Shishir Kallapur

Boston, MA, 02119 | (582)201-8592 | kallapur.shi@northeastern.edu | github.com/shishirkallapur | linkedin.com/in/shishirkallapur

Professional Summary

AI-focused Software Engineer and aspiring Machine Learning Engineer with 2+ years of experience delivering full-stack and intelligent solutions. Skilled in machine learning, reinforcement learning, NLP, transformers, large language models (LLMs), model fine tuning, prompt engineering, vector database integration and cloud-native development. Experienced in building GenAI and RAG pipelines for production-ready solutions. Passionate about translating cutting-edge AI research into robust, production-grade systems that deliver measurable business value.

Education

Northeastern University, Boston, MA — Sept. 2023 – May 2025
Master of Science in Artificial Intelligence — GPA: 3.91
Khoury College of Computer Sciences
Courses: Foundations of AI, Programming Design Paradigm, Algorithms, Machine Learning, Reinforcement Learning, Natural Language Processing, Advanced ML, AI for HCI

The National Institute of Engineering, Mysore, India — Aug. 2017 – Aug. 2021
Bachelor of Engineering in Computer Science and Engineering — GPA: 3.57

Technical Knowledge

Languages: Python, JavaScript, HTML, CSS, Java, SQL, C, C++, Angular
Databases: MySQL, MongoDB
AI/ML: GenAI, LLMs, RAG, Reinforcement Learning, NLP, Transformers, MLOps, ML System Design, Model Fine-Tuning
Frameworks: PyTorch, TensorFlow, Scikit-Learn, OpenCV, Spring, Streamlit, JUnit, NumPy, Matplotlib
Tools: Git, Docker, Pinecone, Gspread, AWS (EC2, S3, Lambda), JIRA, ServiceNow
Certifications: AWS Cloud Practitioner, ServiceNow Certified System Administrator

Work Experience

Amplifier Security — May 2024 – Aug. 2024
AI Product Intern
- Spearheaded a comprehensive benchmarking initiative for GPT models(GPT-3.5, GPT-4, GPT-4o) significantly enhancing Ampy's response accuracy, speed, and overall performance.
- Implemented guardrails and prompts that boosted topical relevance by 35%, reducing hallucinations.
- Automated response evaluation with custom Python scripts, improving testing speed by 3x.
- Implemented a Retrieval-Augmented Generation (RAG) system with LangChain using Pinecone as Vector DB, enabling contextual replies from proprietary unstructured data.

JP Morgan Chase & Co., Bangalore, India — Sept. 2021 – Aug. 2023
Software Engineer
- Overhauled ServiceNow Knowledge module, enhancing request resolution speed by 20%.
- Integrated JIRA with ServiceNow to automate SDLC tracking and reporting, incorporating CI/CD automation best practices and reducing manual effort by 40%.
- Delivered 5 reusable UI macros to streamline HR documentation workflows; improved HR team's document update efficiency by 45%,
- Introduced and deployed catalog automation features, reducing request handling time by 30%.

MiQ Digital, Bangalore, India — Jan. 2021 – July 2021
Software Developer Intern
- Implemented full-stack features using Spring Boot and AngularJS, enhancing platform performance and UX.
- Integrated DSPs (Xandr, DV360) into internal tools, streamlining campaign activation and data processing.
- Enabled DV360 as a viable DSP option, accelerating project timelines and increasing platform utility.

Projects

Relating Physical Activity to Problematic Internet Use in Youths — Sept. 2024 – Dec. 2024
- Developed a ML pipeline to identify at-risk youths, leveraging physical activity data to promote digital welfare.
- Used transformer autoencoders and Random Forest based imputers to preprocess noisy, incomplete data.
- Achieved 72% mean QWK score using a voting classifier that combined XGBoost, LightGBM, and CatBoost, effectively addressing dataset complexity and imbalance.

nGPT, BART and PEGASUS: A Comparative Study — Sept. 2024 – Dec. 2024
- Designed a benchmarking pipeline to compare nGPT, BART, and PEGASUS for abstractive summarization tasks.
- Fine-tuned nGPT for text summarization, validating faster convergence and lower resource usage with optimized hyperparameters.
- Analyzed training loss trends, inference times, and ROUGE scores to confirm nGPT's efficiency in model deployment.

Enhancing Bipedal Robot Locomotion using RL with Reference Signal Integration — Jan. 2024 – April 2024
- Developed a RL pipeline to enable a bipedal robot to achieve natural, energy-efficient locomotion in a simulated environment.
- Designed a unique reward function to optimize joint parameter control, driving improvements in balance and energy efficiency.
- Integrated control systems with RL methodologies to refine the robot's navigation along predefined paths for efficient locomotion.