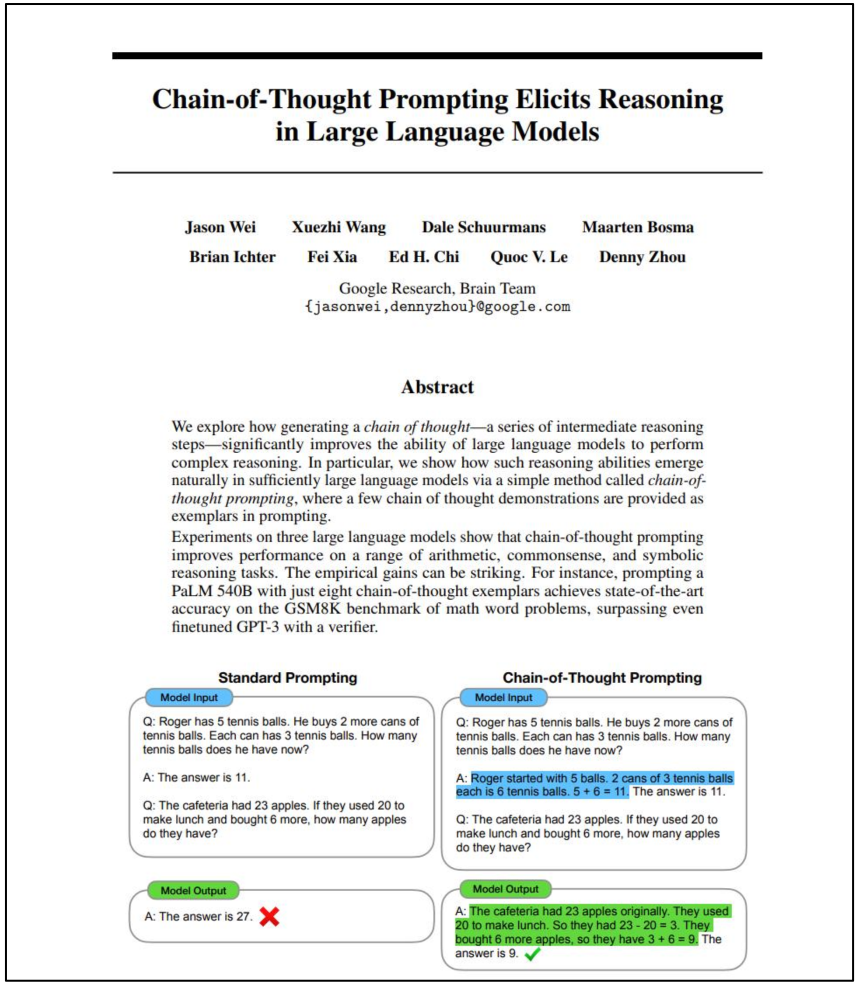
Reasoning with Foundation Models

Zhenglin Pan | MEng 22-Fall

Based on the 2 papers

- Chain-of-thought prompting elicits reasoning in large language models[1]
- A Survery of Reasoning with Foundation Models [2]



Chain-of-Thought Prompting Elicits Reasoning in Large Language Models

Jason Wei Xuezhi Wang Dale Schuurmans Maarten Bosma
Brian Ichter Fei Xia Ed H. Chi Quoc V. Le Denny Zhou

Google Research, Brain Team
{jasonwei,dennyzhou}@google.com

Abstract

We explore how generating a *chain of thought*—a series of intermediate reasoning steps—significantly improves the ability of large language models to perform complex reasoning. In particular, we show how such reasoning abilities emerge naturally in sufficiently large language models via a simple method called *chain-of-thought prompting*, where a few chain of thought demonstrations are provided as exemplars in prompting.

Experiments on three large language models show that chain-of-thought prompting improves performance on a range of arithmetic, commonsense, and symbolic reasoning tasks. The empirical gains can be striking. For instance, prompting a PaLM 540B with just eight chain-of-thought exemplars achieves state-of-the-art accuracy on the GSM8K benchmark of math word problems, surpassing even finetuned GPT-3 with a verifier.

