Bennett Li

bli4@andrew.cmu.edu | 512-751-6230 | linkedin/bento25 | github/SEGV1

Education

Carnegie Mellon University Pittsburgh, PA | May 2026

Bachelor of Science in Computer Science, Concentration in Human-Computer Interaction — *#1 CS program, U.S News*

Coursework: Algorithm Design & Analysis, Distributed Systems, Natural Language Processing, Software Foundations of Security & Privacy, Computer Systems, Human-Centered Software Design, Design of AI Products, Foundations of Software Engineering, Interaction Design Fundamentals, Probability Theory

Skills

Languages:	C++, C, Python, Java, TypeScript, Scala, SQL
Frameworks & Tools:	React, Apache Airflow, Great Expectations, OpenCV, MediaPipe, Protobuf
Platforms:	AWS (CloudWatch, Redshift, S3), GCP, Docker, Linux, Git

Work Experience

Perpay / *Data Engineering Intern* *Philadelphia, PA / May 2024 – Aug 2024*

- Diagnosed systemic silent failures across 20+ Airflow DAGs with zero observability; deployed a modular monitoring layer using Airflow callbacks and custom CloudWatch metrics, giving the team real-time pipeline visibility for the first time
- Built TypeScript/React dashboards surfacing DAG-level query cost attribution from Redshift STL logs, enabling self-serve identification of expensive queries and informing WLM queue prioritization without DBA involvement
- Refactored a high-traffic ingestion service with async batching and exponential backoff, reducing third-party API rate-limit failures by **73%** under sustained peak load
- Introduced column-level data contracts via Great Expectations at DAG entry points, catching upstream schema drift before it reached production and eliminating a recurring class of data quality incidents

ARPA-E Repair Project / *Research Assistant* *Pittsburgh, PA / May 2023 – June 2024*

- Built a real-time OpenCV vision pipeline (adaptive Canny edge detection, diameter estimation, fisheye distortion correction) feeding pipe geometry directly into locomotion control for robots navigating with no prior environment map
- Replaced a threshold-based anomaly detector with a geometry-aware classifier tuned to cylindrical surface curvature, improving fault detection accuracy from **85% to 92%**
- Developed C++ adaptive locomotion strategies adjusting wheel torque and stride in real time, enabling traversal across pipes with up to **40% diameter variance** without manual recalibration

Ivyland Academy / *Lead Teaching Assistant, ML Curriculum* *Austin, TX / June 2025 – Aug 2025*

- Designed targeted walkthroughs on ensemble methods (Random Forest, AdaBoost, XGBoost) after identifying consistent student misconceptions around the bias-variance tradeoff; evaluated 15+ ML projects with structured feedback on model selection and feature engineering

Projects

Velox — Userspace TCP/IP Stack | link · *C, C++, Protobuf, Linux TUN*

- Built an LD_PRELOAD interceptor overriding the full POSIX socket API to route AF_INET calls to a userspace TCP/IP daemon; non-AF_INET calls fall through to glibc via dlsym, with FD range negotiation at startup to avoid collision with kernel-assigned descriptors
- Designed a Protobuf oneof-dispatched RPC layer with a Session state machine using 4-tuple best-match scoring (handles wildcard binds) and correct errno propagation across fork()'d processes

Real-Time Beauty Filter | link · *Python, OpenCV, MediaPipe*

- Implemented a frequency-separation pipeline decomposing each frame into low (tone) and high (edge/pore residual) layers, smoothing only the low layer and recomposing — preserving every edge while removing blemishes; MediaPipe 478-point landmarks drive a feathered skin mask with convex-hull exclusions for eyes and lips
- Replaced ring-average spot healing with Poisson-equation seamless cloning: score donor patches via Sobel gradient energy, run solver once per spot (<80ms), cache result — decoupling the expensive solve from the per-frame render loop

Academic Projects (CMU) — source code withheld per academic integrity policy

Scalable Three-Tier Web Service · *Java, Java RMI*

- Dynamically auto-scaling service with independent FE/app-tier scaling thresholds, cold-start mitigation via pre-launched VMs, and tail-drop load shedding with 12-retry forwarding tolerating ∼6s boot lag

Remote File System · *C, Linux*

- LD_PRELOAD-based interposition transparently redirecting 9 POSIX syscalls to a remote TCP server; fake-fd table offset at 1,000,000 ensures no collision with kernel-assigned descriptors