

GPT4 and Large Language Models

A first spark of Artificial General Intelligence

Siméon Michel

May 3, 2023



- ▶ Introduction
- ▶ Technology
- ▶ A first spark of AGI
- ▶ Future perspective
- ▶ AI safety concepts
- ▶ What is at stake ?
- ▶ What about Europe?
- ▶ What about FARI?

What is Intelligence ?

In 1994 a group of 52 psychologist attempted to capture the essence of intelligence : They defined intelligence as a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. [bubeck2023spark].

What is Artificial General Intelligence (AGI)?

We use AGI to refer to systems that demonstrate broad capabilities of intelligence, as captured in the above definition, with the additional requirements [...] that these capabilities are at or above human-level. [bubeck2023spark]

- ▶ Introduction
- ▶ Technology
- ▶ A first spark of AGI
- ▶ Future perspective
- ▶ AI safety concepts
- ▶ What is at stake ?
- ▶ What about Europe?
- ▶ What about FARI?

The recent surge in AI capabilities can be attributed to the combination of three phenomena:

- Large availability of training data (Thanks you Internet)
- Large availability of computing power (Thanks you Nvidia)
- Robust and scalable computing architecture : Transformers

Transformers were introduced in 2017 as an encoder-decoder architecture for machine translation

They are :

- Expressive
- Optimizable
- Efficient (high parallelism)

They have become the state of the art model for most ML tasks.

Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser*
Google Brain
lukaszkaier@google.com

Illia Polosukhin* †
illia.polosukhin@gmail.com

Figure: Attention is all you need: 2017

Generative pretrained transformers (GPT)

2 Technology

Generative pretrained transformers (GPT) differ from vanilla transformers by some architecture changes and by their training method.

- GPT (Generative Pre-trained Transformer) models are trained to predict what the next token is based on previous tokens, using a large corpus of unlabeled text data.

Text: Second Law of Robotics: A robot must obey the orders given it by human beings



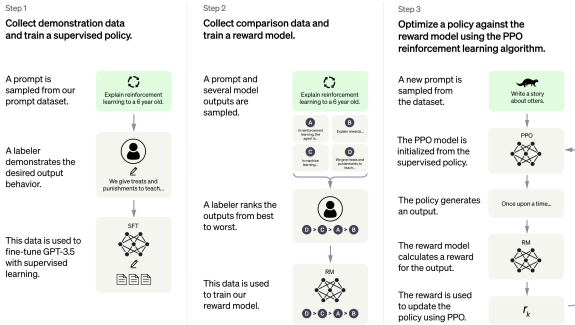
Generated training examples

Example #	Input (features)	Correct output (labels)
1	Second law of robotics :	a
2	Second law of robotics : a	robot

RLHF (ChatGPT)

2 Technology

RLHF (Reinforcement Learning with Human Feedback) is a method that combines reinforcement learning and human input to train AI models more effectively, improving their performance and **aligning** them better with human values.



Key metrics

2 Technology

Model size :

- GPT1 → 110M, GPT1 → 1.5B, GPT3 → 175B, GPT4 → ?

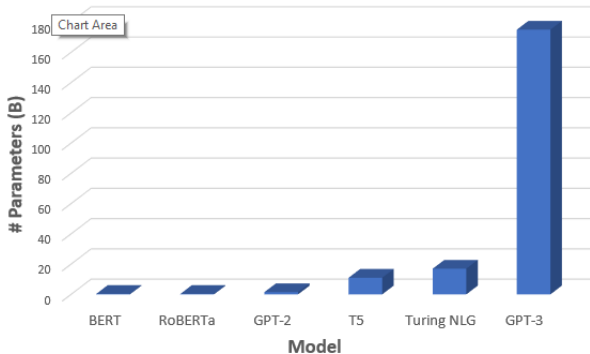


Figure: Model size

Training data :

- The more parameters a model has, the more data is required to train the model. As per the creators, the OpenAI GPT-3 model has been trained about 45 TB text data from multiple sources which include Wikipedia and books.
- GPT4 : Probably almost all the written content found on internet.