Standard Prompting

Model 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

Model 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 Survey of Reasoning with Foundation Models

Jiankai Sun[1], Chuanyang Zheng[1], Enze Xie[§2], Zhengying Liu[§2], Ruihang Chu[1], Jianing Qiu[1], Jiaqi Xu[1], Mingyu Ding[3], Hongyang Li[4], Mengzhe Geng[1], Yue Wu[2], Wenhai Wang[1], Junsong Chen[2,6], Zhangyue Yin[11], Xiaozhe Ren[2], Jie Fu[5], Junxian He[5], Wu Yuan[1], Qi Liu[3], Xihui Liu[3], Yu Li[1], Hao Dong[7], Yu Cheng[1], Ming Zhang[7], Pheng Ann Heng[1], Jifeng Dai[8,4], Ping Luo[3,4], Jingdong Wang[9], Ji-Rong Wen[10], Xipeng Qiu[11], Yike Guo[5], Hui Xiong[12], Qun Liu[2], Zhenguo Li[2]

[1]The Chinese University of Hong Kong.
[2]Huawei Noah's Ark Lab.
[3]The University of Hong Kong.
[4]Shanghai AI Lab.
[5]Hong Kong University of Science and Technology.
[6]Dalian University of Technology.
[7]Peking University.
[8]Tsinghua University.
[9]Hefei University of Technology.
[10]Renmin University of China.
[11]Fudan University.
[12]Hong Kong University of Science and Technology (Guangzhou).

Abstract

Reasoning, a crucial ability for complex problem-solving, plays a pivotal role in various real-world settings such as negotiation, medical diagnosis, and criminal investigation. It serves as a fundamental methodology in the field of Artificial General Intelligence (AGI). With the ongoing development of foundation models, e.g., Large Language Models (LLMs), there is a growing interest in exploring their abilities in reasoning tasks. In this paper, we introduce seminal foundation models proposed or adaptable for reasoning, highlighting the latest advancements in various reasoning tasks, methods, and benchmarks. We then delve into the potential

§Project Lead. Email: {xie.enze,liuzhengying2}@huawei.com

1

There're two take-aways

- What can we do with Foundation Models (and use them in your research)?
- How could we prompt Foundation Models?

This is NOT a lesson

I am just sharing...

Paper 1

Jan 2024

A Survey of Reasoning with Foundation Models

Jiankai Sun[1], Chuanyang Zheng[1], Enze Xie[§2], Zhengying Liu[§2], Ruihang Chu[1], Jianing Qiu[1], Jiaqi Xu[1], Mingyu Ding[3], Hongyang Li[4], Mengzhe Geng[1], Yue Wu[2], Wenhai Wang[1], Junsong Chen[2,6], Zhangyue Yin[11], Xiaozhe Ren[2], Jie Fu[5], Junxian He[5], Wu Yuan[1], Qi Liu[3], Xihui Liu[3], Yu Li[1], Hao Dong[7], Yu Cheng[1], Ming Zhang[7], Pheng Ann Heng[1], Jifeng Dai[8,4], Ping Luo[3,4], Jingdong Wang[9], Ji-Rong Wen[10], Xipeng Qiu[11], Yike Guo[5], Hui Xiong[12], Qun Liu[2], Zhenguo Li[2]

[1]The Chinese University of Hong Kong.
[2]Huawei Noah's Ark Lab.
[3]The University of Hong Kong.
[4]Shanghai AI Lab.
[5]Hong Kong University of Science and Technology.
[6]Dalian University of Technology.
[7]Peking University.
[8]Tsinghua University.
[9]Hefei University of Technology.
[10]Renmin University of China.
[11]Fudan University.
[12]Hong Kong University of Science and Technology (Guangzhou).

Abstract

Reasoning, a crucial ability for complex problem-solving, plays a pivotal role in various real-world settings such as negotiation, medical diagnosis, and criminal investigation. It serves as a fundamental methodology in the field of Artificial General Intelligence (AGI). With the ongoing development of foundation models, e.g., Large Language Models (LLMs), there is a growing interest in exploring their abilities in reasoning tasks. In this paper, we introduce seminal foundation models proposed or adaptable for reasoning, highlighting the latest advancements in various reasoning tasks, methods, and benchmarks. We then delve into the potential

[§]Project Lead. Email: {xie.enze,liuzhengying2}@huawei.com

1

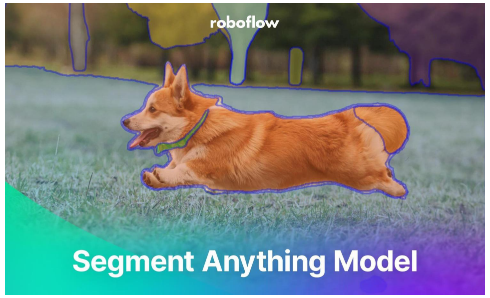
☜ A 90-page Survey of recent Foundation Models

Introduces many FMs in various research area

- Visual
- Multimodal
- Mathematical
- Agent
- Casual
- more...

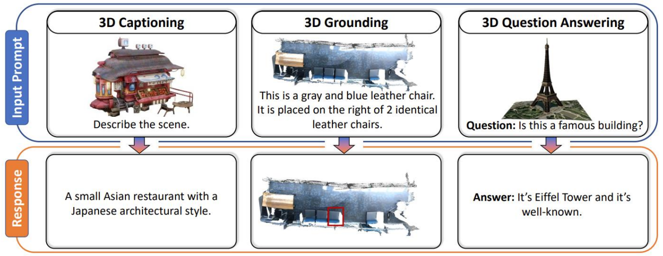
Miscellaneous! We only cover Visual and Multimodal here!

