

L'IA parle-t-elle vraiment et son intelligence est-elle générale?

16 septembre 2024

Jean-Gabriel Ganascia

Professeur Sorbonne Université – LIP6

ex-président du COMETS (comité d'éthique du CNRS)

Président du comité d'éthique de France Travail

Président du comité d'orientation du CHEC (Cycle des Hautes Études de la Culture)

Synoptique

1. Intelligence artificielle — Le chat GPT ne doit pas cacher la forêt

2. IA générative

a. GAN

b. Modèles de langage

c. Limites des modèles

3. Les chatbots sont-ils doués de parole ?

4. Jusqu'où?

a. L'IA est-elle générale? ou forte?

b. Modèles de langage

5. L'IA expliquée en dialogue





1

INTELLIGENCE ARTIFICIELLE
LE CHAT GPT NE DOIT PAS CACHER LA FORÊT

Intelligence – latin *inte(r)* – *legere**

**Legere*: cueillir, lier, rassembler

Quatre significations au sens classique:

1. Esprit, être spirituel

Philosophie de l'esprit

2. Facultés de compréhension, d'imagination (*ingenium*)

créativité, inventivité, ingéniosité, ...

3. Entente, connivence

« être en bonne intelligence »,

« Intelligence avec l'ennemi »

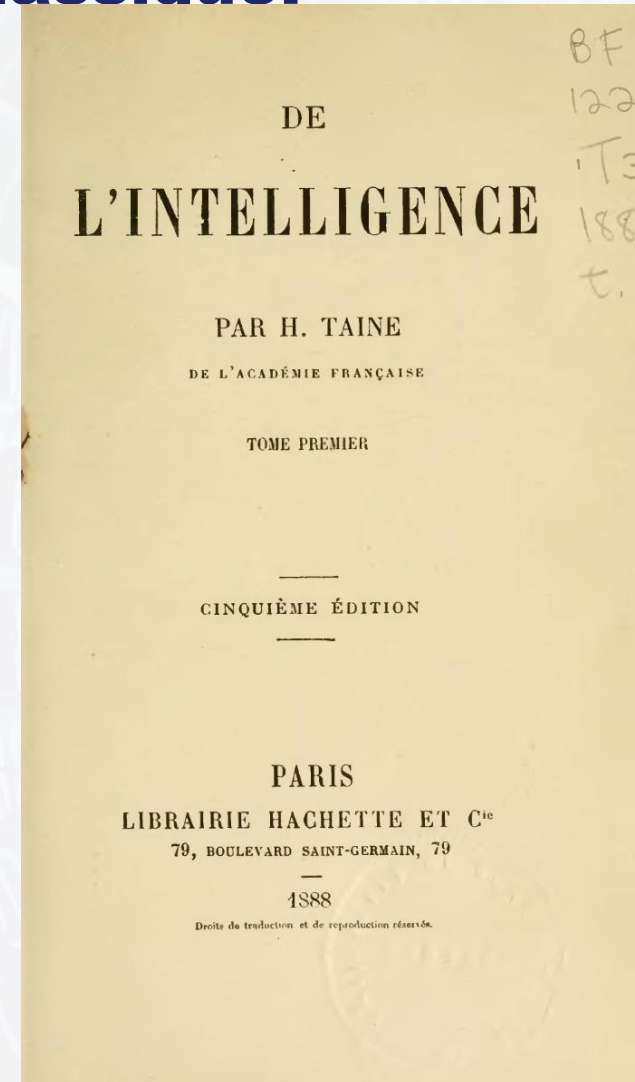
Accord secret — *Intelligence Service*

4. Capacité à connaître

Fonctions cognitives

Psychologie, sciences cognitives

Intelligence artificielle



Ancienneté du projet

Blaise Pascal et Leibniz (*XVII^e siècle*)

Charles Babbage et Ada Lovelace (*début XIX^e siècle*)

Stanley Jevons (*XIX^e siècle*)

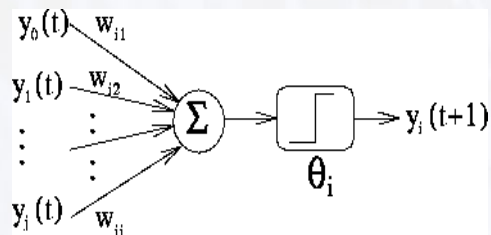
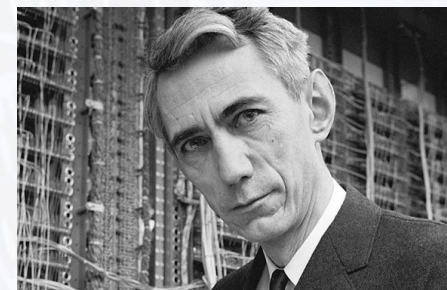
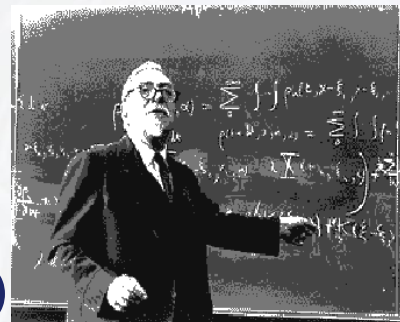
Warren McCulloch and Walter Pitts (1943)

• **Norbert Wiener (1943 et 1946 —
cybernétique)**

Claude Shannon (1948)

Alan Turing (1950)

