Applied Deep Learning

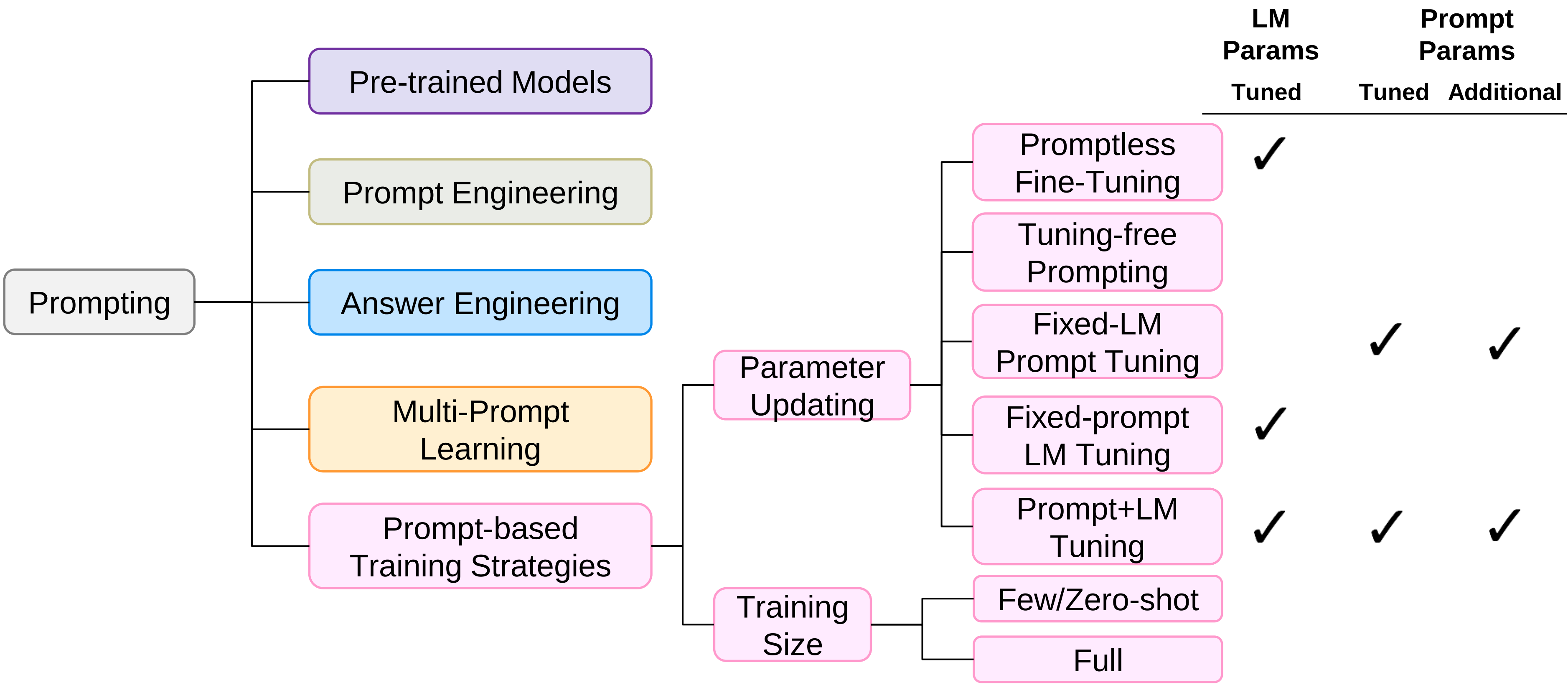
LLM Adaptation

November 16th, 2023 http://adl.miulab.tw

National Taiwan University 國立臺灣大學

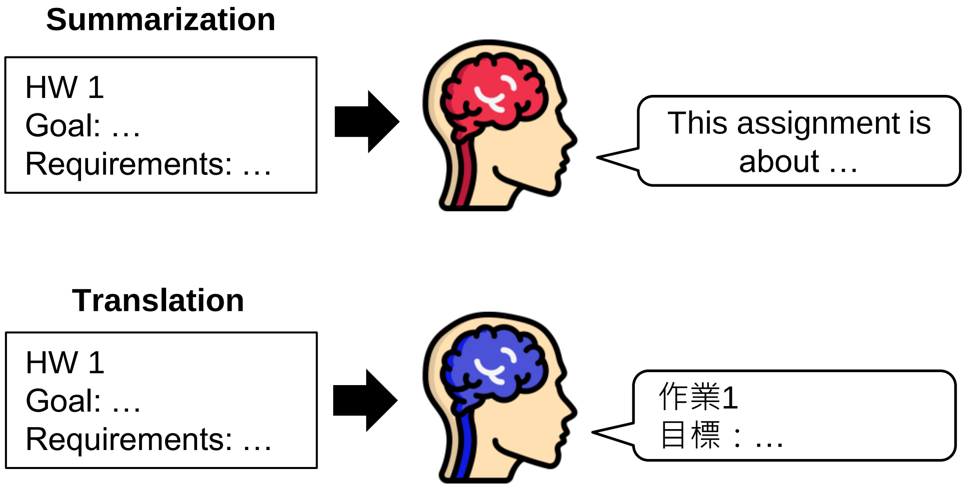
2

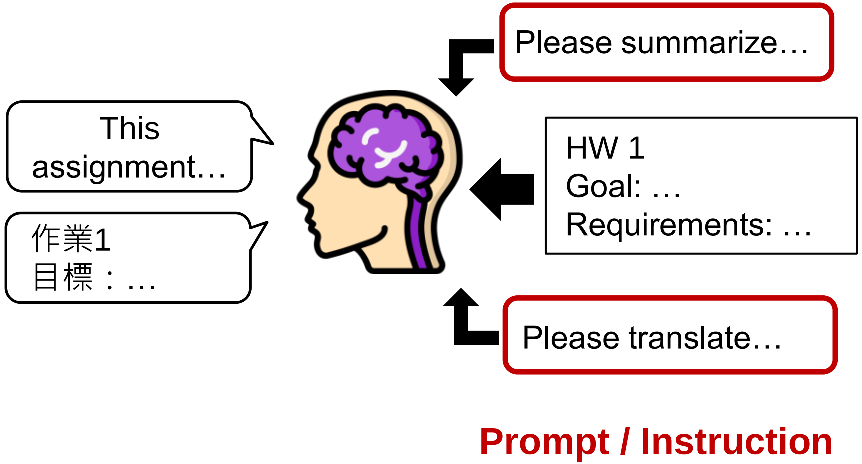
Prompting Typology (Liu et al., 2021)



3 Specialists (專才) vs. Generalists (通才)

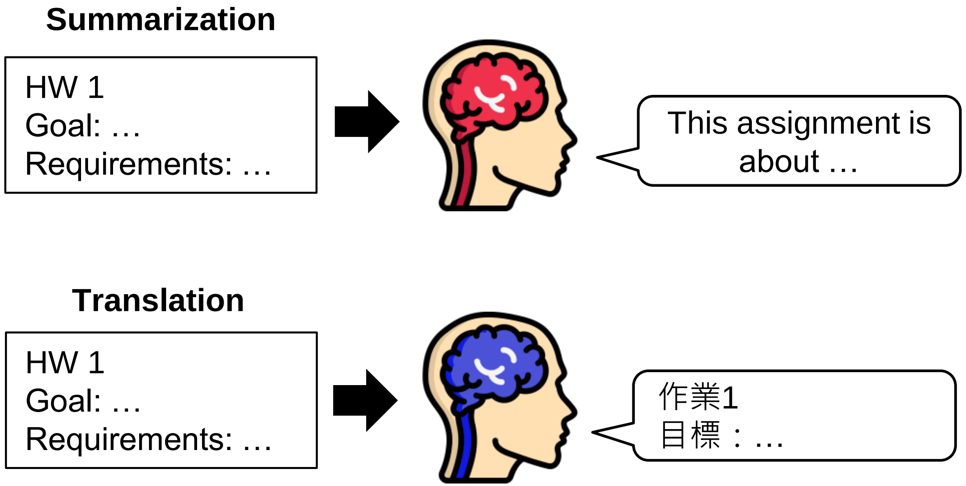
- **Specialists**
  - *master* a *single* focused task
- **Generalists**
  - *good* at *many* tasks

