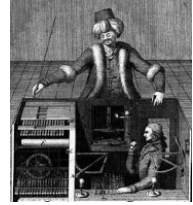


# Les Intelligences Artificielles

## Un peu d'histoire

De nombreux mythes et légendes rapportent la création d'êtres intelligents par l'homme. C'est le cas par exemple de la légende de **Galatée** créée par Pygmalion de la mythologie grecque ou bien des **golems** de la mythologie juive.

Au cours de l'histoire, les automates font figure de précurseurs de l'intelligence artificielle. On peut citer par exemple les automates humanoïdes d'**Al-Jazari** au XII<sup>e</sup> siècle ou bien le **Turc mécanique** joueur d'échecs au XVIII<sup>e</sup> siècle.



Parallèlement, philosophes et scientifiques théorisent sur le processus de la pensée humaine. En 1950 Alan Turing propose le **test de Turing** qui évalue la capacité d'une machine à se faire passer pour un humain dans une conversation, interrogeant la capacité d'une machine à penser.

### 1956 : La conférence de Dartmouth

Cette conférence est souvent considérée comme le moment-clé fondateur de l'intelligence artificielle en tant que discipline indépendante. L'expression « intelligence artificielle » fut adoptée pour la première fois lors de cet événement.

### 1956 – 1974 : L'Age d'or

Les années qui suivent la conférence de Dartmouth sont une ère de découverte et d'enthousiasme. Les progrès semblent prometteurs avec les premiers programmes capables de jouer aux échecs, de traiter le langage naturel et de tenir une conversation avec un humain. On prophétise que d'ici 20 ans on construira une machine aussi intelligente qu'un humain.

Deux approches se dégagent dans cette période : faire un esprit ou modéliser le cerveau.

L'approche **symbolique** essaye de reproduire la pensée humaine par la programmation basée sur la logique algorithmique.

L'approche **connexionniste** qui essaye de modéliser le cerveau humain à l'aide de réseau de « neurones artificiels ».

### 1974 – 1980 : le premier « hiver de l'IA »

Après cet âge d'or, arrive les premières désillusions : les recherches s'essoufflent face à des difficultés techniques et les capacités des ordinateurs et des programmes IA sont trop limitées pour être vraiment utiles. Frustrés par le manque de progrès les investisseurs publics et privés coupent pratiquement tous les fonds de recherche fondamentale en IA.

### 1980 – 1987 : Le boom des systèmes experts

Dans les années 80 les systèmes experts commencent à être progressivement adoptés au sein des entreprises en tant qu'outil d'aide à la décision.

Un système expert est un logiciel capable de répondre à des questions, en effectuant un raisonnement à partir de faits et de règles connues. Parmi les premiers systèmes experts on peut citer **Dendral** qui permettait d'identifier les éléments chimiques d'un matériau ou **Mycin** un système d'aide au diagnostic de maladie du sang et de prescription de médicaments.



## 1987-1993 : Le deuxième hiver de l'IA

L'engouement des investisseurs dans l'intelligence artificielle et les systèmes experts a gonflé puis chuté dans les années 1980 en suivant le schéma d'une bulle économique. Les recherches scientifiques se poursuivent malgré les coupes budgétaires.

### 1997 : Deep Blue contre Kasparov

Pour la première fois un superordinateur bat le champion incontesté des échecs Garry Kasparov. Ce succès retentissant marque le renouveau de l'IA et un retour timide des investissements.



### 2010 le renouveau des IA

La montée en puissance des ordinateurs et l'avènement des GPU (des processeurs spécialisés dans l'exécution de nombreux calculs simultanément) permettent d'outrepasser les difficultés techniques en particulier pour les réseaux de neurones.

En 2011 Apple annonce la sortie de **Siri**, un assistant virtuel pour Iphone dopé à l'intelligence artificielle. Dans les mois qui suivent, des assistants virtuels concurrents sont proposés comme **Google Now**, Microsoft **Cortana** ou **Alexa** d'Amazon.

