



KOMPASS

BASES DE L'INTELLIGENCE ARTIFICIELLE

Fiche thématique

Bases de l'intelligence artificielle

Qu'est-ce que l'intelligence artificielle ?

L'intelligence artificielle (IA) désigne un ensemble de technologies numériques qui permettent d'exécuter automatiquement des tâches en utilisant des méthodes statistiques et algorithmiques. Ces systèmes reposent sur le traitement de données et permettent notamment l'analyse du langage, la classification d'images ou la génération de recommandations.

L'idée d'une intelligence artificielle remonte à l'Antiquité : certains mythes déjà évoquaient des « humains artificiels » ou des automates mécaniques. L'origine technique et scientifique de l'IA moderne se situe cependant au XX^e siècle.

Dans les années 1930 et 1940, des scientifiques comme Alan Turing ont posé les bases avec leurs théories sur la calculabilité et les machines capables de « penser ». En 1950, Turing a proposé le célèbre « test de Turing » : une machine peut-elle, par le langage, paraître aussi convaincante qu'un humain ?

Le terme « intelligence artificielle » (*Artificial Intelligence*) a été officiellement introduit en 1956 lors d'une conférence au Dartmouth College (États-Unis). Des chercheurs tels que John McCarthy, Marvin Minsky, Allen Newell et Herbert Simon ont développé les premiers programmes capables de résoudre des problèmes formalisés, notamment en mathématiques.

À partir des années 1980, l'accent davantage été mis sur les systèmes capables d'apprendre : des ordinateurs pouvant reconnaître de manière autonome des modèles dans des données. Avec l'explosion des volumes de données et le développement d'ordinateurs toujours plus puissants, l'apprentissage automatique a connu un essor notable au cours des dernières décennies (notamment depuis les années 2010), grâce aux réseaux neuronaux artificiels (voir page 2).

Aujourd'hui, l'IA est intégrée dans de nombreux services numériques : systèmes de recommandation, analyse automatisée d'images, outils de traduction, génération de contenus, etc. Elle ne « comprend » pas ses résultats, mais produit des sorties plausibles sur la base de calculs probabilistes. Ses performances dépendent largement de la qualité des données d'entraînement et des modèles utilisés.

Ainsi, l'IA doit être comprise comme un ensemble de techniques algorithmiques probabilistes, et non comme une « intelligence » au sens humain du terme.

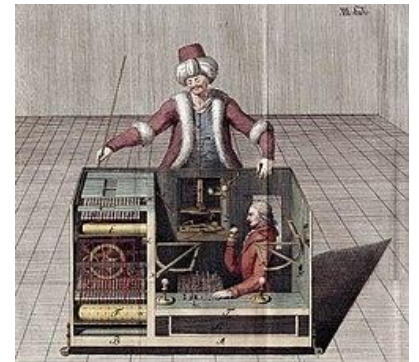


Figure 1: Le Turc, un robot d'échecs construit en 1769, secrètement manipulé par un humain.

