

Lundi de l'IA
5 juin 2023

Jules Pondard



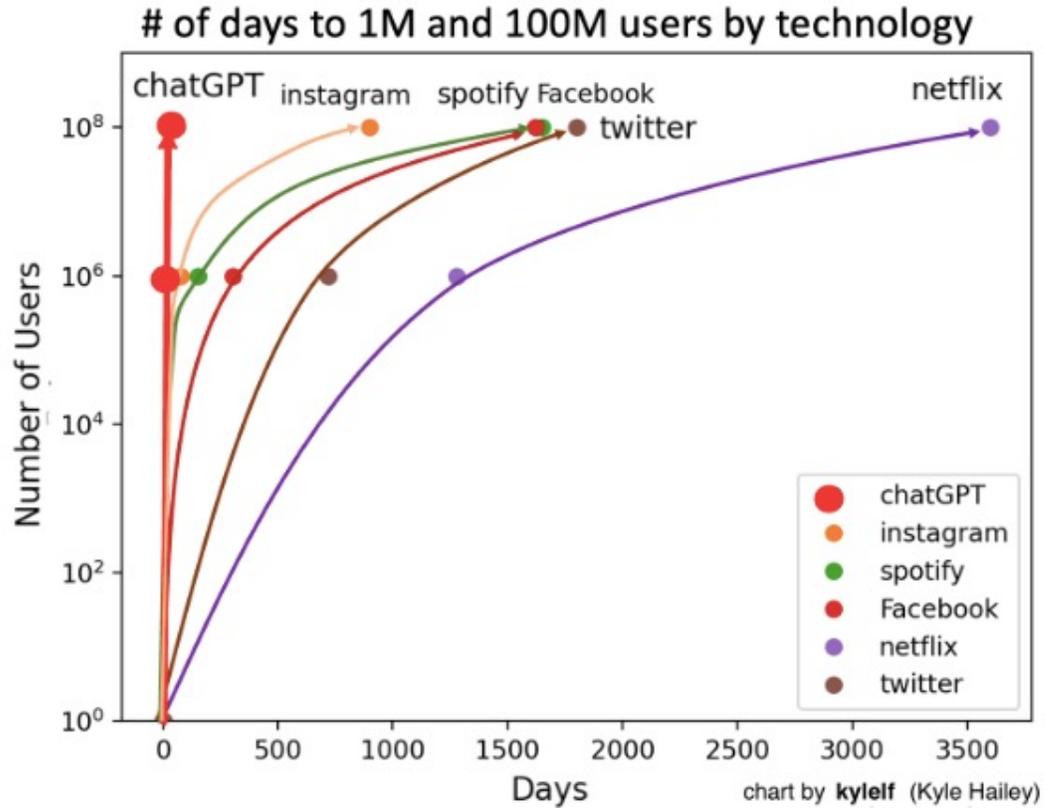
Pôle FINTECH-INNOVATION (ACPR)

Introduction à ChatGPT et ses enjeux

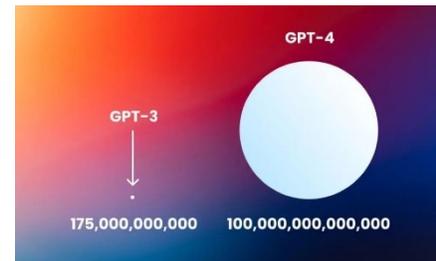
Plan

- ChatGPT : contexte & actualités
- Comment marchent les « LLM » / ChatGPT
- Questions ouvertes & controverses

CHATGPT, CONTEXTE & ACTUALITÉS



Nombre d'utilisateurs (de 1 million à 100 millions) en fonction du temps pour différents produits





UNE CONTROVERSE PUBLIQUE

ChatGPT « n'a rien de révolutionnaire » selon Yann LeCun

Technologie : Si le public perçoit ChatGPT comme révolutionnaire, Yann LeCun n'a pas la même vision. Selon le responsable de l'IA chez Meta, les techniques utilisées ne sont pas différentes de ce qui existe déjà dans de nombreux laboratoires de recherche.



Ces derniers temps, on a beaucoup parlé de l'énorme promesse du programme ChatGPT, d'OpenAI, pour générer des énoncés en langage naturel dans des styles et sur des thématiques variés.

