

LLM Prompting With Database

questions

Blueprint

Motivation

Build an automatic pipeline that scraps websites given links to files. Those files combined and pushed into a data base. The database is connected to Large Language Model (LLM) for following questioning.

Tools

- Web scrapping with Beautiful Soup
- Prompting with OpenAI LLM text-davinci-003
- Storing and indexing data with ChromaDB
- Combining altogether with LangChain framework

Workflow and Results

4 of 6 links downloaded successfully

3 of 4 links parsed successfully to txt files

All the files are combined into a single txt file

The single txt file attached to a database and indexed

Machine learning questions-answers prompted with plain LLM and based on the texts

Prompts

First prompt:

The following is an ML engineer interview. Interviewee should ask 3 the most wide-spread technical questions about machine learning.

Interviewee: Hello, my name is Artem and I am a software engineer.

Interviewer:

Second prompt:

The following is an ML engineer interview. Interviewee asked 3 the most wide-spread technical questions about machine learning. Interviewer should answer all the questions - {questions} - 25 words maximum for each answer.

Interviewee: {questions}

Interviewer:

NOTE: underlined Interviewee was written with mistake - “Enterviewee”, and LLM treated it.

Workflow and Results

Guided queries – general

1. What is the difference between supervised and unsupervised learning?
Supervised learning uses labeled data to train a model, while unsupervised learning uses unlabeled data to identify patterns and clusters.
2. What is the purpose of a neural network?
Neural networks are used to identify patterns and relationships in data, and can be used for classification and regression tasks.
3. What is the difference between deep learning and traditional machine learning?
Deep learning is a subset of machine learning that uses multiple layers of neural networks to learn from data, while traditional machine learning uses simpler algorithms.

Workflow and Results

Guided queries – with sources

1. What are the different types of machine learning?
Supervised, Unsupervised, and Reinforcement Learning.
2. What is bias and variance in machine learning?
Bias is the difference between the expected and actual values of the model. Variance is the measure of how the model's predictions vary for different data points.
3. What is clustering in machine learning?
Clustering is an unsupervised learning technique that groups data points into clusters based on their similarity.

References

Links

- <https://www.simplilearn.com/tutorials/machine-learning-tutorial/machine-learning-interview-questions>
- <https://www.interviewbit.com/machine-learning-interview-questions/>
- <https://www.mygreatlearning.com/blog/machine-learning-interview-questions/>
- [https://www.javatpoint.com/machine-learning-interview-questions",](https://www.javatpoint.com/machine-learning-interview-questions)
- <https://www.springboard.com/blog/data-science/machine-learning-interview-questions/>
- <https://intellipaat.com/blog/interview-question/machine-learning-interview-questions/>

The end

Thank you for your attention!