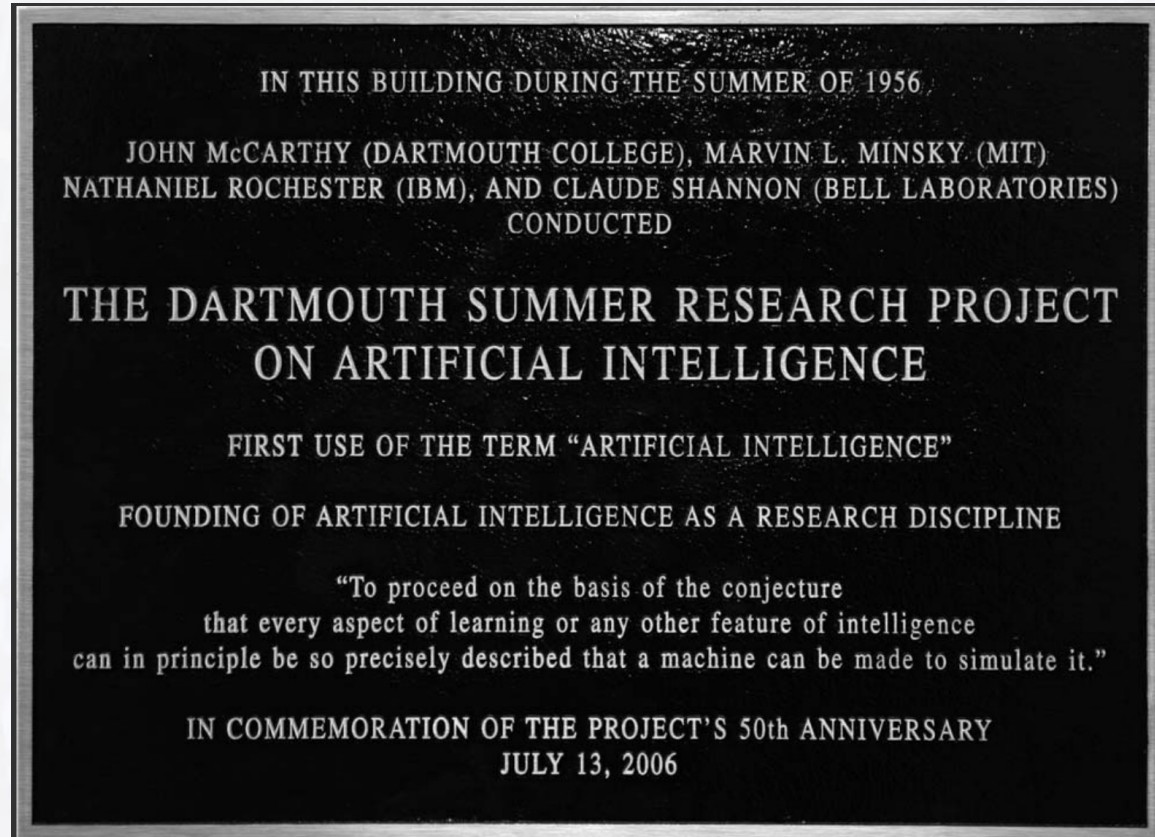
• ...



Une dame âgée de 68 ans...



août 1956: organisation
d'une école d'été par John
McCarthy pour étudier l'IA



Approches de l'IA

Fonctions mentales

Perception

Vision, reconnaissance faciale, ...

Reconnaissance de la parole, des accents, des émotions

Mémoire et apprentissage

Représentation des connaissances et ontologies

SVM, forêts aléatoires, réseaux de neurones, etc.

Raisonnement

Systèmes à base de connaissances

Démonstration automatique

Fusion de données

Fonction expressives

Traitement et Compréhension du langage naturel

Agents conversationnels

Fonctions exécutives

Robots

Sources...

- Psychisme
- Cerveau
- Évolution
- Société — culture
- Physique
- ...

Outils:

Mathématiques

Logique mathématique

Statistiques

Optimisation

...

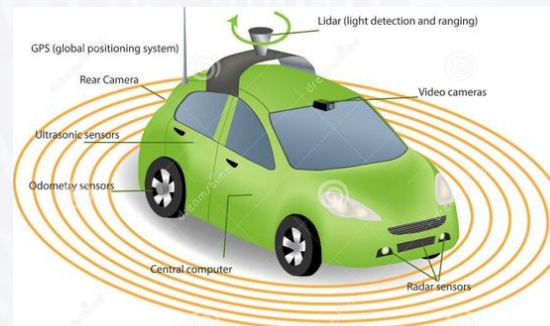
Psychologie, linguistiques, ...



Empan de l'IA

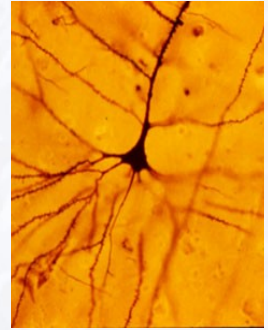
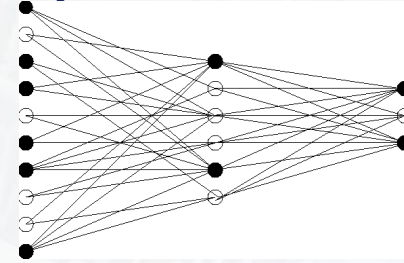
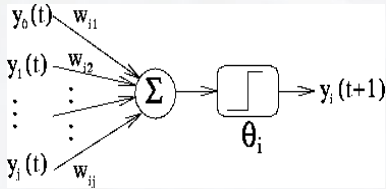
Nous sommes comme Monsieur Jourdain qui faisons de l'IA tous les jours sans le savoir

- **Web** — couplage internet et hypertexte
- **Économie du web** (publicité ciblées)
- **Biométrie** (reco. faciale)
- **Vision artificielle**
- **Reconnaissance de la parole**
- Traitement et Compréhension du **langage naturel** (traduction, moteur recherche, résumé, génération de textes, ...)
- **Robots** (et bots)
- **Véhicules « autonomes »**
- **Sciences *in silico***
- ...

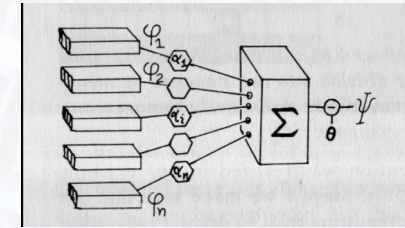


Une approche ancienne, toujours en vogue: les réseaux de « neurones formels »

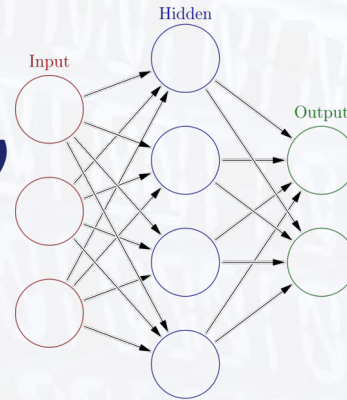
Neural Networks – 1943 (McCulloch & Pitts)



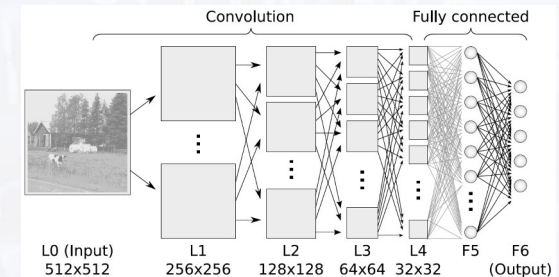
Perceptron – 1958 (Rosenblatt)



Perceptrons multi-couches – 1986 (Rumelhart, Hinton and McClelland)



Apprentissage profond et réseaux de neurones convolutifs – 2010 (Yann Lecun, ...)