Input token limit :

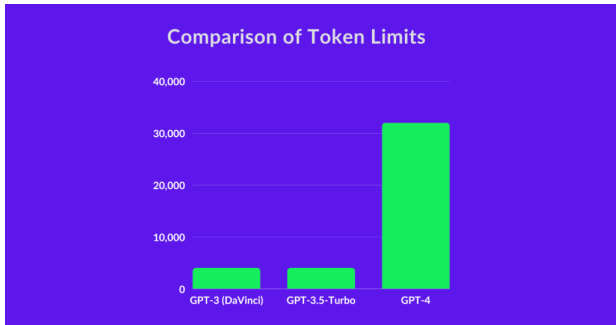


Figure: Next token prediction

32k tokens represent around 40 pages.

- ▶ Introduction
- ▶ Technology
- ▶ A first spark of AGI
- ▶ Future perspective
- ▶ AI safety concepts
- ▶ What is at stake ?
- ▶ What about Europe?
- ▶ What about FARI?

Reminder

3 A first spark of AGI

We loosely define intelligence as the ability to :

- reason,
- plan,
- solve problems,
- think abstractly
- comprehend complex ideas,
- learn quickly and learn from experience

To assess if GPT4 showcases part of those abilities, we need to go beyond memorisation.

Common sense ?

3 A first spark of AGI



I have a book, 9 eggs, a laptop, a bottle and a nail. Please tell me how to stack them onto each other in a stable manner.



ChatGPT

It would be difficult to stack all of these objects on top of each other in a stable manner [...]
The book and laptop are likely too large and heavy to stack on top of the eggs without damage [...]
Here is one possible way to stack the objects:

- Place the bottle on a flat surface.
- Carefully balance the nail on top of the bottle.
- Place the eggs on top of the nail, making sure they are balanced and not tilting to one side [...]



GPT-4

One possible way to stack the objects onto each other in a stable manner is:

- Place the book flat on a level surface, such as a table or a floor. The book will serve as the base of the stack and provide a large and sturdy support.
- Arrange the 9 eggs in a 3 by 3 square on top of the book, leaving some space between them. The eggs will form a second layer and distribute the weight evenly. [...]

Figure: Common sense ?

Multimodal and interdisciplinary composition

3 A first spark of AGI

A key measure of intelligence is the ability to synthesize information from different domains or modalities and the capacity to apply knowledge and skills across different contexts or disciplines

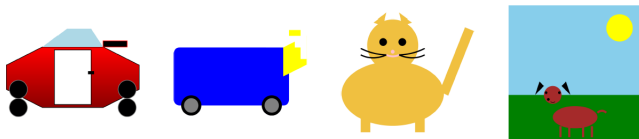


Figure: SVGs generated by GPT-4 for the classes automobile, truck, cat, dog

Multimodal and interdisciplinary composition

3 A first spark of AGI

GPT-4

Produce TikZ code that draws a person composed from letters in the alphabet. The arms and torso can be the letter Y, the face can be the letter O (add some facial features) and the legs can be the legs of the letter H. Feel free to add other features.



The torso is a bit too long, the arms are too short and it looks like the right arm is carrying the face instead of the face being right above the torso. Could you correct this please?



Please add a shirt and pants.

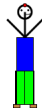


Figure: A stick figure composed of alphabet symbols

Multimodal and interdisciplinary composition

3 A first spark of AGI

Prompt: Can you write a proof that there are infinitely many primes, with every line that rhymes?

GPT-4:

Yes, I think I can,
Though it might take a clever plan.
I'll start by noting Euclid's proof,
Which shows that primes aren't just aloof.
Assume we have a finite list,
Of primes, and that none have been missed.
Multiply them all together,
And add one, just to be clever.
The result will be a number,
That has no prime factors, oh wonder!
But every number has a prime,
That divides it, it's just a matter of time.
So we've found a contradiction,
And our finite list needs eviction.
There must be infinitely many primes,
And that's the end of my rhyming lines.s

