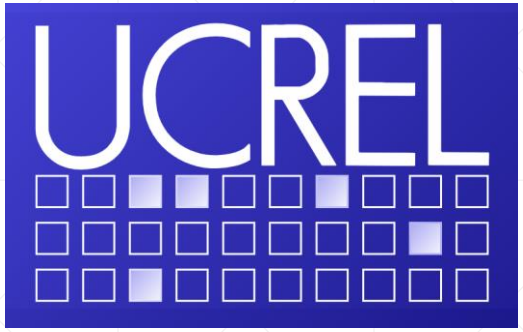

UCREL NLP Summer School 2024

Session 5: Large Language Models (LLMs)

Dr Hansi Hettiarachchi & Isuri Anuradha

July 25, 2024

Language Models (LMs)

- LMs aim to model the generative likelihood of word sequences, so as to predict the probabilities of future (or missing) tokens [1].

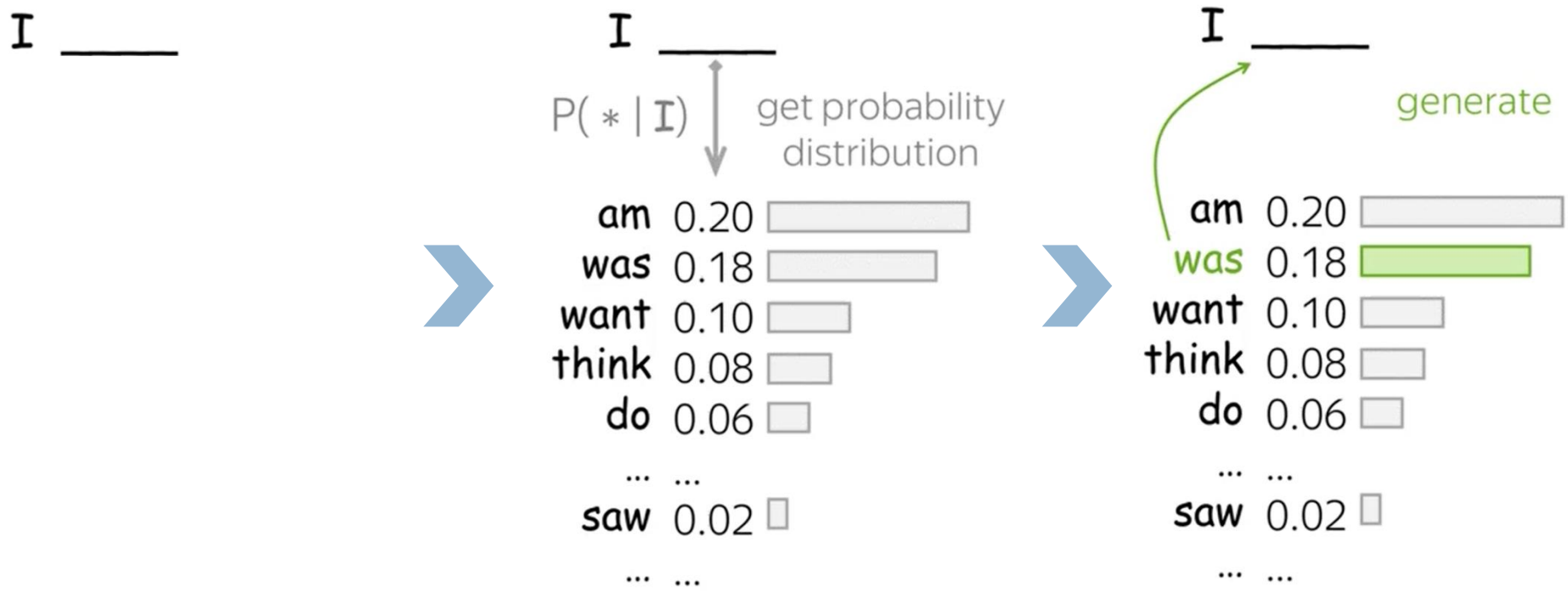

Image source

Language Models (LMs)

- LM aims to model the generative likelihood of word sequences, so as to predict the probabilities of future (or missing) tokens [1].