ChatGPT fascine à tel point les foules que beaucoup pensent que ce programme est unique en son genre. Pourtant, les spécialistes de l'IA ne sont pas tout à fait d'accord.

« En termes de techniques sous-jacentes, ChatGPT n'est pas particulièrement innovant », a déclaré Yann LeCun, responsable scientifique de l'IA chez Meta, lors d'une petite réunion de la presse et des dirigeants sur Zoom la semaine dernière. « Il n'y a rien de révolutionnaire, même si c'est la façon dont il est perçu par le public », a déclaré le pionnier du deep learning.

TECH / GOOGLE / ARTIFICIAL INTELLIGENCE

'Godfather of AI' quits Google with regrets and fears about his life's work



Geoffrey Hinton (foreground) has left Google to speak out on the dangers of AI. Image: Getty

/ Geoffrey Hinton who won the 'Nobel Prize of computing' for his trailblazing work on neural networks is now free to speak about the risks of AI.

By Thomas Rickar, a deputy editor and Verge co-founder with a passion for human-centric cities, e-bikes, and life as a digital nomad. He's been a tech journalist for almost 20 years.

May 1, 2023, 12:58 PM GMT-2 | 53 Comments (53 New)



Pause Giant AI Experiments: An Open Letter

We call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4.

[View this open letter online.](#)

Published	PDF created	Signatures
March 22, 2023	May 5, 2023	27565

AI systems with human-competitive intelligence can pose profound risks to society and humanity, as shown by extensive research¹ and acknowledged by top AI labs.² As stated in the widely-endorsed [Asilomar AI Principles](#), Advanced AI could represent a profound change in the history of



Pour Bill Gates, ChatGPT est la plus grande révolution depuis les années 1980

Dans un post de blog, le cofondateur de Microsoft explique que l'intelligence artificielle est aussi révolutionnaire que l'ordinateur, Internet ou le téléphone portable. Le milliardaire détaille des exemples d'applications de ChatGPT, notamment pour réduire les inégalités, mais aussi les risques.



Yuval Noah Harari: 'I don't know if humans can survive AI'

For the first time, we've invented something that takes power away from us, says the Israeli historian and author of Sapiens

By Harry de Quetteville
23 April 2023 - 9:00am



Harari says he keeps his phone switched off, in a drawer, and is on an "information diet" | credit: david bailey

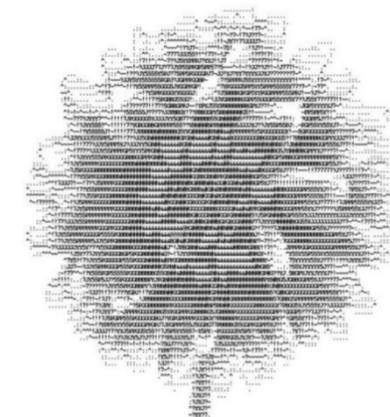
Stories have always been vital to Yuval Harari, the Israeli historian-cum-philosopher. The unique capacity of our species to be bound and united by intangible narratives, even across oceans, was central to Sapiens, his mind-bogglingly popular chronicle of our species' rise, which catapulted him to star-like status after its publication in English nearly a decade ago. Perhaps that is why he is so profoundly concerned today about the rise of a challenger to our tale-telling mastery - artificial intelligence (AI)...



OPINION
GUEST ESSAY

Noam Chomsky: The False Promise of ChatGPT

March 8, 2023



Elon Musk rêve d'un TruthGPT, un ChatGPT qui défendrait la « vérité maximale »

[On a hâte de discuter avec lui](#)

⌚ Temps de lecture : 2 min

Nicolas Lellouche





CHATGPT SUSCITE DES RÉACTIONS POLITIQUES...

Danemark : la première ministre prononce un discours en partie écrit par ChatGPT

Par Le Figaro avec AFP
Publié le 31/05/2023 à 11:00, mis à jour le 31/05/2023 à 11:34

Copier le lien

Écouter cet article 00:00/02:32



Mette Frederiksen, à Helsinki, en Finlande, le 26 mai 2023. RITZAU SCANPIX / REUTERS

Pour souligner l'aspect révolutionnaire mais aussi les risques engagés par cette technologie, Mette Frederiksen a fait rédiger la première partie de son discours par l'intelligence artificielle.

