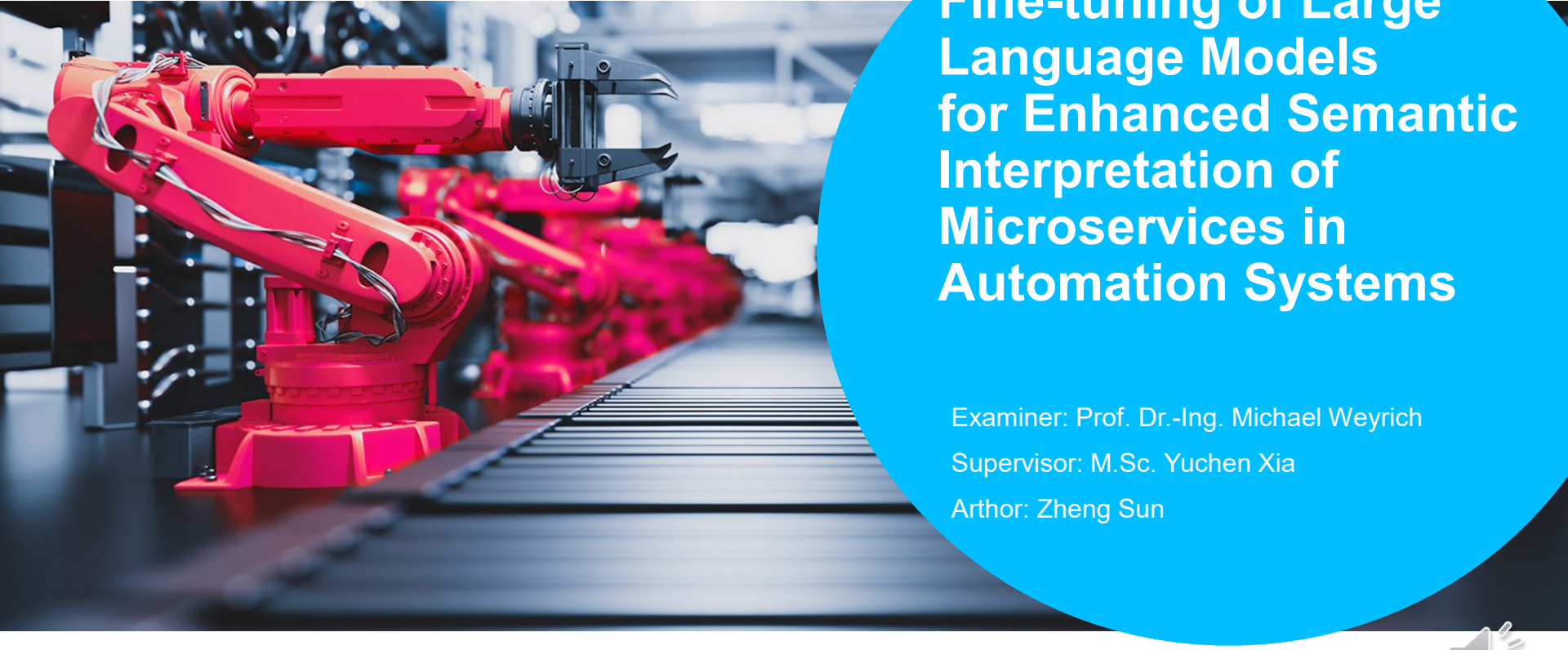
University of Stuttgart
Institute of Industrial Automation
and Software Engineering



Fine-tuning of Large Language Models for Enhanced Semantic Interpretation of Microservices in Automation Systems

