

ABHISHEK BOTE

Jersey City, NJ · (551)-344-9129 · abhishekbote@gmail.com · [LinkedIn](#) · [GitHub](#) · [Portfolio](#) · [LeetCode](#)

TECHNICAL SKILLS

Languages: Python, C++, R, SQL, TypeScript, JavaScript

AI/ML & GenAI: PyTorch, TensorFlow, Hugging Face Transformers, LangChain, LangGraph, CrewAI, LoRA/PEFT, OpenCV, Scikit-learn

LLM & Agentic AI: RAG, Prompt Engineering, Vector Databases, Fine-tuning, Multi-Agent Systems, OpenAI API, Groq, Perplexity API, Gemini, LLaMA

Backend & APIs: FastAPI, Node.js, Next.js, REST APIs, Microservices, Serverless Architecture

Cloud & MLOps: AWS (Lambda, S3, Amplify), Azure AI Foundry, GCP, Docker, Kubernetes, Kafka, Databricks, Snowflake, GitHub Actions, CI/CD

Specializations: Generative AI, NLP, Computer Vision, Deep Learning, Multimodal AI, Distributed Systems

WORK EXPERIENCE

Namaste Global

Jersey City, NJ · July 2025 - Aug 2025

AI Engineer Intern

Python, LangChain, Perplexity API, AWS Lambda, S3, Amplify

- Architected and shipped a production **AI customer support agent** on **LangChain** and **Perplexity-powered retrieval**, processing **100K+ user requests** and matching them to vetted experts across **3 business verticals**.
- Engineered a **multi-step NLP pipeline** that converted unstructured user queries into structured intents, fully automating expert routing and eliminating manual triage from the support workflow.
- Designed an **intent taxonomy** around the top 3 categories driving **70% of user demand**, focusing routing capacity on high-value queries and lifting expert-match quality.
- Launched a **serverless AWS stack** (Lambda, S3, Amplify) serving **300+ daily active users** at **sub-2-second latency**, reducing infrastructure cost by **40%** versus an EC2 deployment.

Nebula Technology

Pune, India · April 2021 - July 2023

AI Software Engineer

Python, FastAPI, GCP, TensorFlow, OpenCV, GitHub Actions

- Architected a **multimodal AI assistant** unifying speech, vision, and OCR into a single interface, scaling voice, text, and image interactions to **5K+ daily user sessions**.
- Developed an **ML-powered workforce automation system** with a skill-matching algorithm trained on employee profiles, cutting task-assignment time by **70%** and eliminating manual allocation for a **50+ person team**.
- Analyzed **20K+ chatbot conversations** to surface intent-classification bottlenecks and drop-off points, then retrained the **NLP pipeline** to lift conversation-completion rates by **25%**.
- Established **MLOps infrastructure** on GCP with **GitHub Actions**, automating **CI/CD across 3 AI products** and accelerating deployments from **4 hours to 20 minutes**.

Nebula Technology

Pune, India · April 2019 - March 2020

Junior Software Engineer

Python, SQL, Power BI, SharePoint, ETL, Excel

- Built **SQL-driven reporting pipelines** and Excel models tracking customer-service SLAs across business units, automating ticket monitoring to cut backlog by **23%** and save **20+ hours/month** of manual reporting effort.
- Engineered an automated **Python + SharePoint ETL flagging system** that detected missing risk and compliance fields in real time, replacing manual data audits and preventing downstream reporting delays.

PROJECTS

PolyLingua: Multilingual NLP Framework

[GitHub](#)

Python, PyTorch, Hugging Face Transformers, LoRA/PEFT, Gemma, LLaMA-3

- Designed an **English → Indic LLM translation system** targeting low-resource Indian languages, training on a multilingual corpus of **~1.5M sentences** aggregated from 3+ open-source datasets.
- Fine-tuned **Gemma 2B, Gemma 7B, and LLaMA-3 8B** with **PEFT/LoRA**, achieving a **~20% improvement** on low-resource translation benchmarks over baseline LLMs.

BLUD AI - AI Career Roadmap Platform

[Live Demo](#)

Next.js 16, TypeScript, Supabase, Groq LLM, Vercel AI SDK

- Shipped a **full-stack AI career platform** that generated **80+ personalized roadmaps** from **35K+ user queries** via a multi-step **Groq LLM** pipeline with deterministic resource ranking across **5 learning phases**.
- Deployed on **Vercel + Supabase** with **Row-Level Security** protecting **300+ users**, sustaining **99% uptime** through the 8-week production launch.

EDUCATION

Pace University - M.S. in Information Systems, New York, NY - **GPA: 3.85/4.0**

Sept 2024 - May 2026

Savitribai Phule Pune University - M.Sc. in Computer Science, Pune, India - **GPA: 9.15/10**

June 2021 - March 2023

Savitribai Phule Pune University - B.Sc. in Computer Science, Pune, India

June 2017 - May 2021

CERTIFICATIONS

Google Cloud - Advanced Generative AI Agents Developer · **NVIDIA** - Fundamentals of Deep Learning · **HackerRank** - Python (Advanced)

LEADERSHIP

Google Developers Group - *Organizer & Technical Lead*: Led technical execution for one of Google's flagship developer conferences, supporting **500+ attendees** across workshops, speaker sessions, and live demos.

Perplexity Campus Partner - Selected among the **top 5% of applicants** to drive on-campus adoption of Perplexity AI; organized workshops and demos reaching **300+ students** and growing platform usage by **200%** in one semester.