

Krish Nagaral

Boston | nagaralkrish@gmail.com | USA: +1(857) 379-3112 | [LinkedIn](#) | [GitHub](#)

Summary

AI/ML Engineer with experience in developing production-ready AI systems, fine-tuning large language models, and building scalable ML solutions. Published researcher with practical expertise in LLM optimization, RAG pipelines, and real-time AI applications, achieving high accuracy and performance.

Skills

AI/ML: Model Training & Fine-Tuning (LoRA), Natural Language Processing, RAG Pipelines, Deep Learning, Generative AI, Computer Vision, CNN, RNNs, Reinforcement Learning, Machine learning, LlamaIndex

Frameworks & Libraries: PyTorch, TensorFlow, Hugging Face Transformers, Scikit-learn, XGBoost, CI/CD Pipelines

Programming: Python, Java, SQL, C++, React, Nodejs, JSON, Data visualization, Pattern Recognition, Jupyter Notebooks

MLOps & Cloud: AWS (SageMaker, EC2, S3, lambda, Cloudwatch), Azure, Docker, Github, MLflow, Kubernetes, Redis, RESTful APIs, APIs integration, Pandas, NumPy, Agentic AI, Ai Strategy

Tools & Platforms: Flask, FastAPI, LangChain, Excel, Voice Synthesis Libraries, Automation Tools Evaluation, Debugging

Experience

AI Research & Tool Developer Intern, Coutloot & Flyads.ai, - Mumbai May 2025 – Aug 2025

- Built AI automation tools increasing workflow efficiency by 85%, fine-tuned LLaMA/Gemma models using LoRA, achieving 92% performance improvement.
- Developed RAG pipelines using LangChain for context-aware automation with 97% accuracy.
- Researched, trained, and deployed AI cloning models for human avatar generation, integrated into ad creation pipelines.

Software Tester Intern, Coutloot - Mumbai

Apr 2024 – Jun 2024

- Designed testing frameworks, improving application stability by 88% with 99.8% product reliability before deployment.
- Recorded and reported test cases, ensuring traceability and actionable insights for developers.
- Improved application performance by systematically identifying and resolving bugs.

Cloud Engineer Intern, Tata Communications - Mumbai,

Sep 2023 – Mar 2024

- Contributed to team efforts in authoring deployment documentation and training materials to enhance onboarding efficiency.
- Supported live systems by troubleshooting incidents and collaborating with cross-functional teams.
- Applied best practices in AWS EC2, S3, and Lambda to deliver resilient cloud solutions

Projects

- JARVIS Voice Assistant** - Python, LLM Models, Voice Synthesis Developed an AI assistant with natural language understanding, achieving 98.5% speech recognition accuracy and sub-150ms response time. Implemented real-time voice synthesis with multi-modal interaction capabilities, processing 5,000+ daily queries.
- LegalLens AI Platform** - Python, NLP, Machine Learning. Built a legal analysis system processing 50,000+ documents with 96% automated citation accuracy. Developed an AI debate generator creating balanced legal arguments with 94% legal precedent accuracy verified by legal experts.
- Offline ChatGPT Clone** - Python, LLM Integration, GUI Created offline conversational AI supporting multiple models (Llama, CodeLlama, GPT-OS), achieving 87% faster response times with optimized resource management and 99.2% uptime reliability.
- Augmented Reality Navigation for Visually Impaired**- Developed an AI-based AR navigation system with real-time obstacle detection, achieving 94% obstacle recognition accuracy and 88% navigation success rate for visually impaired users in urban environments. Tools Used: Python, Computer Vision, AR SDK, AI Models & DL Models
- Credit Risk Scoring System** - Python, XGBoost, SHAP, Flask API Built ML model predicting loan default risk with 91% prediction accuracy, reducing financial losses by 89%. Implemented explainable AI using SHAP for regulatory compliance with 95% model interpretability.

Education

Khoury College of Computer Science, Northeastern University, Boston

Aug 2025 – May 2027

MS in Computer Science

Vellore Institute of Technology, Chennai

Aug 2021 – May 2025

B.Tech Computer Science and Engineering

Certifications

AWS Solution Architect Associate | AWS AI Practitioner | Deep learning.ai Prompt Engineering | Deep learning.ai Machine learning | Huggingface MCP | Nvidia RAG | IBM Cybersecurity