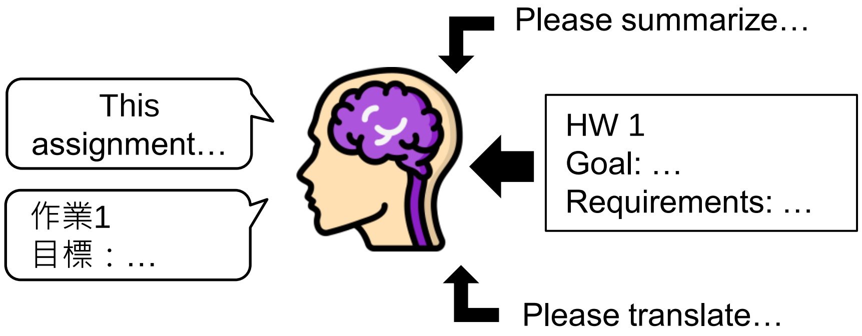




Specialists (專才) vs. Generalists (通才)

- Specialists
  - ***master*** a *single* focused task
- Generalists
  - *good* at *many* tasks





Task Master

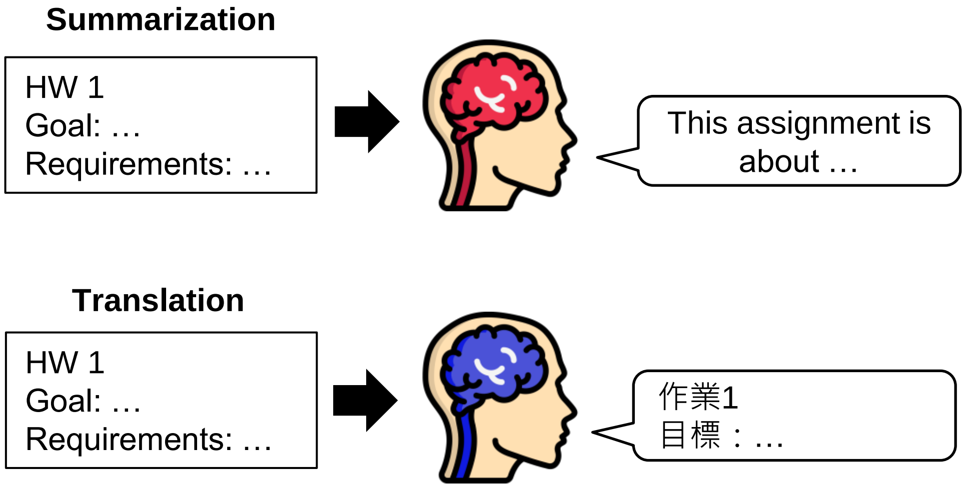
- Machine translation comparison between WMT and GPT

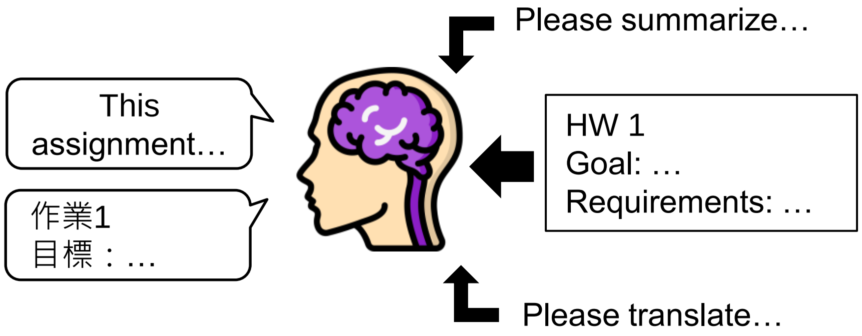
System	COMET-22	COMETkiwi	ChrF	BLEU	COMET-22	COMETkiwi	ChrF	BLEU
	DE-EN				EN-DE			
WMT-Best	**85.0**	**81.4**	**58.5**	**33.4**	**87.2**	**83.6**	**64.6**	**38.4**
text-davinci-002	73.2	73.1	46.1	23.3	82.0	79.0	56.0	28.6
text-davinci-003	84.8*	81.2*	56.8	30.9	85.6*	82.8*	60.2*	31.8*
ChatGPT	84.8*	81.1	58.3*	33.4*	84.2	81.0	59.6	30.9
	ZH-EN				EN-ZH			
WMT-Best	81.0	77.7	**61.1**	**33.5**	**86.7**	**82.0**	**41.1**	**44.8**
text-davinci-002	74.1	73.1	49.6	20.6	84.0	79.0	32.1	36.4
text-davinci-003	**81.6***	**78.9***	56.0*	25.0	85.8*	81.3*	34.6	38.3
ChatGPT	81.2	78.3	56.0	25.9*	84.4	78.7	36.0*	40.3*
	RU-EN				EN-RU			
WMT-Best	**86.0**	**81.7**	**68.9**	**45.1**	**89.5**	**84.4**	**58.3**	**32.4**
text-davinci-002	77.5	76	58.7	34.9	85.4	80.9	51.6	25.1
text-davinci-003	84.8*	81.1*	64.6	38.5	86.7*	82.2*	54.0*	27.5*
ChatGPT	84.8*	81.0	66.5*	41.0*	77.6	70.4	41.1	19.0
	FR-DE				DE-FR			
WMT-Best	**89.5**	**80.7**	**81.2**	**64.8**	**85.7**	79.5	**74.6**	**58.4**
text-davinci-002	66.6	67.9	45.8	25.9	64.2	67.6	44.6	24.5
text-davinci-003	84.6	77.9	65.7*	42.5*	78.5	76.1	58.9	35.6
ChatGPT	84.7*	78.5*	65.2	42.0	81.6*	**79.8***	60.7*	37.3*

Jiao et al., "Is ChatGPT A Good Translator? Yes With GPT-4 As The Engine," *arXiv preprint arXiv:2301.08745*.
Hendy et al., "How Good Are GPT Models at Machine Translation? A Comprehensive Evaluation," *arXiv preprint arXiv:2302.09210*.

6 Specialists (專才) vs. Generalists (通才)

- Specialists
  - *master* a *single* focused task
- Generalists
  - *good* at ***many*** tasks





