

HEMANTH P

SENIOR AI ENGINEER | MACHINE LEARNING ENGINEER | GENERATIVE AI & AGENTIC AI

840-877-5892 | hemanthreddypolu@gmail.com

PROFESSIONAL SUMMARY

- Senior AI Engineer with **12+ years of overall IT experience**, including **10+ years of hands-on experience in Artificial Intelligence, Machine Learning, Generative AI, NLP, and intelligent application development**.
 - Extensive experience designing and developing **AI/ML solutions, predictive models, NLP applications, deep learning models, machine learning pipelines, and production-ready AI platforms** across financial services, healthcare, logistics, and telecommunications domains.
 - Strong expertise in **Generative AI and Large Language Models (LLMs)**, including **RAG, prompt engineering, embeddings, vector search, semantic retrieval, LLM integration, and enterprise AI application development**.
 - Hands-on experience building **Agentic AI and Copilot platforms**, including AI agents, tool/function calling, workflow orchestration, multi-step AI workflows, enterprise knowledge retrieval, conversational AI, and LLM-powered business solutions.
 - Proficient in **Python, LangChain, LangGraph, Azure OpenAI, AWS Bedrock, Hugging Face, TensorFlow, PyTorch, Scikit-learn, MLflow, vector databases, and modern AI/ML frameworks and platforms**.
 - Experienced in developing and operationalizing **machine learning pipelines and MLOps solutions**, including model training, validation, evaluation, deployment, monitoring, containerization, CI/CD, Docker, Kubernetes, and cloud-based AI/ML services.
 - Strong software engineering foundation with hands-on experience in **Java, SQL, Spring, REST APIs, enterprise application development, relational databases, distributed systems, and cloud technologies**, supporting scalable integration of AI capabilities into enterprise applications.
 - Experienced in collaborating with **data scientists, software engineers, architects, product teams, and business stakeholders** to translate complex business requirements into scalable, secure, reliable, and production-ready AI/ML solutions.
 - Strong understanding of the complete **AI/ML lifecycle**, including data preparation, feature engineering, model development, experimentation, evaluation, deployment, optimization, monitoring, and production support.
 - Proven ability to work on modern **enterprise AI transformation initiatives**, integrating traditional machine learning with **Generative AI, LLMs, RAG, Agentic AI, and Copilot experiences** to improve automation, knowledge discovery, decision support, and business processes.
-

TECHNICAL SKILLS

Category	Technologies
Programming Languages	Python, Java, SQL, Bash/Shell Scripting

Artificial Intelligence	Artificial Intelligence, Generative AI, Agentic AI, AI Agents, Intelligent Automation, Conversational AI, AI-Assisted Applications
Machine Learning	Supervised Learning, Unsupervised Learning, Classification, Regression, Clustering, Feature Engineering, Model Training, Model Evaluation, Predictive Analytics
Deep Learning & NLP	Deep Learning, Neural Networks, Natural Language Processing (NLP), Text Classification, Text Embeddings, Transformers, Sentiment Analysis, Named Entity Recognition
Generative AI & LLMs	Large Language Models (LLMs), Generative AI, Prompt Engineering, LLM Integration, Embeddings, Context Engineering, Function/Tool Calling, AI Orchestration
Agentic AI	AI Agents, Agent Workflows, Multi-Agent Patterns, LangGraph, LangChain, Tool Calling, Function Calling, Workflow Orchestration, Agent Memory, Enterprise Copilots
RAG & Knowledge Systems	Retrieval-Augmented Generation (RAG), Semantic Search, Vector Search, Document Retrieval, Chunking, Embeddings, Knowledge Bases, Hybrid Search
AI/ML Frameworks	Scikit-learn, TensorFlow, PyTorch, Hugging Face, Pandas, NumPy, SciPy
LLM / AI Platforms	Azure OpenAI, AWS Bedrock, OpenAI Models, Hugging Face Transformers, Enterprise LLM Platforms
Vector Databases	Pinecone, FAISS, ChromaDB, Azure AI Search, Vector Search
MLOps	MLflow, Model Deployment, Model Versioning, Model Monitoring, Experiment Tracking, ML Pipelines, CI/CD, Model Evaluation
Data Engineering & Processing	Apache Spark, PySpark, Pandas, NumPy, ETL/ELT, Data Preprocessing, Data Transformation, Data Quality
Cloud Platforms	Microsoft Azure, Amazon Web Services (AWS), Azure AI Services, AWS AI/ML Services
Containers & DevOps	Docker, Kubernetes, Git, Jenkins, GitHub Actions, CI/CD, Containerized AI/ML Applications
APIs & Integration	REST APIs, JSON, Microservices, API Integration, Web Services, Python APIs, Java APIs
Databases	PostgreSQL, MySQL, Oracle, SQL Server, NoSQL Databases
Java Technologies	Java, Spring, Spring Boot, Hibernate, J2EE, JDBC
AI Evaluation & Observability	Model Evaluation, LLM Evaluation, Response Quality Assessment, AI Application Monitoring, Model Performance Monitoring, Logging
Development Tools	Jupyter Notebook, VS Code, Git, GitHub, Postman
Operating Systems	Linux, Unix, Windows

