Shashank Gupta

Data Science Enthusiast
shashank96gpt@gmail.com | +918233718798

EDUCATION

JECRC College
B.tech in Information Technology
2014-2018
CGPA: 7.6/10

ADVANCED TECHNOLOGY

TECHNICAL SKILLS

Programming: Python
Machine Learning & NLP: ML, Deep Learning, Advanced NLP
LLM Models: GPT-4, LLAMA2, gpt3.5-turbo, ChatGPT, Claude, Bard, Mistral
Prompt Engineering: Zero Shot Prompt, Few Shot Prompt, Chain of Thoughts, Tree of Thoughts, Forest of Thoughts, React Prompting, Multimodal COT prompting, Graph Prompting
LLM Fine-tuning: LLAMA2, gpt3.5-turbo, GPT-3
Optimization: Hallucination Optimization using DPO (Direct Performance Optimization)
AI Frameworks: LangChain, Semantic Kernel, Guidance.ai, Neo4j
Prompt Frameworks: CO-STAR, RACE, STAR, CLEAR, SWOT, AUTOMAT

LIBRARIES AND FRAMEWORKS

- LangChain
- Semantic Kernel
- Gensim
- Pandas
- Semantic Router
- Llama
- Neo4J
- Seaborn
- Guidance.ai
- Elastic Search

CERTIFICATES

- AIMLRL (End to End Course from Statistics to ML to DL by Jaskirat Singh)
- Microsoft Azure AZ-100
- Microsoft Azure AI-102
- Microsoft Azure DP-100
- DataCamp (- Python Data Science Tool Box, - Intermediate Python, -Data Manipulation with Pandas)

PERSONAL STATEMENT

Experienced AI Engineer with expertise in AI, OpenAI, machine learning, data science, and Python programming. Seeking opportunities to utilize technical skills in developing and deploying AI solutions. Proficient in OpenAI technologies, prompt optimization, prompt engineering, and LLM models. Passionate about leveraging these tools to solve complex problems and drive innovation in artificial intelligence.

EXPERIENCE

| Senior Data Scientist | Sept 2021 - Present

CELEBAL TECHNOLOGIES

PROJECTS

| Mar 2024 | Role: Senior Data Scientist

Advanced Knowledge Mining System

Description-Leveraged LLM and Python for developing an Advanced Knowledge Mining Search Engine, planned to efficiently retrieve answers from user-inputted data, facilitating seamless information access.

Roles and Responsibilities:

- Performed robust data extraction mechanisms from diverse file types including PDFs, Word documents, Excel spreadsheets, PowerPoint presentations, videos, audios, and images. Employed entity recognition, summarization, and question-answering techniques utilizing Azure OpenAI LLM.
- Applied prompt engineering methodologies such as CoT and ReACT, along with frameworks like CO-STAR, to ensure accuracy and efficiency.
- Performed Retrieval Augmented Generation Techniques to elevate the quality of generated responses and query recommendations within the KM system, leveraging Azure AI Vector Search. Contributed to the development and enhancement of LangChain LLM Agents, leveraging advanced Natural Language Processing (NLP) techniques to augment language understanding and generation capabilities.
- Ensured the accuracy of generated responses by leveraging the RAGAS Library, thereby guaranteeing high-quality and reliable answers for users.
- Architected a multi-level admin hierarchy for efficient management. Utilized Evaluation Metrics for improving response accuracy by 95%
- Enacted security measures including Azure Defender and JWT token for enhanced data protection and user authentication, ensuring robust security standards are met. Integrated security measures, ensuring 99.9% system uptime and data protection.

Prompt Portal

| Jan 2024 | Role: Senior Data Scientist

Description - Led the Prompt Portal Project, introducing a comprehensive platform for streamlined content management and personalized user experiences, aimed at revolutionizing prompt engineering and maximizing AI utilization using LLM's and Python.

