

AI Platform Overview

The Infrastructure Redefining Quantitative Market Research

RICCHE LTD

ricche.ai

Powered by the NVIDIA Accelerated Computing Stack

Why Ricche Exists

The quantitative finance landscape is at an inflection point. Traditional approaches -- hand-crafted models, limited datasets, CPU-bound computation -- are hitting fundamental limits. The firms that will lead the next decade are those building on AI-native infrastructure: GPU-accelerated, data-rich, and algorithmically sophisticated.

Ricche is building that infrastructure from the ground up. Not by bolting AI onto legacy systems, but by designing every layer -- from data ingestion to model serving -- for the scale and speed that modern machine learning demands.

We believe the next breakthrough in quantitative research will not come from a better algorithm -- it will come from better infrastructure.

47x

GPU Acceleration

<2ms

Inference Latency

5

Asset Classes

24/7

Monitoring

What We Are Building

Machine Learning Research

NVIDIA CUDA-accelerated model training that compresses weeks of experimentation into hours. PyTorch environments with full experiment tracking, automated hyperparameter search, and model versioning. Deep learning, gradient boosting, transformers, and ensemble methods -- all GPU-accelerated.

Simulation Engines

GPU-parallelised Monte Carlo and agent-based simulations that stress-test models across thousands of market scenarios before they ever see live data. Regime changes, liquidity shocks, flash crashes -- if a model cannot survive the simulation gauntlet, it does not progress.

Data Infrastructure

Multi-asset market data spanning equities, futures (including commodities), options, FX, and fixed income from major global exchanges. Real-time streaming and historical batch pipelines. GPU-accelerated feature engineering via RAPIDS. Institutional-grade quality assurance.

Operations & Monitoring

Real-time Control Room dashboard tracking GPU utilisation, active experiments, simulation queues, and pipeline throughput. Automated anomaly detection and alerting. Complete audit trail for every experiment, every model, every data transformation -- full research governance out of the box.

Powered by NVIDIA

We are building Ricche on the NVIDIA accelerated computing stack because there is no substitute for raw GPU performance when it comes to financial ML. The architecture targets speak for themselves:

NVIDIA CUDA

The foundation of everything we do. CUDA-accelerated training and simulation deliver dramatic speedups over CPU-only infrastructure. Mixed-precision training (FP16/BF16) maximises GPU throughput without sacrificing model fidelity.

RAPIDS cuDF

GPU-accelerated DataFrames that make pandas feel like a relic. Feature engineering on billion-row financial time-series in seconds, not hours. End-to-end GPU pipelines eliminate the CPU bottleneck entirely.

TensorRT

Production inference optimised to sub-2ms latency through layer fusion, kernel auto-tuning, and INT8 quantisation. When markets move in milliseconds, this is the difference between leading and lagging.

NVIDIA NIM

Optimised model inference microservices with one-click deployment and dynamic scaling. Built-in orchestration will handle multi-model serving with intelligent batching. The production layer designed to scale with our ambition.

Global Market Coverage

Our data infrastructure spans the world's most important financial markets, with institutional-grade feeds from major exchanges and data providers:

- Global Equities -- developed and emerging markets across 30+ exchanges
- Futures & Commodities -- energy, metals, agriculture, financial futures
- Options -- full listed chains with Greeks, volatility surfaces, and skew data
- Foreign Exchange -- spot, forwards, and crosses for major and EM pairs
- Fixed Income -- government bonds, credit markets, and interest rate instruments
- Alternative Data -- NLP-processed news, satellite, sentiment, and web data

The Opportunity

The global quantitative finance market is growing rapidly as AI adoption accelerates across the financial sector. Traditional quant firms are racing to modernise their infrastructure. New entrants need institutional-grade tools from day one. The demand for GPU-accelerated research infrastructure has never been higher.

Ricche sits at the intersection of three powerful trends: the explosion of available market data, the maturation of GPU-accelerated computing, and the proven superiority of machine learning approaches to financial modelling. We are not building for today -- we are building for the decade ahead.

Who Should Partner With Ricche

- Technology providers looking to showcase GPU computing in finance
- Data vendors seeking a showcase platform for their market data products
- Research institutions exploring AI applications in financial markets
- Quantitative firms needing modern, GPU-native research infrastructure
- Investors looking at the convergence of AI, finance, and high-performance computing

Ready to Explore What We Are Building?

Whether you are a potential partner, investor, or researcher, we would love to show you what GPU-accelerated market research looks like in practice.

info@ricche.ai | ricche.ai