PROFESSIONAL EXPERIENCE

Client: MasterCard - St. Louis, MO

Role: Senior AI Engineer – Agentic AI & Copilot Platform

Duration: Jan 2024 – Present

Project Description: Worked on an enterprise **AI and Copilot platform** focused on developing Generative AI and Agentic AI capabilities for intelligent knowledge retrieval, conversational assistance, workflow automation, and integration with enterprise applications. The project involved building LLM-powered applications, RAG pipelines, AI agents, tool-enabled workflows, and scalable AI services with emphasis on reliability, security, evaluation, and production deployment.

Environment: Python, Generative AI, Agentic AI, Large Language Models (LLMs), RAG, LangChain, LangGraph, Azure OpenAI, AWS Bedrock, Hugging Face, Prompt Engineering, Context Engineering, AI Agents, Tool Calling, Function Calling, Pinecone, FAISS, ChromaDB, Azure AI Search, REST APIs, Microservices, Docker, Kubernetes, MLflow, Git, CI/CD, Jupyter, Linux

Responsibilities

- Designed and developed enterprise **Generative AI and Agentic AI solutions** using Python, LLMs, RAG, and AI orchestration frameworks for intelligent automation, knowledge discovery, and Copilot applications.
- Developed **AI agent workflows** using LangChain and LangGraph, implementing tool calling, function calling, workflow orchestration, agent memory, and multi-step task execution patterns.
- Built end-to-end **Retrieval-Augmented Generation (RAG)** pipelines covering document ingestion, preprocessing, chunking, metadata enrichment, embedding generation, vector indexing, retrieval, context construction, and grounded response generation.
- Integrated **Azure OpenAI and AWS Bedrock** capabilities into enterprise applications for LLM inference, prompt-based applications, structured responses, conversational experiences, and AI-powered workflows.
- Developed enterprise **Copilot capabilities** combining conversational interfaces, LLM reasoning, knowledge retrieval, business rules, and integrations with internal applications and approved enterprise services.
- Developed Python-based **AI services and REST APIs** to integrate LLM applications, AI agents, vector search services, enterprise data sources, and downstream business applications.
- Implemented **semantic and vector search** using Pinecone, FAISS, ChromaDB, and Azure AI Search to improve retrieval accuracy and provide relevant context to LLM applications.
- Applied **prompt engineering and context engineering** techniques to improve response quality, consistency, relevance, and grounding across different enterprise use cases.
- Implemented **tool and function calling** capabilities that allowed AI agents to interact with approved APIs, databases, and enterprise services through controlled execution workflows.
- Developed evaluation approaches for **LLM and RAG applications**, assessing response relevance, retrieval quality, grounding, hallucination risk, latency, and overall application performance.

- Built and maintained **production deployment pipelines** for AI applications using Docker, Kubernetes, Git, CI/CD, and cloud infrastructure across development, testing, and production environments.
 - Implemented application logging, monitoring, exception handling, and performance tracking to support **production reliability and troubleshooting** of LLM and Agentic AI applications.
 - Integrated AI capabilities with existing **microservices and enterprise APIs**, ensuring scalable communication between AI applications and downstream business systems.
 - Collaborated with data scientists, software engineers, product teams, security teams, and business stakeholders to translate business requirements into scalable and production-ready **AI/ML and GenAI solutions**.
 - Participated in technical design, proof-of-concept development, code reviews, testing, deployment, troubleshooting, and production support for enterprise AI initiatives.
 - Contributed to continuous improvement of AI applications by analyzing production feedback and optimizing **prompts, retrieval strategies, agent workflows, model configurations, and application performance**.
-

Client: Bank of America — Charlotte, NC

Role: Senior AI Engineer

Duration: Jan 2022 – Dec 2023

Project Description: Worked on enterprise AI initiatives focused on **Machine Learning, Generative AI, NLP, and LLM-powered applications** for intelligent automation, knowledge discovery, document processing, and decision-support use cases. Developed scalable AI services and RAG-based solutions while integrating machine learning capabilities with enterprise applications and cloud platforms.