ChatGPT : Bruno Le Maire a testé l'IA pour écrire un discours sur la Chine

PROUESSE TECHNOLOGIQUE Le ministre de l'Economie est fasciné par ChatGPT

L.B. | Publié le 03/04/23 à 17h23

31 COMMENTAIRES 118 PARTAGES



Le ministre de l'Economie Bruno Le Maire à l'Élysée, le 15 mars 2023. — Jacques Witt / SIPA

Écouter cet article ChatGPT : Bruno Le Maire a testé l'IA pour écrire un discours sur la Chine 00:00

ChatGPT fascine même jusqu'au ministre de l'Economie. Et si on en croit Bruno Le Maire,

À LIRE AUSSI

- 10/01/23 | COLÈRE « C'est de la merde » : Nick Cave fustige une chanson écrite par ChatGPT
- 31/03/23 | MENACE Données personnelles : L'Italie bloque le robot conversationnel ChatGPT
- 10/02/23 | IA ChatGPT a failli réussir un examen de médecine américain

D'ACTU



Publié le 28 avril 2023 à 23h53

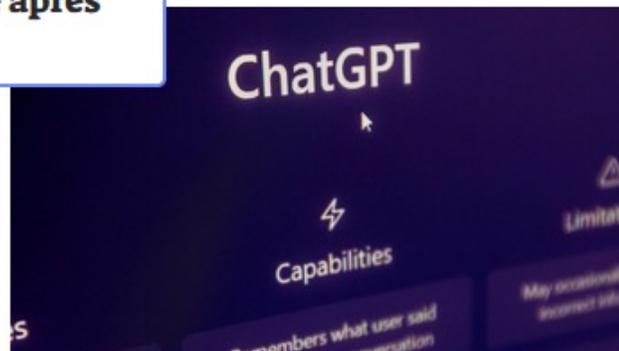
ChatGPT revient en Italie après une exclusion d'un mois

Il ritorno

Temps de lecture : 2 min

Julien Lausson

Tech Web Vie privée et données personnelles



TECH > INTELLIGENCE ARTIFICIELLE

CHATGPT: LE GOUVERNEMENT NE VEUT PAS DE RÉGULATION "EXCESSIVE"

Raphaël Grably Le 29/05/2023 à 11:08



Le ministre Jean-Noël Barrot le 31 août 2022 à l'Élysée - LUDOVIC MARIN / AFP

Le ministre délégué au numérique estime que la réglementation européenne telle qu'elle se dessine risque d'être trop contraignante. Il vise notamment de potentielles obligations de transparence.

Ne pas se laisser distancer par les Américains et les Chinois en optant pour une réglementation trop radicale. C'est en substance le discours du ministre délégué au Numérique, Jean-Noël Barrot, ce 29 mai sur Franceinfo. Il est notamment revenu sur le cas des intelligences artificielles génératives, qui permettent de créer du texte, comme ChatGPT, ou des images, comme Midjourney.

Pour Jean-Noël Barrot, "à ce stade, la position que le Parlement européen a prise sur l'IA est excessive et risque de faire sortir l'Union européenne de l'histoire technologique".

Ne pas se faire distancer

Le ministre fait référence au projet de réglementation européenne de l'intelligence artificielle, en cours de discussion au Parlement européen, et qui vise à imposer de nouvelles contraintes aux principaux acteurs du secteur à partir de 2025.



...ET ALERTE LA « BIG TECH »

Grâce à l'IA, le géant des puces Nvidia rejoint les GAFAM dans le club des entreprises à 1.000 milliards de dollars

EXPLOSION L'action du leader des puces graphiques a presque triplé en un an, face à la puissance de calcul galopante nécessaire pour faire fonctionner les nouveaux systèmes d'intelligence artificielle générative

20 Minutes avec AFP | Publié le 31/05/23 à 05h04

Microsoft investit 10 milliards de dollars dans OpenAI

Dominique Filippone, publié le 23 Janvier 2023

ChatGPT : Microsoft réclame la création d'une autorité de régulation de l'intelligence artificielle



[ANTOINE BARSACO](#)

, le 26 mai 2023 17:36

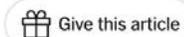
[Accéder aux articles rédigés par cet auteur](#)

Ce sont maintenant les premiers concernés par une régulation de l'IA qui réclame que le gouvernement prenne ses responsabilités. Le vice-président de Microsoft a appelé lors d'un discours à la création d'une autorité de régulation de l'intelligence artificielle aux États-Unis.



