif de simuler le cerveau de mammifères[48]. En 2012, le réseau neuronal convolutif AlexNet a lancé l'utilisation de processeurs graphiques pour entraîner des réseaux de neurones, décuplant ainsi les capacités de calcul dédiées à l'apprentissage[49]. En 2016, un programme a gagné quatre des cinq parties de go jouées contre Lee Sedol, l'un des meilleurs joueurs au monde. Des organisations visant à créer une intelligence artificielle générale ont vu le jour, comme DeepMind en 2010[50] et OpenAI en 2015[51]. Dès les années 2010, des outils d'intelligence artificielle (spécialisée ou générative) ont accompli des progrès

spectaculaires, mais restent loin des performances du vivant dans beaucoup de ses aptitudes naturelles, en particulier sur son aptitude à apprendre rapidement à partir d'un faible volume d'information (par induction), selon le magazine *Slate* en 2019[52].

Dans les années 2010, les assistants personnels intelligents sont l'une des premières applications grand public de l'intelligence artificielle.

En 2017, des chercheurs de Google ont proposé l'architecture transformeur, qui a servi de base aux grands modèles de langage. En 2018, Yann Le Cun, Yoshua Bengio et Geoffrey Hinton ont remporté le prix Turing pour leurs travaux sur l'apprentissage profond[53, 54].

En 2022, des programmes générant des images à partir de descriptions textuelles, comme Midjourney ou DALL-E 2, se sont popularisés[55]. La même année, l'agent conversationnel ChatGPT a connu une croissance inédite, gagnant un million d'utilisateurs en seulement cinq jours[56] et cent millions d'utilisateurs en deux mois[57], ce qui a accentué un phénomène de « course » à l'IA[58]. En 2023, les progrès rapides de l'IA ont suscité des inquiétudes quant à un potentiel risque d'extinction de l'humanité[59]. Des modèles de fondation « multimodaux », c'est-à-dire capables de traiter simultanément plusieurs modalités (texte, images, son) ont émergé, tels que Google Gemini[60] et GPT-4o[61].

De nouvelles infrastructures matérielles se sont développées, exploitant notamment les processeurs graphiques Blackwell et les TPU. Des recherches explorent également les potentielles applications de l'ordinateur quantique en AI. Le concept d'« usine d'IA » s'est également concrétisé, qui comprend des infrastructures intégrées combinant production et entraînement de modèles à grande échelle avec des supercalculateurs et des centres de données dédiés. En 2025, plusieurs projets de ce type étaient annoncés, notamment aux États-Unis et en Europe, comme le partenariat entre l'Allemagne et Nvidia[62].

Intelligence artificielle générale

L'intelligence artificielle générale (IAG) comprend tout système informatique capable d'effectuer ou d'apprendre pratiquement n'importe quelle tâche cognitive propre aux humains ou autres animaux[63]. Elle peut alternativement être définie comme un système informatique surpassant les humains dans la plupart des tâches ayant un intérêt économique[64].

L'intelligence artificielle générale a longtemps été considérée comme un sujet purement spéculatif[65]. Certains travaux de recherche ont déjà décrit GPT-4 comme ayant des « étincelles » d'intelligence artificielle générale[66, 67]. Les experts en intelligence artificielle affichent de larges désaccords et incertitudes quant à la date potentielle de conception des premières intelligences artificielles générales (parfois appelées « intelligences artificielles de niveau humain »), leur impact sur la société, et leur potentiel à déclencher une « explosion d'intelligence »[68].

Un sondage de 2022 suggère que 90 % des experts en IA pensent que l'IAG a plus d'une chance sur deux d'être réalisée dans les 100 ans, autour d'une date médiane de 2061[69].

Une superintelligence artificielle est un type hypothétique d'intelligence artificielle générale dont les capacités intellectuelles dépasseraient de loin celles des humains les plus brillants[70]. Le philosophe Nick Bostrom note que les machines disposent de certains avantages par rapport aux cerveaux humains,

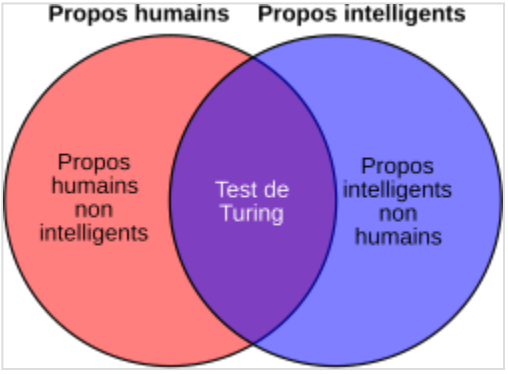
notamment en ce qui concerne la mémoire, la vitesse (la fréquence des processeurs étant de l'ordre de dix millions de fois plus élevée que celle des neurones biologiques) et la capacité à partager des connaissances[71].

Tests

Dans ce contexte, un test est un moyen d'évaluer les capacités d'une intelligence artificielle à imiter certains comportements et raisonnements humains.

Test de Turing

Dans le test de Turing, une machine et un humain répondent textuellement aux questions d'un interrogateur humain. L'interrogateur ne les voit pas mais doit déterminer à partir des réponses textuelles lequel des deux est la machine. Pour passer le test, la machine doit parvenir une bonne partie du temps à tromper l'interrogateur. Ce test a été conçu par Alan Turing en 1950 dans l'article « *Computing Machinery and Intelligence* ». Initialement appelé le « jeu de l'imitation », son but était de fournir une expérience concrète pour déterminer si les machines peuvent penser[72].



Le test de Turing évalue la capacité à se faire passer pour un humain dans un dialogue écrit, mais incite donc aussi à imiter les défauts humains.

Test du café

Imaginé par Steve Wozniak, le test du café consiste à placer un système intelligent dans un habitat américain moyen et à lui demander de faire un café[73]. La réussite du test implique donc plusieurs tâches comme l'orientation dans un environnement inconnu, déduire le fonctionnement d'une machine, trouver les ustensiles nécessaires…

Test de l'étudiant

Proposé par Ben Goertzel, le test de l'étudiant évalue la capacité d'un robot à s'inscrire dans un établissement d'enseignement supérieur, suivre les cours, passer les examens et obtenir le diplôme final[74].

Test de l'embauche

Proposé par le chercheur Nils John Nilsson, le test de l'embauche consiste à faire postuler un système intelligent à un travail important pour l'économie, où il doit travailler au moins aussi bien qu'un humain[75].

Personnalités

Prix Turing

Plusieurs prix Turing (ACM Turing Award) ont été attribués à des pionniers de l'intelligence artificielle, notamment :