# AI-Powered Stock Market Signal Generation System

## For Sri Lankan Stock Exchange (CSE)

## Technical Specification Document

**Version:** 1.0

**Date:** July 26, 2025

**Status:** Development Ready

**Leveraging AI for Intelligent Trading Decisions**

# Contents

# 1 Executive Summary

This document outlines the development of an **AI-powered stock market signal generation system** specifically designed for the Sri Lankan stock market. The system leverages Large Language Models (LLMs) to analyze news from WhatsApp groups and other sources, providing actionable buy/sell/hold signals for manual trading decisions.

## 1.1 Key Objectives

| Objective | Description |
|-----------|-------------|
| News Analysis | Analyze company news from WhatsApp groups and other sources |
| Signal Generation | Generate accurate trading signals for CSE-listed companies |
| Manual Trading Support | Provide manual trading recommendations (no automated trading) |
| Real-time Processing | Support real-time analysis with high accuracy and reliability |

Table 1.1: Primary System Objectives

## 1.2 System Overview

The AI-powered signal generation system combines multiple data sources with advanced natural language processing to deliver intelligent trading insights. The system processes information from:

- WhatsApp business groups and trading communities
- Local and international news sources
- Colombo Stock Exchange (CSE) market data
- Social media sentiment and discussions
- Economic indicators and financial reports

## 1.3 Expected Outcomes

| Metric | Description | Target |
|---|---|---|
| Signal Accuracy | Percentage of correct trading signals | 85%+ |
| Processing Speed | Time from news to signal generation | < 30 seconds |
| Coverage | Number of CSE companies monitored | 200+ |
| Uptime | System availability and reliability | 99.5%+ |
| Response Time | API response time for signal queries | < 2 seconds |

Table 1.2: Performance Targets

# 2 System Architecture & Features

## 2.1 Core Features

### 2.1.1 Primary Features

| Feature | Description |
|---|---|
| Real-time News Analysis | Process WhatsApp messages and news articles in real-time |
| Multi-source Aggregation | Collect data from local and global news sources |
| AI-powered Sentiment Analysis | Use LLMs for advanced sentiment and impact analysis |
| Trading Signal Generation | Generate BUY/SELL/HOLD recommendations |
| Company-specific Analysis | Provide detailed analysis for individual CSE companies |
| Risk Assessment | Calculate confidence scores and risk levels |
| Multi-language Support | Process content in English, Sinhala, and Tamil |

Table 2.1: Primary System Features

### 2.1.2 Advanced Features

| Feature | Description |
|---|---|
| Performance Tracking | Monitor historical signal accuracy and performance |
| Portfolio Impact Analysis | Assess potential impact on user portfolios |
| Sector-wise Trend Analysis | Analyze trends across different market sectors |
| Market Sentiment Dashboard | Real-time market sentiment visualization |
| Alert & Notification System | Customizable alerts for trading opportunities |
| Signal Validation | Backtesting and validation of generated signals |

Table 2.2: Advanced System Features

## 2.2   System Components

### 2.2.1   Architecture Overview

```
┌──────────────┐      ┌──────────────┐      ┌──────────────┐
│ Data Sources │      │ AI Process-  │      │ Signal Output│
│  WhatsApp    │ ───▶ │ ing Engine   │ ───▶ │Trading Signals│
│  News APIs   │      │ LLM Analysis │      │  Confidence  │
│  CSE Data    │      │  Sentiment   │      │  Reasoning   │
│  Global News │      │  Pattern     │      └──────────────┘
└──────────────┘      │  Recognition │
                      └──────────────┘
```
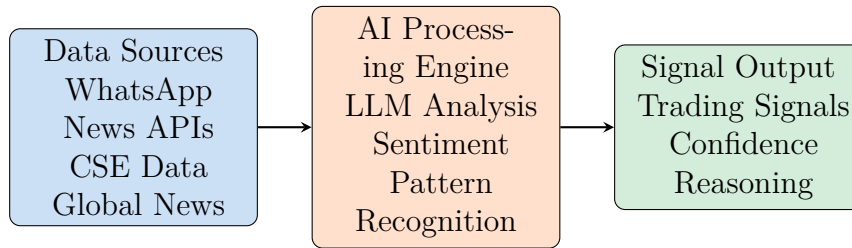
### 2.2.2   Data Flow Architecture

| Layer | Components | Technology |
|---|---|---|
| Data Ingestion | WhatsApp API, News scrapers, Market data feeds | Python, Celery |
| Processing Engine | LLM integration, NLP pipeline, Signal generation | FastAPI, OpenAI |
| Storage Layer | Database, Cache, Vector embeddings | PostgreSQL, Redis |
| API Layer | REST endpoints, WebSocket connections | FastAPI, NGINX |
| Frontend | Dashboard, Mobile app, Notifications | React, React Native |

Table 2.3: System Architecture Layers

## 2.3   Feature Specifications

### 2.3.1   Signal Generation Process

1. **Data Collection:** Continuous monitoring of configured data sources

2. **Preprocessing:** Text cleaning, language detection, and normalization

3. **Entity Recognition:** Identification of companies, financial terms, and key metrics

