TEXT-DRIVEN QUANTITATIVE RISK INTELLIGENCE SYSTEM

BY : OM BAROT

B.TECH CSE – VIT VELLORE (2026)

A Natural Language to Institutional-Grade Risk Intelligence Pipeline

---

Abstract

Modern financial institutions increasingly rely on quantitative risk systems to manage portfolio exposure, regulatory capital, and tail-risk events. However, a fundamental gap persists between **how humans communicate risk requirements** (natural language) and **how quantitative systems consume them** (structured numerical inputs). This project proposes and implements a **Text-Driven Quantitative Risk Intelligence System**, a five-phase pipeline that transforms informal portfolio descriptions into **Basel-compliant Value-at-Risk (VaR), Expected Shortfall (ES), backtesting diagnostics, and executive-level risk intelligence**.

The system demonstrates a full production-grade workflow: 1. Natural language intent parsing (Phase 1) 2. Market data engineering and return construction (Phase 2) 3. Multi-model risk metric computation (Phase 3) 4. Regulatory backtesting and validation (Phase 4) 5. LLM-powered executive risk interpretation (Phase 5)

The design emphasizes determinism, transparency, regulatory relevance, and explainability, making the system suitable both as a **research contribution** and as a **real-world deployable risk engine**. This document presents a complete descriptive explanation of the system, intended to be fully understandable to readers with **zero prior background in quantitative finance**.

Keywords

Quantitative Risk Management, Value at Risk, Expected Shortfall, Basel III, Natural Language Processing, Financial Engineering, LLM Risk Intelligence

---

1. Introduction

1.1 Motivation

In real financial markets, portfolio managers, traders, and risk officers rarely communicate using mathematical notation. Instead, instructions are given in natural language:

> "Create a NIFTY50-heavy portfolio and compute 99% 10-day VaR"

However, quantitative risk engines require **precise numerical inputs**: - Asset identifiers - Portfolio weights - Time horizons - Confidence levels - Risk models

This mismatch creates operational friction, increases error probability, and limits accessibility to advanced risk analytics. The proposed system addresses this gap by **making quantitative finance conversational, deterministic, and auditable**.

1.2 Research Objective

The primary objective is to design a **complete end-to-end system** that: - Accepts natural language portfolio descriptions - Converts them into structured, machine-readable contracts - Computes industry-standard risk metrics - Validates models using regulatory backtesting - Translates quantitative results into executive-level decision intelligence

1.3 Contributions

This project contributes: - A deterministic NLP-to-quant interface without machine learning training - A production-resilient market data engine with fallback guarantees - A multi-model VaR and ES computation framework - A Basel-compliant backtesting and traffic-light system - A novel LLM-based CRO (Chief Risk Officer) decision intelligence layer

---

2. Background Concepts (For Zero-Finance Readers)

2.1 What Is Financial Risk?

Financial risk refers to the **possibility of losing money** due to market movements. Markets fluctuate daily, and risk quantifies how severe those fluctuations could be under adverse conditions.

2.2 Returns vs Prices

Prices alone are not useful for risk measurement. Risk is measured using **returns**, which represent percentage changes in price:

[R_t =]

This allows fair comparison between assets of different price levels.

2.3 Logarithmic Returns

The system uses **log returns**, defined as:

[r_t = ()]

Log returns are preferred because: - They are additive over time - They are symmetric for gains and losses - They align with continuous-time financial models

---

3. System Architecture Overview

The system follows a strict sequential pipeline:

```
Natural Language Input
        ↓
Phase 1: Intent Parsing
        ↓
Phase 2: Market Data & Returns
        ↓
Phase 3: Risk Metrics (VaR, ES)
        ↓
Phase 4: Basel Backtesting
        ↓
Phase 5: Executive Risk Intelligence
```

Each phase consumes validated outputs from the previous phase, ensuring correctness, traceability, and auditability.

---

4. Phase 1 – Natural Language Intent Parsing Engine

4.1 Purpose

Phase 1 converts unstructured human language into a structured quantitative contract. This is the foundation of the entire system.

