Abhishek Mamdapure
AI Engineer

✉ mamdapureabhishek@gmail.com   📞 +91-9421592524   in Linkedin

AI & ML Engineer | Generative AI Specialist | Innovator in Large-Scale AI Solutions

Results-driven AI Engineer with **5+ years** of specialized experience in Data Science & Generative AI, delivering measurable business outcomes. Reduced operational costs by millions through custom LLM implementations and scaled AI systems for Fortune 500 company. Expert in **production-ready NLP, SLM, LLM models and RAG solutions** that directly enhance user experience metrics.

💼 WORK EXPERIENCE

Citibank, *Officer - GenAI Development* 🔗   01/2024 – present | Pune, India
  - Spearheaded the development and deployment of a groundbreaking Generative AI project (**Google Gemini** based) within a large-scale customer-facing program (**IVR**).
  - Successfully led the exploration and evaluation of various AI/ML frameworks, ultimately developing a proprietary solution tailored to serve a vast customer base.
  - Built the **end-to-end AI pipeline**, including backend services using frameworks like **FastAPI**, ensuring scalable, efficient, and production-ready architecture.
  - Built and mentored a high-performing team to execute the project vision, ensuring successful delivery and exceeding expectations.
  - Delivered **multi-million dollar cost savings** through automation and AI-driven efficiencies in customer service workflows.
  - Pioneered the adoption of Generative AI at **Citi**, establishing foundational systems and standards for future AI initiatives across the organization.

Pienomial (Previously Wynum) 🔗

Lead Data Scientist   04/2022 – 12/2023 | Pune, India
  - Drove critical AI/ML initiatives from **prototype to production**, collaborating closely with the **CTO** and cross-functional teams.
  - Developed and deployed embeddings-based search engines using **FAISS** and commercial vector databases like **Pinecone** and **MongoDB Atlas**, powering fast and intelligent information retrieval across domains. (RAG Architecture)
  - Built and fine-tuned NLP models to extract **medical entities** (e.g., diseases, inhibitors) from biomedical abstracts, enabling downstream analytics for healthcare and life sciences.
  - Designed and implemented a **knowledge graph pipeline** over massive unstructured life sciences datasets, structuring complex biomedical relationships for advanced query and discovery use cases.
  - Created an **AI chatbot** that generates **contextual graphs and plots** from user queries—particularly impactful for pharma sales reps needing quick visual summaries during client interactions.
  - Leveraged **Apple Silicon chips** (M1/M2) for model inference workflows, significantly optimizing performance and reducing reliance on GPU-based environments.
  - Architected and maintained **scalable data pipelines** for ingesting, processing, and storing large structured and unstructured datasets across multiple platforms.
  - Transitioned 5+ high-impact projects to production under tight deadlines, balancing performance, scalability, and cost-efficiency.
  - Handled technical communication with 3 major enterprise clients, simplifying complex ML concepts for business leaders and contributing to strategic decision-making.
  - Mentored and coached a team of 6+ junior data scientists, fostering a culture of code quality, experimentation, and robust system design.

Data Scientist   07/2020 – 04/2022 | Pune, India
  - Analyzed large datasets to identify trends, patterns, and relationships.
  - Designed and implemented data pipelines to collect, store, and process large datasets.
  - Developed dashboards and reports to visualize and communicate data insights to stakeholders.
  - Stayed up-to-date with the latest trends and techniques in the field of data analysis.
  - Participated in the continuous improvement of data analysis processes and tools.

Affine Analytics Private Limited, *Data Science Intern*   01/2020 – 07/2020 | Bangalore, India

    During my tenure at Affine, I was responsible for designing and implementing an NLP system for converting natural language questions into SQL queries for a specific database.

🎓 EDUCATION

Masters of Technology in Mathematical Modeling and Simulation,   2020 | Pune, India
Pune University ✐

    Subjects: Machine Learning, Operations Research, Optimization, Numerical Computing.
    Alumini Profile Link ✐

Bachelors of Engineering in Electronics, *Pune University* ✐   2016 | Pune, India

🧠 SKILLS

Python Programming, **Machine Learning**, **Deep Neural Networks**, **PyTorch**, **Statistical Analysis**, **Natural Language Processing**, **Large Language Models (LLMs)**, **Transformers**, **Amazon Web Services (AWS)**, **Productionizing Models**, **MongoDB**, **Customer Segmentation**, **SDLC and Agile Methodology**, **Stakeholder Communication**, **Team Management**, **Generative AI** — RAG │ Usage of open-source and commerical LLMs (Llama, gpt's, anthropic, gemini, mistral) │ Prompt Engineering (few shot, chain of thought)

📑 PAPERS & CERTIFICATES

AI Engine for research and modelling in field of Immunotherapy ✐ — ● **NPTEL - Deep Learning** — *Indian Institute of Technology - Kharagpur* ● **NPTEL - Practical Machine Learning With Tensorflow** — *Indian Institute of Technology - Madras*