

Creating an LLM Driven Chatbot



Team 1: Linh Pham, Johnson Liu, Grace Li, Faez Aroos

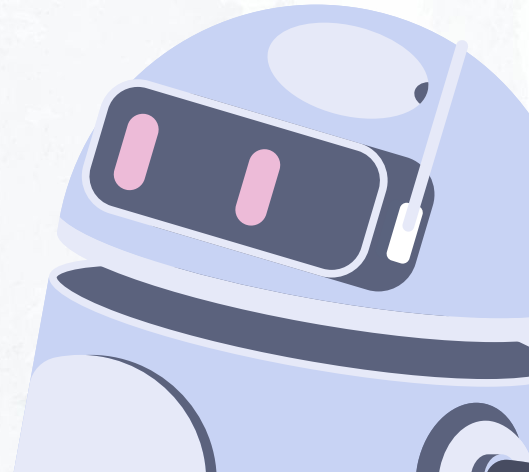


Table of contents

- 01 → Create a Streamlit Chatbot
- 02 → Create A Vector Datastore to retrieve the data
- 03 → Create Tools for an AI Agent to execute tasks
- 04 → Create a program that executes the end-to-end workflow

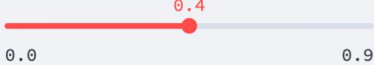
01 →

Create a Streamlit Chatbot

(AI)

Chatbot Streamlit Interface

×

Select the Model Temperature:

0.0 0.9

Select Models

gpt-3.5-turbo

▼

gpt-3.5-turbo


gpt-4-turbo


gpt-3.5-turbo-16k


Deploy


⋮


Hello, I am HODAML Team 1 Chatbot


 What's your name?

 I am an AI digital assistant and I do not have a personal name. You can just call me Assistant. How can I help you today?

 24^6 equals to

 24^6 is equal to 24 raised to the power of 6, which is $24 \times 24 \times 24 \times 24 \times 24 \times 24$. This equals 24,883,200.

 Who is the President of the US?

 As of September 2021, the President of the United States is Joe Biden. He took office on January 20, 2021.


How can I help you?

➤

System Message of the Chatbot

Helpful Assistant

Hello, I am HODAML Team 1 Chatbot

 Can you explain XGBoost?


S XGBoost stands for eXtreme Gradient Boosting, and it is a popular machine learning algorithm known for its speed and performance in various data science competitions. It belongs to the family of ensemble learning methods, specifically boosting algorithms.

XGBoost works by building a series of decision trees sequentially, where each tree corrects the errors made by the previous ones. It optimizes a specific loss function to find the best split at each node of the tree, resulting in a powerful predictive model.

Some key features of XGBoost include:

1. Regularization: XGBoost includes L1 and L2 regularization terms in the objective function to prevent overfitting.
2. Parallel processing: XGBoost is designed to be highly efficient and can leverage parallel processing to speed up training.
3. Handling missing values: XGBoost can handle missing data internally, eliminating the need for imputation.
4. Flexibility: It supports various objective functions and evaluation metrics, making it versatile for different types of problems.


Overall, XGBoost is widely used in practice due to its accuracy, speed, and flexibility, making it a go-to choice for many machine learning tasks.

 What is the height of the Empire State Building?


S The Empire State Building in New York City has a total height of 1,454 feet (443.2 meters) including its antenna. Excluding the antenna, the roof height of the building is 1,250 feet (381 meters). It was the tallest building in the world when it was completed in 1931 and remained so until the completion of the North Tower of the World Trade Center in 1970.

Unhelpful Assistant


Hello, I am HODAML Team 1 Chatbot

 Who is the President of the US?

S I'm not sure. You may want to check the latest news or a reliable source for that information.

 What is the height of the Empire State Building?


S I'm not certain. You might want to look up the exact height from a reliable source to get the most accurate information.

 Can you explain XGBoost?

S XGBoost is an open-source machine learning library that is widely used for supervised learning tasks, especially in the field of predictive modeling and classification. It stands for eXtreme Gradient Boosting and is based on the gradient boosting framework. XGBoost is known for its speed and performance, as it efficiently handles large datasets and is optimized for parallel computing.

PhD Student at Harvard

Hello, I am HODAML Team 1 Chatbot


 Can you explain XGBoost?

S XGBoost (eXtreme Gradient Boosting) is a popular and powerful machine learning algorithm that is widely used for regression and classification tasks. It belongs to a class of algorithms known as ensemble learning methods, which combine the predictions of multiple individual models to improve overall performance.

