Bernard Shen
blshen@uchicago.edu | (704) 713 - 4607
bernardshen.com | linkedin.com/in/bern-shen | github.com/bern-shen

EDUCATION

The University of Chicago Chicago, IL
Bachelor of Science in Computer Science with Specialization in Machine Learning Expected, June 2026
Bachelor of Arts in Economics with Specialization in Data Science
Cumulative GPA: 3.82
Relevant Coursework: Machine Learning, Theory of Algorithms, Database Systems, Operating Systems, Linear Algebra, Econometrics, Causal Inference, Statistical Models and Methods

EXPERIENCE

University of Chicago Booth Business School Chicago, IL
Research Assistant, InterviewAI Project – AI and Labor Economics May 2025 – Present

- **Senior Thesis (ML Model Development):** Engineered LoRA fine-tuning pipeline for LLaMA-2 to predict worker retention from interview transcripts; designed three-way forecasting contest comparing traditional ML (Random Forest, Logit), fine-tuned LLMs, and human recruiters on held-out test set. Implemented stratified splits with cross-validation, evaluating models on retention accuracy, calibration, and feature importance.
- **NLP Feature Engineering:** Developed modular NLP pipeline using SBERT embeddings, UMAP dimensionality reduction, and HDBSCAN clustering to extract interview themes from transcripts; implemented LLM-based prompt engineering to partition transcripts into structured question-answer pairs with JSON-encoded metadata.
- **Causal ML & Feature Engineering:** Designed econometric models with hierarchical specifications to analyze mediating factors in 70,000-person RCT (study found AI interviews increased job offers by 12%, retention by 17%); engineered 16-variable linguistic feature set across conversation dynamics, interpersonal connection, information quality, and language complexity.
- **Data Engineering & Infrastructure:** Optimized GCP pipelines for 4GB+ datasets, implementing memory-efficient streaming and schema enforcement to enable stable processing of 12,000+ transcripts.

Red Ventures Fort Mill, SC
Data Analyst Intern, Allconnect team – Sales Operations June 2025 – August 2025

- Engineered modular SQL pipeline in Databricks parsing 1,500+ product bundles via temporal joins, creating centralized variable-cost model for $21M channel; developed Python reconciliation script validating outputs against raw data, resolving 0.4% error rate and eliminating 12 analyst-hours monthly.
- Led stakeholder UAT to validate edge-case logic, establishing pipeline as authoritative source for commission payouts and activation-adjusted revenue forecasting.

TECHNICAL SKILLS

Languages: Python, SQL, R, C/C++, Rust
ML & Data Science: PyTorch, Scikit-Learn, Pandas, NumPy, Statistical Modeling, A/B Testing, Feature Engineering
NLP & Clustering: Sentence Transformers (SBERT), UMAP, HDBSCAN, LLM Fine-Tuning, Prompt Engineering
Tools & Cloud: Git/GitHub, GCP (Vertex AI, Google Cloud Storage), Databricks, Tableau, Unix/Linux
Certifications: Data Scientist (DataCamp, 2025)

TECHNICAL PROJECTS

Pintos OS | C November 2025

- Architected virtual memory subsystem using hash-based supplemental page tables for lazy loading and demand paging; implemented bitmap-backed swap partition and priority scheduler with nested priority donation.
- Implemented system call handlers with dual-layer memory validation for process control; extended x86 kernel file system with multi-level indexed inodes supporting 8.5MB extensible files and hierarchical subdirectories.

MiniDB Storage Management System | Rust May 2025

- Architected database engine with slotted-page storage and heap file management, achieving 475µs for 1,000-record insertions; built Volcano-style query execution engine with hash equi-joins (O(N+M)), nested loop joins, and group-by aggregations for end-to-end SQL processing.
- Engineered concurrent buffer pool with RwLock latches and automatic page compaction for space reclamation.