Afiya Ayman

Phone: (346)-7756249, afiya.ayman17@gmail.com, Google Scholar | LinkedIn | Web

SUMMARY

- **Six+ years of experience as an academic researcher,** developing **deep learning and statistical models**, **optimizing** machine learning training processes for efficiency and impact
- Research advances **fundamental questions in task grouping and transfer efficiency** while **delivering data-driven solutions for practical systems such as transportation and energy**.
- Experience in **building scalable ML systems** and **collaboration across disciplines**
- **Computer Scientist**, specializing in **Automated Multi-Task Machine Learning, Data Science, and AI**
- **Strong publication record**, featuring contributions to prestigious ML conferences and journals

RESEARCH FOCUS

- Automated Machine Learning, Multi-task Learning, Data Science, Artificial Intelligence for Social Good

EDUCATION

- ***PhD*** *in Information Sciences & Technology*, **Pennsylvania State University**, University Park, PA **Aug 2022 - Aug 2025**
  Advisor: Dr Aron Laszka, CGPA: 4.00/4.00
- ***MS*** *in Computer Science,* **University of Houston**, Houston, TX **Jan 2019 - Aug 2022**
  Advisor: Dr Aron Laszka, CGPA: 3.95/4.00
- ***BSc*** *in Computer Science & Engineering,* **Mar 2011 - Oct 2015**
  Chittagong University of Engineering and Technology, Bangladesh

ACADEMIC 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 a **two-stage ensemble model** leveraging gradient dynamics to predict MTL performance, enabling cost-effective MTL optimization.
- Designed two complementary gain-prediction approaches to estimate MTL performance before actual MTL training.
- 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 statistical analyses on cybersecurity datasets, uncovering behavioral and security patterns in smart contract development and bug bounty programs.

PhD DISSERTATION: *Automated Multi-Task Machine Learning and Its Application*

- Proposed approaches to automatically identify beneficial task groupings in Multi-Task Learning (MTL)—gradient-based affinity modeling and feature-based MTL gain prediction—improving transfer efficiency and generalization across domains.
- Demonstrated the applicability of affinity-aware MTL to real-world systems, including ridership prediction for public transportation, enabling data-driven, environmentally sustainable, and equitable mobility planning.

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 elevation and distance features.
- Models trained on multi-vehicle telemetry data (6 vehicle types across 8 months), road geometry, traffic, and weather; prediction outcomes informed energy-aware routing and scheduling strategies for transit agencies in Tennessee.
- Integrated into a distributed ML pipeline for real-time energy prediction

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 yielded the best results compared to hand-designed baselines, demonstrating the benefit of customizing model complexity and feature set to task-specific 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** than single-task models and **~7% MSE improvement** over baseline MTL. Validated on real-world public transit ridership data, demonstrating consistent performance gains across task groups.

Failure Dynamics and Anomaly Detection in Traffic Networks:

- Developed an LSTM-based traffic prediction model to capture spatial-temporal dependencies in large-scale urban road networks, achieving high accuracy (MSE $\approx 6.55\times10^{-4}$) in predicting traffic speeds using real data from Nashville, TN.

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

Data Analysis Research: 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+ Tennessee transit boardings** 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.

- 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 transit agencies to prioritize equitable service restoration for future 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, Statistical Modelling,*

TEACHING EXPERIENCE

Teaching Assistant, Pennsylvania State University, University Park, PA **Jan 2025 - May 2025**

- **Object-Oriented Programming** — Spring 2025:
  Assisted in instruction, grading, and student support for undergraduate programming course.

Lecturer, School of Science, Engineering & Technology, East Delta University, Bangladesh **Sep 2016 - Dec 2018**

- Taught foundational CS and engineering courses, including Programming, Data Structures, Discrete Math, AI, and OS.
- Designed course materials and mentored students on research and programming competitions.
- Founded the **EDU Computer Club**, which regularly organized intra- and inter-university coding competitions.

PUBLICATIONS

Conference

- Atefi, S., Sivagnanam, A., **Ayman, A.**, Grossklags, J., & Laszka, A. (2023). The Benefits of Vulnerability Discovery and Bug Bounty Programs: Case Studies of Chromium and Firefox. In *ACM Web Conference 2023 (WWW), pp. 2209–2219.*
- **Ayman, A.**, Martinez, J., Pugliese, P., Dubey, A., & Laszka, A. (2022). Neural Architecture and Feature Search for Predicting the Ridership of Public Transportation Routes. In *8th IEEE International Conference on Smart Computing (SMARTCOMP).*
- Sivagnanam, A., **Ayman, A.**, Wilbur, M., Pugliese, P., Dubey, A., & Laszka, A. (2021). Minimizing Energy Use of Mixed-Fleet Public Transit for Fixed-Route Service. In *35th AAAI Conference on Artificial Intelligence (AAAI).*
- **Ayman, A.**, Wilbur, M., Sivagnanam, A., Pugliese, P., Dubey, A., & Laszka, A. (2020). Data-driven Prediction of Route-Level Energy Use for Mixed-Vehicle Transit Fleets. In *2020 IEEE International Conference on Smart Computing (SMARTCOMP), pp. 41–48.*
- Basak, S., **Ayman, A.**, Laszka, A., Dubey, A., & Leao, B. (2019). Data-driven Detection of Anomalies and Cascading Failures in Traffic Networks. In *Annual Conference of the PHM Society 2019, 11(1)*.
- Polly, F. P., Shil, S. K., Hossain, M. A., **Ayman, A.**, & Jang, Y. M. (2018, January). Detection and Classification of HGG and LGG Brain Tumor using Machine Learning. In *2018 International Conference on Information Networking (ICOIN) (pp. 813-817). IEEE.*

