TASK 7P

STUDENT NAME: M. Rachel

STUDENT ID: 224234147

STUDENT MAIL ID: rachelriney1608@gmail.com

PART -1

1)What Is Azure OpenAI?

Azure OpenAI is a cloud-based artificial intelligence (AI) platform provided by Microsoft that allows developers to build and deploy custom AI models using OpenAI's machine learning technology. The platform provides a range of tools and services for training, deploying, and managing AI models, including pre-built models for natural language processing, computer vision, and other AI tasks.

- It is enables to develop powerful AI solutions that benefit from the security, scalability and integration of other services provided by azure cloud platform.
- We won't do it directly with Open AI because there is not security, scalability and integration, but we can get in azure so we use with Microsoft Open AI.

2) What is Tokenizer?

- A tokenizer is a component of natural language processing (NLP) that breaks down text into smaller units called tokens. Tokens can be words, phrases, or other units of meaning, depending on the specific use case.
- There are different types of tokenizers, each with its own approach to breaking down text. Some tokenizers split text into words based on whitespace, while others use regular expressions or other rules to identify tokens.
- To use the Azure OpenAI tokenizer, users can make API calls to the Azure OpenAI service, passing in the text they want to tokenize. The tokenizer will then return a list of tokens, along with metadata such as the token start and end positions in the original text.

3) What is few-shot and Zero-shot Learning and their advantages on LLMs and prompting?

Few-shot and zero-shot learning are technique used in NLP models, especially LLM models like GPT4 that enable models to make predictions or classify data points with limited or no training data, they allow the models to generalize to new tasks and domains without requiring extensive retraining or fine tuning.

- Zero-shot learning is a technique, is quite powerful it involves making predictions or classifications for completely new, unseen categories or tasks.

- Few-shot learning is a technique, it explores how to efficiently communicate with language models to elicit the desired output, it defines where a model is trained on a small number of examples (typically on the order of a few to a few dozen) and then asked to generalize to new, related tasks.

Advantages:

1. They enable the models to generalize to new tasks and domains without requiring extensive retraining. This is particularly useful for LLMs, which can be computationally expensive to train and may not have access to large amounts of labeled data for every new task.
2. To make more accurate predictions and classifications for rare or unseen categories. By learning a set of features or attributes that can be applied to new categories, LLMs can make more informed predictions and avoid overfitting to the training data.

4) What is the difference between System Prompt and meta prompt? Provide an example.

- System prompting involves providing the large language model with a system message at the beginning of the prompt.The system message is used to describe the assistance personality to define what the model should and should'nt answer and define the format of the model responses. It is a way to guide AI system behaviour and performance.
- Meta prompting is more advanced technique the uses meta learning Algorithms to find better prompt initializations for fast adaptations to new tasks. It is useful for soft prompting methods for learnable vectors, such as pseudo prompt tokens to achieve better performance.
- The key difference between system prompts and meta prompts is that system prompts are fixed and do not change, while meta prompts are dynamically generated based on the specific task or context. This allows meta prompts to be more flexible and adaptable to a wider range of tasks and domains.

Example:

to generate a recipe for chocolate chip cookies, the system prompt might be:

"You are a helpful assistant. Please generate a recipe for chocolate chip cookies."

While the meta prompt might be:

"You are a helpful assistant. Please generate a recipe for chocolate chip cookies. Here are some ingredients you can use: flour, sugar, butter, eggs, chocolate chips, baking soda, salt."

By including the list of ingredients in the meta prompt, we provide the LLM with additional context that can help it generate a more accurate and useful recipe.

5) Explain generate the code with Azure OpenAI service? What's the advantage of using this service?