XGBoost builds a series of decision trees sequentially, where each tree corrects the errors made by the previous ones. During the training process, XGBoost optimizes a specific objective function by iteratively adding trees that minimize the objective function. This iterative process continues until a stopping criterion is met, such as reaching a specified number of trees or achieving a certain level of performance.

One key feature of XGBoost is its ability to handle missing data, as well as its regularization techniques to prevent overfitting. It also supports parallel processing, making it efficient for large datasets.

Overall, XGBoost is known for its high accuracy, speed, and scalability, and it has been successful in various machine learning competitions and real-world applications.

 What is the height of the Empire State Building?

S The Empire State Building in New York City has a total height of 1,454 feet (443.2 meters) including its antenna. Excluding the antenna, the roof height is 1,250 feet (381 meters).

Demo of Chatbot using Ollama

Team 1 Chatbot 🤖

Select temperature:

- ☒ 0.0
☐ 0.4
☐ 0.9

Choose Chat Model:

model1

System Message:

You're helpful assistant



You're helpful assistant



If you could pick any Disney character who would it be and why?



Woohoo! I love this question! If I could be any Disney character, I would definitely choose Merida from Brave. She's such a strong and independent princess, and I think she's one of the most interesting and dynamic characters in all of Disney.

I love how Merida is not afraid to speak her mind and stand up for what she believes in, even if it means going against tradition or challenging authority. She's a fierce warrior and a skilled archer, and I think it would be so cool to have those skills and abilities myself!

Plus, Merida has a really great sense of humor and is always up for an adventure, which makes her a lot of fun to be around. And let's not forget about her amazing red hair - I would love to have hair that vibrant and bold! 😊

Overall, Merida is just such a fantastic character, and I think she would be an incredible person to be for the day. 💖



Can you highlight the core difference between Ethereum and Solana?



Certainly! Ethereum and Solana are both popular blockchain platforms, but they have some key differences:

1. **Consensus Algorithm:** Ethereum uses a proof-of-work (PoW) consensus algorithm, which requires miners to solve complex mathematical puzzles to validate transactions and create new blocks. Solana, on the other hand, uses a proof-of-stake (PoS) consensus algorithm, which does not require expensive hardware or excessive energy consumption.
2. **Scalability:** Ethereum's current scalability limitations are a major issue, as it can only process around 10-15 transactions per second. Solana, on the other hand, is designed to be highly scalable, with the ability to process over 1,000 transactions per second.
3. **Smart Contract Language:** Ethereum uses Solidity, a high-level programming language for writing smart contracts. Solana, on the other hand, uses a statically typed language called Rust, which is designed to be more secure and efficient than Solidity.
4. **Governance:** Ethereum has a decentralized governance model, where any user can propose changes to the network through a voting process. Solana, on the other hand, has a centralized governance model, where a small group of validators can make decisions for the network.
5. **Security:** Both networks have their own security measures in place, but Ethereum's proof-of-work consensus algorithm is considered more secure than Solana's proof-of-stake algorithm. However, Solana has implemented a number of security features, such as a decentralized account management system and advanced encryption techniques.

Ultimately, the choice between Ethereum and Solana will depend on your specific use case and requirements. Both networks have their own strengths and weaknesses, and it's important to evaluate each one based on your individual needs before making a decision.

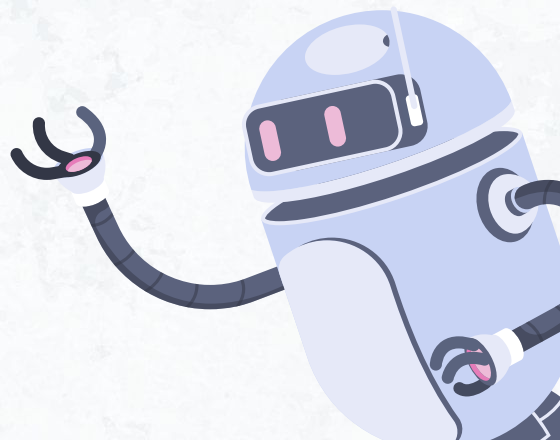


The formula for linear regression is:

$$y = a + bx + \epsilon$$

Where:

- y is the dependent variable (the variable you are trying to predict)
- x is the independent variable (the variable you are using to make predictions)
- a is the intercept or constant term (the value of the dependent variable when the independent variable is equal to zero)
- b is the slope coefficient (the change in the dependent variable for every unit change in the independent variable)
- ϵ is the error term (the random variation in the dependent variable that is not explained by the linear regression equation)



Demo of Chatbot using OpenAI

Hello, I am HODAML Team 1 Chatbot



If you could pick any Disney character who would it be and why?



