

Effective use of generative AI in creating exam questions

Tsunenori Ishioka



formerly of **Nat. Ctr. for
Univ. Entrance Exam.**

1

Introduction

- In Nov. 2022, **ChatGPT**, an excellent **generative AI**, was released, surprising the world.
- Additional learning using **own data and knowledge**
 - Incremental, Fine-tuning, Transfer
- A specific application: **test question generation** for **education purposes**

2

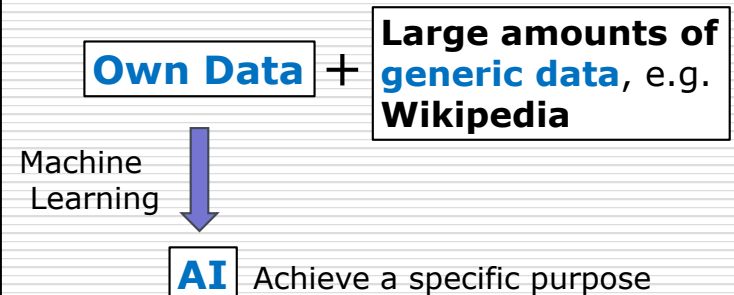
Outline

- Three methods of **additional learning** in generative AI
- Three **components** and three **aspects** to consider in **test writing/creating**
- Discussion



