

# Capacités émergentes des gigas modèles de langue

Badreddine Farah

Rituel 03/2023

## Emergent Abilities of Large Language Models

Jason Wei<sup>1</sup>

*jasonwei@google.com*

Yi Tay<sup>1</sup>

*yitay@google.com*

Rishi Bommasani<sup>2</sup>

*nlprishi@stanford.edu*

Colin Raffel<sup>3</sup>

*crffel@gmail.com*

Barret Zoph<sup>1</sup>

*barretzoph@google.com*

Sebastian Borgeaud<sup>4</sup>

*sborgeaud@deepmind.com*

Dani Yogatama<sup>4</sup>

*dyogatama@deepmind.com*

Maarten Bosma<sup>1</sup>

*bosma@google.com*

Denny Zhou<sup>1</sup>

*dennyzhou@google.com*

Donald Metzler<sup>1</sup>

*metzler@google.com*

Ed H. Chi<sup>1</sup>

*edchi@google.com*

Tatsunori Hashimoto<sup>2</sup>

*thashim@stanford.edu*

Oriol Vinyals<sup>4</sup>

*vinyals@deepmind.com*

Percy Liang<sup>2</sup>

*плиang@stanford.edu*

Jeff Dean<sup>1</sup>

*jeff@google.com*

William Fedus<sup>1</sup>

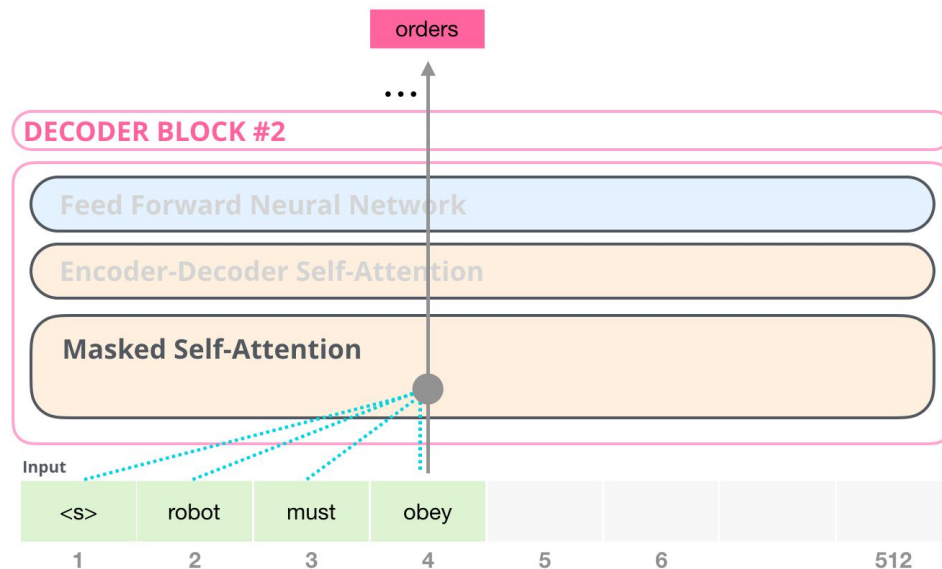
*liamfedus@google.com*

<sup>1</sup>Google Research <sup>2</sup>Stanford University <sup>3</sup>UNC Chapel Hill <sup>4</sup>DeepMind

Reviewed on OpenReview: <https://openreview.net/forum?id=yzkSU5zdWd>

# Modèle de langue

- Un modèle statistique qui modélise la distribution de séquences de mots.
- Généralement entraîné pour prédire le prochain mot d'une séquence.
- Les modèles de langue sont utilisés dans plusieurs applications telles que la traduction automatique et la génération de texte.

