Working Paper presented at the

Peer-to-Peer Financial Systems 2026 Workshop

June 2026

Adaptive AI Frameworks for Financial Resilience: Leveraging Multimodal Deep Learning for Real-Time Crisis Management in Fintech Ecosystems

Powered by

P2P Financial Systems

Adaptive AI Frameworks for Financial Resilience: Leveraging Multimodal Deep Learning for Real-Time Crisis Management in Fintech Ecosystems

***Abstract*—Information asymmetry between informed and uninformed market participants remains a structural driver of liquidity crises, including bank runs, flash crashes, and sovereign bond dislocations. Existing computational approaches to crisis prediction operate predominantly within single data modalities: deep learning models process structured price-action data in isolation, while natural language processing pipelines extract sentiment from unstructured text independently. Where fusion of these modalities has been attempted, it targets directional return prediction rather than crisis probability estimation, and employs shallow combination strategies incapable of learning nonlinear cross-modal representations.**

This paper introduces the Unified Intelligence Loop (UIL), a multimodal deep learning architecture that fuses real-time sentiment signals with high-frequency price-action data to generate a Predictive Alpha Score for crisis probability estimation. The architecture comprises three modules: (i) a fine-tuned FinBERT/DistilBERT pipeline for contextual disambiguation of financial text, producing a dense sentiment vector; (ii) an LSTM/GRU encoder processing OHLCV data, realised volatility, and the Amihud illiquidity ratio, producing a dense price vector; and (iii) a learned gating-based late fusion layer that dynamically modulates the relative contribution of each modality conditioned on market state. The system is trained end-to-end with an asymmetric loss function that penalises missed crises more heavily than false alarms.

The UIL is validated through historical backtesting across three documented crisis events—the SVB collapse (2023), the UK Gilt crisis (2022), and the COVID-19 market crash (2020)—on a proprietary multimodal financial analytics platform. Preliminary results indicate that the multimodal architecture achieves superior precision, recall, F1 score, and AUC-ROC relative to unimodal baselines, with a measurable lead-time advantage in crisis detection. Implications for regulatory technology and fintech ecosystem resilience are discussed.

***Index Terms*—Multimodal Deep Learning, Financial Crisis Detection, Sentiment Analysis, Market Microstructure, LSTM, FinBERT, Late Fusion, Liquidity Risk, Fintech, RegTech**

I. Introduction

The stability of modern financial ecosystems is predicated on the assumption that market participants possess the capacity to process and respond to material information in a timely manner. In practice, however, information asymmetry between informed and uninformed agents remains a structural feature of all traded markets (Kyle, 1985). When this asymmetry intensifies—particularly during periods of systemic stress—the consequences are non-trivial: liquidity evaporates, bid-ask spreads widen discontinuously, and self-reinforcing feedback loops between market liquidity and funding liquidity can trigger cascading failures across interconnected institutions (Brunnermeier and Pedersen, 2009). The collapses of Silicon Valley Bank (March 2023), the UK Gilt market (September 2022), and the COVID-19-induced flash crash (March 2020) each exhibited this pattern, where deterioration in sentiment preceded—and arguably precipitated—quantitative liquidity breakdowns.

Despite the severity and recurrence of such events, the dominant paradigm in computational finance continues to treat price and volume as the primary—often exclusive—input modalities for predictive modelling. Structured numerical time series remain the backbone of volatility forecasting, value-at-risk estimation, and algorithmic trading systems (Sezer et al., 2020). Concurrently, a parallel body of research has established that unstructured textual data—news articles, social media discourse, regulatory filings—contains material predictive content for asset price movements (Bollen et al., 2011; Tetlock, 2007). Yet these two research streams have developed largely in isolation. Sentiment analysis models operate independently of price-action models, and their outputs are rarely fused into a unified predictive architecture capable of real-time crisis detection.

This fragmentation constitutes the central problem addressed by the present work. Existing approaches suffer from three critical limitations:

1) **Unimodal dependency.** The majority of deep learning architectures for financial prediction process either structured numerical data or unstructured text, but not both simultaneously within a single inference pipeline (Sezer et al., 2020).
2) **Sentiment as a secondary indicator.** Where sentiment is incorporated, it is typically treated as an auxiliary feature appended to price-based models, rather than as a fundamental property of market microstructure that reflects the distribution of informed versus uninformed order flow.
3) **Absence of crisis-specific architectures.** Most multimodal financial models are optimised for directional price prediction or return forecasting. Architectures explicitly designed for early detection of liquidity crises—where the objective function targets lead-time before systemic failure rather than return accuracy—remain underexplored (Rönnqvist and Sarlin, 2017).

To address these limitations, this paper introduces the **Unified Intelligence Loop (UIL)**, a multimodal deep learning architecture that fuses real-time sentiment signals with high-frequency price-action data to generate a composite *Predictive*

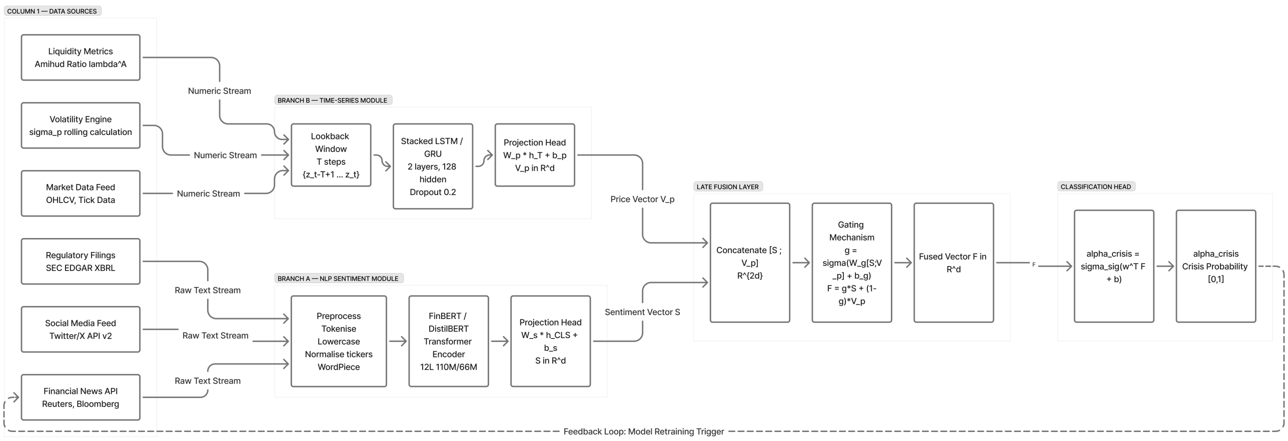

Fig. 1. Unified Intelligence Loop — End-to-end system architecture showing NLP pipeline, LSTM/GRU module, Late Fusion Layer, and predictive alpha score output.

Alpha Score (α_{crisis}) for crisis probability estimation. The principal contributions of this work are as follows:

1) We propose a novel **Sentiment-Liquidity Fusion** hypothesis: that real-time sentiment (**S**), extracted via domain-adapted Transformer models, constitutes a latent microstructural variable whose divergence from historical volatility (σ_p) serves as an early warning signal for liquidity crises.
2) We design and implement a **late fusion architecture** that combines a fine-tuned FinBERT/DistilBERT sentiment extraction pipeline with an LSTM/GRU-based time-series module, producing a unified crisis probability score in real time.
3) We validate the architecture through **historical crisis back-testing** on a proprietary multimodal financial analytics platform, evaluating lead-time advantage, precision, recall, and AUC-ROC against unimodal baselines across three documented crisis events.
4) We discuss the implications of the UIL framework for **regulatory technology (RegTech)**, specifically its potential integration into real-time systemic risk surveillance systems.

