Prabhath Chellingi

ML Research Engineer — LLM Systems — Agentic AI — Knowledge Graphs

+91-9963771971 prabhathchellingi2003@gmail.com in/prabhath-chellingi github.com/Prabhath003

About

ML Research Engineer with a strong focus on **LLM reasoning**, **Agentic workflows**, and **Knowledge Graphs**. Experienced in fine-tuning SLMs and designing **graph-augmented RAG systems** to reduce hallucinations. Hands-on in building **scalable inference pipelines**, **autonomous agents**, and **retrieval systems**. Interested in advancing **trustworthy**, **interpretable**, and **efficient AI systems**.

Education

Indian Institute of Technology, Hyderabad

B.Tech. in Computer Science and Engineering | Grade 8.52 2020 - 2024

Minor in Entrepreneurship | Grade 8.92 2022 - 2024

Research Publications

Prajwal Ganugula, Y. S. S. S. Santosh Kumar, N. K. Sagar Reddy, **Prabhath Chellingi**, Avinash Thakur, Neeraj Kasera, and C. Shyam Anand

Multi-Object Segmented Arbitrary Stylization Using CLIP

ICCV Workshop, 2023

Experience

Stealth Startup - GiKA.ai

Founding ML Engineer Jun 2024 - Present

- **Agentic AI Systems & Reasoning Platform**: Developed an **in-house agentic framework (LangGraph-like)** supporting multi-step LLM reasoning, planning, tool invocation, and memory management for autonomous execution of complex workflows.
- Designed and deployed **agentic planners and pipelines** with terminal and web access for autonomous data gathering, real-time analytics, and **knowledge graph construction and traversal** from heterogeneous sources for **fin-tech**.
- **Model Optimization & Knowledge-Aware Learning**: Fine-tuned **Small Language Models** (LLaMA-3.1-8B, Gemma-2-9B) for **knowledge graph–aware reasoning**, improving structured inference and factual grounding in downstream tasks.
- **Commercial AI Applications & Impact**:
  - Developed an **Agentic JSON Builder** for large-scale structured data completion
  - Autonomously populated **8,000+ fields** via context-aware retrieval, reducing manual effort by **∼95%**
  - Built a **low-latency Voice AI Agent** for an edtech sales platform
  - Achieved **sub-300ms retrieval latency**, **90%+ intent accuracy**, covering **1,500+ courses**
- **Knowledge-Driven Search & Web Intelligence**: Engineered a **knowledge graph–powered semantic search engine** enabling intent-based product discovery with improved retrieval accuracy.
- Developed scalable **web intelligence infrastructure**, including a universal HTML-to-structured-data parser and a Playwright/Selenium-based crawler supporting dynamic interactions (**∼20s/page**).
- Researched and implemented advanced **coreference resolution** and **entity linking** pipelines (GPT-4, LLaMA3, SpanBERT, LingMess) to validate the superiority of **Knowledge Graphs over vector databases** in **RAG systems**.

Indian Institute of Technology, Hyderabad

Research Assistant(Mini Project) under Prof. Saketha Nath 2024

- RamayanaGPT- A Knowledge Retrieval Approach:
  - Worked on designing a **context-aware retrieval system** for Ramayana-based QA.
  - Built a **graph-based retrieval engine** linking database chunks via semantic entities and edges to reduce LLM hallucination.

OnePlus/Oppo(OPLUS) Mobiles India R&D

Innovation Research Intern Jan 2023 - June 2023

- Contributed to **MOSAIC** (Multi-Object Segmented Arbitrary Image Stylization), accepted at **ICCV'23 workshop**; performed extensive ablation studies on key models (e.g., ITNet, StyTr2, HDRNet).
- Compressed Dumoulin-style CNN networks for edge deployment, achieving **8x size reduction** and **11x FLOPs optimization**.

- Developed **ODASH**, a debugging tool for R&D teams, implementing a full-stack web app with **KNN search on log embeddings**.

Personifwy

AI Intern Trainee Jan 2022 - Mar 2022

- Executed end-to-end **ML projects** in **image processing** and **NLP**—including news classification, handwritten digit recognition, and fashion object detection—using **CNNs**, OpenCV, scikit-learn, and TensorFlow.

Projects

AgenticRAG | *Autonomous Retrieval-Augmented Generation System* **2025**

- Developed an **agentic RAG framework** enabling **chunk-level traversal** and **document jumping**, mimicking human-like reasoning and exploration across large knowledge corpora.
- Engineered for large-scale retrieval tasks (10K+ docs) with adaptive context management, and multi-hop reasoning improving factual accuracy by 25% and contextual depth in generation.

LLM Inference Server | *GPU management* **2025**

- Open-sourced a **GPU-optimized inference server** for HuggingFace models with dynamic **model offloading/loading**.
- Designed to act as a central access node within MCP protocols, serving 10+ concurrent GPU tasks across distributed inference servers, offering an alternative to OLLAMA servers.

Finance Time Series Forecasting with MixerMLP | *Python, Pytorch, ANNs, Finetuning* **2023**

- Adapted the **MixerMLP architecture** for financial time series prediction with a novel input format for tabular data.
- Achieved a **low MSE of** 1.3×10^{-3} on benchmark datasets.

Achievements & Certifications

Ethical Hacking and Cyber Security Masterclass - *Udemy* **Jan 2025**

- Comprehensive training in ethical hacking, penetration testing, network security, vulnerability assessment, and cybersecurity defense mechanisms.

MongoDB Developer Toolkit - *GeeksforGeeks* **Dec 2024**

- Advanced MongoDB skills: replication, sharding, indexing, aggregation framework, data modeling, NoSQL database design and optimization with Atlas.

IBM Statistics 101 Badge - *IBM Developer Skills Network* **2022**

Top 5 Finalist - Novus X.0 Electronic Prototype Competition **2021**

JEE Advanced All India Rank 1689 — JEE Mains All India Rank 2345 **2020**

Technical Skills

Languages: Python | C++ | Java | JavaScript | SQL | HTML/CSS
Frameworks & Libraries : PyTorch | Tensorflow | Flask | Celery | OpenCV | ROS | FastAPI
Tools & Platforms: Git | Docker | pgAdmin4 | Android Studio | OpenSearch | Neo4j | Playwright | MongoDB
Familiar With: LLM Frameworks (HuggingFace, LangChain) | vLLM | BitsAndBytes | Linux Shell Scripting

Leadership / Extracurricular

- Core Member – **Robotix (IITH Robotics Club)**
- Publicity Volunteer – **Elan & Envision (IITH Tech Fest)**