

# ATHARVA YEOLA

+1 8582203508 | [ayeola@ucsd.edu](mailto:ayeola@ucsd.edu) | [in/atharvayeola](https://www.linkedin.com/in/atharvayeola) | [/atharvayeola](https://github.com/atharvayeola) | [/atharvayeola](https://www.atharvayeola.com)

## EDUCATION

### University of California San Diego (UCSD)

Master of Science (M.S) – Machine Learning & Data Science

Sep 2023 – Present

San Diego, USA

### Pune Institute of Computer Technology (PICT)

Bachelor of Engineering (B.E.) – Electronics & Telecommunication

Aug 2019 – Jun 2023

Pune, India

## WORK EXPERIENCE

### Data Scientist

Jul 2024 – Present

Qualcomm Institute - UC San Diego Division of Calit2

San Diego, USA

- Spearheaded the **causal analysis** of U.S. sports betting's impact on gambling addiction via **synthetic control** and time series forecasting; quantifying a 23% nationwide increase in addiction.
- Engineered 'Waldo', an open-source NLP tool by **fine-tuning** the **RoBERTa LLM** on 360,000+ Reddit posts to detect Adverse Events with **99.7% accuracy**, surpassing GPT-3.5 Turbo by 57% in F1-score.

### Research Intern - Computer Vision

Dec 2022 – Aug 2023

Indian Institute of Technology (IIT) Patna

Remote

- Created a **synthetic dataset** using image augmentation, expanding training data to approximately 100,000 images (**30% increase**), improving model robustness to detect occluded traffic signs in autonomous vehicles.
- Achieved **85.3%** accuracy using an ensemble model (ViT + EfficientNet) by tackling edge-case scenarios in adverse conditions, **outperforming the state-of-the-art** F1 score by 9%.

### Data Scientist - Recommender Systems

Aug 2022 – Apr 2023

RhythmFlows Solutions Pvt. Ltd

Pune, India

- Led **A/B testing** of a hybrid music recommendation system; proposed model (treatment) resulted in a **33% higher** click-through rate (CTR), **20% longer** session duration, and a **25% increase in app adoption** among target users.
- Deployed** the recommendation system into the live application using **AWS Sagemaker** with continuous updates for improved performance, collaborating cross-functionally to ensure seamless integration and scalability.

### Data Scientist - Product

Jan 2022 – Jun 2022

Atomic Loops

Pune, India

- Spearheaded a team of three** in building a YOLOv5-based real-time object detection model for an automated food delivery conveyor belt, achieving **98%** detection precision and **increasing customer footfall by 30%**.
- Collaborated** with the hardware engineering team to synchronize the belt's speed with the model's processing time using Raspberry Pi, **reducing false negatives by 15%**; later reused for inventory tracking, **reducing food waste by 20%**.

## PUBLICATIONS & CONTRIBUTIONS

- Authored** an original statistical investigation in JAMA Internal Medicine (acceptance rate: 4%) on gambling addiction trends post the legalization of U.S. sports betting.
- Led and **authored** a novel research study in IEEE Xplore (CVMI 2024) on occluded traffic sign detection.
- Co-developed** HIVTrends.org, an interactive public health analytics dashboard for monitoring HIV test sales.

## PROJECTS

### End-to-End Text Summarizer | Transformers, NLP, GitHub Actions

- Built a text summarizer for concise summaries, managing the entire workflow from configuration to deployment.
- Automated CI/CD** with AWS and GitHub Actions, handling Docker, EC2, ECR, and runner setup.

### Retail Vision Enhancement | YOLOv8, SuperGlue, Docker

- Implemented YOLOv8** for object detection, achieving over 90% accuracy in labeling on-shelf retail products.
- Introduced SuperGlue for precise product identification, and created a Dockerfile to ensure cross-platform replicability.

## TECHNICAL SKILLS

Programming – **SQL**, **Python**, R, Pandas, NumPy, SciPy, PySpark, Scikit-learn, dplyr, **PyTorch**, TensorFlow

Data Management – MySQL, **ETL**, **BigQuery**, Hadoop, Apache Spark, Periscope, Redshift, Google Analytics

Visualization Tools – **DBT**, RShiny, ggplot2, Plotly, Matplotlib, Seaborn, Pyplot, **Tableau**, PowerBI, Metabase

Cloud & Deployment – **AWS**, Azure, GCP, Git, **Docker**, Kubernetes, DataBricks, GitHub Actions, CI/CD

Other – **Probability**, **Statistics**, A/B Testing & Experimentation, **LLMs**, NLP, Machine Learning