Journal

- Wilbur, M., **Ayman, A.**, Sivagnanam, A., Ouyang, A., Poon, V., Kabir, R., Vadali, A., Pugliese, P., Freudberg, D., Laszka, A., et al. (2023). Impact of COVID-19 on Public Transit Accessibility and Ridership. *Transportation Research Record, 2677(4), 531–546.*
- **Ayman, A.**, Sivagnanam, A., Michael, W., Pugliese, P., Dubey, A., & Laszka, A. (2021). Data-driven Prediction and Optimization of Energy Use for Transit Fleets of Electric and ICE Vehicles. *ACM Transactions on Internet Technology (TOIT), 22(1), 1-29.*

Workshop

- **Ayman, A.**, Roy, S., Alipour, A., & Laszka, A. (2020). Smart Contract Development from the Perspective of Developers: Topics and Issues Discussed on Social Media. In *Financial Cryptography and Data Security: FC 2020 International Workshops, AsiaUSEC, CoDeFi, VOTING, and WTSC, Revised Selected Papers, pp. 405–422. Springer.*
- Sivagnanam, A., Atefi, S., **Ayman, A.**, Grossklags, J., & Laszka, A. (2021). On the Benefits of Bug Bounty Programs: A Study of Chromium Vulnerabilities. In the *Workshop on the Economics of Information Security (WEIS).*

Book-Chapter

- Wilbur, M., Sivagnanam, A., **Ayman, A.**, Samaranayake, S., Dubey, A., & Laszka, A. (2023, August). Artificial intelligence for Smart Transportation. In Y. Vorobeychik & A. Mukhopadhyay (Eds.), *Artificial Intelligence and Society* (book chapter). ACM Press.

Submitted Manuscripts

- **Ayman, A.,** Mukhopadhyay, A., & Laszka, A. (2025). "Ensemble Prediction of Task Affinity for Efficient Multi-Task Learning." Submitted (under review).
- **Ayman, A.** & Laszka, A. (2025). "Identification of Task Affinity for Multi-Task Learning Based on Divergence of Task Data." Submitted (under review).

Preprint

- **Ayman, A.**, Mukhopadhyay, A. & Laszka, A., (2023). Task Grouping for Automated Multi-Task Machine Learning via Task Affinity Prediction. *arXiv preprint* arXiv:2310.16241.
- **Ayman, A.**, Aziz, A., Alipour, A. & Laszka, A., (2019). Smart Contract Development in Practice: Trends, Issues, and Discussions on Stack Overflow. *arXiv preprint* arXiv:1905.08833.

CONFERENCE PRESENTATION, POSTER & TALKS

- **Ayman, A.** (2022). *Data-Driven Prediction and Optimization of Energy Use for Transit Fleets of Electric and ICE Vehicles.* Poster presentation at the **ICDS Symposium, Pennsylvania State University, Fall 2022**.
- **Ayman, A.,** Roy, S., Alipour, A., & Laszka, A. (2020). *Smart Contract Development from the Perspective of Developers: Topics and Issues Discussed on Social Media*. **Presented at the Financial Cryptography and Data Security (FC 2020) Workshops – WTSC**, Kota Kinabalu, Malaysia.
- **Ayman, A.**, Martinez, J., Pugliese, P., Dubey, A., & Laszka, A. (2022). *Neural Architecture and Feature Search for Predicting the Ridership of Public Transportation Routes*. Oral presentation at the **8th IEEE International Conference on Smart Computing (SMARTCOMP 2022)**, Helsinki, Finland.
- Ayman, A., Wilbur, M., Sivagnanam, A., Pugliese, P., Dubey, A., & Laszka, A. (2020). *Data-Driven Prediction of Route-Level Energy Use for Mixed-Vehicle Transit Fleets*. Oral presentation at the **6th IEEE International Conference on Smart Computing (SMARTCOMP 2020)**, Bologna, Italy.
- **Ayman, A.** (2020). *Data-Driven Prediction of Route-Level Energy Use for Mixed-Vehicle Transit Fleets*. Presented at the *Harvard CRCS Rising Stars Workshop: AI for Social Good, Center for Research on Computation and Society (CRCS), Harvard University.*

HONORS/AWARDS

- **Scholarship: Graduate Tuition Fellowship – Pennsylvania State University** (Fall 2022 - Summer 2025)
- **U.S. Permanent Residency** under **National Interest Waiver (NIW)** – recognized for nationally significant AI research
- **UH NSM Alumni Scholarship** (2021-2022) & (2022-2023)
- **GHC Scholar** awarded Student Scholarship for attending **Grace Hopper Celebration (GHC'20)**.
- **Rising Star 2020** at **CRCS Workshop on AI for Social Impact, Harvard**
- **Scholarship: Graduate Tuition Fellowship – University of Houston** (Spring 2019 - Summer 2022)

ACADEMIC SERVICE

- **Program Committee Member,** AAAI Conference on Artificial Intelligence (AAAI 2025)

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.

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,
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, 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

- Managed large-scale experiments via SLURM batch scheduling and screen-based remote sessions
- Parallelized model training pipelines using multiprocessing and threading

REFERENCES

Dr. Aron Laszka (PhD Supervisor)
Assistant Professor, Pennsylvania State University
Email: aql5923@psu.edu
Phone: 814-865-1551
Website: https://aronlaszka.com

Dr. Abhishek Dubey (Long Time Collaborator)
Associate Professor, Vanderbilt University
Email: abhishek.dubey@vanderbilt.edu
Phone: 615-322-8775
Website: https://abhishekdubey.bio

Dr. Amin Alipour, (UH CS Faculty and Collaborator)
Associate Professor, University of Houston
Email: maalipou@central.uh.edu
Phone: 713-743-7462
Website: https://www.aminalipour.com/