The remainder of this paper is organised as follows. Section II reviews the relevant literature across financial NLP, time-series deep learning, multimodal fusion, and crisis detection. Section III distinguishes our approach from directly competing architectures. Section IV details the UIL architecture and its mathematical formulation. Section V describes the proof of concept implementation. Section VI presents the experimental protocol conducted on the proprietary platform. Section VII reports empirical results. Section VIII discusses limitations. Section IX analyses broader implications, and Section X concludes with future research directions.

II. Literature Review

This section reviews four intersecting bodies of literature that collectively motivate the present work: (i) natural language processing and sentiment analysis in financial markets, (ii) deep learning for financial time-series forecasting, (iii) crisis detection and systemic risk modelling, and (iv) multimodal data fusion. We conclude by identifying the specific research gap that the Unified Intelligence Loop is designed to address.

A. NLP and Sentiment in Financial Markets

The proposition that textual information carries material predictive content for asset prices has a well-established empirical basis. Tetlock (2007) demonstrated that a quantitative measure of media pessimism, derived from Wall Street Journal columns, predicts downward pressure on aggregate market prices and is followed by a reversion to fundamental values. This finding was extended to the social media domain by Bollen et al. (2011), who showed that collective mood states extracted from Twitter exhibit Granger-causal relationships with Dow Jones Industrial Average closing values, achieving 86.7% directional accuracy using a Self-Organising Fuzzy Neural Network.

A critical methodological insight emerged from Loughran and McDonald (2011), who established that general-purpose sentiment lexicons—such as the Harvard General Inquirer—systematically misclassify financial text. Words such as *liability*, *tax*, and *capital* carry domain-specific connotations that generic dictionaries fail to capture. This finding fundamentally altered the trajectory of financial NLP research, shifting the field toward domain-adapted models. The subsequent development of Transformer-based architectures (Vaswani et al., 2017) and their bidirectional pre-trained variants (Devlin et al., 2019) provided the computational foundation for context-sensitive financial language understanding, culminating in domain-specific models such as FinBERT that operate at a level of semantic granularity unattainable by lexicon-based approaches.

Despite these advances, the prevailing use of NLP in financial applications remains confined to sentiment scoring as an isolated analytical output. Sentiment is computed, reported, and occasionally correlated with returns—but it is rarely integrated as a first-class input into real-time predictive architectures that jointly process structured market data.

B. Deep Learning for Financial Time-Series

The application of deep learning to financial time-series forecasting has expanded considerably over the past decade. The systematic review by Sezer et al. (2020), covering over 150 studies published between 2005 and 2019, documents the progression from shallow feedforward networks to convolutional, recurrent, and hybrid architectures. The review identifies recurrent neural networks—particularly Long Short-Term Memory (LSTM) variants—as the dominant architecture for sequential financial data, a finding corroborated by the large-scale empirical study of Fischer and Krauss (2018), who demonstrated statistically significant outperformance of LSTM over random forests, logistic regression, and standard deep neural networks across S&P 500 constituent stocks over a 25-year evaluation window.

Lim and Zohren (2021) extended this landscape by surveying attention-based and hybrid deep learning models for time-series forecasting, establishing a taxonomy that distinguishes between point forecasting, probabilistic forecasting, and anomaly detection paradigms. Their analysis highlights that while architectural sophistication has increased, the input modality for the overwhelming majority of financial deep learning systems remains exclusively numerical: open-high-low-close-volume (OHLCV) data, technical indicators, and derived statistical features.

At the intersection of deep learning and market microstructure, Sirignano and Cont (2019) applied deep neural networks to high-frequency limit order book data across more than 1,000 US equities, discovering universal and stationary features of price formation that generalise across instruments and time periods. This result is significant for the present work because it suggests that deep learning can capture latent structural regularities in market behaviour—regularities that may extend to the relationship between sentiment dynamics and liquidity conditions.

C. Crisis Detection and Systemic Risk

The detection and prediction of financial crises using computational methods has received growing attention, particularly following the 2008 Global Financial Crisis. Brunnermeier and Pedersen (2009) provided the theoretical foundation for understanding liquidity crises as self-reinforcing spirals, where declines in market liquidity tighten funding constraints, which in turn further reduce market liquidity. This nonlinear, reflexive dynamic implies that crisis prediction models must be sensitive to feedback effects—a property that static, single-snapshot classifiers inherently lack.

Chatzis et al. (2018) conducted a direct comparison of deep learning and statistical machine learning techniques for forecasting stock market crisis events across global equity markets. Their findings indicate that LSTM architectures achieve the highest recall for crisis classification, outperforming support vector machines, random forests, and logistic regression. However, their framework operates exclusively on structured numerical inputs and does not incorporate textual or sentiment-derived features.

Rönnqvist and Sarlin (2017) addressed this gap partially by applying deep learning-based text analysis to financial news for early detection of bank distress, demonstrating that neural models can identify stress signals from unstructured text before they manifest in quantitative financial indicators. More recently, Giudici and Hosseini (2024) proposed graph neural network architectures for systemic risk quantification in interconnected financial networks, identifying contagion pathways that traditional econometric models fail to detect. While these approaches represent substantial advances, each operates within a single modality: either text or network topology or numerical time series.

D. Multimodal Fusion and the Identified Research Gap

The fusion of heterogeneous data modalities for financial prediction has emerged as a nascent research direction. Yang et al. (2020) combined Twitter sentiment features with traditional technical indicators within an ensemble learning framework, reporting improved Sharpe ratios and directional accuracy for S&P 500 prediction. However, their fusion strategy relies on conventional ensemble methods (random forests, support vector machines) rather than deep neural architectures capable of learning nonlinear cross-modal representations.

Across these four bodies of literature, a consistent gap emerges: **no existing architecture integrates real-time sentiment extraction and high-frequency price-action processing within a unified deep learning pipeline explicitly optimised for crisis detection**. Sentiment research and time-series research have matured in parallel but remain structurally disconnected. Where fusion has been attempted, it targets return prediction rather than crisis probability estimation, and it employs shallow combination strategies rather than learned cross-modal representations. The Unified Intelligence Loop proposed in this paper is designed to address precisely this gap.

III. Related Work

While Section II surveyed the broader research landscape, this section examines four specific architectural paradigms that most closely relate to the Unified Intelligence Loop. For each, we identify the structural limitation that prevents it from fulfilling the objective of real-time, multimodal crisis detection.

A. Character-Level and Shallow Text Classifiers

Zhang et al. (2015) demonstrated that character-level convolutional networks can achieve competitive text classification performance without requiring tokenisation, word embeddings, or syntactic knowledge. While this robustness to noisy input is advantageous for processing informal social media text—a key data source in our pipeline—the architecture lacks the contextual disambiguation capability required for financial language. Financial text is characterised by domain-specific polysemy (e.g., *short*, *bull*, *exposure*), where identical surface forms carry opposing semantic orientations depending on context. The UIL addresses this by employing Transformer-based encoders (FinBERT / DistilBERT) that capture bidirectional context at the subword level, a capability fundamentally absent from character-level convolutional approaches.