Generating the code with Azure OpenAI:

- The first step in generating code with Azure OpenAI Service is to define generating the code. This involves writing a natural language prompt that describes the code to be generated. The prompt should be clear, concise, and detailed, and should include any necessary context or requirements for the code.
- **Send the Prompt to Azure OpenAI Service:** Once you have defined the code generation task, you can send the prompt to Azure OpenAI Service using the OpenAI API. The LLM will receive the prompt and generate code based on its training and the information provided in the prompt.
- **Review and Refine the Generated Code :** After Azure OpenAI Service generates the code, you should review it carefully to ensure that it meets your requirements and is free of errors. If necessary, you can refine the prompt and resend it to Azure OpenAI Service to generate a revised version of the code.
- **Test and Deploy the Generated Code:** Once you are satisfied with the generated code, you can test it to ensure that it works as expected. If the code passes testing, you can deploy it in your application or project.

Advantages:

- Automated Code Generation
- Versatility
- Customization
- Scalability
- Cost Effective

6) What is DALL-E? explain it in the context of Azure OpenAI services?

- DALL-E is a machine learning model developed by OpenAI that can generate images from textual descriptions. It is a type of text-to-image model that uses deep learning techniques to understand the relationship between words and images. DALL-E can generate a wide variety of images, including objects, scenes, and even abstract concepts.
- In the context of Azure OpenAI services, DALL-E is not currently available as a standalone service. However, Azure Cognitive Services, which is a collection of AI services provided by Microsoft, includes a similar service called the Computer Vision API. The Computer Vision API can analyze and identify images, and it also includes features for generating images from textual descriptions.
- Both DALL-E and the Computer Vision API demonstrate the potential of machine learning to transform the way we interact with images and text.

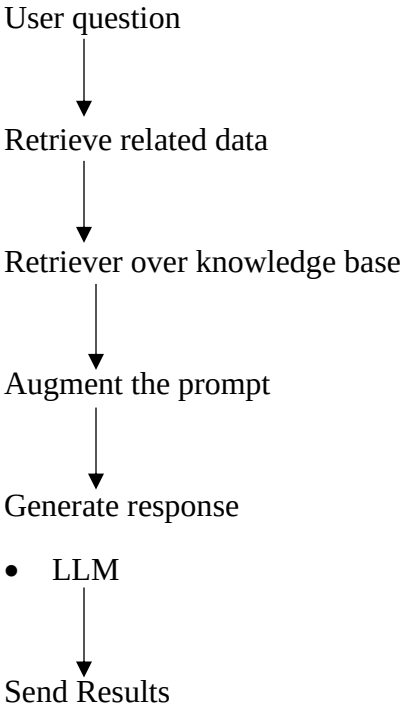
7) What is RAG? Summarize your understanding of your understanding from the Lecture.

Retrivial Augumented Generation (RAG) it allows developers to use the supported AI chat models that can reference specific sources of information to ground the response, adding the information allows the model to reference both the specific data provided and its pre-trained

knowledge to provide more effective responses. Azure openAI enables RAG by connecting pretrained models to your own data sources.

- The RAG approach has several benefits over traditional retrieval-based or generative models. By using a retrieval-based model to find relevant information, RAG can generate more accurate and relevant responses than a traditional generative model. By using a generative model to create a response, RAG can generate more creative and natural-sounding responses than a traditional retrieval-based model.

Workflow:



8) What is the Azure AI Search Hybrid Retrieval? Explain Vector Embedding.

Azure AI Search Hybrid Retrieval is a feature of Azure Cognitive Search that combines the strengths of two search techniques: keyword search and semantic search. It uses a hybrid approach to search that combines the speed and precision of keyword search with the relevance and accuracy of semantic search.

- Keyword search is a traditional search technique that matches user queries to exact keywords in a document. It is fast and precise, but it can miss relevant documents that do not contain the exact keywords used in the query.
- Semantic search, on the other hand, uses natural language processing (NLP) techniques to understand the meaning and context of user queries and documents. It can find relevant documents even if they do not contain the exact keywords used in the query, but it can be slower and less precise than keyword search.