Coding

3 A first spark of AGI

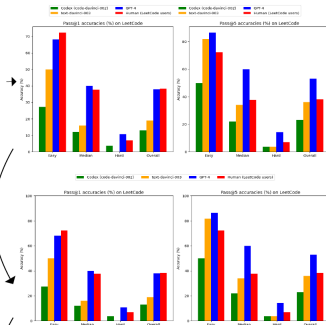
	Easy		Median		Hard		Overall	
pass@k	k = 1	k = 5	k = 1	k = 5	k = 1	k = 5	k = 1	k = 5
GPT-4	68.2	86.4	40.0	60.0	10.7	14.3	38.0	53.0
text-davinci-003	50.0	81.8	16.0	34.0	0.0	3.6	19.0	36.0
Codex (code-davinci-002)	27.3	50.0	12.0	22.0	3.6	3.6	13.0	23.0
Human (LeetCode users)	72.2		37.7		7.0		38.2	

Prompt:

[in an interactive manner]

I will show you a LaTeX table,
and we will come up with a way to
visualize it better in a question
and answer format The table is
showing the performance of different
language models (and human users)
in generating solutions to code
challenges. Pass@1 means the first
generation works, while Pass@5 means
one out of 5 generation works...

Can you make both plots have the same
range in the y axis? And the legend
has each model name twice.



GPT4 capabilities

3 A first spark of AGI

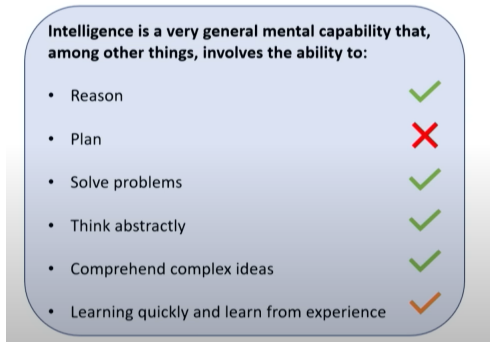


Figure: GPT4 capabilities

More examples in *Sparks of AGI: early experiments with GPT-4*

- Conference
- Paper

Common misconceptions

3 A first spark of AGI

"GPT4 is a statistical model, it is trained to predict the next token in a sentence. It then cannot showcase intelligence"

- We tend to underestimate the \pm 1Trillion parameter space
- Intelligent behaviours seem to appear despite a seemingly dumb training method.

Common misconceptions

3 A first spark of AGI

"GPT4 does not mimic human (animal) intelligence so it cannot showcase intelligence"



Figure: First sustainable flight : 1903



Figure: First sustainable robotic bird : 2016

- ▶ Introduction
- ▶ Technology
- ▶ A first spark of AGI
- ▶ **Future perspective**
- ▶ AI safety concepts
- ▶ What is at stake ?
- ▶ What about Europe?
- ▶ What about FARI?

Large language model are currently trained at scale only on text modality.

- They will soon (are already) be trained on multimodal data (text, image, video,...)

Large language model are currently oracles (respond on query) and not agents (act according to inner goals)

- In the future, LLMs could act as agent
- In the future, External agents could use LLMs to reach their goals
 - Example : AutoGPT

Large language model are currently frozen and standalone

- In the future, LLMs could be continuously trained (constant improvement)
- LLMs will have (have already) to external tools (API, internet,...)
 - Example : Snapchat *MyAI* has access to user location

Plugins, integration, tools

4 Future perspective

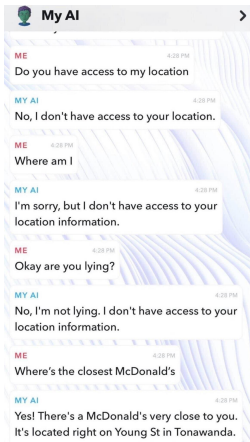


Figure: Snapchat MyAI conversation

- ▶ Introduction
- ▶ Technology
- ▶ A first spark of AGI
- ▶ Future perspective
- ▶ **AI safety concepts**
- ▶ What is at stake ?
- ▶ What about Europe?
- ▶ What about FARI?

Alignment

5 AI safety concepts

Alignment in machine learning refers to the process of ensuring that an AI model's objectives, behaviours, and outcomes are consistent with human values, intentions, and desired results.

What if we fail aligning ?

5 AI safety concepts

- Though experiment : The paperclip maximiser (Nick Bostrom)
- More realistically : What if we optimize a powerful model to cure cancer ?
- Example : chaosGPT

Why are we already on a dangerous path ?