2

IA GÉNÉRATIVE



Composition musicale avec ordinateurs

Pierre Barbaud (1911-1990)
1^{ère} tentatives: 1957



Lejaren Hiller (1924-1994)



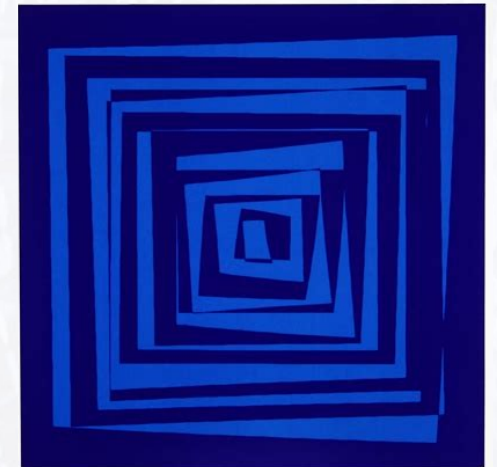
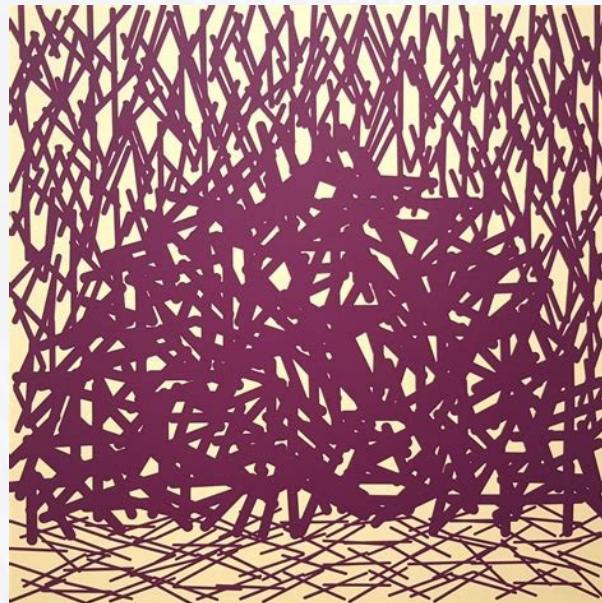
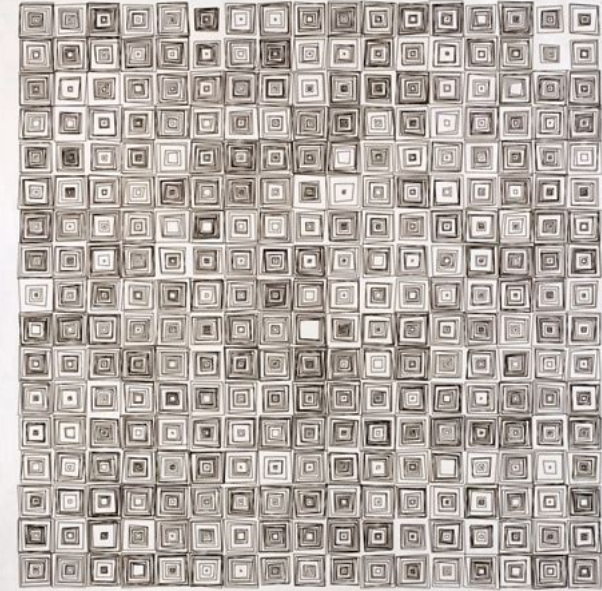
Improvisation en Jazz:

Trio rythmique → Génération de la ligne de basse (1992)

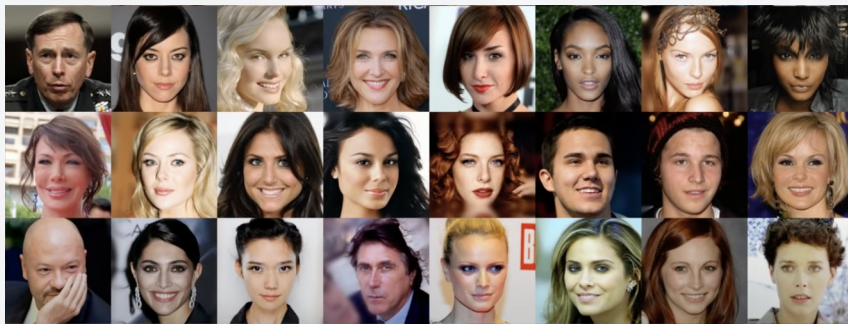


Dessin et peinture: "MOLNART"