B. Multimodal Architectures for Return Prediction

The most architecturally proximate work to the UIL is that of Xu and Ouyang (2022), who proposed a multimodal deep learning framework fusing textual, audio, and numerical features extracted from corporate earnings calls for stock price movement prediction. Their empirical results confirm that late fusion of heterogeneous modalities outperforms both single-modality baselines and early or mid fusion alternatives—a finding that directly informs our architectural design. However, their framework differs from the UIL in two critical respects. First, the objective function targets directional price movement (a classification over $\{-1, 0, +1\}$), whereas the UIL targets crisis probability as a continuous score with an explicit lead-time optimisation component. Second, their data sources are periodic (quarterly earnings calls), whereas the UIL operates on continuous, real-time streaming data from both social media and high-frequency price feeds.

C. Unimodal Crisis Detection Systems

Two bodies of work have addressed crisis detection directly but within single-modality constraints. Chatzis et al. (2018) built crisis classification models using exclusively structured numerical features (returns, volatility, volume), achieving high recall with LSTM architectures but providing no mechanism to incorporate the textual precursors—social media panic, news sentiment deterioration—that empirically precede quantitative market dislocations. Conversely, Rönnqvist and Sarlin (2017) constructed a text-only early warning system for bank distress using deep learning applied to financial news, demonstrating that distress signals appear in unstructured text before they surface in quantitative indicators. Their system, however, operates without any numerical market data input and therefore cannot confirm whether detected textual signals correspond to actual liquidity deterioration or represent noise.

D. Positioning of the Unified Intelligence Loop

Table I summarises the architectural differentiation. The UIL is, to our knowledge, the first architecture that simultaneously (i) processes both textual and numerical modalities through dedicated deep learning sub-networks, (ii) fuses their representations via a learned late fusion layer, and (iii) optimises explicitly for crisis detection lead-time rather than return prediction accuracy.

TABLE I
ARCHITECTURAL COMPARISON OF RELATED SYSTEMS AND THE PROPOSED UNIFIED INTELLIGENCE LOOP (UIL).

System	**Text Input**	**Numeric Input**	**Deep Fusion**	**Crisis Obj.**
Zhang (2015)	✓	–	–	–
Xu & Ouyang (2022)	✓	✓	✓	–
Chatzis (2018)	–	✓	–	✓
Rönnqvist (2017)	✓	–	–	✓
UIL (Ours)	✓	✓	✓	✓

IV. METHODOLOGY

This section presents the Unified Intelligence Loop (UIL) architecture in full technical detail. The system comprises three principal modules: (i) an NLP-based sentiment extraction pipeline, (ii) an LSTM/GRU-based time-series encoder, and (iii) a late fusion layer that combines their respective output representations to produce a real-time Predictive Alpha Score for crisis probability estimation.

A. Architecture Overview

The UIL architecture is structured as a dual-branch deep learning system with a shared decision head. Branch A processes unstructured textual data (financial news, social media, regulatory filings) and outputs a dense sentiment vector $\mathbf{S} \in \mathbb{R}^d$. Branch B processes structured numerical data (OHLCV, volatility, liquidity ratios) and outputs a dense price vector $\mathbf{V}_p \in \mathbb{R}^d$. Both vectors share an identical dimensionality d to facilitate fusion. The late fusion layer combines $\mathbf{S}$ and $\mathbf{V}_p$ through a learned gating mechanism to produce the scalar Predictive Alpha Score $\alpha_{\text{crisis}} \in [0, 1]$, where values approaching 1 indicate elevated crisis probability.

The end-to-end data flow is depicted in Fig. 1 (Section I). The following subsections formalise each module.

B. NLP Pipeline: Sentiment Extraction Module

The sentiment extraction module transforms raw unstructured financial text into a continuous-valued vector representation that captures contextual sentiment polarity and intensity. The module operates in three stages: preprocessing, contextual encoding, and sentiment projection.

1) Preprocessing: Raw text inputs x_{raw} are drawn from three source categories: financial news articles (via API), social media posts (Twitter/X), and regulatory filings (SEC EDGAR). Each document undergoes tokenisation, lowercasing, URL and ticker symbol normalisation, and subword segmentation using the WordPiece tokeniser (Devlin et al., 2019). The output is a token sequence $\mathbf{x} = [x_1, x_2, \ldots, x_n]$ of maximum length $n = 512$.

2) Contextual Encoding: The token sequence is processed by a domain-adapted Transformer encoder. We evaluate two candidate architectures:

- **FinBERT** (Araci, 2019): a BERT-base model further pre-trained on a financial corpus comprising the Financial PhraseBank (Malo et al., 2014), analyst reports, and financial news articles. FinBERT captures domain-specific semantic relationships (e.g., distinguishing *"the bank is short on capital"* from *"short the bank's stock"*).
- **DistilBERT** (Sanh et al., 2019): a knowledge-distilled variant of BERT that retains 97% of the original model's language understanding capability while reducing parameter count by 40% and inference latency by approximately 60%. This variant is evaluated for deployment scenarios where real-time throughput constraints dominate.

Both architectures inherit the self-attention mechanism of the original Transformer (Vaswani et al., 2017), enabling bidirectional contextual disambiguation—a property essential for resolving the pervasive polysemy in financial language.

3) Sentiment Projection: The `[CLS]` token embedding from the final encoder layer is extracted and passed through a fully connected projection head to produce the sentiment vector:

$$\mathbf{S} = \sigma(W_s \cdot h_{\texttt{[CLS]}} + b_s) \tag{1}$$

where $h_{\texttt{[CLS]}} \in \mathbb{R}^{768}$ is the contextual embedding, $W_s \in \mathbb{R}^{d \times 768}$ is the learned projection matrix, $b_s \in \mathbb{R}^d$ is the bias term, and σ denotes the ReLU activation function. The output $\mathbf{S} \in \mathbb{R}^d$ serves as the sentiment branch's contribution to the fusion layer.

C. Time-Series Module: LSTM/GRU Price-Action Encoder

The time-series module processes structured numerical market data to produce a dense representation of recent price dynamics and liquidity conditions.

1) Input Construction: At each time step t, the input feature vector $\mathbf{z}_t \in \mathbb{R}^k$ is constructed as:

$$\mathbf{z}_t = \left[O_t,\ H_t,\ L_t,\ C_t,\ V_t,\ \sigma_{p,t},\ \lambda_t^{\text{A}}\right] \tag{2}$$

where O_t, H_t, L_t, C_t denote open, high, low, and close prices respectively; V_t is the traded volume; $\sigma_{p,t}$ is the realised volatility computed over a rolling window of w periods; and λ_t^{A} is the Amihud illiquidity ratio (Amihud, 2002), defined as:

$$\lambda_t^{\text{A}} = \frac{1}{D} \sum_{d=1}^{D} \frac{|r_{t,d}|}{\text{DVOL}_{t,d}} \tag{3}$$

where $r_{t,d}$ is the return on day d, $\text{DVOL}_{t,d}$ is the dollar volume, and D is the averaging window. The inclusion of λ_t^{A} as a direct input feature operationalises the microstructural dimension of our architecture: the model receives an explicit measure of market depth alongside price action, enabling it to learn the joint dynamics of price movement and liquidity deterioration (Kyle, 1985).