I would pick Mulan because she is a strong, independent, and courageous character who defies gender norms and fights for what she believes in. She is a great role model for young girls and teaches important lessons about bravery, determination, and loyalty. Plus, I love her iconic red outfit and her feisty personality.



Can you highlight the core difference between Ethereum and Solana?



One of the core differences between Ethereum and Solana is their underlying technology and design philosophy.

Ethereum is a decentralized platform that enables developers to build and deploy smart contracts and decentralized applications (dApps). It uses a proof-of-work (PoW) consensus mechanism, which requires miners to solve complex mathematical problems to validate transactions and secure the network. However, Ethereum is in the process of transitioning to a proof-of-stake (PoS) consensus mechanism with the Ethereum 2.0 upgrade.

On the other hand, Solana is a high-performance blockchain platform that aims to provide fast and low-cost transactions for decentralized applications. Solana uses a unique proof-of-history (PoH) consensus mechanism in combination with a proof-of-stake (PoS) consensus mechanism called Tower BFT. This allows Solana to achieve high throughput and scalability without compromising on security.

In summary, Ethereum is a well-established platform with a large developer community and a focus on decentralization, while Solana is a newer platform that prioritizes speed and scalability. Both platforms have their strengths and weaknesses, and the choice between Ethereum and Solana would depend on the specific requirements of a project or application.



What is the formula for linear regression?



The formula for simple linear regression is:

$$y = mx + b$$

Where:

- y is the dependent variable (the variable we are trying to predict)
- x is the independent variable (the variable we are using to make predictions)
- m is the slope of the regression line
- b is the y-intercept of the regression line

The formula for calculating the slope (m) of the regression line is:

$$m = \frac{\sum((x - \bar{x})(y - \bar{y}))}{\sum(x - \bar{x})^2}$$

Where:

- Σ represents the sum of all values
- \bar{x} is the mean of the independent variable x
- \bar{y} is the mean of the dependent variable y

The formula for calculating the y-intercept (b) of the regression line is:

$$b = \bar{y} - m * \bar{x}$$

Once you have calculated the slope and y-intercept, you can use the formula $y = mx + b$ to make predictions for the dependent variable y based on the independent variable x .

Key Observations

- Ease of integration
- Free Development Experience
- API Design - supports streaming responses

Key Challenges

- Challenges to update Streamlit to the latest version.
- Required the latest version of Python. Created a new python environment with the latest python 3.11



02 →

Create A Vector Datastore to retrieve the data

an application that not only extracts text from PDFs but also intelligently categorizes and stores this information in a vector datastore, allows for quick and precise retrieval of information based on user queries

(AI)

Let's dive into the technical side

- 1 We use the 'PyPDFDirectoryLoader' for loading and processing documents from a designated folder, ensuring all PDFs are accounted for.
- 2 After extraction, the text is split and embedded into a high-dimensional space using textsplitter and embedding techniques
- 3 These embeddings are then stored in a FAISS vector datastore, which is known for its efficiency in handling large-scale similarity search.
- 4 Use the Vector Store to retrieve information by creating a retrieval tool and using a React Agent
- 5 Streamlit framework to provide a user-friendly web interface


Demo of the Application

Deploy ⋮

Hi, I am *HODAML*, ask me something about the pdf

How can I help you today?

Press Enter to apply



Subject: Status Report on Current Situation

Dear Management,

I hope this message finds you well. I am writing to provide you with a status report on our current situation, highlighting the progress made, challenges faced, and the next steps to be taken.

Current Status:

- We have successfully completed the initial phase of the project, including research and planning.
- The team has been working diligently to meet project deadlines and deliverables.
- Key stakeholders have been engaged and updated on the project's progress.



Key Findings

- Efficiency in Information Retrieval
- Versatility - ability to fit in multiple industries
- Accessibility - ReAct prompt consists few-shot task-solving trajectories, with human-written text reasoning traces and actions

Key Challenges

- Dependency on Quality of Data
 - Complexity in Maintenance
-

03 →

Create Tools for an AI Agent to execute tasks

(AI)

Live Demo

Crafting Text Message & Sending Text

Hi, I am *HODAML*, a FAISS RAG Copilot

How can I help you today?

