Afiya Ayman

afiya.ayman17@gmail.com, Google Scholar | LinkedIn | Web

SUMMARY

- **Six+ years of research experience** in **developing deep learning and statistical models,** providing **data-driven solutions** for **real-world systems** in **collaboration** with **interdisciplinary partners.**
- **Computer Scientist** with a **PhD** majoring in **Machine Learning, Data Science, and AI**
- Led **NSF, DOE-funded ML research projects**, contributing to successful grant proposals and **a strong publication record**
- Experience in **building scalable ML systems** and **collaboration across disciplines**
- **Skilled** at **preprocessing** and **analyzing large-scale text and tabular datasets** to build robust machine learning models.
- **Research Focus:** Automated Machine Learning, Multi-task Learning, Data Science, AI for Social Good

EDUCATION

- ***PhD*** *in Information Sciences & Technology*, **Pennsylvania State University**, University Park, PA **Aug 2022 - Aug 2025**
- ***MS*** *in Computer Science,* **University of Houston**, Houston, TX **Jan 2019 - Aug 2022**
- ***BSc*** *in Computer Science & Engineering,* **CUET,** Bangladesh **Mar 2011 - Oct 2015**

RESEARCH EXPERIENCE

Pennsylvania State University – University Park, PA

Research Assistant **- Applied Artificial Intelligence Lab, Advisor: Dr Aron Laszka** **Aug 2022 - Aug 2025**

- Advanced **multi-task learning (MTL)** research by developing **task-affinity–driven grouping strategies**, improving gain prediction accuracy, and **outperforming baseline MTL gain prediction methods by a wide margin,** achieving up to **4x higher correlation with ground-truth gains** across computer vision, tabular, time-series, and transportation benchmarks.
- Designed **scalable MTL training pipelines**, including a **two-stage ensemble model leveraging** gradient dynamics to predict and optimize MTL performance.
- Applied automated MTL frameworks to **real-world public transit forecasting**, enabling data-driven decision support for high-impact transportation systems.

University of Houston – Houston, TX

Research Assistant **- Resilient Networks and Systems Lab, Advisor: Dr Aron Laszka** **Jan 2019 - Jul 2022**

- Led DOE- and NSF-funded ML research in transportation, energy, and security, producing **multiple peer-reviewed publications** and influencing transit agency operations.
- Built energy prediction models for electric and diesel transit fleets that reduced MSE by up to **33% compared to baselines**, informing fleet scheduling and routing.
- Developed **neural architecture search (NAS)** methods for transit ridership prediction, jointly optimizing prediction error and model complexity per task, achieving **9% prediction error reduction** over generally optimized models.
- Conducted **large-scale NLP and text-mining studies** (topic modeling, embedding analysis, anomaly detection) to uncover developer behavior and security risks, demonstrating expertise in language data preprocessing and analysis.

SELECTED PUBLICATIONS

- **Ayman, A.** et al. (2022). Neural Architecture and Feature Search for Predicting the Ridership of Public Transportation Routes. In *8th IEEE International Conference on Smart Computing.*
- Sivagnanam, A., **Ayman A.** et al. (2021). Minimizing energy use of mixed-fleet public transit for fixed-route service. In *35th AAAI Conference on Artificial Intelligence (AAAI)*.

TECHNICAL SKILLS & TOOLS

Programming Languages: Python, Java (basic), C++ (basic), SQL, Bash
Machine Learning & AI: PyTorch, TensorFlow, Keras, Scikit-learn, Transformers, Lightgbm, XGBoost, Sci-Py, NLTK, MLOps
Data Science & Analysis: Data preprocessing, visualization (Pandas, NumPy, Matplotlib), clustering, statistical analysis, data mining
Methods: Deep learning, multi-task learning, AutoML, neural architecture search, NLP, statistical modeling, embeddings, optimization, Bayesian optimization, Random Forests, Decision Trees, transfer learning, representation learning, performance analysis, exploratory data analysis, object-oriented development, distributed computing
Tools & Platforms: Git, Jupyter, VSCode, Linux, CUDA, AWS (basic), GCP (basic)
Computational Infrastructure: Docker, SLURM, Ray (parallel model training), MongoDB, High-Performance Computing (HPC) — managed large-scale experiments via SLURM batch scheduling and screen-based remote sessions and cluster job submission; parallelized model training pipelines using multiprocessing and threading

SELECTED PROJECTS

Machine Learning Research: Led multiple **NSF- and DOE-funded research projects** in **collaboration** with city agencies and transit operators, designed machine learning frameworks across diverse domains for solving real-world problems, significantly enhancing the accuracy and efficiency of computational tasks.