4.2 Input Example

```
"Create NIFTY50 (40%), BANKNIFTY (35%), IT ETF (25%) portfolio. Daily
data 2018-2025. 99% 10-day VaR. Backtest and explain."
```

4.3 Processing Layers

1. **Text Normalization** – cleans casing, spacing, symbols
2. **Regex Extraction** – identifies assets, weights, years, risk keywords
3. **Schema Validation** – enforces data types and constraints
4. **Business Rules** – ensures weights sum to 100%, valid years, defaults

4.4 Output Contract

```json
{
  "assets": ["NIFTY50", "BANKNIFTY", "IT_ETF"],
  "weights": [0.4, 0.35, 0.25],
  "start_year": 2018,
  "end_year": 2025,
  "confidence": 0.99,
  "horizon_days": 10,
  "models": ["historical", "varcov", "monte_carlo"],
```

```
  "backtesting": true,
  "explain": true
}
```

4.5 Market Relevance

In real institutions, this mirrors how traders communicate with risk desks verbally or via chat systems. Phase 1 makes such communication machine-actionable.

---

5. Phase 2 – Market Data Engine and Returns Computation

5.1 Objective

Phase 2 transforms intent into numerical time series required for risk modeling.

5.2 Data Acquisition

- Primary source: live financial data providers
- Fallback: synthetic data via calibrated stochastic processes

This ensures **system reliability even during data outages**, a common real-world problem.

5.3 Synthetic Data Justification

Synthetic data is generated using geometric Brownian motion, the standard assumption in financial modeling:

[P_t = P_{t-1} (t + Z_t)]

This preserves statistical realism for risk demonstrations.

5.4 Portfolio Returns

Portfolio returns are computed using linear algebra:

[R_{portfolio,t} = w^T R_t]

This formulation scales seamlessly from single-asset to multi-asset portfolios.

5.5 Outputs

- Aligned price series
- Log return matrix
- Portfolio return series
- Covariance matrix

These are the **numerical primitives of risk**.

---

6. Phase 3 – Risk Metrics Computation

6.1 Value at Risk (VaR)

VaR answers the question:

> “What is the maximum expected loss over a given horizon at a given confidence level?”

Formally:

[P(Loss > VaR) = 1 -]

6.2 Historical VaR

- Uses empirical distribution of past returns
- No distributional assumptions

6.3 Variance–Covariance VaR

Assumes normality:

[VaR = Z__p]

Where: - (Z_): standard normal quantile - (_p): portfolio volatility

6.4 Monte Carlo VaR

- Simulates thousands of correlated return paths
- Captures nonlinear and tail behavior

6.5 Expected Shortfall (ES)

ES measures the **average loss beyond VaR**, making it a coherent risk measure:

[ES = E[Loss | Loss > VaR]]

6.6 Real Market Significance

Banks, hedge funds, and regulators rely on these metrics daily for capital allocation, stress testing, and risk limits.

---

7. Phase 4 – Basel Backtesting and Regulatory Validation

7.1 Why Backtesting Matters

Risk models must be validated against realized outcomes. Basel regulations mandate this.

7.2 Kupiec Test

Tests whether observed VaR breaches match expected frequency.

7.3 Traffic Light Framework

- Green: model acceptable
- Yellow: increased monitoring
- Red: capital penalty or model rejection

7.4 Interpretation Philosophy

Regulators prioritize **capital safety over statistical elegance**. Zero breaches indicate conservatism, not perfection.

---

8. Phase 5 – LLM-Powered Executive Risk Intelligence

8.1 Motivation

Executives do not consume equations—they consume narratives.

8.2 Role of LLMs

The LLM: - Interprets risk numbers - Explains regulatory implications - Produces CRO-level recommendations

8.3 Output Example

```
EXECUTIVE SUMMARY:
Portfolio exhibits moderate tail risk with conservative VaR
calibration.

REGULATORY STATUS:
Basel Green Zone – no capital penalty.

CRO RECOMMENDATION:
Approve Monte Carlo VaR as primary model with quarterly review.
```

This closes the loop between quantitative analytics and decision-making.

---

9. End-to-End Market Impact

This system enables: - Faster portfolio risk assessment - Reduced human error - Regulatory audit readiness - Democratization of quantitative finance

It demonstrates how **modern AI and classical finance can coexist responsibly**.

---

10. Conclusion

The Text-Driven Quantitative Risk Intelligence System represents a complete, production-grade framework that bridges human language and institutional risk analytics. By integrating deterministic parsing, robust data engineering, rigorous quantitative modeling, regulatory validation, and LLM-based interpretation, the project delivers both **technical depth** and **practical relevance**.

This work is suitable as: - A final-year engineering project - A research paper foundation - A prototype for real financial institutions

---