

www.masteringllm.com

Top 100+ Interview Questions for Large language models (LLM)



Mastering LLM
@mastering-llm-large-language-
model

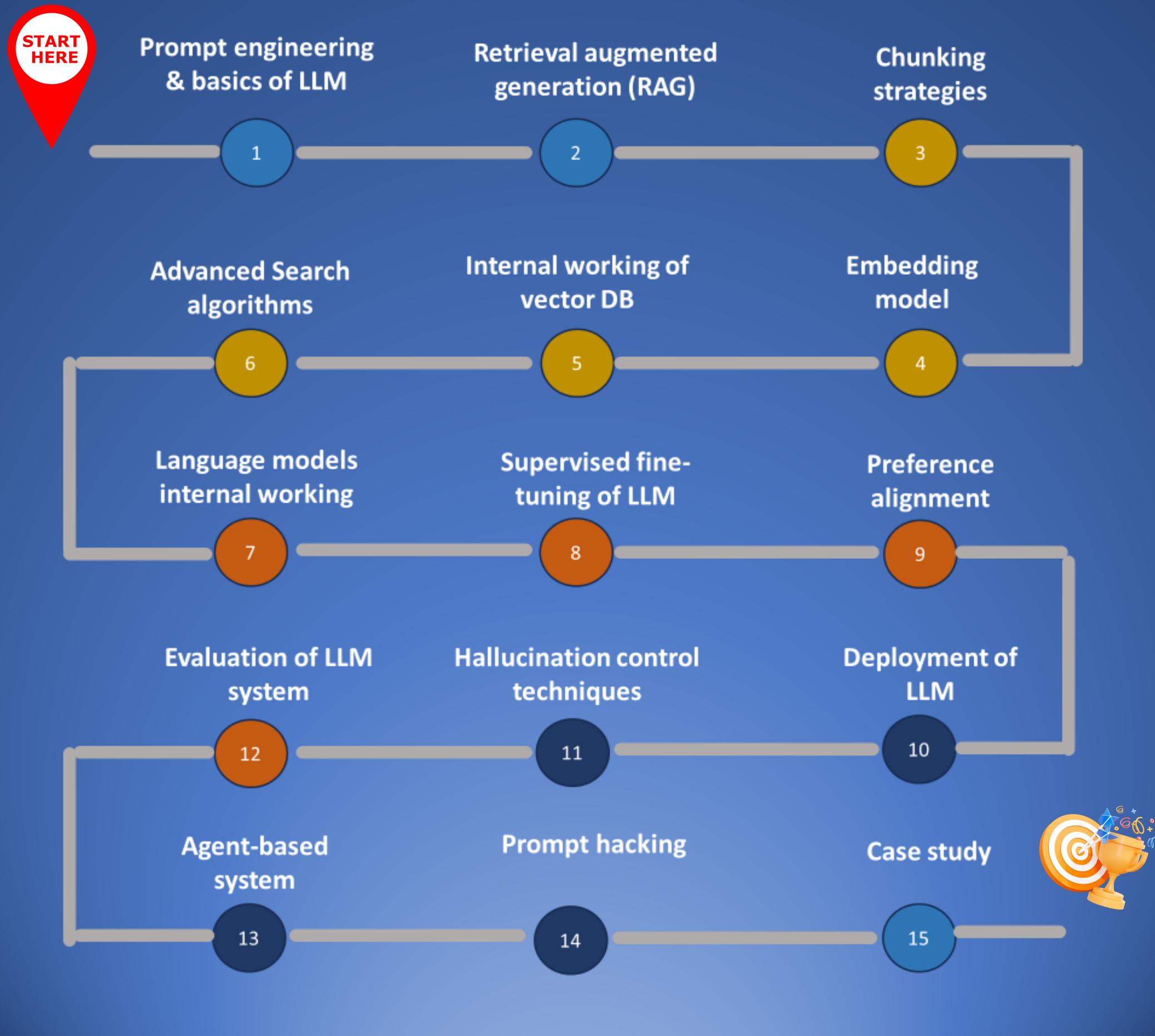


What will this document cover?