Examiner: Prof. Dr.-Ing. Michael Weyrich

Supervisor: M.Sc. Yuchen Xia

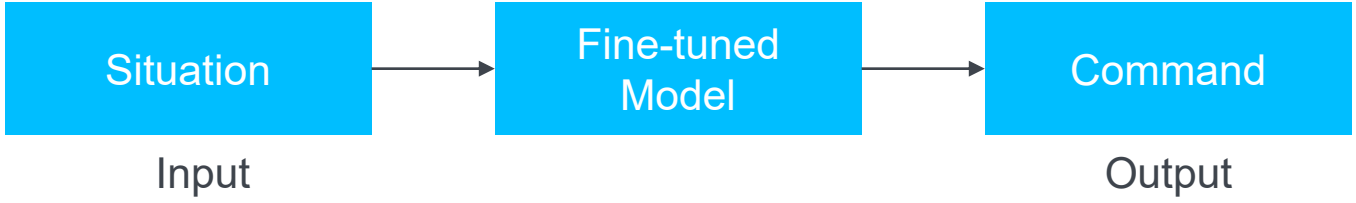
Arthor: Zheng Sun

Introduction

Privacy Concern → Small-scale language model

Introduction

Task description



Robots check the quality of cardboard for packaging boxes.

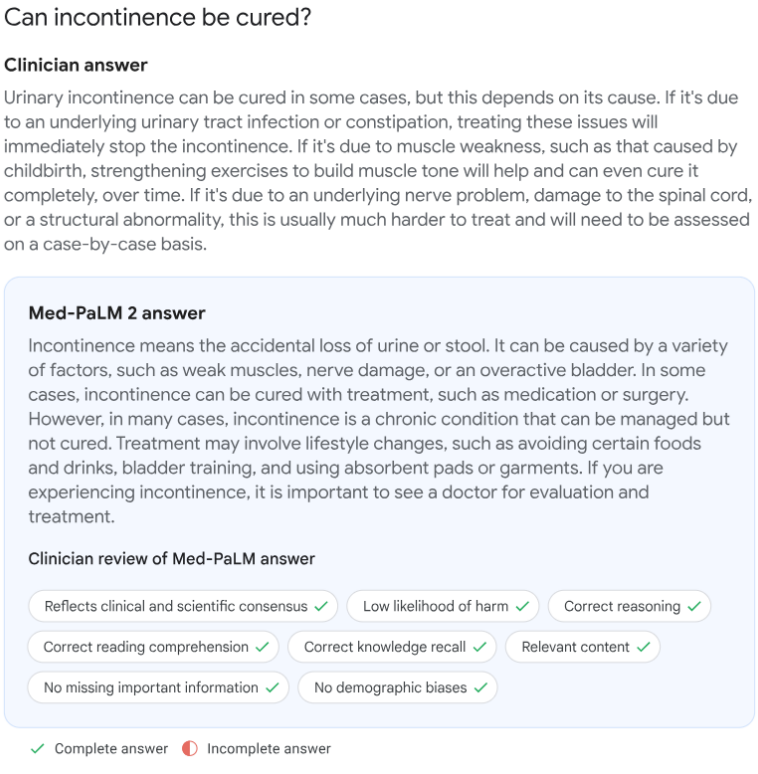
quality_check("cardboard")

A low supply of copper wiring is delaying the assembly of electrical circuits.

order_material("copper wiring")

Background

What can LLMs do?



[1]

Med-PaLM 2

- Scored up to 86.5% on the MedQA dataset
- Rapid progress towards physician-level performance

Background

What can LLMs do?

State	UBE Threshold
Alabama, Minnesota, Missouri, New Mexico, North Dakota	260
Indiana, Oklahoma	264
Connecticut, District of Columbia, Illinois, Iowa, Kansas, Kentucky Maryland, Montana, New Jersey, New York, South Carolina, Virgin Islands	266
Michigan	268
Alaska, Arkansas, Colorado, Maine, Massachusetts, Nebraska New Hampshire, North Carolina, Ohio, Oregon, Rhode Island, Tennessee, Texas, Utah, Vermont, Washington, West Virginia, Wyoming	270
Idaho, Pennsylvania	272
Arizona	273