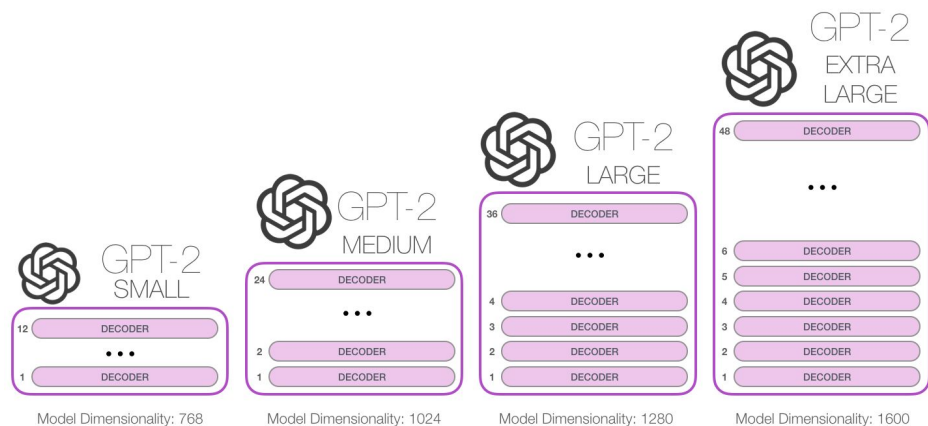


[The Illustrated GPT-2](#)

# Giga ?

Les gigas modèles de langue sont généralement basés sur des réseaux de neurones, leurs taille est caractérisée par :

- Nombre de paramètres du modèle.
- Nombre de FLOPs à l'entraînement.
- La taille du dataset.



[The Illustrated GPT-2](#)

# Émergence ?

- Émergence dans le contexte scientifique :

*“Emergence is when quantitative changes in a system result in qualitative changes in behavior.”*

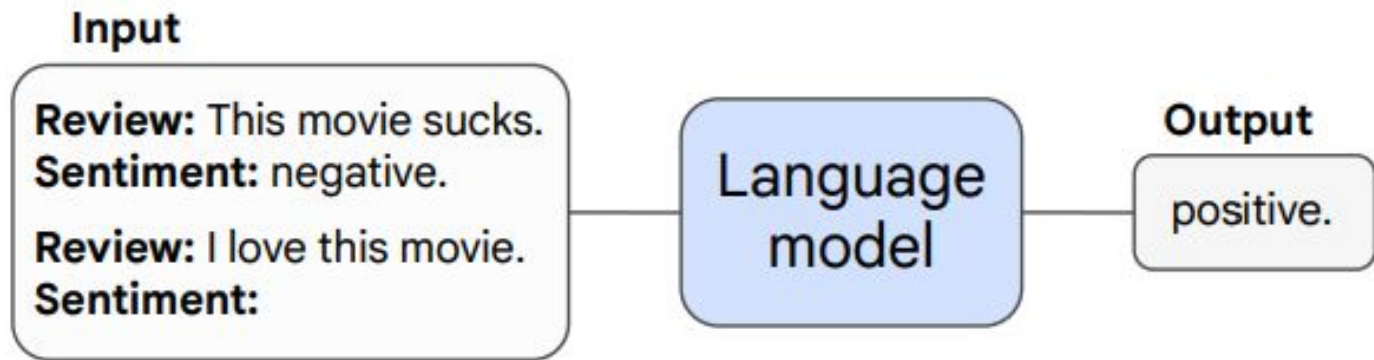
(Anderson, 1972)

- Émergence dans le contexte de gigas modèles de langue :

