

CHARANPREET NARULA

Software Engineer (Backend/MLOps)

+15193199026 • charanpreet9000@gmail.com • github.com/joy-narula • Vancouver, BC, Canada

Summary

To develop robust and innovative solutions in the field of AI and backend infrastructures which shape the future and build exciting, disruptive and impactful technology. My work style is collaborative and like a visionary who redefines what's possible, pushing boundaries and improving everyday.

Education

Western University, Ontario, Canada

Master's Degree (MEng) in Electrical and Computer Engineering (with specialization in Robotics and Control)

09/2022 - 12/2023

- **M.Eng. project:** Visual-Inertial Pose Estimation with landmark detection using 3D depth camera

Delhi Technological University, Delhi, India

Bachelors Degree (B.Tech) in Electronics and Communication Engineering

09/2017 - 08/2021

- **Capstone Project and Publication:** An Efficient Deep Neural Networks Based Framework for Automated Attendance System In University. Published a paper in PCCDS 2021 Conference.

TECH STACK

Programming Languages: C++ • C# • Java • Python • .NET

Distributed Systems: Kubernetes • Docker • Apache Spark • Hadoop

Databases: PostgreSQL • Azure Cosmo DB • Azure SQL DB • MySQL

Cloud Platforms: Microsoft Azure • AWS

Tools and Frameworks:

Pytorch • Tensorflow • OpenCV • ROS2 • Movelt • Scikit-learn • Pandas • NumPy • TensorRT • Jenkins • MLflow • HuggingFace • CUDA • Triton • Github • Apache Airflow • React • NextJS • Kafka • Redis • DeepSpeed

Experience

TikTok

Vancouver, Canada

Backend Engineer (MLOps)

10/2024 - 11/2024

Social Media App

- Worked in the Trust and Safety team, to detect and remove harmful content. Planned to worked on the **distributed model training pipeline using Kubernetes and DeepSpeed.**

Roboads Inc

Palo Alto, United States

ML Backend Software Engineer

03/2022 - Present

Mobile Advertising Robots

- **Trailblazing Robotic Advertising:** Lead architect behind the software for patented world's first Mobile Advertising Robot with a 200 kg payload and 3 digital signage screens, orchestrating an end-to-end development of a large scale backend and that positioned Roboads to multiple investments.
- **Conversational AI using LLM backend:** Implemented end-to-end MLOps Pipeline, Speech-to-text using Deepgram converted into embeddings using BERT, used **LLM based GPT-4o-mini and LLama model from HuggingFace with PyTorch** for response generation, and back to speech using ElevenLabs. The pipeline was containerized using **Docker and orchestrated deployment on Kubernetes cluster with GPU acceleration.** Integrated **Redis caching layers** to minimize response time and **Kafka for message queuing.** **Apache Airflow was used** for managing the **workflow using DAG.** Used **Azure Cosmo DB** for storing queries and **implemented B+ Tree structures** for faster query performance.
- **Machine Learning Operations (MLOps) Implementation for ads recommendation:** Established an MLOps to facilitate continuous training, integration, and deployment of Ad recommendation model, using **collaborative filtering on using Scikit-learn, pandas and numPy** to predict user preferences accurately. Containerized the application with **Docker** and orchestrated deployment on a **Kubernetes** cluster. **Used Apache Spark for data integration and real-time analytics.**
- **Enhanced Navigation & Human Interaction:** Amplified robot perception through the integration of VSLAM and sensor fusion (EKF), equipped with YOLO-based obstacle detection and human interaction features (head pose & eye tracking).
- **User-Centric Control:** Used 3D depth and Stereo vision for 3D reconstruction of maps to Point cloud. Performed camera calibration, to get better stereo matching and disparity computation results. **Sensor Integration and Data Fusion:** Used TensorRT to optimized models for Jetson Orin Nano, used model quantization and pruning to reduce model size and inference time, CUDA GPU pipelines programmed and memory optimized using C++

