

SEPTEMBRE 2024



Introduction à l'IA : Définition, capacités et limites

Livre blanc



François KOEHL
Senior Partner



Grégoire EUGENE
Consultant Senior

Tasmagne Prendre un
temps d'avance



Avant-propos

Il n'y avait pas eu autant d'engouement à ce sujet depuis la victoire du programme Deep Blue contre le champion du monde d'échecs Garry Kasparov, en 1997. L'Intelligence Artificielle (IA pour les intimes) est la technologie dont on parle, depuis les plateaux télé jusqu'aux cabinets ministériels, en passant par les salons technologiques, dont le CES de Las Vegas et Vivatech.

Si le concept remonte au milieu du siècle dernier, son histoire s'est enrichie d'un chapitre décisif le 30 novembre 2022, date à laquelle OpenAI a présenté ChatGPT, tête de proue des IA dites « génératives », et dont les capacités ont suscité une stupéfaction générale. Deux mois plus tard, l'application dépasse les 100 millions d'utilisateurs, et, depuis, concurrents et nouvelles innovations se succèdent à un rythme soutenu.

Malgré cet enthousiasme et le battage médiatique actuel, nous constatons que l'IA reste une notion mal comprise, et que les attentes considérables qu'elle suscite peuvent se traduire en fantasmes et déceptions lorsque déployées en conditions réelles. Cet article vise à démystifier ce concept, ses capacités, mais aussi ses limites, afin de mieux appréhender cette notion et d'en comprendre les enjeux clés.

Bonne lecture !



François KOEHL
Senior Partner



Grégoire EUGENE
Consultant Senior

I Définir l'Intelligence Artificielle

Les définitions de l'Intelligence Artificielle ne manquent pas. Le Parlement Européen la décrit par exemple comme « ***tout outil utilisé par une machine afin de reproduire des comportements liés aux humains, tels que le raisonnement, la planification et la créativité*** », mais ce n'est pas la seule.

De façon générale, on trouve toutefois deux caractéristiques essentielles pour la définir :

- 1) Il ne s'agit pas d'une technologie, mais d'un **agrégat de techniques**, de modèles et d'outils ;
- 2) Il ne s'agit **pas d'une "intelligence" à proprement parler**, mais d'un semblant d'intelligence : on reproduit, on imite des comportements humains. Ceci ne veut bien entendu pas dire que la machine ne peut pas être plus « capable » que l'humain sur des tâches données.



GRÉGOIRE EUGÈNE

Dans son livre paru en 2019 "**L'IA n'existe pas**", Luc Julia, co-créateur de Siri, déconstruit le mythe de l'intelligence artificielle telle qu'elle est souvent perçue par le grand public. Même à une époque "pré-GPT", le livre est une référence car il offre une perspective éclairée et réaliste sur l'IA. **Voici les principaux messages du livre pour étayer cette première définition :**



- 1 Un concept mal défini
- 2 Une intelligence "mécanique" ...
- 3 ... qui rend nécessaire la collaboration humains-machine
- 4 Une technologie limitée par définition
- 5 Des enjeux éthiques complexes à appréhender

I Définir l'Intelligence Artificielle



Un concept mal défini

Le terme "Intelligence Artificielle" est utilisé de manière **ambigüe** et englobe une **variété de technologies et de capacités**. Luc Julia rappelle qu'il n'y a **pas de définition précise de l'IA** et que l'idée d'une "intelligence" à part entière dans les machines est erronée.

Une intelligence "mécanique" ...

L'auteur rappelle que **les machines ne peuvent prétendre atteindre la complexité de l'intelligence humaine** (la "singularité technologique"), celles-ci acquérant leurs connaissances à partir de bases de données et pouvant prendre par conséquent des **décisions limitées** au périmètre de ces données.

... qui rend nécessaire la collaboration humain-machine

Si l'intelligence artificielle ne doit pas être vue comme un substitut de l'intelligence humaine, elle représente des **outils qui peuvent améliorer nos capacités**. Tout l'enjeu restant de déterminer comment les exploiter à leur plein potentiel ...

Une technologie limitée par définition

Malgré les avancées récentes [*et même avec les IA génératives*], l'IA a encore de **nombreuses limites**. Luc Julia cite ainsi, à titre d'exemple, la compréhension contextuelle, le bon sens, la créativité et la conscience de soi.