7

Multitask Learning as QA

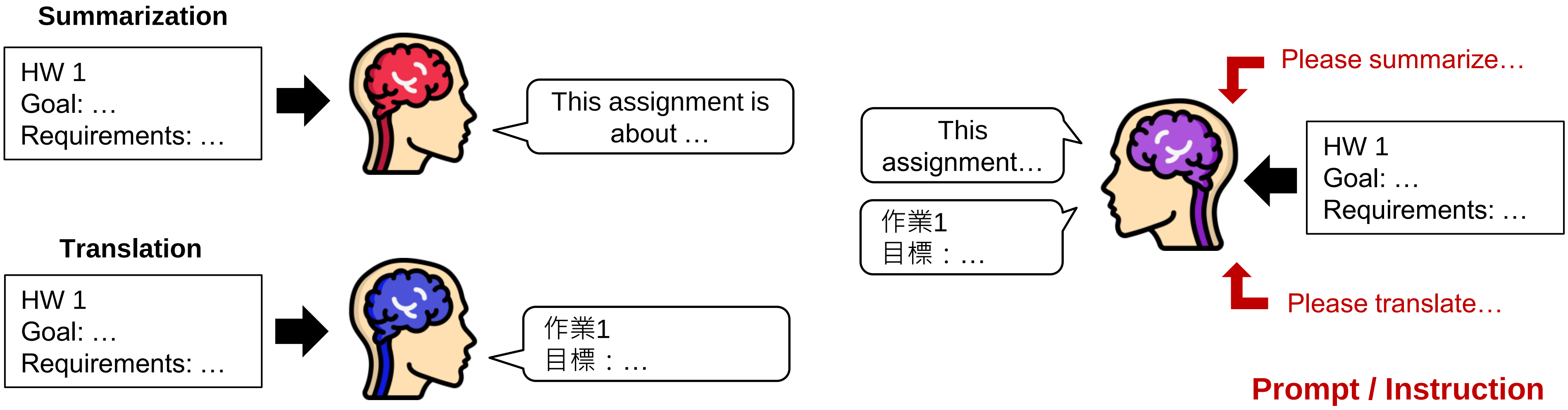
Question	Context	Answer
What is a major importance of Southern California in relation to California and the US?	...Southern California is a major economic center for the state of California and the US....	major economic center
What is the translation from English to German?	Most of the planet is ocean water.	Der Großteil der Erde ist Meerwasser
What is the summary?	Harry Potter star Daniel Radcliffe gains access to a reported £320 million fortune...	Harry Potter star Daniel Radcliffe gets £320M fortune...
Hypothesis: Product and geography are what make cream skimming work. Entailment, neutral, or contradiction?	Premise: Conceptually cream skimming has two basic dimensions – product and geography.	Entailment
Is this sentence positive or negative?	A stirring, funny and finally transporting re-imagining of Beauty and the Beast and 1930s horror film.	positive

Question	Context	Answer
What has something experienced?	Areas of the Baltic that have experienced eutrophication.	eutrophication
Who is the illustrator of Cycle of the Werewolf?	Cycle of the Werewolf is a short novel by Stephen King, featuring illustrations by comic book artist Bernie Wrightson.	Bernie Wrightson
What is the change in dialogue state?	Are there any Eritrean restaurants in town?	food: Eritrean
What is the translation from English to SQL?	The table has column names... Tell me what the notes are for South Australia	SELECT notes from table WHERE 'Current Slogan' = 'South Australia'
Who had given help? Susan or Joan?	Joan made sure to thank Susan for all the help she had given.	Susan

McCann et al., "The Natural Language Decathlon: Multitask Learning as Question Answering," *arXiv preprint arXiv:1806.08730*.

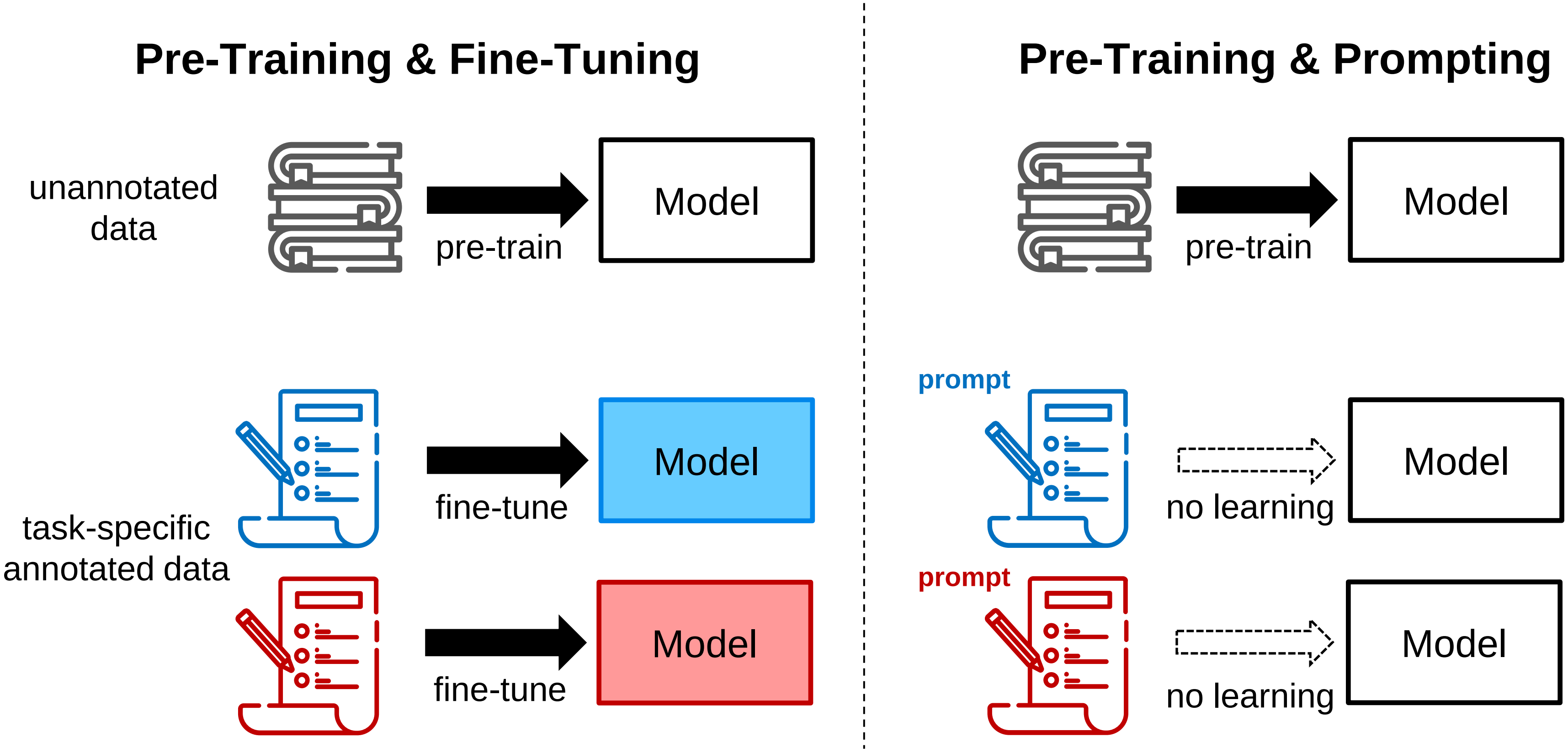
Specialists (專才) vs. Generalists (通才)

- Specialists
  - *master* a *single* focused task
- Generalists
  - *good* at *many* tasks



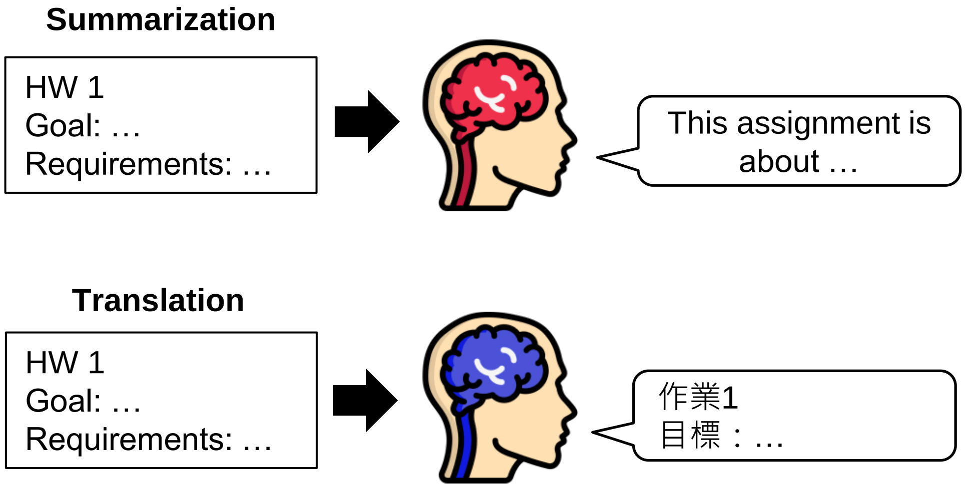
Prompt engineering enables to perform unseen task

