

---

**EDUCATION**

---

**School of Computer Science, Carnegie Mellon University: Rising Senior graduating in 2026****Concentration – Machine Learning**

---

**TECHNICAL SKILLS**

---

- Languages - Python, Java, C/C++, R, JavaScript, SML (Functional Programming Language), Protobuf, HTML, CSS
- IDE's: Visual Studio Code, PyCharm, IntelliJ, RStudio
- Others – Unix, SQL, Machine Learning, PySpark, PyTorch, CI/CD, APIs, RPC, Math Modeling, Probability, GitHub, Socnetv, LaTeX

---

**INTERNSHIP / WORK EXPERIENCE**

---

**Google SWE Intern - Workspace AI Platform Team (May 2025 - Aug 2025)**

Designed and developed end-to-end LLM-based topic modeling system to effectively assign topics to any given input document context. This enabled analysis of live user data without direct human review; expanded the capabilities of Google's internal Gen AI analysis service; and improve the quality of synthetic evaluation datasets for Gemini.

- Engineered a robust Python data pipeline to preprocess synthetic document context data, effectively removing noise and utilizing coherence classifier for a higher-quality dataset, and established a human-labeled benchmark subset.
- Implemented a comprehensive topic modeling workflow, covering topic generation, clustering, and assignment, developing colabs for full end-to-end execution.
  - Iteratively developed and optimized LLM prompts (zero-shot, one-shot, and few-shot variations) for topic generation, leveraging the Gemini API to produce candidate topics and descriptions from over 600 document contexts. Conducted rigorous human and automated evaluations to achieve high accuracy & consistency.
  - Refined and curated the generated topic list by applying embedding models (Gecko), KMeans clustering, and optimized LLM prompts to effectively select representative topic labels for each cluster.
  - Improved topic assignment efficacy through iterative prompt refinement and comprehensive evaluations (human and automated), selecting the most effective strategy for classifying new contexts.
- Designed and implemented a new RPC classification API within the analysis service, including Mendel flags, proto definitions, RPC handler logic (with C++), and Piccolo configs, followed by thorough testing.
- Used internal tools including Moma, Buganizer, Piper, Blaze, Critique, Codesearch, and Fig to manage code, documentation, feedback loops, and other tasks.

**Research Assistant – Robotics Institute, CMU (May 2024 – Jul 2024)**

Contributing to a project to enhance the physics-solving capabilities of large language models (LLMs).

- Developed a DSL format in YAML to represent various physics topics, such as forces and Newton's Laws problems.
- Simulated spring problems using the MuJoCo engine and employed the mujoco-py library to model and solve dynamic spring systems.

**ML Intern - Infinite Solutions (Jun 2024 – Aug 2024)**

Worked with data on equipment failures to develop a machine learning model aimed at predicting future issues and informing maintenance plans. My responsibilities included

- Extracting and cleaning the data using Python libraries such as pandas and numpy
- Splitting data into training and testing sets
- Developing the predictive model. Generated outputs and created files to feed the model's predictions back into ERP for the development of effective maintenance plans.

**Intern - Machine Learning Lab at University of Delaware (Jun 2023 - Aug 2023)**

Vision-based Neural Solver for Combinatorial Optimization over Graphs - Teaming with a PhD student, contributed to the development of a neural solver to determine if a given network is Non-Hamiltonian or Hamiltonian using vision-based machine learning techniques. Utilized keras, matplotlib, numpy and other ML related libraries

- Developed an algorithm to generate 3D embeddings for nodes from 2D adjacency matrices and represented edges as points proportional to their lengths.

- Graphed these coordinates using matplotlib to form 3D spherical point cloud representations.
- Edited the PointNet model to accept point clouds as inputs and classify networks as Hamiltonian or Non-Hamiltonian. Conducted training and testing of the PointNet model using point cloud datasets to evaluate its accuracy.

### **Research Intern – University of Delaware Computational Data Science Lab (Jun 2021 - Aug 2021)**

- Completed an independent research project on NLP and social network visualization.
- Developed algorithms and code to construct, visualize, and analyze social networks of Reddit forums using Python libraries such as gensim, pandas, requests, nltk, and numpy

---

## **OTHER IMPORTANT PROJECT EXPERIENCE**

---

### **Machine Learning – Model Compression with Pruning Techniques (Apr 2025)**

- Implemented and compared three pruning techniques—magnitude-based pruning, L1-norm filter pruning, and ThiNet pruning—on a GPT model with ~593K parameters to remove low-importance weights and filters
- Fine-tuned pruned models with binary masks to maintain sparsity and recover accuracy; plotted the Pareto frontier to visualize accuracy-sparsity trade-offs.

### **Deep Learning - Machine Translation Model for Low-Resource Language (Nov 2024 - Dec 2024)**

Developed a machine translation model to translate sentences from English to Chichewa. Tested several baseline models, including Helsinki-NLP's MarianMT, Cohere, and Meta's LLaMA before fine-tuning. Fine-tuned the MarianMT model by integrating LoRA (Low-Rank Adapters) to enable parameter-efficient fine-tuning & also implemented a sequence-to-sequence (Seq2Seq) model using LSTM-based architectures with an attention mechanism

- Utilized pytorch, transformers, peft, scikit-learn, pandas, and matplotlib libraries to implement and evaluate our translation models
- Used the Chichewa machine translation dataset from Hugging Face, preprocessed and split into training and testing data accordingly.
- Evaluated model performances with BLEU scores and WER (Word Error Rate)

### **Machine Learning - Logistic Regression and KNN Models on Weather (Apr 2024 - May 2024)**

Implemented and compared logistic regression and KNN (K-Nearest Neighbor) models to predict the weather summary (Cloudy, Clear, Foggy, etc) given various inputs such as temperature, humidity, wind speed, and visibility.

- Used PyTorch, scikit-learn, pandas, numpy, and matplotlib to implement models and visualize accuracy
- Randomly split the 11 years of hourly data (96k hours) into 80% training and 20% testing. Filtered out extra data and only used the 12 most popular weather outputs
- For logistic regression, implemented a single statistical model considering all weather type outputs together (multi-class with softmax) to determine the probabilities of a given input falling into each weather category.
- For KNN, optimized the value of k during training of the implemented model to maximize accuracy

### **Software Engineering - PrestoDB Array Splitting and Concatenation (Mar 2024 - Apr 2024)**

Resolved performance issues in PrestoDB's array constructor by enabling support for arrays exceeding 255 elements.

- Rewrote constructor logic to split large arrays into 200-element chunks and reassemble using ARRAY and CONCAT functions
- Created unit tests to validate output and committed changes to the open-source GitHub repository
- Project listed in the CMU course website Hall of Fame

### **Egyptian Ratscrew Game (Nov 2022 - Dec 2022)**

- Developed the popular card game Egyptian Ratscrew using standard Python and Tkinter, incorporating graphics and event-based animations.

---

## **MAJOR ACHIEVEMENTS / AWARDS**

---

- Qualified for AIME (American Invitational Mathematics Examination) every year of high school and scored 11
- Ranked 10th among all 11<sup>th</sup> graders in the National Math League High School Championship held across the US
- Received honors in Online Physics Olympiad Contest by placing top 40 among 340 teams across 45 countries
- Tied for 2nd place nationally in the 2021 Math League Press Contest
- Ranked in the top 5 on the PUMaC (Princeton University Mathematics Competition) Division B Geometry Section
- Selected all four years for Lehigh Valley ARML team (DE, NJ, PA) via highly competitive test.
  - I was part of top 10 (2022 Ice) and top 20 (2021 Lightning) national team finishes
- Participated in Putnam 2022 and was in the top 800