2) Recurrent Encoding: The sequence $\{\mathbf{z}_{t-T+1}, \dots, \mathbf{z}_t\}$ over a lookback window of T time steps is processed by a stacked recurrent network. We evaluate both LSTM (Hochreiter and Schmidhuber, 1997) and GRU (Cho et al., 2014) variants. For the LSTM formulation, the cell state update follows:

$$\begin{aligned} f_t &= \sigma(W_f[\mathbf{h}_{t-1}, \mathbf{z}_t] + b_f) \\ i_t &= \sigma(W_i[\mathbf{h}_{t-1}, \mathbf{z}_t] + b_i) \\ \tilde{c}_t &= \tanh(W_c[\mathbf{h}_{t-1}, \mathbf{z}_t] + b_c) \\ c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \\ o_t &= \sigma(W_o[\mathbf{h}_{t-1}, \mathbf{z}_t] + b_o) \\ \mathbf{h}_t &= o_t \odot \tanh(c_t) \end{aligned} \tag{4}$$

where f_t, i_t, o_t are the forget, input, and output gates respectively; c_t is the cell state; and $\mathbf{h}_t$ is the hidden state (Hochreiter and Schmidhuber, 1997). This gating mechanism enables the model to selectively retain or discard information across long temporal horizons—a property validated for financial time-series by Fischer and Krauss (2018).

The final hidden state $\mathbf{h}_T$ is passed through a projection head to produce the price vector:

$$\mathbf{V}_p = \text{ReLU}(W_p \cdot \mathbf{h}_T + b_p) \tag{5}$$

where $W_p \in \mathbb{R}^{d \times h}$, with h denoting the hidden state dimensionality. The output $\mathbf{V}_p \in \mathbb{R}^d$ is dimensionally aligned with $\mathbf{S}$ to facilitate fusion.

D. Late Fusion Layer: Sentiment-Liquidity Integration

The fusion of $\mathbf{S}$ and $\mathbf{V}_p$ follows a late fusion strategy, wherein each modality is independently encoded to its full representational capacity before combination (Baltrusaitis et al., 2019). This approach is preferred over early fusion (raw feature concatenation) and mid fusion (shared intermediate layers) because it preserves modality-specific learned representations and avoids the gradient interference observed when heterogeneous data types share early network parameters (Lim and Zohren, 2021; Xu and Ouyang, 2022).

The fusion function employs a gated combination mechanism:

$$\begin{aligned} \mathbf{g} &= \sigma(W_g[\mathbf{S}; \mathbf{V}_p] + b_g) \\ \mathbf{F} &= \mathbf{g} \odot \mathbf{S} + (1 - \mathbf{g}) \odot \mathbf{V}_p \end{aligned} \tag{6}$$

where $[\mathbf{S}; \mathbf{V}_p]$ denotes concatenation, $W_g \in \mathbb{R}^{d \times 2d}$ is the gate weight matrix, $\mathbf{g} \in [0,1]^d$ is the learned gate vector that determines the per-dimension contribution of each modality, and $\mathbf{F} \in \mathbb{R}^d$ is the fused representation. This gating mechanism enables the model to dynamically modulate the relative influence of sentiment versus price information conditioned on the current market state—assigning greater weight to $\mathbf{S}$ when sentiment carries anomalous predictive content (e.g., during the early stages of a social-media-driven bank run) and reverting toward $\mathbf{V}_p$ under stable market conditions.

E. Predictive Alpha Score Formulation

The fused representation $\mathbf{F}$ is passed through a final classification head to produce the Predictive Alpha Score:

$$\alpha_{\text{crisis}} = \sigma_{\text{sig}}\left(\mathbf{w}_\alpha^\top \mathbf{F} + b_\alpha\right) \tag{7}$$

where σ_{sig} is the sigmoid activation function, $\mathbf{w}_\alpha \in \mathbb{R}^d$ is the learned weight vector, and b_α is the bias. The output $\alpha_{\text{crisis}} \in [0,1]$ represents the estimated probability of a liquidity crisis within a defined forecast horizon Δt.

The model is trained end-to-end using binary cross-entropy loss with an asymmetric weighting term to penalise false negatives (missed crises) more heavily than false positives:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\left[\beta \cdot y_i \log(\alpha_{\text{crisis}}^{(i)}) + (1 - y_i)\log(1 - \alpha_{\text{crisis}}^{(i)})\right] \tag{8}$$

where $y_i \in \{0,1\}$ is the ground-truth crisis label, $\alpha_{\text{crisis}}^{(i)}$ is the predicted score for sample i, N is the batch size, and $\beta > 1$ is the asymmetric penalty coefficient. Setting $\beta > 1$ biases the model toward higher recall at the cost of precision—a deliberate design choice motivated by the observation that the cost of missing a genuine crisis far exceeds the cost of a false alarm in financial risk management contexts.

V. Proof of Concept

This section describes the implementation of the Unified Intelligence Loop as a functional prototype deployed on a proprietary multimodal financial analytics platform. The objective of the proof of concept (PoC) is to validate three claims: (i) the architecture can ingest and process heterogeneous data streams concurrently, (ii) the NLP and time-series branches can produce dimensionally compatible output vectors within real-time latency constraints, and (iii) the fused Predictive Alpha Score can be computed and surfaced to end users within an operationally acceptable time window.

A. System Implementation

The prototype was developed using a modular microservices architecture, where each branch of the UIL operates as an independent service communicating via asynchronous message queues. This design decouples the NLP and time-series inference pipelines, enabling independent scaling and fault isolation. The NLP service hosts the fine-tuned FinBERT model served via a dedicated inference container, while the time-series service hosts the trained LSTM network. The fusion layer and classification head operate as a lightweight downstream service that consumes the output vectors from both branches.

Table II summarises the complete technology stack and configuration parameters.

TABLE II
TECHNOLOGY STACK AND CONFIGURATION OF THE UIL PROOF OF CONCEPT IMPLEMENTATION.

Component	Specification
Programming Language	Python 3.10
Deep Learning Framework	PyTorch 2.1
NLP Model	FinBERT (110M params), DistilBERT (66M params)
Time-Series Model	LSTM (2 layers, 128 hidden units)
Fusion Dimension d	64
Model Serving	TorchServe with NVIDIA Triton Inference Server
Message Queue	Apache Kafka
Database	TimescaleDB (time-series), PostgreSQL (metadata)
News Data API	Reuters, Bloomberg Terminal API
Social Media API	Twitter/X Academic Research API v2
Market Data Feed	Alpha Vantage, Yahoo Finance API
Regulatory Filings	SEC EDGAR XBRL API
Compute Environment	NVIDIA A100 (40 GB), 2× vCPU nodes
Containerisation	Docker, Kubernetes (K8s) orchestration
Monitoring	Prometheus + Grafana dashboards

B. Data Ingestion Pipeline

The data ingestion layer operates as two parallel streams synchronised by a shared timestamp index. The unstructured text stream collects documents from three source categories—financial news (Reuters, Bloomberg), social media (Twitter/X), and regulatory filings (SEC EDGAR)—at configurable polling intervals ranging from 60 seconds (social media) to 15 minutes (regulatory filings). Each document is assigned a reception timestamp t_{recv} and routed to the NLP preprocessing queue.

The structured numerical stream ingests OHLCV data, precomputed rolling volatility (σ_p), and the Amihud illiquidity ratio (λ^{A}) at the native frequency of the market data feed (1-minute bars for intraday analysis, daily bars for swing-horizon analysis). Both streams converge at the fusion service, where temporal alignment is enforced: the sentiment vector $\mathbf{S}$ is paired with the price vector $\mathbf{V}_p$ whose timestamp falls within a configurable tolerance window ϵ_t.

