

Barathi Kanna Saminathan

Birmingham, United Kingdom, B5 7AJ | +44 7818998783 | barathikanna07@gmail.com | [LinkedIn](#) | [GitHub](#)

SUMMARY

AI & Data Engineering Practitioner with a strong foundation in multimodal AI, Large Language Models (LLMs), and Retrieval-Augmented Generation (RAG) systems. Experienced in designing, deploying, and explaining AI/ML solutions using causal inference, explainable AI techniques (DiCE, SHAP, LIME), and cloud-native architectures. Proven ability to integrate speech, text, and vision pipelines with vector databases for production-grade applications. Background in Oracle database administration, automation, and real-time data engineering in both on-premise and cloud environments (AWS, Azure).

WORK EXPERIENCE

UoB & Neptune Submerge

ML Engineer Intern

July 2025 – August 2025

Remote, United Kingdom

- Designed a Retrieval-Augmented Generation (RAG) voice assistant using Ollama-LLaMA, LangChain, and Nomic embeddings, integrating it into the existing VR system.
- Implemented speech-to-text (Vosk) and text-to-speech (Azure Cognitive Services) for real-time voice interaction.
- Stored knowledge in ChromaDB vector database for high-speed semantic retrieval.
- Built a Tkinter-based prototype to demonstrate AI-driven voice conversations with domain-specific knowledge.

Epsilon

September 2022 – August 2024

Bangalore, India

Oracle Database Administrator (KELK Stack)

- Maintained Oracle databases by performing regular patching, backups, performance tuning, and user access management to ensure secure and stable operations.
- Automated OS patching across 296+ servers using Ansible, reducing manual effort by 80% and saving over 150 engineering hours per quarter.
- Led the migration from Cloudera Kafka to Confluent Kafka, achieving 99.99% uptime, cutting infrastructure costs by 60%, and improving message throughput by 35%.
- Scaled AWS infrastructure to support real-time data flows, resulting in a 25% increase in system responsiveness during peak usage.
- Optimized Elasticsearch clusters and built custom Logstash pipelines, reducing query latency by 40% and improving data ingestion efficiency by 30%.
- Resolved over 100 high-impact database incidents during on-call shifts, enhancing overall system reliability and reducing downtime.

TATA Consultancy Services

May 2021 - September 2022

Bangalore, India

Assistant System Engineer (Business Analyst, Morgan Stanley)

- Certified 35+ high-risk applications for Morgan Stanley's Access Management, reducing security-related incidents by 90%.
- Integrated 70+ applications into secure platforms, streamlining governance workflows and improving access control compliance by 40%.
- Automated real-time Jira reporting using Python, reducing report generation time from 3 hours to 5 minutes (95% time savings).
- Built a dynamic Python dashboard to monitor Agile progress, increasing sprint visibility and improving project delivery efficiency by 25%.
- Maintained strong relationships with 20+ senior stakeholders (VPs to MDs), enabling faster decision-making cycles and greater project alignment.
- Collaborated across tech and compliance teams, contributing to a cross-functional workflow that cut policy validation cycles by 30%.

EDUCATION

University of Birmingham, Birmingham, United Kingdom

2024 - 2025

Master of Science, Artificial Intelligence and Machine Learning

- GPA: 2:1 (Expected)
- Coursework: NLP, Neural Computation, Machine Learning, Data Visualization, Linear Algebra, Probabilities, Game Theory

Amrita Vishwa Vidyapeetham, Coimbatore, Tamil Nadu, India

2017 - 2021

Bachelor of Technology, Electronics and Communication Engineer

- GPA: 7.67/10
- Honours: First class with Merit
- Coursework: Digital Signal Processing, IoT, Soft Computing, Pattern Recognition, VLSI

PROJECTS

Churn360-Ecomm: A Multimodal Explainable AI Framework for E-commerce Customer Churn Prediction

- Designed an AI solution to predict which customers are likely to leave by analysing purchase behaviour and review text on a realistic, causally modelled dataset.
- Used Logistic Regression and XGBoost to uncover the main reasons behind customer attrition.
- Applied DiCE counterfactuals, SHAP, and LIME to make insights clear and actionable for non-technical teams.
- Extended beyond prediction to analyse counterfactual scenarios (e.g., impact of offering discounts) for targeted retention strategies.
- Built an interactive Streamlit dashboard so marketing teams can see churn risk, key drivers, and recommended actions at a glance.

Research Assistant: Scientific Journal Parser

- Developed an AI-powered parser using PyPDF, CSVLoader, and LangChain to process and analyse research PDFs.
- Extracted text, tables, charts, and images into structured ChromaDB storage for retrieval.
- Implemented an RAG(Nomic Embedding) + LLM (llama3) chatbot to answer complex, multi-document research queries, including visual data interpretation.
- Built a Streamlit frontend supporting multi-project article organisation, conversational querying, and analytics.

PUBLICATIONS

Smart Shop Surveillance System

Publisher: Springer

- Built a multimodal AI solution combining IoT-based temperature checks, real-time face mask detection using CNN, and a mobile app for tracking and updating customer data and health records in a central database.
- DOI: <https://doi.org/10.1007/978-981-16-2422-355>

Low-Cost Smart Navigation System for the Blind.

Publisher: IEEE

- Designed a navigation system for the visually impaired using vibration motors for turn indications.
- DOI: <https://doi.org/10.1109/ICACCS51430.2021.9442056>

TECHNICAL SKILLS

- **Programming Languages:** Python, C++, Linux shell, Oracle SQL, SQL
- **Data Engineering:** Kafka, Elasticsearch, Logstash, Kibana, Grafana, Prometheus
- **Database:** Oracle, SQL server
- **Cloud Infrastructure:** AWS (EC2, S3, RDS), Azure Cognitive Services, Terraform, Ansible Tech Stack: Jira, OEM, Ansible, Terraform, Power BI
- **Research Interests:** Speech Foundation Models, Agent-based Systems, Causality, AI Explainability
- **Machine Learning & AI:** LLMs (Ollama–LLaMA), LangChain, DiCE, SHAP, LIME, Scikit-learn, XGBoost, NLP, TTS, Neural Networks
- **Vector Databases & RAG:** ChromaDB, FAISS, Nomic Embeddings
- **Document & Data Parsing:** PyPDF, CSVLoader, PDFMiner, data wrangling
- **Visualization:** Streamlit, Power BI, Matplotlib, Seaborn
- **Other:** Jira, Agile/Scrum, SDLC, DevOps CI/CD