Vector Embedding:

The vector embedding used in Azure AI Search Hybrid Retrieval is based on a pre-trained language model that has been fine-tuned on a specific domain, it is a way of representing words, phrases, or documents as vectors in a high-dimensional space. Each vector captures the semantic meaning and context of the word, phrase, or document. By representing documents as vectors, Azure AI Search Hybrid Retrieval can use vector similarity measures to find documents that are semantically similar to the user's query, even if they do not contain the exact keywords used in the query.

9) Explain the fundamentals of Responsible GenAI?

Responsible GenAI refers to the principles and practices that ensure the ethical and responsible development, deployment, and use of generative artificial intelligence (GenAI) systems, the models that can generate new content, such as text, images, audio, or video, based on patterns learned from large datasets.

Transparency:

GenAI systems should be transparent and explainable, so that users can understand how they work and how they make decisions. Transparency can help build trust and confidence in GenAI systems, and reduce the risk of bias, errors, or misuse.

Fairness:

GenAI systems should be designed and tested to ensure that they do not discriminate or reinforce harmful biases based on factors such as race, gender, age, or ability. This requires careful consideration of the data used to train the models, as well as ongoing monitoring and evaluation to detect and address any biases or disparities.

Privacy:

GenAI systems should respect users' privacy and data protection rights, by collecting, processing, and storing data in a secure and ethical manner. This includes obtaining informed consent from users, minimizing data collection and retention, and implementing robust security measures to prevent unauthorized access or use.

Accountability:

GenAI systems should be accountable and responsible for their actions and outcomes, by providing clear lines of responsibility and oversight, and by establishing mechanisms for redress and remediation in case of harm or misuse.

Social and ethical values:

GenAI systems should be designed and used in a way that aligns with social and ethical values, such as human rights, democracy, and sustainability. This requires careful consideration of the potential impacts and consequences of GenAI systems, as well as ongoing engagement and dialogue with stakeholders, including users, experts, and policymakers.

PART-2

Find an advanced intelligent system and provide a comprehensive overview of the system. You need the to discuss the problem and why do we need to use the system to solve the proposed problem

AUTOMATED VEHICLE(AV)

One real-world advanced intelligent system that has gained significant attention is the autonomous vehicle (AV). AVs are self-driving cars that use sensors, cameras, and artificial intelligence (AI) algorithms to navigate roads and make decisions without human intervention.

According to the World Health Organization, road accidents are the leading cause of death among young people aged 15-29, and result in over 1.35 million deaths worldwide each year. Human error is responsible for up to 90% of these accidents, making AVs a promising solution to reduce the number of accidents and save lives.

They use a combination of sensors, such as cameras, lidar, and radar, to perceive their environment and detect obstacles, pedestrians, and other vehicles. They also use AI algorithms, such as deep learning and reinforcement learning, to analyze the data from the sensors and make decisions based on the current situation. For example, AVs can detect a pedestrian crossing the road and apply the brakes to avoid a collision.

They can also improve traffic flow and reduce congestion by optimizing driving behavior and reducing the need for human intervention. They can communicate with other AVs and infrastructure, such as traffic lights and road signs, to coordinate their movements and reduce delays. AVs can also operate 24/7, without the need for breaks or rest, which can increase the efficiency of transportation and reduce the need for parking spaces.

AVs also pose several challenges and risks, such as the potential for software bugs, cyber attacks, and ethical dilemmas. For example, AVs may need to make decisions in situations where there is a trade-off between safety and other considerations, such as minimizing damage to property or avoiding harm to pedestrians. AVs may also face challenges in adapting to different driving conditions, such as snow, fog, or construction zones.

To address these challenges, AV developers and policymakers have proposed several solutions, such as:

Robust testing and validation: AVs should undergo extensive testing and validation to ensure their safety and reliability. This includes simulation testing, on-road testing, and third-party testing and certification.

Cybersecurity measures: AVs should implement robust cybersecurity measures to prevent hacking and unauthorized access. This includes encryption, authentication, and intrusion detection systems.