1. Road map
2. Prompt engineering & basics of LLM
3. Retrieval augmented generation (RAG)
4. Chunking strategies
5. Embedding Models
6. Internal working of vector DB
7. Advanced search algorithms
8. Language models internal working
9. Supervised fine-tuning of LLM
10. Preference Alignment (RLHF/DPO).
11. Evaluation of LLM system
12. Hallucination control techniques
13. Deployment of LLM
14. Agent-based system
15. Prompt Hacking
16. Case Study & Scenario-based question



RoadMap to prepare for LLM Interview



Mastering LLM
@mastering-llm-large-language-model



www.masteringllm.com

**50%
OFF**

LLM Interview Course



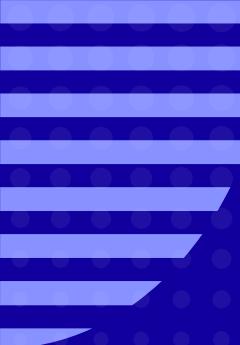
Want to Prepare yourself for an
LLM Interview?

- ✓ 100+ Questions spanning 14 categories
- ✓ Curated 100+ assessments for each category
- ✓ Well-researched real-world interview questions based on FAANG & Fortune 500 companies
- ✓ Focus on Visual learning
- ✓ Real Case Studies & Certification



Coupon Code - LINK50

Coupon is valid till 30th April 2024



www.masteringllm.com

Prompt engineering & basics of LLM



Mastering LLM
@mastering-llm-large-language-
model



1

What is the difference between Predictive/ Discriminative AI and generative AI?



Mastering LLM
@mastering-llm-large-language-
model



2

What is LLM & how LLMs are trained?



Mastering LLM
@mastering-llm-large-language-
model



3

What is a token in the language model?



Mastering LLM
@mastering-llm-large-language-
model



4

How to estimate the cost of running a SaaS-based & Open source LLM model?



Mastering LLM
@mastering-llm-large-language-model



5

Explain the
Temperature
parameter and how
to set it.



Mastering LLM
@mastering-llm-large-language-
model



6

What are different decoding strategies for picking output tokens?



Mastering LLM
@mastering-llm-large-language-model



7

What are the different ways you can define stopping criteria in a large language model?



Mastering LLM
@mastering-llm-large-language-model



8

How to use stop sequence in LLMs?



Mastering LLM
@mastering-llm-large-language-
model

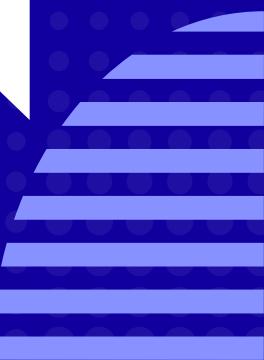


9

Explain the basic structure of prompt engineering.



Mastering LLM
@mastering-llm-large-language-model



10

Explain the type of
prompt
engineering.



Mastering LLM
@mastering-llm-large-language-
model



11

Explain In-Context Learning.



Mastering LLM
@mastering-llm-large-language-model



12

What are some of
the aspects to keep
in mind while using
few-shots
prompting?



Mastering LLM
@mastering-llm-large-language-
model



13

What are certain strategies to write good prompts?



Mastering LLM
@mastering-llm-large-language-model



14

What is
hallucination & how
can it be controlled
using prompt
engineering?



Mastering LLM
@mastering-llm-large-language-
model



15

How do I improve the reasoning ability of my LLM through prompt engineering?



Mastering LLM
@mastering-llm-large-language-
model



16

How to improve LLM reasoning if your COT prompt fails?



Mastering LLM
@mastering-llm-large-language-
model



Retrieval augmented generation (RAG)



Mastering LLM
@mastering-llm-large-language-
model



1

How to increase accuracy, and reliability & make answers verifiable in LLM?



Mastering LLM
@mastering-llm-large-language-model



2

How does Retrieval augmented generation (RAG) work?



Mastering LLM
@mastering-llm-large-language-
model



3

What are some of
the benefits of
using the RAG
system?



