

# Introduction to Utilizing LLMs in Resresearch

---

Hsiu-Hsuan(Jacky) Yeh

2024-01-26

# Outline

LLM

In Context Learning

Few-Shot Learning

Chain-of-Thought

Step-Back Prompting

ReAct

Medprompt+

others

Fine-tuning

LangChain

References

**LLM**

---

# List of LLMs

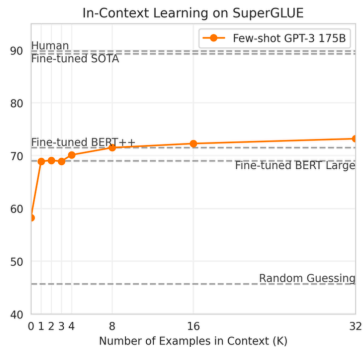
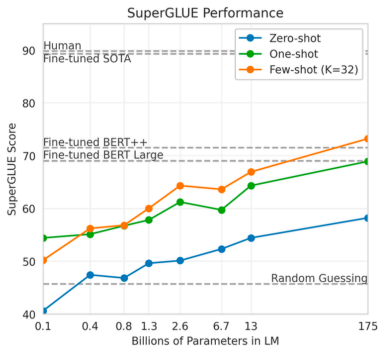
- ChatGPT 3.5/4
- Claude 2
- Gemini Pro 1.5/Ultra
- Llama 2, Touvron et al. (2023)
- Mistral 7B, Jiang et al. (2023)
- Falcon 180B, Almazrouei et al. (2023)

# In Context Learning

---

## Few-Shot Learning, Brown et al. (2020)

LLM can significantly improve task-agnostic few-shot learning performance, achieving strong results on a variety of NLP tasks without task-specific tuning. The study contrasts GPT-3's few-shot capabilities with zero-shot and one-shot settings, showing it can perform competitively with or even exceed state-of-the-art models that have been fine-tuned for specific tasks.



**Table 5:** Evaluation Metrics for Test Set Contained in 7000 News Articles

	(1)	(2)	(3)	(4)
	micro_f1	macro_f1	weighted_f1	precision_0
8 shot with reason	0.707	0.705	0.799	0.707
6 shot with reason	0.672	0.674	0.762	0.677
3 shot with reason (fine-tuned 1000)	0.555	0.598	0.63	0.58
6 shot no reason	0.54	0.585	0.621	0.566
zero shot with reason (fine-tuned 1000)	0.417	0.579	0.59	0.472
zero shot no reason (fine-tuned 1000)	0.372	0.588	0.589	0.438
zero shot with reason	0.401	0.58	0.588	0.459
zero shot no reason	0.415	0.572	0.587	0.469

Notes: ADD NOTES HERE



CoT involves providing the model with a few exemplar prompts that demonstrate the reasoning process step by step, leading to the final answer. It has been shown to improve performance across a variety of tasks, including arithmetic, commonsense, and symbolic reasoning, by making the models' thought process more interpretable and allowing them to tackle multi-step problems more effectively.

### 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. ✓

總統馬英九競選期間提出的募兵支票恐怕面臨跳票？國防部長陳肇敏昨在立法院外交國防委員會中表示，推動募兵制的確有困難，且要推動到全面募兵制，「是國防部嚴峻的考驗」。陳肇敏指出，國防預算佔GDP（Gross Domestic Product，國內生產毛額）百分之三，但以現有國軍員額來看，根本不可能達成全面募兵制，因此國軍要再精進是必然方向。而我國走的是一個海島守勢作戰，各軍種人員配比都要進行檢討，並將繼續推動「再精進案」，將目前二十七萬大軍減至二十到二十五萬人以內。陳肇敏並指出，目前募得許多素質相當不錯的國軍弟兄，主要與過去八年經濟不景氣，工作難找有關，未來經濟好轉後，是否還有誘因？另外，台灣人口有限，國防部會努力推動募兵制，但不能因滿足「量」的需求而減低「質」的要求。至於重大軍購案，陳肇敏強調向美國購買F-16C/D戰機有其必要，但其他軍購案如潛艦，因目前我方連採購潛艇的型號都沒有，這部分會再研究，會就作戰需求再作評估。記者王 <U+70F1> 華

Let's think step by step. First, the news is related to Taiwan's economics as it points out the relation between the quality of voluntary military and the Taiwan's economics condition. Second, the news is related to the policy regarding the voluntary military service. Finally, the news is related to the uncertainty as it discusses some possibilities and possible progress of the policy. This news simultaneously mentions economics, policy and uncertainty. Hence, it should not be excluded when constructing Taiwan's EPU index.

This prompting technique improves reasoning in large language models (LLMs) by leveraging abstraction. This method prompts LLMs to first derive high-level concepts or principles from detailed questions before addressing the original query.

## Case Study

User:

Is this article related to economic policy uncertainty? Answer in 30 words.

{ news }

LLM(Mistral 7B):

This article discusses Taiwan's defense budget and the challenges of implementing a full-scale conscription policy, with Defense Minister Chen Chwen-min stating that it is a severe test for the National Defense Ministry.