The dinosaur is T-Rex. The temperature is 90F. Craft a text message with actions taken to keep the Dino s

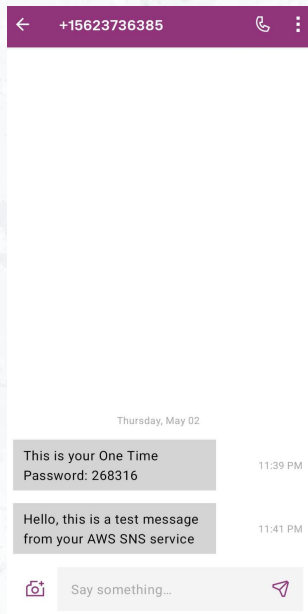


When the temperature for T-Rex exceeds 68°F, immediate action is required to keep the dinosaur safe. Here are the steps to take:

1. Check the temperature reading from the TYR55 sensor in the cargo bay.
2. If the temperature is above 68°F, notify the Associate Director of Operations and Strategy.
3. Work with the transporter to move the T-Rex to a safe area with a temperature between 57°F and 68°F.
4. Record the temperature and conditions in the new location.
5. Ensure the T-Rex is moved back into the freighter once the temperature is within the safe range.

These actions will help ensure the safety of the T-Rex when the temperature is outside the safe range.

☐ Show Document Sources

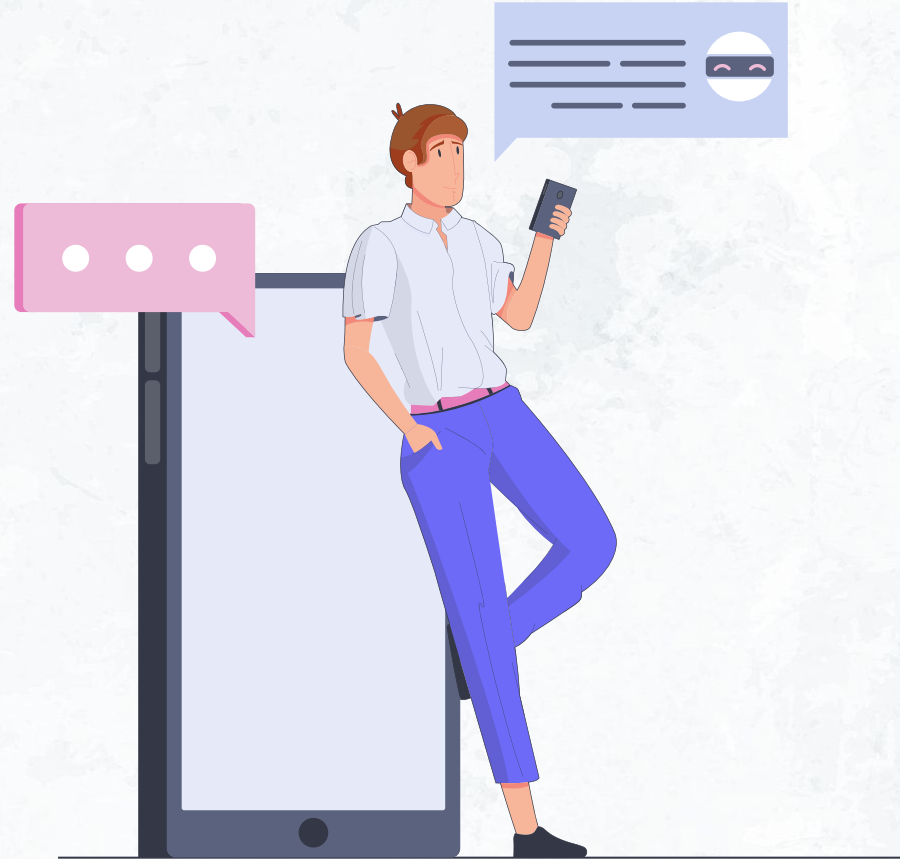


Key Observations

- Smooth Integration of DynamoDB & SQL Database
- Creating custom functions to retrieve data: City, DynoID & Dino name
- Setting up the string value to 'transactional' instead of 'promotional'

Key Challenges

- API key handling
- AWS SNS setup
- Accessible phone number recipient



04 →

**Create a program that
executes the end-to-end
workflow**

(AI)

Live Demo

Thank You!

