

Emily Cheng

emilycheng.3@gmail.com • (908) 210-0521 • www.linkedin.com/in/emily-c-cheng • <https://emilycheng156.github.io/web-portfolio/>

EDUCATION

University of Maryland (UMD)

May 2026

B.S. Mathematics: Statistics and B.S. Computer Science, Minor in General Business

GPA: 3.81

Relevant Courses: Data Science, Data Structures, Algorithms, Databases, Linear Algebra, Probability & Statistics

TECHNICAL SKILLS

Languages: Java, Python (Pandas, Plotly, Tensorflow, Numpy), SQL, MATLAB, SAS, R, C++, HTML/CSS

Developer Tools: Tableau, AWS (CloudWatch, Lambda, S3, Redshift), Power BI, PySpark, Dataiku, Terraform, Jira

EXPERIENCE

RISE (Risk-Informed Solutions in Engineering) Lab at UMD

Aug. 2023 - present

Research Assistant

College Park, MD

- Built statistical models in MATLAB and Python integrating linear and circular storm variables (heading angle, wind speed, radius of maximum wind) to analyze tropical cyclone trajectories and landfall risk
- Conducted distribution fitting and dependence modeling to quantify uncertainty in storm behavior
- Presented findings to a non-technical audience at the UMD Research Symposium, effectively translating statistical results into intuitive insights and earning 3rd place for Best Presentation among 50+ student projects

Johnson & Johnson

May 2025 - Aug 2025

Commercial Data Science Intern

Titusville, NJ

- Designed data transformation pipelines in Dataiku (Python, SQL, AWS Redshift) to integrate multi-source datasets into the enterprise data warehouse
- Evaluated and optimized forecasting models using Prophet and custom anomaly detection, improving prediction accuracy by 8% and identifying speculative-buying trends that saved ~ \$800 K.
- Developed and optimized interactive Tableau dashboards, incorporating data-driven recommendations and visual trend insights to guide supply-chain decisions.
- Partnered with non-technical stakeholders to ensure dashboards met business requirements for global demand forecasting

MassMutual

May 2024 - Aug. 2024

Actuarial Intern

Springfield, MA

- Supported enterprise reporting by querying, cleansing, and validating data from relational databases
- Automated data validation and reporting for reinsurer communication, improving accuracy and streamlined workflows
- Collaborated with data engineers to integrate financial data pipelines aligned with corporate governance standards

Ticketmaster

January 2024

Software Engineer Intern

Reston, VA

- Developed a cloud-native data monitoring system in AWS to detect and resolve survey data exceptions across 400+ entertainment venues, enhancing data integrity and availability for analytics workflows.
- Engineered ETL-style Lambda pipelines in Python to process dead-letter queue (DLQ) messages, automatically identify failures, and stream real-time alerts via Slack, improving data pipeline reliability metrics
- Collaborated in an Agile (SAFe) environment, managing sprint tasks in Jira and participating in stand-ups Supported infrastructure deployment using Terraform and configured CI/CD workflows with GitLab CI YAML

PROJECTS

NLP-Based Statement Classification | *Python (Pandas, Scikit-learn, Numpy)*

- Built and tuned multi-class classification models (Logistic Regression, Random Forest, LightGBM) to predict statement reliability, achieving 79.7% accuracy.
- Developed an end-to-end ML pipeline with data cleaning, feature engineering (TF-IDF, OneHotEncoder), and model evaluation using GridSearchCV and performance metrics

Netflix Analytics Dashboard | *Tableau*

- Designed an interactive BI dashboard visualizing content consumption trends by genre, region, and release period
- Integrated and cleaned streaming data to enable self-serve analytics and insight discovery for data-driven decisions