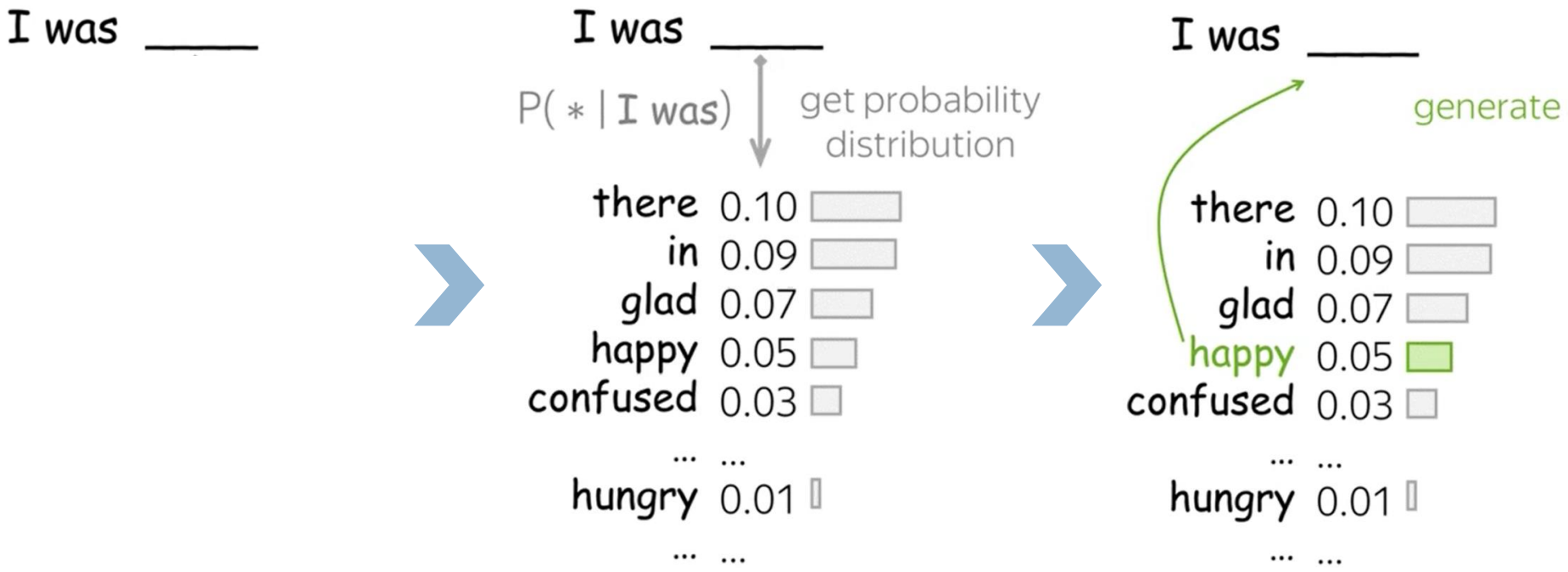

Image source

Language Models (LMs)

