Shreyas Puttaraju

Chicago,IL | 224-389-0281 | shreyas2797@gmail.com | Linkedin | Portfolio

EXECUTIVE SUMMARY

- Engineered three advanced generative AI applications, including AI agents, utilizing Langchain and AWS for scalable AI solutions.
- Expert at writing high-performing prompts for AI systems, ensuring clarity, consistency, and optimized task performance, leading to enhanced LLM response and accuracy.
- Strong foundation in data science and analytics, demonstrated by comprehensive experience in ETL processes, data analysis, visualization, and predictive modeling.
- Led the end-to-end development of generative AI POCs and document intelligence tools using Azure OpenAI, LangChain, and cloud-native architectures to automate financial workflows and scale enterprise AI adoption.

WORK EXPERIENCE

Jewelers Mutual Group

AI Applications Engineer

- Spearheaded the rapid development of four generative AI proof-of-concept (POC) applications, leveraging LLMs from Azure OpenAI to enhance automation and decision-making workflows.
- Engineered both internal and external-facing generative AI tools to optimize business operations, increasing productivity and significantly reducing manual effort across multiple teams.
- Developed robust evaluation frameworks to assess the effectiveness and reliability of generative AI applications, implementing benchmarking strategies to ensure consistent performance.
- Applied advanced prompt engineering techniques to fine-tune LLM outputs, optimizing accuracy, coherence, and consistency in financial document processing tasks such as extraction, summarization, and analysis.
- Integrated AI-driven document processing solutions using LangChain, Python, and cloud-native architectures, streamlining the extraction and analysis of financial data at scale.

PeritusHub (Stealth Start-Up)

AI Engineer

- Conceptualized and implemented a scalable multi-agent system, pioneering an agentic design approach using Langchain and AWS Bedrock to solve complex, multidimensional conversational AI tasks.
- Integrated a supervisory mechanism for AI agents to adopt various personas, ensuring the optimal agent is engaged for each task segment.
- Designed system prompts for 7+ agents using advanced prompt engineering techniques, ensuring consistent agent response and actions across the system.
- Built 6 custom tools for AI agents using Python, adding new functionalities and increasing agents' performance.
- Architected the ETL pipeline for the multi-agent system using AWS, ensuring efficient data integration and management to support complex AI functionalities and system scalability.
- Constructed and managed the CI/CD pipeline using GitHub Actions, ensuring streamlined deployments and continuous integration of code changes across projects.
- Engineered an advanced Document Management System incorporating RAG with OpenAI LLMs for enhanced information retrieval and processing.
- Pioneered a Candidate Scoring System leveraging LLMs and AWS Lambda for serverless computing.

H-E-B Groceries

Data Scientist Intern

- Engineered a comprehensive system for extracting product entities and attributes from 80,000 product images.
- Utilized Azure OpenAI LLM combined with advanced computer vision techniques for precision and accurate entity extraction.
- Achieved heightened attribute detection and extraction capabilities by integrating Paddle OCR and GPT-3.5-Turbo within the LangChain framework, resulting in enhanced performance.
- Formulated consistent and clear prompts for LLMs to enhance task comprehension and efficiency.
- This innovation boosted product attribute management efficiency and customer engagement by up to 8%.

Austin, TX(Remote)

San Antonio, TX

Neenah, WI(Remote)

- Crafted a web-based user interface for the extraction system utilizing HTML, JavaScript, and Flask, improving accessibility and ease of use for end users.
- Partnered with MLOps and Software Engineering teams to seamlessly integrate the new extraction system into the existing infrastructure, exemplifying strong teamwork and technical coordination skills.

VOLUNTEERING EXPERIENCE

Change The Present Organisation

Volunteer AI Engineer Intern

- Automated social media content creation for over 1400 gifts using advanced Prompt Engineering with OpenAI's LLMs and Python scripts, enhancing the efficiency and appeal of marketing campaigns.
- Built content creation pipeline for LinkedIn, Pinterest, Instagram, and Facebook to streamline multi-platform publishing.

EDUCATION

Master's in Data Science

DePaul University

Bachelor's in Computer Science and Engineering

P.E.S. College Of Engineering

SKILLS

Generative AI:	Large Language Models, Langchain, HuggingFace, GANs, AI Agents, Retrieval Augmented Generation (RAG), Conversational AI, Prompt Engineering
Programming Languages:	Python (with PyTorch and Scikit-Learn), R
Data Processing & ETL:	SQL, Data Cleaning, Data Wrangling, Data Preprocessing, ETL Pipeline Design
Statistical Techniques:	Statistical Analysis, Hypothesis Testing, A/B Testing, Predictive Modeling
Visualization Tools:	Matplotlib, Tableau, and Power BI
Machine Learning:	Supervised and Unsupervised Learning(Regression Analysis, Clustering, Decision Trees, Random Forest), Ensemble Methods, XGBoost, Neural Networks(CNN, RNN, LSTM)
Cloud Platforms and CI/CD:	AWS, Microsoft Azure, Docker, Kubernetes, CI/CD Pipelines
Other Relevant Skills:	Project Management, Agile Methodologies, Cross-functional Collaboration, Strategic Planning, Effective Communication

New York, NY(Remote)

Chicago, IL

India