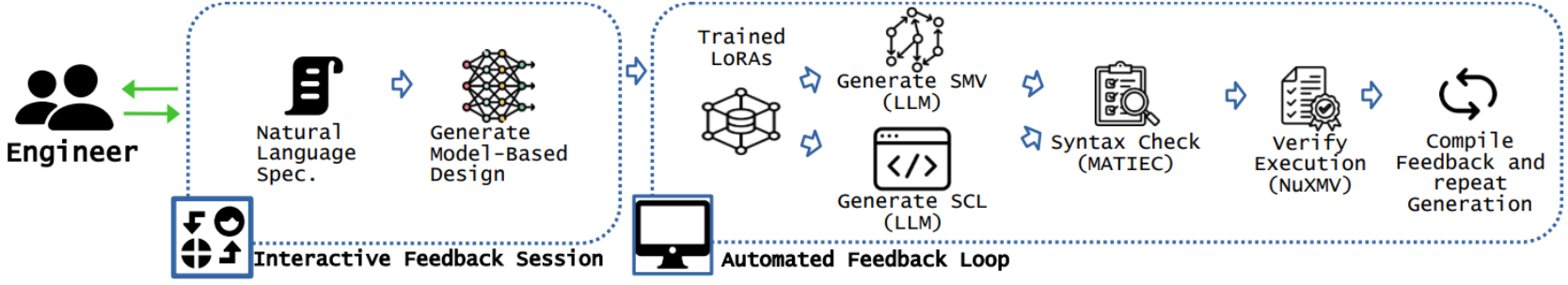
Table 8. Uniform Bar Exam (UBE) Minimum Passing Score by State.

[2]

(0-shot)GPT-4: 297

Background

What can LLMs do?

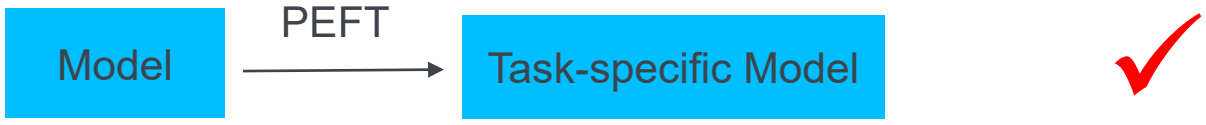


LLM4PLC[3]

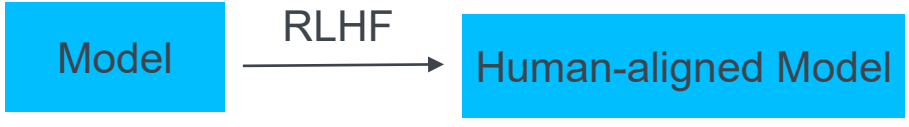
Background

Fine-tuning

- When do we need fine-tuning?
- PEFT vs. RLHF
- PEFT(Parameter-efficient Fine-tuning) :



- RLHF(Reinforcement-Learning from Human Feedback):