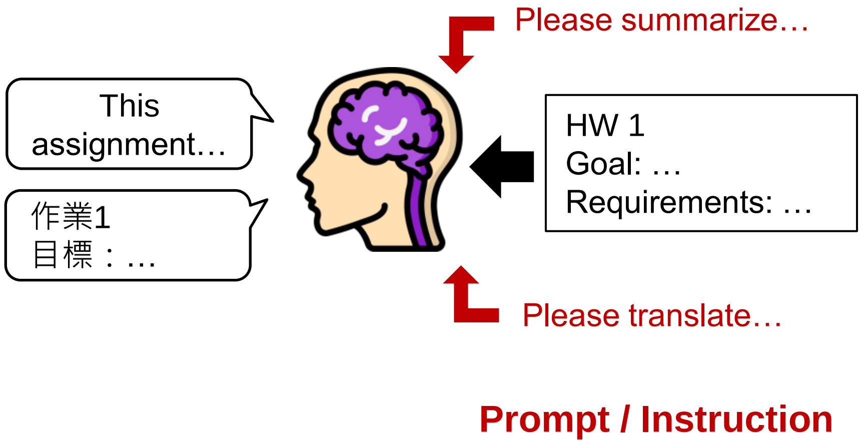
Fine-Tuning vs. Prompting



Specialists (專才) vs. Generalists (通才)

- Specialists
  - *master* a *single* focused task
- Generalists
  - *good* at *many* tasks

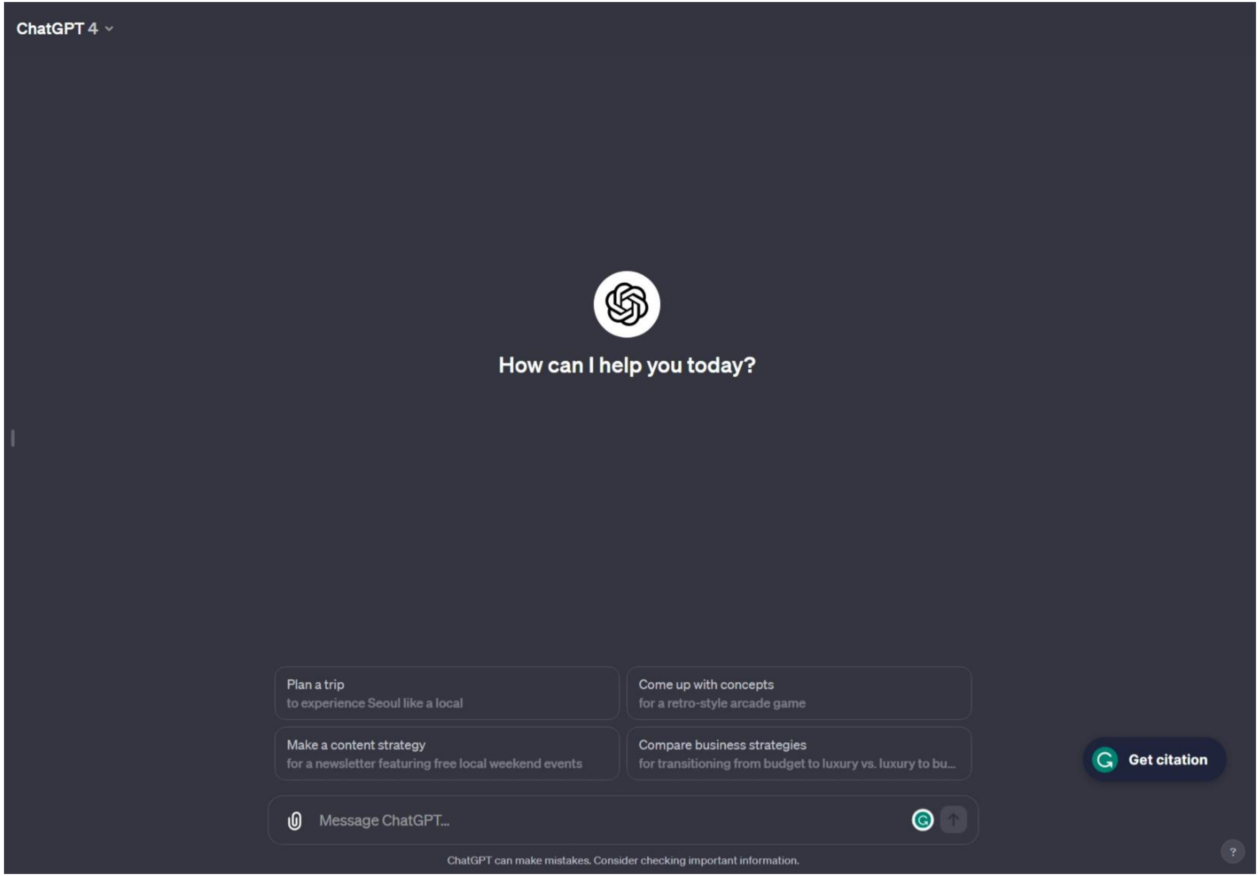




→ Fine-tuning

→ Prompting

GPT Data Fine-Tuning?

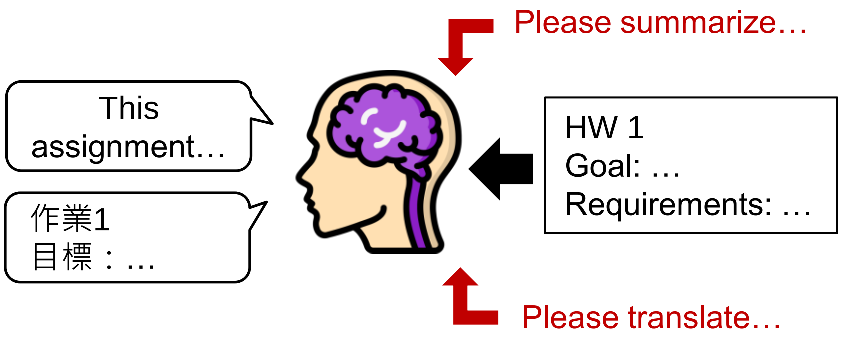


LLM: Large Language Model

- How to train a good generalist that is _good_ at _many_ tasks
  - Large pre-trained data
  - Large model size

  } **emergent ability**

- Further improvement
  - Learning to perform well on ***known*** tasks
    - Prompt tuning / engineering
    - LM tuning