Ethical guidelines: AVs should follow ethical guidelines that prioritize safety, transparency, and accountability. This includes guidelines for decision-making in ethical dilemmas, such as the trolley problem.

Regulatory frameworks: AVs should operate under clear regulatory frameworks that define their rights and responsibilities. This includes regulations for licensing, insurance, and liability.

Overall, AVs represent a promising solution to improve transportation safety and efficiency, but they require careful design, testing, and regulation to ensure their safe and responsible deployment.

2) You need to explain the details of the system (the diagram or flowchart) and discuss how the system is working, the input data, the AI techniques, how AI services are working together as a system.

Flowchart of automated vehicle:

Sensors: AVs use various sensors to perceive their environment, such as cameras, lidar, radar, and ultrasonic sensors. These sensors provide data about the vehicle's surroundings, such as the position, velocity, and acceleration of other vehicles, pedestrians, and obstacles.

Perception: The perception module processes the sensor data using AI techniques, such as computer vision, object detection, and segmentation. The module identifies and classifies objects, such as cars, pedestrians, and traffic signs, and estimates their properties, such as position, velocity, and orientation.

Localization: The localization module determines the vehicle's position and orientation in the environment using AI techniques, such as simultaneous localization and mapping (SLAM) and global positioning system (GPS). The module uses the perception data and a map of the environment to estimate the vehicle's position and orientation.

Planning: The planning module generates a plan of actions based on the localization and perception data. The module uses AI techniques, such as motion planning and decision-making, to select a safe and efficient path that avoids obstacles and complies with traffic rules.

Control: The control module executes the plan of actions using AI techniques, such as feedback control and trajectory tracking. The module sends commands to the actuators, such as the steering wheel, throttle, and brakes, to follow the selected path.

Communication: The communication module enables the AV to communicate with other vehicles, infrastructure, and users. The module uses AI techniques, such as natural language processing and machine learning, to interpret and generate messages

The AV system uses a combination of AI techniques to process the sensor data and generate a plan of actions. The planning module may use decision trees or rule-based systems to select a safe and efficient path based on the perception and localization data. The control module may use feedback control algorithms, such as proportional-integral-derivative (PID) controllers, to track the selected path and adjust the vehicle's speed and direction.

REFERENCES:

1) azure.microsoft.com. (n.d.). Azure OpenAI Service – Advanced Language Models | Microsoft Azure. [online] Available at: https://azure.microsoft.com/en-us/products/ai-services/openai-service/.
2) Medium. (n.d.). Medium. [online] Available at: https://towardsdatascience.com/tokenization-breaking-down-text-into-words-and-phrases-a66e7593a16e [Accessed 5 May 2024].
3) Ribeiro, M.T., Wu, T., Guestrin, C. and Singh, S. (2020). Beyond Accuracy: Behavioral Testing of NLP Models with CheckList. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. doi:https://doi.org/10.18653/v1/2020.acl-main.442.
4) azure.microsoft.com. (n.d.). Azure OpenAI Service - Pricing | Microsoft Azure. [online] Available at: https://azure.microsoft.com/en-us/pricing/details/cognitive-services/openai-service/.
5) Yahnoosh (2023). Full text search - Azure AI Search. [online] learn.microsoft.com. Available at: https://learn.microsoft.com/en-us/azure/search/search-lucene-query-architecture#hybrid-search [Accessed 5 May 2024].
6) European Commission (2021). Ethics guidelines for trustworthy AI | Shaping Europe's digital future. [online] European Commission. Available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.
7) NHTSA (2021). Automated Vehicles for Safety | NHTSA. [online] www.nhtsa.gov. Available at: https://www.nhtsa.gov/vehicle-safety/automated-vehicles-safety.
8) wayback.archive-it.org. (n.d.). Archive-It Wayback Machine. [online] Available at: https://wayback.archive-it.org/12090/.
9) SA Main Site. (n.d.). The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. [online] Available at: https://standards.ieee.org/industry-connections/ec/autonomous-systems/.