A New Chat Bot Is a 'Code Red' for Google's Search Business

A new wave of chat bots like ChatGPT use artificial intelligence that could reinvent or even replace the traditional internet search engine.



Google Calls In Help From Larry Page and Sergey Brin for A.I. Fight

A rival chatbot has shaken Google out of its routine, with the founders who left three years ago re-engaging and more than 20 A.I. projects in the works.

DES PREMIERS RÉSULTATS PROMETTEURS

90% in Bar Exam, 88% in LSAT: GPT-4 beats 90% of humans in world's toughest exam

1 min read • 15 Mar 2023, 06:05 PM IST

[Livemint](#)

In a demonstration, the engineers at OpenAI noted that the enhanced version of ChatGPT had outperformed 90 per cent of humans in some of the toughest exams in the US.

Bientôt les premiers cas d'usage concrets ?

- **Forces** : rédaction, synthèse, code
- **Faiblesses** : raisonnement, rapport à la vérité, capacité à planifier...

Microsoft Researchers Claim GPT-4 Is Showing "Sparks" of AGI

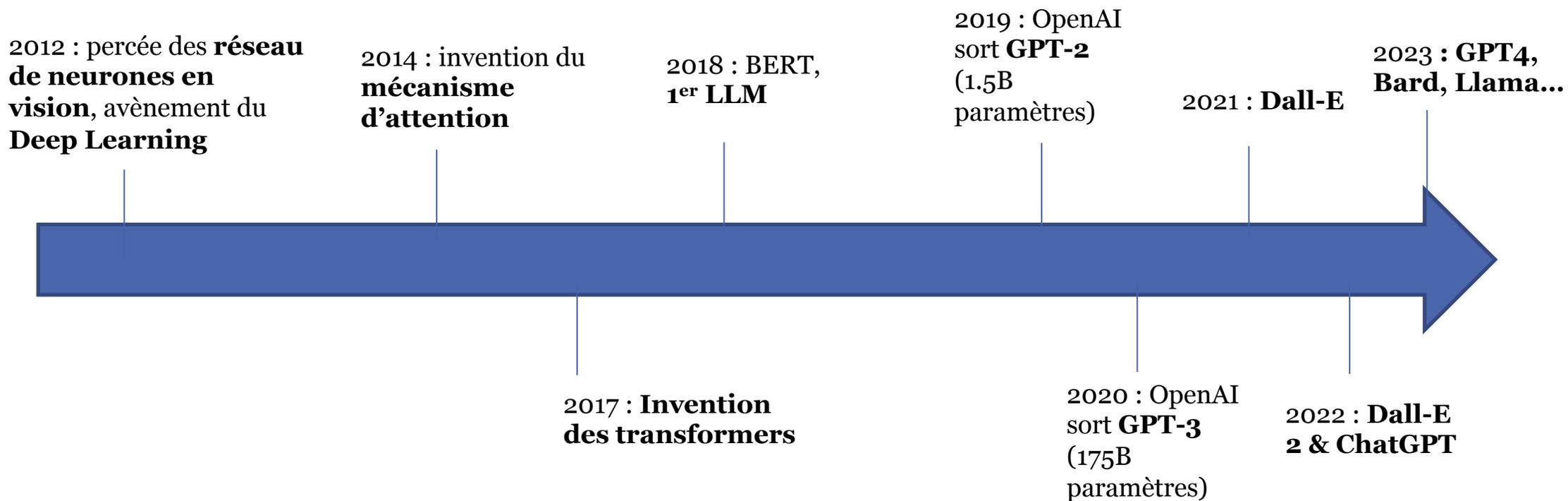
"We believe that GPT-4's intelligence signals a true paradigm shift in the field of computer science and beyond."