En 2012, le modèle **AlexNet**, obtient les meilleures performances à la campagne d'évaluation internationale ImageNet de reconnaissance d'images. Ce modèle surpasse largement le deuxième et popularise ainsi les approches par apprentissage profond en vision par ordinateur.

En 2015, le programme **AlphaGo**, un modèle neuronal profond qui a « appris » à jouer au jeu de go grâce à l'apprentissage par renforcement, bat le champion européen Fan Hui sur le score de 5 – 0. Il n'aura fallu que 5 ans pour développer cette IA.

En 2021, la société OpenAI propose **DALL-E** un programme d'intelligence artificielle générative capable de créer des images à partir de leur description (ou prompt). D'autres programmes le concurrence comme Midjourney, Stable Diffusion ou Flux.

En 2022 cette même société OpenAI propose **ChatGPT**, un agent conversationnel (ou chatbot) capable de tenir une conversation, de répondre à des questions, faire des traductions... Dès son lancement cet outil connaît une forte popularité atteignant 100 millions de comptes en quelques mois.



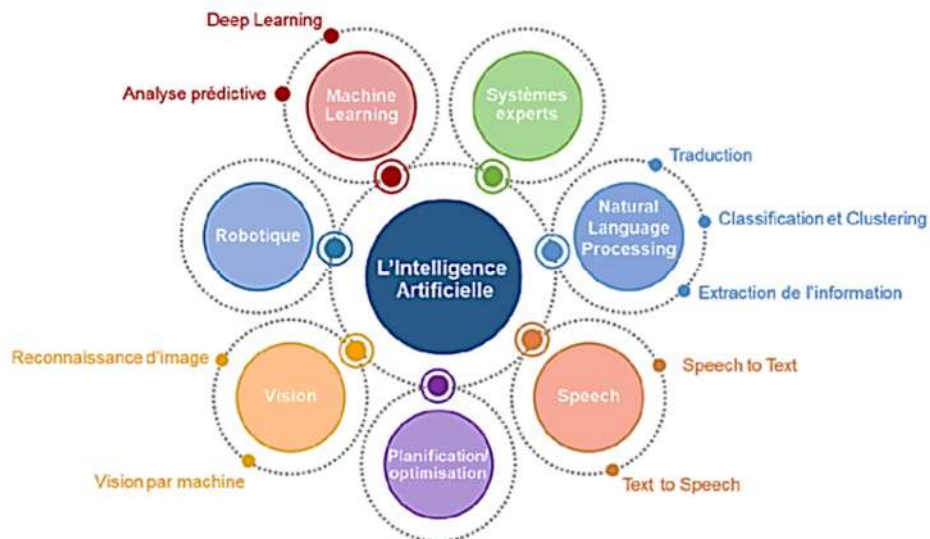
En 2025, on commence à voir apparaître des IA spécialisées dans la génération de vidéos. On peut citer par exemple **Sora** d'OpenAI mais également Google **Veo** ou **Vibes** de Facebook.

Aujourd'hui l'IA est omniprésente : chaque géant de la tech s'est doté de son IA de Google **Gemini** sur les téléphones à Microsoft **Copilot** sur les ordinateurs. De manière plus discrète, les systèmes de détections de fraude, de recommandations de contenu sur Internet, d'ajustement des prix des billets d'avion s'appuient sur des algorithmes et de l'IA pour fonctionner.

## Un peu de vocabulaire

L'**Intelligence Artificielle** désigne la possibilité pour une machine de reproduire des comportements liés aux humains, tels que le raisonnement, la planification et la créativité. (*Définition du Parlement Européen*).

Ses domaines d'application sont très nombreux et couvrent entre autres : les moteurs de recherche, les systèmes de recommandation, la robotique et les voitures autonomes, les outils de génération de texte et d'images, les outils d'aide à la décision et au diagnostic...