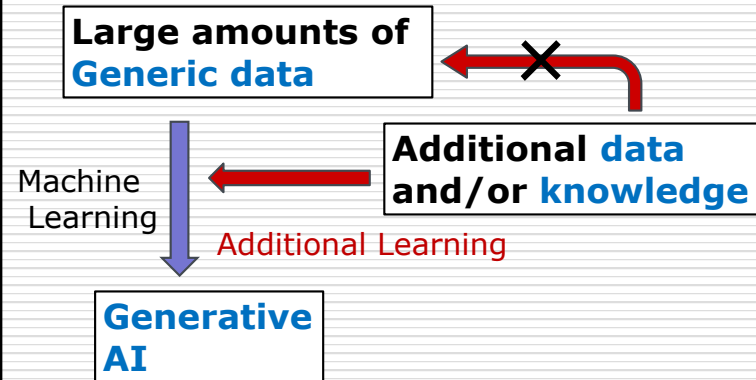
3

How to generate an AI model



4

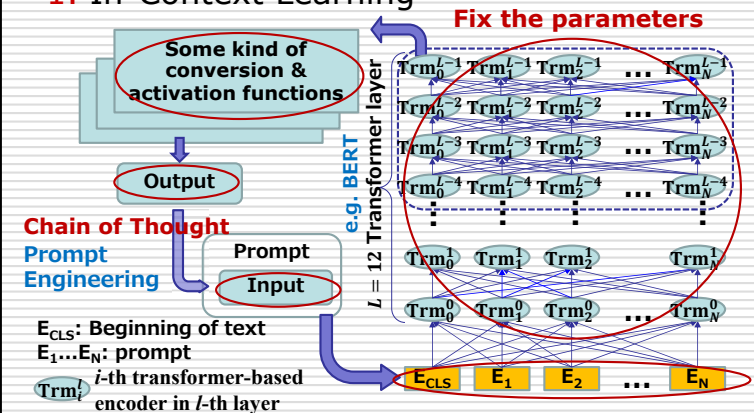
Generative AI



5

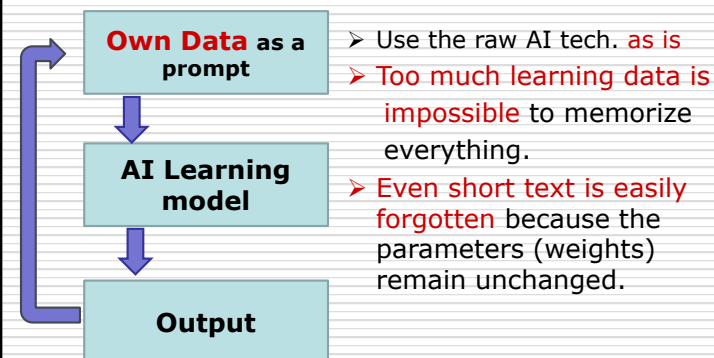
Three ways to learn more

1. In-Context Learning



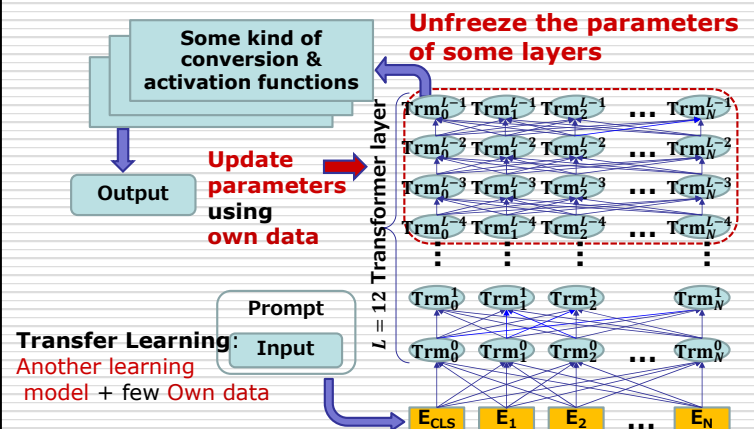
6

In-Context Learning



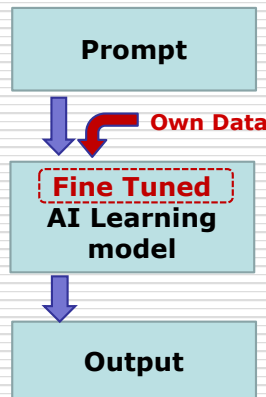
7

2. Fine-tuning



8

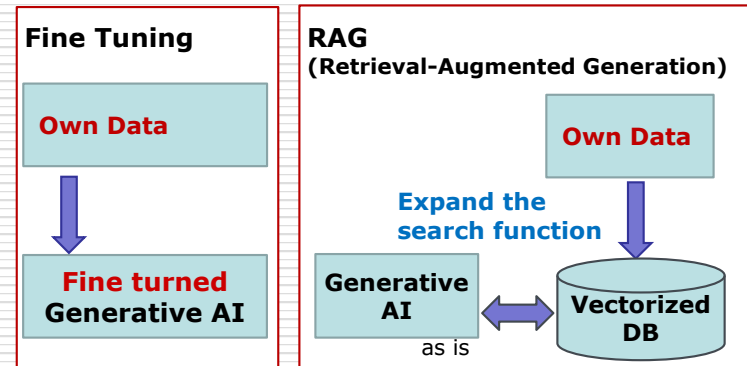
Disadvantages of fine-tuning



- Learning becomes **more demanding**.
- Additional learning materials don't provide a **satisfactory answer**.
- Output is **uncertain**

9

3. RAG & Embedding



10

LLM Orchestration Framework

 LangChain

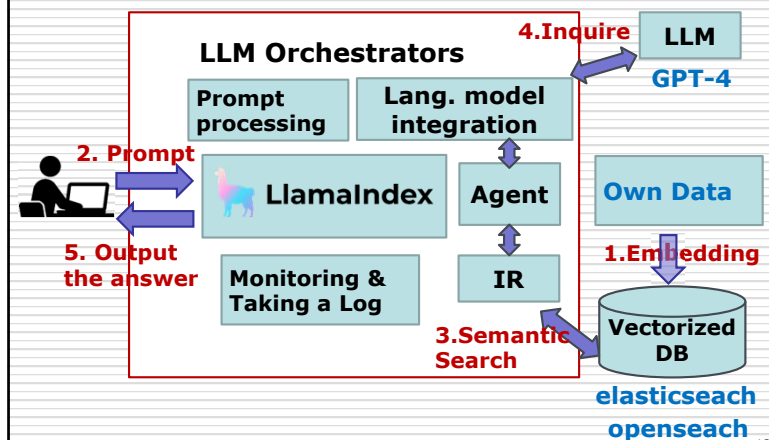
 LlamaIndex

Semantic Kernel by Microsoft

 orkes

11

Example of RAG structure



12

Agent RAG

1. Set the task objective 
2. Search and retrieve relevant information 
3. Generate and evaluate the initial answer 
4. Perform additional searches or calculations as needed (agent intervention)  
5. Output the final answer 

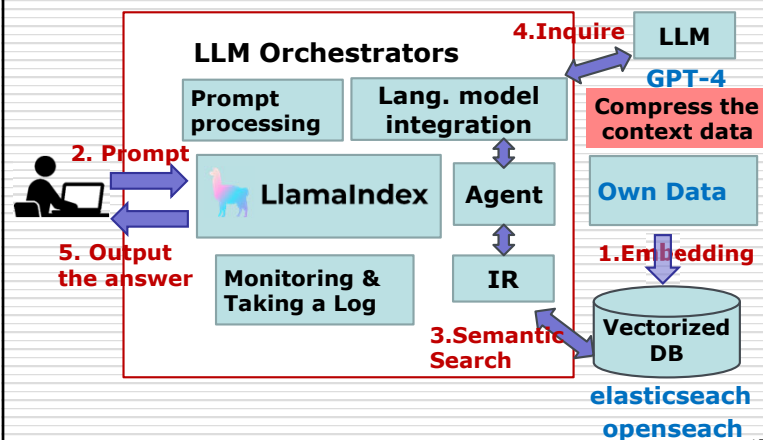
13

RAG Latest research

- Compress the context data passed to RAG to the **absolute minimum**.

