# Jiaqi (Scarlett) He

hjqscarlett@gmail.com | 4479022527 | linkedin.com/in/jiaqihe-stats
Github: github.com/scarletthejiaqi | Personal Website: scarletthejiaqi.github.io

## EDUCATION

**University of Chicago**                                                                              **Sep. 2024 - Dec. 2026**
Bachelor of Science, *Statistics*                                                                              GPA: 3.83/4.00
**Courses**: Modern Methods in Applied Statistics (M.S. Course), Probability and Statistics, Bayesian Statistics

**University of Illinois Urbana-Champaign**                                                              **Aug. 2022 - May. 2024**
Bachelor of Science, *Statistics & Computer Science*                                                              GPA: 3.98/4.00
**Courses**: Machine Learning, Statistical Modeling, Time Series Analysis, Stochastic Process, Data Structure

## SKILLS

- **Programing Languages:** Python (Pytorch, Pandas, NumPy, Scikit-Learn), R, SQL, C++
- **Tools:** Excel, PowerPoint, Microsoft Office, PowerBI, Tableau, AWS
- **Technical Skills:** Natural Language Processing, Machine Learning, Statistical Modeling, Data Visualization
- **Interests:** AI Safety, Badminton, Table Tennis, Pickleball, Hiking, Baking, Cooking

## PROFESSIONAL EXPERIENCE

**AntGroup** | *Python*                                                                              **Shanghai, China**
*Data Science Intern*                                                                              *Aug. 2025 – Sep. 2025*
- Extracted and engineered 12140+ Alipay questionnaire data, segmented age-specific consumption patterns, built reproducible data pipeline to retrain consumption preference model quarterly
- Built econometric+ML pipeline to quantify AI's impact on Total Factor Productivity (TFP) across industries

**UChicago Data Science Institute** | *Pytorch, HuggingFace*                                                                              **Chicago, IL**
*Research Assistant*                                                                              *Jun. 2025 – Aug. 2025*
- Designed schema and led annotation of ~2.5k jokes; fine-tuned 5 LLMs for offense detection, achieving 94% accuracy
- Conducted error analysis on humor classification models, reduced false positives by 22% with data augmentation
- Benchmarked OpenAI few-shots, showed fine-tuned models 54% increase in F1 score, documented training pipelines

**Marmon Holdings – A Berkshire Hathaway Company** | *Python, SQL, PowerBI*                                                                              **Skokie, IL**
*Business Intelligence Intern*                                                                              *Jun. 2024 – Aug. 2024*
- Forecasted 110% warehouse overflow within 3 months with ARIMA time-series, prevented supply chain disruptions
- Automated POS data processing with a python pipeline, reduced manual workload by 75%
- Implemented an ETL process to centralize outsourced data, support cross-unit performance comparisons for leadership

**PricewaterhouseCoopers** | *Excel, Python*                                                                              **Shanghai, China**
*Core Assurance Intern*                                                                              *Jul. 2023 – Aug. 2023*
- Automated data cleaning and validation processes using Python, improving data quality, cutting auditor review time
- Built 5+ matplotlib sales anomaly visualizations, supporting audit investigation

## PROJECTS

**Nutrition Analysis Tool with Generative AI** | *Git, Shell, AWS*                                                                              *Jan. 2024 – Aug. 2024*
- Built iOS nutrition assistant chatbot (image recognition + LLM), AWS backend, deployed to 10+ beta users
- Implemented calorie-estimation model mapping predictions, validated against 500 manually labeled meals

**Interactive Voice Response System Optimization** | *Python*                                                                              *Mar. 2024*
- Trained Random Forest/XGBoost on 200k+ call logs, improved routing precision for agent transfer from 73% to 89%
- Designed new routing that cut extra escalations by 3%, reduced staffing needs and improved call center efficiency

## LEADERSHIP INVOLVEMENT

**Uchicago Table Tennis Club**                                                                              **Chicago, IL**
*Vice President, Treasurer*                                                                              *May. 2025 – Present*
- Enhanced club visibility by managing social media, resulting in a 30% year-over-year membership growth
- Led fundraising initiatives, generating over $2,000 to support club events and participation in NCTTA competitions