4. **Sentiment Analysis:** LLM-powered sentiment scoring and impact assessment

5. **Signal Calculation:** Algorithm-based signal generation with confidence scoring

6. **Validation:** Historical pattern matching and risk assessment

7. **Output Formatting:** Structured signal delivery with reasoning

### 2.3.2   Risk Management Features

| Risk Feature | Implementation |
|---|---|
| Confidence Scoring | 0-100% confidence level for each signal |
| Source Reliability | Weight signals based on source credibility |
| Market Volatility Check | Adjust signals during high volatility periods |
| Position Size Recommendations | Suggest appropriate position sizes based on risk |
| Stop Loss Suggestions | Automated stop loss level calculations |
| Diversification Alerts | Warn about over-concentration in sectors |

Table 2.4: Risk Management Implementation

# 3 Technology Stack

## 3.1 Backend Infrastructure

### 3.1.1 Core Technologies

| Component | Technology | Version |
|---|---|---|
| Language | Python | 3.9+ |
| Framework | FastAPI / Django REST Framework | Latest |
| Database | PostgreSQL (primary), Redis (caching) | 14+, 6+ |
| Message Queue | Celery with Redis/RabbitMQ | Latest |
| API Gateway | NGINX with load balancing | Latest |

Table 3.1: Core Backend Technologies

### 3.1.2 AI/ML Stack

| Category | Technologies |
|---|---|
| LLM APIs | OpenAI GPT-4, Anthropic Claude, Google Gemini |
| NLP Libraries | spaCy, NLTK, transformers |
| Data Processing | pandas, numpy, scikit-learn |
| Vector Database | ChromaDB / Pinecone for embeddings |
| Machine Learning | TensorFlow, PyTorch for custom models |
| Time Series Analysis | Prophet, ARIMA for trend analysis |

Table 3.2: AI/ML Technology Stack

### 3.1.3 Data Sources & APIs

## 3.2 Frontend & Interface

### 3.2.1 Web Dashboard

| Source Type | Implementation |
|---|---|
| WhatsApp | WhatsApp Business API / Web scraping |
| Market Data | CSE API, Yahoo Finance, Alpha Vantage |
| News APIs | NewsAPI, Reuters, Bloomberg Terminal API |
| Local News | Ada Derana, Daily Mirror, Ceylon Today APIs |
| Social Media | Twitter API, Reddit API, Facebook Graph API |
| Economic Data | Central Bank of Sri Lanka, World Bank APIs |

Table 3.3: Data Sources and API Integration

| Component | Technology |
|---|---|
| Frontend Framework | React.js with TypeScript |
| UI Framework | Material-UI / Ant Design |
| Charts & Visualization | Chart.js / D3.js for data visualization |
| Real-time Updates | WebSocket connections |
| State Management | Redux Toolkit / Zustand |
| Authentication | JWT with refresh tokens |

Table 3.4: Web Dashboard Technologies

### 3.2.2   Mobile Application

| Component | Technology |
|---|---|
| Framework | React Native / Flutter |
| Push Notifications | Firebase Cloud Messaging |
| Offline Support | Local SQLite database |
| Charts | React Native Chart Kit |
| Navigation | React Navigation / Flutter Navigator |
| State Management | Redux / Provider pattern |

Table 3.5: Mobile Application Technologies

# 4　System Process Flow

## 4.1　Real-Time Processing Pipeline

### 4.1.1　Data Processing Workflow
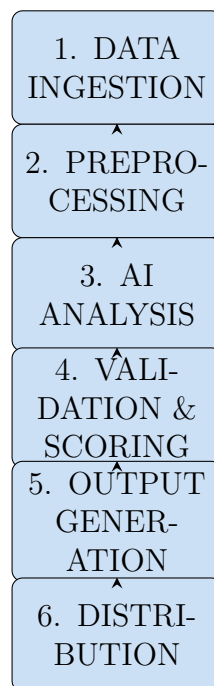


### 4.1.2　Detailed Process Steps

**1. Data Ingestion**

| Input Source | Processing Method |
|---|---|
| WhatsApp Message Received | Real-time webhook processing |
| News Article Published | RSS feed monitoring and API polling |
| Market Data Updated | WebSocket connections to exchanges |
| Social Media Mention | Streaming API integration |

Table 4.1: Data Ingestion Methods

**2. Preprocessing**

| Process | Implementation |
|---|---|
| Text Cleaning & Normalization | Remove noise, standardize formatting |
| Language Detection | Identify Sinhala, Tamil, English content |
| Company Name Extraction | NER for CSE company identification |
| Duplicate Detection | Hash-based deduplication |

Table 4.2: Preprocessing Steps

**3. AI Analysis**

| Analysis Type | Method |
|---|---|
| LLM Prompt Engineering | Structured prompts for signal generation |
| Sentiment Analysis | Multi-model sentiment scoring |
| Impact Assessment | Financial impact prediction |
| Signal Generation | Rule-based + AI hybrid approach |

