Dhruv Kamalesh Kumar

Boston, MA | +1 (737) 296-4694 | kamaleshkumar.d@northeastern.edu

linkedin.com/in/dhruvkamaleshkumar | github.com/DB-25 | Portfolio - **Streamlit App**

AI Developer with 3+ years of experience in mobile app development and a strong focus on Gen AI, LLMs, and prompt engineering. Led high-impact projects like the GenAI Sandbox for the Massachusetts government and built 20+ apps across Google and Apple platforms. Graduating with a Master's in AI in Dec 2024, targeting Gen AI roles starting Jan 2025

EDUCATION

Northeastern University, Boston, MA   **Expected December 2024**
Khoury College of Computer Sciences - M.S. in Artificial Intelligence (GPA - 4.0)
- Key Coursework: Pattern Recognition and Computer Vision, Algorithms, Artificial Intelligence, Programming Design Paradigm, Linear Algebra and Probability, Human-Computer Interaction

Acharya Institute of Technology, India   **July 2022**
Bachelor of Engineering in Computer Science

SKILLS

- **Programming Languages**: Python, Dart, C, C++, Java, SQL
- **Generative AI & NLP**: LLMs, Prompt Engineering, RAG, Guardrails, Deep Learning, NLP, Langchain, OpenAI, SpaCy, NLTK, TensorFlow, PyTorch, SkLearn, Keras, Pinecone, Agents
- **Cloud & AI Deployment**:
  - AWS (Bedrock, Lambda, S3, Kendra, DynamoDB, CloudFormation, EC2, Cognito)
  - Serverless Architecture
  - Security Best Practices (OIDC, Azure AD to Cognito Integration)
- **Development Tools**: Flutter, Git, REST API, Streamlit, Jupyter Notebook
- **Databases**: MySQL, MongoDB, Firebase

EXPERIENCE

The GovLab, Boston, MA   **July 2024 - Present**
Generative AI Engineer (Gen AI, AWS, RAG, LLM)
- Oversaw 7 Massachusetts state generative AI projects, including direct responsibility for the GenAI Sandbox and EOED initiatives. Leveraged AWS Bedrock, Lambda, S3, Kendra, DynamoDB, and Guardrails to optimize project efficiency by 20% and reduce delivery time by 15%
- Managed and mentored 13 co-ops across these projects, providing technical guidance in Gen AI, LLMs, and RAG setups, which enhanced team competency by 40% and ensured timely delivery of all initiatives
- Implemented robust security measures and best practices within a serverless AWS architecture, ensuring compliance and protection of user data. Integrated SSO using OIDC from Azure AD to Amazon Cognito, streamlining authentication while enhancing the security and scalability of the tool

The Burnes Center for Social Change, Boston, MA   **January 2024 - June 2024**
Generative AI Product Development Co-op (Gen AI, AWS, RAG)
- Built a human-centered, highly scalable GenAI Sandbox for Massachusetts state employees, enhancing work efficiency and AI technology exploration, and presented the project to the Governor of Massachusetts
- Developed 'Smart Model' technology, designed to assist beginner users by automatically selecting the optimal model from 14 AWS Bedrock options (Titan, Claude, Jurassic, Llama, Mistral) based on the task. This innovation reduced operational costs by 40% while enhancing efficiency and user satisfaction, making advanced AI more accessible to non-experts
- Engineered and packaged a reusable framework for RAG setups, integrating CI/CD deployment in a full-stack serverless architecture, improving accuracy by 20% with a prompt engineering template and data preprocessing layer
- Spearheaded technical guidance and conducted code reviews across 5 state projects, increasing system efficiency by 25%

Acharya Institutes, India   **September 2021 - August 2023**
Lead Software Engineer (Flutter)
- Redesigned the app in Flutter for optimal performance across iOS and Android, reducing load times and memory usage, and supporting the university's ERP system for over 20,000 daily users

- Iteratively enhanced the app based on user feedback, raising its rating from 1.2 to 4.5 out of 5. Key improvements included better navigation, class schedules, and data reporting for students and faculty
- Successfully deployed the app to the Play Store and App Store, enabling the university's full ERP migration. Ensured high scalability and reliability, supporting thousands of users daily for critical tasks

WeSource Company, India   **May 2021 - September 2022**
Machine Learning Intern (Computer Vision, Machine Learning, Flutter)
- Developed a robust fraud detection system using Python, TensorFlow, and Scikit-learn, achieving an 87.34% accuracy rate and saving the company $20,000
- Applied techniques like Random Forest and SVM for classification and leveraged data preprocessing methods to improve model performance
- Redesigned and developed two cross-platform mobile apps using Flutter, incorporating user-centric design principles and state management tools like Provider, resulting in a 35% boost in sales and a 20% increase in user engagement

SM Creatives, India   **January 2022 - May 2022**
Software Engineering Intern (Flutter)
- Spearheaded development and deployment of 5 distinct cross-platform applications across multiple clients
- Enhanced a kids' educational app by making it more user-friendly and simple to use, driving a 20% increase in engagement among toddlers and preschoolers
- Improved application efficiency by 25% through advanced state management tools like Provider and GETX

RESEARCH / PROJECTS

ME (Gen-AI, RAG, Agents) ([Link](#))   **October 2023 - November 2023**
- Developed an interactive AI portfolio using GPT-4, Pinecone, and Streamlit, showcasing skills in creating real-time, personalized responses with modular AI agents
- Leveraged Retrieval-Augmented Generation (RAG) and vector similarity search for precise information retrieval, highlighting advanced AI capabilities
- Built a scalable, user-friendly interface with flexibility for future enhancements, positioning it as an evolving, all-in-one personal assistant

Plastic Usage Classification (Computer Vision) ([Video](#))   **March 2023 - April 2023**
- Engineered a precise product image classifier for plastic content into four categories using Python and PyTorch
- Boosted model performance by 30% through fine-tunning ResNet50 and MLP head integration, achieving an 88.77% F1 score with hyperparameter optimization via Optuna

Real-time 2-D Object Recognition (Computer Vision) ([Video](#))   **February 2023 - March 2023**
- Formulated and executed a real-time 2D object recognition on a white surface with a top-down camera setup
- Utilized C++ and OpenCV to Threshold, Segment live video feed to compute feature vector of significant regions
- Implemented and compared Nearest Neighbors and K Nearest Neighbors for the classification of over 12 items

Sequential Sentence Classification in Medical Abstracts (NLP, Deep Learning)   **March 2022 - July 2022**
- Established a Deep Learning NLP model to sequentially classify sentences into five categories
- Leveraged PubMed 200k RCT dataset for training and validation over the medical research paper abstracts
- Applied transfer learning using pre-trained embeddings for an 85% F1 score

EXTRACURRICULAR ACTIVITIES

AWS x RIOT VCT Hackathon | Finalist ([Video](#))   **September 2024 - October 2024**
- Created a generative AI Agent using Amazon Bedrock and LangChain to assist with esports management, answering complex queries about VALORANT players, and providing team-building support
- We used Retrieval-Augmented Generation (RAG) to pull contextualized data from Riot Games resources (1 TB)
- Deployed on AWS (Bedrock, Lambda, DynamoDB), ensuring scalability and efficiency, and achieved finalist recognition for innovative LLM and prompt engineering applications

Smart India Hackathon | State-Level Winner for Deep Learning Project   **August 2019 - September 2019**
- A state-of-the-art app leveraging deep learning and computer vision to detect driver drowsiness in real-time, ensuring safety for truck drivers