14

Latest RAG structure



15

Latest RAG Example

- **REFRAG**: Superintelligence Labs (Meta), Sep. 2025; Make RAG **30 times** faster
- **xRAG**: Peking Univ./Microsoft, Neurips 2024, May.
- Trade-off between **#of char.** entered and the **speed/accuracy** of the response

16

Q Generation for Education Purposes

- Essential Components
- **Context** = Prompt; Input text or information; ≠ “discourse context”
- **Question** = Output;
 - *factoid* or *non-factoid*
 - *single-hop* or *multi-hop*
 - *standalone* or *sequence*
 - *free text* or *multiple choice*
- **Answer.** *answer-aware* (=given) or *answer-unaware*

17

Neural Question Generation Survey (2016~)

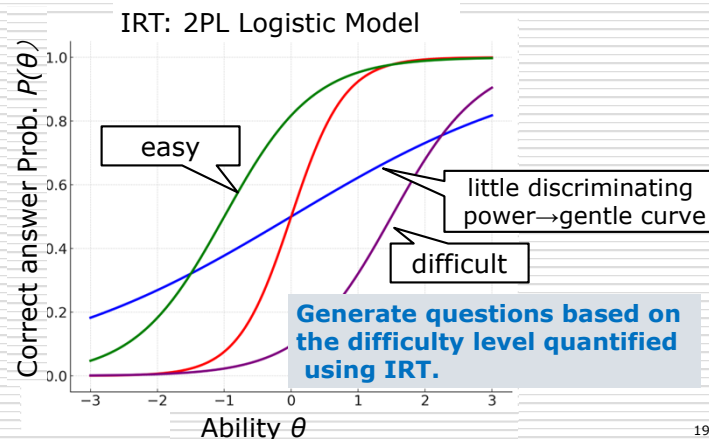
- Al-Farby, et al. (2024), Review on Neural Question Generation for Education Purposes

Inter. J. of Artificial Intelligence in Education, **34**, 1008–1045 [Open access](#)

- 390 = 302 (201 Google Scholar, 34 ACL, 16 ACM, 51 DBLP) + 88 cited
→ 224 papers

18

Points to consider when creating exam questions – 1. Difficulty adjustment



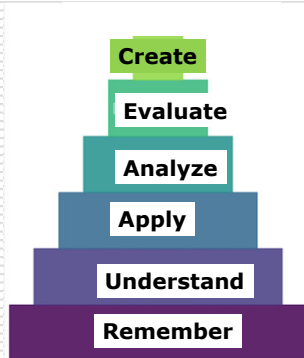
19

Difficulty of a question

- **Answerability:** If a question is **challenging** for a QA system to answer accurately, it is considered **more difficult** (Gao et al., 2019).
- **Textual complexity:** # of different parts or passages of text that need to be considered to answer the question (Cheng et al., 2021).
- **Answer length:** Questions with **longer and more detailed answers** may be considered more difficult (Murakhovska et al., 2021).

20

2. Bloom's Taxonomy(Rev.2001)



- A framework for classifying learning objectives into 6 levels

Under providing the definition of Bloom's Taxonomy, Chain of Thought Approach was used.

Scaria, et.al, 2024 , AIED2024

3. Distractor Generation

- Criteria for a good incorrect answer
 - unambiguously wrong
 - should not be obviously incorrect
 - = effective as a means of distraction
- A major challenge in automated question generation (Taslimipoor et al. 2024)

22

Latest Research of Distractor Generation

- Taslimipoor et al.,(2024), LREC-Coling'24
 - "creation model that creates correct and incorrect answers" vs "answer model that selects the correct answer."
- Shimei et al.(2025), AIED'25
 - Provide students' error patterns data with LLM

23

Conclusion & Future work

- Neural methods for QG
 - Making fluent & related questions from a limited length of context is OK.
 - Making high-level questions has not yet.
 - We need
 - more specialized datasets,
 - Enhancing the effectiveness of automatic evaluation
- To generate abstractive, long, and useful questions.

24

Discussion

■ ChatBOT UI

- ChatGPT can be used locally

■ Local LLM tool

- GPT4All
- LM Studio
- llama.cpp

25

THANK YOU!

26