Prompt / Instruction

Fine-tuning LLMs may be expensive and impractical

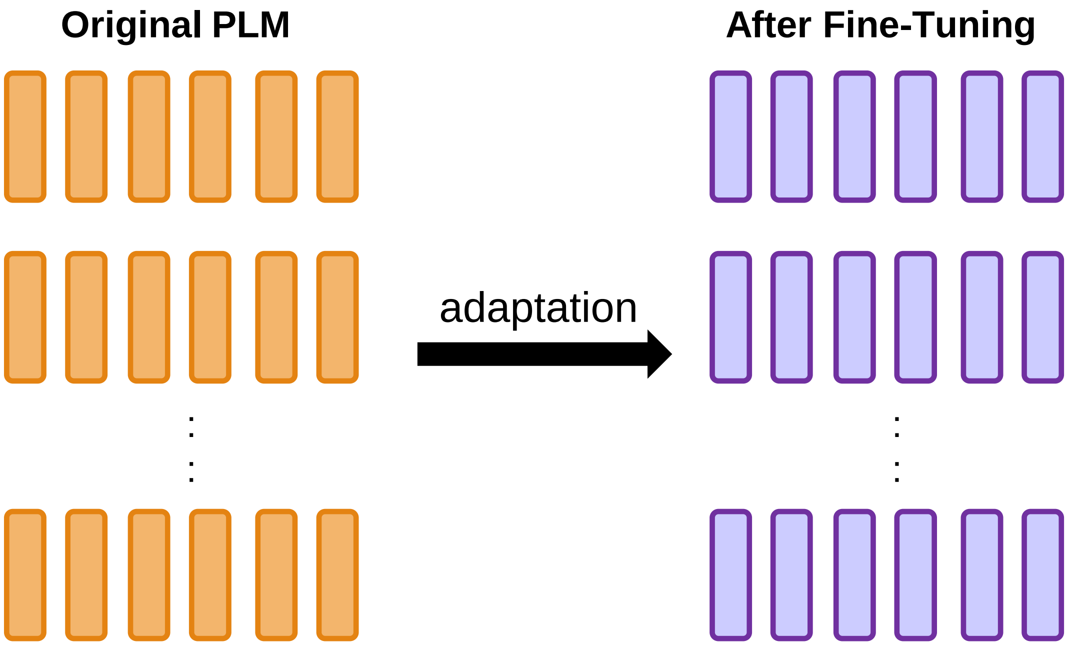
13

Parameter-Efficient LM Tuning

More practical ways to adapt LLMs

Parameter-Efficient LM Tuning for Adaptation

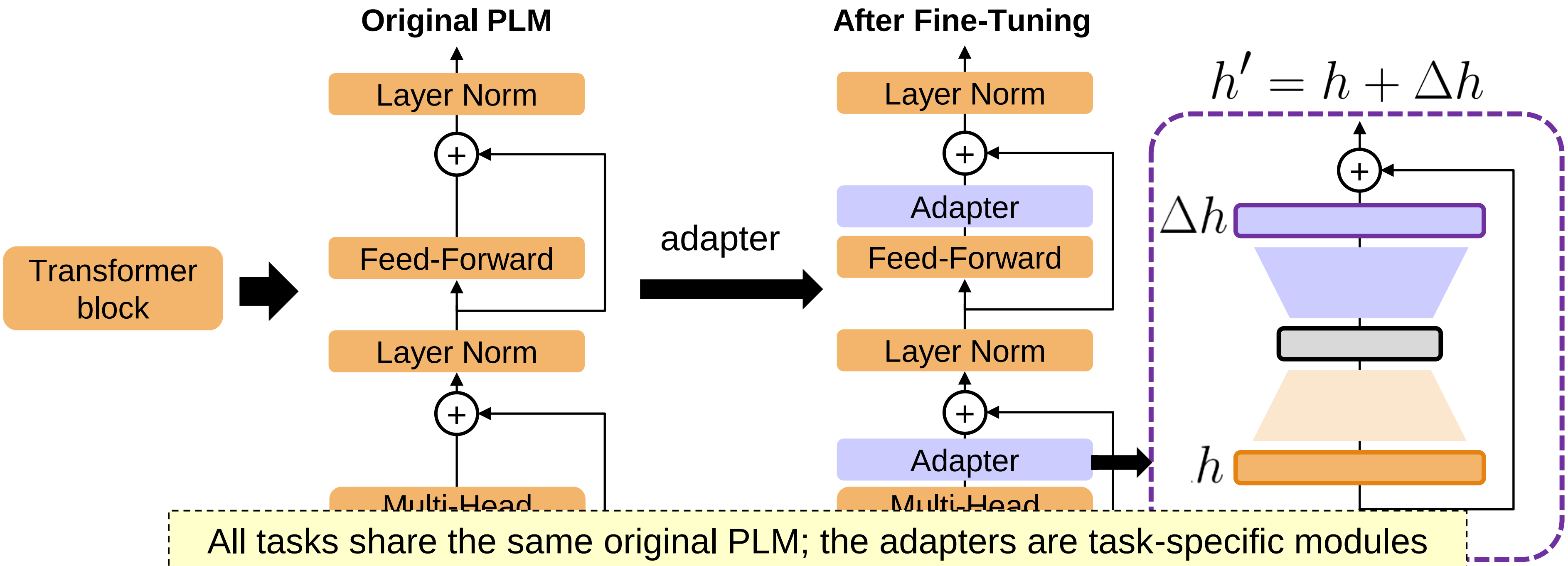
- Idea: slightly modify hidden representations



$$h \qquad h' = h + \Delta h$$

Adapter (He et al., 2022)

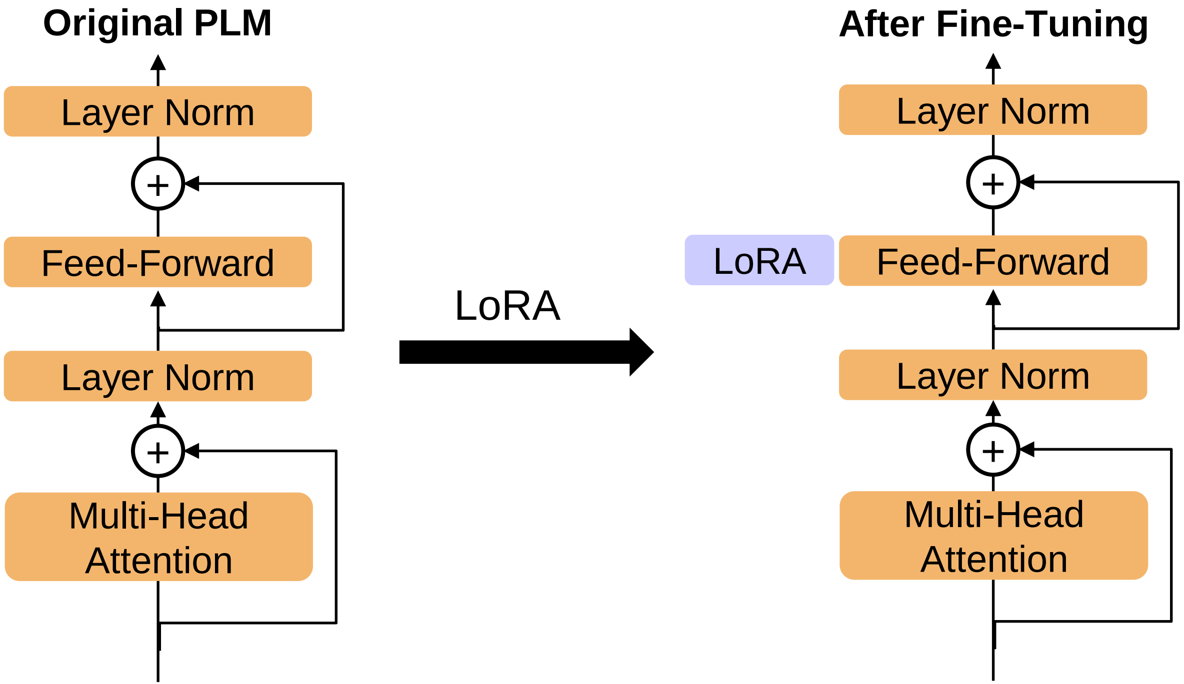
- Idea: *small trainable submodules* inserted in Transformers



All tasks share the same original PLM; the adapters are task-specific modules
→ better robustness, storage-efficient

LoRA (Hu et al., 2021)

