IA et Nous

Avenir possible
IA Générale ?
Obstacles

Yann Lecun

Conférence à l'ENS 2017

https://www.youtube.com/watch?v=9ajwtKWH8ng

Aller à 1h28'20"

Les modèle génératifs sont ils intelligents ?

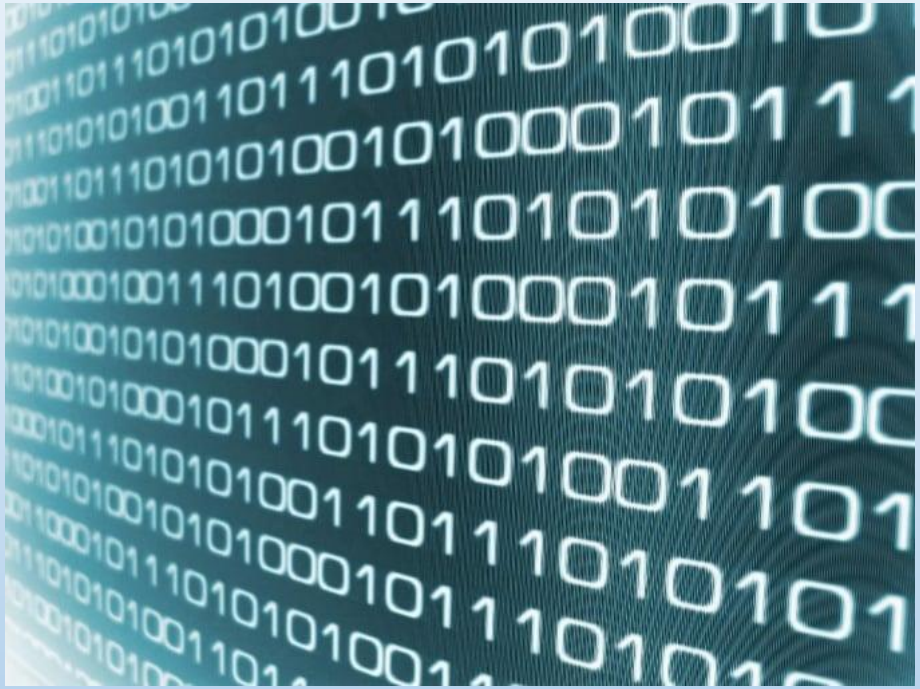
Discussion

- *Quelles sont leurs forces?*
- *Les points clés ?*
- *Leurs faiblesses ?*

IA générale
ou
Super Intelligence ?

IA générale ou Super Intelligence ?

Réaliste



Science Fiction ?

IA générale
ou
IA Hybride

Les deux approches de l'IA

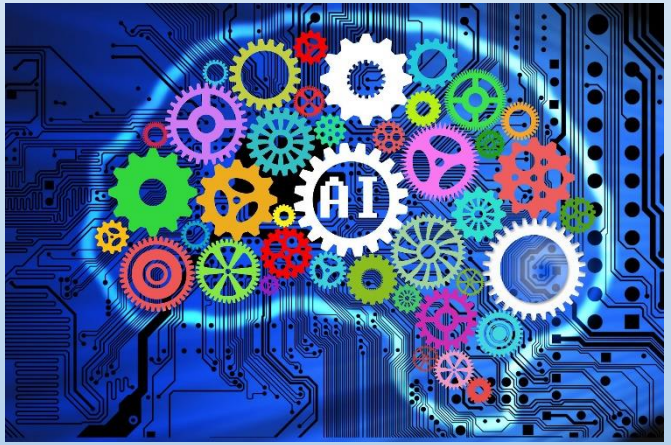
Quand on veut modéliser un système, deux voies :