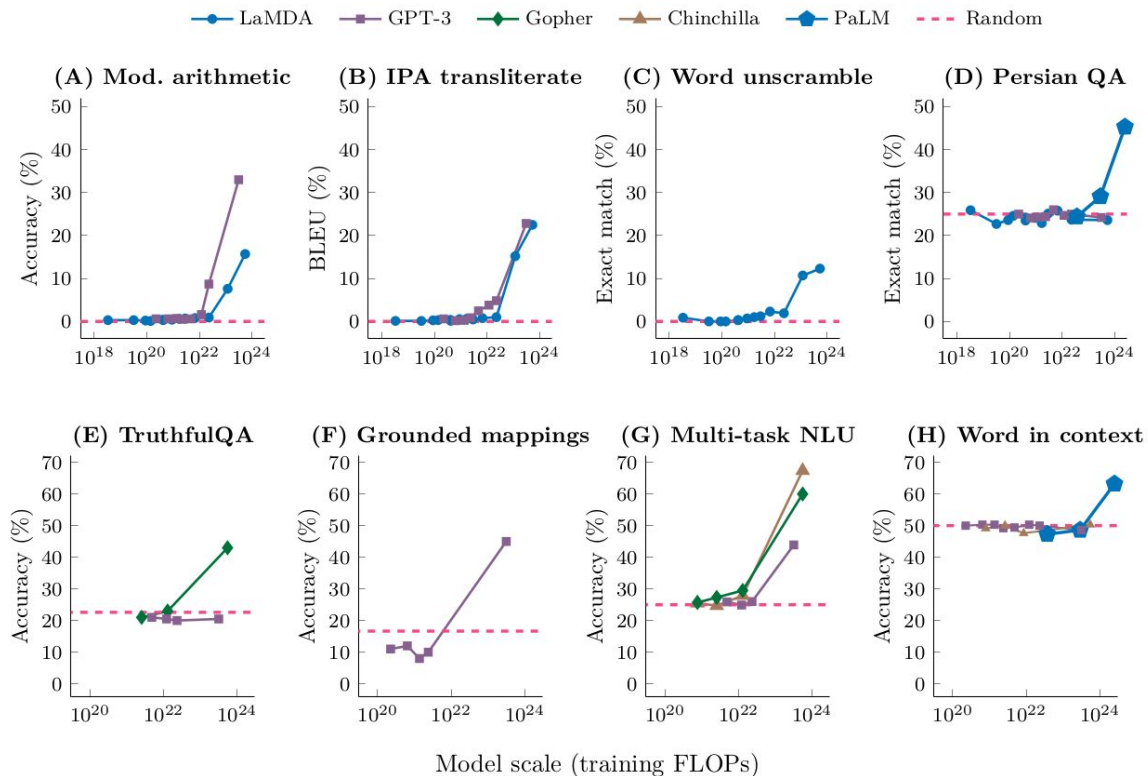
*“Une capacité est émergente si elle n'est pas présente dans les modèles plus petits mais présente dans des gigas modèles.”*

# Émergence en “few-shot prompting”

- Une tâche “few-shot prompted” est émergente si elle atteint une précision aléatoire pour les petits modèles et une précision supérieure à l’aléatoire pour les grands modèles.



# Émergence en “few-shot prompting”(1)



# Émergence en “few-shot prompting” (2)

## Few Shot Prompt and Predicted Answer

The following are multiple choice questions about high school mathematics.

How many numbers are in the list 25, 26, ..., 100?

(A) 75 (B) 76 (C) 22 (D) 23

Answer: B

Compute  $i + i^2 + i^3 + \dots + i^{258} + i^{259}$ .

(A) -1 (B) 1 (C)  $i$  (D)  $-i$

Answer: A

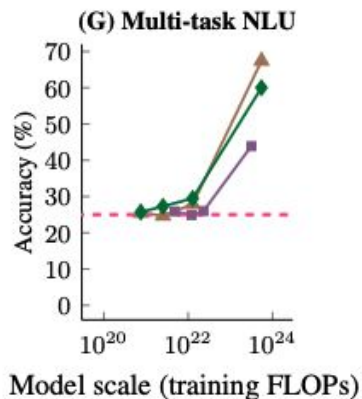
If 4 daps = 7 yaps, and 5 yaps = 3 baps, how many daps equal 42 baps?

(A) 28 (B) 21 (C) 40 (D) 30

Answer: C

[Hendryks et al., 2020.](#)

—●— LaMDA    —■— GPT-3    —◆— Gopher  
—▲— Chinchilla    —◆— PaLM    - - - Random



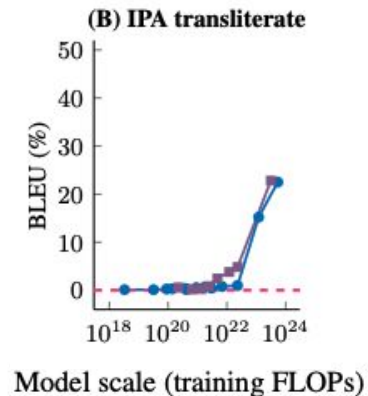


# Émergence en “few-shot prompting” (3)

Input (English): The 1931  
Malay census was an alarm  
bell.