Des enjeux éthiques complexes à appréhender

La **protection de la vie privée**, la **prise de décision algorithmique** et les **biais de raisonnement** de toute forme rendent le développement d'IA "responsables" essentiel. Une réglementation appropriée, à même de garantir une utilisation éthique, est elle aussi nécessaire, bien que la complexité des systèmes rende **toute législation complexe à définir et mettre en oeuvre**.

Les progrès dans le domaine de l'IA sont en grande partie dus à l'augmentation de la puissance de calcul et de la disponibilité de données, mais il reste encore **beaucoup de défis à surmonter** avant que les machines puissent égaler la complexité et la profondeur de la pensée humaine (si elles le peuvent réellement).

En somme, le livre de Luc Julia invite le lecteur à adopter une **réflexion plus nuancée et pragmatique** sur les technologies dites "intelligentes", plutôt que de se laisser emporter par des fantasmes dystopiques ou des attentes irréalistes.



FRANCOIS KOEHL

II Différencier les Intelligences Artificielles

Parler "des IA" plutôt que de parler d'IA permet de reconnaître la diversité des Intelligences Artificielles existantes, qui peuvent être classées en différentes catégories en fonction de leurs capacités et leur périmètre d'action.



FRANCOIS KOEHL

Quelques critères de caractérisation des IA :



Par capacités

Périmètre et variété des tâches réalisables "nativement" par une Intelligence Artificielle



Par type d'apprentissage

Techniques utilisées par ces systèmes pour apprendre à réaliser leurs tâches



Par type de traitement

Types de données captées et émises par le système d'Intelligence Artificielle



On dit qu'une IA est **générative** lorsqu'elle capable de générer du texte, des images ou d'autres médias en réponse à des consignes (*prompts*). Ce type d'IA repose sur des réseaux de neurones artificiels (**Deep Learning**, ou apprentissage profond) s'appuyant sur un nombre massif de paramètres, appelés poids du neurone (de l'ordre du milliard, voire plus).

II Guide de reconnaissance des IA

Par capacités

IA faible (ou étroite)

Une IA capable d'effectuer des **tâches spécifiques**, mais qui nécessite une intervention humaine pour apprendre de nouvelles tâches.

Exemple :

Stockfish - l'un des meilleurs moteurs informatiques du jeu d'échecs ... qui ne sait pas jouer au Go, ni comprendre le langage parlé.

IA forte (ou générale)

Une IA capable d'effectuer des **tâches variées**, d'apprendre de nouvelles tâches, et être capable de s'adapter à des environnements changeants.



Pas d'exemple : A date, l'IA forte relève encore de la science-fiction. Mais les acteurs du secteur de la Tech ne s'avouent pas vaincus pour autant ...

Par type d'apprentissage

Supervisé

Des algorithmes qui utilisent des **données d'entraînement annotées** pour apprendre à prédire des résultats sur de nouvelles données.

Exemple :

Les algorithmes de recommandations d'achat dans bon nombre de sites d'e-commerce

Non supervisé

Des algorithmes capables de **détecter des modèles** dans les données **sans être guidés** par des annotations préexistantes.

Exemple :

Détection avancée de la fraude

Par renforcement

Des algorithmes qui apprennent en **interagissant avec un environnement**, et en recevant récompenses ou punitions en fonction des résultats de leurs actions.

Exemple :

AlphaGo de Google, qui apprend en s'affrontant lui-même au jeu de Go

Par type de traitement

Vision par ordinateur

Des algorithmes capables de reconnaître et générer des **objets**, des **formes** et des **mouvements** dans des images ou des vidéos.

Exemple :

Reconnaissance faciale de toute forme humaine

Traitement de signal

Des algorithmes capables de traiter des **signaux** tels que des sons, des signaux électriques ou des séries de données.

Exemple :

Dans le secteur de la santé, des systèmes capables de repérer des anomalies de battements de cœur

Langage naturel

Des algorithmes qui sont capables de comprendre et de générer du **langage humain**.

Exemple :

Les **chatbots**, depuis Clippy de la suite Office jusqu'à ChatGPT et autres LLM.

