

NICHOLAS CHEN

nchen06@berkeley.edu | +1 510-529-6747 | linkedin.com/in/nchen06 | github.com/nnicholas-c | nicholaschen.dev

EDUCATION

University of California, Berkeley

Berkeley, CA

B.S. in Electrical Engineering & Computer Science | GPA: 3.8 / 4.0

Expected May 2028

Relevant Coursework: Optimization Models (EECS 127), Probability & Discrete Math (CS 70), Deep Learning (CS 182); In progress: Probability & Random Processes (EE 126), Linear Algebra (Math 110), Algorithms (CS 170), Machine Learning (CS 189).

HONORS & AWARDS (SELECTED)

- **USACO Gold**; International Olympiad in Informatics (IOI) 2024 National Camp Member.
- Australian Mathematics Competition — **Medal (top 1 / 5,000)** & Top in Auckland; 3× High Distinction.
- NZ Mathematical Olympiad Camp Member; NZ Senior Mathematics Competition — 1st Place.
- **UC Berkeley Chandra Research Fellow** (URAP) — competitive funded summer research fellowship in AI-assisted decision-making.
- Science olympiad golds: NZ Chemistry (2× Gold, 1× Silver, National Exam full score), UK Chemistry, British Physics, British Biology; IChO 2024 Qualifier.

EXPERIENCE

UC Berkeley — Operations & Behavioral Analytics Lab

Berkeley, CA

Quantitative Researcher

Sep 2025 – Present

- Modeled order-bundling as a sequential decision problem (state = order menu, action = chosen bundle, reward = earnings per unit time) and built a DQN agent (reward shaping, experience replay, target networks), benchmarked against human choices.
- Clustered per-worker decision features into **5 statistically distinct policies** (8× earnings spread, $F=276$) across 85 participants and 1,000+ rounds; built the full SvelteKit / Firebase experiment pipeline and ran offline policy evaluation against behavior-cloning baselines.
- Showed payout salience causally raises suboptimal choices by 71 percentage points (fixed-effects causal inference, bootstrap CIs). **First author** of two papers in preparation for ACM CHI 2027 and ICML 2027.
- Found no measurable learning across 15 unaided rounds, motivating adaptive support; now building contextual-bandit recommendation policies that condition on each worker's diagnosed decision policy, evaluated on decision quality, adoption, and response time.

Shenwan Hongyuan Securities

Hong Kong SAR

Quantitative Data Engineering Intern

Jan 2025 – Feb 2025

- Owned QA and reconciliation for equities market-data pipelines feeding quant research — multi-terabyte tick and daily-bar feeds (billions of rows) — with schema, null, outlier, duplicate/gap, and symbol/timestamp-alignment checks for clean, correctly time-aligned research data.
- Cut transform runtime 20% via partition pruning and vectorized Pandas / NumPy / Spark operations over partitioned Parquet; built idempotent, backfill-safe ETL (Python / SQL / Airflow) with dependency-managed DAGs, raising ingestion throughput 10%.
- Cut production query latency 30% by tuning PostgreSQL (partition-aware access, query-plan optimization); added Redis caching for hot datasets and containerized services with Docker.

MIT — Pentelute Lab

Cambridge, MA

Machine Learning Researcher

Feb 2023 – Apr 2024

- Built a deep generative + RL stack in PyTorch (diffusion sampling, policy / value methods) for molecular design using RFDiffusion/ProteinMPNN/AlphaFold3 — framed as constrained, multi-objective optimization over a large discrete search space, maximizing predicted affinity while penalizing toxicity / aggregation.
- **First-authored** a peer-reviewed paper on AI-driven molecular design (DOI: 10.36838/v7i4.29); designed the reward objective, tracked training / RL dynamics (loss / grad norms, reward curves), and ran sensitivity analyses on generation quality (top-k, diversity).

University of Oxford

Oxford, UK

Computational Research Program — **Ranked 1st of 45 Researchers**

Dec 2023 – Jan 2024

- Built Monte Carlo simulations and stochastic-process / time-series models of nonlinear dynamical systems (Python / C++ / MATLAB), with parameter sweeps and stability / sensitivity analysis.
- Built high-dimensional statistical-learning pipelines (PCA, k-means, random forests) for candidate scoring and uncertainty / risk assessment.

PROJECTS

Cross-Sectional Equity Signal Research

Python · pandas · scikit-learn

- Built a survivorship-controlled, point-in-time S&P 500 study (~614 names) testing six signals net of costs — purged/embargoed walk-forward CV, Newey–West t-stats, factor-neutral alpha, a deflated Sharpe, and a LinUCB contextual-bandit allocator.
- **Caught and killed a survivorship artifact** — a +8%/yr alpha ($t \approx 3.6$) vanished under point-in-time membership — and reported the honest null, backed by an independent re-derivation and 50 unit tests.

CivicGrid — Voice-to-Structured-Data Pipeline

CalHacks 2025

- Built a full-stack pipeline (TypeScript / React, Python / Flask) converting call audio (Deepgram STT) into structured records via an LLM classification layer; featured at the CalHacks 2025 YC afterparty.

SKILLS

Languages: Python, C++, Java, SQL, JavaScript / TypeScript, C#, MATLAB

Quant & ML: Probability & statistics, stochastic processes, time-series, optimization, Monte Carlo; reinforcement learning; PyTorch, NumPy, Pandas, scikit-learn, statsmodels, factor models

Data & Tools: PostgreSQL, Spark, Airflow, Redis, Docker, Git, Linux / Bash, Hugging Face