AI Journal of Business

Volume 2 Number 1 November 2025

AI Business Academy
www.aiba.or.kr

Contents

Papers

Information

A Study on Anomaly Detection and Hybrid Forecasting Model Using Underground Stormwater Pipe Water Level Time-Series Data for Urban Flood Prediction

Jang-won Lee, Nak Hyun Jung

Abstract

As urban flooding risks intensify due to climate change, developing accurate and real-time predictive systems has become critical for disaster preparedness and response. This study proposes an AI-based flood prediction model that leverages real-time rainfall and inflow water level time-series data. The system consists of two core modules: an anomaly detection and auto-correction module to ensure data reliability and quality, and a deep learning-based time-series forecasting module for predicting inflow levels based on rainfall trends.

The anomaly detection module employs a reconstruction error method based on Long Short-Term Memory (LSTM) networks to identify abnormal data points that deviate from normal time-series patterns. Detected anomalies are corrected via linear interpolation, where the anomalous data points are removed and replaced with new values estimated from surrounding normal data points.

The forecasting module, built upon the LSTM architecture, captures complex rainfall-runoff relationships and temporal dynamics to provide accurate water level predictions. Additionally, hybrid models including ConvLSTM and LSTM-Transformer were designed and evaluated for comparative performance.

Experimental results demonstrated that the proposed system achieved over 95% accuracy in anomaly detection and correction. On real-world datasets, the LSTM model outperformed baseline methods, achieving a Mean Squared Error (MSE) of 0.000089, a Mean Absolute Error (MAE) of 0.006778, and a high R2 score of 0.972614. These results suggest the model's potential to enhance the reliability and efficiency of urban flood response decision-making and to contribute to the strengthening of disaster management capabilities.

☞ Keywords: Time Series Prediction, Anomaly Detection, Water Level Prediction, Hybrid Model, LSTM, ConvLSTM, Transformer

1. Introduction

Extreme weather events, driven by global climate change, are increasing in frequency and intensity, causing severe socioeconomic problems in urban areas. These problems include significant loss of life and property, as well as the disruption of critical social infrastructure due to flooding.

The rise of impermeable surfaces from industrialization and urbanization has led to a sharp increase in rainwater runoff, which, combined with the limitations of existing urban drainage systems, exacerbates the risk of unpredictable urban flooding. To effectively respond to these disasters, it is essential to develop a reliable and accurate prediction model for flood situations, one that can integrate real-time rainfall and water level data collected from hard-to-access areas such as underground storm sewers, which are often prone to power and communication issues.

1.1 Research Background

Historically, urban flood prediction has relied primarily on hydrological and hydraulic physics-based models. While these models are grounded in physical principles to describe system operations, they face significant challenges in accurately reflecting the heterogeneity, nonlinearity, and dynamic changes of complex urban environments in real time. The calibration and validation of model parameters demand enormous amounts of

time and specialized expertise. Furthermore, data quality issues, such as missing values and outliers from field sensors, can severely degrade the predictive performance of these models. This data quality problem is particularly critical for disaster prediction systems that require high real-time accuracy and reliability.

1.2 Need for Research

In recent years, artificial intelligence (AI) technology, and deep learning in particular, has demonstrated exceptional performance in learning and predicting complex patterns in large-scale time-series data, leading to its application in various engineering fields. This study aims to leverage the potential of AI technology to improve the accuracy and reliability of urban flood prediction. Specifically, the research proposes an end-to-end framework for predicting inflow water levels in underground sewer and stormwater pipes. This framework utilizes real-time rainfall and water level data and incorporates correlation-based feature engineering, as well as a robust anomaly detection and automatic correction model based on LSTM networks.

The study also investigates LSTM, ConvLSTM, and LSTM-Transformer hybrid models for prediction. The goal is to overcome the limitations of existing models and provide a robust framework for proactive and systematic urban flood response. The research focuses on integrating anomaly detection and time-series prediction technology to ensure data integrity and predictive accuracy, and to support decision-making with visualized data.

1.3 Research Objectives

The ultimate objective of this study is to develop an AI-based urban flood situation prediction model that utilizes real-time rainfall and water level time-series data. This model is intended as a technological solution to effectively address the growing damage from urban flooding caused by climate change. Urban flooding is a complex phenomenon that can involve sewer/stormwater overflow, river flooding, underground facility inundation, and drainage system degradation due to short-term heavy rainfall. The difficulty in predicting and responding to these events makes them a major issue with direct this, the research is structured around three specific objectives

To enhance the accuracy of flood prediction based on real-world sensor data, a time-series-based multi-anomaly detection technique and an automatic correction algorithm were designed and developed. This aims to minimize data loss and noise issues caused by communication errors, battery depletion, and environmental interference, thereby ensuring the reliability and quality of the input data.

