

HIMANSHU PAHADIA

☎ +1 480-791-6581

✉ himanshu.pahadia@gmail.com

🌐 [pahadiahimanshu](https://pahadiahimanshu.com)

🏠 himanshupahadia.me

Education

Arizona State University

Master of Science in Computer Science

Aug 2021 – May 2023

Tempe, Arizona

- **Thesis** - SKoPe3D: Synthetic Keypoint Perception 3D Dataset for Vehicle Pose Estimation
- **GPA** - 4.00 (with distinction); **Awards** - MORE Fellowship (Spring 2022), Graduate Scholarship (2022 - 2023)

IIT Delhi

Bachelor of Technology in Computer Science and Engineering

Aug 2014 – May 2018

Delhi, India

Technical Skills

Languages: Python, PySpark, Java, C#, Typescript, JavaScript

Technologies/Frameworks: PyTorch, Keras, Tensorflow, MLLib, OpenCV, Scikit-learn, Transformers, Langchain, Numpy, Pandas, Matplotlib, MLFlow, XGBoost, Django, Prophet, ROS, Flask, NLTK, NetworkX, Elasticsearch

Databases, DevOps and Cloud: MongoDB, MySQL, PostgreSQL, Azure ML Studio/Synapse, GCP, AWS, Docker, Git

Work Experience

Data Scientist, Boston Consulting Group, Boston

Jul 2023 – Present

- Conducted a robust assessment of centerline recommendations for a client by generating setpoint recommendations and analyzing uplift in key metrics on live tests. Improved the system to gain over 5% uplift in throughput.
- Designed and integrated a tabular chat agent that can ingest tabular data sources and answer queries via LLM models. Used Langchain and vector stores to integrate the RAG-based system. The system was showcased to 10+ clients.
- Deployed an FP-Growth Recommendation Engine using PySpark and MLLib to recommend frequently bought-together products for a major motorcycle manufacturer, successfully achieving a 3% jump in accuracy.

Consultant II (Data Scientist), Hitachi Vantara, Pune

Sep 2019 – Jun 2021

- Spearheaded the development of a road surveillance and analytics engine, **earning second prize in a national-level hackathon**. Fine-tuned a Yolo-based ensemble model on GRDD 2020 dataset and object detection model
- Developed a state-of-the-art solution for entity extraction from documents. Devised algorithms - spatial entity extraction, contextual OCR correction, checked-entity extraction - that resulted in 25% higher extraction accuracy.
- Forecasted query traffic on an in-house HR chatbot system by analyzing time series chat data using FB Prophet and spaCy. Successfully reduced almost 30% of tickets by incorporating more knowledge into the chatbot.

Associate Innovation Engineer, ZenAIR Labs, Zensar Technologies, Pune

Jun 2018 - Aug 2019

- **Filed two innovation patents** on multi-modal hand gesture recognition & unmanned aerial vehicle.
- Engineered custom Drive-by-Wire systems for steering, acceleration, and braking on a self-driving golf cart, **winning the Team Eureka Award 2019**. Collected real-time encoder data and trained the steering angle prediction ConvNet.
- Developed a visual analytics system for a retail client with age, gender, and emotion detection modules (VGGNet and ResNet). Devised a bias correction matrix to remove facial expression bias, improved age prediction accuracy by 7%.

Projects

SKoPe3D: Synthetic Keypoint Perception 3D Dataset for Vehicles (Thesis)

Aug 2022 - Mar 2023

- Created a one-of-its-kind synthetic keypoint perception 3D dataset from roadside perspective with 4.9M Keypoints, 151K 3D/2D bounding boxes for vehicles. Achieved exceptional 84.13% of correct keypoints (avg) using a Keypoint R-CNN model. Paper published in IEEE ITSC 2023.

ARGOS Vision - Active Perception Group, ASU

Aug 2021 - May 2022

- Designed a scalable object detection and trajectory mapping system for vehicles at intersections using Yolov5, NetworkX and Kafka. Integrated a video stabilization module using sparse optical flow, securing a 5% improvement in accuracy.
- Implemented a monocular 3D object detection model that uses multi-bin pose estimation and tight 2D/3D constraint, resulting in a 0.58 AOS score on the KITTI dataset (80-20 split). (MORE Fellowship)

Analysis of CGM time series data, Data Mining, ASU

Sep 2021 - Nov 2021

- Implemented and fine-tuned various classification algorithms, such as KNN, SVM, and Logistic Regression, as well as supervised-clustering algorithms like K-Means and DBScan to work on time-series data analysis of CGM data. Extracted features using Fourier and power transform, achieving a remarkable 70% accuracy that exceeded the baseline of 60%.

Movie Recommendation Engine, ASU

Oct 2021 - Nov 2021

- Trained a collaborative filtering model using matrix factorization, K-means, Spectral Clustering and XGBoost on the MovieLens 100K dataset. Achieved 0.771 RMSE using XGBoost on 80-20 splits.