Mastering LLM
@mastering-llm-large-language-
model



4

What are the architecture patterns you see when you want to customize your LLM with proprietary data?



Mastering LLM
@mastering-llm-large-language-model



5

When should I use Fine-tuning instead of RAG?



Mastering LLM
@mastering-llm-large-language-
model



www.masteringllm.com

Chunking strategies



Mastering LLM
@mastering-llm-large-language-
model



1

What is chunking and why do we chunk our data?



Mastering LLM
@mastering-llm-large-language-
model



2

What are factors influences chunk size?



Mastering LLM
@mastering-llm-large-language-model



3

What are the
different types of
chunking methods
available?



Mastering LLM
@mastering-llm-large-language-
model



4

How to find the ideal chunk size?



Mastering LLM
@mastering-llm-large-language-model



www.masteringllm.com

Embedding Models



Mastering LLM
@mastering-llm-large-language-
model



1

What are vector embeddings? And what is an embedding model?



Mastering LLM
@mastering-llm-large-language-model



2

How embedding model is used in the context of LLM application?



Mastering LLM
@mastering-llm-large-language-
model



3

What is the
difference between
embedding short
and long content?



Mastering LLM
@mastering-llm-large-language-
model



4

How to benchmark embedding models on your data?



Mastering LLM
@mastering-llm-large-language-
model



5

Walk me through the steps of improving the sentence transformer model used for embedding.



Mastering LLM
@mastering-llm-large-language-model



Internal working of vector DB



Mastering LLM
@mastering-llm-large-language-
model



1

What is vector DB?



Mastering LLM
@mastering-llm-large-language-
model



2

How vector DB is different from traditional databases?



Mastering LLM
@mastering-llm-large-language-
model



3

How does a vector database work?



Mastering LLM
@mastering-llm-large-language-model



4

Explain the
difference between
vector index, vector
DB & vector plugins.



Mastering LLM
@mastering-llm-large-language-
model

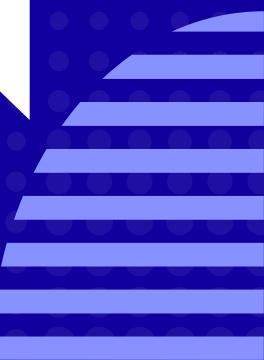


5

What are different vector search strategies?



Mastering LLM
@mastering-llm-large-language-model



6

How does clustering reduce search space? When does it fail and how can we mitigate these failures?



Mastering LLM
@mastering-llm-large-language-
model



7

Explain the Random projection index.



Mastering LLM
@mastering-llm-large-language-
model



8

Explain the Locality-sensitive hashing (LHS) indexing method?



Mastering LLM
@mastering-llm-large-language-model



9

Explain the product quantization (PQ) indexing method.



Mastering LLM
@mastering-llm-large-language-model



10

Compare different Vector indexes and given a scenario, which vector index you would use for a project?



Mastering LLM
@mastering-llm-large-language-model

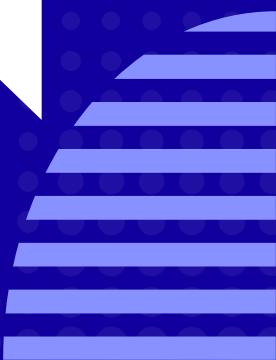


11

How would you decide on ideal search similarity metrics for the use case?



Mastering LLM
@mastering-llm-large-language-model



12

Explain the different types and challenges associated with filtering in vector DB.



Mastering LLM
@mastering-llm-large-language-model

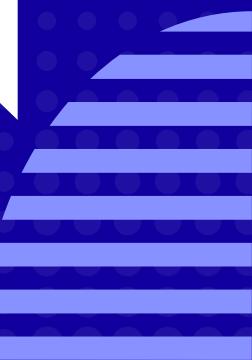


13

How do you
determine the best
vector database for
your needs?