A nuancer : pas d'évaluation indépendante à ce stade



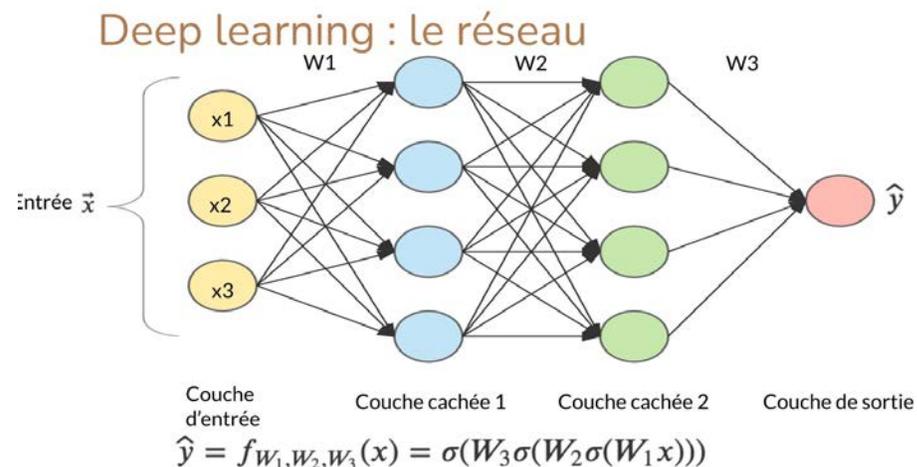
HISTORIQUE





CHATGPT

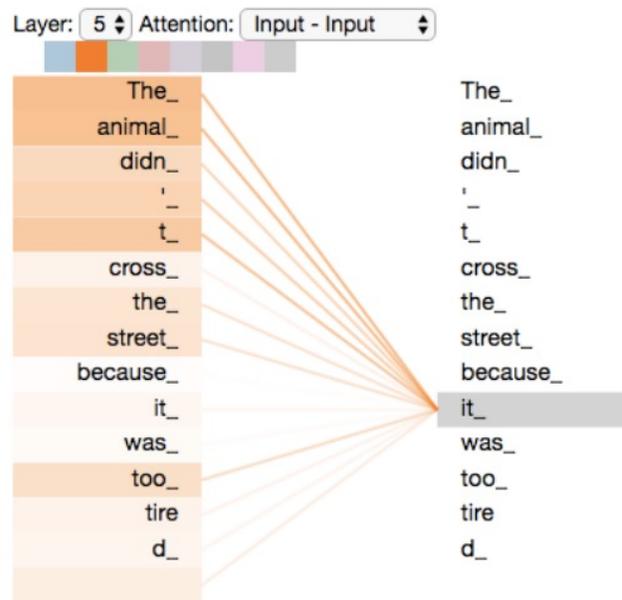
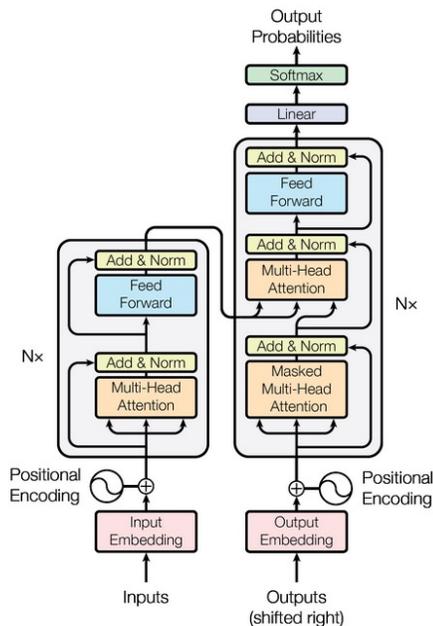
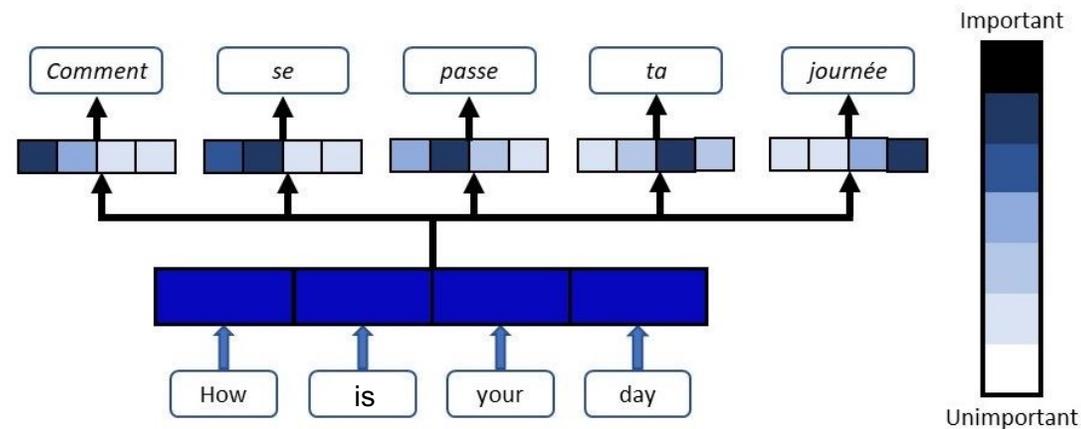
- **ChatGPT** est constitué de **trois blocs** :
 - Un « **LLM** » (« Large Language Model »), parfois aussi appelé « modèle de fondation ». Ces LLM sont des réseaux de neurones particuliers, appelés « **Transformers** ». Dans le cas d'OpenAI, ces LLM sont nommés « **GPT** » (« Generative Pre-trained Transformers »).
 - Une méthode dite « SFT » (« Supervised Fine-Tuning »)
 - Une méthode dite « **RLHF** » (« Reinforcement Learning from Human Feedback »)
- Ainsi, **les LLM sont de grands « Transformers »**
- Un « Transformer » est un type particulier de réseau de neurones qui s'appuie sur le « mécanisme d'attention »
- Rappel : Qu'est-ce qu'un réseau de neurones ?





