

# Kewal Jayshankar Mishra

Worcester, MA | LinkedIn | Github | Portfolio

## EDUCATION

---

**Worcester Polytechnic Institute** Jan 2023 - Dec 2024  
M.S. Data Science GPA: **3.88**

**Relevant Courses:** MLOps, On-Device Deep Learning, Natural Language Processing, Statistical Methods

**Dwarkanadas J. Sanghvi College of Engineering** Aug 2016 - Oct 2020  
Bachelor of Engineering in Electronics and Telecommunication GPA: **3.50**

**Relevant Courses:** Neural Networks, Computer Vision, Database Management Systems, Object Oriented Programming

## EXPERIENCE

---

**GlaxoSmithKline** Collegeville, PA  
*Data Science Co-op* April 2024 – Present

- Developed a multi-label, multi-class neural network classification model for automating the gating process of high-dimensional Flow Cytometry data, achieving an F1 score of 0.97. Implemented robust class-balancing strategies to handle class imbalance.
- Engineered a cell segmentation model using **ResNet50** backbone coupled to a feature pyramid to identify cancerous cells.

**Worcester Polytechnic Institute** Worcester, MA  
*Artificial Intelligence Researcher (XTRA Sensing Limited.)* Jan 2024 – May 2024

- Built a **hybrid anomaly detection** method that combines **multimodal-feature** extraction and a transformer autoencoder for pump systems achieving an MCC score of 0.966.

*Graduate Research Assistant (Prof. Emmanuel Agu)* Jan 2023 – March 2024

- Led implementation and deployment of CNNs including RetinexNet, UNet to analyze the healing progress of wounds from pairs of wound photographs and thermal images. Migrated on-prem system to **AWS** and created a **MLOps** CI/CD pipeline.
- Built a Denoising **Diffusion** Probabilistic Model for data augmentation of the wound photographs and thermal image pairs.

*Machine Learning Researcher (Availity, LLC.)* Jan 2023 – Dec 2023

- Created a corpus of annotated medical notes in collaboration with EHR specialists and physicians. Benchmarked the performance of state-of-the-art solutions and developed a custom NER model for medical note parsing, with a 0.93 F1 score.

**Think360.ai** Mumbai, Maharashtra  
*Senior Data Science Associate* March 2021 – Dec 2022

- Developed a Dynamic Fleet Optimization and Real-time Monitoring System for a global logistics firm, enhancing operational efficiency across 47,000 containers and 160+ ships. Developed core optimization algorithms, Power BI dashboards for real-time monitoring, and machine learning models for demand forecasting and competition analysis, deployed on **Azure DevOps**.
- Developed a Credit Scoring Model using machine learning and alternate data, processing 50M+ customer IDs and identifying 230M+ banking domain incidents, enhancing lending decision accuracy for Algo360. Tuned efficient **SQL** queries.
- Led a team of 3 to develop a **time series** sales **forecasting** model for a U.S. CPG company, attaining a MAPE of 9.87%.

*Data Science Associate* Aug 2020 – Feb 2021

- Engineered cloud-native data solutions, integrating sources like EDGAR and GDELT. Crafted sophisticated data pipelines using **Scrapy** and NLP. Empowered users with insights into companies' standings and **competitive analysis**.

## ACADEMIC PROJECTS

---

**Generative AI with Large Language Models** Jan 2024 - March 2024

- Fine-tuned **FLAN-T5** for dialogue summarization, achieving a 22% improvement in **ROUGE** scores with full fine-tuning; demonstrated **PEFT**'s effectiveness with marginally lower metrics but enhanced efficiency.
- Leveraged **Proximal Policy Optimization** to detoxify FLAN-T5, reducing toxic content generation by 40% according to Facebook's hate speech model.

**Mobile NeRF: Real-Time On-device Neural Radiance Field** Aug 2023 – Dec 2023

- Optimized **MobileNeRF** framework using **tiny-cuda-nn** for real-time mobile operation, achieving 20 FPS on iPad with 93% size reduction and improved latency using **Knowledge Distillation** and structured **pruning** of channels and layers.

**Detecting Machine-Generated Text** Feb 2023 – May 2023

- Conducted an in-depth analysis on the detection of machine-generated text using BERT, pioneering the application of various finetuning techniques. Leveraging the HC3 corpus, achieved a 5% performance improvement over conventional methods.

**Digital Content Recommendation System** Dec 2020 - Mar 2021

- Developed a low-latency content recommendation engine for the Isha Foundation, a wellness NGO led by Sadhguru. Utilized a retrieve and re-rank strategy with MiniLM, achieving a response time of 2 seconds and a throughput of 100 requests/second.

## SKILLS

---

- Programming Languages:** Python, R, MySQL, Java, MATLAB, C++, C#
- Frameworks & Libraries:** Pyspark, PyTorch, TensorFlow, HuggingFace, cuDNN, JAX, TensorRT, Scikit-Learn, Flask
- Data Science:** Pandas, Numpy, Statistical Modelling, Feature Engineering, Dashboard Design, A/B test, Time Series
- Tools & Platforms:** AWS, Azure, Databricks, Git, CI/CD, Docker, Kubernetes, Airflow, Neo4j, Power BI, Tableau
- Natural Language Processing:** NER, LLMs, LLAMA, PaLM, GPT, Transformer, RAG