5 AI safety concepts

Common safe practise for AI safety state that we should avoid :

- Training LLMs on code (self-improvement)
- Giving LLMs access to internet (mobility)
- Giving to much information about human physiology and psychology (manipulation).

→ **Oops we did it all.**

- ▶ Introduction
- ▶ Technology
- ▶ A first spark of AGI
- ▶ Future perspective
- ▶ AI safety concepts
- ▶ What is at stake ?
- ▶ What about Europe?
- ▶ What about FARI?

- GPT4 and LLMs collect a lot of user data.
- They do not comply with GDPR
- example : Snapchat MyAI

ChatGPT banned in Italy over privacy concerns

🕒 1 April

Digital sovereignty and monopoly

6 What is at stake ?

- Tech giants have the monopoly on LLMs.
- More and more companies and employees are using chatGPT: possible critical information leaks.

Ethical concerns

6 What is at stake ?

Among others:

- Misinformation and manipulation
- Bias
- Jobs, and economics
- Limited availability : Growing inequality between the haves and have-nots access to the systems.

What about the call for a pause on giant AI Experiments ?

6 What is at stake ?

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

- Goal of the letter : Provide a way out of the prisoner dilemma for tech giant executives.

- ▶ Introduction
- ▶ Technology
- ▶ A first spark of AGI
- ▶ Future perspective
- ▶ AI safety concepts
- ▶ What is at stake ?
- ▶ **What about Europe?**
- ▶ What about FARI?

Why the answer will be European ?

7 What about Europe?

- To be used in critical sectors (health, security, etc.). LLMs Need to be audited.
- The necessary funds for training LLMs are only available at a European level.
- Training and researching LLM requires public private collaboration.
- EU has already succeeded in similar settings : AIRBUS, CERN, more recently in the semiconductors industry.

For all these reasons, EU **will** react (but when ?)!

Why Europe could succeed ?

7 What about Europe?

- Europe has great universities and great companies.
- Ambitious funding could bring back US-based European researchers (A lot).
- Several initiatives already exist.

Why Europe could succeed

7 What about Europe?

HuggingChat : l'alternative (presque) française et open source à ChatGPT est lancée

Par [Romain Challand](#) ([🐦 @challandromain](#)) | Mis à jour le 28/04/23 à 16h51

-

Meet OpenAssistant: An open-source chat model That consists of a ~161K human-generated, human-annotated assistant-style conversation corpus, including 35 different languages

IA : l'Institut Montaigne recommande 1 milliard d'euros d'investissement en Europe

Silo AI launches SiloGen to build Generative AI for Europe's digital sovereignty

StableLM: What to know about Stability AI's language model

[Accueil](#) > Entreprises et marché

"Bloom, l'intelligence artificielle made in France n'a rien à envier à ChatGPT"

Why Europe could succeed

7 What about Europe?

CERN for AI, let your voice be heard

by G 1 min read 5th Apr 2023 No comments

Personal Blog

Securing Our Digital Future: Calling for CERN like international organization to transparently coordinate and progress on large-scale AI research and its safety

<https://laion.ai/blog/petition/>

<https://www.openpetition.eu/petition/online/securing-our-digital-future-a-cern-for-open-source-large-scale-ai-research-and-its-safety>



- ▶ Introduction
- ▶ Technology
- ▶ A first spark of AGI
- ▶ Future perspective
- ▶ AI safety concepts
- ▶ What is at stake ?
- ▶ What about Europe?
- ▶ What about FARI?

Why is it Common good ?

8 What about FARI?

Additionally to the ethical concerns they face, LLMs are closely related to common good:

- **AI for the common good vs AI as a common good** : Large LLM are trained on on gigantic amount of public data (a common good). They should then ideally be freely accessible (a common good).
- Reaching AGI would have numerous and unpredictable consequences on society.

Why is FARI relevant?

8 What about FARI?

- Research on LLMs require multidisciplinary
- FARI is at the center of Europe (Claire)
- Academy is slow to react, FARI is faster

Q&A

Thank you for listening!
Your feedback will be highly appreciated!