We have so many FMs

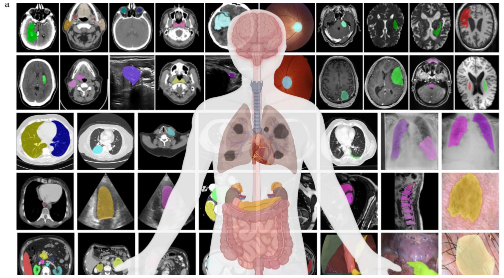


Researchers are using FMs in their reseach

3D-LLM[10]

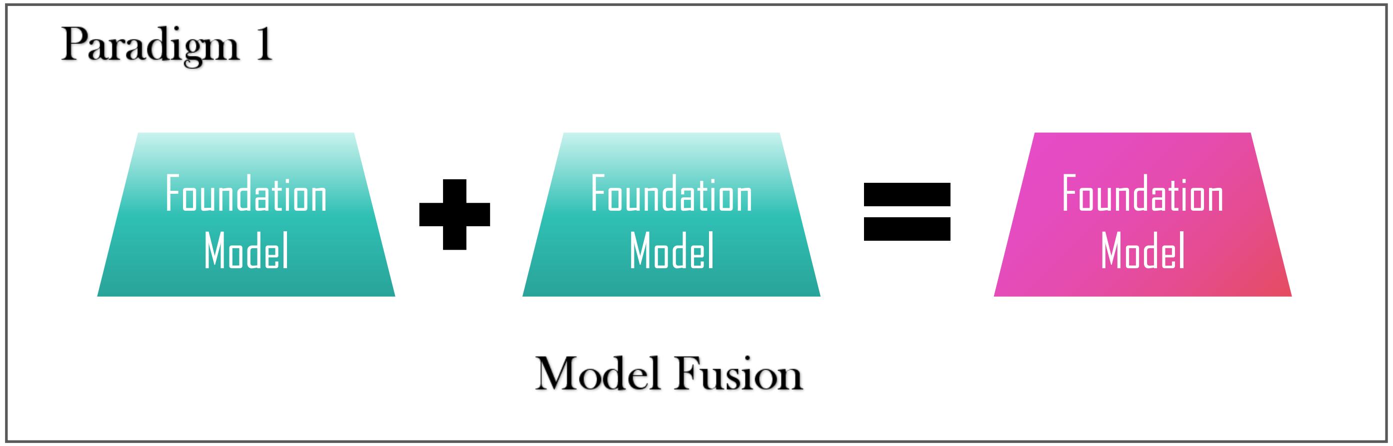


MedSAM[12]