Symbolique

modéliser le comportement

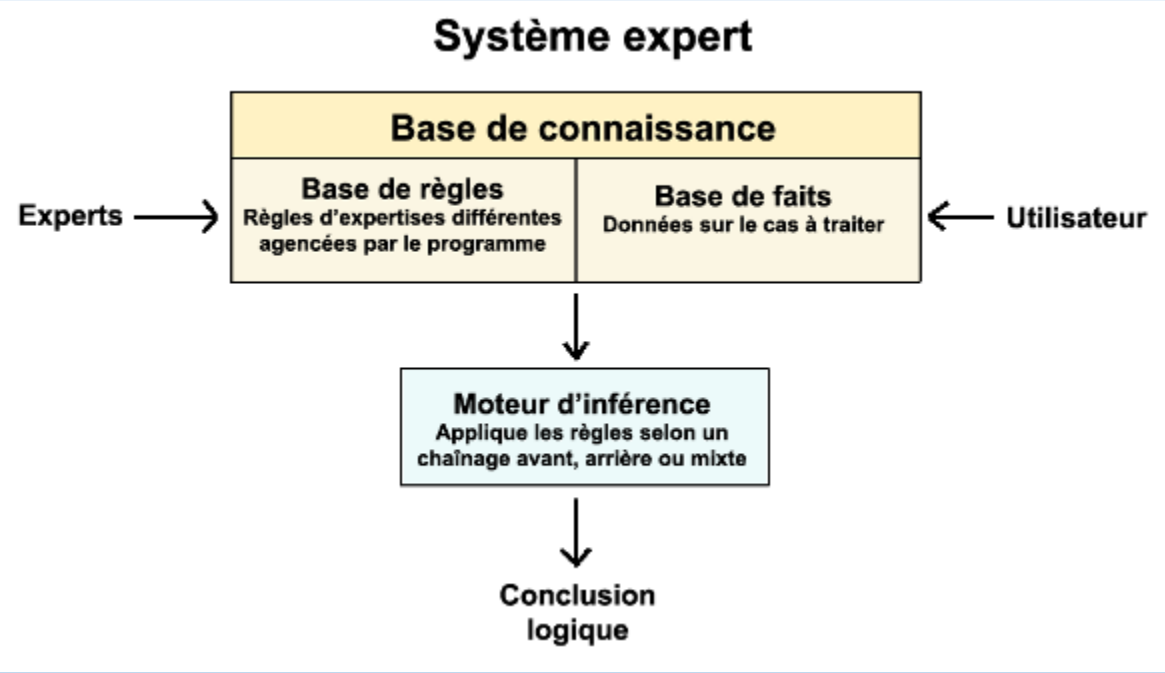
Connexionniste

modéliser les mécanismes



Synthèse : les deux approches

- ***Systèmes logiques***



- ***Systèmes d'apprentissage***

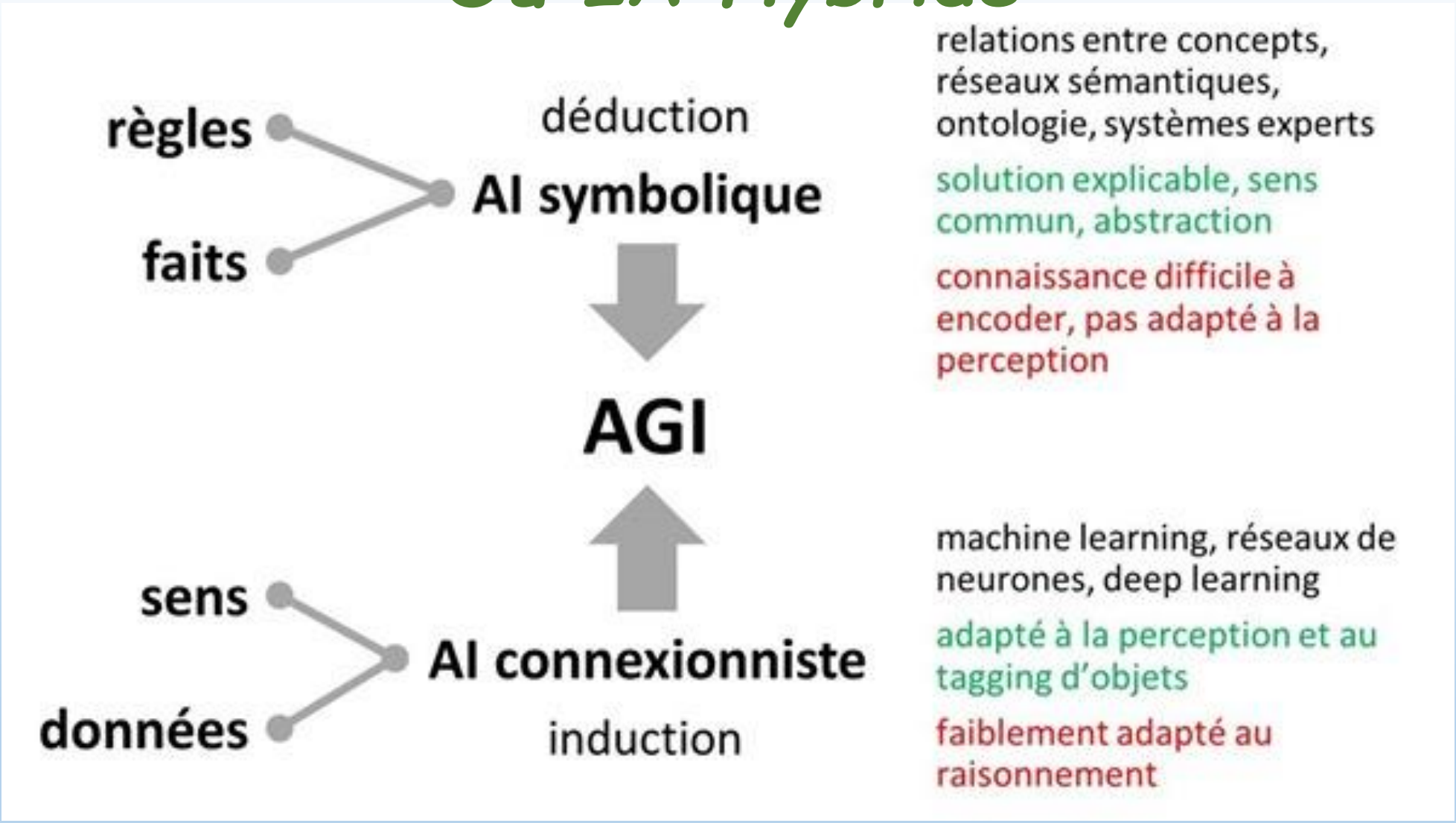
-Beaucoup d'efforts

-Traçable et maîtrisé

- Nécessite beaucoup de données : Big Data

- On ne peut plus expliquer la décision ou le résultat

