

# Johann Lee

johannchrislee@gmail.com | Ithaca, NY 14850 | <https://github.com/xforeverjlx>

## EDUCATION

**Cornell University, College of Arts and Sciences**

Ithaca, New York

B.A., Major in Computer Science. **GPA:** 3.97/4.0

*December 2024 (expected)*

- **Honors:** Milstein Program in Technology & Humanity Scholar (25 per cohort), Dean's List Honors
- **Relevant Courses:** Object-Oriented Programming, Data Structures, Functional Prog, Backend Web Development

## TECHNICAL SKILLS

**Programming Skills:** Java, Python, C, OCaml

**Database Frameworks:** AWS, SQLite, MongoDB, Git

**Web Development:** Django, Flask, Nginx, JS, Vue.js, PHP

**Machine Learning:** PyTorch, TensorFlow

## WORK EXPERIENCE

**Bank of America**

Charlotte, North Carolina

**Software Engineer Intern**

*Jun - Aug 2023*

- Developed automated test suite for generative Large Language Models, reducing 800+ instances of human labeling
- Built MVP of enhancing Erica chatbot with LLMs, business unit estimated \$1Mil+ cost savings per year if deployed
- Working on improving NLP machine's word embeddings to be contextual, performance increase TBD

**Nextstep Health Tech**

New York City, New York

**Backend Software Engineer Intern**

*Jul - Aug 2022*

- Automated data processing pipeline of mental health app's 10,000+ check-ins, reducing analysis time spent by 500%
- Scripted analysis investigating driving factors of positive emotions, presented findings to 50+ people (*Python*)
- Developed methods in a test-driven way (random, glass & black box testing), with function and class documentation

**Alibaba**

Hangzhou, China

**Algorithms Team Software Engineer Intern**

*Dec 2020 - Jan 2021*

- Built an auto-regressive Recurrent Neural Network to predict sales across 80+ malls, preprocessed 3 years of data
- Achieved 5% test error predicting 2 weeks in advance, enabling preemptive staff deployment across malls

**Hubel Labs (Software R&D Lab)**

Hong Kong

**Software Engineer Intern**

*Jun - Aug 2020*

The following features were released to all users of a language learning app for children:

- Designed algorithms to personalize a user's vocabulary lists based on books they would like to read (*AWS*)
- Automated the generating of word-tracing worksheets, now customized to a user's vocabulary list (*OpenCV*)
- Implemented update-emails that periodically summarize a user's progress (*Python, Firebase*)
- Automated vocabulary collection from textbook scans (*OCR, NLP*), increasing amount of data by 300% (*MongoDB*)

## PROJECTS & INTERESTS

**MathSearch: Web-based PDF Search Engine** ([Github](#))

*Aug 2022 - Present*

Given a math equation, the website searches for PDF pages with similar math equations

- Built backend Flask + Nginx + Gunicorn listening ports for requests packets and data retrieval from S3 intermediary
- Developed data storing process and algorithms to determine similarity between query and database entries efficiently
- Currently working on message queuing (for concurrent users) and speeding up search through hierarchical clustering

**Freelance Project: Course Search Website for Cornell MCSA** ([Github](#))

*Jun - Sep 2022*

- Web-scraped 5 sites containing information about courses to aggregate data (*requests*), then cleaned data (*pandas*)
- Developed backend RESTful API for routing data about the courses to the front end (*Django*)
- Designed frontend interactive web application and implemented content search and sort (*Vue.js, HTML, CSS*)

**Interests:** Certified Ski Instructor, Tennis, Poker (beginner)