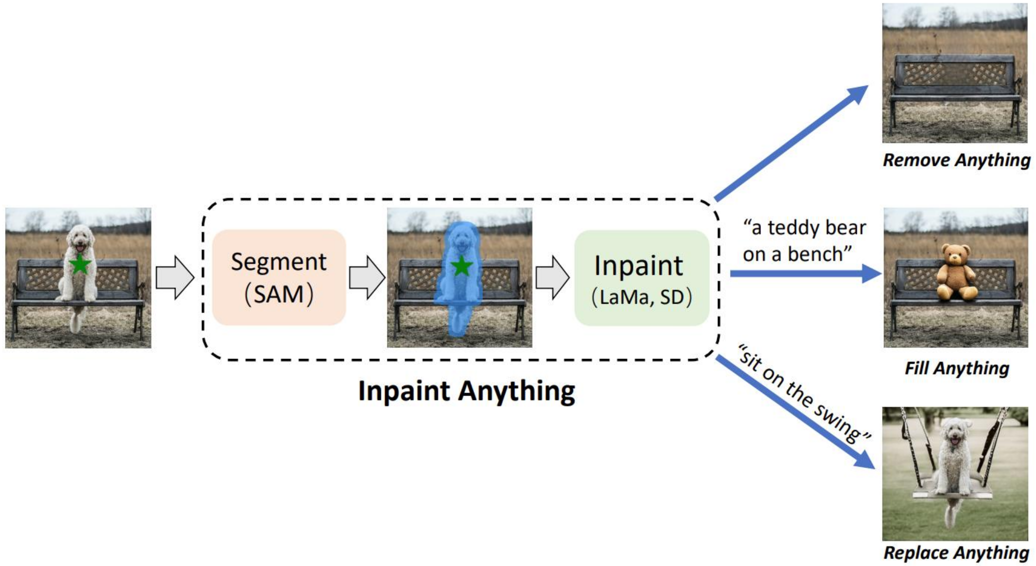


Briefly, there're 3 practical paradigms in using FMs

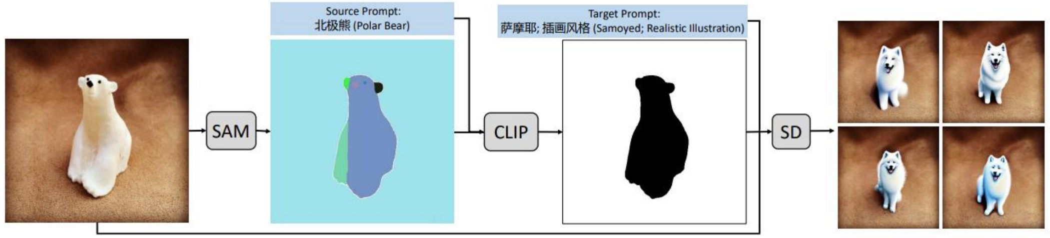
FM Paradigm 1



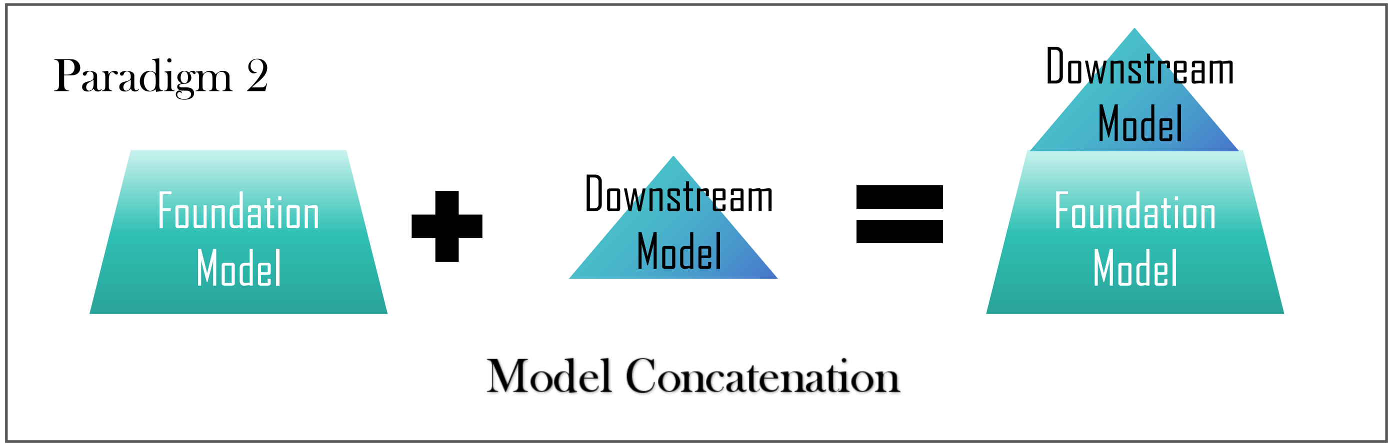
Inpaint Anything[3] = SD[4] + SAM[5]



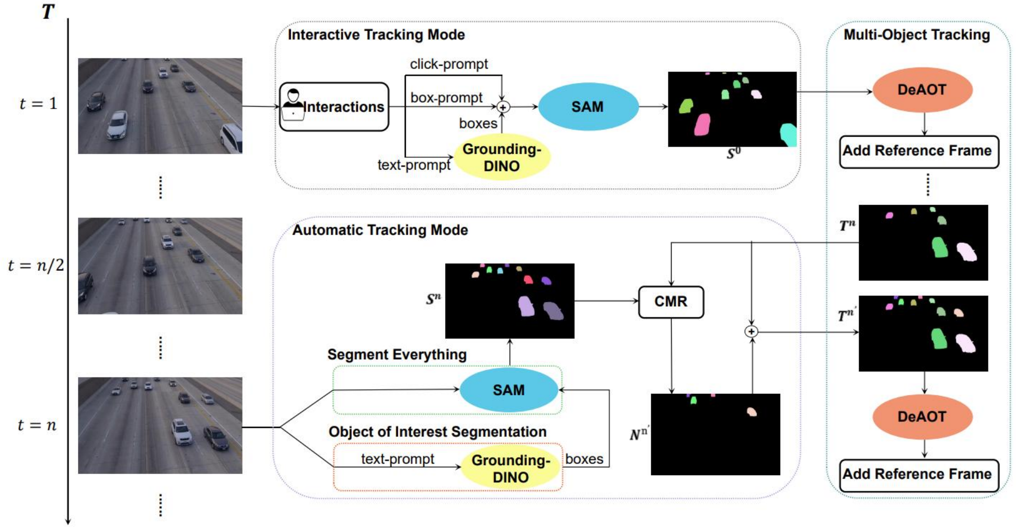
Edit Everything[6] = SAM[5] + CLIP[7] + SD[4]



