

# OMKAR MAHESH GARAD

New York, NY | 332-273-8613 | US Citizen | [omg22@cornell.edu](mailto:omg22@cornell.edu)  
<https://github.com/OMGarad> | <https://www.linkedin.com/in/omkargarad/> | [omgarad.github.io](https://omgarad.github.io)

## EDUCATION

**Cornell University**, New York, NY Aug 2023 - May 2025  
*MS in Information Systems* | GPA: 3.76  
William & Barbara Dahl Master's Fellow, Merit Scholar

**BITS Pilani**, Goa, India Aug 2019- May 2023  
*Bachelor of Engineering in Computer Science* | *Minor in Data Science* | CGPA: 8.52/10  
Merit Scholar

**Relevant Coursework:** Deep Learning, Computer Vision, Machine Learning, Artificial Intelligence, Foundations of Data Science, Natural Language Processing, Applied Statistical Methods, Data Structures & Algorithms, Database Systems

## TECHNICAL SKILLS

**Languages:** Python, C++, C, Java, SQL

**Tools and Frameworks:** PyTorch, TensorFlow, Pandas, Azure Databricks, AWS, GCP, Spark, Kafka, Hadoop, Docker, Git, MLOps (MLflow, CI/CD), Tableau, Agile Methodologies

## EXPERIENCE

**Honeywell, AI/ML Intern** May 2024 - Aug 2024

- **Integrated** ML predictions into smart-building controls, boosting energy savings by 35%
- Built an automated Delta Lake pipeline in Azure Databricks, generating features for 1M+ time-series records
- Optimized input windows to deliver average macro-F1 scores of xLSTM: 97%, LSTM: 95%, 1-D CNN: 90%, DNN: 89%, XGBoost: 87%, RF: 83%
- Implemented MLflow for experiment tracking and artifact/version management, ensuring full reproducibility

**UGent LT3, NLP Research Intern** Aug 2022 - Jan 2023

Research group at Ghent University that conducts fundamental research on different aspects of Natural Language Processing

- Developed a joint-learning system for complex and short-context multilingual named entity recognition (NER)
- Fine-tuned multiple language-specific XLM-RoBERTa models, a transformer-based LLM, to detect the domain of named entities after detecting the span, and deployed domain-based XGBoost models to identify fine-grained tags
- Obtained **average macro-F1 scores** for English: **80.3%**, Hindi: **76.3%**, Farsi: **73%**, Bangla: **71.6%**
- Created custom data preprocessing scripts to handle large-scale text datasets in the above languages effectively

**IBM, Machine Learning Intern** May 2022 - Aug 2022

- Engineered an end-to-end ML pipeline to ingest, clean, and feature-engineer large-scale GSI engagement datasets
- Trained and tuned an XGBoost classifier in scikit-learn, achieving an 86% F1-score for revenue-opportunity prediction
- Collaborated with a cross-functional teams and presented PoC demos to senior leadership, securing buy-in for production rollout
- Integrated SHAP explainability to generate feature-importance insights, driving data-backed stakeholder decisions

## PROJECTS

**Personalized Nutritional Deficiency and Dietary Advisory System** Aug 2024 - Dec 2024

- Fine-tuned T5 with FAISS-backed RAG (LangChain) in an AWS Bedrock pipeline, serving sub-second personalized nutrition guidance via a Streamlit front-end

**Enhancing Histopathology Classification** Aug 2024 - Dec 2024

- Built a PyTorch pipeline integrating Gigapath embeddings with attention pooling and contrastive learning, achieving a 5% higher AUROC and 4% higher F1-score than CTransPath and ViT-WSI on TCGA and SICAPv2 histopathology datasets.

**Keyphrase Extraction from Hindi Texts** (PyTorch, Pandas, Google Colab, Matplotlib) Dec 2021- May 2022

- Applied multilingual knowledge distillation to train a student XLM-RoBERTa transformer model to align English and Hindi sentence embeddings from a monolingual English teacher distillBERT transformer model
- Extracted keywords with student model using cosine similarity between the document and N-gram word embeddings