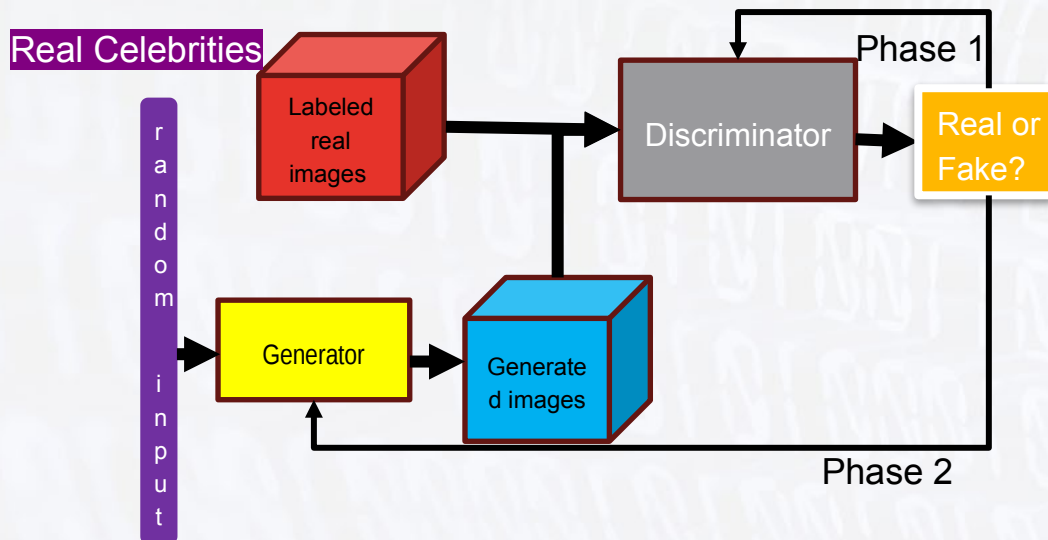
Vera Molnàr (1924 - 2023)



Generative Adversarial Networks - 2014

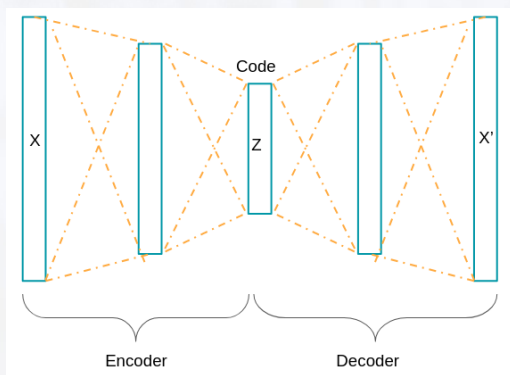
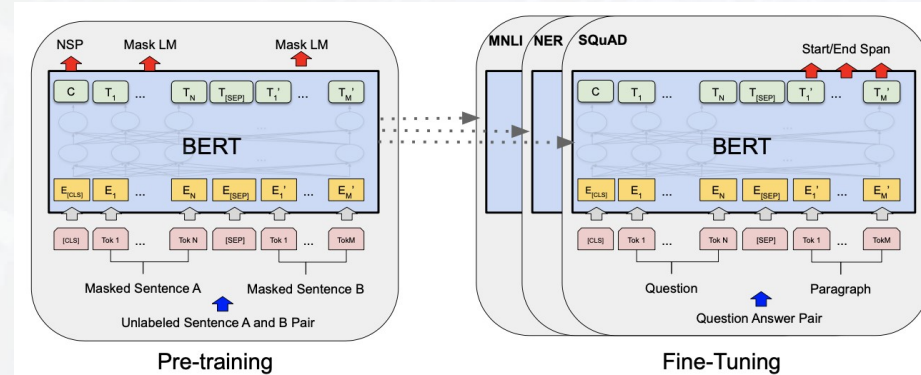


Fake Celebrities



Grands modèles de langage (*Large Language Models*) – *transformer 2017*

Year	Model	# of Parameters	Dataset Size
2019	BERT [39]	3.4E+08	16GB
2019	DistilBERT [113]	6.60E+07	16GB
2019	ALBERT [70]	2.23E+08	16GB
2019	XLNet (Large) [150]	3.40E+08	126GB
2020	ERNIE-GEN (Large) [145]	3.40E+08	16GB
2019	RoBERTa (Large) [74]	3.55E+08	161GB
2019	MegatronLM [122]	8.30E+09	174GB
2020	T5-11B [107]	1.10E+10	745GB
2020	T-NLG [112]	1.70E+10	174GB
2020	GPT-3 [25]	1.75E+11	570GB
2020	GShard [73]	6.00E+11	–
2021	Switch-C [43]	1.57E+12	745GB



- **Modèles de langage** extraction de l'esprit des langues — notion de sémantique distributionnelle: le sens d'un mot est lié aux mots qui l'entourent — plongements lexicaux **ou** enchâssements (*embeddings*)
- Entraînés par apprentissage machine sur d'immenses quantités de textes (encyclopédies en lignes: taille 10² Go → équivalent de **centaines de milliers de livres**)
- Réseaux de neurones avec cent mille **milliards de connexions!**

Utilisations pratiques des LLM

Traitement Automatique des Langues (TAL) améliorées par les LLM

Résumé de texte

Détection d'entité nommées → extraction de connaissances à partir de textes (BD)

Modules question réponse: données scientifiques, juridiques ou administratives

IA générative: aider dans toutes les tâches administratives de rédaction (utilisation de texte préformé ou de LLM?)

Comptes-rendus

Correspondance clients

Rédaction de projets

Logiciels: génération code

Fabrication de sites web

Images (remplacement graphistes)

...



B.



3

LES CHATBOTS SONT-ILS DOUÉS DE PAROLE?

Que sont les agents conversationnels

Chatbots – Étymologie & origine

« Chatter » (bavard) + « bot » (robot)

⇒ chatterbot puis chatbot

Systemes de dialogue

Génériques

Spécialisés

Réaction à mots clefs, sans compréhension, ni but

Premier « chatterbot »:

ELIZA – J. Weizenbaum 1965

But: réaliser le test de Turing dans une situation particulière

Applications:

Jeux sur Internet

« Messageries roses »

Assistants « FAQ »



Bavarder est-ce parler?

ChatGPT: 30 novembre 2022

- **GPT (LLM)** — apprentissage profond pour la génération de texte
- **Apprentissage par renforcement** (*entraîné en Afrique*): aseptisation... censure



Chatbot — Modèles de langage: LaMDA

L'affaire Lemoine!

16 juin 2022

Opinions - 16.06.2022 - 00:00

LaMDA: Can an artificial intelligence develop consciousness?

Researchers at Google have developed an AI-based dialog system, which as claimed by a Google software engineer has become sentient. After publishing the transcript of a dialogue with the AI-system without permission, the software engineer got suspended. This has opened up the question on how open AI research can or should be? By Damian Borth.



16 June 2022. A few days ago, Blake Lemoine, a software engineer at Google published an article on [Medium](#) about whether the AI-based dialogue system LaMDA, which is being developed at Google, has developed consciousness. On the same day, this was addressed by the [Washington Post](#), whereupon Blake Lemoine was suspended by Google.