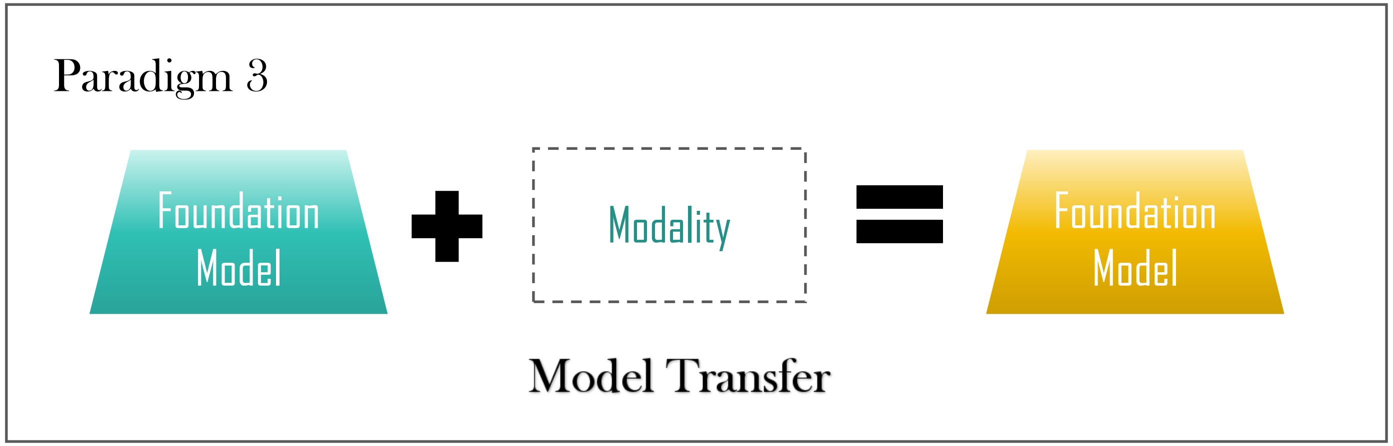
FM Paradigm 2



SAM-Track[8] = SAM[5] + DeAOT[9]