Environment: Python, Generative AI, Machine Learning, NLP, LLMs, RAG, Transformers, Hugging Face, Azure OpenAI, AWS, LangChain, Prompt Engineering, Embeddings, Vector Search, Pinecone, FAISS, Scikit-learn, PyTorch, TensorFlow, MLflow, Apache Spark, PySpark, REST APIs, Docker, Kubernetes, Git, CI/CD, Linux

Responsibilities

- Designed and developed **AI/ML solutions** using Python for enterprise use cases involving predictive analytics, NLP, intelligent automation, and knowledge-based applications.
- Developed **LLM-powered applications** and integrated foundation models into enterprise workflows to support conversational AI, content generation, summarization, classification, and information extraction.
- Built **Retrieval-Augmented Generation (RAG)** solutions combining document processing, embeddings, vector search, retrieval pipelines, and LLM-based response generation.
- Implemented **NLP and text-processing pipelines** for document classification, entity extraction, text similarity, semantic analysis, and unstructured data processing.
- Worked with **Hugging Face Transformers, PyTorch, TensorFlow, and Scikit-learn** for developing, fine-tuning, evaluating, and deploying machine learning and NLP models.

- Integrated **Azure OpenAI and cloud-based AI services** with Python applications and REST APIs to enable enterprise Generative AI capabilities.
 - Implemented **prompt engineering techniques** and reusable prompt patterns to improve the accuracy, consistency, relevance, and quality of LLM-generated responses.
 - Developed embedding and retrieval workflows using **vector databases and search technologies**, including Pinecone and FAISS, to support semantic search and enterprise knowledge retrieval.
 - Built scalable **Python REST APIs and microservices** to expose AI/ML capabilities and integrate models, LLM applications, data services, and enterprise applications.
 - Used **MLflow and MLOps practices** for experiment tracking, model versioning, model evaluation, deployment workflows, and maintaining reproducible machine learning pipelines.
 - Containerized AI/ML applications using **Docker** and supported deployment through **Kubernetes and CI/CD pipelines**, following enterprise development and release practices.
 - Collaborated with data scientists, software engineers, business analysts, and stakeholders to translate business requirements into **production-ready AI/ML and Generative AI solutions**, including testing, troubleshooting, optimization, and production support.
-

Client: FedEx — Memphis, TN

Role: Senior Machine Learning Engineer

Duration: Jan 2019 – Dec 2021

Project Description: Worked on enterprise **Machine Learning and AI initiatives** supporting logistics, operational analytics, forecasting, customer intelligence, and process optimization. Developed machine learning and deep learning solutions for structured and unstructured data, built scalable data and ML pipelines, and supported the deployment and monitoring of models in production environments.

Environment: Python, Machine Learning, Deep Learning, NLP, Scikit-learn, TensorFlow, PyTorch, Pandas, NumPy, Apache Spark, PySpark, MLflow, AWS, Azure, Docker, Kubernetes, REST APIs, Microservices, SQL, PostgreSQL, Oracle, Git, Jenkins, CI/CD, Linux

Responsibilities

- Designed and developed **machine learning models** for classification, regression, clustering, forecasting, and predictive analytics using Python and Scikit-learn.
- Developed **deep learning solutions** using TensorFlow and PyTorch for complex prediction, classification, and unstructured data processing requirements.
- Built data preprocessing and **feature engineering pipelines** to clean, transform, normalize, and prepare large datasets for machine learning model development.
- Developed **NLP-based solutions** for processing and analyzing text data, including text classification, entity extraction, similarity analysis, and document categorization.
- Used **Pandas, NumPy, Apache Spark, and PySpark** to process and analyze large datasets and prepare training data for machine learning workflows.