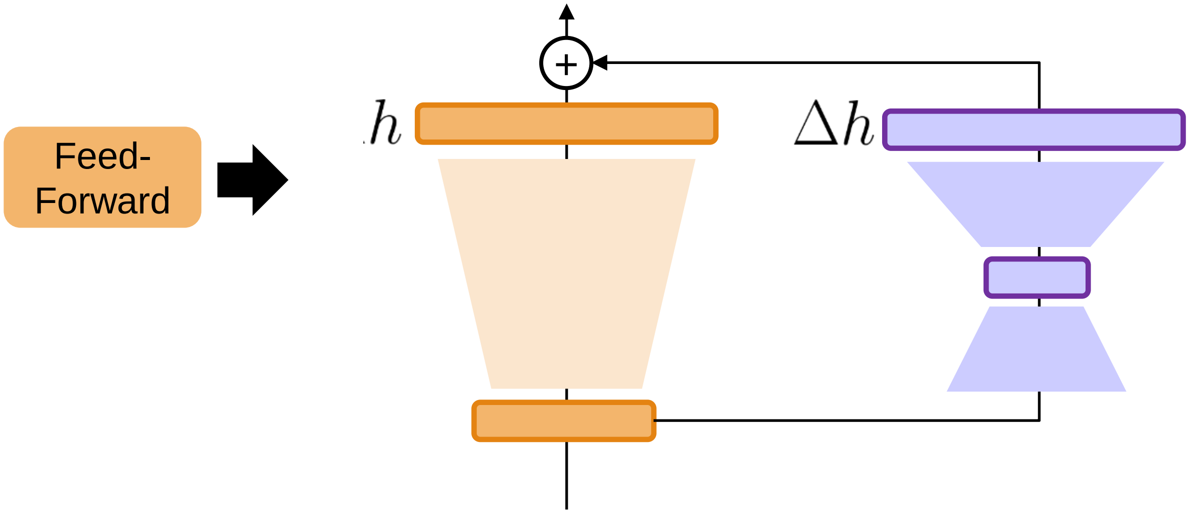
- Idea: low-rank adaptation



LoRA (Hu et al., 2021)

- Idea: low-rank adaptation

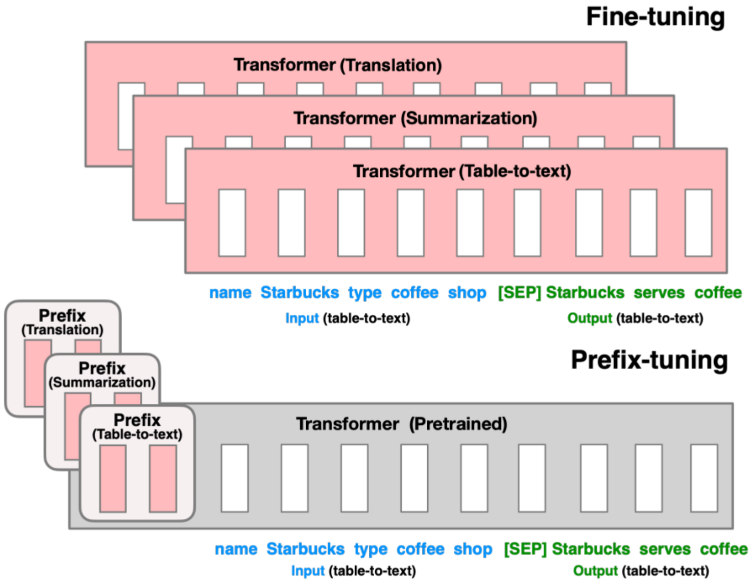
$$h' = h + \Delta h$$

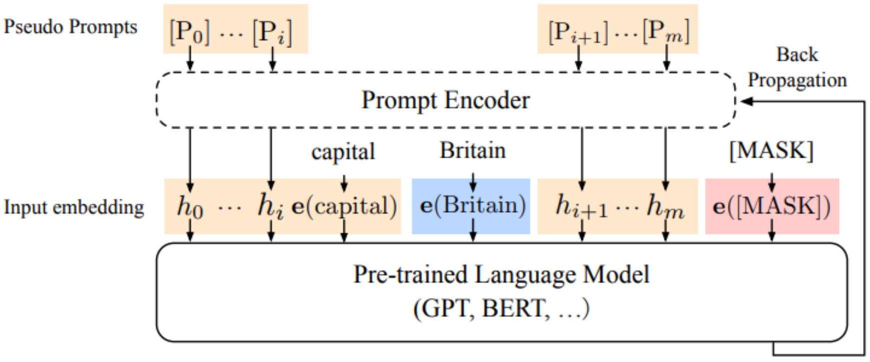


Weight updates for downstream fine-tuning give a low intrinsic rank

Prompt Tuning

- Prefix-tuning & soft prompt-tuning are parameter-efficient adaptation





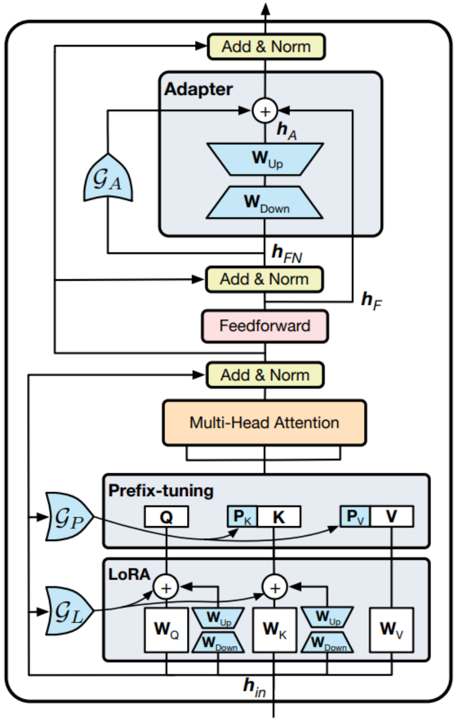
(b) P-tuning

Parameter-Efficient Tuning

- Which one is better? (Mao et al., 2022)

Method	SST-2	MRPC	CoLA	RTE	QNLI	STS-B	MNLI	QQP	Avg.
[$K = all$] Best Performance on GLUE Dev									
Fine-tuning	91.63	90.94	**62.08**	66.43	89.95	**89.76**	83.23	**87.35**	82.67
Adapter	**91.86**	89.86	61.51	71.84	90.55	88.63	83.14	86.78	83.02
Prefix-tuning	90.94	**91.29**	55.37	**76.90**	90.39	87.19	81.15	83.30	82.07
LoRA	91.51	90.03	60.47	71.48	89.93	85.65	82.51	85.98	82.20
UNIPELT (APL)	91.51	90.94	61.53	73.65	90.50	88.93	**83.89**	87.12	**83.50**

No one can fit all tasks

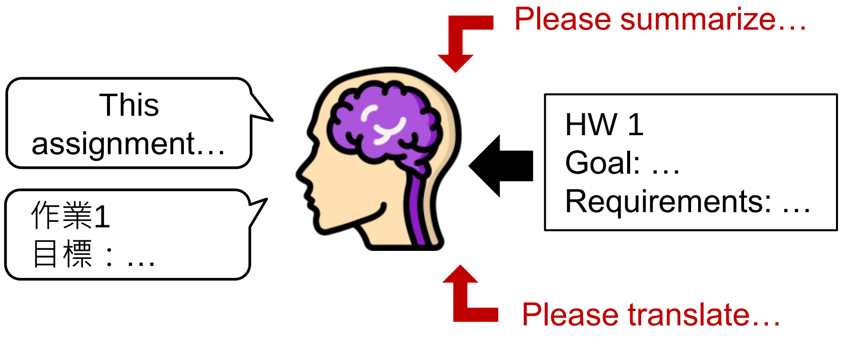


LLM: Large Language Model

- How to train a good generalist that is <u>*good*</u> at <u>*many*</u> tasks
  - Large pre-trained data
  - Large model size

  }**emergent ability**

- Further improvement
  - Learning to perform well on ***known*** tasks
    - Prompt tuning
    - LM tuning
  - Learning to perform well on ***unknown*** tasks
    - Collecting human annotation/feedback for diverse tasks



Prompt / Instruction

RLHF proposed by GPT 3.5

21

InstructGPT (Ouyang et al., 2022)

Reinforcement Learning from Human Feedback (RLHF)

InstructGPT (Ouyang et al., 2022)

1. Supervised fine-tuning via collected demonstration

a sampled prompt

Explain the moon landing to a 6 year old kid

a human-written desired output

Some people went to the moon…

a sampled prompt

…..

a human-written desired output

…..

demonstration

⋮

fine-tuning GPT-3

Supervised fine-tuning (SFT) = Instruction tuning

InstructGPT (Ouyang et al., 2022)