Qu'est-ce que parler?

La parole soulève plus de terre que le fossoyeur ne le peut.

TROIS RESPIRATIONS, René Char,
in *Recherche de la base et du sommet*,
Œuvres complètes, la Pléiade, p.652

Le Golem ne parle pas!

Golem: créature informe

Légende du Golem: Rabbi Löwe, le « Maharal » de Prague, est réputé avoir construit une statue d'argile qui s'anime lorsque l'on place un message derrière ses dents et se charge des tâches ménagères...

Mais il ne parle pas! Il n'est pas doué de parole

Faisons nous mieux que Rabbi Löwe?



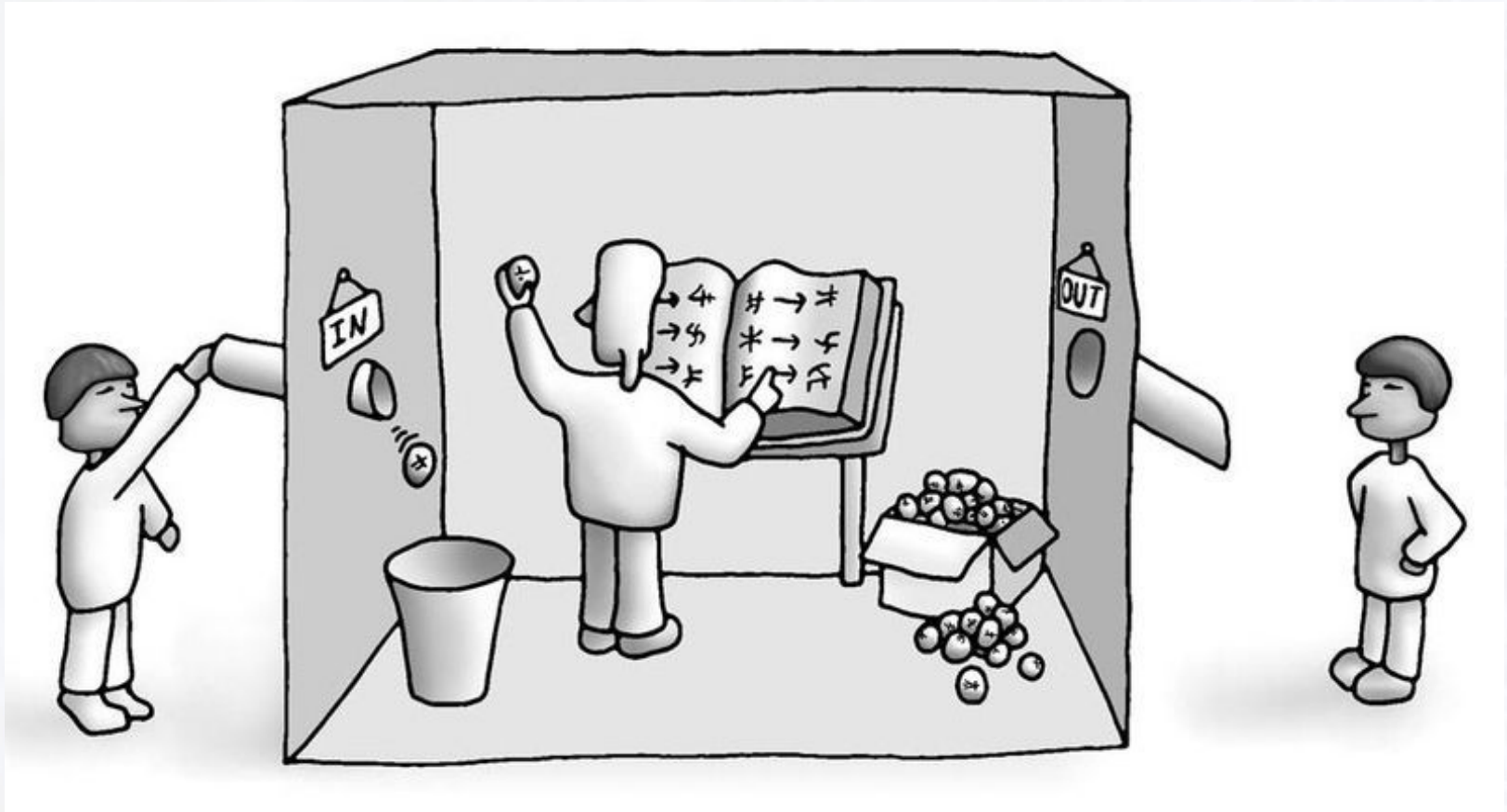
Nos chatbots soulèvent-ils plus de terre que les fossoyeurs?

4

JUSQU'OU?

