

# Justin Yoo

[justineyoo1.github.io/PersonalPortfolio](https://justineyoo1.github.io/PersonalPortfolio)

(919) 820-5290 | [jeyoo@unc.edu](mailto:jeyoo@unc.edu) | [linkedin.com/in/justineyoo](https://linkedin.com/in/justineyoo) | [github.com/justineyoo1](https://github.com/justineyoo1)

## EDUCATION

University of North Carolina at Chapel Hill

Aug 2023 – May 2027

B.S. Computer Science | B.S. Statistics

- **Relevant Coursework:** Discrete Mathematics, Calculus I-III, **Data Structures and Algorithms**, Systems Fundamentals, Introduction to Python Programming, **Object-Oriented Programming**, Computer Organization, Web Design and Development, Linear Algebra for Applications, **Algorithms and Analysis**

## TECHNICAL SKILLS

**Languages:** Python, SQL, JavaScript, Java

**Web Dev:** HTML, CSS, Flask, Streamlit

**AI/ML:** PyTorch, TensorFlow, Scikit-learn

**Data Tools:** Jupyter, Snowflake, dbt

**Libraries:** Matplotlib, Pandas

**Systems:** Git, Docker, Kubernetes, Linux

## EXPERIENCE

Red Hat INC. | IBM Subsidiary

May 2025 – Dec 2025

Data Science Intern

Raleigh, NC

- Developed a full-stack financial automation web app with Python, Streamlit, and Google Drive APIs, reducing reporting time by **90%** and automating **\$50M+** in engineering expense allocations across **8+** product lines.
- Built and deployed internal PDF processing tool to automate structured data extraction from complex PDFs for the Hybrid Commitment Spend team, saving stakeholder **20+ hours** weekly.
- Migrated Redshift data pipelines to Snowflake and implemented scalable **dbt** models, refactoring legacy Python ETL into maintainable **SQL** workflows.

AIMING Lab | UNC CS Dept.

Aug 2025 – Present

Machine Learning Research Assistant

Chapel Hill, NC

- Supporting research on adaptive machine learning systems, with a focus on agent generalization, alignment, and continuous learning under **Dr. Huaxiu Yao**
- Built and benchmarked **3** gradient-based token attribution strategies (Batch, Sequence, Combined) for small LLMs (GPT-2), visualizing token-level importance and assessing model interpretability.

UNC CS + Social Good | Project Team

Aug 2024 – May 2025

Backend Developer

Chapel Hill, NC

- Collaborated with a team of **6** developers to design and build a full-stack web application serving **30+** users.
- Developed backend scraping pipelines and RESTful APIs, automating data updates and reducing load time by **30%**.

## PROJECTS

SyllabusSync | FastAPI | Docker | Celery | Python | Postgres (pgvector) | OpenAI

- Developed a full-stack Retrieval-Augmented Generation (**RAG**) system for syllabus uploads and question answering, handling parsing, embedding, and retrieval; delivered answers in under **10s** on average, even with **50k+** chunks.
- Improved answer quality and reliability by introducing versioning, diversification, and scoped prompting, increasing precision by **50%** and reducing hallucinations by **two-thirds** while cutting LLM costs by **30%**.

ML Movie Recommender System | Scikit-learn | Flask | Python | Optuna

- Engineered **3** collaborative filtering algorithms (**SVD matrix factorization**, **user-based/item-based CF**) on **100K** ratings dataset with automated hyperparameter optimization using Optuna.
- Built scalable Flask REST API with TMDb integration serving real-time recommendations with confidence scoring, supporting datasets from **100K** to **20M+** ratings.

Premier League Match Predictor | RandomForest | FastAPI | Streamlit | Python | Pandas

- Built an end-to-end ML system predicting Premier League outcomes with **0.428 log-loss** on a **3** season dataset.
- Deployed a FastAPI + Streamlit app for real-time predictions with leakage-safe rolling statistics and Elo ratings.

## CERTIFICATIONS

MTA: Introduction to Programming Using Python | Microsoft

Feb 2022