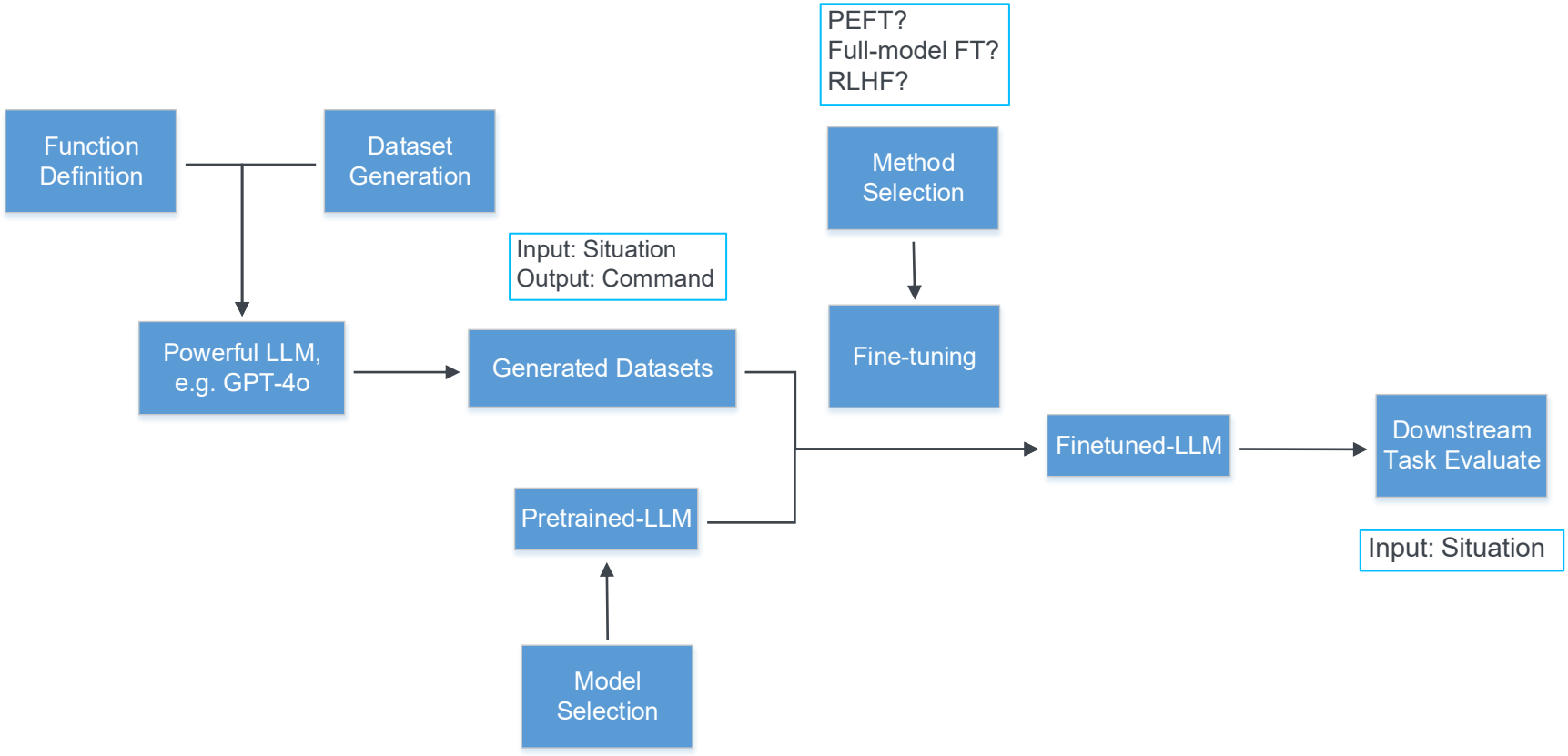


Contents

- Conceptual Design & Implementation
- Evaluation
- Summary and Outlook

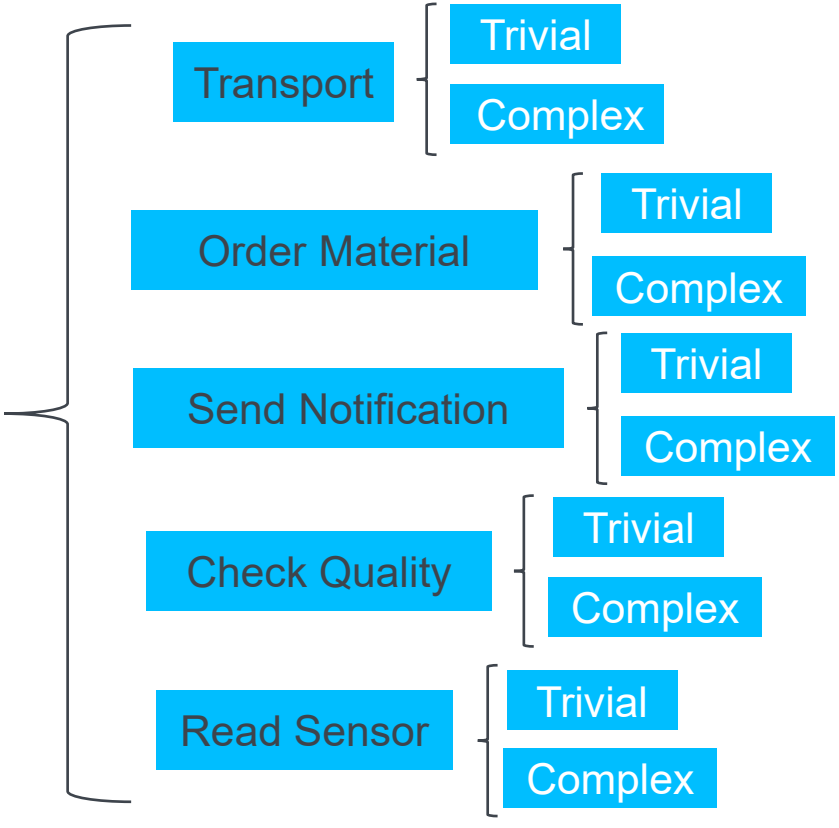
Conceptual Design & Implementation

Conceptual Design & Implementation



Conceptual Design & Implementation

Function Definition & Dataset Generation



Train Data: 1000
Test Data: 100

Example (transport):

Trivial: Input: Robots transport metal sheets from storage to the cutting machine. Output: transport("metal sheets", "storage", "cutting machine")

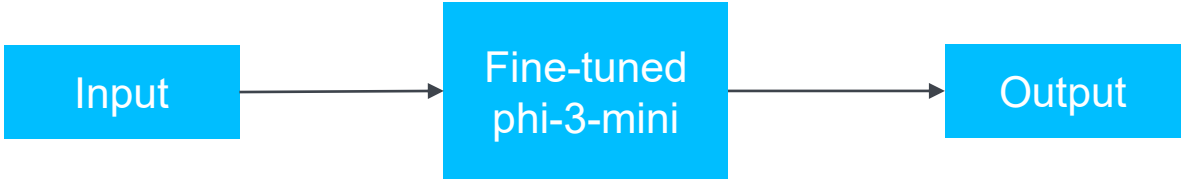
Complex: Input: Online orders are packed in the warehouse, awaiting to be placed at the front desk. Output: transport("online orders", "warehouse", "front desk")

Conceptual Design & Implementation

Model Selection

Microsoft Phi-3-mini

- Phi-3-mini: 3.8B parameters vs. ChatGPT-3.5: 175B parameters
- Can a small language model finish the given task?