Un **algorithme** est une suite d'instructions et d'opérations à appliquer dans un certain ordre pour résoudre un problème.

L'**apprentissage automatique** (machine learning en anglais) est une sous-catégorie de l'IA qui utilise des algorithmes pour permettre à un ordinateur d'apprendre par lui-même, à l'aide de données fournies sans être explicitement programmé pour accomplir cette tâche. L'**apprentissage profond** est un type d'apprentissage tirant partie des **réseaux de neurones** artificiels pour effectuer son apprentissage.

Un **grand modèle de langage** (LLM) est un algorithme d'apprentissage profond capable de reconnaître et d'interpréter le langage humain ou d'autres types de données complexes. Le terme « grand » signifie que ces modèles sont entraînés avec d'énormes quantités de données. Souvent, ces données proviennent d'Internet et peuvent représenter des milliers, voire des millions, de giga-octets de texte. Le plus connu de ces LLM est le modèle GPT.

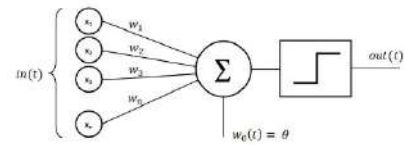
Un **modèle de diffusion** est un type de modèle capable de générer des données similaires auxquelles il a été entraîné. On utilise principalement ces modèles dans la génération d'images et de vidéos mais aussi dans la synthèse sonore.

Une **IA faible** (ou étroite) est une IA spécialisée sur un type de tâche : génération de texte, d'image, reconnaissance vocale, traduction... La totalité des IA actuelles sont des IA faible. À l'inverse une **IA forte** serait une IA capable de reproduire toutes les capacités cognitives humaines. À l'heure actuelle, une IA forte relève de la science-fiction.

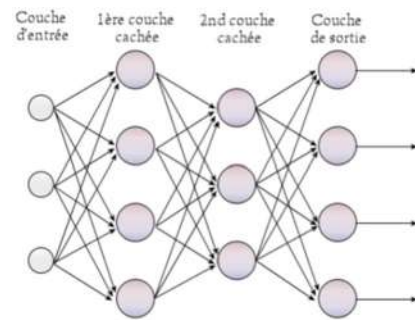
## Une IA, comment ça marche ?

Les réseaux de neurones sont à l'origine du renouveau des IA. Ces neurones artificiels appelés « **perceptron** » sont une modélisation mathématique du fonctionnement d'un neurone biologique.

Un perceptron reçoit plusieurs signaux en entrée pondéré par un coefficient (ou poids). Si la somme de ces signaux pondérés dépasse un certain seuil, alors le perceptron émet à son tour un signal.



Dans un réseau de neurones, ceux-ci sont connectés par couches successives et transmettent les signaux d'une couche à la suivante. La force de ces connexions, représentée en poids, détermine l'influence de la sortie d'un neurone sur l'entrée d'un autre neurone. Un réseau de neurones peut comporter plusieurs centaines de couches avec plusieurs dizaines de milliers de neurones par couches.



L'entraînement d'un réseau de neurone consiste donc à ajuster progressivement les poids de ces connexions à partir d'une grande quantité de données.

### Comment fonctionnent les LLM ?

Les LLM utilisent des réseaux neuronaux particuliers, de type « **transformers** », capables de traiter des séquences entières de données simultanément. Cela permet aux transformers de capturer plus efficacement les relations entre les mots dans un texte et de leur apporter un contexte.

Dans un premier temps, le texte est décomposé en « **tokens** » (des mots ou des syllabes) ces tokens sont ensuite converti en groupe de chiffres. Ces chiffres seront d'autant plus proches que les mots sont dans le même champs lexicaux. Dans son traitement dans le transformer, certains tokens seront d'avantage pris en considération en fonction du contexte et des mots traités précédemment.