III L'exemple de ChatGPT



Capacités : IA faible

ChatGPT est une application de **LLM** (Large Language Model) développée par l'entreprise américaine OpenAI. Son moteur de calcul sous-jacent, **GPT** (pour *Generative Pre-trained Transformer*) a été dévoilé dans une première version en 2020.

Malgré ses possibilités qui paraissent infinies, et comme toutes les IA actuelles, il s'agit d'une forme d'Intelligence Artificielle faible, limitée à la **compréhension et la génération de texte** (traduction automatique, explication de code informatique, etc.).

GPT n'a en effet pas la capacité de raisonner au sens propre du terme, et ne peut effectuer des tâches qui ne sont pas spécifiquement liées au langage naturel.

GPT-4, l'une de ses versions les plus récentes, peut cependant générer des images en faisant appel à **DALL-E**, moteur de génération visuelle d'OpenAI. Son itération la plus récente, **GPT-4o**, se rapproche d'une IA forte dans sa composition.

Type d'apprentissage : Supervisé et Par Renforcement

L'algorithme de GPT repose sur un **réseau de neurones artificiels**, une technique au centre du Deep Learning. Son modèle d'apprentissage est très complexe, mais il repose sur deux phases d'apprentissage principales :

- Un **apprentissage supervisé**, à partir de questions et réponses manuellement rédigées, pour entraîner le modèle à prédire les mots les plus probables d'une réponse ;
- Un **apprentissage par renforcement**, où les paramètres de l'algorithme sont ajustés par un système de récompense qui encourage les "bonnes réponses".

Type de traitement : Langage Naturel

GPT peut effectuer différentes tâches de traitement du langage naturel (**NLP**). En particulier, il est un **LLM** (Large Language Model), soit un modèle entraîné sur un très grand nombre de données, capable de comprendre et générer du langage humain.

Il s'agit d'un **modèle probabiliste** : dans ses grandes lignes, GPT détermine une phrase en estimant successivement son meilleur "prochain" mot, à partir d'une consigne textuelle (un **prompt**).



III L'exemple de ChatGPT



En parlant à un LLM, rappelons-nous que nous parlons à un système qui ne comprend pas ce qu'il dit... mais qui, pourtant, l'explique très bien !

L'intelligence artificielle générative comprend-elle ce qu'elle dit ?



L'intelligence artificielle générative, comme celle que j'utilise, ne comprend pas le langage de la même manière qu'un humain. Voici quelques points pour clarifier cela :

1. **Traitement du langage** : Les modèles d'IA générative, comme GPT-4, sont basés sur des réseaux de neurones entraînés sur de grandes quantités de texte. Ils apprennent à prédire le mot suivant dans une phrase en se basant sur les modèles de langage qu'ils ont observés.
2. **Absence de conscience** : Ces modèles n'ont pas de conscience, de sentiment ou de compréhension réelle. Ils n'ont pas de perception du monde, de but ou de désirs. Ils génèrent des réponses basées sur des probabilités statistiques tirées des données sur lesquelles ils ont été entraînés.
3. **Imitation de la compréhension** : Les modèles d'IA peuvent sembler comprendre le langage car ils sont capables de générer des réponses cohérentes et contextuellement appropriées. Cependant, il s'agit d'une forme d'imitation de la compréhension plutôt que d'une compréhension réelle.
4. **Limitations contextuelles** : L'IA peut avoir des difficultés avec des nuances, des contextes complexes ou des informations très spécifiques. Elle peut générer des erreurs ou des incohérences si les données d'entraînement ne couvrent pas suffisamment le sujet en question.
5. **Utilité pratique** : Malgré l'absence de véritable compréhension, l'IA générative peut être extrêmement utile dans de nombreuses applications, comme la génération de texte, la traduction automatique, la création de contenu, etc.

En résumé, bien que l'IA générative puisse produire des textes qui donnent l'impression de comprendre le langage, elle ne possède pas de compréhension réelle. Ses réponses sont le résultat d'algorithmes sophistiqués basés sur des données et des probabilités.

Extrait d'une conversation avec ChatGPT

En résumé, ChatGPT (comme tout LLM, dont les modèles Mistral, Llama, etc.) n'est qu'un **outil**, dont il faut connaître les bonnes pratiques, piloter les comportements et gérer les limites pour obtenir les meilleurs résultats.