Mastering LLM
@mastering-llm-large-language-
model



Advanced search algorithms



Mastering LLM
@mastering-llm-large-language-
model



1

Why it's important to have very good search



Mastering LLM
@mastering-llm-large-language-
model



2

What are the architecture patterns for information retrieval & semantic search, and their use cases?



Mastering LLM
@mastering-llm-large-language-model



3

How can you
achieve efficient
and accurate
search results in
large scale
datasets?



Mastering LLM
@mastering-llm-large-language-
model



4

Explain the keyword-based retrieval method.



Mastering LLM
@mastering-llm-large-language-model



5

How to fine-tune re-ranking models?



Mastering LLM
@mastering-llm-large-language-
model



6

Explain most common metric used in information retrieval and when it fails?



Mastering LLM
@mastering-llm-large-language-model



7

I have a recommendation system, which metric should I use to evaluate the system?



Mastering LLM
@mastering-llm-large-language-model

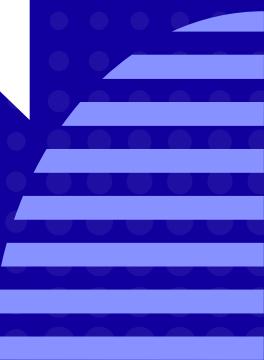


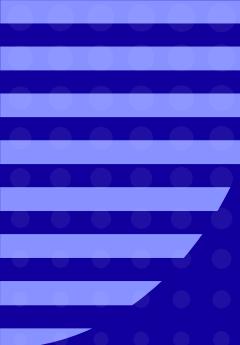
8

Compare different information retrieval metrics and which one to use when?



Mastering LLM
@mastering-llm-large-language-
model





www.masteringllm.com

Language models internal working



Mastering LLM
@mastering-llm-large-language-
model

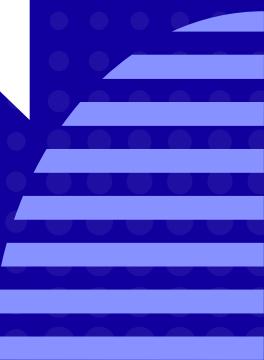


1

Detailed understanding of the concept of self-attention.



Mastering LLM
@mastering-llm-large-language-model



2

Overcoming the disadvantages of the self-attention mechanism.



Mastering LLM
@mastering-llm-large-language-
model

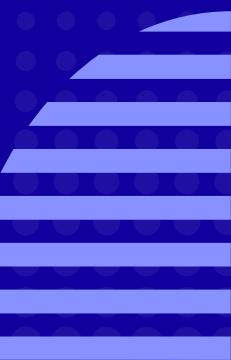


3

Understanding positional encoding



Mastering LLM
@mastering-llm-large-language-
model



4

Detailed explanation of Transformer architecture.



Mastering LLM
@mastering-llm-large-language-model



5

Advantages of using a transformer instead of LSTM.



Mastering LLM
@mastering-llm-large-language-
model



6

Difference between local attention and global attention.



Mastering LLM
@mastering-llm-large-language-
model



7

Understanding the computational and memory demands of transformers.



Mastering LLM
@mastering-llm-large-language-model



8

Increasing the context length of an LLM.



Mastering LLM
@mastering-llm-large-language-model

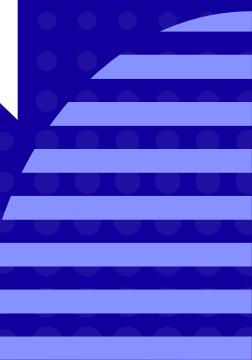


9

How to Optimizing transformer architecture for large vocabularies.



Mastering LLM
@mastering-llm-large-language-
model

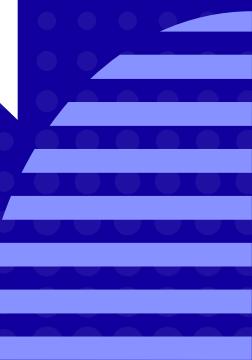


10

What is a mixture of expert models?



