

# Yu-An(Aaron) Chen

412-214-2260 • [Personal Website](#) • [yuanche2@alumni.cmu.edu](mailto:yuanche2@alumni.cmu.edu) • [GitHub](#) • [LinkedIn](#)

## Education

**Carnegie Mellon University, GPA: 3.81/4.00**

**9/2020 - 12/2023**

B.S. in Statistics and Machine Learning • Minor in Computer Science

Pittsburgh, PA

## Work Experience

**Analytics Engineer | Insurify**

**2/2024 - Present**

*Insurance Comparison Startup (\$600M Valuation)*

Cambridge, MA

- Scaling departmental initiatives by developing new data pipelines/transformations/reports while ensuring timely and reliable delivery of all data products for strategic decision making
- Directing MLOps in real-time ad bidding (\$9MM+ monthly spend) and customer LTV estimation by improving feature engineering and post-processing methods, increased traffic volume by 65% in 5 months
- Standardized performance reporting for advertisers to offer strategic bid-up recommendations, contributed to an **eleven-fold** increase in company revenue in 2024
- Leading upgrade of revenue attribution framework by implementing robust Marketing Mix Models (MMMs), resulting in expansion of television advertising from \$100k to \$2MM+ monthly spend
- Managing 1000+ dbt models and 100+ Airflow DAGs by conducting quality assurance, developing cross-referential tests, and implementing refactors to improve quality of existing tech stack

**Teaching Assistant(s) | Carnegie Mellon University**

**8/2023 - 12/2023**

*Machine Learning Dept., Dept. of Statistics & Data Science*

Pittsburgh, PA

- Facilitated [Machine Learning with Large Datasets](#) (graduate/PhD-level, 80+ students) by developing assignments and exams, grading student work, hosting office hours, and directing recitations
  - Key topics: Spark, distributed ML, GPUs, hashing, deep learning optimization, dimensionality reduction
- Facilitated [Modern Regression](#) (senior/graduate-level, 230+ students) by grading student work, hosting office hours, providing feedback, and proctoring exams
  - Key topics: regression theory, model inference & diagnostics, feature & model selection, regularization

**Data Science Intern | Federated Hermes**

**5/2023 - 8/2023**

*Investment Manager (\$800B+ AUM)*

Pittsburgh, PA

- Engineered 120+ features from sales and CRM data (19+ million rows, ~10.5 GB) with aggregation and NLP methods to determine optimal client contact strategies to increase mutual fund inflows
- Built and automated data pipelines with PySpark on Databricks to streamline ML development, established benchmark model performance of 93.7% accuracy predicting purchase activity & suggesting next steps
- Segmented client base based on purchase history with regression and determined ideal combination of interactions across client types for maximizing conversion rates

**Machine Learning Intern | Behavior**

**5/2022 - 8/2022**

*Addiction Recovery via AI*

Pittsburgh, PA

- Built frameworks for evaluating craving-predicting ML models through F-score and K-fold cross-validation, facilitating context-specific model evaluation
- Parallelized hyperparameter tuning process of CNN and XGBoost models by incorporating multiprocessing, decreased evaluation runtime by up to 800%
- Employed and tested multiple feature engineering techniques (Gaussian Smoothing, Autoencoding, etc.) to enhance model performance, increased cross-validation accuracy by 3.5%

**Research Assistant | Learnlab**

**1/2021 - 5/2023**

*CMU Research Lab (School of Computer Science)*

Pittsburgh, PA

- Developed data pipelines & reporting involving NLP, mixed-effects modeling, and hypothesis testing to reveal associations between course material presentation methods and student performance
- Led the data analysis team of Podsie, an educational start-up, and analyzed the efficacy of its products under low-sample size constraints, collectively leading to renewed lab funding in both review cycles

## Skills & Personal Projects

**Programming Languages:** Python, R, SQL, C

**Tools/Frameworks:** dbt, Airflow, Databricks, Apache Spark, Redshift, AWS, sklearn, Hightouch, Stitch, PowerBI

**Certifications:** Financial Engineering, Risk Management, Finance/Financial Markets, AWS Technical Essentials

**Projects:** [Optimal ETF Allocation Solver \(R\)](#) • [Portfolio Exposure Attribution \(R\)](#) • [Neural Wavelets for Time Series Compression \(Python\)](#) • [Distributed ML with Big Data \(PySpark\)](#) • [Government Bias in Affordable Housing \(R\)](#)