Le modèle prédit ensuite le mot suivant. Pour ce faire, il génère une distribution de probabilité sur le vocabulaire pour le token suivant. Par exemple pour la phrase : « *Aujourd'hui il fait très \_\_\_\_\_* » le modèle pourra produire des probabilités comme : [« **beau** » : 0,4, « **chaud** » : 0,2, « **froid** » : 0,1, ...]. Finalement le modèle ajout le mot le plus probable et le processus est répété pour d'autres mots si nécessaire : *Aujourd'hui il fait très beau.*

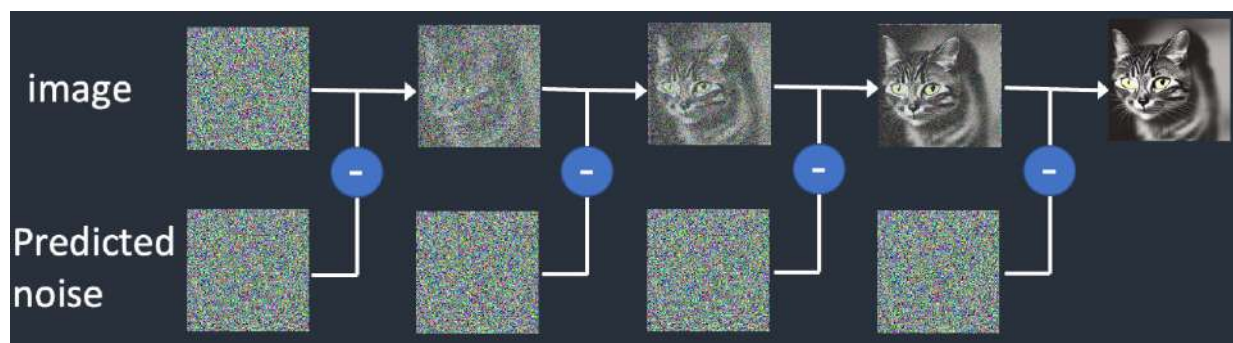
### Comment fonctionnent les modèles de diffusion ?

Les modèles de diffusion sont largement utilisés pour la génération d'images où ils sont particulièrement efficaces.

Ces modèles fonctionnent en 2 étapes : l'ajout de bruit (ajout de pixels aléatoires) et l'enlèvement de bruit (restaurer une image dégradée). Ces étapes sont itératives, c'est-à-dire qu'on ajoute et retire le bruit progressivement.

L'ajout de bruit permet d'entraîner le modèle sur une image pour un mot-clé donné.

Le retrait de bruit se fait étape par étape à partir d'une image composée à 100% de pixels aléatoires. En associant un modèle de diffusion avec un LLM, on se retrouve donc avec un outil capable de générer une image à partir de sa description (ou **prompt**).



## Les enjeux de l'IA

### Une opacité de fonctionnement

Un problème inhérent des réseaux de neurones vient du fait qu'une fois entraîné, même son créateur est incapable d'expliquer en détail son fonctionnement. Impossible en effet de déterminer quels neurones ou groupe de neurones est responsable de telle ou telle décision. Dès lors, comment faire confiance à une IA dont nous ne comprenons pas le fonctionnement ? Comment l'auditer, la réguler ou la réparer en cas de problème ?

### Des IA défectueuses ?

Des webcams qui détectent mal les personnes à la peau foncée ([HP](#)), un algorithme de prédiction de délit qui cible en priorité une population pauvre ou racisée ([police de Los Angeles](#)) ou encore un outil de recrutement qui écarte systématiquement les CV de femmes ([Amazon](#))... Tout cela est le résultat d'IA mal conçues entraînées sur des données biaisées ou incomplètes. Ces biais sont les reflets de nos propres biais, souvent inconscients et particulièrement difficiles à détecter.