FM Paradigm 3

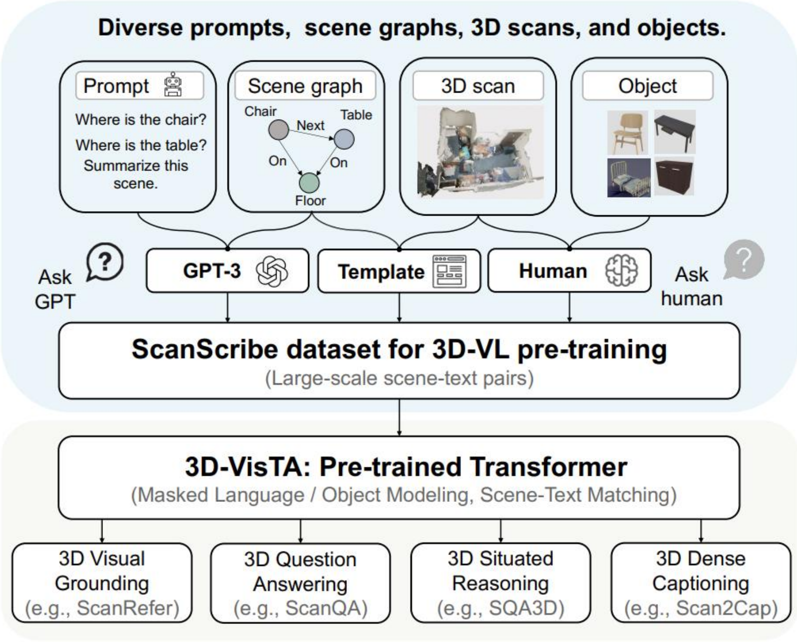


No downstream small models

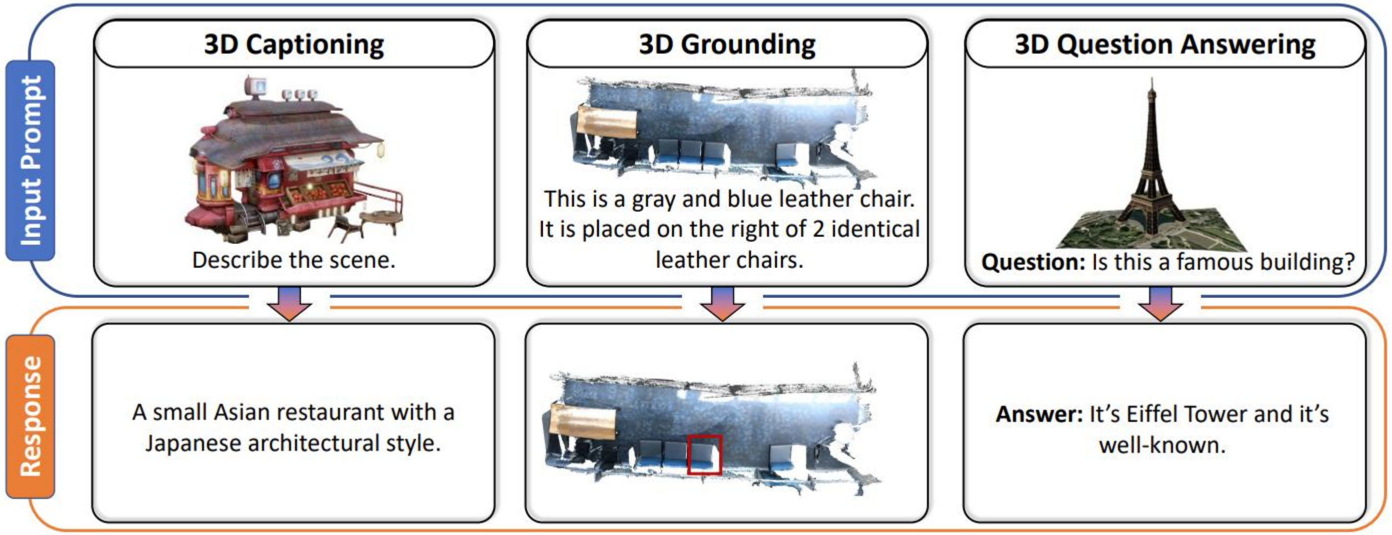
it's basically introducing new task

i.e transfering knowlege from FMs

3D-VisTa[11] = PointCloud + CLIP[7]