Data-driven Energy Optimization for Multi-Modal Transit Agencies (Project Summary):

- Designed and developed a framework for predicting energy consumption for various transit vehicle types using multi-month sensor data, outperforming classic learning algorithms (decision trees and linear regression) by ~ **33% in MSE reduction**.
- Achieved **<5% prediction error** for 6-hour trips by aggregating sample-level predictions across time-series segments.
- Built a decision tree–based **map-matching** module linking noisy GPS to road geometry with 90% accuracy, enabling **spatial features** such as elevation, road curvature, and distance.
- Trained models on multi-vehicle telemetry data (6 vehicle types~8 months), combined with **geospatial layers** (road network, traffic, and weather), prediction outcomes informed energy-aware routing and scheduling strategies for transit agencies in Tennessee.
- Integrated into **distributed ML pipelines** for real-time prediction, with reproducible training/evaluation workflows
- **Collaborated** with **data engineers and transit agency analysts** to translate research models into actionable insights.

Neural Architecture & Feature Search for Transit Ridership Prediction (Project Summary):

- Developed a neural architecture and feature search framework for route-specific ridership prediction using Automatic Passenger Count and weather data - jointly optimizing prediction error and model complexity
- Evaluated on real-world transit ridership data across ten routes, showing that route-specific neural network architectures and features outperform generally optimized models in prediction accuracy (**9% lower error**).
- Architectures optimized per route-task consistently yielded the best results compared to hand-designed baselines, demonstrating the benefit of customizing both model complexity and features to **spatial and behavioral patterns** in real-world transit systems.

Automated Multi-Task Machine Learning for Ridership Prediction of Public Transportation Routes:

- Proposed an efficient affinity-driven MTL framework that pre-selects task groups to maximize MTL performance gains.
- Achieved **15% lower prediction error** vs. single-task models and **~7% MSE improvement** over baseline MTL; validated on real-world transit ridership data, demonstrating consistent performance gains across task groups.

Data-Driven Detection of Anomalies and Cascading Failures in Traffic Networks:

- Developed an LSTM-based traffic prediction model to capture spatial-temporal dependencies in large-scale urban road networks, achieving high accuracy in predicting traffic speeds using real data from Nashville, TN.

Relevant Skills: *Python, Pytorch, TensorFlow, Sci-kit Learn, Deep Learning, Transformers, Embeddings, Docker*

Data Analysis Studies: Collected, integrated, and analyzed data from multiple sources to conduct exploratory data analysis (EDA), topic modeling, and statistical studies.

Impact of COVID-19 on Public Transit Accessibility and Ridership (Project Summary):

- Analyzed **3.3M+ transit boarding events** from Nashville and Chattanooga by integrating farebox, GPS, and telemetry data; performed temporal, spatial, and socio-economic analyses to assess ridership declines across demographics, locations, and time-of-day patterns.
- Identified **persistent COVID-19 impacts** on transit accessibility, informing transit agencies' strategies for equitable service restoration; findings published in *Transportation Research Record*.
- Findings informed adaptive strategies for agencies to prioritize equitable service restoration and plan for disruptions.

Smart Contract Security Awareness Analysis: Investigated security concerns and awareness in the smart contract developer community by analyzing Q&A discussions, blog posts, and associated source code from multiple platforms.

Bug Bounty Ecosystem Analysis: Examined Chromium bug bounty program data, including activity logs and rules descriptions, to characterize participant incentives, behaviors, and vulnerability reporting processes.

- Preprocessed large-scale text datasets (forums, blogs, code repositories), applying NLP methods (topic modeling, embeddings, clustering), demonstrating transferable experience for tokenization and representation learning.

Relevant Skills: *Python, NLTK, Matplotlib, Numpy, Pandas, Sci-Kit Learn*

Software & Systems Development Experience

- Developed **full-stack web and mobile applications** (restaurant management, real estate listing, online examination systems) using CorePHP, MySQL, JavaScript, HTML/CSS, Bootstrap, and jQuery.
- Built **ML-powered analytics pipelines** integrating Python, PyTorch, and TensorFlow with large-scale sensor, GPS, and weather data for automated transit data processing and prediction.

HONORS/AWARDS

UH NSM Alumni Scholarship (2021-2022)
GHC Scholar awarded Student Scholarship for attending **Grace Hopper Celebration (GHC'20)**.
Rising Star 2020 at **CRCS Workshop on AI for Social Impact, Harvard**