ATTENTION & TRANSFORMER

- **Mécanisme d'attention** (2014, Bahdanau et al.)
- **Transformer** (2017, Google Brain)





COMMENT MARCHE UN LLM ? (GPT)

Transformers

- Données d'entraînement (exemple GPT-3)

Sources de données



- Common Crawl
- Books1
- Wikipedia
- WebText2
- Books2

- Entraînés à prédire le mot suivant $P(w_n | w_1, \dots, w_{n-1})$

Exemple : « La capitale de la France est [?] »

- Peut conduire le modèle à « **comprendre** », « **raisonner** » et « **synthétiser** ».

Exemple : Donner les trois quarts d'un roman policier à un LLM puis ajouter « Ainsi le tueur est (?) »



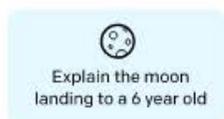
TRANSFORMATION EN CHATBOT

ChatGPT = LLM + **SFT** + **RLHF**

Step 1

Collect demonstration data, and train a supervised policy.

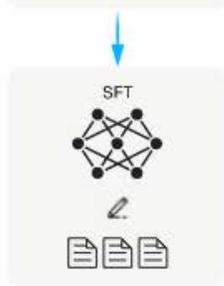
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



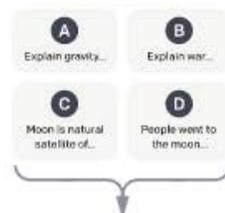
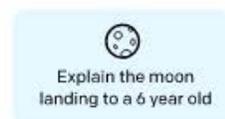
This data is used to fine-tune GPT-3 with supervised learning.



Step 2

Collect comparison data, and train a reward model.

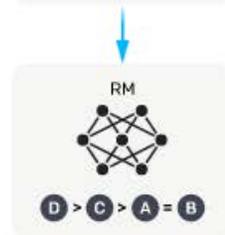
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



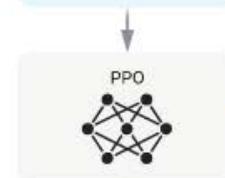
Step 3

Optimize a policy against the reward model using reinforcement learning.

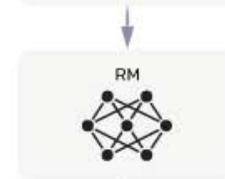
A new prompt is sampled from the dataset.



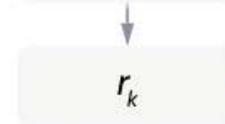
The policy generates an output.



The reward model calculates a reward for the output.



The reward is used to update the policy using PPO.