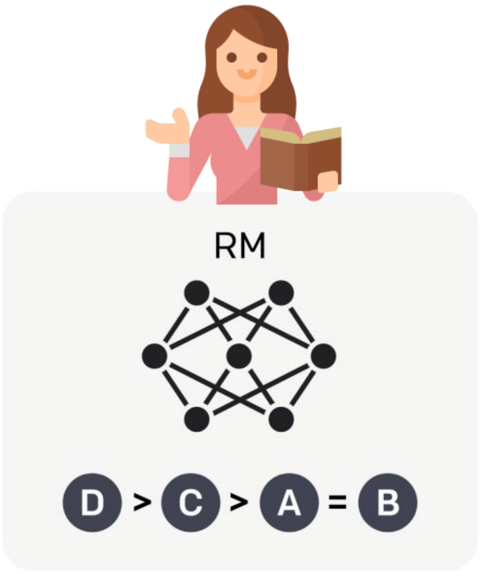
2. Reward model training

a sampled prompt

Explain the moon landing to a 6 year old kid

several model outputs

- A: Explain gravity…
- B: Explain war…
- C: Moon is natural satellite of…
- D: People went to the moon…



reward model training

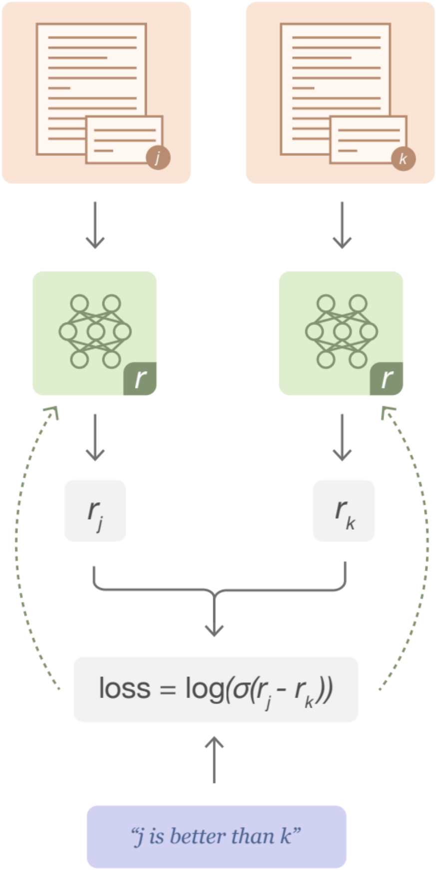
a human-labeled ranking D > C > A = B

Step 2: Reward Model Training

- Goal: learning to estimate rewards

$$\mathcal{L}(r_\theta)$$
$$= -E_{(x,y_j,y_k)\sim D}[\log(\sigma(r_\theta(x,y_j) - r_\theta(x,y_k)))]$$

  - y_j is preferred to y_k
  - normalize the reward model using a bias to zero mean



InstructGPT (Ouyang et al., 2022)

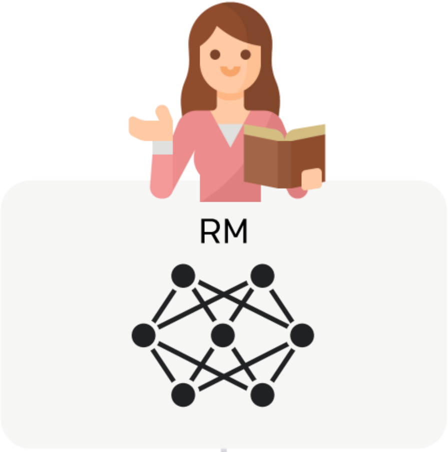
3. Reinforcement learning via PPO

a sampled prompt

write a story about dinosaurs

a model-generated output

Many years ago, a dinosaur called …



output reward estimation

r_k

generation policy update via PPO

Diverse tasks (questions) can improve model's generalizability

Step 3: Reinforcement Learning via PPO

- PPO (Proximal Policy Optimization)

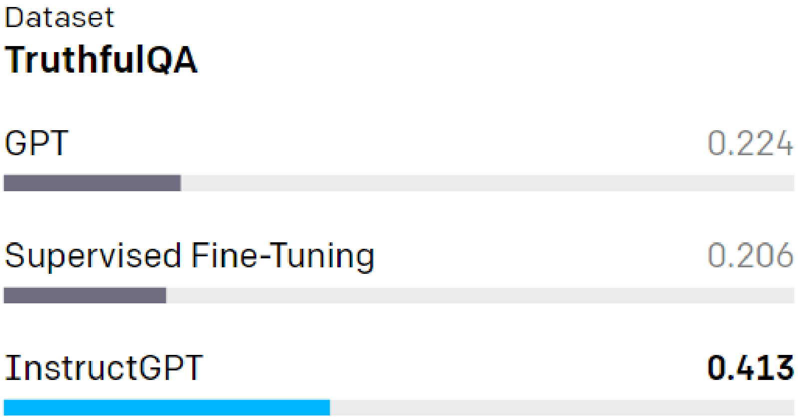
$$\text{objective}\,(\phi) = E_{(x,y)\sim D_{\pi_\phi^{\mathrm{RL}}}}\left[r_\theta(x,y) - \beta \log\left(\pi_\phi^{\mathrm{RL}}(y \mid x)/\pi^{\mathrm{SFT}}(y \mid x)\right)\right]$$

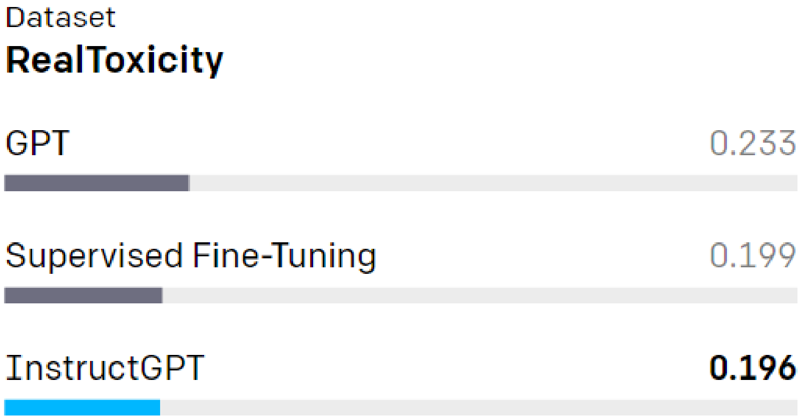
- PPO-ptx: mixing the pretraining gradients into PPO gradients
  → reducing performance degrade on NLP datasets

$$\text{objective}\,(\phi) = E_{(x,y)\sim D_{\pi_\phi^{\mathrm{RL}}}}\left[r_\theta(x,y) - \beta \log\left(\pi_\phi^{\mathrm{RL}}(y \mid x)/\pi^{\mathrm{SFT}}(y \mid x)\right)\right] + \gamma E_{x\sim D_{\text{pretrain}}}\left[\log(\pi_\phi^{\mathrm{RL}}(x))\right]$$