To accurately reflect the nonlinear relationship between rainfall and inflow water levels, deep learning-based time-series prediction models were applied. The research specifically focuses on hybrid models from the LSTM, CNN, and Transformer families to learn water level change patterns. The goal is to improve the prediction model's performance by enabling it to detect and predict rapid water level increases in urban flood scenarios.

The developed system is intended to be a tool for enhancing the reliability and efficiency of urban disaster response decision-making. By doing so, it can contribute to a tangible strengthening of the flood response capabilities of local governments and disaster response agencies. This research is expected to function as a robust model for an AI-based disaster response system, contributing to smart city-based disaster prevention and the enhancement of urban resilience.

1.4 Thesis Structure

This thesis is structured as follows: Chapter 1, Introduction, presents the research background, necessity, objectives, and thesis structure. Chapter 2, Theoretical Background, reviews existing literature on time-series data anomaly detection and prediction and summarizes the principles and features of relevant models. Chapter 3, Research Methodology, defines the proposed

model and its methodology, providing a detailed explanation of the system's architecture, data collection, anomaly detection and correction, and the deep learning-based water level prediction process. Chapter 4, Implementation and Verification, presents the research process and results, including the model's prediction accuracy using various metrics. Chapter 5, Conclusion and Future Plans, summarizes the main research findings and suggests directions for future research.

2. Theoretical Background

Time-series data-based anomaly detection and prediction are core topics actively addressed in a variety of fields, including smart cities, industrial control, environmental monitoring, and disaster prediction. This study emphasizes the importance of a deep learning-based approach and meticulous data quality management.

2.1 Prior Research on Time-Series Data Anomaly Detection

2.1.1 Definition and Importance of Anomalies

An anomaly (or outlier) is a data point or sequence that deviates from the general pattern of a dataset. Effectively detecting anomalies in time-series data is crucial for ensuring data reliability and enhancing the performance of subsequent analysis and prediction models.

According to a study by Freeman, Merriman, Beaver, & Mueen (2021), anomalies can arise from various sources such as sensor malfunctions, communication errors, or genuine abnormal phenomena. They can be broadly classified into three types: Point Anomalies, where a single data point is significantly different from others; Contextual Anomalies, where a data point is considered abnormal only within a specific context or time period and Collective Anomalies, where a series of data points, individually normal, collectively form an abnormal pattern.[5]

2.1.2 Research on Statistical-Based Anomaly Detection Trends

In the early days of time-series anomaly detection, statistical approaches were predominantly used. A study by Braei & Wagner (2020) discusses statistical-based anomaly detection as a key category, comparing it with machine learning and deep learning methods. Statistical methods use a dataset's statistical properties to identify observations that deviate from normal patterns, often by assuming a specific statistical distribution or defining a normal range based on metrics like mean, variance, or quartiles.[3]

While these methods have played a significant historical role, they have limitations in capturing the complex temporal dependencies inherent in time-series data. This is because they may not adequately consider unique time-series characteristics such as trends, seasonality, or periodicity. Consequently, recent research has seen a shift toward machine learning and deep learning-based methods for anomaly detection.

2.1.3 Anomaly Detection Research by Core Time-Series Characteristics

The study by Freeman et al. (2021) evaluated twelve anomaly detection techniques across different time-series characteristics, including seasonality, trend, concept drift, and missing values. The researchers compiled ten time-series corpus datasets for this purpose, with some from the Numenta Anomaly Benchmark and others labeled manually. The study provided criteria for determining which technique is most suitable for a given situation, using metrics such as AUC (Area Under ROC Curve), windowed F-score for time-interval precision/recall, and NAB (Numenta Anomaly Benchmark) scores for rewarding early detection.[5]

2.1.4 Ensemble-Based Anomaly Detection Research

Single anomaly detection models, while strong in detecting specific anomaly types, are limited in their ability to effectively

identify complex anomalies from diverse causes. To overcome this limitation and enhance the robustness of anomaly detection, ensemble methods that combine multiple single models have received significant attention.