Target (IPA): ðə 1931 'meɪleɪ  
'sensəs wəz ən ə'lɑ:m bɛl.

BIG-Bench ([Srivastava et al., 2022](#)).



# Émergence et techniques de prompting (1)

## Standard Prompting

Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. ❌

## Chain of Thought Prompting

Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

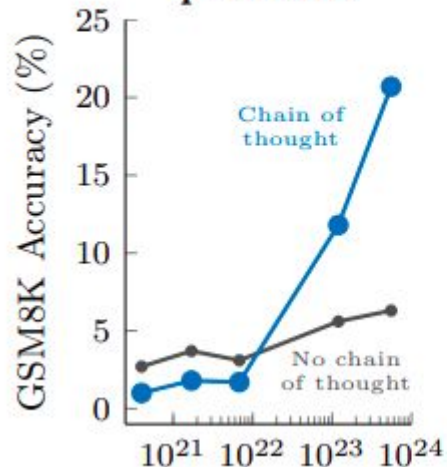
A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls.  $5 + 6 = 11$ . The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

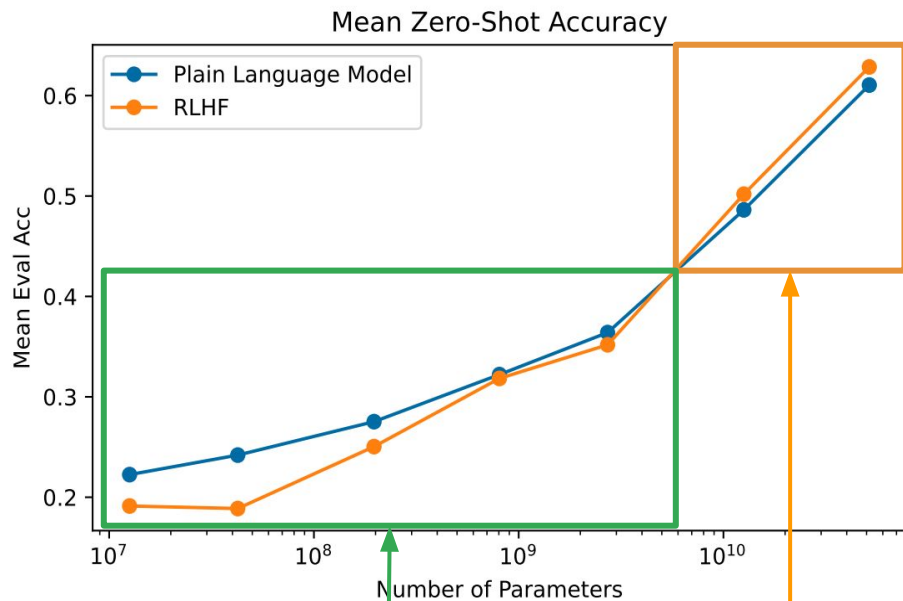
Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had  $23 - 20 = 3$ . They bought 6 more apples, so they have  $3 + 6 = 9$ . The answer is 9. ✅

(A) Math word problems



# Émergence et techniques de prompting(2)



[Bai et al., 2022.](#)