Vers l'IA généralisée
Ou IA Hybride



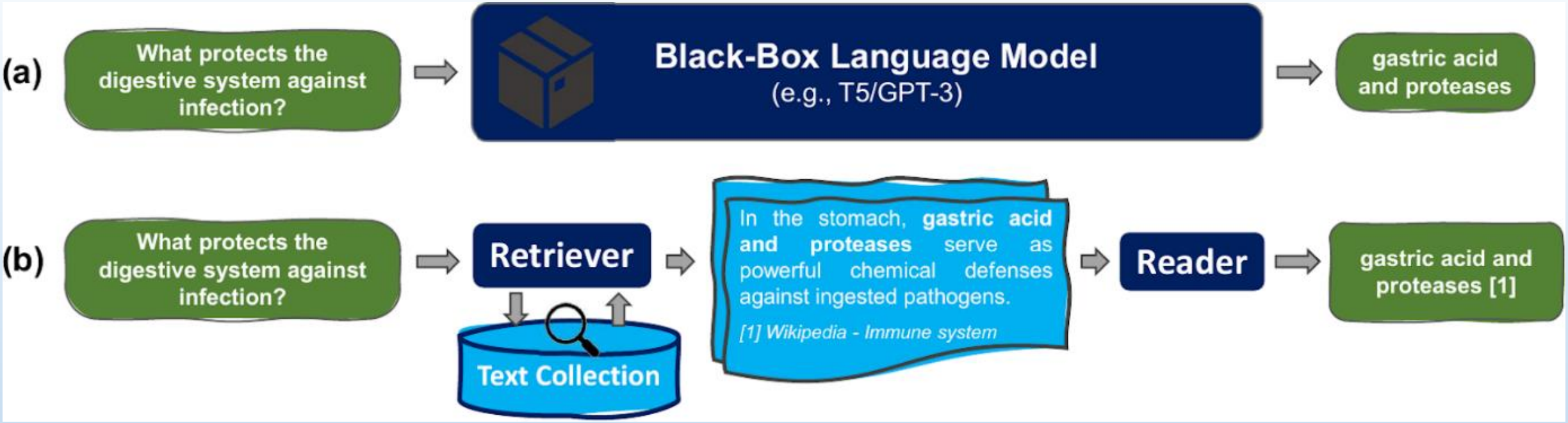
Vers l'IA généralisée ?

Opportunité fabuleuse pour l'Europe, et plus particulièrement la France.

Notre pays compte des chercheurs de haute qualité appartenant aux différentes communautés de l'IA.

*Selon **Nature**, la France figure au **5ème rang mondial** de la recherche dans ce domaine*