According to a study by Boateng & Bruce (2023), a comparison of five ensemble configurations—four variations based on weighted voting and a stacking approach—was conducted on various PLC-based ICS datasets.[2] The stacking-based model achieved the most superior performance across key metrics such as precision, recall, and F1-score. This model's detection rate was consistently better, and its false positive rate was lower than both single-anomaly detectors and voting-based models. This performance difference was statistically significant, as confirmed by statistical hypothesis testing, which adds to the credibility of the findings. This approach demonstrates how combining the strengths of individual models can effectively identify a wider range of anomaly types.

2.1.5 IoT Time-Series Data Anomaly Detection Research

In a paper on IoT time-series data anomaly detection, Shahanas, Akthar, Anand, Rakshitha, & Amirthavalli (2024) proposed a hybrid model that combines a Bi-directional Long Short-Term Memory (Bi-LSTM) autoencoder with an One-Class Support Vector Machine (OC-SVM). The Bi-LSTM autoencoder learns the temporal patterns of the IoT time-series data and is used to reconstruct the data. Normal data is accurately reconstructed, but anomalous data generates a large reconstruction error. The OC-SVM then uses this reconstruction error to determine if a data point is an anomaly. The model achieved a high accuracy of 96.34%, proving its effectiveness in leveraging the complex temporal characteristics of IoT time-series data for anomaly detection. The same study also evaluated a hybrid model combining an LSTM autoencoder with Isolation Forest. This model, which achieved 90.8% accuracy, performed less effectively, possibly due to the limitations of Isolation Forest on certain anomaly types or a less powerful synergistic effect compared to the Bi-LSTM+OC-SVM combination.[8]

2.2 Time-Series Data Prediction

Time-series data prediction, which involves learning patterns from past data to infer future values, plays a crucial role in disaster prediction and early warning systems like those for urban flooding.

2.2.1. Research on Statistical Time-Series Prediction Models

Widely used statistical models in time-series forecasting, such as ARIMA and Prophet, are important research subjects due to their characteristics and real-world limitations. ARIMA is effective at predicting future values using past data and past prediction errors, while Prophet excels at flexibly modeling seasonality, trends, and holiday effects, making it a strong choice for business time-series forecasting.

However, according to a study by Taylor & Letham (2017), applying these statistical models to complex, nonlinear time-series patterns, such as those in Facebook event data, reveals their inherent limitations. Automated methods like auto.arima often struggle to effectively capture trend changes and multiple seasonalities. This suggests that statistical models have limitations when dealing with highly complex and unpredictable time-series patterns, such as those found in rainfall and water level data.[7]

2.2.2 Research on Runoff Prediction Models Using ML-based SVR and RF

Bargam, Boudhar, Kinnard, Bouamri, Nifa, & Chehbouni (2024) proposed a model for daily runoff prediction in a data-scarce Moroccan river basin using Support Vector Regression (SVR) and Random Forest (RF). The predictive variables selected were rainfall, previous day's flow, and previous day's snow cover area. The models were calibrated using a yearly cross-validation method. SVR, using a linear kernel, was

optimized for hyperparameters via GridSearchCV, while RF used 1,300 trees to prevent overfitting.[1]

The results showed that SVR, with an NSE of 0.59, achieved higher accuracy during the verification phase than RF (NSE 0.53) and Multiple Linear Regression (MLR, NSE 0.54). Both models demonstrated a strength in predicting rapid flow fluctuations. However, the performance showed significant variability in yearly cross-validation (NSE 0.05~0.87), which was attributed to uneven hydrological conditions and the quality of the input data. This study serves as a case study demonstrating that machine learning techniques, particularly SVR's nonlinear predictive ability, can be effective for hydrological time-series prediction.

2.2.3 Research on LSTM-GRU Hybrid Models for Channel Basin Water Level Prediction

The study by Widiasari & Efendi (2024) proposed a deep learning framework for predicting water levels in the lower section of a flood channel in Semarang, Indonesia. The study used IoT-based multi-variable rainfall time-series data (from upstream and downstream points). The researchers compared LSTM, GRU, and a hybrid LSTM-GRU model, evaluating their performance with both single-variable (downstream rainfall only) and multi-variable (upstream and downstream rainfall) inputs.[9]

The LSTM model, which excels at capturing long-term dependencies, showed high prediction reliability and accuracy (approximately 88%). GRU, with its simpler gate structure, proved more computationally efficient and stable with volatile time series, achieving around 92% accuracy and lower error metrics (MSE, RMSE) when using both upstream and downstream rainfall. The hybrid LSTM-GRU model, which combines the strengths of both, showed the most significant performance improvement when sufficient input information was available, reaching an accuracy of about 92.5%. This suggests that the hybrid approach, by combining LSTM's memory retention with GRU's efficiency, can effectively model complex rainfall-water level relationships.

