

MICHAEL D. SONG

Boston, MA | 650-804-9449 | songmichael11@gmail.com | github.com/songmichael11 | michael-song.vercel.app

EDUCATION

Northeastern University | Boston, MA

Sep 2024 – Expected Dec 2027

Candidate for *Bachelor of Science in Computer Science and Biology*

GPA: 3.97/4.0

Coursework: *Object-Oriented Design* | *Database Design* | *Fundamentals of Computer Science 1 & 2* | *Biostatistics*

Boston University | Boston, MA

Sep 2022 – May 2024

High School Dual Enrollment, 48 Credit Hours

GPA: 4.0/4.0

Coursework: *Linear Algebra* | *Intro to Computer Science 1 & 2* | *Cell Biology* | *Genetics*

TECHNICAL SKILLS

Languages: Python, SQL, TypeScript, Java, R, C, HTML

Frameworks & Libraries: React, Pandas, NumPy, Flask, Tailwind CSS, Tidyverse, Streamlit, Drizzle ORM

Tools & Technologies: Git, GitHub Actions, AWS (Athena & S3), MySQL, Postman, Jupyter Notebooks, PostgreSQL, Docker, Vercel, Figma, RStudio, LaTeX

EXPERIENCE

Project Lead & Software Developer | *Sandbox Software Consultancy*

Jan 2025 – Present

- Lead a team of **10** developers and designers to create a co-op review platform enabling **10,000+** Northeastern students to view and submit reviews for co-op roles utilizing the **T3 tech stack**
- Collaborate in an Agile environment, utilizing standups, weekly code reviews, and a **GitHub Projects** ticket board
- Extended **PostgreSQL** data model to include role and company requests using a **Drizzle ORM** with **TypeScript**, enabling users to request additional roles and companies
- Redesigned **7** mobile-responsive **React** UI components using **TypeScript** and **Tailwind CSS** to ensure full functionality on all devices

Data Analyst | *Kaerwell Inc*

Jul 2025 – Sep 2025

- Implemented and tested automated market basket analysis using **PySpark**'s FP-Growth algorithm to automatically recommend relevant products based on current shopping cart state
- Built an **AWS Lambda**-ready FP-Growth workflow leveraging hashed, pickled files in **S3** to reduce recommendation time from **~100** milliseconds to **<1** millisecond each

Data Analyst | *MedTech Analytics*

Jul 2024 – Apr 2025

- Architected a **Python** data pipeline with **AWS Athena** and **S3** integration to ingest **50M+** data points, automating DDL generation and execution for large-scale analysis of Duchenne muscular dystrophy (DMD)
- Developed **Presto SQL** queries to analyze **150M+** unique rows of healthcare data, identifying common symptom patterns and flagging **1,000+** misdiagnoses of Lambert-Eaton Myasthenic Syndrome (LEMS)
- Translated technical findings into clinical insights during biweekly meetings with healthcare professionals

Research Assistant | *National Institutes of Health*

May 2023 – Aug 2023

- Conducted an independent **Python** research project applying molecular fingerprinting algorithms and training regression models on **1,600+** drugs to predict half-lives from molecular structures using **Pandas** and **NumPy**
- Communicated complex findings by writing a **5,000**-word thesis and presenting to an audience of **300+** with data visualization generated using **Matplotlib**

PROJECTS

Consensus | *Python, MySQL, Flask, Docker, Streamlit, CSS, Scikit-learn, TensorFlow*

May 2025 – Jun 2025

- Built a full-stack economic policy proposal social media application that utilized regression models and deep neural networks (DNNs), enabling users to support their posts with real data and predictions
- Implemented **20 RESTful API** endpoints using **Flask** and **MySQL**, enabling complex user-post interactions and custom predictions from integrated machine learning models
- Created **85%** of the app's pages with **Streamlit** and **CSS**, ensuring a cohesive theme and user-friendly experience
- Designed the relational database schema and generated **MySQL DDL** scripts, structuring **20** interconnected tables
- Led team product planning through persona and user stories, wire-framing pages, and conducting a 20-minute demo

AI Bias Study | *R, RStudio, dplyr, ggplot2, readr*

Aug 2025

- Analyzed whether AI detectors were more likely to flag text written by non-native versus native English speakers as AI using **R** and **RStudio**