Intelligence artificielle forte — faible

La chambre chinoise – John Searle



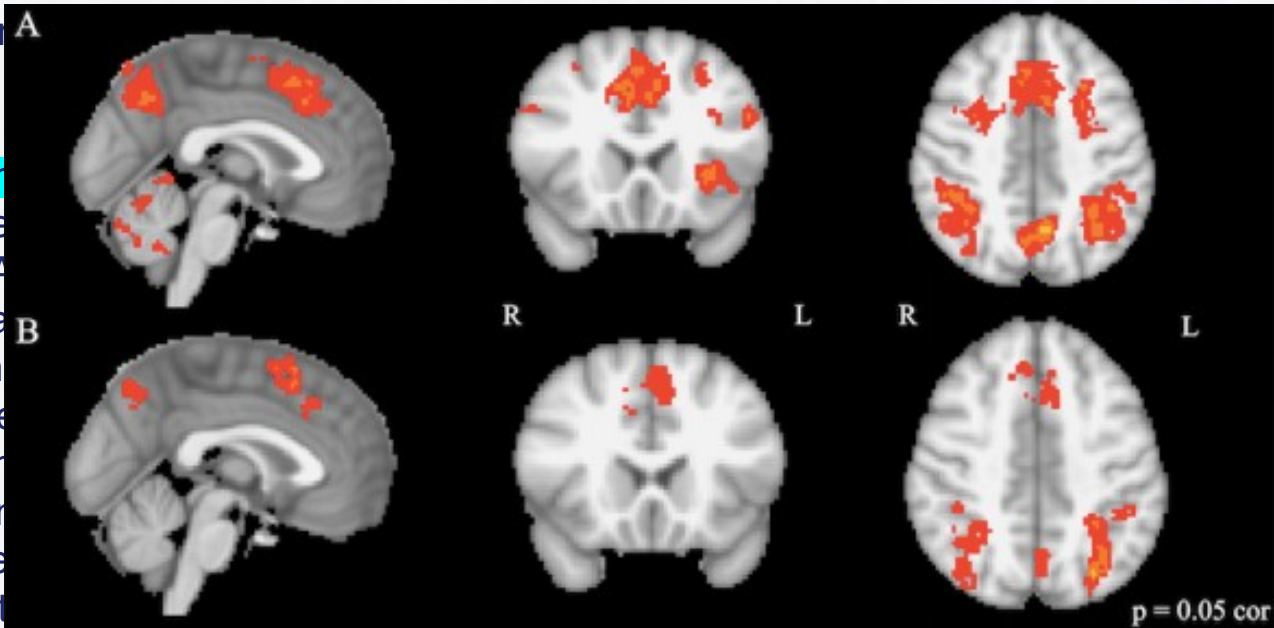
Intelligence artificielle Générale

Il n'y a pas d'esprit dans les machines!

Pause Giant AI Experiment

We call on all AI labs to pause their efforts to train models more powerful than GPT-4.

AI systems with **human-level** capabilities, as shown by extensive reasoning, planning and management, should be planned for a controlled, out-of-control race to develop. AI creators - can understand the risks. Contemporary AI systems are not. ourselves: Should we let AI systems we automate away all the jobs that might eventually outnumber, outsmart, obsolete and replace us? Should we risk loss of control of our civilization?



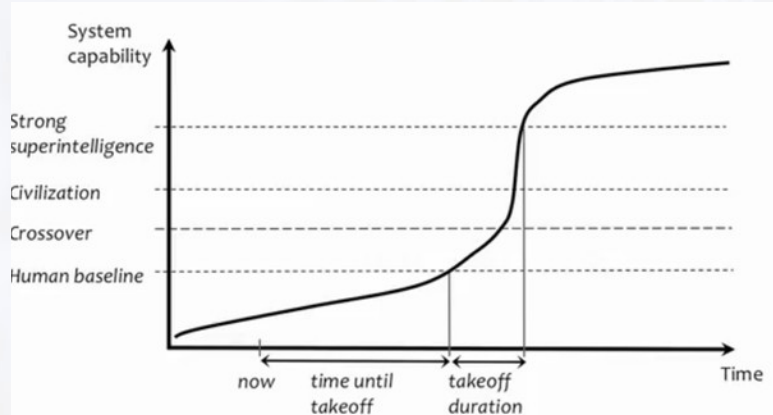
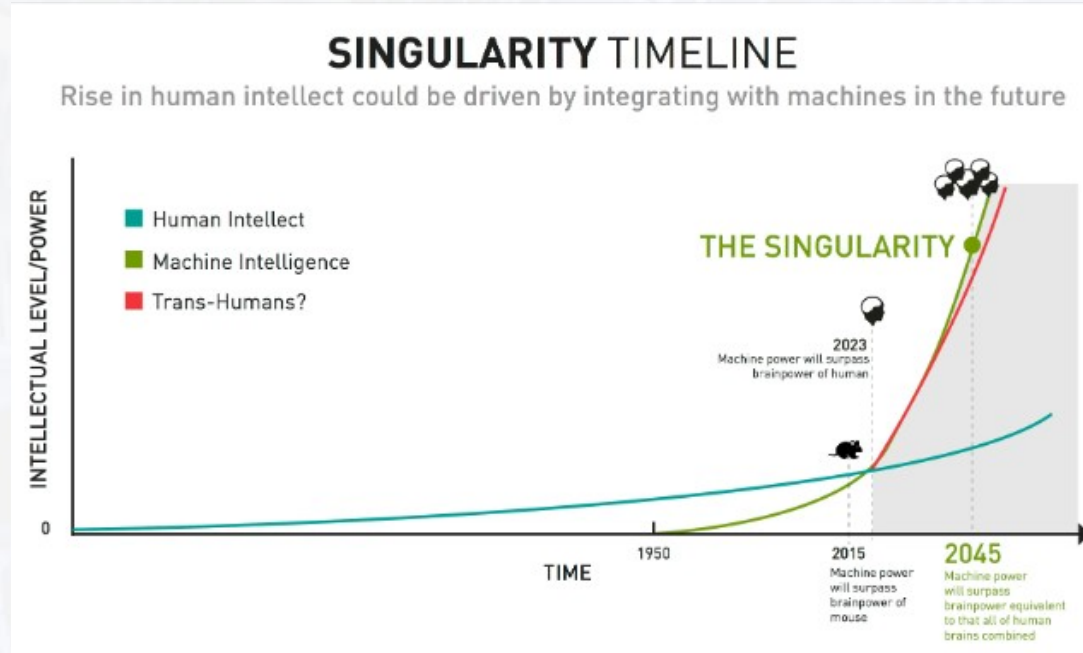
Such decisions must not be delegated to unelected tech leaders. Powerful AI systems should be developed only once we are confident that their effects will be positive and their risks will be manageable. This confidence must be well justified and increase with the magnitude of a system's potential effects. OpenAI's recent statement regarding **artificial general intelligence**, states that "At some point, it may be important to get independent review before starting to train future models. We have advanced efforts to agree to limit the rate of growth of compute used for creating AI models. This point is now."