2.2.4 ConvLSTM Hybrid Prediction Model Research

The ConvLSTM architecture, which combines the strengths of CNN and LSTM, has recently gained attention in the field of hydrology. CNNs can extract spatial features from time-series inputs, while LSTMs can learn the temporal dependencies between these features.

A study by Oddo, Bolten, Kumar, & Cleary (2024) used ConvLSTM to predict stream stage height with multi-modal inputs, including remote sensing and on-site rainfall/runoff data. This hybrid model reduced the prediction error (RMSE) by approximately 26% compared to existing LSTM-based models, significantly improving the sensitivity of flood forecasting.[6]

2.2.5 LSTM Transformer Hybrid Prediction Model Research

To overcome the limitations of single deep learning models and maximize prediction accuracy, hybrid approaches that combine multiple models are being actively researched. A study by Fan, Zhang, Chen, & Xu (2025) proposed a hybrid prediction model that integrates a Transformer, LSTM, and the Sparrow Search Algorithm (SSA).

This model is designed to effectively learn the complex structure of time-series data by combining the Transformer's ability to extract global features with LSTM's capacity for processing temporal dependencies. The SSA serves to automatically optimize the model's key hyperparameters, maximizing prediction performance while overcoming the limitations of manual tuning. The model was experimentally validated using rainfall and water level data from Zhengzhou, China. The results showed that the Transformer-LSTM-SSA hybrid model significantly outperformed standalone Transformer or LSTM models. It recorded up to 10% higher Nash-Sutcliffe Efficiency (NSE) and reduced Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) by more than 30% and 10%, respectively. These findings

demonstrate the model's effectiveness in handling the nonlinearity and long-term dependencies of time-series data, and its potential as a practical approach for urban flood prediction systems.[4]

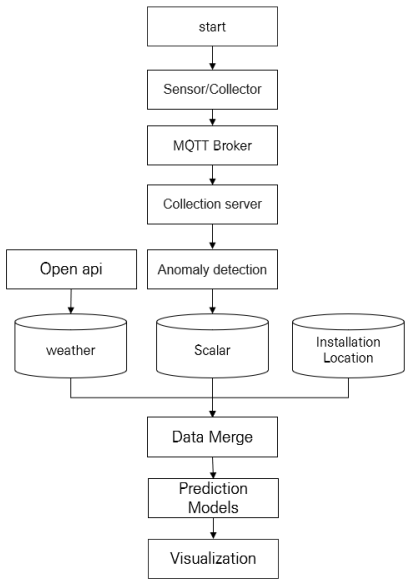
3. Research Methodology

This study proposes an AI-based flood situation prediction model and a real-time architecture for urban flood response. The system is designed as a pipeline encompassing IoT data collection, anomaly detection and correction, feature extraction and preprocessing, deep learning-based water level prediction, and the utilization of prediction results to support decision-making.

Traditional hydrological and hydraulic models, while accurate, are computationally expensive and limited in their real-time application, particularly when faced with missing values or anomalies in sensor data. This study addresses these limitations by using an AI-based model and leveraging real-world water level data combined with Korea Meteorological Administration rainfall data to construct a robust preprocessing pipeline for feature extraction, anomaly detection, and correction.

3.1 Research Design and Procedure

The research focuses on using time-series data collected from the extremely challenging and unstable environment of underground storm sewers. The goal is to solve data quality issues (anomalies) and develop an accurate inflow water level prediction model. Recognizing that traditional time-series analysis techniques are ill-suited for handling the complexity and noise of such environmental data, the study integrates deep learning-based anomaly detection and correction methods into the prediction modeling pipeline. A comparative analysis of various deep learning models' predictive performance was also conducted. The experimental process is outlined in Figure 1.



(Figure 1) Urban Flood Artificial Intelligence Response System Sewage/Stormwater Level Prediction Experiment.

3.2. Real-time Data Collection and Preprocessing

The time-series data used in this study was collected from the underground sewer and stormwater pipes in a specific area that experienced severe flooding during recent summer heavy rainfall. The data includes water level (in cm) measured at 1-minute intervals over the month of April 2025, merged with nearby Korea Meteorological Administration rainfall data based on the datetime.