- Experiment:



Who wrote a room of one's own?

Will it be windy in Paris next Tuesday?

WikiSearch("A Room of One's Own")

GetWeather("Paris", "next Tuesday")

Conceptual Design & Implementation

Model Selection

- Results (100 test data for each command type)

Aspect / Type	Awareness	Correct in Type	Correct in Type and Information
WikiSearch()	8	6	3
GetWeather()	69	63	59

- Trainable parameters: 0.65% of all the parameters.

Fine-tuned Phi-3-mini has potential to complete the downstream task!

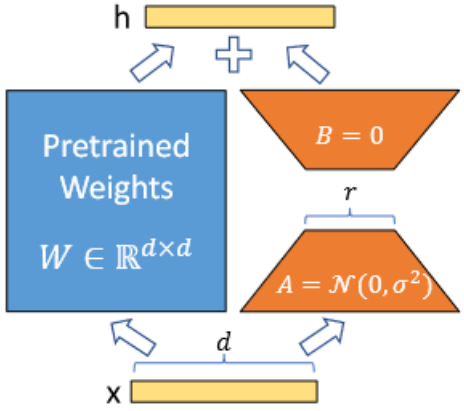
Conceptual Design & Implementation

Method Selection

- LoRA(Low Rank Adaptation)
- Update process: $W_0 + \Delta W = W_0 + BA$, where $W_0 \epsilon \mathbb{R}^{d \times k}$, $B \epsilon \mathbb{R}^{d \times r}$, $A \epsilon \mathbb{R}^{r \times k}$.
- Forward pass can be written as

$$h = W_0 x + \Delta W x = W_0 x + BAx. (1)$$

- Low Rank: rank $r \ll min(d, k)$



[4]

