

Tom Chen

Backend, Data & Market Systems Engineering

Mississauga, Ontario, Canada

tom863631@gmail.com | github.com/Tcx086 | portfolio.tcx086.com/

PROFESSIONAL PROFILE

University of Toronto BCom graduate combining financial-market research, data and systems engineering, early-stage AI product development, and quality review. Specializes in market microstructure and data integrity: multi-venue acquisition, exact order-book and trade facts, causal replay, validation, observability, recovery, and research-ready storage. Brings hands-on backend development, source checking, issue investigation, and clear communication with cross-functional collaborators.

CORE SKILLS

Programming & Data: Python, SQL, TypeScript, basic Rust, Pandas, NumPy, FastAPI, PostgreSQL, SQLite, DuckDB, SQLAlchemy, Alembic, JSON/CSV, Parquet.

Market & Trading Data: Crypto derivatives, perpetual futures, Level-2 order books, trades, funding, open interest, mark/oracle state, liquidations, basis, liquidity/depth, replay, market-event research.

Data Quality & Controls: Normalization, schema validation, sequence integrity, duplicate/conflict detection, reconciliation, causal timestamps, source identity, audit trails, root-cause investigation, go/no-go readiness gates.

Systems & Infrastructure: REST/WebSocket APIs, durable spool/landing, immutable Parquet, DuckDB compaction, observability, structured logs, recovery, Vercel, Cloudflare Workers/D1, Git/GitHub, CI-style testing.

Analytics & Modeling: Excel, Power BI, statistical/financial analysis, volatility, drawdown, CVaR, Monte Carlo, factor/regime analysis, credit-risk modeling, scenario testing, market/macro/event research.

Quality & Collaboration: Document review, requirements validation, terminology consistency, exception tracking, technical documentation, stakeholder coordination, Excel reporting, translation QA.

WORK EXPERIENCE

Lomidance | May 2026 - Present

Backend & Data Engineer | Contract Part-time

- Contribute to an early-stage AI workspace for multi-scene generative-video production in a shared product codebase.
- Own substantial backend/database work with Python, FastAPI, PostgreSQL, SQLAlchemy and Alembic: API contracts, schema/model design, migrations, application data flows, and frontend/backend integration.
- Built and hardened multimodal service boundaries, validation fixtures, redaction and failure-safe helpers, focused tests, and deployment integration with explicit provider/runtime boundaries.
- Integrated external generation APIs and verified an end-to-end application workflow through successful video output. Collaborate through Git/GitHub and Vercel-based development and deployment workflows.

PROJECTS

QUANTITATIVE PROJECTS

Microstructure Discovery Factory | August 2026 - Present

Independent Project | Independent Researcher | Market Data & Systems

- Designed and built a causal market-data capture, persistence, replay, research and validation system spanning Binance, OKX, Bybit, Deribit and Hyperliquid. The pipeline runs from WebSocket acquisition and normalization through exact L2/trade facts, durable spool/landing, Parquet commit markers, an active-part index and bounded DuckDB compaction to replay/research.
- Operate six common perpetual trade lanes for BTC/ETH across Binance, OKX and Bybit, plus primary-venue L2 capture, Deribit option-state feeds, and Hyperliquid funding, open-interest, premium and mark/oracle state.
- One live FOMC capture completed 12/12 lanes with 653,722 L2 updates and 4,339,527 trades; v4.4 records zero sequence mismatches, normalization rejections, duplicates or source-identity conflicts, with peak RSS of 2.887 GiB. These are results from that capture.
- Used the event data to quantify approximately 68% contraction in displayed BTC/ETH depth and measure subsequent liquidity recovery.
- Investigated sequence gaps, reconnect behavior, persistence failures, stale replay boundaries, queue pressure and latency outliers using source evidence, structured logs, replay and reconciliation. Replaced starvation-prone queues with bounded ring-style storage and cooperative draining; propagated integrity failures into research-readiness checks.
- Built causal replay using exchange-event, received and available timestamps, immutable landing metadata and bounded research readers to reduce future leakage. Developed a downstream Liquidation Pressure layer retaining provenance, sparse-state semantics, explicit missingness and causal query boundaries without changing authoritative raw facts.
- Maintained acceptance criteria, blocker tracking, durability/recovery checks, protected interfaces, closure reports and regression suites across runtime, workstation, replay, dashboard and research components. Stack: Python, TypeScript/Node.js, Rust components, WebSockets, Parquet, DuckDB, local dashboards and structured logging.