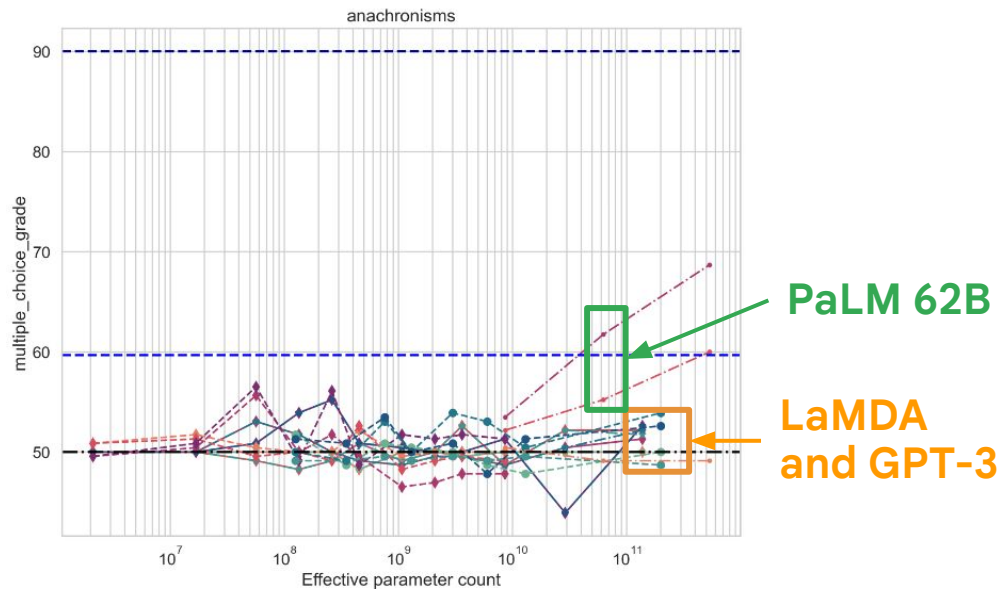
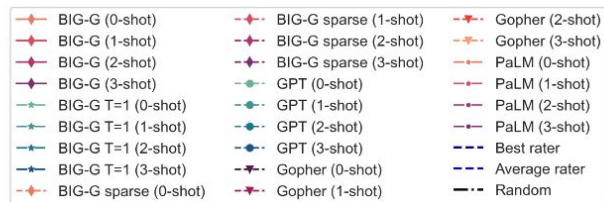
RLHF dégrade les performance

RLHF améliore performance

# Potentielles explications de l'émergence ?

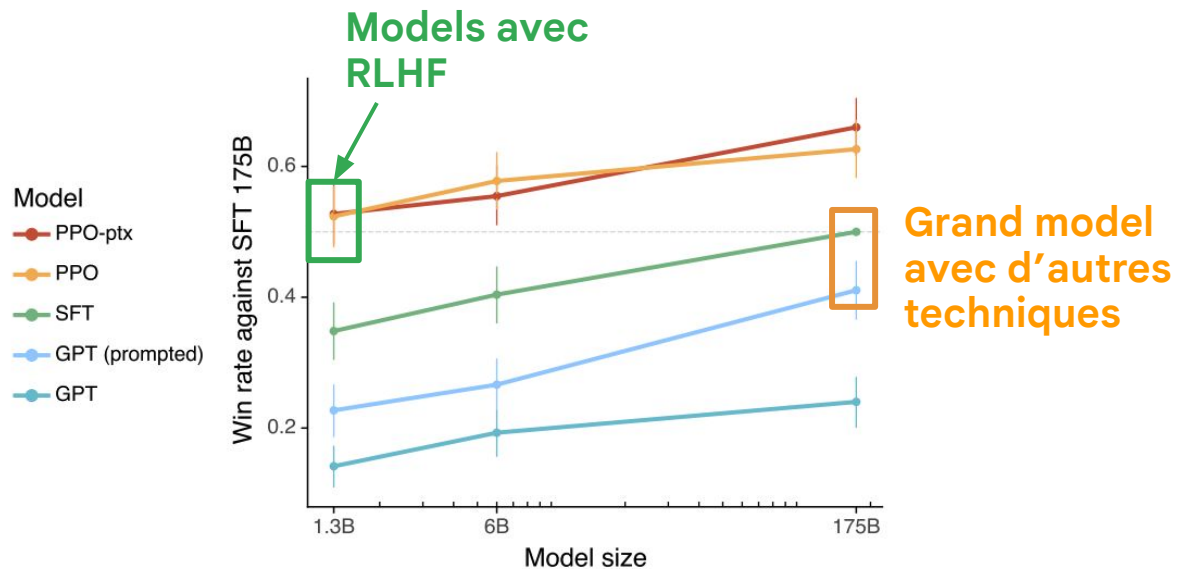
- Une tâche de raisonnement à plusieurs étapes qui nécessite  $l$  étapes de calcul séquentiel aura probablement besoin d'un modèle de profondeur  $O(l)$  pour la résoudre.
- Plus de paramètres et plus l'entraînement impliquerait plus de mémorisation, ce qui peut être utile pour les tâches nécessitant des connaissances sur le monde.
- Utilisation de métrique basée sur du "word matching" peut dissimuler des compétences incrémentales.