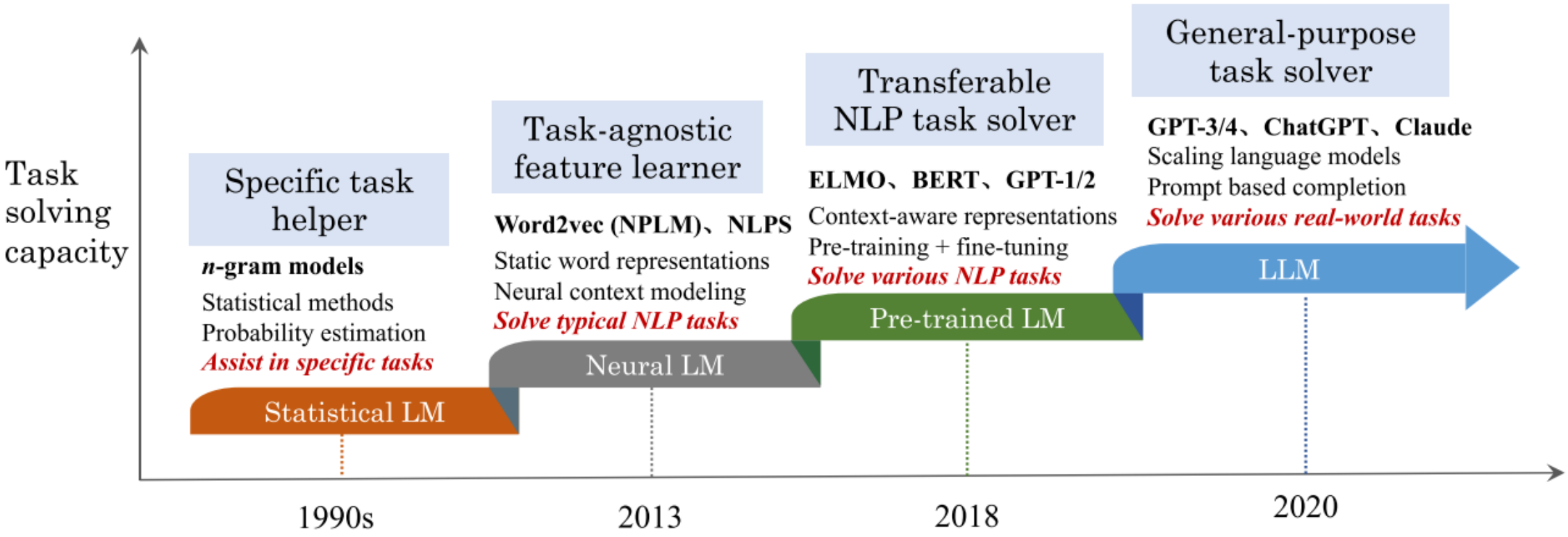

Image source: [1]

Large Language Models (LLMs)

- LLMs are transformer-based neural language models that contain tens to hundreds of billions of parameters, which are pre-trained on massive text data [2]