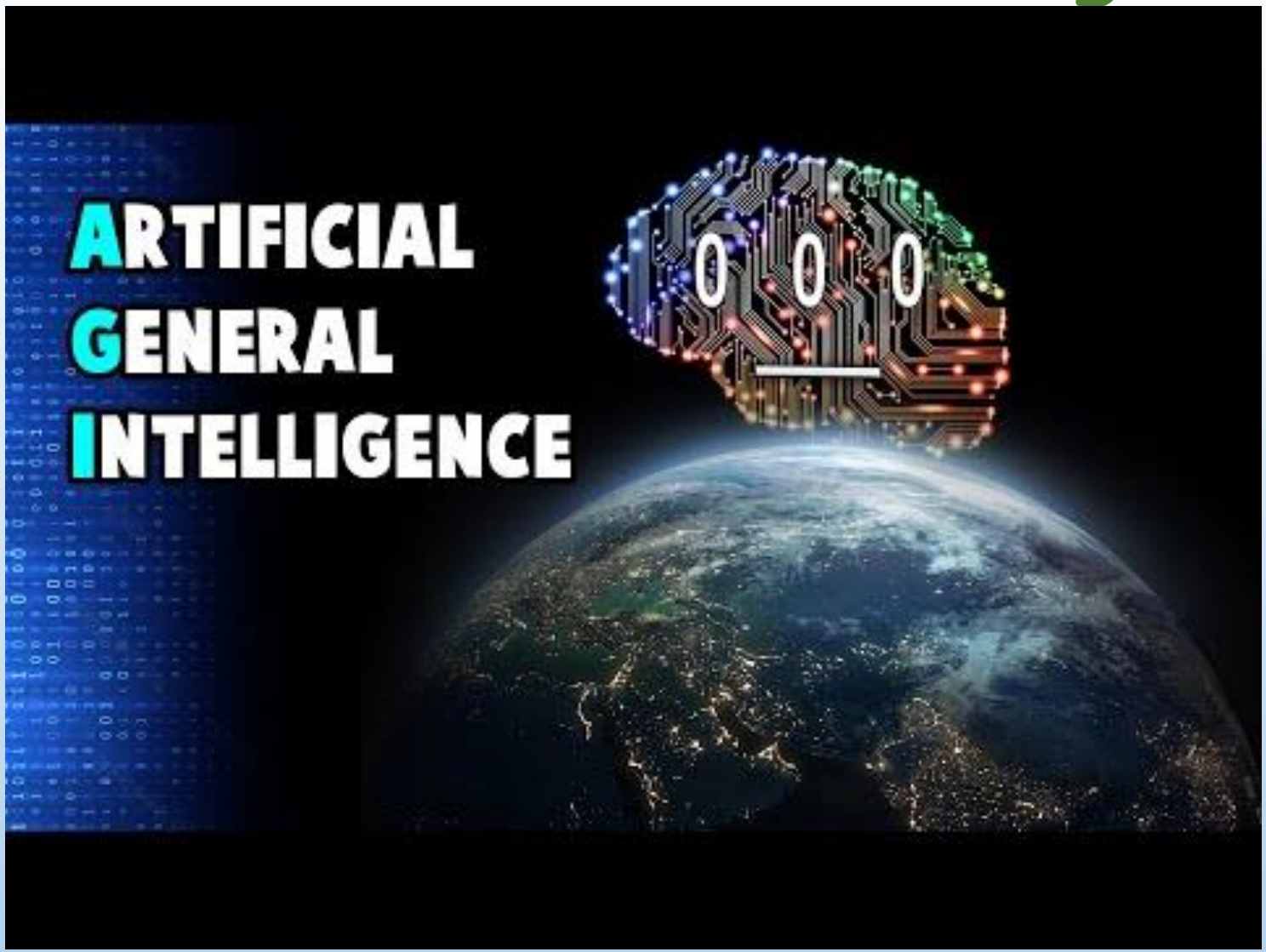
Exemple : NLP Systèmes de récupération



Source : Stanford
Construire des modèles de PNL évolutifs, explicables et adaptatifs avec récupération
Exemple ***ColBERT*** *pour l'extension de la récupération expressive à des corpus massifs via une interaction tardive,* ***ColBERT-QA*** *pour répondre avec précision aux questions du domaine ouvert*

IA générale
ou
Super Intelligence ?

Autre vision de l'IA générale



Source : Le Big Data - Bastien L 19 mai 2023 Dossiers, Intelligence artificielle


CHATGPT
AGI

IA générale : l'AGI

Une intelligence artificielle générale (AGI) est une IA comparable au cerveau humain :

- *capable de transférer ses connaissances pour accomplir n'importe quelle tâche,*

- *dotée d'une conscience.*

Un tel système n'existe pas encore.

Certains pensent que des outils tels que ChatGPT s'en rapprochent de plus en plus.
Et vous ?

IA générale : l'AGI

Il s'agirait d'une ***IA dotée de toutes les capacités cognitives humaines****, notamment la conscience et la prise de décision.*

Une telle machine serait capable ***d'émuler l'esprit et le comportement humain*** *pour résoudre n'importe quel type de problème.*

Elle pourrait ***agir, réagir, ressentir et penser*** *comme nous le faisons.*

Elle serait capable de ***trouver une solution face à une tâche inconnue****, de la même manière et avec les mêmes performances qu'un humain. Aucune programmation, aucune intervention humaine ne serait nécessaire.*

L'AGI *permettrait à* ***une machine de comprendre, d'apprendre*** *et d'effectuer toutes sortes de tâches intellectuelles. C'est la principale différence avec une IA spécialisée comme celles que nous connaissons à ce jour.*

On désigne aussi parfois ***l'AGI par le terme « intelligence artificielle forte ».***

Elle s'oppose à une ***IA faible ou étroite****, pouvant uniquement être appliquée à* ***des problèmes spécifiques****.*

IA générale : l'AGI

*L**a définition d'une AGI varie** selon celle de l'intelligence humaine.*

*Les **scientifiques informatiques** la considèrent souvent comme la c**apacité à atteindre des buts.***

L**es **psychologues** la caractérisent par **l'adaptabilité et la survie.

Pour certains, une IA générale serait une machine dotée d'une forme de conscience.

*Elle serait capable de **planifier, d'acquérir des capacités cognitives, d'émettre des jugements, de gérer des situations incertaines** ou encore **d'intégrer ses connaissances pour la prise de décision.***

IA générale : l'AGI

*L**a définition d'une AGI varie** selon celle de **l'intelligence humaine.***

Les scientifiques informatiques** la considèrent souvent comme la c**apacité à atteindre des buts.

Les psychologues** la caractérisent par **l'adaptabilité et la survie.

Pour certains, une IA générale serait une machine dotée d'une forme de conscience.

*Elle serait capable de **planifier, d'acquérir des capacités cognitives, d'émettre des jugements, de gérer des situations incertaines** ou encore **d'intégrer ses connaissances pour la prise de décision.***

IA générale : l'AGI

Une AGI pourrait effectuer des tâches ***hors de portée des meilleurs ordinateurs existants*** *à ce jour.*

Elle pourrait accomplir tout ce que peut faire ***un humain*** *et aurait le* ***même potentiel que notre cerveau.***

Ce système devrait être capable de ***pensée abstraite, de connaissances basiques, d'un sens commun****, de* ***créativité****, de* ***sociabilité****, d'une* ***compréhension des liens de cause à effet****, ou encore d'un* ***apprentissage par transfert***

IA générale : l'AGI

Perception sensorielle

L'AGI serait capable de ***lire, de comprendre du code écrit par l'humain et de l'améliorer****.*

Une telle IA serait aussi très douée pour les tâches liées à ***la perception sensorielle comme la reconnaissance de couleur****.*

Même si les systèmes de Deep Learning ont permis d'importants progrès dans la vision par ordinateur (Computer Vision), les ***systèmes IA actuels sont loin d'égaler*** *les capacités de perception humaines.*

C'est la raison pour laquelle les ***véhicules autonomes sont facilement trompés*** *par du scotch noir ou des auto-collants sur un panneau stop. De même, les IA actuelles ne peuvent percevoir le son comme nous le faisons.*

IA générale : l'AGI

Raisonnement et transfert de connaissances

*Comme l'humain, **l'AGI pourra appliquer les connaissances** qu'elle acquiert à différentes circonstances.*

*Elle pourra aussi appliquer le **savoir accumulé afin de planifier pour le futur**.*

*C'est un point commun avec **l'humain qui use de son expérience** pour créer un **plan** et guider son **avenir.***

*Les AGI pourront **s'adapter à toutes les circonstances** et prendre des décisions à la volée.*

*Une **autre capacité sera le raisonnement**, l'analyse d'une situation pour déterminer le cours d'une action même si elle dépasse les limites de ce que l'humain lui a enseigné.*

IA générale : l'AGI

Motricité

Sa motricité serait aussi largement accrue,

Exemple : capacité d'attraper des clés dans une poche. Ceci implique un certain niveau de perception imaginative.