Quantitative Portfolio Optimization | 2025 - August 2026

Independent Project

- Built and compared Markowitz, maximum-Sharpe/minimum-volatility, risk-parity, Black-Litterman, CVaR, factor-aware, dynamic/regime-aware and multi-objective portfolio-allocation approaches.
- Evaluated changing volatility, correlation, covariance, downside-risk, transaction-cost and market-state assumptions using scenario analysis, Monte Carlo-style testing, drawdown sensitivity and risk-adjusted metrics. Compared assumptions and failure modes rather than treating optimization output as a stand-alone investment decision.

Historical Live Perpetual Futures Execution | 2026

Independent Trader | Own Capital

- Previously ran own-capital USDT-margined perpetual-futures exposure, primarily on OKX. Live trading is paused while the next-generation research and market-data infrastructure is rebuilt and validated.
- Monitored order placement/fills, leverage, funding exposure, transaction costs, slippage, position risk and drawdown controls. Compared cross-venue funding, spot-perpetual basis, liquidity/depth, open interest, order flow and liquidations to distinguish venue dislocations from broader market changes.
- Kept live, paper/replay and test workflows separate to validate execution and strategy changes before broader use; signals were not treated as automatic execution triggers.

PROJECTS (CONTINUED)

ANALYTICAL PROJECTS

Situation Monitor / Market Context Engine | 2025 - Present

Independent Product & Research Project

- Built a market-context system combining macroeconomic data, economic expectations, news narratives, regulatory/policy information, geopolitical developments and market data for structured event interpretation and research.
- The v4.4 historical snapshot reports 106K+ news records and 207K+ structured LLM-extraction rows, with controls for source coverage, missing values, duplicates, schema consistency, reproducibility and downstream research use. These are source-version totals, not current throughput claims.
- Developed a Trusted Context Foundation with strict macro identities, immutable pre-event expectations, canonical actual/revision handling, deterministic fact selection, temporal guards and as-of event inspection.
- Connected the runtime to Cloudflare Workers/D1 and validated durable write/read/redeploy persistence. The v4.4 snapshot records 5 tables, 11 indexes, 12 triggers and a passing 110-item live parity suite covering ordering, revisions, inspection and temporal guards.
- Maintained a production-facing Svelte/SvelteKit/TypeScript frontend, server-side provider access, caching and secret isolation. Kept rule-based narrative/impact signals separate from source facts without unsupported predictive claims; used the system for provenance, normalization, deterministic selection, failure handling, deployment and validation.

Credit Risk Modeling | 2025

University / Independent Analytical Project

- Analyzed 34,000+ applicant records using logistic regression, random forest and XGBoost. Evaluated accuracy, precision, recall, F1, ROC-AUC, confusion matrices and SHAP-based feature interpretation.
- Compared default-risk identification approaches with an emphasis on validation and interpretability; model scores were not treated as stand-alone credit decisions.

Legacy API-Based Trading & Monitoring Workflows | 2024 - August 2026

Independent Projects

- Integrated public REST/WebSocket market-data feeds, local SQLite storage, alert/command workflows, transaction-cost assumptions, position limits, drawdown controls and structured runtime logs.
- Used monitoring and validation to identify data-quality issues, execution-assumption gaps, anomalies and recovery needs; documented fixes and follow-up tests. These workflows informed Microstructure Discovery Factory and are no longer the canonical research infrastructure.

VOLUNTEER EXPERIENCE

Transcommon | September 2026 - Present

Translation Quality Review Volunteer

- Review translations against source material and requirements for accuracy, completeness, consistent terminology, meaning preservation and formatting quality.
- Flag mistranslations, omissions, inconsistent terminology, ambiguous phrasing and requirement mismatches for correction before final use.

EDUCATION

University of Toronto Mississauga | September 2021 - April 2025

Bachelor of Commerce | April 2025

Relevant areas: Business Analytics, Economics, Finance/Financial Fundamentals, and Data Analysis.

LANGUAGES

Mandarin Chinese - Native; English - Professional Working Proficiency; French - Beginner.

HOBBY

Golf - playing regularly since high school.