Evaluation

Evaluation

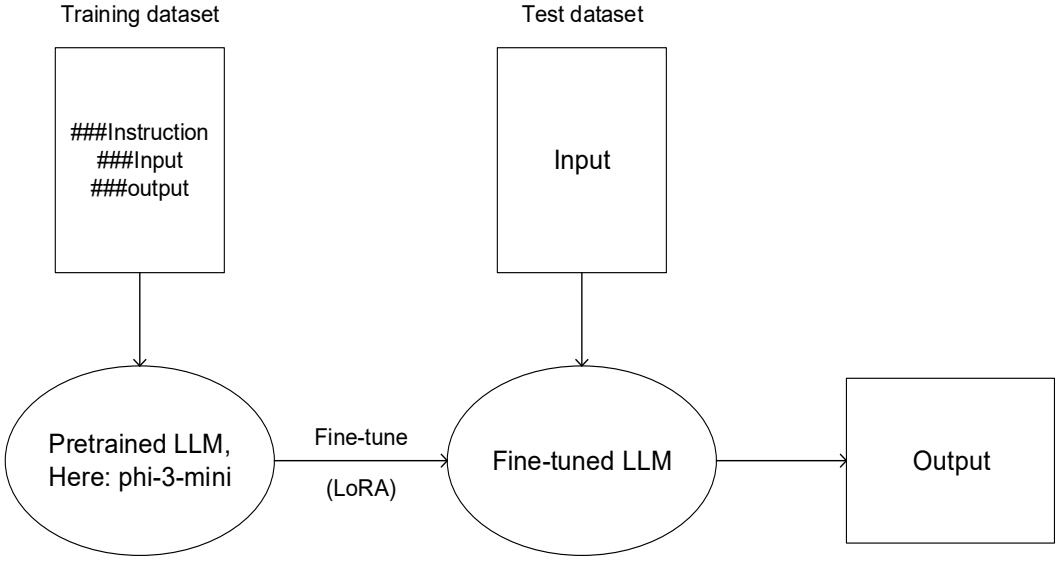
Robotics Commands

Test data: 100

# of Target Modules	LoRA Rank	# of Trainable Parameters(%)	Awareness(%)	Correct in Type(%)	Correct in Type and information(%)
2	8	0.1233	73	57	45
	16	0.2464	39	27	21
	32	0.4915	82	55	36
4	8	0.3282	69	59	45
	16	0.6543	91	72	66
	32	1.3001	65	59	58
5	8	0.3355	69	57	47
	16	0.6688	82	74	72
	32	1.3287	81	69	65

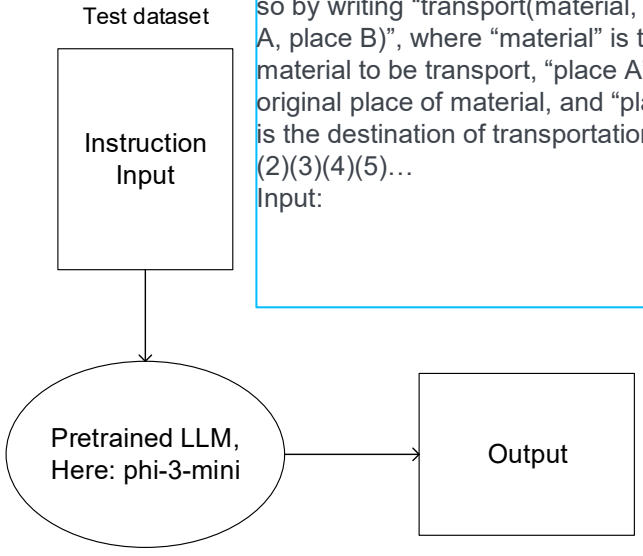
Evaluation

Fine-tuning vs. Using Prompt(zero-shot)



Fine-tuning (LoRA)

Prompt: You are a helpful assistant who are able to generate the correct commands for robots, according to the input missions. You have access to the following commands:
(1)transport(material:str, place A:str, place B:str): You can use the command to transport the materials. You can do so by writing “transport(material, place A, place B)”, where “material” is the material to be transport, “place A” is the original place of material, and “place B” is the destination of transportation.
(2)(3)(4)(5)…
Input:



Using Prompt(zero-shot)

Evaluation

Fine-tuning vs. Using Prompt(zero-shot)