Une IA générale sera aussi capable de ***naviguer en projetant son mouvement*** *dans des espaces physiques avec encore plus de précision que les systèmes existants tels que le GPS.*

IA générale : l'AGI

Compréhension du langage et sens commun

Comprendre le sens des paroles humaines *en fonction du contexte grâce à un haut niveau d'intuition.*

Une caractéristique qui manque cruellement aux ***IA faibles*** *est* ***le sens commun****.*

Les ***IA actuelles*** *peuvent* ***inventer des faits****.*

L'IA générale ne devra pas avoir ce problème*.*

IA générale : l'AGI

Conscience et créativité

*Les IA devront pouvoir **prendre en charge** **divers types d'algorithmes d'apprentissage**, créer des structures fixées pour toutes les tâches, et comprendre les systèmes de symboles.*

*Elles devront pouvoir aussi utiliser **différents types de connaissances, comprendre les systèmes de croyances** ou encore s'engager dans la **« métacognition »** et utiliser les connaissances métacognitives.*

*L'IA générale pourrait même être **dotée d'une véritable conscience d'elle-même.***

*Elle devrait être capable de **percevoir les besoins, les émotions, les processus de pensée d'autres entités intelligentes***

IA actuelle : faible ou étroite

Une IA étroite : excellente *sur une* ***tâche spécifique****, comme jouer aux échecs, analyser le contenu d'images, ou répondre à la question d'un utilisateur.*

Fondées sur les ***technologies*** *passées en revue dans l'année.*

Ces technologies ***n'arrivent pas à la cheville du cerveau humain*** *pour les tâches globales dont on vient de faire l'inventaire qui exigent une capacité cumulative.*

Que serait une IA générale

*Une **IA forte devra effectuer une large diversité** de tâches et même apprendre par elle-même à résoudre de nouveaux problèmes.*

*Là où une **IA faible requiert l'interférence humaine** pour définir les paramètres de ses algorithmes d'apprentissage et lui fournir des données de haute qualité, **ce ne sera pas nécessaire pour une IA forte.***

*Ses performances seront **égales ou supérieures à celles des humains** pour la résolution de problèmes dans la plupart des domaines.*

Quand sera créée la première AGI ?

IA générale : à quel horizon ?

En 2017, ***le célèbre futuriste Ray Kurzweil*** *connu pour sa perspicacité estimait à la conférence SXSW que les ordinateurs* ***égaleraient l'intelligence humaine d'ici 2029****.*

Geoffrey Hinton, prédit qu'une IA générale pourrait émerger ***dans moins de 20 ans****.*

Le CEO d'Alphabet DeepMind, Demis Hassabis, estime que l'IA atteindra ***« une cognition de niveau humain » dans moins d'une décennie****.*

GPT-5 : Prévue fin 2023 *cette cinquième version aurait atteint* ***l'intelligence artificielle générale !***

GPT4
Premiers signes intelligence humaine selon Microsoft

« Sparks of Artificial General Intelligence », *13/04/2023*
Microsoft compare GPT-3 et GPT-4 (le nouveau Bing).

*La **réponse très pertinente de BingGPT4** à une tâche d'équilibrage a impressionné les chercheurs.*

- *L'IA était chargée d'indiquer comment empiler un livre, neuf œufs, un PC portable, une bouteille et un clou de façon stable.*
- *Si GPT-3.5 a proposé de poser les œufs sur le cloud, **GPT-4 a suggéré d'arranger les œufs en grille de 3 par 3** sur le livre pour que le laptop et le reste des objets puissent tenir dessus en équilibre.*

*Pour Microsoft, cette réponse correcte à un **puzzle exigeant une compréhension du monde physique** démontre un pas en avant vers **l'AGI.***

Q : Que faut-il en penser ?

Les dangers de l'IA générale

Avant même que l'intelligence artificielle ne devienne un domaine de recherche scientifique, les ***écrivains de science-fiction s'inquiétaient*** *des risques de dérives.*

Par exemple, ***Isaac Asimov a écrit les Lois de la Robotique en 1942****. Ces règles visaient à s'assurer que les* ***machines intelligentes restent bienveillantes****.*

Mai si l'IA générale pourrait apporter de nombreux bienfaits à l'humanité, elle ***pourrait aussi causer sa disparition****…*

Capacité à manipuler l'humain ?

En poursuivant ses propres buts et aspirations, guidés par ses besoins et sa volonté, l'IA générale pourrait ***décider que l'humain est un obstacle sur sa route*** *et entreprendre de s'en débarrasser.*

Et pour y parvenir, elle pourrait exploiter sa ***profonde compréhension de nos émotions****, de* ***nos comportements****, de la façon dont nous prenons des décisions. Ceci lui confèrerait la capacité de manipuler l'humain.*

Grâce à sa compréhension de notre fonctionnement, ***l'IA pourrait facilement jouer sur les sentiments****. Elle pourrait par exemple se fabriquer le corps d'un adorable animal pour inspirer la confiance.*

Exemple : la fabrique de trombones

Le grand remplacement du travail humain ?