...

Regarding major procurement projects, Minister Chen emphasized the necessity of purchasing F-16 C/D fighter jets from the US, but other procurement cases like submarines will be studied further based on operational needs.

User:

What is an economic policy uncertainty(EPU) index.

LLM(Mistral 7B):

An Economic Policy Uncertainty (EPU) index measures the level of uncertainty in the economy due to potential government actions that may impact the economy. It is calculated based on newspaper articles and other sources, using a specific methodology developed by Scott R. Baker, Nicholas Bloom, and Steven J. Davis. The index is constructed as a composite measure of uncertainty related to fiscal policy, monetary policy, and regulatory policy. Higher values of the EPU index indicate greater economic policy uncertainty.

User:

Is this article related to EPU? Answer in 30 words.

{ news }

LLM(Mistral 7B):

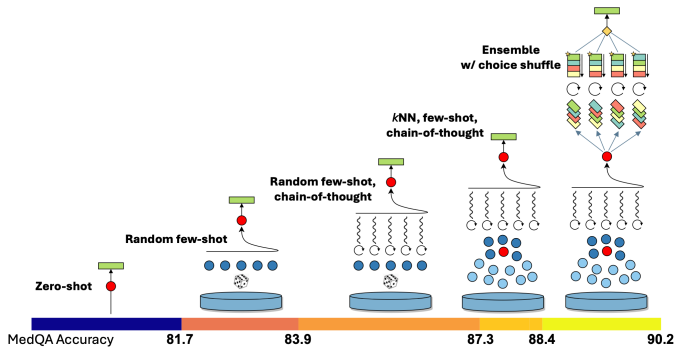
The Economic Policy Uncertainty (EPU) index is a measure of the level of policy uncertainty in an economy, calculated using data from newspapers and other sources to identify the number of articles mentioning economic policy uncertainty. This article does not directly discuss the EPU index, but rather Taiwan's defense budget and military conscription issues during the presidential election period.



ReAct prompts LLMs to interleave reasoning (generating verbal reasoning traces) with action (making task-specific actions), thus enabling dynamic adjustment of plans based on both internal reasoning and external information.

Question	What profession does Nicholas Ray and Elia Kazan have in common?
Thought 1	I need to search Nicholas Ray and Elia Kazan, find their professions, then find the profession they have in common.
Action 1	Search[Nicholas Ray]
Observation 1	Nicholas Ray (born Raymond Nicholas Kienzle Jr., August 7, 1911 - June 16, 1979) was an American film director, screenwriter, and actor best known for the 1955 film Rebel Without a Cause.
Thought 2	Professions of Nicholas Ray are director, screenwriter, and actor. I need to search Elia Kazan next and find his professions.
Action 2	Search[Elia Kazan]
Observation 2	Elia Kazan was an American film and theatre director, producer, screenwriter and actor.
Thought 3	Professions of Elia Kazan are director, producer, screenwriter, and actor. So profession Nicholas Ray and Elia Kazan have in common is director, screenwriter, and actor.
Action 3	Finish[director, screenwriter, actor]

This prompting technique combines dynamic few-shot example selection, self-generated chain-of-thought reasoning, and choice shuffle ensembling. MedPrompt showcases the potential of generalist models to achieve specialist-level performance through carefully engineered prompts, demonstrating a versatile approach applicable across various domains beyond medicine.



- Principles Of Prompting, Bsharat et al. (2024)
- Self-Consistency, Wang et al. (2022)
- Tree Of Thoughts, Yao et al. (2023a)
- Self-Discover, Zhou et al. (2024)

# Fine-tuning

---

In deep learning, fine-tuning is an approach to transfer learning in which the weights of a pre-trained model are trained on new data.

Fine-tuning improves on few-shot learning by training on many more examples than can fit in the prompt, letting you achieve better results on a wide number of tasks. Once a model has been fine-tuned, you won't need to provide as many examples in the prompt.



Documentation

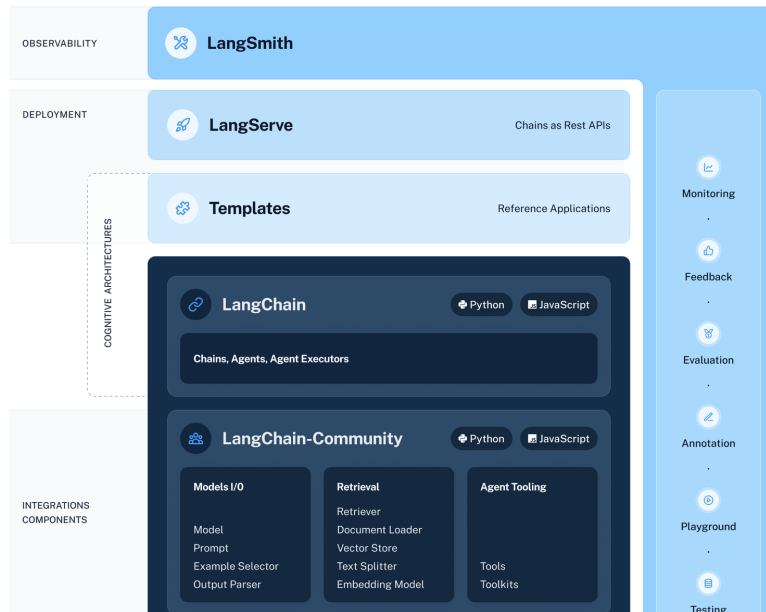
API reference

# LangChain

---

The main component of LangChain-Core is LCEL, LangChain Expression Languages, which involves several sub-components: prompt template, model, output parser, etc.

