Benjamin Chang

📞 (669)-291-0140
📍 Cupertino, CA

benjaminchang.dev • bchang26@jhu.edu

in benjaminchang7
○ benchang323

EDUCATION

The Johns Hopkins University **Baltimore, MD**
B.S. Computer Science, B.S. Applied Mathematics & Statistics (Dean's List) Expected 05/25
- **Focus Areas:** Software Engineering, Natural Language Processing, Statistics and Statistical Learning
- **Teaching (OOP – C/C++):** Guided 350+ students; Assessed 300+ assignments; Scripted 5+ autograding tests; Provided OH debugging support

PROFESSIONAL EXPERIENCE

Palantir **Washington, DC**
Software Engineering Intern 08/24 – Present
- Implemented a scalable, fault-tolerant retry queue system for high-volume real-time data processing using Java, Spring Boot, and Kafka, boosting system reliability by 27% and reducing user wait times by 19% through distributed message queuing and asynchronous execution
- Redesigned fault-tolerant real-time data pipeline using dynamic configuration and microservices, reducing configuration errors by 34% and setup time by 26% while implementing horizontal scaling, load balancing, and parallel processing for high-throughput distributed handling
- Improved system architecture by implementing priority-based job scheduling, streamlining API endpoints, and smart task cancellation, resulting in 15% improvement in reliability and performance through enhanced error handling, resource allocation and load balancing

Intel **Hillsboro, OR**
Open-Source Software Engineering Intern 05/24 – 08/24
- Engineered 5+ functional testing tools (screen record/capture, render diff, etc.) for 3D graphics benchmarking and optimization using Python (OpenCV, FFmpeg, PIL, MP) and C++ (Vulkan, OpenGL, Kernel Interface) with Mesa Graphics drivers, increasing testing coverage by 47+%
- Designed a synchronized data collection system mapping frames with I/O events, creating a 10k+ input database for training gameplay scripts
- Contributed to a semi-supervised ML system using RLHF and transfer learning with PyTorch, performing feature engineering and data augmentation to automate gameplay script generation, reducing time by 70%, increasing accuracy by 30%, and expanding test suite by 10+
- Integrated 2 profiling tools into the functional suite enhancing shader profiling techniques (runtime analysis, shader execution, etc.) using RenderDoc and Perf, expanding API record/replay for improved performance analysis, optimizing the graphics rendering pipeline by 15+%

Scale AI **San Francisco, CA**
Artificial Intelligence Model Trainer 01/24 – 05/24
- Crafted 100+ high-quality training datasets to enhance sophisticated AI models to accurately understand and process complex math concepts
- Evaluated and ranked model responses for accuracy, relevance, and factuality using statistical scripts, boosting model capabilities by 12+%

Google **Mountain View, CA**
Research Scholar 09/23 – 12/23
- Gained insight into the R&D cycle, emphasizing on CI/CD pipelines, code QA, and the reliability/performance of multi-distributed ML systems
- Engaged in 10+ panels on computing research, delving into ML system architecture design, optimization, and scalable software solutions

StudyFind **New York, NY**
Software Developer (Full-Stack) Intern 03/23 – 09/23
- Refined 3+ RESTful microservice architectures for real-time distributed data processing, increasing transaction throughput accuracy by 18%
- Architected 4 server-side internal tools using Node.js with Express, using A/B testing to refine real-time database interactions and workflow
- Led code reviews on backend reliability, resolved 7+ critical Firebase Realtime Database API issues, and optimized Jenkins CI/CD pipelines

SoKat **Woodstock, MD**
Machine Learning Engineer Intern 06/23 – 08/23
- Coded 5+ Azure RESTful APIs with Nginx load balancing/Circuit Breaker fault tolerance, ensuring instant financial forecasting for 2K+ firms
- Developed 3+ robust internal tools for automated financial data scraping and aggregation from 10+ sources (SEC filings, 8K, etc.) using Python (BeautifulSoup, Scrapy), employing SQL indexing, NoSQL sharding, and query caching/batching techniques to boost backend speed by 18+%
- Boosted backend throughput by 35%+ using Redis in-memory caching, asynchronous tasks, and exponential backoff to enhance reliability
- Researched and fine-tuned 20+ financial sentiment analysis, summarization, and QA transformer models, applying ML metrics to embed top-performing models into existing backend architecture with Docker and Kubernetes, thereby improving accuracy and analytical scope by 18+%

RESEARCH EXPERIENCE

Johns Hopkins University – Center for Language and Speech Processing **Baltimore, MD**
Undergraduate Researcher 01/23 – Present
- Implemented multi-GPU processing across distributed cluster environments using Slurm scripts and CUDA, leveraging parallel computing to train/test/validate transformer-based neural networks in PyTorch and generate 10+ model checkpoints for NLP information retrieval tasks
- Programmed a scalable NLP pipeline using Tevatron/Elasticsearch, yielding a 20%+ ETL performance boost, to process a large-scale dataset of 80M+ web articles from 50+ languages that facilitates the development of a cross-lingual information retrieval model for citation generation
- Crafted 5+ human-inspected 2k+ passage datasets using HuggingFace for instruction-following LLM tasks and benchmark experiments

Kennedy Krieger Institute – F.M. Kirby Research Center for Functional Brain Imaging **Baltimore, MD**
Undergraduate Researcher 09/22 – 01/23
- Optimized data architecture and ETL for fcMRI studies, developing automation scripts to purge 55%+ redundant entries from 5+ years of archives, improving efficiency by 30%+ through data cleansing, NoSQL refinement, and enforced pseudonymization for data security
- Architected multi-process pipeline in CONN, integrating 3+ advanced statistical packages and proprietary algorithms to preprocess fcMRI images, ensuring accurate generation of research-grade visualizations, metrics and figures, streamlining neuroimaging workflows

TECHNICAL SKILLS

- **Languages:** Python, Java, C++, HTML, CSS, JavaScript, C, MATLAB, Bash, SQL, TypeScript, Assembly
- **Selected Libraries & Frameworks:** BeautifulSoup, Scrapy, PyTorch, Django, Flask, NLTK, tkinter, Requests, PyTest, JUnit, Apache, STL, Bootstrap, PyData (NumPy, Pandas, scikit-learn, etc.), MERN (MongoDB, Express.js, React.js, Node.js), Next.js
- **Database & Cloud:** MySQL, PostgreSQL, Firebase, SQLite, Redis, Azure, AWS, Google Cloud Platform (GCP), Heroku
- **Tools & Platforms:** CUDA, Docker, Slurm, Kubernetes, Postman, Git, Jira, Databricks, Airflow, Jenkins, Terraform, Ansible