Harish Sista (He/Him)

(Permanent Resident)

(201) 310-5683
harish2sista@gmail.com
Brooklyn, NY-11217

SUMMARY

AI Research Scientist, pursuing Ph.D. in Machine Learning (ML) and Natural Language Processing (NLP), with over 7 years of practical experience dealing with ML models, Big Data, and Cloud Computing platforms. I have 4+ years of professional experience working collaboratively in team environments. Developed and deployed innovative models such as an Evidence Extraction and Fact Verification model employing Prompt Engineering and a Black Box Optimization model for real-time hyperparameter fine-tuning using concave analysis.

SKILLS

AI Tools & Frameworks	Data Engineering	Azure AI Solutions	Python Libraries	Cloud Platforms	Programming Languages
• Transformers • LangChain • SciKitlearn • PyTorch • TensorFlow • NumPy • NLTK	• Kubernetes • Docker • Databricks • SQLAlchemy • PostgreSQL • MongoDB	• Azure AI Foundry • Azure Machine Learning Studio • Lakehouse AI • PowerApps • Power Automate • Copilot Studio	• FastAPI • TypeScript • Dash • Multiprocessing • NumPy • Pandas • Matplotlib	• Azure • AWS • Google Cloud • GIT	• Python • SQL • NoSQL • C++ • Swift

EXPERIENCE

AI Research Scientist/Software Analyst (Contract), Eplus – New York, NY 03/2025 – Present

- Built system designs for various applied financial workflows by understanding the client's needs and the technology specifications and limitations from the developers.
- Analyze technical challenges and recommend SOTA scholarly research solutions in designing an efficient Agentic AI workflow using prompt engineering.
- Represent the client's needs and interpret the technical designs, solution and challenges in weekly meetings and oversee developer's tasks in building Azure workflows and RAG system design.
- Collaborate with several client teams in generating development and test data for AI models and facilitated this process by building use case specific survey templates using Python and Dash web applications.
- Built various orchestration agents using OpenAI agent tools and hosted them using FastAPI endpoints.
- Built a MicroRAG system for effective long document processing using Azure Databricks – Lakehouse AI, Spark and Azure AI Foundry agents.
- Built various agentic architectures using Copilot Studio for interfacing with SEC.gov, salesforce APIs, and Microsoft 365 products.

AI Research Scientist (Consultant), Talentix (Remote) – Fremont, CA 06/2024 – 03/2025

- Collaborate with the founding team in building solutions to the core problem and design the ML infrastructure.
- Oversee and assign tasks to the Data Engineering team in developing the server-client infrastructure for ML models using AWS (EC2, SQS, S3, Lambda, Docker, EKS).
- Architect Python backend workflows and database schemas using PostgreSQL for data manipulation and processing
- Conduct research on serverless LLMs through OpenAI and Azure AI to optimize cost and infrastructure efficiency.
- Utilize Python ML libraries such as Transformers, PyTorch, and Azure AI for building, training, and fine-tuning LLMs and DNNs.
- Innovate GenAI solutions by creating agentic prompts integrated with RAG data to drive Agile solution development.

Research Assistant, Stevens Institute of Technology – Hoboken, NJ 10/2018 – 09/2019

- Developed a chemistry literature dataset optimized for similarity search with enhanced query capabilities.
- Applied data preprocessing techniques, including PDF conversion and noise reduction via the NLTK library, organizing data using Pandas.
- Engineered a Novel Generative Ranking Model for relevant paper extraction based on topic and key phrase analysis, implemented with SciKitlearn and Networkx libraries.

AI Engineer/iOS Developer, Fresh Digital Group – NY, NY 07/2016 – 06/2017

- Collaborated with Content Writers and Creative Team to individually design and deploy 30 chat applications for Amazon-Alexa which won appreciation from tech newsletters.
- Architected solutions with AWS SDKs, deploying resources using EC2 instances, S3 Buckets, and Lambda functions.

- Contributed to achieving preferred partnership statuses with Amazon-Alexa and Microsoft-Cortana.

Ph.D. Candidate Infinity Lab, Stevens Institute of Technology – Hoboken, NJ 01/2018 – Present

- Work with various opensource and closed LLMs to address the challenges of Fact Verification using Prompt Engineering, Text Generation, Classification and Evidence Extraction Tasks.
- Designed LLM & ML frameworks to make complex the API calls and built customized input data structures using Requests and asyncio.
- Write LLM & ML workflows to implement the tasks of prompting engineering, model training, and fine-tuning on local and on cloud platforms using AWS, Azure AI and Transformers libraries.
- Built various Python Libraries using PyTorch and Multiprocessing for hyper-parameter fine-tuning, and automated fact verification.
- Produced various novel Bayesian inferencing algorithms using Torch, NumPy and Pandas libraries and implemented various SOTA modeling functions from SciKitlearn, Gensim, and scikit-spatial libraries.
- Showcased the feature and performance metrics using Matplotlib and Seaborn libraries.
- Surveyed and implemented various Attention Networks, DNNs and Generative Models to perform NLP tasks of Data Extraction, Modeling, Regression, Ranking, and Classification.
- Organized hardware and software for the lab-server and built a personal machine learning RIG.

EDUCATION

Ph.D. in Machine Learning (Computer Engineering department) 01/2018 – 05/2025
Stevens Institute of Technology, Hoboken

M.Eng. in Software Engineering (Electrical Engineering department) 01/2014 – 12/2015
Stevens Institute of Technology, Hoboken

B.Eng. in Electronics and Communication Engineering 08/2009 – 05/2013
M.V.S.R Engineering College, Hyderabad, TG, India

Publications

EEFactUPP: Evidence Evaluation and Fact Verification using User Perspective Prompting
AIR-RES 2025(SpringerNature) 04/16/2025

PERSONAL PROJECTS

Resume Assistant 10/01/2024

- Created an AI-driven application employing advanced GPT-LLMs and prompt engineering techniques for personalized resume generation.
- Engineered a custom GPT-API framework to handle both text and image processing.
- Project published as an open-source Python library on PyPI, achieving 3000 downloads within the first three weeks.

AI for NYC meetup - November - Prompt Engineering Techniques 10/30/2024

- Conducted a seminar analyzing hallucination patterns in LLMs and presenting state-of-the-art prompt engineering solutions.
- Delivered practical examples through Jupyter Notebooks during an open seminar attended by 50+ participants.

CERTIFICATIONS

- Microsoft Certified: Azure AI Fundamentals 05/27/2025
- Databases and SQL for Data Science with Python – IBM, Coursera 09/09/2024