# Ecosystem



# Prompt Template

prompt template example

```

163  ✓ def request_batch(self,
164         prompt: Prompt,
165         query_file_path: str,
166         fewshot_examples_path: Optional[str] = None,
167         sleep: int = 3,
168         ) -> None:
169     self.prompt = prompt
170     _i, n_query, data = self.__hook_process(query_file_path)
171     query_list = [i.get(prompt.query_key) for i in data]
172
173     if fewshot_examples_path is not None:
174         message_prompt = prompt.few_shot(fewshot_examples_path)
175         mlflow.log_artifact(fewshot_examples_path)
176     else:
177         message_prompt = prompt.zero_shot()
178
179     logger.info('start the query process')
180     with self.llm_response_file_path.open('ab') as f:
181         for i in range(n_query, position=0, leave=True):
182             chain = (
183                 message_prompt.partial(**{prompt.query_key: query_list[_i+i]})
184                 | self.llm
185                 | prompt.parser
186             )
187             res = self.request_instance("", chain, _i+i)
188             f.write(orjson.dumps(res, option=orjson.OPT_APPEND_NEWLINE))
189             time.sleep(sleep)
190
191     logger.info('finish the query process')
192     self.mlflow_logging(data)

```

## More Examples

- few-shot examples
- query v.s. response: an example of translating English Address to Chinese
- my packages, llm-research

## References

---



**Almazrouei, Ebtesam, Hamza Alobeidli, Abdulaziz Alshamsi, Alessandro Cappelli, Ruxandra Cojocaru, Mérouane Debbah, Étienne Goffinet, Daniel Hesslow, Julien Launay, Quentin Malartic, Daniele Mazzotta, Badreddine Noune, Baptiste Pannier, and Guilherme Penedo, “The Falcon Series of Open Language Models,” November 2023.**

**Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford,**

**Ilya Sutskever, and Dario Amodei**, “Language Models Are Few-Shot Learners,” *Advances in Neural Information Processing Systems*, 2020, 33, 1877–1901.

**Bsharat, Sondos Mahmoud, Aidar Myrzakhan, and Zhiqiang Shen**, “Principled Instructions Are All You Need for Questioning LLaMA-1/2, GPT-3.5/4,” January 2024.

**Jiang, Albert Q., Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, L  lio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timoth  e Lacroix, and William El Sayed**, “Mistral 7B,” October 2023.

**Nori, Harsha, Yin Tat Lee, Sheng Zhang, Dean Carignan, Richard Edgar, Nicolo Fusi, Nicholas King, Jonathan Larson, Yuanzhi Li, Weishung Liu, Renqian Luo, Scott Mayer McKinney, Robert Osazuwa Ness, Hoifung Poon, Tao Qin, Naoto Usuyama, Chris White, and Eric Horvitz, “Can Generalist Foundation Models Outcompete Special-Purpose Tuning? Case Study in Medicine,” November 2023.**

**Touvron, Hugo, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar**

Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom, “Llama 2: Open Foundation and Fine-Tuned Chat Models,” July 2023.

**Wang, Xuezhi, Jason Wei, Dale Schuurmans, Quoc V. Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou**, “Self-Consistency Improves Chain of Thought Reasoning in Language Models,” in “The Eleventh International Conference on Learning Representations” September 2022.

**Wei, Jason, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc V. Le, and Denny Zhou**, “Chain-of-Thought Prompting Elicits Reasoning in Large Language Models,” *Advances in Neural Information Processing Systems*, December 2022, 35, 24824–24837.

**Yao, Shunyu, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan**, “Tree of Thoughts: Deliberate Problem Solving with Large Language Models,” *Advances in Neural Information Processing Systems*, December 2023, 36.

— , Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao, “ReAct: Synergizing Reasoning and Acting in Language Models,” March 2023.

**Zheng, Huaixiu Steven, Swaroop Mishra, Xinyun Chen, Heng-Tze Cheng, Ed H. Chi, Quoc V. Le, and Denny Zhou,** “Step-Back Prompting Enables Reasoning Via Abstraction in Large Language Models,” in “The Twelfth International Conference on Learning Representations” October 2023.

**Zhou, Pei, Jay Pujara, Xiang Ren, Xinyun Chen, Heng-Tze Cheng, Quoc V. Le, Ed H. Chi, Denny Zhou, Swaroop Mishra, and Huaixiu Steven Zheng,** “Self-Discover: Large Language Models Self-Compose Reasoning Structures,” February 2024.