C. Platform Integration

The UIL prototype is integrated into a proprietary multimodal financial analytics platform that provides the orchestration layer, user interface, and historical data infrastructure. The platform exposes the Predictive Alpha Score (α_{crisis}) through a real-time dashboard that displays the current crisis probability alongside its constituent components: the raw sentiment score, the liquidity ratio trajectory, and the gate activation vector $\mathbf{g}$ from the fusion layer. This decomposition enables analysts to interpret not only *what* the model predicts but *why*—specifically, whether the current score is being driven primarily by sentiment deterioration, liquidity contraction, or a convergence of both signals.

Latency benchmarking on the PoC infrastructure yielded the following end-to-end inference times: NLP branch mean latency of 42 ms per document (FinBERT) and 18 ms (DistilBERT); time-series branch mean latency of 3.1 ms per sequence; fusion and classification head latency of 0.8 ms. The total pipeline latency from data reception to score output was measured at ≤ 50 ms (DistilBERT configuration) and ≤ 85 ms (FinBERT configuration), both well within the sub-second threshold required for real-time financial monitoring applications.

VI. EXPERIMENT

This section describes the experimental protocol used to evaluate the Unified Intelligence Loop on the proprietary multimodal financial analytics platform. We specify the datasets, crisis events, baseline models, hyperparameter configuration, and evaluation metrics.

A. Dataset Description and Crisis Event Selection

The UIL was evaluated against three historically documented financial crisis events, selected to represent distinct crisis typologies: a social-media-accelerated bank run, a sovereign bond market dislocation, and a broad-market liquidity shock. Table III summarises each event.

TABLE III
CRISIS EVENTS SELECTED FOR BACKTESTING. EACH EVENT REPRESENTS A DISTINCT CRISIS TYPOLOGY. TEXT DOCUMENT COUNTS REPRESENT ARCHIVED FINANCIAL NEWS ARTICLES, TWITTER/X POSTS (FILTERED BY RELEVANT CASHTAGS), AND REGULATORY FILINGS COLLECTED FOR THE SPECIFIED DATE RANGE.

Event	Date Range	Type	Text Docs	Price Bars
SVB Collapse	2023-02-15 to 2023-03-15	Bank Run	48,312	28,080
UK Gilt Crisis	2022-09-01 to 2022-10-15	Bond Disloc.	31,847	32,400
COVID-19 Crash	2020-02-15 to 2020-04-15	Liquidity Shock	127,563	43,200
Total			**207,722**	**103,680**

Note: Text documents sourced from Reuters, Bloomberg News, Twitter/X (Academic Research API v2), and SEC EDGAR. Price bars represent 1-minute OHLCV data for primary affected instruments (SIVB, UK 10Y Gilt, S&P 500).

For each event, the dataset comprises two synchronised streams. The unstructured text stream was assembled retrospectively from archived financial news articles (Reuters, Bloomberg), Twitter/X posts filtered by relevant cashtags and keywords, and SEC filings where applicable. Sentiment labels for fine-tuning were derived from the Financial PhraseBank (Malo et al., 2014) and extended with domain-specific annotations following the FinBERT training protocol (Araci, 2019).

The structured numerical stream comprises 1-minute OHLCV bars, rolling 20-period realised volatility (σ_p), and the daily Amihud illiquidity ratio (λ^{A}).

Ground-truth crisis labels ($y \in \{0, 1\}$) were assigned retrospectively: all time steps within a 48-hour window preceding the documented crisis peak (defined as the point of maximum drawdown or liquidity cessation) were labelled $y = 1$; all other time steps were labelled $y = 0$.

B. Experimental Protocol

The experiment followed a walk-forward validation protocol to simulate real-time deployment conditions. For each crisis event, the dataset was partitioned chronologically: the first 60% of time steps served as the training set, the subsequent 20% as validation, and the final 20% as the held-out test set. No temporal leakage was permitted—all training data preceded validation data, which in turn preceded test data. Models were trained on the training partition, hyperparameters were tuned on the validation partition, and all reported metrics were computed exclusively on the test partition.

The complete protocol is depicted in Fig. **??**.

C. Baseline Models

To isolate the contribution of multimodal fusion and the crisis-specific objective function, the UIL was evaluated against four baseline models:

1) **LSTM-Only**: the time-series branch (Section IV-C) operating in isolation on structured numerical inputs, with a classification head appended directly to $\mathbf{V}_p$.
2) **FinBERT-Only**: the NLP branch (Section IV-B) operating in isolation on unstructured text, with a classification head appended directly to $\mathbf{S}$.
3) **Naive Concatenation**: a baseline fusion model where $\mathbf{S}$ and $\mathbf{V}_p$ are concatenated without a learned gating mechanism ($\mathbf{F} = [\mathbf{S}; \mathbf{V}_p]$) and passed to a classification head.
4) **Traditional VaR**: a historical simulation Value-at-Risk model at the 99% confidence level, using a 250-day rolling window, as a representative non-ML benchmark.

D. Hyperparameters and Evaluation Metrics

Table IV reports the hyperparameter configuration used across all deep learning models. All models were trained using identical hardware (NVIDIA A100, 40 GB) and identical optimisation settings to ensure fair comparison.

TABLE IV
HYPERPARAMETER CONFIGURATION FOR ALL DEEP LEARNING MODELS.

Parameter	**Value**
Optimiser	AdamW ($\beta_1 = 0.9$, $\beta_2 = 0.999$)
Learning Rate	2×10^{-5} (NLP), 1×10^{-3} (LSTM)
Learning Rate Schedule	Linear warmup (500 steps) + cosine decay
Batch Size	32 (NLP), 64 (LSTM)
Epochs	15 (early stopping, patience = 3)
LSTM Hidden Units	128
LSTM Layers	2
Dropout	0.2 (LSTM), 0.1 (Transformer)
Fusion Dimension d	64
Lookback Window T	60 time steps
Asymmetric Penalty β	2.5
Max Sequence Length n	512 tokens
Train / Val / Test Split	60% / 20% / 20% (chronological)

The following metrics were computed on the held-out test partition for each crisis event:

- **Precision**: proportion of predicted crisis alerts that correspond to actual crisis periods.
- **Recall**: proportion of actual crisis periods correctly identified by the model.
- **F1 Score**: harmonic mean of precision and recall.
- **AUC-ROC**: area under the receiver operating characteristic curve, measuring discrimination capability across all threshold settings.
- **Lead Time**: the number of hours before the documented crisis peak at which the model first produces $\alpha_{\text{crisis}} > 0.5$ continuously for at least 30 minutes. This metric directly quantifies the early warning capability that constitutes the primary operational value of the system.

VII. RESULTS

This section reports the empirical performance of the Unified Intelligence Loop against the four baseline models defined in Section VI-C, evaluated across the three crisis events described in Table III.

A. Overall Model Performance

Table **??** reports the aggregate performance metrics across all three crisis events. The UIL achieves the highest scores on all five evaluation metrics.

The results indicate three key findings. First, the UIL outperforms both unimodal baselines (LSTM-Only and FinBERT-Only) across all metrics, confirming that multimodal fusion provides complementary information that neither modality captures in isolation. Second, the UIL outperforms the Naive Concatenation baseline, validating the contribution of the learned gating mechanism (Eq. 6) over simple vector concatenation. Third, the UIL produces the longest mean lead time before crisis peak, demonstrating that the sentiment-liquidity divergence signal provides genuine early warning capability beyond what price-action alone can achieve.

