



[Interview Question 4 - #LLM]

Question: What are some of the strategies to reduce hallucination in large language models (LLMs)?

Mastering LLM (Large Language Model)



1

Hallucination Reduction

1. Prompt Level
2. Model Level
3. Self-check

Mastering LLM (Large Language Model)



2

Prompt Level



Mastering LLM (Large Language Model)



3

Include instructions to the model

Include instruction to the model not to make up stuff on its own. For e.g. “Do not make up stuff outside of given context”

Mastering LLM (Large Language Model)



4

Repeat

Repeat most important instructions multiple times.

Mastering LLM (Large Language Model)



5

Position

Position most important instruction at the end,
making use of latency effect

Mastering LLM (Large Language Model)



6

Parameter

Keep temperature to 0

Mastering LLM (Large Language Model)



7

Restrict

Restrict output to a confined list instead of free float text

Mastering LLM (Large Language Model)



8

Add: Add CoT type instructions

For GPT models “Lets think step by step” works for reasoning task, For PaLM – “Take a deep breath and work on this problem step by step” outperforms.

Mastering LLM (Large Language Model)

Add: Use Few shot examples

Use domain/use case specific few shot examples, also there is a recent study on new technique called "Analogical Prompting" where model generates its own examples internally which outperforms few shot prompting.



10

Add: In Context Learning

Use In-context learning to provide better context to the model

Mastering LLM (Large Language Model)



11

Add: Self-consistency/Voting

Generating multiple answers from the model and selecting the most frequent answers.

Mastering LLM (Large Language Model)



12

Model Level



Mastering LLM (Large Language Model)



13

DoLa

Decoding by Contrasting Layers Improves Factuality in Large Language Models: Simple decoding strategy in large pre-trained LLMs to reduce hallucinations

<https://github.com/voidism/DoLa>

Mastering LLM (Large Language Model)



14

Fine-Tuned model on good quality data

Fine-tuning small LLM model on good quality data has shown promising results as well as help reduce hallucinations

Mastering LLM (Large Language Model)



15

Self-check



Mastering LLM (Large Language Model)



16

Chain of verification

Methods like chain of verification can help reduce the hallucinations to a great extent, read more here <https://tinyurl.com/masteringllm>

Mastering LLM (Large Language Model)



Save this post

Save this post for later review, like, reshare & comment for maximum reach.

Mastering LLM (Large Language Model)



What's Your Take?

Have you used any other strategies to reduce hallucination in LLMs? Share your experiences in the comments!

Mastering LLM (Large Language Model)