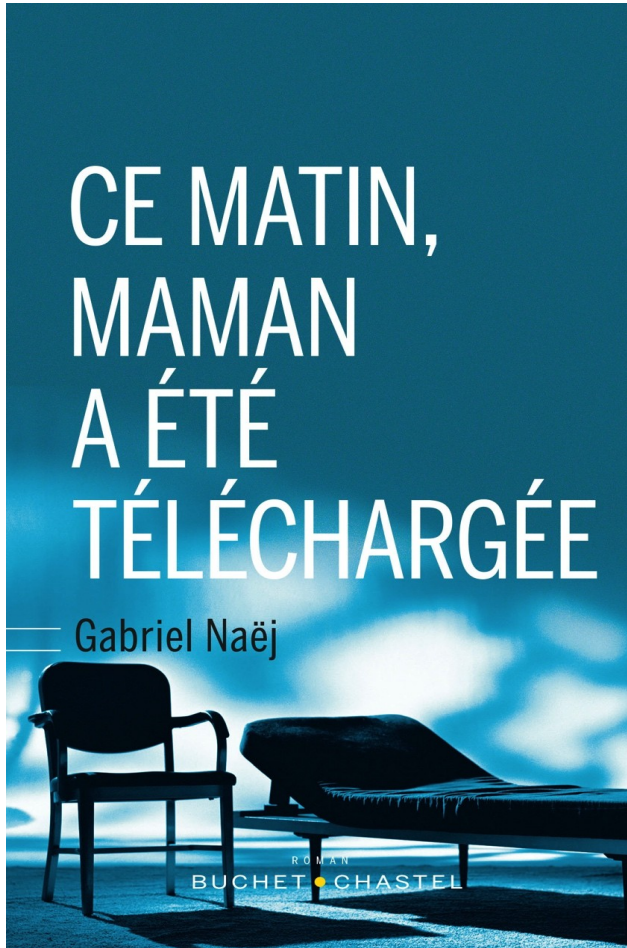
L'intelligence n'est pas générale!

Super-intelligence et Singularité technologique

Advenue d'une intelligence supérieure


Future
of Humanity
Institute
UNIVERSITY OF OXFORD





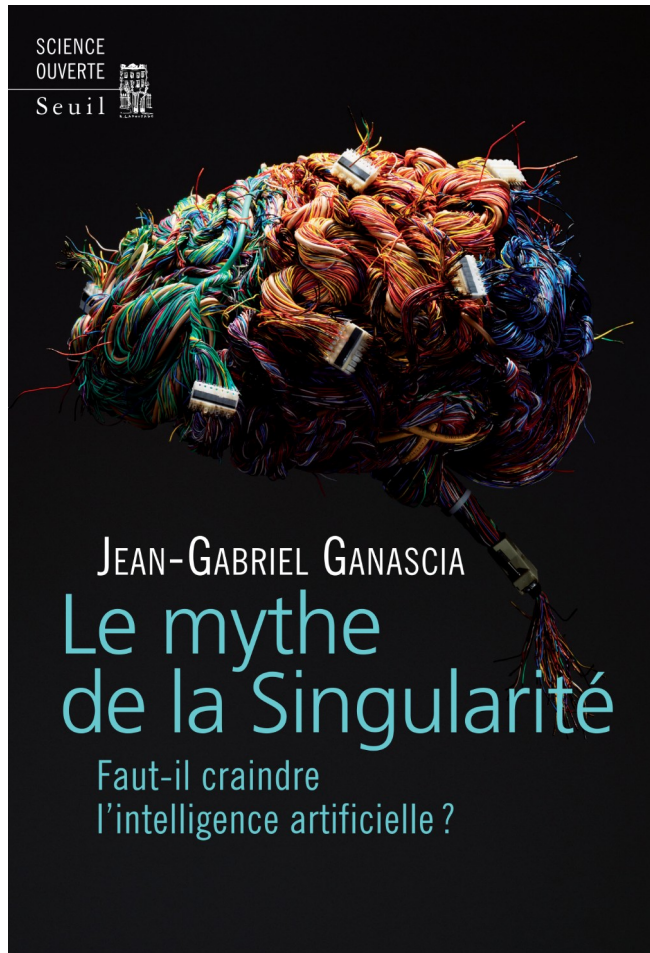
La Singularité est advenue

**Et, comble de malheur pour lui,
sa maman, sa vraie mère décide
de procéder à une
métempsychose* numérique.**

**Et en plus, elle s'est réincarnée
dans un robot sexuel ce qui le
trouble beaucoup**

**Et, elle fait irruption à tout
moment dans sa vie...**

* passage de l'âme d'un corps dans un autre



Réponses: La loi de Moore

- Limitations de la loi de Moore
- Puissance de calcul \neq intelligence (au sens d'esprit...)

L'apprentissage

- Limites de l'induction

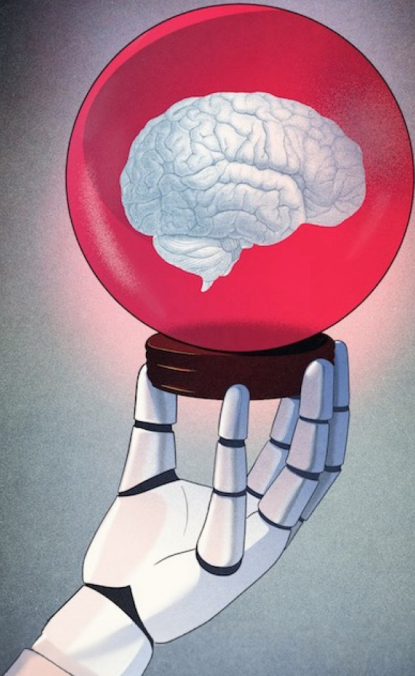
Gare aux pompiers pyromanes!

Google, Nokia, Cisco, ...

Elon Musk, Bill Gates, ...

JEAN-GABRIEL GANASCIA

L'IA. expliquée aux humains



Seuil

5

L'IA EXPLIQUÉE EN DIALOGUE

Dialogue avec 4 élèves du collège Raymond Queneau:
Soukeyna,
Jade,
Naël et
Louis
chargés de réaliser un numéro spécial des
Querelles de Queneau sur l'intelligence artificielle

Réponses aux questions des collégiens

JEAN-GABRIEL GANASCIA

L'IA expliquée aux humains



Seuil

Définitions des principaux termes

Intelligence, algorithme, calcul, ...

Succès et applications

Internet, santé, reconnaissance de la parole, ...*

Comment ça marche ?

Sources d'inspiration, différentes techniques, IA gén, ...

Les machines excèdent-elle l'Homme ?

Qu'est-ce que nous faisons mieux qu'elles ?

Sciences in silico, « Big Data », Création artistique...