ConeTec

Burnaby, BC, Canada

Backend Software Engineer

01/2024 - 09/2024

Geotechnical and Mining Site Investigation

- Spearheaded the development of the project and led the team to successfully deploying the first **Semi-Autonomous Operated Robot for mining.**
- Using ROS and moveit to program the robot chassis movements and UR10e Robotic Arm with a Robotiq Gripper to pick and place objects employing deep reinforcement learning algorithms to optimize trajectory planning.
- Estimated the pose of miniature objects in a scene with different backgrounds and pose estimation CNN network, using TensorFlow and OpenCV. Used C++ and CUDA to deploy the algorithm on Nvidia GPU, using parallel processing and multi-threading and the use of TensorRT for GPU acceleration for increased performance.
- Designed a DAG based distributed Workflow for seamless and reliable flow of operations of the robot, node manager to monitor synchronization of all the nodes and using **Apache Tez / Apache Aitflow** for wotkflow management using **C# for backend service integration.**

Experience

Mirnah Technology Systems

Machine Learning Engineer

Riyadh, Saudi Arabia

08/2021 - 02/2022

Enterprise Solutions

- Deployed a microservice for **Real-time Loan approval for customers of Emirates National Bank in Dubai**, built an end to end pipeline for Training and Inference of the model using MLflow on over million data entries. **EDA was performed for data cleaning, feature selection and outlier removal. XGBoost model was deployed stacked with Random Forest. Optuna was used for hyperparameter tuning and SHAP analysis** to understand important features selected by the model. Created **REST End Points using FastAPI** and handling requests.
- Developed a smart product identification API using **Yolov5 Model on Pytorch, creating augmented datasets using Roboflow implemented a distributed computing environment using Hadoop and Spark for handling large-scale data.**
- Implemented an AI-driven facial recognition system for employee registration using deep learning models and advanced face recognition techniques. **Used High-Performance Computing using C++ for improved processing speed.**
- **Employed RNNs and transformer models for a Speech Recognition System, enhancing user interaction with mobile apps. Worked with SQL and NoSQL databases** for efficient data management and extraction.
- **Utilized Azure SQL DB for efficient querying and managing big data stored in distributed systems. Leveraged SparkSQL for data processing and analytics tasks on big data sets. Set up a robust End to End MLOps pipeline using MLflow, Docker, and Jenkins, streamlining the machine learning model lifecycle from development to deployment and monitoring.**

Padmini VNA Mechatronics

Computer Vision Engineer

Gurugram, India

06/2019 - 07/2020

Automobile Manufacturing

- **Machine Vision Experience: Developed and trained object detection (Yolov4) and semantic segmentation models UNet using TensorFlow to detect defects in auto parts, achieving a 95% accuracy rate with high IoU in quality inspection and a significant reduction in operational costs.**
- **Automated Optical Inspection:** Implemented traditional computer vision algorithms for AOI using OpenCV in C++. Algorithms involved, Hough Circle/Line Detection, De-noising, Morphological Transformation, Smoothing, Region detection using Contours, HSV based contour detection.
- **Multidisciplinary Collaborator & Communicator:** Detected 7 kinds of defects in controlled light conditions with 95% accuracy

Projects

Advanced Road Lane Finder for Mobile Robots at IGVC Competition

Michigan, USA

04/2019 - 06/2019

Developed a sophisticated lane detection system using LSTM networks to handle dynamic environmental conditions such as lighting and shadows, enhancing the robot's navigational accuracy. Used computer vision algorithms such as imaging warping and polynomial curve detection.

Data Acquisition Software for Formula Student UK Competition

Silverstone, UK

09/2017 - 04/2018

Designed a multi-sensor data acquisition system integrating machine learning models to analyze and predict vehicle performance metrics, leading to data-driven decision-making in race strategies. Used protocols like I2C, SPI, UART. Fostered data-driven decision-making in race strategy, sensors includes, Accelerometers, IMU, Strain Gauges, Engine Throttle Sensor, Temperature sensors.

Key Achievements

💎 Placed in the Top 10 international teams in Intelligent Ground Vehicle Competition 2019, Michigan, USA

First team ever to represent the University in the international competition and achieved commendable results.

💎 2nd Place in Formula Bharat National level Competition and placed in the top 30 international teams in Formula Student UK 2017 international competition.

Contributed to a young team of outstanding engineers, purpose driven and problem solvers

Publications

Springer books under the Conference - International Conference on Paradigms of Communication, Computing and Data Sciences, 2021

05/2021 - Present

An Efficient Deep Neural Network based framework for Real-Time Face Recognition at Scale

Charanpreet Singh, Rahul Thakur, Harshit Singh

Training / Courses

Generative AI Nanodegree (Focussed on NLP and LLMs) — Udacity