Roles and Responsibilities:

- Cooperated closely with stakeholders to meticulously gather detailed functional and technical requirements for the prompt platform.
- Integrated OpenAI GPT models to deliver context-aware answers and conduct prompt testing, thereby amplifying platform capabilities. Capitalized on Azure Cognitive Search for flexible prompt testing on user-added documents.
- Elevated user engagement by implementing interactive features such as commenting, rating and feedback. Employed CosmosDB for efficient storage and rapid data retrieval, increasing user engagement by 40%.
- Engineered a robust categories management system and optimized workflow through version control mechanisms. Made the most of Azure SQL DB for meticulous log storage.
- Pioneered hyper-personalization Key Performance Indicators (KPIs) to tailor user experiences. Introduced prompt sharing functionalities to foster coordination among users.
- Managed a dynamic notification and announcement creation system leveraging Azure PubSub, ensuring users and administrators remained abreast of updates. Formulated a dedicated section for users to share success stories, fostering a sense of community and celebrating prompt creation achievements.
- Developed a leaderboard and gamification system, boosting user interaction by 25%. Established an approval pipeline facilitated by category owners for prompt validation. Facilitated the creation of bots from top-performing prompts for universal utilization, thereby maximizing platform efficacy.

VECTOR DATABASES

- Azure AI Vector Search
- Redis
- FAISS
- ChromaDB
- Pinecone
- PGVector

NLP TOOLS

- Semantic Chunking
- Agentic Chunking
- MapReduce
- Long-Context Reorder
- MultiVector Retriever
- Parent Document Retriever

TOOLS

- Cognitive Services
- Git & GitHub
- Azure OpenAI
- Azure Form Recognizer
- Azure DevOps
- Azure ML
- Azure App Service
- Azure Kubernetes
- Azure Speech Services
- Azure AI Vision
- Azure Functions
- Azure Logic Apps
- Azure CosmosDB
- Azure Data Factory
- MongoDB
- Azure SQL Server
- Azure Synapse

TeraDataSQL Query Generator | Dec 2023 | Role: Senior Data Scientist

Description-Led the development of a TeraDataSQL Query Generator, utilizing NLP techniques, Python and advanced search technologies for efficient query crafting.

Roles and Responsibilities:

- Executed data extraction pipelines to extract information from excel files, ensuring smooth data flow. Updated indexers post data embedding to improve search capabilities by incorporating relevant fields. Enabled user interaction through natural language queries, facilitating intuitive search processes.
- Build a dynamic TeraDataSQL queries based on user inputs and selected tables, utilizing LLM models and vector search techniques, reducing query generation time by 95%
- Promoted collaboration by enabling users to share queries and responses, fostering knowledge sharing among stakeholders.
- Automated the generation of missing column descriptions, enhancing metadata completeness and system usability, improving metadata completeness by 60%.. Devised administrative features to track user queries, manage accounts, and enforce secure access controls.
- Used search history storage and presentation features using Azure CosmosDB, enabling seamless retrieval of past searches for users.

AI-Powered BRD ASSISTANT | Nov 2023 | Role: NLP Engineer

Description- Orchestrated the development of the AI-Powered BRD Assistant using LLM and Python, aimed at streamlining BRDs creation for government and public services reducing BRD creation time by 20%.

Roles and Responsibilities:

- Proactively engaged with clients to gather detailed functional and technical requirements, ensuring alignment with organizational goals.
- Designed a scalable and secure architecture plan for the BRD Assistant, leveraging Azure services for optimal performance.
- Harnessed OpenAI Large Language Models (LLM) for automatic context creation and customization, employing advanced prompt engineering techniques. Carried out BRD summarization and Q&A functionalities using AI vector search, enabling users to extract key insights and address queries efficiently.
- Managed the creation and deployment of Docker images to Azure Container Registry, streamlining containerized deployment processes, reducing deployment time by 30%.
- Operationalized the solution on Azure Container Instances, integrating continuous integration and continuous deployment (CI/CD) methods for seamless deployment and maintenance.

RESEARCH PAPERS

Development of IoT for Smart Agriculture | 2018

Description: Explores the application of IoT in agriculture using sensors, actuators, and embedded microcontrollers.

- Source: ICETEAS, Springer Publication
- Link: https://link.springer.com/chapter/10.1007/978-981-13-2285-3_50

Improve UIQI and PSNR of Noised Colored Images using DWTT Filter| 2019

Description: Investigates the use of Discrete Wavelet Transform and Thresholding technique to remove gaussian noise from digital images.

- Source: GUCON, IEEE Publication
- Link: https://ieeexplore.ieee.org/document/8674928