

Dipen Kumar Maheshwari

+1-780-708-5514 | dipenkum@ualberta.ca | linkedin.com/in/dipen-kumar | github.com/dipenkumarr | dipenkumar.vercel.app

EDUCATION

University of Alberta

Bachelor of Science Honors in Computer Science

Edmonton, AB

Expected: December 2026

Relevant Coursework: Software Engineering, Data Structures & Algorithms, Relational Database Management, Machine Learning, Artificial Intelligence

EXPERIENCE

Machine Learning Engineer Intern

SAP

Jan 2025 – Present

Vancouver, BC

- Engineered an end-to-end data processing and unsupervised **NLP** pipeline on **Azure Databricks** (PySpark) to process ~**38k** raw tickets reducing message volume by ~**85%** via **FAISS**-optimized semantic de-duplication.
- Implemented comprehensive data quality workflows using **NLTK/Spacy**, reducing noisy data by **15%** and significantly improving data consistency for model training.
- Leveraging **LLMs (GPT-4o/Gemini)** with structured Pydantic outputs to evaluate and categorize tickets, projecting for nearly **60%** in reduction in manual triage time.
- Prototyping a multi-format **RAG pipeline** using **LangChain** and **ColBERT**, featuring dual-indexing strategy, significantly reducing context contamination and boosting response relevance.

Software and Research Intern

University of Alberta - ANCL

May 2024 – June 2024

Edmonton, AB

- Implemented the **containerization** of autonomous navigation systems using **Docker**, reducing development environment setup time by **50%** and enhancing system reliability.
- Refactored core navigation systems in **ROS, C++, and Python** and established robust **Git and GitHub** version control workflows, significantly improving code maintainability, stability, and team collaboration.

PROJECTS

Hooked/ai | *FastAPI, TypeScript, Next.js, Python, Cohere API, Modal GPU, AWS S3, Inngest, SQLite* | [GitHub](#)

- Engineered a **full-stack** platform to automate podcast-to-short-form video conversion, reducing manual editing time by about **95%** for content creators.
- Architected a multi-stage **AI pipeline** using **WhisperX** for transcription, **Cohere Command A** for viral moment detection, and an Active Speaker Detection model for intelligent video cropping.
- Built a scalable, serverless backend using **Python (FastAPI)** and **Modal** for **GPU-accelerated** video processing, orchestrated via an **Inngest** event-driven queueing system.
- Managed storage and delivery via **AWS S3**; built a responsive **Next.js + TypeScript** frontend to track jobs and manage assets.

SpotLotto - Event Lottery System | *Java, Android Studio, Firebase, JUnit, Git, GitHub, XML, MVC* | [GitHub](#)

- Spearheaded a **6-member** Agile team as the **project lead**, owning the product roadmap, managing sprint planning, and resolving team conflicts to ensure on-time project delivery.
- Engineered the core backend in **Java** for a **multi-role (user/organizer)** system using Firebase, implementing key features like **ML Kit QR scanning** while also serving as the team's lead troubleshooter by **debugging** complex code and resolving critical Git merge conflicts.

UniWays - GHacks Winner | *Next.js, React, TypeScript, Prisma, PostgreSQL, Google Maps APIs* | [GitHub](#)

- Developed a **full-stack** campus navigation web application using **Next.js**, and **TypeScript**, achieving a **20% reduction** in estimated student travel time through **Google Maps API** integration and real-time route optimization.
- Designed **REST APIs** with **Prisma** and **PostgreSQL**, incorporating **OAuth** for robust user authentication and enhanced data protection.

TECHNICAL SKILLS

Languages: Python, TypeScript, JavaScript, Go, Java, C, SQL, HTML/CSS

Frameworks/Libraries: Next.js, React, Node.js, LangChain, Pandas, NumPy, PyTorch

Cloud & Tools: Git, GitHub, Docker, Azure Databricks, AWS, Firebase, PostgreSQL, MongoDB, Linux, Jira