# SIYING (ELLIE) YANG

Cambridge, MA

617- 821- 9334 | ellie.yang02siying@gmail.com| https://ellie-yang-siying.github.io/ | Expected Graduation: August 2025

## TECHNICAL SKILLS

- Python, R, SQL, Julia, C++, MATLAB, Tableau, Power BI, SPSS, LaTeX, ArcGIS, Google Analytics, Databricks, AWS
- **Natural Language Processing, Computer Vision, Time Series Analysis, Forecasting, Classification, Data Visualization, Multi-Modality, Optimization, Experimentation (A/B testing), Statistical and Causal Inference, Customer Analysis**

## EXPERIENCE

**Data Scientist:** *Boost 'fill-in-the-form' conversion by redesigning MassMutual's website navigation* — Cambridge, MA
*MassMutual / MIT Analytics Lab* — Fall 2024

- Cleaned over 2M data for observed customer website journey path data and finished **feature engineering**
- Applied **K-means** algorithm to identify 3 distinct customer groups; used **Optimal Classification Tree (OCT)** and **LIME** to interpret clustering results; guided journey orchestration strategies and presented to Head of Marketing Team
- Optimized product-specific journey paths through **counterfactual analysis**; **fine-tuned** a **RoBERTa-based NLP neural network** to extract sequential path embeddings and identify high-conversion pathways; increased conversion by 1.5x and ROI by $7M

**Data Scientist Intern** — Shanghai, China
*POIZON (Leading Chinese e-commerce Gen Z platform)* — Summer 2023

- Performed **ETL** process; Queried 100M data with **SQL**; developed **performance metrics framework**, including conversion efficiency and customer satisfaction metrics; established data standards and SOPs for product team
- **Segmented customer** using **K-means** clustering and purchase behavior model; generated **customer analysis** report discussing characteristics of different groups and purchasing routes; guided product design for each group
- Performed **A/B testing**, attribution and funnel analysis; implemented App page redesign strategies; increased CTR by 15%, lifted UV value by 0.5%, and boosted NPS by 13%; developed **Tableau dashboard** and presented to Head of Product Team

**Analyst, Part-Time Assistant** — Beijing, China
*Oliver Wyman* — Spring 2023

- Enhanced risk measurement for a bank by developing **time-series predictive warning models** with **LSTM** and **XGBoost**; engineered 120 features and conducted **feature selection** through SFA, permutation and business relevance
- **Forecasted** volatility of 10 interest rates over 3, 5, and 10 years; attained low average missing rate of 4%
- Enhanced model interpretability by analyzing historical trends of key predictive factors; created client-facing presentation

**Quantitative Research Intern** — Shanghai, China
*Shanghai J. Royal Asset Management Co., Ltd* — Summer 2022

- Built **time-series LightGBM classification** model to predict stock prices 3 ticks ahead, achieving out-of-sample precision of over 90%; generated trading signals and crafted high frequency T+0 **trading strategy**, with commission considered
- Designed automated back-test programs to execute trades based on predictions and strategy; **visualized** trading process including signals and returns; executed 20-day back-tests on 102 stocks, and 86 yielded positive returns

**Data Scientist:** *Estimate global resource needs for global clinical trial* — Cambridge, MA
*Thermo Fisher Scientific / MIT Capstone Project* — Feb 2025 - **Present**

- Develop ML models to develop global clinical development resourcing estimators for drug candidate portfolio

**Data Scientist:** *Improve retail media marketing in e-commerce sector by leveraging Gen-AI tools* — Cambridge, MA
*The Campbell's Company / MIT Generative AI Lab* — Feb 2025- **Present**

- Design and implement an AI chatbot capable of answering marketing-related user questions leveraging **LLMs** and **Retrieval-Augmented Generation (RAG)** framework

## EDUCATION

**MASSACHUSETTS INSTITUTE OF TECHNOLOGY, SLOAN SCHOOL OF MANAGEMENT** — Cambridge, MA
*Candidate for Master of Business Analytics (Data Science), Operations Research Center, August 2025* — 2024 - Present

- *ML Project:* Developed a **multi-modal stacking LightGBM** to enhance demand prediction in online marketplace; integrated enriched feature sets by extracting **text embeddings** using **TF-IDF, SpaCy, FastText** and **image embeddings** by **ResNet50**

**TSINGHUA UNIVERSITY** — Beijing, China
*Double Major: Bachelor of Science in Mathematics and Physics, Bachelor of Engineering in Energy* — 2020 - 2024

- Academic: 3.86 / 4.0 GPA, Academic Excellence Scholarship (top 5%)
- Leadership: President of Student Union's Learning Department; organized 20 events for 1300 students