Les chatbots ne savent pas dire « je ne sais pas ». Lorsqu'ils n'ont pas suffisamment de données sur un sujet, les IA inventent de toute pièces des réponses présentées avec aplomb : elles « [hallucinent](#) ». C'est ainsi qu'une IA imagine l'existence [d'œufs de vache](#), invente une [jurisprudence](#) ou propose de mettre de la [colle dans les pizzas](#).

### Les menaces sur notre vie privée numérique

L'entraînement des IA est extrêmement gourmand en données et les géants de la tech ne se gênent pas pour utiliser nos données pour alimenter celui-ci. Il est facile alors de concevoir des IA capables de nous profiler à partir de nos achats en ligne, de déterminer notre état de santé à partir de simples photos de profil, etc...

Comme souvent, les réglementations peinent à suivre le rythme de l'innovation et des pratiques, créant un vide juridique propice aux dérives.

## L'épidémie de la désinformation

Les IA génératives sont capables de créer à un rythme inédit des contenus toujours plus réalistes. C'est du pain bénit pour les créateurs de fake news. Si on associe à cela la tendance des réseaux sociaux à prioriser l'engagement à la vérité nous avons les conditions parfaites pour de la manipulation de masse.

## Des usages malveillants

Les IA sont des outils et comme tous les outils ils peuvent être utilisés à bon ou mauvais escient. Lorsqu'elle est utilisée pour nuire les IA peuvent faire des ravages :

- Des [arnaques](#) toujours plus sophistiquées avec l'aide des IA de traduction ou des chatbots.
- Des campagnes de désinformation à la carte avec des « [fermes de bots](#) » pour inonder les réseaux sociaux de contenus et commentaires.
- Des régimes autocratiques s'appuient sur la [surveillance algorithmique](#) pour contrôler leur population.

## Tous accros à l'IA ?

Avec le succès fulgurant de ChatGPT, l'IA est largement adoptée dans le monde universitaire : par exemple, 8 étudiants de [Sciences Po](#) sur 10 utilisent déjà l'IA générative, et 40 % l'utilisent quotidiennement. Si cet outil s'avère très pratique pour bien des usages (résumer un texte, reformuler une phrase, etc...) il est aussi très tentant de tout déléguer à l'IA. Et ce, d'autant plus que celle-ci est omniprésente dans nos usages du quotidien : sur Internet, nos téléphones, ordinateurs...). A trop déléguer nos décisions aux IA ne devenons nous pas dépendant de ces dernières ? D'autre part il ne faut pas oublier que derrière ces outils se trouvent des géants de la tech aux intérêts discutables.

## Des impacts sociétaux difficiles à évaluer

En contrôlant les infrastructures d'IA les plus performantes ces entreprises tech ont créés un nouveau monopole entraînant des inégalités entre ceux qui peuvent les payer et les autres. Un peu partout on annonce des vagues de licenciement liés au remplacement de postes par de l'IA sans création équivalente d'emplois.

En parallèle, on assiste à une nouvelle forme d'exploitation humaine : l'entraînement et l'alimentation des IA. Pour cela une armée de « [travailleurs du clic](#) » réalisent des tâches répétitives (identifier des objets, traduire du texte, nettoyer les données...) pour un salaire dérisoire, sans contrat, statut ou droit du travail.

Le monde artistique autrefois relativement épargné subit de plein fouet ces évolutions : une plateforme de diffusion musicale ([Spotify](#)) diffuse de la musique générée par IA au détriment des artistes, [ChatGPT](#) est capable de générer des dessins dans le style d'artistes connus. Est-ce vraiment l'usage que l'on souhaite pour l'IA ?

En raison de leur complexité et du volume de données à traiter, l'entraînement d'une IA à une [consommation énergétique colossale](#), toujours à la hausse : 1,5% de la demande mondiale en 2024 et le double prévu pour 2030. Ajoutons à cela la pollution liés à la fabrication des puces et équipements électroniques des ordinateurs faisant fonctionner ces IA, ce qui alourdi encore plus un bilan carbone préoccupant.

*Date de dernière mise à jour : novembre 2025*