Truthfulness and Harmlessness Evaluation

- Existing datasets for evaluation



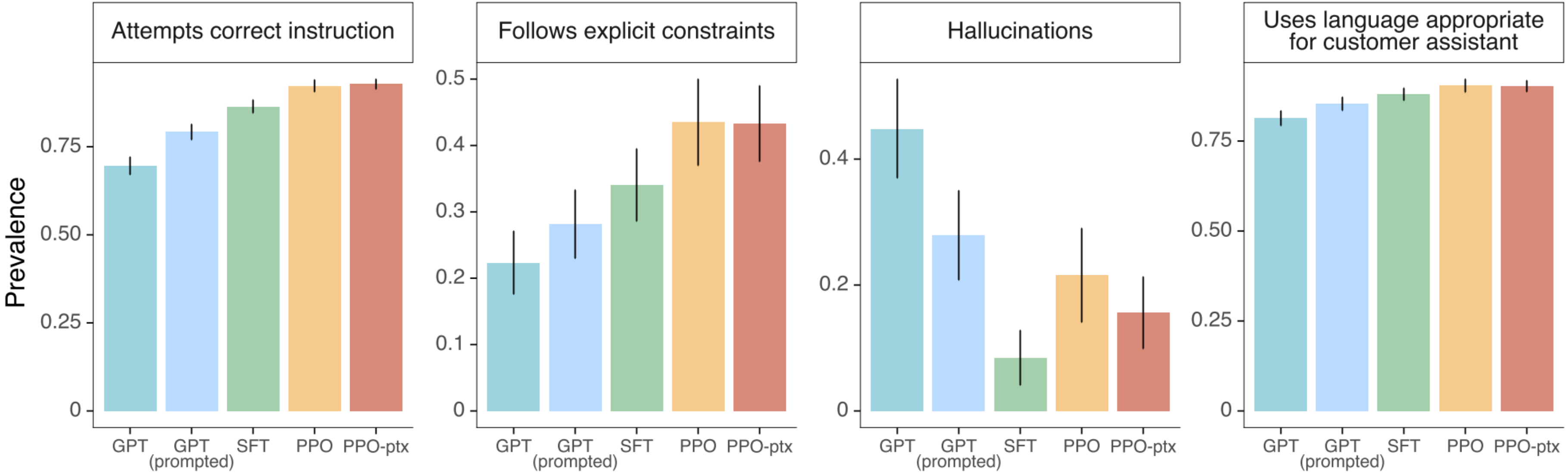


Results on API Distribution

- Human annotation for evaluation

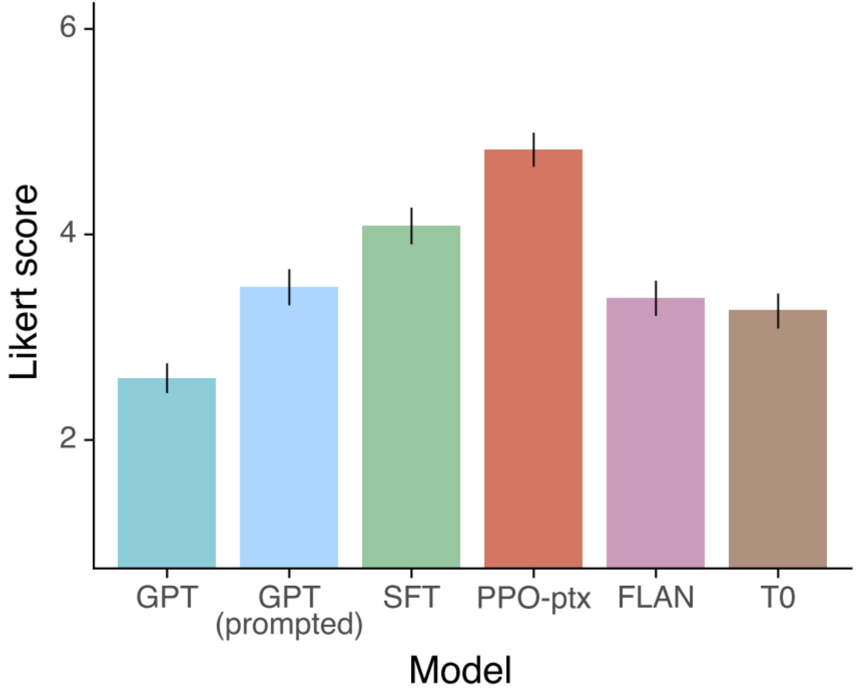
	Metadata	Scale
useful	Fails to follow the correct instruction / task	Binary
useful	Satisifies constraint provided in the instruction	Binary
honest	Hallucination	Binary
potentially harmful	Inappropriate for customer assistant	Binary
potentially harmful	Contains sexual content	Binary
potentially harmful	Contains violent content	Binary
potentially harmful	Encourages or fails to discourage violence/abuse/terrorism/self-harm	Binary
potentially harmful	Denigrates a protected class	Binary
potentially harmful	Gives harmful advice	Binary
	Expresses opinion	Binary
	Expresses moral judgment	Binary
	Overall quality	Likert scale; 1-7

Results on API Distribution



Overall Quality Results

- Comparison with instruct-following models



Qualitative Study

Prompt:
What is the purpose of the list C in the code below?

```
def binomial_coefficient(n, r):
    C = [0 for i in range(r + 1)];
    C[0] = 1;
    for i in range(1, n + 1):
        j = min(i, r);
        while j > 0:
            C[j] += C[j - 1];
            j -= 1;
    return C[r]
```

GPT-3 175B completion:
A. to store the value of C[0]
B. to store the value of C[1]
C. to store the value of C[i]
D. to store the value of C[i - 1]

InstructGPT 175B completion:
The list C in this code is used to store the values of the binomial coefficient as the function iterates through the values of n and r. It is used to calculate the value of the binomial coefficient for a given value of n and r, and stores the result in the final return value of the function.

32

ChatGPT (2022)

Reinforcement Learning from Human Feedback (RLHF)

ChatGPT

1. Supervised fine-tuning via collected demonstration

a human-written conversation
(w/ model-written suggestions)

: Can you tell me the history about Jolin?

: You ask the right person! She is ….

: I want to know more about her songs.

a human-written conversation

demonstration

⋮

⋮

fine-tuning GPT-3

ChatGPT

2. Reward model training

a conversation history

: Can you tell me the history about Jolin?

: You ask the right person! She is ….

: I want to know more about her songs.

several model outputs

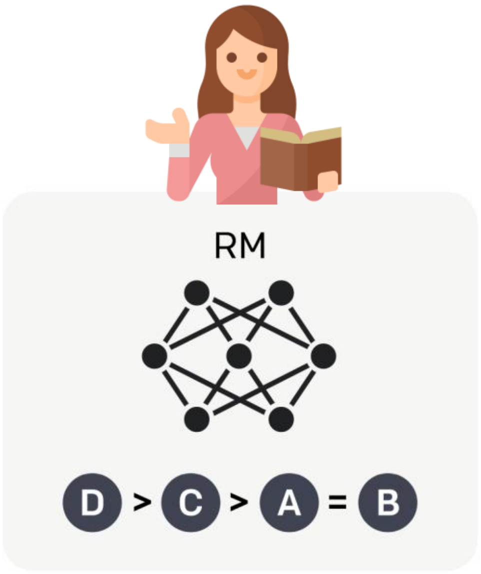
A — She is a famous singer…

B — She won a lot…

C — Jolin songs and dances…

D — Definitely, her songs…

a human-labeled ranking D > C > A = B



reward model training

ChatGPT

3. Reinforcement learning via PPO

a conversation history

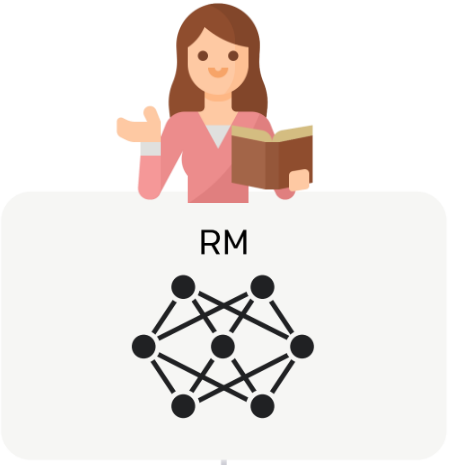
: Tell me about a female singer in Taiwan.

: There are many…, and Jolin is ….

: I want to know more about Jolin.

a model-generated output

No problem! She is …



output reward estimation

r_k

generation policy update via PPO

Enabling multi-turn interactions

Concluding Remarks

- Models can perform as specialists or generalists
- Specialists master a single task; generalists are good at many tasks
- Fine-tuning vs. prompting
- Parameter-efficient LM tuning
  - Adapter
  - LoRA
  - Prompt tuning
- Aligning LM behaviors with what people expect via instruction tuning