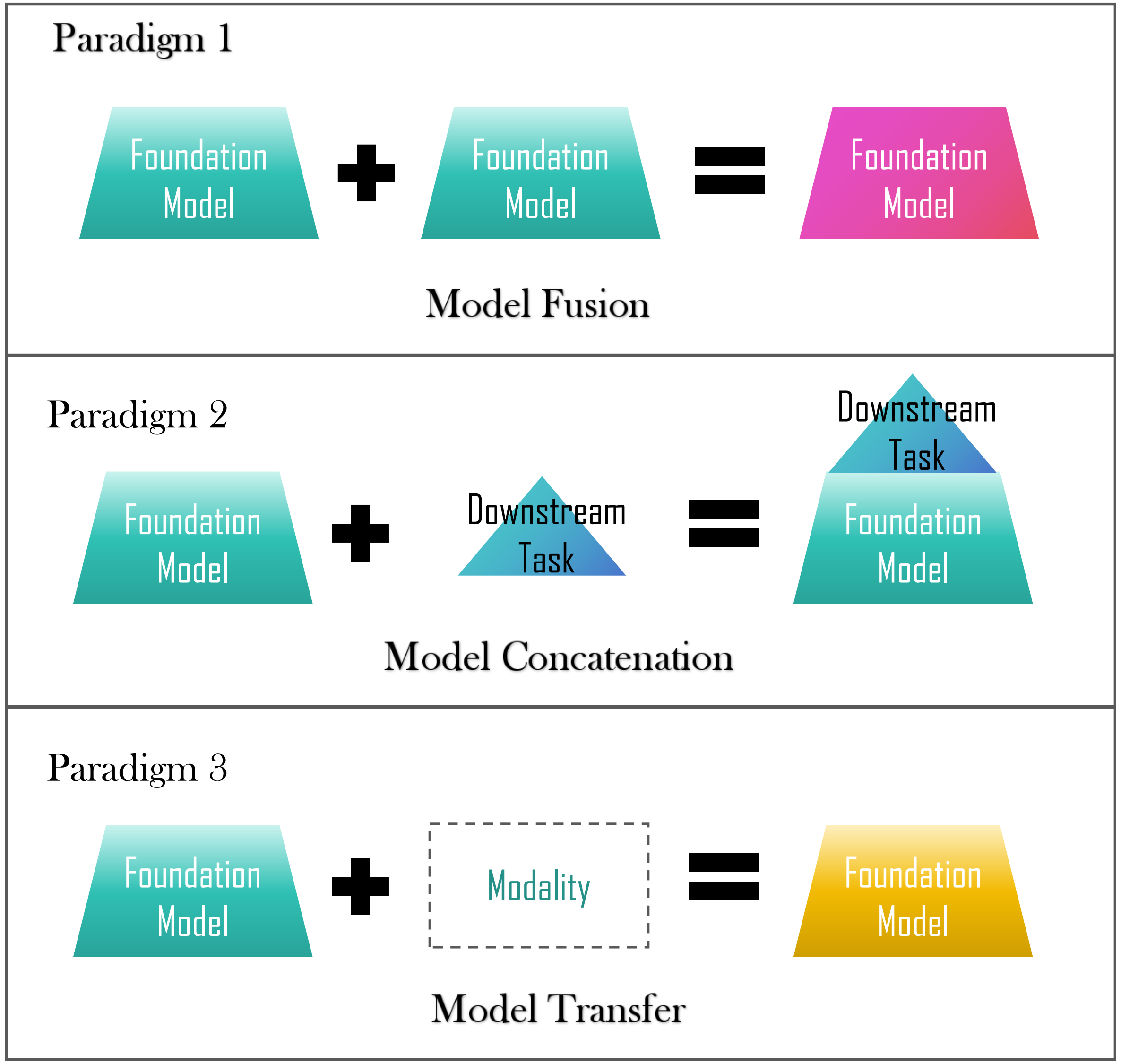


3D-LLM[10] = 3D objects + LLM

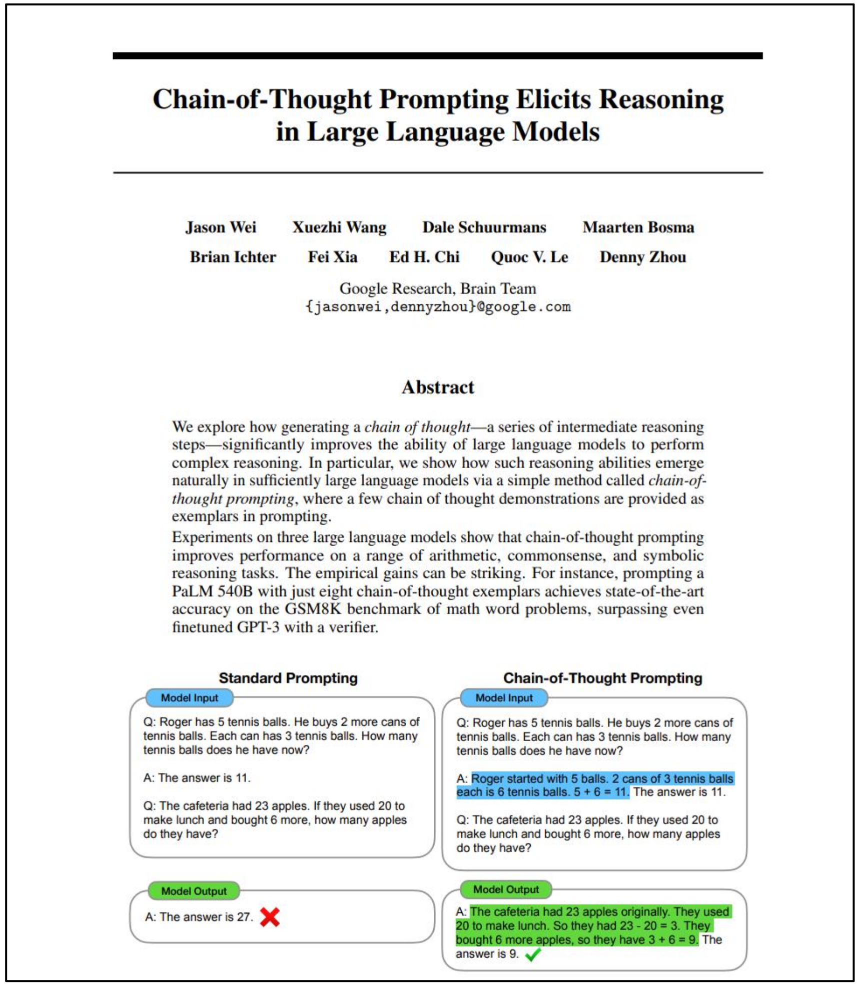


Take-aways

- What can we do with Foundation Models (and use them in your research)?



Paper 2



Chain-of-Thought Prompting Elicits Reasoning in Large Language Models

Jason Wei Xuezhi Wang Dale Schuurmans Maarten Bosma
Brian Ichter Fei Xia Ed H. Chi Quoc V. Le Denny Zhou

Google Research, Brain Team
{jasonwei,dennyzhou}@google.com

Abstract

We explore how generating a *chain of thought*—a series of intermediate reasoning steps—significantly improves the ability of large language models to perform complex reasoning. In particular, we show how such reasoning abilities emerge naturally in sufficiently large language models via a simple method called *chain-of-thought prompting*, where a few chain of thought demonstrations are provided as exemplars in prompting.