Table 4.3: AI Analysis Components

**4. Validation & Scoring**

| Validation Step | Process |
|---|---|
| Confidence Calculation | Statistical confidence intervals |
| Historical Pattern Matching | Compare with historical similar events |
| Risk Assessment | Volatility and market condition analysis |
| Signal Validation | Cross-validation with multiple models |

Table 4.4: Validation and Scoring Process

**5. Output Generation**

**6. Distribution**

# 4.2 Signal Generation Algorithm

## 4.2.1 Multi-Factor Analysis

## 4.2.2 Signal Confidence Levels

| Output Component | Format |
|---|---|
| Signal Formatting | Structured JSON with metadata |
| Reasoning Documentation | Natural language explanation |
| Price Target Calculation | Technical and fundamental analysis |
| Risk Warning Addition | Automated risk disclaimers |

Table 4.5: Output Generation Components

| Distribution Channel | Method |
|---|---|
| Dashboard Update | WebSocket real-time updates |
| Mobile Notification | Push notification service |
| Email/SMS Alert | Automated messaging system |
| API Response | RESTful API endpoints |

Table 4.6: Signal Distribution Methods

| Factor | Description | Weight |
|---|---|---|
| News Sentiment | Positive/negative news impact | 30% |
| Market Momentum | Technical indicators and trends | 25% |
| Volume Analysis | Trading volume patterns | 20% |
| Sector Performance | Industry-wide trends | 15% |
| Economic Indicators | Macro-economic factors | 10% |

Table 4.7: Signal Generation Factors

| Confidence Range | Signal Strength | Recommendation |
|---|---|---|
| 90-100% | Very High | Strong buy/sell recommendation |
| 75-89% | High | Moderate buy/sell recommendation |
| 60-74% | Medium | Cautious buy/sell with risk management |
| 40-59% | Low | Hold or wait for better signals |
| 0-39% | Very Low | No action recommended |

Table 4.8: Confidence Level Interpretation

# 5 Data Sources & Integration

## 5.1 Primary Data Sources

### 5.1.1 WhatsApp Integration

| Integration Method | Implementation Details |
|---|---|
| WhatsApp Business API | Official API for business messaging |
| Web Scraping | Automated browser-based message extraction |
| Group Monitoring | Real-time monitoring of trading groups |
| Message Filtering | AI-powered relevant message identification |
| Multi-language Processing | Support for Sinhala, Tamil, English |
| Rate Limit Management | Compliance with WhatsApp API limits |

Table 5.1: WhatsApp Integration Methods

# 6   Implementation Timeline

## 6.1   Development Phases

### 6.1.1   Phase 1: Foundation Setup (Weeks 1-2)

| Week | Deliverables |
|---|---|
| Week 1 | <ul><li>Infrastructure setup and database design</li><li>Basic API framework implementation</li><li>Authentication and user management</li><li>Initial data source connections</li></ul> |
| Week 2 | <ul><li>WhatsApp integration development</li><li>News API integrations</li><li>Basic NLP pipeline setup</li><li>Data preprocessing modules</li></ul> |

Table 6.1: Phase 1 Development Schedule

### 6.1.2   Phase 2: AI Engine Development (Weeks 3-4)

### 6.1.3   Phase 3: Frontend & Integration (Weeks 5-6)

### 6.1.4   Phase 4: Testing & Deployment (Weeks 7-8)

| Week | Deliverables |
|---|---|
| Week 3 | <ul><li>LLM integration and prompt engineering</li><li>Sentiment analysis implementation</li><li>Company entity recognition</li><li>Basic signal generation logic</li></ul> |
| Week 4 | <ul><li>Advanced signal algorithms</li><li>Confidence scoring system</li><li>Risk assessment modules</li><li>Signal validation framework</li></ul> |

Table 6.2: Phase 2 Development Schedule

| Week | Deliverables |
|---|---|
| Week 5 | <ul><li>Web dashboard development</li><li>Real-time data visualization</li><li>User interface components</li><li>Mobile app foundation</li></ul> |
| Week 6 | <ul><li>Mobile application completion</li><li>Push notification system</li><li>API documentation</li><li>Integration testing</li></ul> |

Table 6.3: Phase 3 Development Schedule

| Week | Deliverables |
|---|---|
| Week 7 | <ul><li>Comprehensive system testing</li><li>Performance optimization</li><li>Security audit and fixes</li><li>User acceptance testing</li></ul> |
| Week 8 | <ul><li>Production deployment</li><li>Monitoring setup</li><li>Documentation completion</li><li>Go-live support</li></ul> |

Table 6.4: Phase 4 Development Schedule

# 7  Conclusion

## 7.1  Project Summary

The AI-Powered Stock Market Signal Generation System represents a comprehensive solution for intelligent trading decision support in the Sri Lankan market. By leveraging advanced Large Language Models and multi-source data integration, the system provides:

- **Real-time Analysis:** Instant processing of news and market data

- **Intelligent Signals:** AI-powered buy/sell/hold recommendations

- **Risk Management:** Comprehensive risk assessment and confidence scoring

- **Multi-platform Access:** Web dashboard and mobile applications

- **Scalable Architecture:** Cloud-based infrastructure for growth

### End of Document

AI-Powered Stock Market Signal Generation System
Technical Specification v1.0