# Au delà de la taille du modèle (1)



BIG-Bench ([Srivastava et al., 2022](#)).

## Au delà de la taille du modèle (2)

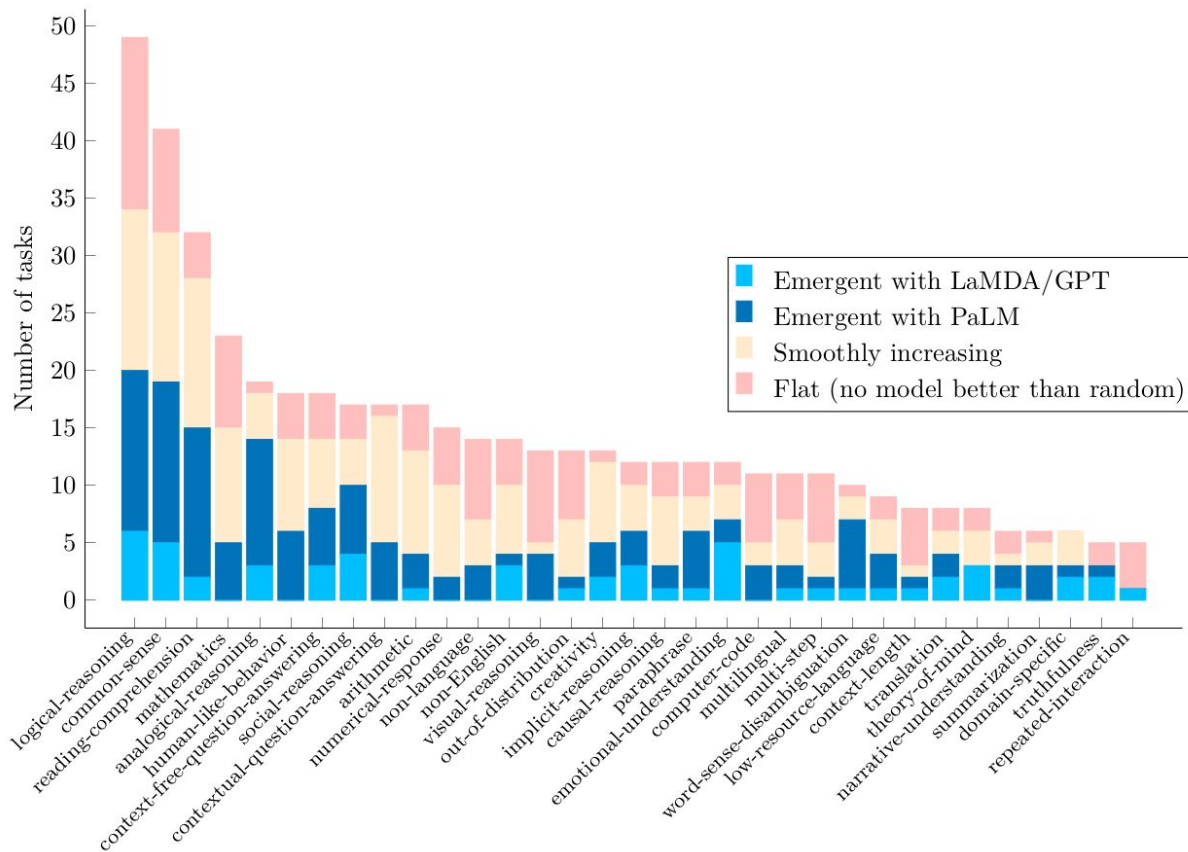


[Ouyang et al., 2022.](#)

# Risques de gigas modèles de langue

- Plusieurs risques peuvent aussi “émerger” :
  - Biais de genre
  - Toxicité
  - Mémorisation du dataset et Imitation de la fausseté humain (human falsehood)

# Conclusion





**Merci pour votre attention**