Les sous-domaines de l'intelligence artificielle

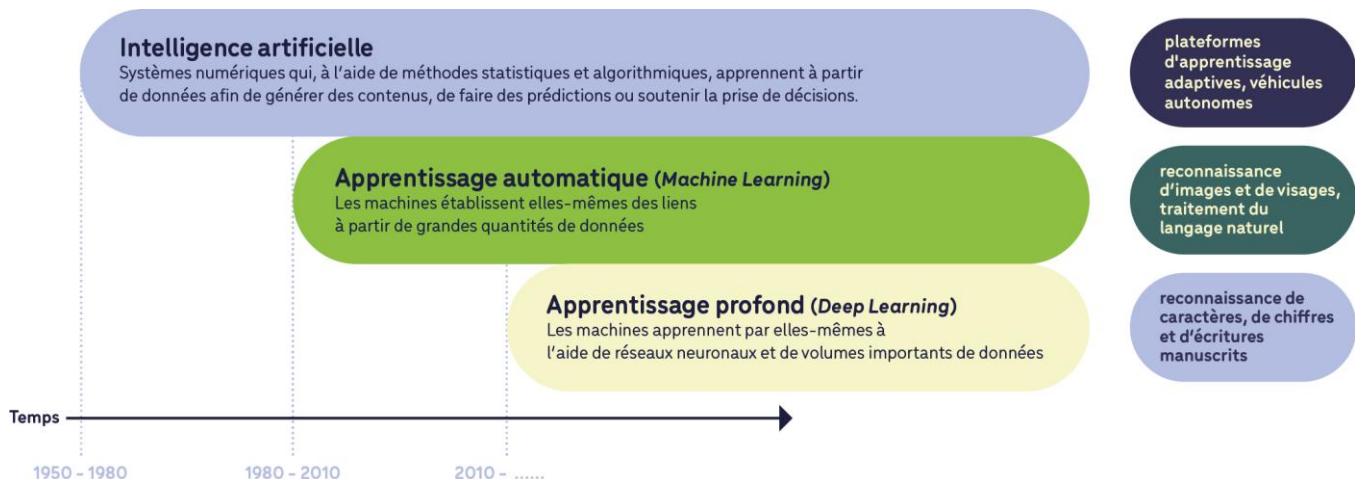


Figure 2: Développement historique des sous-domaines de l'IA

Apprentissage automatique	<p>L'apprentissage automatique regroupe différentes méthodes, parmi lesquelles :</p> <ul style="list-style-type: none"> • Apprentissage supervisé (<i>supervised learning</i>) : le système reçoit des données d'entraînement avec la « bonne réponse » (par exemple, une image montrant un chien, associée au terme « chien ») et apprend ainsi à les reconnaître. • Apprentissage non supervisé : le système détecte de manière autonome des similitudes ou des groupes dans les données (par exemple, des clients ayant des comportements d'achat similaires). • Apprentissage par renforcement : le système apprend par le biais de récompenses ou de sanctions, comme un robot qui trouve son chemin dans un labyrinthe en essayant différentes options.
Apprentissage profond	<p>L'apprentissage profond (<i>deep learning</i>) fait référence à des réseaux neuronaux artificiels constitués de nombreuses couches intermédiaires (<i>hidden layers</i>). Ces réseaux s'inspirent du fonctionnement du cerveau humain. Ils se composent de trois types de couches :</p> <ul style="list-style-type: none"> • une couche d'entrée (<i>input layer</i>), qui reçoit les stimuli de l'environnement sous forme de millions de points de données, • au moins une, mais généralement plusieurs couches intermédiaires cachées (<i>hidden layers</i>), qui réalisent les calculs, • et une couche de sortie (<i>output layer</i>), qui fournit le résultat.

Les trois types d'IA

Type	Description
IA faible	Exécute des tâches spécifiques (par exemple, assistants vocaux, reconnaissance d'images), mais ne dispose d'aucune autonomie.
IA forte	Capable d'accomplir diverses tâches et de transférer ses connaissances à de nouveaux domaines non prévus initialement. Par exemple, une IA conçue pour extraire du texte à partir d'images peut ensuite extraire du texte à partir de vidéos sans entraînement supplémentaire.
Superintelligence	À ce niveau, l'intelligence artificielle serait capable de s'améliorer elle-même. Elle pourrait se fixer ses propres valeurs et objectifs et s'adapter de manière flexible à différentes situations et environnements.

Cependant, l'intelligence artificielle actuelle n'en est encore qu'au stade de l'IA faible. Il existe toutefois de nombreux projets de recherche visant à atteindre une IA forte.

Pourquoi est-il important de s'informer sur l'IA ?

L'IA fait déjà partie intégrante de notre quotidien, par exemple dans les smartphones, les moteurs de recherche, les réseaux sociaux ou les systèmes de navigation.

Une compréhension fondamentale de l'IA est aujourd'hui indispensable pour utiliser les technologies de manière responsable dans l'enseignement et pour permettre aux apprenants d'utiliser l'IA de manière consciente et critique. Cette approche est également appelée « AI Literacy » (compétences en IA).

En ignorant les bases de l'IA, on court le risque de :

- ne plus être en mesure de comprendre certaines décisions : lorsque des algorithmes déterminent quel contenu est affiché ou qui obtient un crédit.
- céder son autonomie : sans connaissances de base, les recommandations ou les évaluations sont moins remises en question.
- accorder une confiance excessive : de nombreux systèmes d'IA semblent « intelligents », mais ils sont imparfaits – ils donnent parfois des résultats faux, injustes ou biaisés.
- adopter inconsciemment des préjugés : l'IA reprend des modèles issus des données. Sans en avoir conscience, des stéréotypes ou des discriminations peuvent être transmis involontairement (voir la fiche d'information sur les biais).
- divulguer des données personnelles : les systèmes d'IA collectent souvent des données dont le traitement ultérieur est difficile à évaluer, ce qui peut entraîner des problèmes en matière de protection des données.
- être manipulé : celui qui ne sait pas que les fils d'actualité des réseaux sociaux sont personnalisés par l'IA, par exemple, est plus réceptif à une influence ciblée ou aux bulles de filtres.

