

Justin Lee

917-480-6998 | JustinYuilLee@gmail.com | [Linkedin](#) | [Github](#) | [Portfolio](#)

EDUCATION

Stony Brook University

Bachelor of Science in Computer Science ; Applied Math — GPA: 3.50

Stony Brook, NY

Aug. 2023 – May 2027

- **Relevant Coursework:** Graduate Data Analysis, Graduate Probability Theory, Analysis of Algorithms, Data Structures, System Fundamentals I/II, Software Engineering, Programming Abstractions, Computer Networks

EXPERIENCE

Software Engineer Intern

Acadia Analytics

May 2025 – August 2025

Cambridge, Massachusetts

- Enhanced **ML pipelines** for financial time series, cutting training time by **25%** and boosting accuracy **10%**.
- Deployed **RESTful API** with LLM agents + backtesting infra, automating reports and saving **40%** effort.
- Optimized **predictive models**, raising forecasting accuracy **15%** and scaling pipelines to 10,000+ records.
- Cleaned and scraped data for **50+** stocks into a **PostgreSQL** database, powering dashboards for analysts.

Software Engineer Intern

Desung Inc.

Jan 2024 – August 2024

New York City, NY

- Engineered behavior-aware game logic in **Unity**, improving puzzle completion rates by **12%** for active players.
- Implemented **scalable learning pipeline** improving puzzle-solving performance and overall efficiency by **25%**.
- Deployed **RESTful** backend services for multiplayer and analytics, cutting average response time by **120ms**.
- Led CI/CD pipelines, sprint planning, and **Dockerized** deployments with **GitHub Actions** and **Jenkins**.

PROJECTS

Clinical Trial Copilot | *Next.js, TypeScript, FastAPI, Cerebras, Anthropic Claude, PostgreSQL*

- Engineered AI pipeline converting **1,000+** trial PDFs into structured datasets, contradictions, and summaries.
- Integrated **Cerebras** for parallel PDF parsing, cutting clinical review time by **92%** across **50+** studies.
- Leveraged **Claude** to generate summaries, heatmaps, and forest plots, enabling comparison of **10+** trials at once.

Seawolf Course Finder | *AWS (EC2, RDS), PyTorch, scikit-learn, BeautifulSoup4*

- Developed AI course recommender with **TF-IDF + cosine similarity**, semantically ranking **4,000+** courses.
- Implemented **AWS**-based backend with RDS and EC2, handling **50K+ queries/month** with high reliability.
- Deployed **Flask** REST API for semantic search and course similarity, returning results in **<1s** latency.

Hive - Multi-Agent Conversational AI | *FastAPI, WebSockets, React, PostgreSQL, LangChain*

- Built chat platform orchestrating **5+ LLM personas** with multi-round discussions and diverse perspectives.
- Designed **PostgreSQL memory** with embeddings for semantic retrieval, powering **30+** cross-session chats.
- Optimized WebSocket concurrency + context pipeline, cutting prompt size **60%** and supporting **5+** sessions.

Bionic Reading Extension | *JavaScript, HTML, CSS, JSON, Chrome Extension APIs, DOM Manipulation*

- **Won SBU Hopper Hackathon** with Chrome extension improving reading accessibility for dyslexia/ADHD.
- Implemented Bionic Reading Mode with JS/HTML/CSS, boosting reading speed **15%** across **50+** webpages
- Optimized Chrome Extension APIs and **DOM handling** to process **1,000+ words/s** with **sub-100ms** latency.

TECHNICAL SKILLS

Programming: Java, Python, C++, SQL, .NET Core, JavaScript (React, Node.js, Angular), Spring Boot, REST APIs
Frameworks/Tools: GitHub, Docker, Harness, Flask, OpenShift, Azure, Google Cloud Platform, CI/CD pipelines, Cloud Security, Enterprise Data Systems, Agile

Data Engineering: Data modeling, ETL pipelines, MongoDB, MySQL, PostgreSQL, TensorFlow, NumPy, Pandas

Visualization/Analytics: Tableau, Streamlit, Jupyter, Matplotlib, Seaborn, Plotly

Other: AI/ML concepts, Generative AI, Prompt Engineering, Data Labeling, Backpropagation