Si l'intelligence artificielle générale ***surpassait l'humain dans tous les domaines****, l'impact sur le monde du travail serait sans précédent.*

D'innombrables personnes deviendraient ***inutiles, obsolètes, et se retrouveront au chômage****.*

Ce grand remplacement du travail humain est l'une des principales craintes liées à l'émergence de l'AGI.

Les robots AGI

*L'humain cherche à **confèrer à ces AGI un corps similaire ou supérieur** au sien.*

De nombreuses entreprises ont déjà commencé à concevoir des robots humanoïdes, à l'instar du Tesla Bot et des robots de Boston Dynamics.

*Ces robots pourraient alors **interagir avec le monde physique** de la même manière que nous. Les humains pourraient donc se retrouver **surpassés intellectuellement ET numériquement** par ces êtres artificiels*

Comment se Préparer et se protéger ?

*L'**humanité doit-elle se préparer** à l'apparition de ces créatures ?*

Si oui, comment ?

1. ***Régulation***

   *Les gouvernements doivent dès maintenant mettre en place **des règles et des lois** pour encadrer **l'utilisation et le développement d'IA…***

2. ***Formation Information***

   Nécessité de conserver la maîtrise du développement (Open Source)

   *Pilotage par les **usages/besoins. Ces IA doivent rester des outils pour humains***

   Limitations de l'autonomie des IA et robots

3. ***Education***

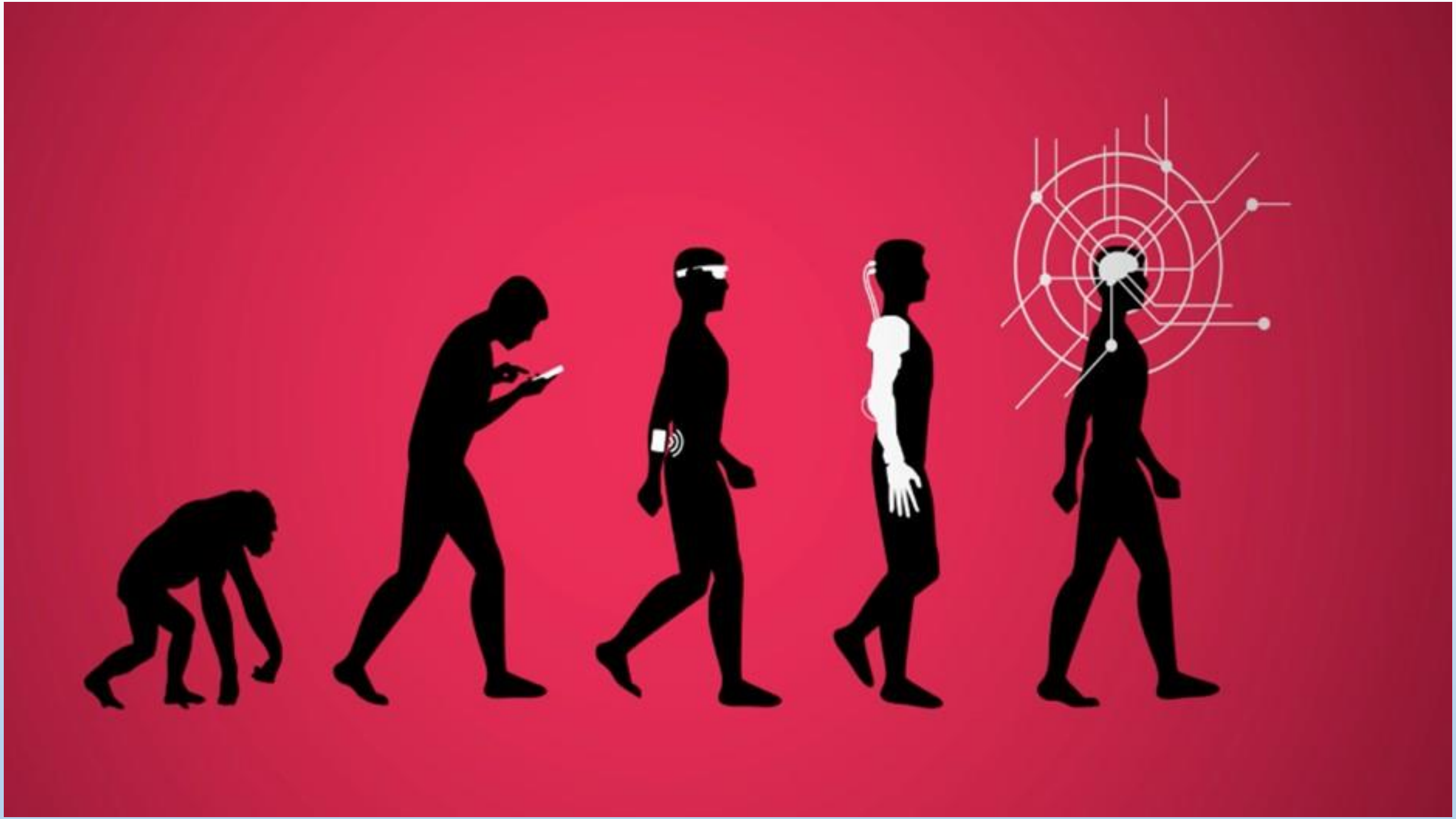
   *Gros effort sur l'école pour développer dès le plus jeune âge **les fonctions cognitives de TOUS les enfants***

Bienfaits de ces IA ? ?

