

Omer Tafveez

• +1 (734) 596 6185 • ometaf@umich.edu • linkedin.com/in/omer-tafveez • github.com/omertafveez-2001

Education

University of Michigan - Ann Arbor

Aug 2025 - May 2027

Ms. Information Sciences - Big Data Analytics

Courses: Foundations of LLMs, Advanced Databases, Web Development, Applied Parallel Computing, Data Mining

Lahore University of Management Sciences

Aug 2021 - May 2025

B.Sc. Economics-Mathematics, Minor: Computer Science - CGPA: 3.59

Courses: Data Science, Convex Optimization, Generative AI, Deep Learning, Advanced Machine Learning, Econometrics II

Professional Experience

Data Science Intern

Jul 2023 - Aug 2023

Lipton Teas & Infusions | Python, GeoPandas, Sklearn, Google Distance API, XGboost, SQL

Karachi, Pakistan

- Performed EDA on delivery logs using SQL, uncovering **~20% idle time and ~30% underutilized fleet capacity**.
- Applied K-Means clustering to segment distributors by distance, vehicle capacity, and delivery frequency for route grouping.
- Trained an XGBoost regressor ($R^2 \approx 0.87$) to predict delivery times within clusters, enabling dynamic daily route scheduling.
- Reduced average delivery time by ~18%, cut transportation costs by ~22%**, compared to the fixed-route baseline.

Data Science Intern

Jun 2022 - Jul 2022

Toyota | Python, Sklearn, Pandas, NumPy, Machine Learning, Econometrics

Karachi, Pakistan

- Achieved 87.3% accuracy** using Ridge Regression to model the impact of T-Bills, inflation, and SBP policy rates on prices.
- Pinpointed key sales drivers with Decision Trees, helping address the underperformance of **5 key vehicles**.
- Ensured stationarity of data via EMA and differencing, validated with ACF plots and ADF tests, improving **forecast accuracy by 18%**.

Research Experience

Sparse-Autoencoder Guided Fitnets

Mar 2025 - Apr 2025

Python, Pytorch, CNNs, Algorithmic Design, Distillation

- Integrated Sparse Autoencoders into FitNets, steering teacher representations via decoder-weighted feature injection, improving ResNet accuracy by **+2% in-distribution** and **+11% on OOD datasets** (Style-Transfer Animals100, TinyImagenette).
- Implemented Top-k SAEs to reduce polysematicity in teacher features, producing more neurons that respond to well-defined concepts.

Decoupled Gradient Knowledge Distillation

Nov 2024 - Mar 2025

Python, Pytorch, CNNs, Vision Transformer, Algorithmic Design, Distillation

- Engineered gradient-based decoupled distillation that maximizes MSE between the gradients of target and non-target class losses.
- The method acts as a **regularizer**, mitigating overconfidence, encouraging fidelity, and accelerating feature compaction.
- Achieved **6% increase** and up to **15% gain on OOD data** over standard decoupled distillation across ViT, MobileNet, ResNet, and VGG.

Key-Driven Grouped Query Attention

Jun 2024 - Aug 2024

Python, Pytorch, Transformers, HuggingFace, GPT, LLaMa, LLMs, Algorithmic Design

- Built "Dynamic Grouped Query Attention" with window-based norm updates, assigning queries to keys by changes in key norms.
- Implemented EMA-based key-norm evolution, smoothing rapid changes and reducing noise, **leading to a +3% accuracy gain** on ViT-Large across CIFAR-100, TinyImageNet, and Food101. Further Evaluated on GPT-NEOx, GPT-2, and T5.

Technical Projects

Transformers for Time-series: A Statistical View | Pytorch, Transformers, Gluonts, Statsmodel, SciPy, Seaborn

- Exposed volatility and misclassifications in Autoformer forecasts by analyzing structural breaks beyond MASE; introduced a Trust Score using 3 OLS tests, where the series averaged **0.37** and **failed 2 of 3 reliability checks**.
- Analyzed segments using heatmaps and bar plots, revealing increased volatility and spurious detections leading to misclassifications.

Content Moderation using LLMs | Pytorch, Matplotlib, NLTK, Transformers, HuggingFace, NLP

- Deployed a fine-tuned LLaMA-2 with QLoRA using Uber's Ludwig for toxic content classification, with an **F1 score of 0.81**.
- Pre-processed 223,500 texts by normalizing, removing unnecessary punctuations, fixing encoding, and filtering emojis.

Mobility Pattern Mining: Explaining Urban Development | NumPy, Pandas, Leaflet, Sklearn, Microsoft Open Data, SQL

- Designed a Linear Regression model to approximate urban development in Beijing with an **R^2 of 0.69** and a **0.91 correlation**.
- Transformed Beijing into 8 zones (as regressors) using KMeans, and visualized each zone's stats with hexplots.

Technical Skills

Languages: C, C++, Python, SQL, TypeScript, R, HTML, JavaScript, CSS, Bash Scripting

Frameworks/Tools: NumPy, PyTorch, LangChain, OracleDB, Git, Sklearn, Pandas, CVXPY, TailwindCSS, ReactJS, ExpressJS, NodeJS, Statsmodels, SciPy, TransformerLens, MongoDB