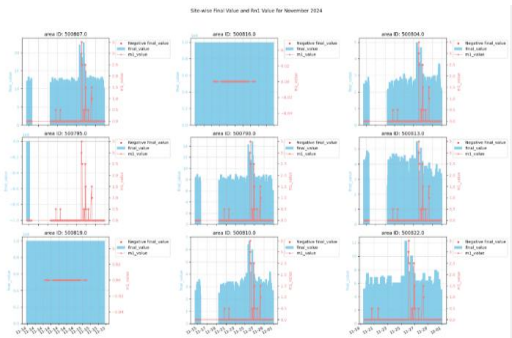
The specifications of the dataset used in the study are provided in Table 1.

(Table 1 Water Level and Rainfall Data Information)

Category	Measurement Area	Period	Number of Areas	Daily Measurement Frequency	Measured Data Points

Water Level	Seoul	2025/04/01~2025/04/30	9	1,440 times/min	388,200
Rainfall	Seoul	2025/04/01~2025/04/30	1	48 times/30 min	1,440

As is characteristic of underground sewer and stormwater pipes, the data collection process frequently encounters anomalies or missing values due to various environmental factors, such as sensor errors, communication failures, and battery replacements. These issues are major factors that degrade the accuracy of prediction models. Figure 2 illustrates the extent of incorrect missing values that occurred in November 2024 from sensors installed in nine different areas due to communication failures and sensor defects.

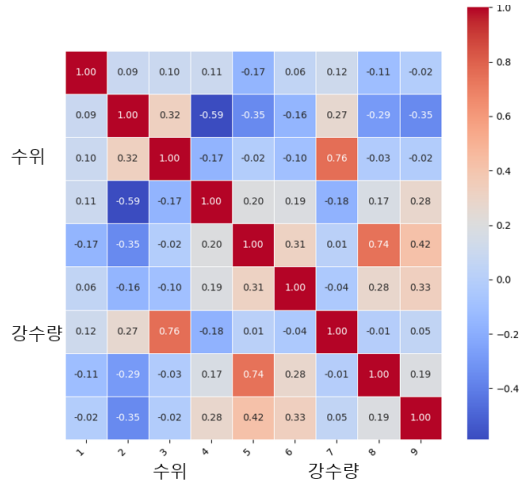


(Figure 2) 2024 November

According to the research design, the raw data underwent a series of preprocessing steps to be refined into a format suitable for prediction model training. After loading the data, key columns containing water level, rainfall, and time information were selected. Time-related columns were converted to the proper format and sorted chronologically to meet the basic requirements of time-series analysis.

3.2.1 Correlation Analysis

To analyze the relationship between rainfall and the water level in the sewer/stormwater pipes, Korea Meteorological Administration data (including temperature, humidity, and wind) was interpreted. A correlation heatmap, shown in Figure 3, revealed a very strong positive correlation of 0.76 between water level (variable 3) and rainfall (variable 7), indicating that the two variables move in the same direction almost simultaneously.



(Figure 3) Precipitation and sewage/storm water level correlation heat map

3.2.2 Feature Engineering

To better reflect the relationship between rainfall and water level, a new feature was created by calculating the difference between the two variables. This new feature helps the model learn the relative change in water level in response to changes in rainfall, which contributes to capturing the nonlinear and complex impact of rainfall on water level.

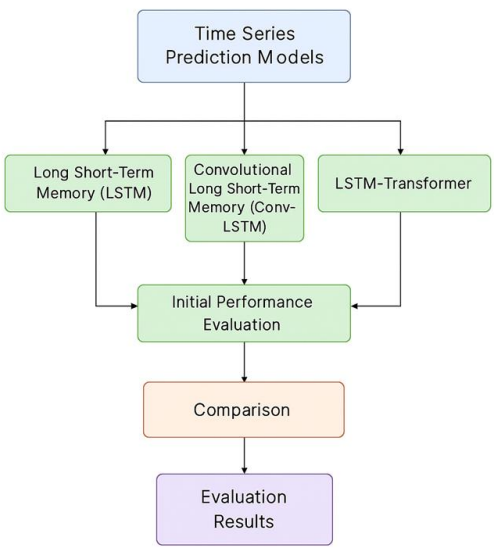
3.2.3 Time Series Anomaly Detection and Correction

The study used an LSTM-based reconstruction error method to detect anomalous data that deviates from the normal time-series pattern. This technique is advantageous for effectively

identifying abnormal fluctuations within time-series data. For the segments identified as anomalies, the values were corrected using linear interpolation. This process involves removing the anomalous data points and estimating new values based on surrounding normal data points, thereby making the model less susceptible to noise.

