Caroline Lui

cel243@cornell.edu | (847)-224-8150 | github.com/cel243

EDUCATION

Cornell University, College of Engineering, Ithaca, NY **Expected May 2022**
B.S., Computer Science
Major GPA 4.30/4.30 (4.28 overall)

Relevant Courses: Data Structures, Algorithms, Functional Programming, Natural Language Processing, Artificial Intelligence, Databases, Computer Vision, Machine Learning, System Security, Formal Verification

WORK EXPERIENCE

Google, Virtual Internship, *Software Engineering Intern* **Jun–Aug. 2021**

- Implemented a C++ server and client command-line tool that simplified user access to metadata for a YouTube Ads data repository. Reduced the need for users to get direct assistance from the data infrastructure team, saving SWE hours on both sides
- Developed a detailed design, prioritizing user needs and devising strategies for addressing them
- Finished the planned project ahead of schedule with time to implement bonus features
- Rapidly learned new technologies to automate ongoing testing and regular releases of the tool
- Created monitoring dashboard that reports latency, QPS, and error rates, as well as fires alerts on abnormal behavior

Google, Virtual Internship, *STEP Intern* **May–Aug. 2020**

- Collaborated with 2 other interns to design, create, and present full stack Recommendations Impact Dashboard application that helps customers visualize the long-term impact of GCP IAM Bindings Recommendations
- Designed data pipeline for automated dashboard data updates, involving 5 Google APIs
- Independently designed and implemented the database layer of the application using Bigquery, Java, and SQL

HCA Healthcare (Parallon), Nashville, TN, *Strategic Data Services, Assoc. Product Analyst* **May–Aug. 2019**

- Designed and developed a chatbot AI using Python and Rasa that decodes complex user messages to generate and run appropriate SQL queries, enabling efficient, intuitive interaction with insurance denials data
- Independently identified data quality issues in a key data table and developed a SQL tool to standardize the data, which reduced the table size by 90% and led to the identification of underutilized hospital resources
- Automated previously manual analysis of 30-million-record table to evaluate collections performance across facilities

PROJECTS

Book Club: Group Book Recommendation Search Engine, *CS 4300: Language & Info, Group Project* **May 2021**
bookclub-4300.herokuapp.com

- Collaborated with a team of five to conceptualize, design, and develop a full-stack book recommender that integrated diverse book and author preferences of a group using similarity scoring techniques
- Independently designed and implemented the recommender's backend data representation, using vector space modeling and machine learning techniques to condense ~20GB of Goodreads book review data into ~100MB
- Project was placed in the class "Hall of Fame," a distinction awarded to only 3 projects out of ~50

AI Music Generation, *CS 4701: AI Independent Project Practicum, Group Project* **Feb-May 2021**
tinyurl.com/ai-music-generation-report

- Designed and trained a variety of N-Gram language models and RNNs on classical piano music to compare probabilistic and neural approaches to music generation
- Our best model received a median score of 5/5 from survey respondents rating the likelihood that the music was human-generated

Story Cloze Test Commonsense Reasoning Classifier, *CS 4740: NLP, Group Project* **Nov. 2019**

- Created NN that predicted the ending of 4-sentence stories, exceeding highest course-defined accuracy threshold
- Designed and optimized pipeline to extract feature vector representation of the data, using sentiment scoring, word embeddings, and part-of-speech tagging

ACTIVITIES/LEADERSHIP

Cornell CIS Department, *Teaching Assistant for Algorithms (CS 4820)* **Aug. 2020–Present**
Cornell CIS Department, *Teaching Assistant for Discrete Structures (CS 2800)* **Aug. 2019–May 2020**
Cornell University Chorus, *Recruitment Chair* **Aug. 2020–May 2021**