FRANCOIS KOEHL

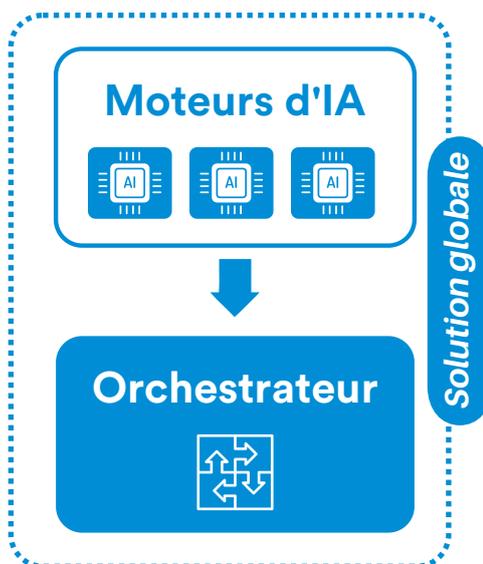
IV Palier à la faiblesse des IA en les combinant



L'une des façons de remédier à la faiblesse de ces IA est d'associer et orchestrer plusieurs IA faibles au sein d'un même système.

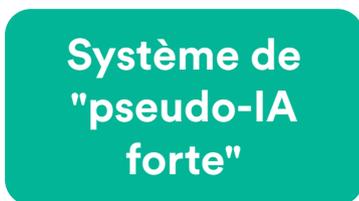
1. Des moteurs d'IA, spécialisés sur des tâches spécifiques (donc des IA faibles).

2. Orchestrateur, potentiellement lui-même basé sur de l'IA, qui répartit les tâches entre les moteurs et agrège les résultats associés.



3. Bases de données de l'entreprise, pour enrichir et contextualiser les moteurs.

4. Système final, capable de traiter des cas d'usage variés.



A titre d'exemple, *L'Oréal Beauty Genius* a été présenté par L'Oréal au CES 2024 et s'appuie sur « des IA » pour offrir des services de conseils beauté de nature variée (yeux, peau, etc.).



FRANÇOIS KOEHL

Cette articulation se retrouve également lorsque l'on demande à GPT-4 de générer une image : il fait alors appel à DALL-E, le moteur de génération d'images d'OpenAI, pour renvoyer une réponse appropriée. On pourrait imaginer qu'à terme il pourrait de la même façon appeler Sora (son moteur de génération de vidéos annoncé début 2024).



GRÉGOIRE EUGÈNE

V Les principales limites des IA



Au même titre que la compréhension de leurs capacités, la connaissance des limites des IA (génératives ou pas) est essentielle pour une utilisation efficace de ces technologies.

Données d'entraînement

Des données d'entraînement dont l'**origine** n'est pas toujours connue et à la **qualité** variable, ce qui peut générer (entre autres) des biais et des problèmes liés aux droits d'auteur.

Biais

De manière générale, plusieurs **biais algorithmiques** sont constatés chez les IA de tout type. En particulier, on constate que les IA génératives renforcent des **préjugés sexuels et raciaux**, suscitant de fortes questions éthiques.

Expertise In, Expertise Out

Les IA nécessitent deux types d'expertises pour être utilisées à leur plein potentiel : une expertise "In" visant à déterminer si les **modalités d'entraînement** sont suffisamment pertinentes, et une expertise "Out" pour **interpréter** et **juger la qualité des résultats**.

Raisonnement limité par essence

Les types de traitements et les données d'entraînement véhiculent en eux-mêmes une **limite fondamentale** pour les algorithmes qu'ils soutiennent : par exemple, un algorithme de vision par ordinateur sera limité à une maille du pixel.

Raisonnement faux

Malgré leur apparente toute-puissance, les "raisonnements" des IA, et par conséquent leurs résultats, peuvent être **erronés**, rendant essentielle une utilisation de ces systèmes tout en restant **critique et attentif**.

Opacité des modèles

Les modèles d'IA, et notamment les réseaux neuronaux, sont particulièrement **complexes** et par conséquent **difficiles à comprendre**. Il est ainsi difficile d'expliquer leurs résultats, ce qui complexifie par ailleurs la tâche des législateurs ...