Mastering LLM
@mastering-llm-large-language-
model



Supervised fine-tuning of LLM



Mastering LLM
@mastering-llm-large-language-model

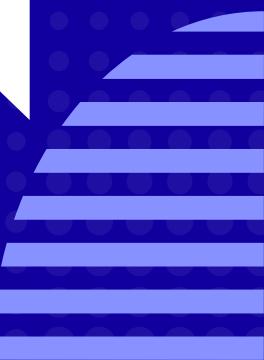


1

What is finetuning and why it's needed in LLM?



Mastering LLM
@mastering-llm-large-language-
model



2

Which scenario do we need to finetune LLM?



Mastering LLM
@mastering-llm-large-language-model



3

How to make the decision of fine-tuning?



Mastering LLM
@mastering-llm-large-language-model

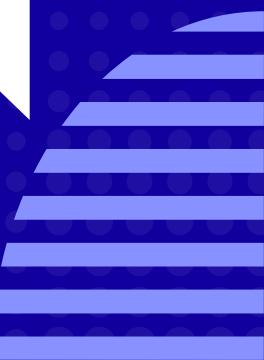


4

How do you create a fine-tuning dataset for Q&A?



Mastering LLM
@mastering-llm-large-language-
model



5

How do you
improve the model
to answer only if
there is sufficient
context for doing
so?



Mastering LLM
@mastering-llm-large-language-
model



6

How to set hyperparameter for fine-tuning.



Mastering LLM
@mastering-llm-large-language-
model



7

How to estimate infra requirements for fine-tuning LLM?



Mastering LLM
@mastering-llm-large-language-
model



8

How do you fine-tune LLM on consumer hardware?



Mastering LLM
@mastering-llm-large-language-model

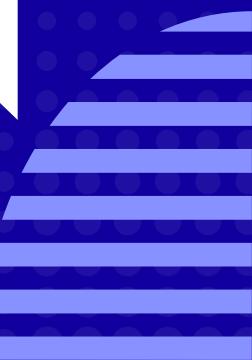


9

What are the
different categories
of the PEFT method?



Mastering LLM
@mastering-llm-large-language-
model



10

Explain different re-parameterized methods for fine-tuning LLM?



Mastering LLM
@mastering-llm-large-language-model



11

What is catastrophic forgetting in the context of LLMs?



Mastering LLM
@mastering-llm-large-language-
model



Preference Alignment (RLHF/DPO)



Mastering LLM
@mastering-llm-large-language-
model

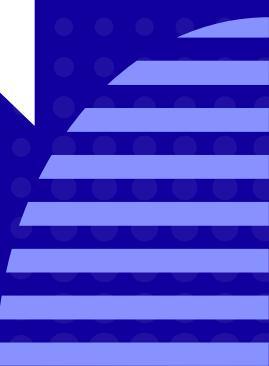


1

At which stage you will decide to go for the Preference alignment type of method rather than SFT?



Mastering LLM
@mastering-llm-large-language-model



2

Explain Different Preference Alignment Methods?



Mastering LLM
@mastering-llm-large-language-model



3

What is RLHF, and how is it used?



Mastering LLM
@mastering-llm-large-language-
model

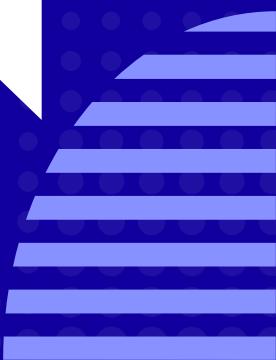


4

Explain the reward hacking issue in RLHF.



Mastering LLM
@mastering-llm-large-language-model



www.masteringllm.com

Evaluation of LLM system



Mastering LLM
@mastering-llm-large-language-
model



1

How do you evaluate the best LLM model for your use case?



Mastering LLM
@mastering-llm-large-language-
model

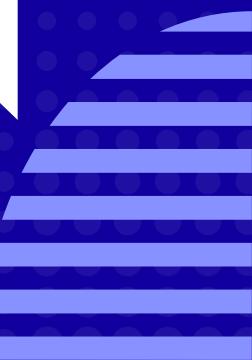


2

How to evaluate the RAG-based system?