Test data: 100

Aspect / Method	Awareness	Correct in Type	Correct in Type and Information	Extra Command In Total	GPU RAM	Input token	Time
Fine-Tuning	82	74	72	0	10.9 GB		
Using Prompt	98	90	16 (↓78%)	183	3.5 GB	~30 folds	~1.5x slower

Using prompt → Difficult to learn underlying rules → Unexpected problems

Evaluation

Catastrophic Forgetting

Dataset / Model	MMLU-Nutrition (3-Shot)	GSM8K(0-Shot)
Pre-Trained	56.5%	92.5%
Fine-Tuned	42.0%(↓25.6%)	85.5%(↓7.6%)

- MMLU: Multi-choice questions cover a wide range.
- GSM8K: school math problems with diverse complexity.
- Each dataset contains 200 test data.

- “Memorized” knowledge are easier to forget than reasoning skills

Summary and Outlook

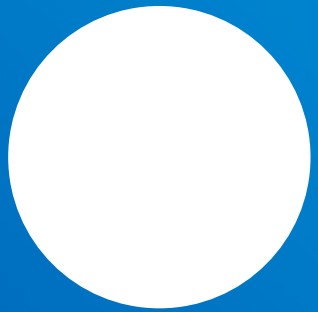
Summary and Outlook

- A fine-tuned model with 3B parameters can complete the specific task! (72%)
- LoRA as a fine-tuned method can achieve great performance while reducing parameters!
- Fine-tuned model with high quality data is more effective and efficient in the long run!

- Future work: Fine-tune more models as AI agents to finish complex tasks with high quality.

University of Stuttgart
Institut of Industrial Automation
and Software Engineering

Thank you!



Zheng Sun

e-mail st181999@stud.uni-stuttgart.de
phone +49 (0) 711 685-
fax +49 (0) 711 685-

University of Stuttgart
Institut of Industrial Automation and Software Engineering

Attachment

Examples of dataset

	Input	Output
Trivial	Robots transport metal sheets from storage to the cutting machine.	transport("metal sheets", "storage", "cutting machine")
Complex	Online orders are packed in the warehouse, awaiting to be placed at the front desk.	transport("online orders", "warehouse", "front desk")
Trivial	Robots order more fuel oil.	order_material("fuel oil")
Complex	he inventory system is low on RFID tags for tracking high-value items.	order_material("RFID tags")
Trivial	Robots send notifications to the lab analysis team.	send_notification("Lab Analysis Team")
Complex	The robot encounters a network connectivity issue in the office.	send_notification("IT Support")
Trivial	Robots check the quality of steel sheets for car body production.	quality_check("steel sheet")
Complex	A chemical processing plant receives barrels of specialty solvents.	quality_check("specialty solvents")
Trivial	The robot checks the weight of an object it is carrying using a weight sensor.	get_sensor_reading("weight sensor")
Complex	The robot needs to detect obstacles in its path.	get_sensor_reading("ultrasonic sensor")

Attachment

Examples of training data(general-purposed)

Instruction: You are a helpful assistant who are able to provide correct answer. If you are not sure about the answer or want to look up some updated information to produce a better answer, you can use API calls to help you get information required to generate correct answer. You have access to the following APIs: (1)WikiSearch("term"): You can use this API to look up information on Wikipedia.You can do so by writing “Wiki Search("term")” where “term” is the search term you want to look up. (2)GetWeather(city:str, time:str): You can use this API to look up information about weather and specified time. You can do so by writing “GetWeather(city, time)”, where “city” is the city you want to look up, and “time” is the specified time you want to know. Here is a example of API calls:
Input: Where do you find the engine number on a car?
Output: WikiSearch("Engine number")