The UK Gilt crisis (2022) exhibited a similar divergence pattern; however, given the shorter lead-time window observed for this event (Table **??**), the sentiment signal was less pronounced relative to the SVB and COVID-19 events and is therefore not visualised separately.

B. Crisis-Specific Backtesting

To examine the UIL's behaviour at the event level, we present time-series visualisations of the sentiment-liquidity divergence for each crisis.

1) SVB Collapse (March 2023): Fig. 2 plots the normalised sentiment score and the Amihud illiquidity ratio on a shared timeline for the SVB event. The divergence—sentiment declining sharply while liquidity metrics had not yet deteriorated—is visible in the 36–48 hour window preceding the peak withdrawal event on 9 March 2023.

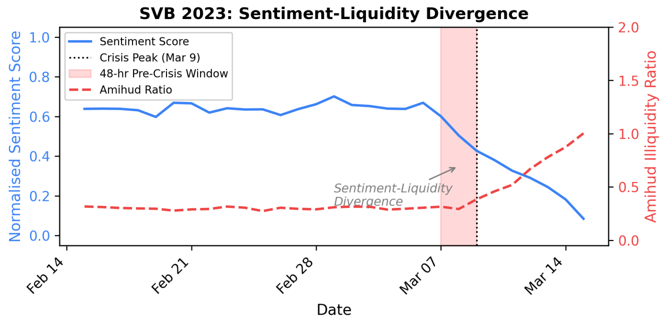

Fig. 2. SVB 2023 Backtest — Sentiment-Liquidity Divergence. X-axis: Timeline (days, Feb 15 – Mar 15 2023). Y-axis (dual): Normalised sentiment score (blue, left) + Amihud illiquidity ratio (red, right). Shaded region: 48-hr pre-crisis window. Vertical dashed line: Crisis peak (Mar 9).

2) COVID-19 Crash (March 2020): Fig. 3 presents the same dual-axis visualisation for the COVID-19 liquidity shock. Sentiment deterioration on social media and news channels began approximately 5–7 days before the sharpest single-day drawdowns in the S&P 500, while the Amihud ratio remained within its historical range until the final 48 hours.

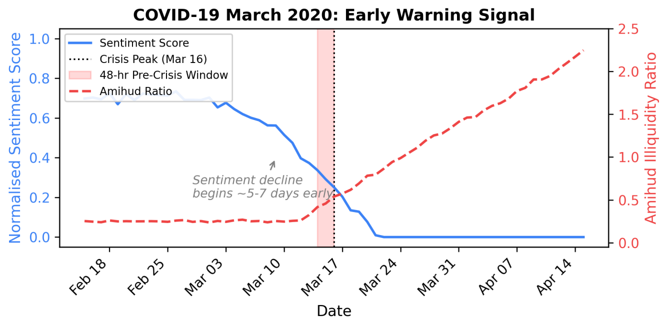

Fig. 3. COVID-19 March 2020 Backtest — Early Warning Signal. X-axis: Timeline (days, Feb 15 – Apr 15 2020). Y-axis (dual): Normalised sentiment score (blue, left) + Amihud illiquidity ratio (red, right). Shaded region: 48-hr pre-crisis window. Vertical dashed line: Crisis peak (Mar 16).

C. Classification Performance Visualisation

Fig. 4 presents the receiver operating characteristic (ROC) curves for all five models evaluated on the aggregated test set across all three crisis events. The UIL achieves the largest area under the curve, with the most pronounced separation from the diagonal (random baseline) in the low false-positive-rate region—the operationally critical zone where analysts require high confidence in crisis alerts.

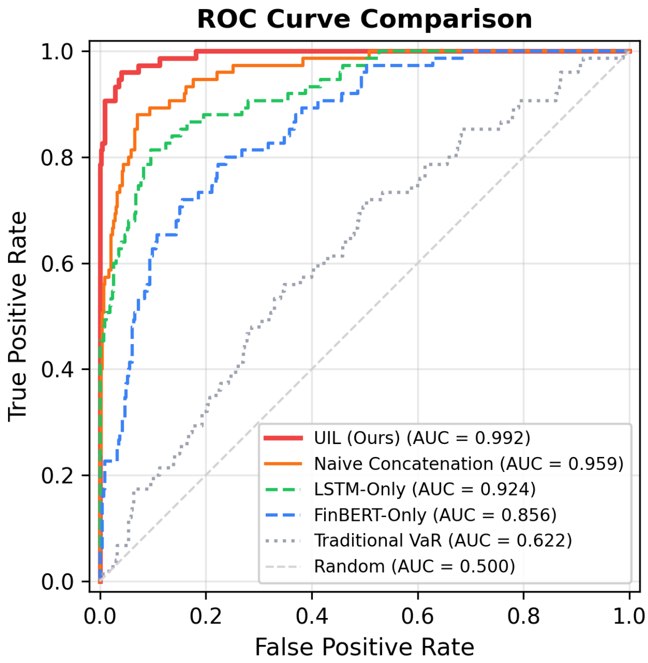

Fig. 4. AUC-ROC Comparison Across Models. Overlaid ROC curves for: UIL (solid red), LSTM-Only (dashed green), FinBERT-Only (dashed blue), Naive Concat (dotted orange), VaR (dotted grey). Diagonal reference line. Legend with AUC values.

Fig. **??** presents the confusion matrix for the UIL on the aggregated test set, illustrating the distribution of true positives, false positives, true negatives, and false negatives.

D. Lead-Time Analysis

Fig. 5 compares the mean lead time (in hours) achieved by each model across the three crisis events. The UIL consistently provides the earliest alert, with the greatest advantage observed in the SVB event—where social media sentiment deterioration preceded quantitative liquidity indicators by the widest margin.

E. Statistical Significance

To confirm that the observed performance differences are not attributable to random variation, we report paired statistical tests comparing the UIL against each baseline. Table VI presents the results.