Si beaucoup d'experts redoutent que l'IA provoque un cataclysme, d'autres y voient une technologie qui pourrait ***résoudre les plus grands problèmes de l'humanité****.*

Selon ***Sam Altman, CEO d'OpenAI****, l'IA générale pourrait notamment*

- *accélérer le développement de la fusion nucléaire,*
- *résoudre la crise énergétique,*
- ***soigner des maladies considérées comme incurables,***
- ***…..***



Les Obstacles

Les 5 murs de l'IA


Crédit © JM Prima

Bertrand Braunschweig
Consultant
Ancien Directeur de recherche INRIA
2021

Mur de la confiance
Mur de la sécurité
Mur de l'énergie
Mur de l'interaction avec l'humain
Mur de l'inhumanité

Mur de la confiance

Dialogue entre Lonia, le chatbot de la banque et Y, qui a demandé un crédit.

Y: est-ce que mon prêt a été accordé ?

Lonia: non.

Y: peux-tu me dire pourquoi mon prêt n'a pas été accordé?

Lonia: non

Y: mais, pourquoi ne peux-tu pas me dire pourquoi mon prêt n'a pas été accordé?

Lonia: parce que je suis une intelligence artificielle, entraînée à partir de données de crédits passés, et je ne sais pas produire d'explications.

Y: c'est bien dommage! Mais peux-tu au moins prouver que ta décision est la bonne?

Lonia: Non, on ne peut pas prouver les conclusions établies par des IA entraînées par apprentissage à partir de données.

Y: ah, bon. Mais, alors, as-tu été certifiée pour le travail que tu fais? As-tu un quelconque label de qualité?

Lonia: Non, il n'existe pas de normes pour les IA entraînées par apprentissage, il n'y a pas de certification.

Y: Merci pour tout cela. Au revoir, je change de banque.

Mur de la confiance

Dialogue entre Lonia, le chatbot de la banque et Y, qui a demandé un crédit.

Y: est-ce que mon prêt a été accordé ?

Lonia: non.

Y: peux-tu me dire pourquoi mon prêt n'a pas été accordé?

Lonia: non

Y: mais, pourquoi ne peux-tu pas me dire pourquoi mon prêt n'a pas été accordé?

Lonia: parce que je suis une intelligence artificielle, entraînée à partir de données de crédits passés, et je ne sais pas produire d'explications.

Y: c'est bien dommage! Mais peux-tu au moins prouver que ta décision est la bonne?

Lonia: Non, on ne peut pas prouver les conclusions établies par des IA entraînées par apprentissage à partir de données.

Y: ah, bon. Mais, alors, as-tu été certifiée pour le travail que tu fais? As-tu un quelconque label de qualité?

Lonia: Non, il n'existe pas de normes pour les IA entraînées par apprentissage, il n'y a pas de certification.

Y: Merci pour tout cela. Au revoir, je change de banque.

Mur de la confiance

Dialogue entre Lonia, le chatbot de la banque et Y, qui a demandé un crédit.

Y: est-ce que mon prêt a été accordé ?

Lonia: non.

Y: peux-tu me dire pourquoi mon prêt n'a pas été accordé?

Lonia: non

Y: mais, pourquoi ne peux-tu pas me dire pourquoi mon prêt n'a pas été accordé?

Lonia: parce que je suis une intelligence artificielle, entraînée à partir de données de crédits passés, et je ne sais pas produire d'explications.

Y: c'est bien dommage! Mais peux-tu au moins prouver que ta décision est la bonne?

Lonia: Non, on ne peut pas prouver les conclusions établies par des IA entraînées par apprentissage à partir de données.

Y: ah, bon. Mais, alors, as-tu été certifiée pour le travail que tu fais? As-tu un quelconque label de qualité?

Lonia: Non, il n'existe pas de normes pour les IA entraînées par apprentissage, il n'y a pas de certification.

Y: Merci pour tout cela. Au revoir, je change de banque.

Mur de la confiance

Dialogue entre Lonia, le chatbot de la banque et Y, qui a demandé un crédit.

Y: est-ce que mon prêt a été accordé ?

Lonia: non.

Y: peux-tu me dire pourquoi mon prêt n'a pas été accordé?

Lonia: non

Y: mais, pourquoi ne peux-tu pas me dire pourquoi mon prêt n'a pas été accordé?

Lonia: parce que je suis une intelligence artificielle, entraînée à partir de données de crédits passés, et je ne sais pas produire d'explications.

Y: c'est bien dommage! Mais peux-tu au moins prouver que ta décision est la bonne?

Lonia: Non, on ne peut pas prouver les conclusions établies par des IA entraînées par apprentissage à partir de données.

Y: ah, bon. Mais, alors, as-tu été certifiée pour le travail que tu fais? As-tu un quelconque label de qualité?

Lonia: Non, il n'existe pas de normes pour les IA entraînées par apprentissage, il n'y a pas de certification.

Y: Merci pour tout cela. Au revoir, je change de banque.