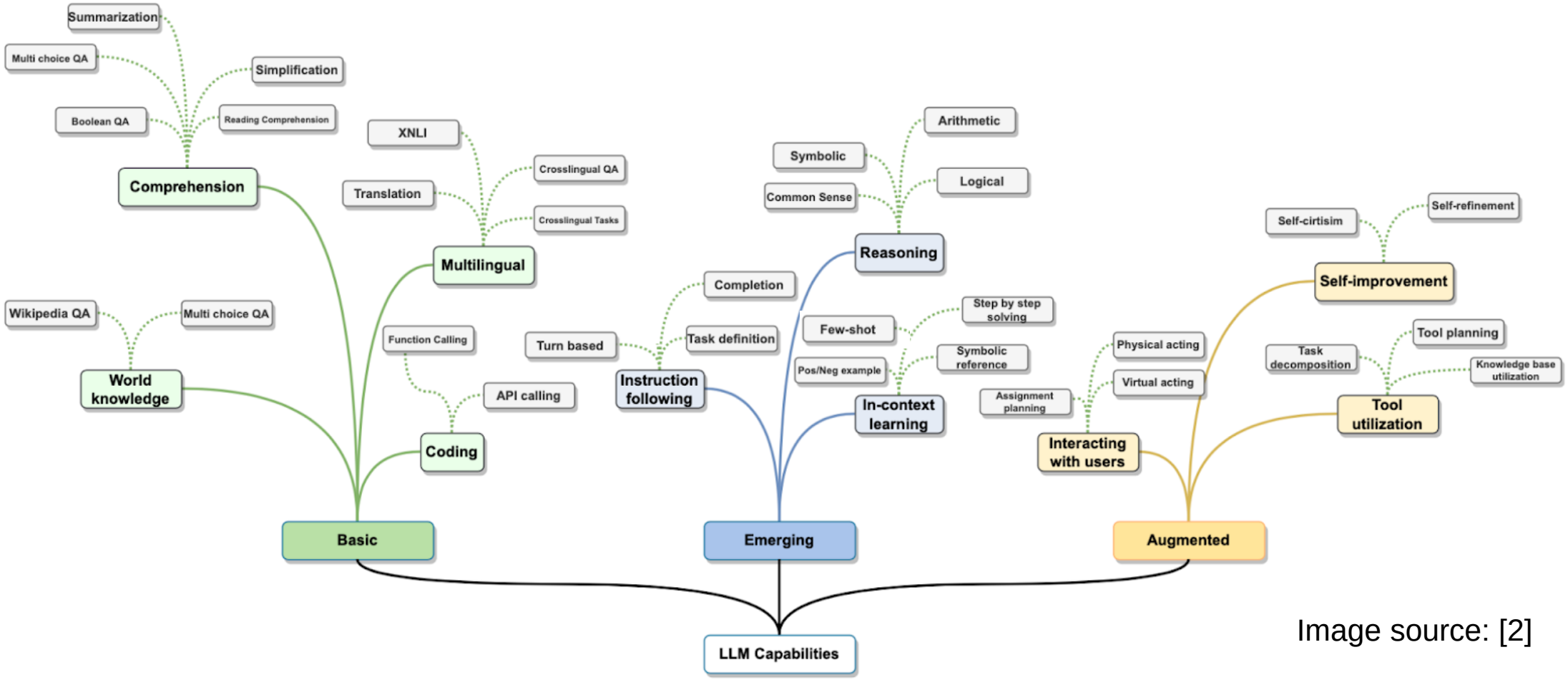

Image source: [2]

LLM Types

Base Models

- Learned from a vast amount of text data
- Continue what has been written so far

Instruction-tuned Models

- Learned from instruction–response pairs
- Carry out the instruction / answer the question

What is the capital of France?

What is the capital of Spain?

Paris

Further reading: [3]