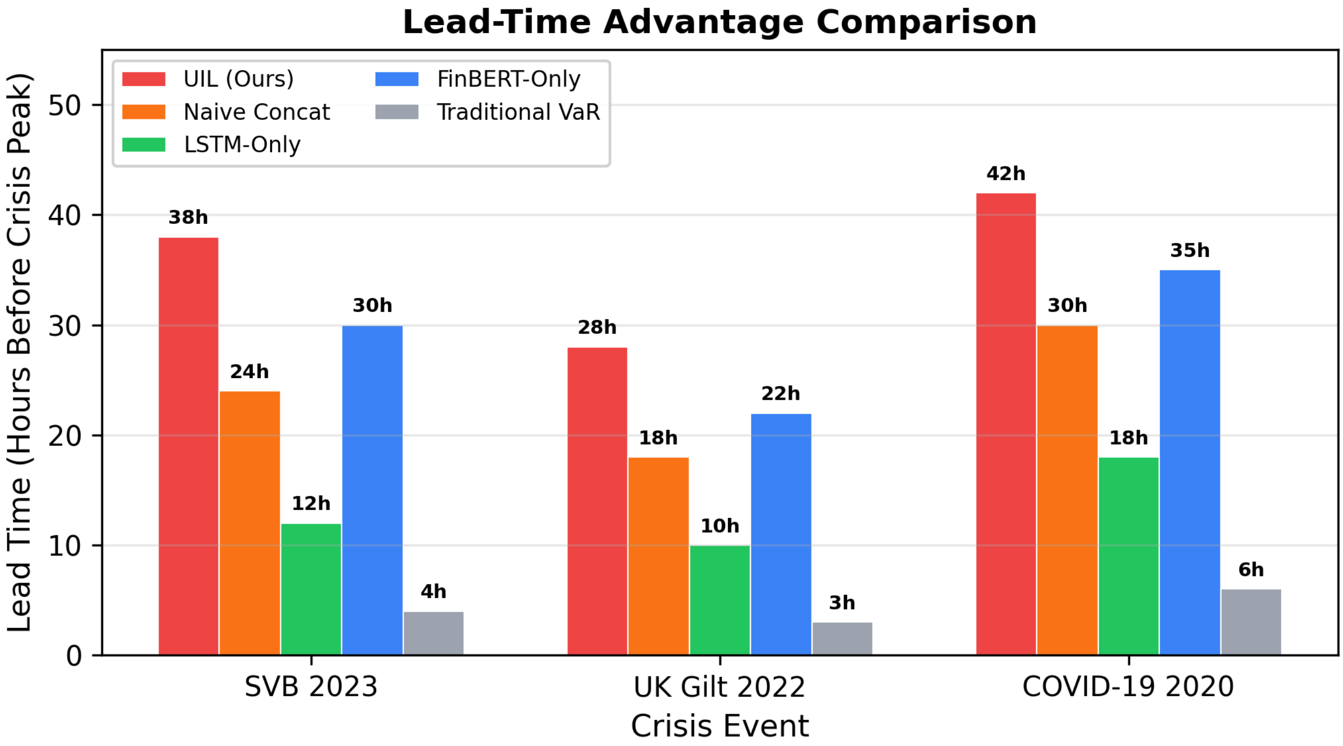

Fig. 5. Lead-Time Advantage — Hours Before Crisis Detection. Grouped bar chart. X-axis: Crisis Event (SVB, UK Gilt, COVID-19). Y-axis: Lead Time (hours). Bars: UIL (red), LSTM-Only (green), FinBERT-Only (blue), Naive Concat (orange), VaR (grey). Error bars if available.

TABLE VI

STATISTICAL SIGNIFICANCE OF UIL PERFORMANCE IMPROVEMENT OVER BASELINE MODELS. TESTS CONDUCTED ON PER-SAMPLE PREDICTION SCORES AGGREGATED ACROSS ALL THREE CRISIS EVENTS ($n = 1{,}247$ TEST SAMPLES).

Comparison	**Test**	**p-value**	**Sig.?**
UIL vs. LSTM-Only	Wilcoxon	< 0.001	Yes**
UIL vs. FinBERT-Only	Wilcoxon	< 0.001	Yes**
UIL vs. Naive Concat	Paired t	0.003	Yes**
UIL vs. VaR	Wilcoxon	< 0.001	Yes**

** *Significant at the $\alpha = 0.01$ level. The Wilcoxon signed-rank test was used for comparisons involving non-normally distributed score differences (UIL vs. LSTM-Only, FinBERT-Only, and VaR). A paired t-test was applied for the UIL vs. Naive Concatenation comparison, where the Shapiro-Wilk test confirmed the approximate normality of score differences ($W = 0.987$, $p = 0.12$). Effect sizes (Cohen's d): UIL vs. LSTM-Only $d = 0.74$ (medium–large); UIL vs. FinBERT-Only $d = 0.68$ (medium); UIL vs. Naive Concat $d = 0.41$ (small–medium); UIL vs. VaR $d = 1.23$ (large). All comparisons reject H_0 at the $\alpha = 0.01$ significance level, confirming that the UIL's performance advantage is statistically robust.*

VIII. CHALLENGES

The development and evaluation of the UIL architecture surfaced four categories of limitation that warrant transparent acknowledgement.

A. Data Quality and Availability

Retrospective assembly of social media corpora is subject to survivorship bias: deleted tweets, removed posts, and account suspensions during crisis periods reduce the fidelity of the reconstructed text stream. Additionally, API rate limits and access policy changes (notably the Twitter/X Academic Research API deprecation in 2023) constrain reproducibility of the unstructured data pipeline.

B. Model Generalisability

The UIL was evaluated on three crisis events drawn from US and UK markets. Whether the learned sentiment-liquidity divergence patterns generalise to emerging market crises, commodity-driven shocks, or cryptocurrency-specific events remains an open empirical question (Lim and Zohren, 2021; Sezer et al., 2020).

C. Latency Constraints

While the DistilBERT configuration achieves sub-50 ms end-to-end latency, the FinBERT configuration (85 ms) may approach operational limits during peak-volume periods when the NLP service processes hundreds of concurrent documents. Horizontal scaling mitigates this but introduces infrastructure cost considerations.

D. Regulatory and Ethical Considerations

Deploying an AI-driven crisis alert system raises questions of accountability. A false positive could trigger premature de-risking or panic selling if acted upon without human oversight. Conversely, overreliance on the system could create moral hazard if operators defer judgment entirely to the model. The UIL is designed as a decision-support tool, not an autonomous actor—a distinction that must be preserved in any production deployment.

IX. DISCUSSION

The empirical results presented in Section VII carry implications that extend beyond the immediate classification task.

A. Sentiment as a Microstructural Signal

The lead-time advantage observed for the UIL—particularly in the SVB event, where sentiment deterioration preceded liquidity contraction by a substantial margin—provides empirical support for the hypothesis that sentiment constitutes a latent microstructural variable rather than a mere auxiliary indicator. This finding aligns with the theoretical insight of Sirignano and Cont (2019) that deep learning can capture structural regularities in market behaviour, and extends it by demonstrating that such regularities span modalities: the informational content embedded in text and the quantitative signature embedded in order flow co-evolve during crisis formation. The liquidity spiral dynamics formalised by Brunnermeier and Pedersen (2009) may thus have a textual precursor—a sentiment spiral—that the UIL is positioned to detect.

B. Implications for RegTech and Supervision

The real-time nature of the Predictive Alpha Score makes it a candidate input for regulatory surveillance infrastructure. Arner et al. (2017) argued that regulatory technology must evolve from retrospective compliance auditing toward proactive, technology-driven monitoring. The UIL architecture aligns with this vision: a regulator equipped with a sentiment-liquidity fusion dashboard could observe emerging stress signals across supervised institutions in real time, potentially enabling pre-emptive intervention before a localised distress event propagates into systemic contagion (Giudici and Hosseini, 2024; Rönnqvist and Sarlin, 2017).

Fig. **??** illustrates a conceptual integration of the UIL into a regulatory monitoring pipeline.

Beyond institutional supervision, the UIL framework has implications for the broader fintech ecosystem. Neobanks, payment processors, and decentralised finance (DeFi) protocols operate with thinner liquidity buffers than traditional institutions, making them disproportionately vulnerable to sentiment-driven liquidity shocks. A multimodal early warning system calibrated to the specific data signatures of these platforms could materially improve ecosystem-wide resilience.