Avoir confiance dans l'IA !

L'intelligence artificielle digne de confiance, robustesse, ...

Risques

Craintes, IA générale, faible vs. forte, conscience, ...

Dangers

Remplacement, responsabilité, biais, discrimination, ...

Comment se prémunir ?

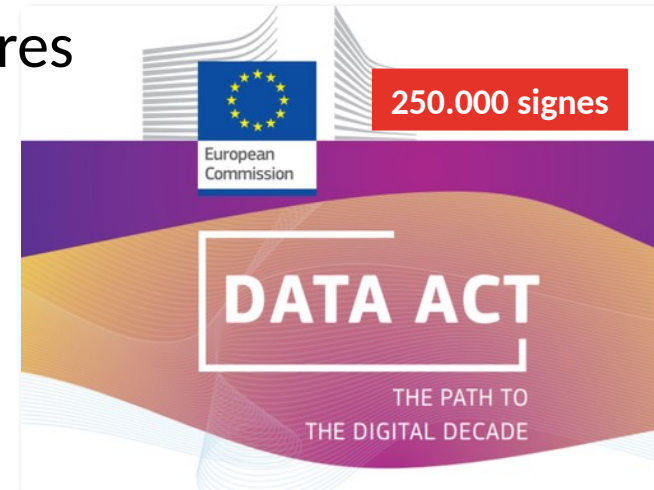
Moratoire? « bouton rouge », réglementation...

La leçon de Laverdure

Régulation frénétique du numérique!

Accélération depuis décembre 2019

Opposition de la commission aux États membres



Fondé sur le risque:

1. **Risque inadmissible:**
interdiction
2. **Risque élevé:**
régulation
3. **Risque modéré:**
transparence
4. **Risque faible**



Risque inadmissible:

- i. images subliminales
- ii. systèmes notation
 - Exception pour les banques et les sociétés privées
 - Ex. permis à point!
- iii. biométrie en temps réel



Risque faible

- iv. filtre à spam
- v. jeux vidéo



Chatbot:

- vi. risque modéré à faible (en 2022)
- vii. introduction réglementation spécifique pour les LLMs en 2023

La « charité ensorcelée » des géants du web



Diversity and Inclusion

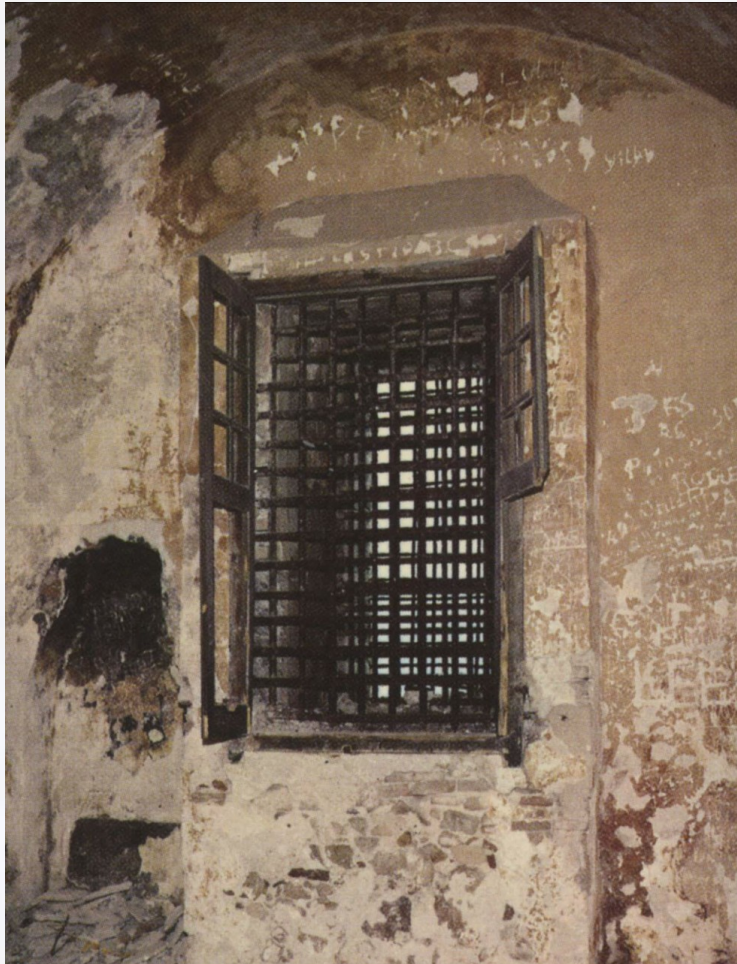
IBM CEO's Letter to Congress on
Racial Justice Reform



IBM annonce qu'il se retire du marché de la reconnaissance faciale étant donné qu'elle sert au profilage racial et appelle à une réforme de la police *Le 9 juin 2020*

Selon des documents examinés par Liberty Investigations et The Verge, IBM a signé un contrat de 69,8 millions de dollars avec le gouvernement britannique pour créer une plateforme biométrique dotée d'une fonctionnalité de reconnaissance faciale le mois dernier. *Le 1^{er} septembre 2023*

« Servitude volontaire » E. de La Boétie



« Le théâtre, les jeux, les farces, les spectacles, les gladiateurs, les bêtes curieuses, les médailles, les tableaux et autres drogues de cette espèce étaient pour les peuples anciens les appâts de la servitude, le prix de leur liberté ravie, les outils de la tyrannie. Ce moyen, cette pratique, ces allèchements étaient ceux qu'employaient les anciens tyrans pour endormir leurs sujets sous le joug. [...] **Mais ils ne font guère mieux ceux d'aujourd'hui qui, avant de commettre leurs crimes les plus graves, les font toujours précéder de quelques jolis discours sur le bien public et le soulagement des malheureux** »

« Don't be Evil! » Google
Rachat de Twitter par Elon Musk pour garantir la liberté d'expression!
Appel au moratoire sur l'intelligence artificielle par ceux qui la produisent!...



The background features a series of wavy, horizontal lines in shades of purple and blue, creating a sense of motion and depth. Below these lines, a grid of small, glowing dots is visible, adding a digital or data-like texture to the overall composition.

**Merci de votre
attention**