Manikantan (Mani) Srinivasan

Boston, MA | +1 (857)-218-2285 | mani.srinivasan2k1@gmail.com | LinkedIn | Portfolio | GitHub

EDUCATION

Northeastern University **Sep 2023 – Dec 2025**
Khoury College of Computer Sciences **(Expected)**
MS in Artificial Intelligence - GPA: **3.7** **Boston, MA**
Relevant Coursework: NLP, AI in Human Computer Interaction, LLM, Machine Learning, Data Science, Unsupervised ML
Teaching Assistant for CS6220 Data Mining – *created assignment sets for graduate-level data mining course for topics like ARM, PCA, Data visualization and Analytics.*
Research Area*– Generative Multimodal AI and Agentic Systems – MCP, Diffusion Transformers, LLMs and Agentic Analytics*

PROFESSIONAL EXPERIENCE

SignoFi **Sep 2024 – Jan 2025**
AI Research Engineer Intern **Raleigh, NC**

- Pioneered an Agentic AI application leveraging Llama 70B and LangChain to retrieve balance sheet data from NoSQL databases, enabling interactive queries and computational analysis for over 50,000 public companies.
- Leveraged RAG and Fine-tuning with adaptive chunking strategies to ground Llama 70B and 90B, boosting the Faithfulness score to **0.84 (84%)** and enhancing contextual precision, recall, and relevancy **(92%)** in factual question answering.
- Architected a custom workflow on Google Cloud Platform using virtual machines, Vertex AI, GCP Buckets, and multi-core GPUs to optimize operations at a portfolio management firm, achieving up to **20%** in cost savings.

KEK **Aug 2022 – Aug 2023**
Software Engineer – R&D **Ibaraki, Japan**

- Developed distributed computing solutions for the prestigious BELLE-2 particle acceleration experiment, officially part of **KEK** and working across **5+** world renowned laboratories including **KEK, Brookhaven and CERN**.
- Created innovative software tools that empowered physicists to experiment with over **10 PB** of data across centers.
- Researched on grid computing techniques and **Graph ML** to help physicists interpret particle data across **127+** institutions and labs improving average job execution time on the grid by **10%** using **AI-based predictive modelling**.

Jio **May 2022 – Jul 2022**
Software Engineer Intern – (Center of Excellence - AI & ML) **Mumbai, India**

- Engineered the BrainOS platform integrating **7+** verticals for Reliance Industries, utilizing knowledge graphs and predictive ML across products like JioTV, Hydrocarbons, and JioMart, achieving **86%** accuracy and **0.79** precision.
- Architected robust knowledge graphs by preprocessing raw data and implemented **100+** CRUD API functions using gRPC to query NoSQL databases, utilizing debugging tools to analyze and refine **1000+** functions.

SKILLS

Languages: Python, Java, C, C++, Go, HTML, CSS, JavaScript, Swift, Scala, R
Databases & DQL: MongoDB, ArangoDB, SQL, NoSQL, Redis, Pinecone, GraphQL
Cloud & DevOps: CI/CD, Docker, AWS, AWS BeanStalk, AWS S3, GCP, Azure, Git, GitHub, GitLab
Data Science: Pandas, Plotly, Seaborn, NumPy, MATLAB, KDP
AIML: PyTorch, HuggingFace, TensorFlow, LangChain, LangSmith, LlamaIndex, A2A, MCP, Reinforcement Learning, Data Analytics

PROJECTS

ClinAI: MCP-based Agentic platform for voice enabled patient record management

- Engineered an agentic AI clinical assistant (ClinAI) leveraging Model Context Protocol and Gemini LLMs to automate extraction and structuring of clinical data from conversations, reducing manual documentation time for doctors by over **60%.**
- Deployed a multi-agent workflow for real-time patient record updating and action item extraction, achieving a low categorisation-error rate of **6.2%** on voice transcriptions and reducing manual documentation workload across 100+ clinical transcripts. - https://github.com/mani2001/clinai-agent

Machine Unlearning: Erase key concepts from LLM's Knowledge base

- Designed and applied **3** machine unlearning techniques **(SAE, LoRA Fine-tuning, Guardrailing)** on a **7B** parameter LLM, ensuring a retained perplexity score under **20** for unaffected knowledge during concept forgetting.
- Performed a comparative analysis with both qualitative and perplexity-based studies. Performed hyper-parameter tuning for PEFT finetune with **1B+** parameters. - https://github.com/mani2001/Machine-Unlearning

BlindSight: AI-Powered Voice-Driven File Navigation for the Visually Impaired

- Developed an **AI-powered voice navigation system**, enabling visually impaired users to perform file management and document editing tasks with **95%** speech recognition accuracy using Open-AI's C++ implementation of **Whisper.**
- Incorporated **LLaMA-70B** with in-context learning (ICL) to deliver task-specific capabilities, such as document analysis and advanced text editing, achieving **85%** accuracy in real-world testing scenarios.
- Designed a multi-threaded architecture that reduced task execution latency by **30%**. - https://github.com/mani2001/BlindSight