Mastering LLM
@mastering-llm-large-language-
model



3

What are the
different metrics
that can be used to
evaluate LLM



Mastering LLM
@mastering-llm-large-language-
model

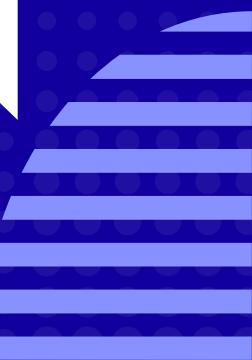


4

Explain the Chain of verification



Mastering LLM
@mastering-llm-large-language-model



www.masteringllm.com

Hallucination control techniques



Mastering LLM
@mastering-llm-large-language-
model



1

What are the different forms of hallucinations?



Mastering LLM
@mastering-llm-large-language-model



2

How do you control hallucinations at different levels?



Mastering LLM
@mastering-llm-large-language-
model



www.masteringllm.com

Deployment of LLM



Mastering LLM
@mastering-llm-large-language-
model

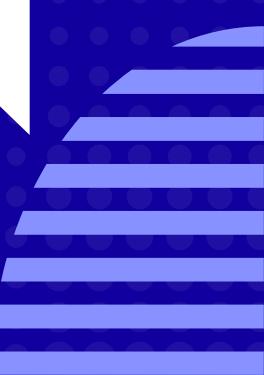


1

Why does quantization not decrease the accuracy of LLM?



Mastering LLM
@mastering-llm-large-language-model



Agent-based system



Mastering LLM
@mastering-llm-large-language-model



1

Explain the basic concepts of an agent and the types of strategies available to implement agents.



Mastering LLM
@mastering-llm-large-language-model



2

Why do we need agents and what are some common strategies to implement agents?



Mastering LLM
@mastering-llm-large-language-model

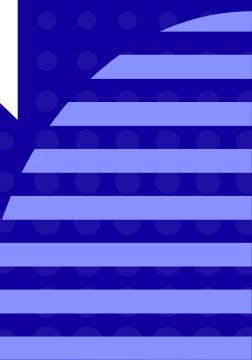


3

Explain ReAct prompting with a code example and its advantages



Mastering LLM
@mastering-llm-large-language-
model



4

Explain Plan and Execute prompting strategy



Mastering LLM
@mastering-llm-large-language-model



5

Explain OpenAI functions with code examples



Mastering LLM
@mastering-llm-large-language-
model



6

Explain the
difference between
OpenAI functions vs
LangChain Agents.



Mastering LLM
@mastering-llm-large-language-
model



www.masteringllm.com

Prompt Hacking



Mastering LLM
@mastering-llm-large-language-
model



1

What is prompt hacking and why should we bother about it?



Mastering LLM
@mastering-llm-large-language-
model

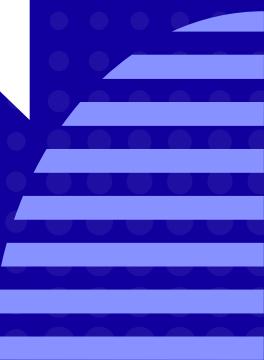


2

What are the different types of prompt hacking?



Mastering LLM
@mastering-llm-large-language-model



3

What are the different defense tactics from prompt hacking?



Mastering LLM
@mastering-llm-large-language-model



Case study & scenario-based Question



Mastering LLM
@mastering-llm-large-language-
model

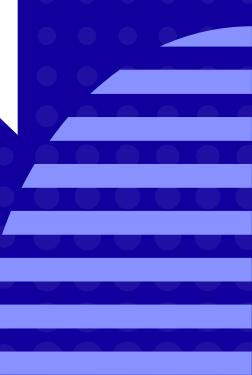


1

How to optimize the cost of the overall LLM System?



Mastering LLM
@mastering-llm-large-language-model



www.masteringllm.com

We can't give away
all our secrets! 😊

We're feeling extra generous,
we're offering a 50% discount!

Use the discount code below

Code: LINK50

Code is valid till 30th April

2024.



Mastering LLM
@mastering-llm-large-language-
model