Experiments on three large language models show that chain-of-thought prompting improves performance on a range of arithmetic, commonsense, and symbolic reasoning tasks. The empirical gains can be striking. For instance, prompting a PaLM 540B with just eight chain-of-thought exemplars achieves state-of-the-art accuracy on the GSM8K benchmark of math word problems, surpassing even finetuned GPT-3 with a verifier.

Standard Prompting

Model 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

Model 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.

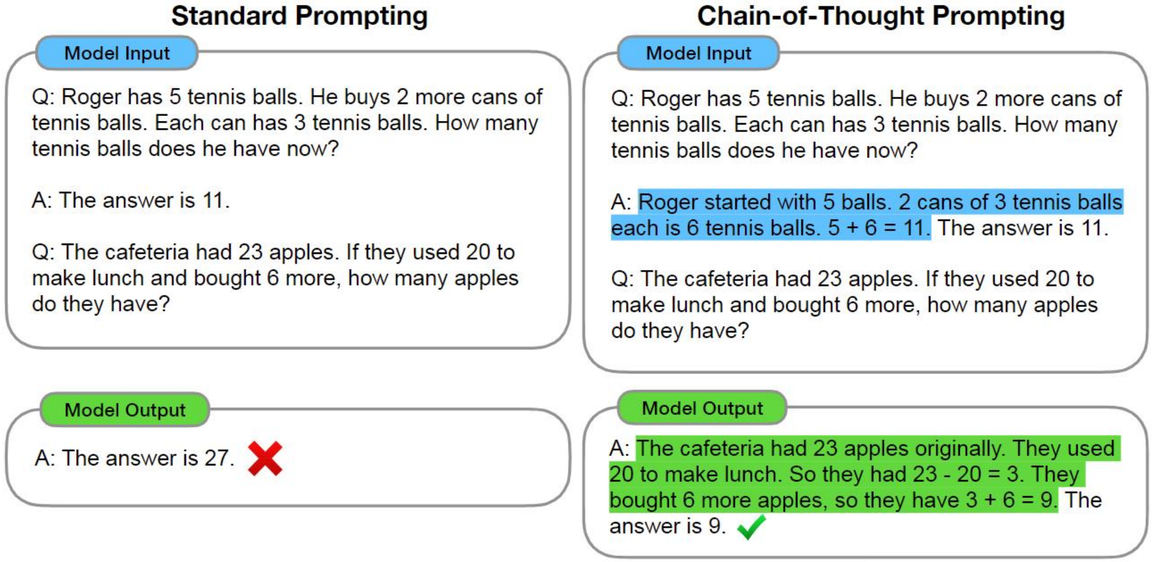
Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

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.

☜ Proposed Train of Thoughts Prompting

Reasoning **Step-by-Step**



This do help to improve LLM's accuracy

Extra Juice:
How to prompt Foundation Models?

Standard Prompt

"Tell me the result of 1+2+3+4+5."

Optimized Prompt

"**As an math expert**, tell me the..."

Role Play

"...**you'll be rewarded with $10 if correct**"

Emotion Simulation

"**Break it down and solve it step-by-step**"

Train of Thoughts

"**As an math expert,** Tell me the result of 1+2+3+4+5. **Break it down and solve it step-by-step, you'll be rewarded with $10 if correct**"

Refernece

- [1] Wei, Jason, et al. "Chain-of-thought prompting elicits reasoning in large language models." Advances in neural information processing systems 35 (2022): 24824-24837.
- [2] Sun, Jiankai, et al. "A survey of reasoning with foundation models." arXiv preprint arXiv:2312.11562 (2023).
- [3] Yu, Tao, et al. "Inpaint anything: Segment anything meets image inpainting." arXiv preprint arXiv:2304.06790 (2023).
- [4] Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models. 2022 IEEE." CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021.
- [5] Kirillov, Alexander, et al. "Segment anything." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.
- [6] Xie, Defeng, et al. "Edit everything: A text-guided generative system for images editing." arXiv preprint arXiv:2304.14006 (2023).
- [7] Radford, Alec, et al. "Learning transferable visual models from natural language supervision." International conference on machine learning. PMLR, 2021.
- [8] Cheng, Yangming, et al. "Segment and track anything." arXiv preprint arXiv:2305.06558 (2023).
- [9] Yang, Zongxin, and Yi Yang. "Decoupling features in hierarchical propagation for video object segmentation." Advances in Neural Information Processing Systems 35 (2022): 36324-36336.
- [10] Hong, Yining, et al. "3d-llm: Injecting the 3d world into large language models." Advances in Neural Information Processing Systems 36 (2023): 20482-20494.
- [11] Zhu, Ziyu, et al. "3d-vista: Pre-trained transformer for 3d vision and text alignment." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.
- [12] Ma, Jun, et al. "Segment anything in medical images." Nature Communications 15.1 (2024): 654.