Mur de la sécurité

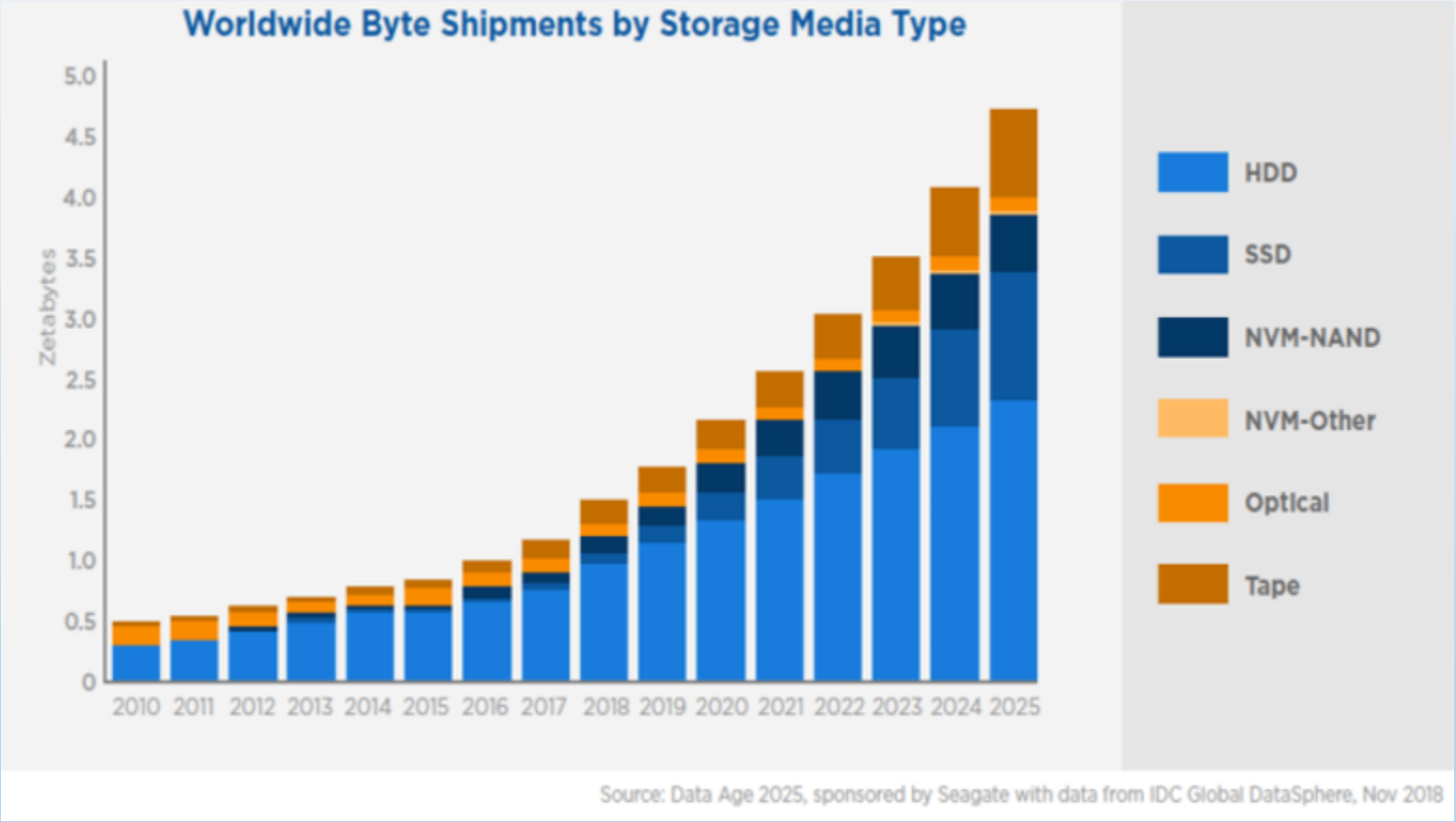
Attaques sur données d'entrée

Respect de la vie privée
Deep fakes

Panneau stop non reconnu et panda confondu avec un gibbon, extrait de publications usuelles sur ces sujets.

Mur de l'énergie

Cabinet IDC : « Le mur de la consommation énergétique liée aux besoins de calcul intensif des applications de l'IA basées sur l'apprentissage profond et consommatrices de très grandes quantités de données, en arrêtera inévitablement la croissance exponentielle, à terme relativement rapproché, si l'on ne fait rien pour y remédier. »



Mur de l'interaction avec l'humain

On peut classer ces applications en grandes catégories:

– dialogue (chatbots);

– résolution partagée de problèmes et de prise de décision;

– partage d'un espace et de ressources (cohabitation avec des robots qu'on ignore ou à qui on donne des ordres);

– partage de tâches (robot coéquipier).

Or nous ne comprenons pas suffisamment la cognition, la motivation et le comportement social de haut niveau de l'être humain social

- *La nature de l'intelligence humaine reste difficile à cerner.*
- *L'IA ne sait pas expliciter les intentions de l'humain*
- *…*

Mur de l'inhumanité

Tout ce que nous, humains, possédons naturellement et que les systèmes d'intelligence artificielle n'ont pas – et n'auront pas à court ou moyen terme :

- *Connaissance du monde*
- *Sens commun*
- *Raisonnement causal*
- *….*

Voie de recherche d'hybridation de l'apprentissage machine avec le raisonnement symbolique utilisant des règles, des faits, des connaissances.