Cameron Cooper

Applied AI Engineer · LLM Post-Training & Evaluation · Document AI

New York, NY · Remote-ready · (917) 912-1264 · czc@umich.edu · linkedin.com/in/czcoop · github.com/camczc · camcooper.com

SUMMARY

AI engineer who fine-tunes and evaluates LLMs that extract structured data from long, messy financial documents. At LightBox I own the training-data, post-training and benchmark pipeline for a model that converts commercial real estate rent rolls into structured workbooks at 99.8% cell-level F1, and I'm helping turn that work into paid AI add-on products for clients.

EXPERIENCE

LightBox New York, NY
Associate AI Research Engineer Aug 2026 – Present

- Fine-tune an open-weight LLM on 2× NVIDIA H200 GPUs to extract entire rent-roll documents into a seven-sheet structured workbook, reaching **99.8% cell-level F1** and **96% strict cell accuracy** on a 122-document held-out set.
- Built the dataset pipeline behind it: versioned training releases, document-stratified train/validation splits and locked answer keys, so every model run is scored against the same frozen data.
- Designed the rent-roll evaluator for CRE Bench, LightBox's internal benchmark for document AI. It has strict and fair scoring modes, run-to-run comparison and audit workbooks, and matched the reference scorer exactly across 270K+ cell decisions, backed by a 165-test suite.
- Partner with product to package extraction models as AI add-ons that clients purchase inside the LightBox platform.

Data Science Engineer (summers 2023–2025, full-time from 2026) May 2023 – Aug 2026

- Shipped production NLP pipelines in Python that extract, structure and validate CRE documents, processing up to 100 files a day and removing 3–4 hours of manual review per document.
- Built LLM-based grading and HUD Fair Market Rent benchmarking that condenses 500+ page documents into structured 50-point summaries at 90–95% accuracy.

Capital Z Partners New York, NY
Data Science Analyst Intern May 2022 – Aug 2022

- Modeled and structured the co-investment raise for a $2B leveraged acquisition of an insurance services business, and supported due diligence and transaction documentation for the firm's founders.

PROJECTS

Kalshi Microstructure Lab · Python, WebSockets, Parquet, zstd 2026

- Built a lossless market-data recorder for Kalshi's 15-minute BTC markets that captures order books, trades and settlement-index ticks alongside Coinbase, Kraken and Binance. Validation run: 304K events with zero sequence gaps.
- Wrote a no-lookahead feature builder, a latency- and fee-aware strategy search and a maker queue-position replay. Strategies are deployed only if they pass a pre-registered gate (600+ markets, multiple-testing-adjusted confidence interval above zero).

AlphaLens & CryptoLens · FastAPI, PostgreSQL, Claude API 2025 – 2026

- Built two full-stack research platforms for equities and crypto: technical signal scoring, a backtesting engine (Sharpe, Sortino, max drawdown, benchmarked against buy-and-hold) and LLM-generated analyst briefs.

ScoutingReport · FastAPI, Supabase, Plotly, MLB Stats API 2025

- Built an MLB analytics app with AI-written scouting reports, pitch-location heatmaps, head-to-head matchup analysis and daily probable-pitcher previews.

EDUCATION

University of Michigan · B.S. Data Science, Minor in Computer Science · GPA 3.7 May 2026
Harvard Data Science Initiative · Certificate, Agentic AI: Contextualized and Applied 2026

SKILLS

ML & LLMs: supervised fine-tuning, LLM evaluation and benchmarking, document AI / information extraction, PyTorch, Hugging Face, RAG, agentic workflows, prompt engineering
Engineering: Python, SQL, C++, TypeScript · FastAPI, PostgreSQL, Supabase · pytest, Git, Linux, uv, GitHub Actions, AWS
Data & Quant: pandas, NumPy, statistical testing, backtesting, Plotly, Streamlit

ADDITIONAL

Early team member at Code Central (2017–2022), a coding school for ages 8–17. Wrote the Python, C++ and robotics curriculum and helped grow it to 400+ students and about $1M in annual revenue.