Inspirations pour l'enseignement

Les « cinq grandes idées en intelligence artificielle » présentent les fondements de l'IA de manière simple. Leurs lignes directrices (voir [AI4K12 – Grade Band Progression Charts](#)) montrent comment cette compréhension de base peut être acquise dès l'âge de trois ou quatre ans.

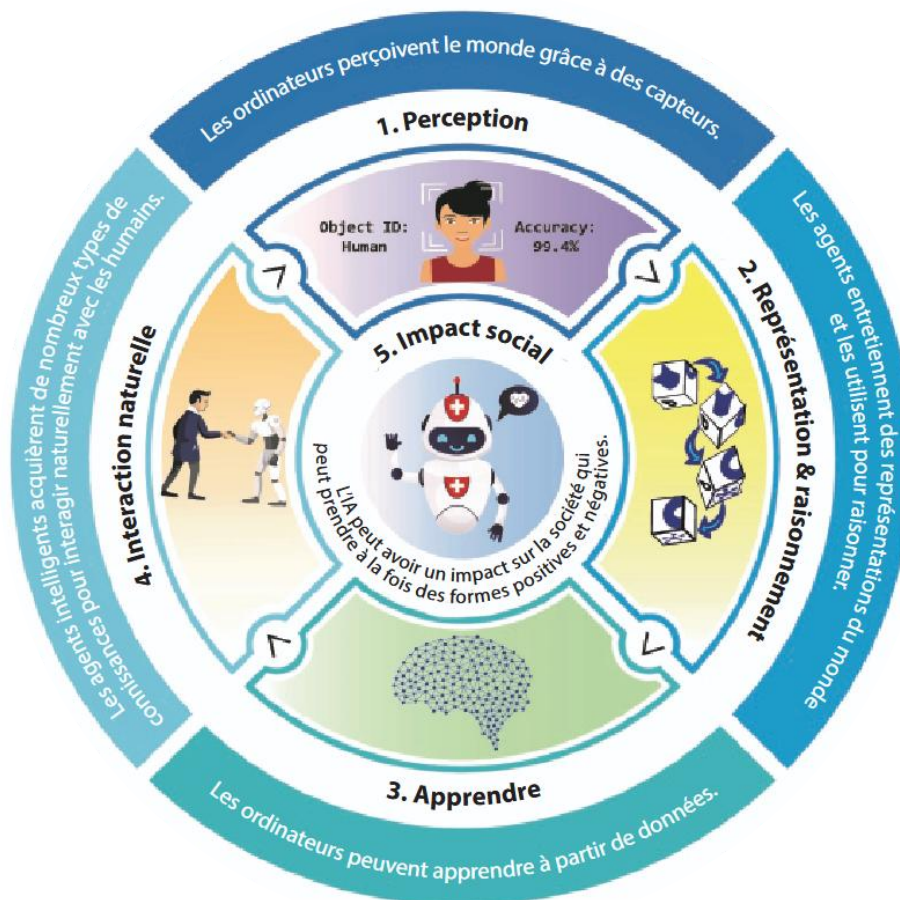


Figure 3: Source de l'image : [Les cinq grandes idées en intelligence artificielle](#)

Un enseignement de base de l'IA peut se faire avec ou sans outils numériques (« unplugged »). Voici quelques exemples :

- Le [jeu de Nim](#) comme illustration analogique de l'apprentissage par renforcement
- Vous trouverez d'autres exemples analogiques sur la classification, le deep learning et le test de Turing sur <https://www.aiunplugged.org>
- La [reconnaissance d'images comme exemple numérique d'apprentissage supervisé](#)

La plateforme [code.org](#) propose entre autres des ressources intéressantes sur l'IA.

Deux exemples pour encourager une réflexion critique sur les répercussions sociales de l'IA :



- <https://www.hownormalami.eu/>
- <https://theyseeyourphotos.com/>

Vous trouverez également de plus amples informations sur l'AI Literacy dans le [Medienkompass](#).