- Implemented model training, validation, hyperparameter tuning, and performance evaluation using appropriate **machine learning metrics and validation techniques**.
 - Developed reusable **ML pipelines** covering data preparation, feature engineering, model training, evaluation, model packaging, and deployment.
 - Used **MLflow** for experiment tracking, model versioning, performance comparison, and maintaining reproducibility across machine learning projects.
 - Developed Python-based **REST APIs and microservices** to expose trained machine learning models and integrate predictive capabilities with enterprise applications.
 - Containerized machine learning applications using **Docker** and supported deployment of scalable ML services using Kubernetes and CI/CD practices.
 - Implemented model and application monitoring, logging, error handling, and performance analysis to support **production ML workloads and ongoing model optimization**.
 - Collaborated with data scientists, data engineers, software developers, business analysts, and stakeholders to translate business requirements into **scalable, production-ready machine learning solutions** and provide ongoing production support.
-

Client: Blue Cross Blue Shield (BCBS) — Chicago, IL

Role: AI/ML Engineer

Duration: Jul 2016 – Dec 2018

Project Description: Worked on enterprise **Artificial Intelligence and Machine Learning initiatives** focused on predictive analytics, healthcare data analysis, NLP, and intelligent automation. Developed machine learning models and data-processing pipelines to analyze structured and unstructured data, improve prediction accuracy, and support business decision-making and operational processes.

Environment: Python, Machine Learning, Artificial Intelligence, NLP, Scikit-learn, TensorFlow, Pandas, NumPy, SciPy, Apache Spark, PySpark, SQL, Oracle, PostgreSQL, REST APIs, Flask, Jupyter, Git, Jenkins, Linux

Responsibilities

- Developed and implemented **machine learning models** using Python and Scikit-learn for classification, regression, clustering, and predictive analytics use cases.
- Performed **data preprocessing, cleansing, transformation, normalization, and feature engineering** on structured and unstructured datasets to prepare data for model development.
- Built predictive models to identify patterns and trends in large datasets and support **data-driven business and operational decisions**.
- Developed **NLP solutions** for processing unstructured text, including text classification, document categorization, text similarity, and information extraction.
- Used **Pandas, NumPy, SciPy, and SQL** for exploratory data analysis, statistical analysis, data manipulation, and preparation of machine learning datasets.
- Developed and trained **neural network and deep learning models** using TensorFlow for selected classification and prediction requirements.

- Applied model evaluation and validation techniques, including **cross-validation, accuracy, precision, recall, F1-score, ROC-AUC**, and other appropriate metrics.
 - Performed **hyperparameter tuning and model optimization** to improve model performance, stability, and generalization.
 - Used **Apache Spark and PySpark** to process and analyze larger datasets and support scalable machine learning data pipelines.
 - Developed Python-based services and **REST APIs** to expose machine learning functionality and integrate predictive models with enterprise applications.
 - Created reusable **data and machine learning workflows** covering data preparation, feature engineering, model development, evaluation, and deployment support.
 - Collaborated with data scientists, software engineers, data analysts, business stakeholders, and QA teams to understand requirements, develop AI/ML solutions, validate results, troubleshoot issues, and support models in production.
-

Client: AT&T — Dallas, TX

Role: Java Developer / SQL Developer

Duration: Jan 2014 – Jun 2016

Project Description: worked on enterprise application development and database initiatives supporting telecommunications business operations. Developed and maintained Java-based applications, REST services, backend components, and SQL-driven data solutions while collaborating with business and technical teams on application enhancements, production support, and system integration.

Environment: Java, J2EE, Spring, Spring MVC, Hibernate, JDBC, SQL, Oracle, SQL Server, REST APIs, Web Services, XML, JSON, Maven, JUnit, Git, Jenkins, Linux/Unix, Shell Scripting, Eclipse

Responsibilities

- Developed and maintained **Java/J2EE applications** supporting enterprise telecommunications business processes and internal applications.
- Designed and implemented application components using **Java, Spring, Spring MVC, Hibernate, and JDBC** following established enterprise development practices.
- Developed **SQL queries, stored procedures, joins, views, and database scripts** for data retrieval, validation, reporting, and application functionality.
- Worked extensively with **Oracle and SQL Server databases**, supporting data analysis, troubleshooting, performance improvements, and application integration.
- Developed and consumed **REST APIs and web services** to enable communication between Java applications and downstream enterprise systems.
- Implemented backend business logic and data-access components using **Spring, Hibernate, and JDBC**.
- Performed application and database troubleshooting by analyzing **logs, SQL queries, application errors, and production issues**.
- Participated in requirements analysis and translated functional requirements into **technical designs and application components**.

- Developed unit tests using **JUnit** and performed functional validation to ensure application components met defined requirements.
- Used **Maven, Git, and Jenkins** to support application builds, source-code management, and automated deployment processes.
- Worked in **Linux/Unix environments**, using shell scripting and standard system utilities for application support, log analysis, and operational tasks.
- Collaborated with developers, QA engineers, database teams, business analysts, and stakeholders throughout the **development, testing, deployment, and production-support lifecycle**.