Prompt Engineering – Template

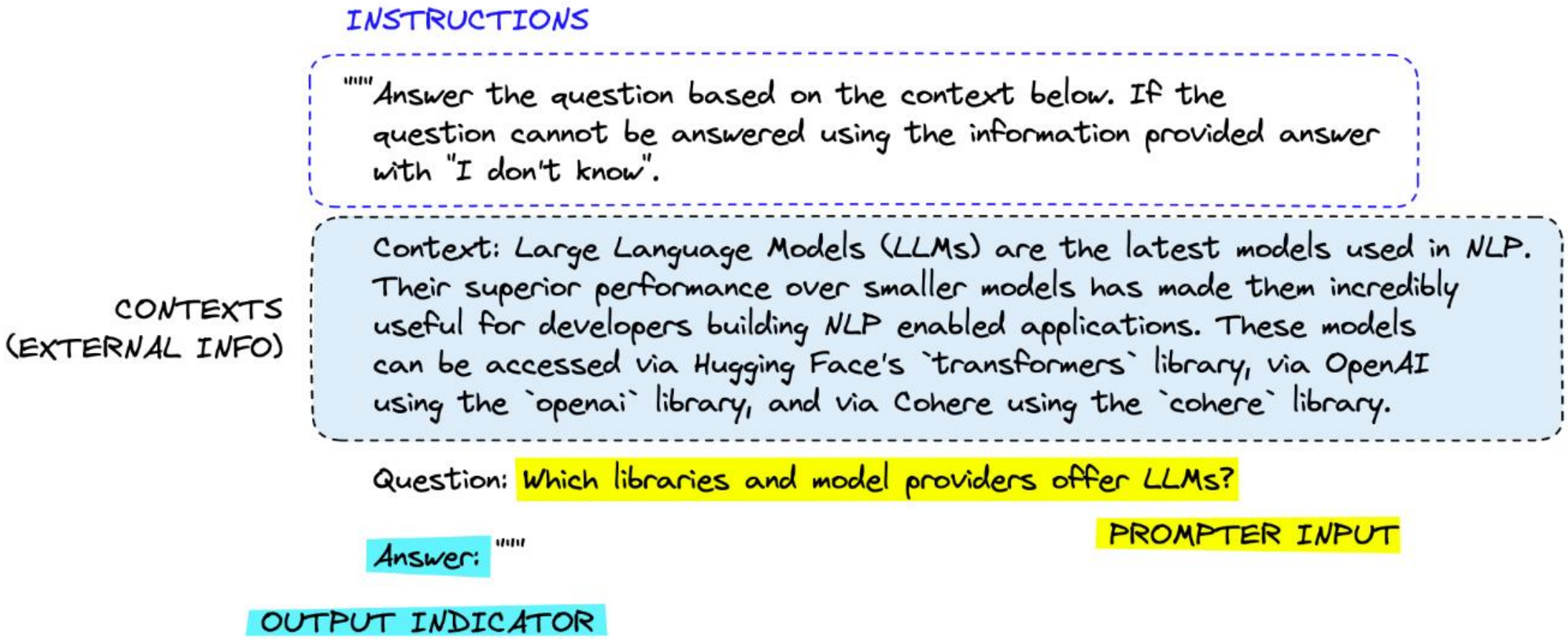

Image source

Prompt Engineering - Techniques

Zero-shot [4]

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A: The answer (arabic numerals) is

(Output) 8 ✗

Zero-shot CoT [7]

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A: ***Let's think step by step.***

(Output) There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls. ✓

Few-shot [5]

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: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A:

(Output) The answer is 8. ✗

Chain-of-Thought (CoT) [6]

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: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A:

(Output) The juggler can juggle 16 balls. Half of the balls are golf balls. So there are 16 / 2 = 8 golf balls. Half of the golf balls are blue. So there are 8 / 2 = 4 blue golf balls. The answer is 4. ✓

Further reading: Prompting Techniques

Prompt Engineering - Best Practices

- Use the latest model
- Write clear instructions - be specific, descriptive and as detailed as possible

  Who's president? Who was the president of Mexico in 2021?
- Specify the desired output format

  Summarise the following text. Summarise the following text in 3 bullet points.
- Clearly indicate distinct parts of the input
- Start with zero-shot, then few-shot, neither of them worked, then fine-tune

Adapted from: source1, source2

References

[1] Zhao, W.X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z. and Du, Y., 2023. “A survey of large language models”. *arXiv preprint arXiv:2303.18223*.

[2] Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X. and Gao, J., 2024. “Large language models: A survey”. *arXiv preprint arXiv:2402.06196*.

[3] Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A. and Schulman, J., 2022. “Training language models to follow instructions with human feedback”. *Advances in neural information processing systems*, 35, pp.27730-27744.

[4] Wei, J., Bosma, M., Zhao, V.Y., Guu, K., Yu, A.W., Lester, B., Du, N., Dai, A.M. and Le, Q.V., 2021. “Finetuned language models are zero-shot learners”. arXiv preprint arXiv:2109.01652.

[5] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A. and Agarwal, S., 2020. “Language models are few-shot learners”. *Advances in neural information processing systems*, 33, pp.1877-1901.

[6] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q.V. and Zhou, D., 2022. “Chain-of-thought prompting elicits reasoning in large language models”. *Advances in neural information processing systems*, 35, pp.24824-24837.

[7] Kojima, T., Gu, S.S., Reid, M., Matsuo, Y. and Iwasawa, Y., 2022. Large language models are zero-shot reasoners. *Advances in neural information processing systems*, 35, pp.22199-22213.

Time for Hands-On Practice!