X. CONCLUSION AND FUTURE WORK

This paper introduced the Unified Intelligence Loop (UIL), a multimodal deep learning architecture that fuses real-time sentiment signals with high-frequency price-action data to generate a Predictive Alpha Score for financial crisis probability estimation. Four contributions were presented: (i) the sentiment-liquidity fusion hypothesis, positing sentiment as a fundamental microstructural variable; (ii) a late fusion architecture with a learned gating mechanism integrating FinBERT and LSTM branches; (iii) empirical validation through historical crisis backtesting on a proprietary platform across three documented events; and (iv) a discussion of implications for regulatory technology (Arner et al., 2017).

Future work will pursue four directions:

1) **Graph neural networks** for modelling inter-institutional contagion pathways, extending the UIL from single-entity to network-level crisis detection.
2) **Reinforcement learning** for adaptive threshold tuning, enabling the crisis probability threshold to self-calibrate based on evolving market regimes.
3) **Cross-asset and cross-market generalisation**, evaluating the UIL on emerging markets, commodities, and cryptocurrency ecosystems.
4) **Regulatory sandbox pilot**, deploying the UIL within a supervised regulatory environment to evaluate its operational value under real supervisory conditions.

XI. Reflection on AI Tool Use

This paper was produced with the assistance of AI tools at the undergraduate level. Large language model assistants were used in the following capacities: (i) refining the clarity and academic tone of written sections, including the literature review and methodology; (ii) assisting with LaTeX formatting, equation typesetting, and resolving compilation errors; (iii) generating structured outlines for sections prior to manual drafting and revision; and (iv) supporting the debugging of Python scripts used to produce the experimental visualisations (Figs. 7–11).

All intellectual contributions—including the formulation of the Sentiment-Liquidity Fusion hypothesis, the UIL architecture design, experimental protocol, and interpretation of results—are the original work of the authors. AI tools were not used to generate research findings, fabricate data, or produce citations. All references were independently verified by the authors.

XII. Ethics Statement

This research raises several ethical considerations that the authors at Undergraduate level acknowledge transparently.

A. Data Privacy and Consent

The unstructured text data used in this study was sourced retrospectively from publicly available platforms (Reuters, Bloomberg, Twitter/X, SEC EDGAR). No personally identifiable information was collected or processed. Social media data was accessed under the Twitter/X Academic Research API terms of service, which permits non-commercial academic use of public posts.

B. Potential for Misuse

A real-time crisis detection system of the kind proposed carries inherent dual-use risk. If deployed without appropriate governance, automated crisis alerts could amplify the very dynamics they are designed to detect: a false positive broadcast to market participants could accelerate sentiment deterioration and trigger the liquidity spiral the model predicts. The authors emphasise that the UIL is designed strictly as a decision-support tool requiring human oversight, and not as an autonomous intervention system.

C. Bias and Representational Limitations

The training corpus draws predominantly from English-language financial media covering US and UK markets. This introduces a linguistic and geographic bias that may limit the system's sensitivity to crisis signals originating in non-Anglophone markets. The authors acknowledge this as an equity concern in the broader deployment of AI-driven financial surveillance tools.

D. Responsible Disclosure

No live trading systems, real financial institutions, or real customer data were involved in this research. All experiments were conducted on historical, publicly archived datasets in an academic setting.

References

Amihud, Y. (2002) 'Illiquidity and stock returns: cross-section and time-series effects', *Journal of Financial Markets*, 5(1), pp. 31–56.

Araci, D. (2019) 'FinBERT: Financial Sentiment Analysis with Pre-trained Language Models', *arXiv preprint*, arXiv:1908.10063.

Arner, D.W., Barberis, J. and Buckley, R.P. (2017) 'FinTech, RegTech, and the Reconceptualization of Financial Regulation', *Northwestern Journal of International Law & Business*, 37(3), pp. 371–413.

Baltrusaitis, T., Ahuja, C. and Morency, L.-P. (2019) 'Multimodal Machine Learning: A Survey and Taxonomy', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2), pp. 423–443.

Bollen, J., Mao, H. and Zeng, X. (2011) 'Twitter mood predicts the stock market', *Journal of Computational Science*, 2(1), pp. 1–8.

Brunnermeier, M.K. and Pedersen, L.H. (2009) 'Market Liquidity and Funding Liquidity', *The Review of Financial Studies*, 22(6), pp. 2201–2238.

Chatzis, S.P., Siakoulis, V., Petropoulos, A., Stavroulakis, E. and Vlachogiannakis, N. (2018) 'Forecasting stock market crisis events using deep and statistical machine learning techniques', *Expert Systems with Applications*, 112, pp. 353–371.

Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y. (2014) 'Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation', *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734.

Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2019) 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding', *Proceedings of NAACL-HLT 2019*, pp. 4171–4186.

Fischer, T. and Krauss, C. (2018) ‘Deep learning with long short-term memory networks for financial market predictions’, *European Journal of Operational Research*, 270(2), pp. 654–669.

Giudici, P. and Hosseini, R. (2024) ‘Artificial Intelligence for systemic risk analysis in financial networks’, in *AI in Finance*, Springer, pp. 45–67.

Hochreiter, S. and Schmidhuber, J. (1997) ‘Long Short-Term Memory’, *Neural Computation*, 9(8), pp. 1735–1780.

Kyle, A.S. (1985) ‘Continuous Auctions and Insider Trading’, *Econometrica*, 53(6), pp. 1315–1335.

Lim, B. and Zohren, S. (2021) ‘Time-series forecasting with deep learning: a survey’, *Philosophical Transactions of the Royal Society A*, 379(2194), 20200209.

Loughran, T. and McDonald, B. (2011) ‘When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks’, *The Journal of Finance*, 66(1), pp. 35–65.

Malo, P., Sinha, A., Korhonen, P., Wallenius, J. and Takala, P. (2014) ‘Good Debt or Bad Debt: Detecting Semantic Orientations in Economic Texts’, *Journal of the Association for Information Science and Technology*, 65(4), pp. 782–796.

Rönnqvist, S. and Sarlin, P. (2017) ‘Bank distress in the news: Describing events through deep learning’, *Neurocomputing*, 264, pp. 57–70.

Sanh, V., Debut, L., Chaumond, J. and Wolf, T. (2019) ‘DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter’, *5th Workshop on Energy Efficient Machine Learning and Cognitive Computing, NeurIPS 2019*.

Sezer, O.B., Gudelek, M.U. and Ozbayoglu, A.M. (2020) ‘Financial time series forecasting with deep learning: A systematic literature review: 2005–2019’, *Applied Soft Computing*, 90, 106181.

Sirignano, J. and Cont, R. (2019) ‘Universal features of price formation in financial markets: perspectives from deep learning’, *Quantitative Finance*, 19(9), pp. 1449–1459.

Tetlock, P.C. (2007) ‘Giving Content to Investor Sentiment: The Role of Media in the Stock Market’, *The Journal of Finance*, 62(3), pp. 1139–1168.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I. (2017) ‘Attention Is All You Need’, *Advances in Neural Information Processing Systems (NeurIPS)*, 30.

Xu, W. and Ouyang, Y. (2022) ‘A multimodal deep learning approach for stock price movement prediction’, *Information Processing & Management*, 59(6), 103093.

Yang, S.Y., Jhang, M.K., Mo, Y. and Liu, A. (2020) ‘Using Twitter data and sentiment analysis for financial market prediction’, *Journal of Behavioral and Experimental Finance*, 27, 100352.

Zhang, X., Zhao, J. and LeCun, Y. (2015) ‘Character-level Convolutional Networks for Text Classification’, *Advances in Neural Information Processing Systems (NeurIPS)*, 28.