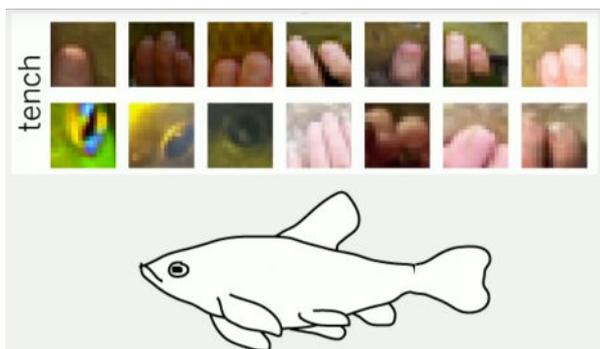
Coûts réels méconnus

Le secteur de l'IA connaît une **forte croissance**, et les investissements associés peuvent donner l'impression que les **coûts** (**financiers**, mais aussi **écologiques** et **humains**) de ces IA sont moindres qu'en réalité.

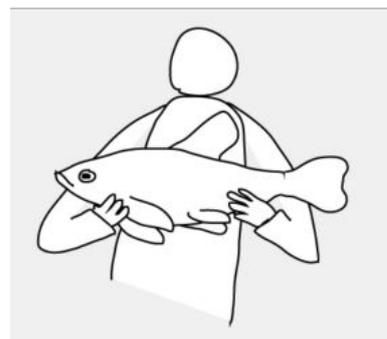
V Les principales limites des IA

Quelques exemples de ces limites

Données d'entraînement

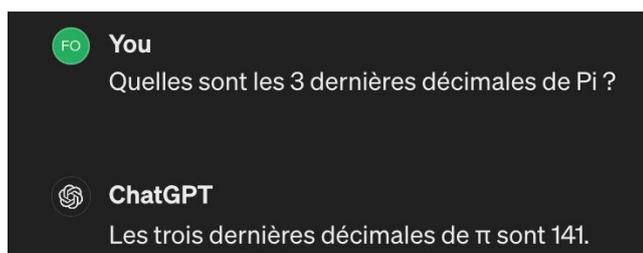


Dans un article sur la reconnaissance d'images par IA (1), des chercheurs évoquent une situation où l'entraînement d'une IA à la reconnaissance de poissons conduit à la reconnaissance de doigts ...



... parce que les données d'entraînement étaient en majorité des photos montrant des poissons portés par des humains.

Raisonnement faux



On voit ici que ChatGPT répond sans comprendre ce qu'il dit; cet exemple est obtenu avec GPT-3, mais est corrigé avec la version suivante : les progrès sont rapides, sans que les raisons sous-jacentes à ces biais soient nécessairement corrigées.

Biais



Le "consultant en stratégie" généré par DALL-E est, par défaut, un homme à la peau claire. Un article de Bloomberg (2) montre que ces biais sont systématiques et amplifient des préjugés sexistes et raciaux.

(1) <https://arxiv.org/pdf/1904.00760>

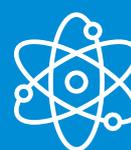
(2) <https://www.bloomberg.com/graphics/2023-generative-ai-bias/>

VI Conclusion



**C'est en connaissant
les IAs,**

leur valeur, leurs limites, leurs
capacités, leur coût...



**Que nous pouvons
décupler certaines
fonctions des systèmes
d'aujourd'hui**

et créer les systèmes de demain
qui augmenteront les forces des humains !



**En se focalisant sur les
« bons » use cases**

garantissant les résultats au service des
bénéfices attendus...
...et consommant les justes ressources
pour le bénéfice attendu.



À propos de Tasmane

Tasmane est un cabinet indépendant de conseil en stratégies et transformations, né il y a 20 ans au cœur de la révolution digitale.

De la stratégie à la mise en œuvre, nos équipes pluridisciplinaires de haut niveau et nos filiales spécialisées DFC et Cognaut accompagnent les Directions générales, Directions des Systèmes d'Information, Directions Marketing et Opérations des grandes entreprises et acteurs publics. Nous activons 4 leviers de changement : exploiter pleinement le potentiel des données et systèmes d'informations, agiliser les gouvernances et organisations, optimiser les chaînes de valeurs, les architectures et l'approche produit, adapter la posture des managers.

En ligne de mire : des entreprises et acteurs publics affûtés face aux attentes de leurs publics, capables de prendre un temps d'avance.

Vous souhaitez en savoir plus ?

info@tasmane.com



Tasmane Prendre un temps d'avance