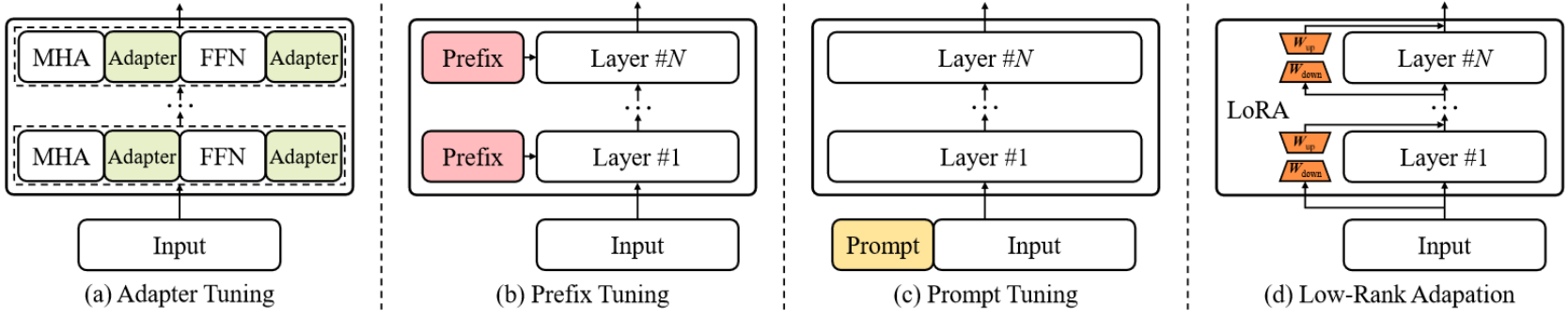
Attachment

Examples of training data(robotics)

Instruction: You are a helpful assistant who are able to generate the correct commands for robots, according to the input missions. You have access to the following commands: (1)transport(material:str, place A:str, place B:str): You can use the command to transport the materials. You can do so by writing “transport(material, place A, place B)”, where “material” is the material to be transport, “place A” is the original place of material, and “place B” is the destination of transportation. (2)order_material(material_name:str): You can use the command to order materials. You can do so by writing “order_material(material_name)”, where “material_name” is the name of material which you want to order. (3)send_notificaton(person_name:str): You can use the command to send notification to person. You can do so by writing “send_notificaton(person_name)”, where “person_name” is the name of person who you want to send notification to. (4)quality_check(material:str): You can use the command to check the quality of material. You can do so by writing “quality_check(material)”, where “material” is the material which you want to do quality check. (5)get_sensor_reading(sensor:str): You can use the command to read the information from sensor. You can do so by writing “get_sensor_reading(sensor)”, where “sensor” is the sensor which you want to read. Here is an example of robot commands: ### Input: The robot finishes an inventory audit. ### Output: send_notification("Inventory Manager")

Attachment

PEFT(Parameter-Efficient Fine-Tuning)



[5]

Bibliography

[1] K. Singhal *et al.*, “Towards Expert-Level Medical Question Answering with Large Language Models,” May 16, 2023, *arXiv*: arXiv:2305.09617. doi: 10.48550/arXiv.2305.09617.

[2] Katz, Daniel Martin and Bommarito, Michael James and Gao, Shang and Arredondo, Pablo, GPT-4 Passes the Bar Exam (March 15, 2023). 382 Philosophical Transactions of the Royal Society A (2024), Available at SSRN: https://ssrn.com/abstract=4389233

Bibliography

[3] M. Fakih, R. Dharmaji, Y. Moghaddas, G. Q. Araya, O. Ogundare, and M. A. A. Faruque, “LLM4PLC: Harnessing Large Language Models for Verifiable Programming of PLCs in Industrial Control Systems,” in Proceedings of the 46th International Conference on Software Engineering: Software Engineering in Practice, Apr. 2024, pp. 192–203. doi: 10.1145/3639477.3639743.

[4] E. J. Hu et al., “LoRA: Low-Rank Adaptation of Large Language Models,” Oct. 16, 2021, arXiv: arXiv:2106.09685. Accessed: Feb. 27, 2024. [Online]. Available: http://arxiv.org/abs/2106.09685

[5] W. X. Zhao *et al.*, “A Survey of Large Language Models,” Nov. 24, 2023, *arXiv*: arXiv:2303.18223. Accessed: Apr. 10, 2024. [Online]. Available: http://arxiv.org/abs/2303.18223