QUESTIONS OUVERTES ET DÉBATS (1/3)

- Recherche / académique, technique
 - **Qu'apporte réellement le RLHF** ? Est-il **nécessaire** ? *Pas certain : LIMA.*
 - **Comment et pourquoi** ces modèles marchent-ils aussi bien ? *Retard de la théorie sur les approches empiriques*
 - Ces modèles marchent-ils suffisamment pour des applications réelles ? **Quelles mesures (métriques) inventer pour valider leur applicabilité** ? Comment les évaluer / les auditer ?
 - Cette typologie de modèles comporte-t-elle **des limites fondamentales** ?
 - Manque de réelle compréhension (causale ?) du langage ?
 - Manque d'ancrage dans le réel / sens commun ?
 - Pas de capacité à planifier ?



QUESTIONS OUVERTES ET DÉBATS (2/3)

- **Économie, monopoles & souveraineté**
 - **Cas d'usages et produits à construire** ; méthodes d'évaluation.
 - Marché dominé par **quelques acteurs importants**, qui vont s'accaparer la valeur ?
 - Place de l'**open-source** (ex : contributions de Meta, Huggingface) ?
 - **Acteurs américains** ou **émergence d'acteurs français/européens** (ex : projet Mistral) ?
- **Enjeux de données**
 - **Données personnelles** et **données des entreprises**
 - Enjeux de **droit d'auteur**
 - Tropicisme **américain et anglophone** dans les jeux de données utilisés



QUESTIONS OUVERTES ET DÉBATS (3/3)

- **Enjeux de société**
 - **Faut-il faire une pause**, le temps de mieux comprendre et réguler ces technologies ? (**moratoire**)
 - A-t-on un **risque de perturbation de nos sociétés** ? (démocratie, informations, prolifération de « fake news », ...)
 - A-t-on un **risque de non-alignement**, par exemple entre le plan proposé par une IA pour résoudre un problème et nos valeurs éthiques ?
 - **Quelle transformation sur le marché du travail** ? A court ? Moyen ? Long terme ? Quelques premières études sur le sujet. Remplacement ? Hybridation ?
 - **Comment réguler ces algorithmes** ?
- **Autres sujets éthiques**
 - Emploi de travailleurs pauvres au Kenya pour aider ChatGPT à être moins toxique (< 2\$/h)
 - Biais de représentation dans les données, et par l'échantillon de personnes choisi pour SFT / RLHF
 - Impact écologique

**Merci pour
votre attention !**



Pôle FINTECH-INNOVATION (ACPR)



LEXIQUE

- Deep learning : ensemble de méthodes d'apprentissage à base de réseaux de neurones « profonds » ; souvent synonyme d'IA dans la communauté académique.
- NLP : « Natural Language Processing », domaine qui applique les techniques d'apprentissage au traitement du langage naturel (français, anglais, ...)
- Mécanisme d'attention : algorithme permettant à un réseau de neurones de se focaliser sur une partie de l'entrée à la fois
- **Transformers** : architecture de réseau de neurones particulière, consistant en un empilement successif de mécanismes d'attention (self-attention)
- **LLM** : « Large Language Models », grands modèles de langage de type transformers, entraînés sur de vastes quantités de données
- **GPT** : « Generative Pre-Trained Transformers », acronyme utilisé par OpenAI pour nommer ses modèles de langage.
- Modèles de fondation : analogue aux LLM, grands modèles entraînés en vue d'être utilisés comme fondement à la résolution de plusieurs types de tâches. Terme plus rare, mais retenu dans discussions au Parlement européen.
- Prompt : texte donné en entrée d'un modèle.
- **IA générative** : terme générique désignant les algorithmes qui, à partir d'une entrée (généralement un prompt), génèrent une création originale en sortie (texte, image, son...)
- **RLHF** : « Reinforcement Learning from Human Feedback », technique qui vient s'ajouter en complément des transformers (GPT) pour constituer ChatGPT. (voir slides suivantes)

We investigate the optimal model size and number of tokens for training a transformer language model under a given compute budget. We find that current large language models are significantly undertrained, a consequence of the recent focus on scaling language models whilst keeping the amount of training data constant. By training over 400 language models ranging from 70 million to over 16 billion parameters on 5 to 500 billion tokens, we find that for compute-optimal training, the model size and the number of training tokens should be scaled equally: for every doubling of model size the number of training tokens should also be doubled. We test this hypothesis by training a predicted compute optimal model, *Chinchilla*, that uses the same compute budget as *Gopher* but with 70B parameters and 4x more data. *Chinchilla* uniformly and significantly outperforms *Gopher* (280B), GPT-3 (175B) Jurassic-1 (178B), and Megatron-Turing NLG (530B) on a large range of downstream evaluation tasks. This also means that *Chinchilla* uses substantially less compute for fine-tuning and inference, greatly facilitating downstream usage. As a highlight, *Chinchilla* reaches a state-of-the-art average accuracy of 67.5% on the MMLU benchmark, greater than a 7% improvement over *Gopher*.

1. Introduction

Recently a series of *Large Language Models* (LLMs) have been introduced (Brown et al., 2020; Lieber et al., 2021; Rae et al., 2021; Smith et al., 2022; Thoppilan et al., 2022), with the largest dense language models now having over 500 billion parameters. These large autoregressive transformers (Vaswani et al., 2017) have demonstrated impressive performance on many tasks using a variety of evaluation protocols such as zero-shot, few-shot, and fine-tuning.

The compute and energy cost for training large language models is substantial (Rae et al., 2021; Thoppilan et al., 2022) and rises with increasing model size. In practice, the allocated training compute budget is often known in advance: how many accelerators are available and for how long we want to use them. Since it is typically only feasible to train these large models once, accurately estimating the best model hyperparameters for a given compute budget is critical (Tay et al., 2021).

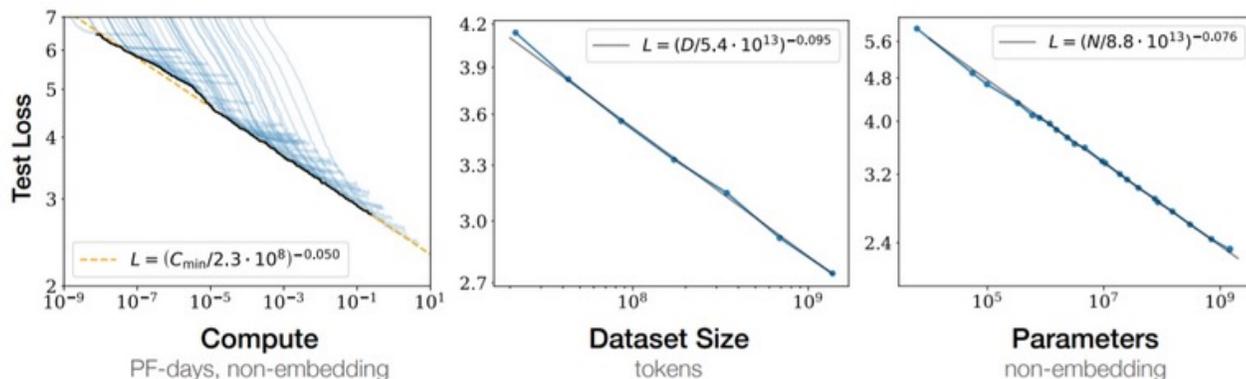
Kaplan et al. (2020) showed that there is a power law relationship between the number of parameters in an autoregressive language model (LM) and its performance. As a result, the field has been training larger and larger models, expecting performance improvements. One notable conclusion in Kaplan et al. (2020) is that large models should not be trained to their lowest possible loss to be compute optimal. Whilst we reach the same conclusion, we estimate that large models should be trained for many more training tokens than recommended by the authors. Specifically, given a 10x increase computational budget, they suggests that the size of the model should increase 5.5x while the number of training tokens should only increase 1.8x. Instead, we find that model size and the number of training tokens should be scaled in equal proportions.

Following Kaplan et al. (2020) and the training setup of GPT-3 (Brown et al., 2020), many of the recently trained large models have been trained for approximately 300 billion tokens (Table 1), in line with the approach of predominantly increasing model size when increasing compute.

Scaling laws

- **Scaling Laws** refer to the observed trend of some machine learning architectures (notably [transformers](#)) to scale their performance on predictable power law when given more compute, data, or parameters (model size), assuming they are not bottlenecked on one of the other resources. This has been observed as highly consistent over more than six orders of magnitude.

$$L(N, D) = \underbrace{\frac{406.4}{N^{0.34}}}_{\text{finite model}} + \underbrace{\frac{410.7}{D^{0.28}}}_{\text{finite data}} + \underbrace{1.69}_{\text{irreducible}}$$



Scaling laws graph from [Scaling Laws for Neural Language Models](#)

arXiv:2203.15556v1 [cs.CL] 29 Mar 2022