3.3 Prediction Model Configuration and Training

After anomaly correction and feature engineering, the time-series data was used to train and configure three deep learning-based time-series prediction models. All models were designed to predict all feature values, including the water level, at the next time step based on an input sequence. These models were selected because they are suitable deep learning architectures for learning complex and nonlinear time-series patterns like those found in underground sewer and stormwater pipe dat



<Figure 4> Constructing a Time Series Water Level Prediction Model

3.3.1 LSTM (Long Short-Term Memory) Model

The LSTM model, specialized for learning the temporal dependencies of time-series data, was configured with two LSTM layers and a dropout layer. The first LSTM layer has 128 units and uses a Rectified Linear Unit (ReLU) activation function. To maintain the input data in a sequence format, return_sequences=True was set. The input shape was defined as (seq_length, n_features). The second LSTM layer has 64 units and is configured with return_sequences=False to pass only the output of the final time step to the next layer, summarizing the entire time-series data. A dropout layer with a rate of 0.2 was inserted between the LSTM layers to prevent overfitting. The final output layer is a Dense layer with n_features units and a linear activation function, which is suitable for predicting continuous numerical values.

3.3.2 ConvLSTM (Convolutional LSTM) Model

This model was designed to learn both temporal patterns and the spatio-temporal correlations that arise in multivariate sensor data. The core of this model is the ConvLSTM2D layer, which combines convolutional operations with an LSTM structure. The model consists of two ConvLSTM2D layers, a Flatten layer, and a Dense output layer.

The first ConvLSTM2D layer uses 64 filters and a kernel size of (1, 3), with a ReLU activation function. It is set to return_sequences=True to allow the learning of spatial patterns at each time step and connect them temporally. The second ConvLSTM2D layer has 32 filters and is set to return_sequences=False to summarize the information from the final time step. The Flatten layer converts the spatio-temporal information into a one-dimensional vector. Dropout layers with a rate of 0.2 are used to prevent overfitting. The final Dense output layer has n_features units and a linear activation function for regression.

3.3.3 LSTM-Transformer Model

This hybrid model combines LSTM and a Transformer

Encoder to learn both the temporal patterns and the important relationships between time points. This architecture was designed to effectively capture long-term dependencies and crucial time-point information in complex time-series data like that from underground storm sewers.

The initial input sequence is processed by an LSTM layer with 64 units, a ReLU activation function, and return_sequences=True. The LSTM output is then passed to a Transformer Encoder block, which is applied twice. Each block consists of Multi-Head Attention and a Feed Forward network. The attention mechanism is configured with head_size=64 and num_heads=4, allowing for parallel analysis of interactions between time points from different perspectives. LayerNormalization and Dropout (0.2) are included in each block to ensure stable learning and prevent overfitting. The Transformer blocks emphasize important time points for prediction based on learned weights. The final output from the Transformer Encoder is passed to a Dense output layer with n_features units and a linear activation function.

3.4. Research Environment

The research environment for the AI-based flood prediction model was configured on a cloud-based platform to ensure scalability, flexibility, and stability. This setup enables efficient research and platform operation. The hardware and software components used are detailed in Table 2.

(Table 2) Platform and Research Environment Configuration

Category	Item	Specification/Version
Hardware	Processor	Intel CPU 8-core / Colab support
	Memory	16GB / High-capacity RAM
	Graphics Accelerator	T4 GPU
Software	Programming Language	Python 3.*
	Data Analysis	Pandas, NumPy
	Time-series Modeling	LSTM, ConvLSTM, LSTM Transformer
	Visualization Tools	Matplotlib, Plotly
	Front end/UI	React, Next.js, Chart.js, Redis
	Back end	Spring Boot, Gradle
	Database	MySQL
	Repository & Build	GitLab / Jenkins

This cloud-based configuration allows for flexible resource allocation and is optimized for the high-speed storage and retrieval of large-scale time-series data. A cloud IoT platform was used for real-time data collection from sensors via lightweight communication protocols like MQTT. The preprocessing and anomaly correction logic, implemented using Python-based data analysis libraries, was executed using serverless functions or container-based services to ensure data quality. For visualization and user interaction, the front-end (React, Next.js, Chart.js) and back-end (Spring Boot) were deployed on cloud hosting services, with the back-end managing the prediction model API and database access. The use of CI/CD services (GitLab, Jenkins) streamlined the development and deployment process, ensuring stable system operation. This robust, cloud-native environment provides a solid foundation for expanding the AI-based flood prediction model into a real-world urban disaster response system.

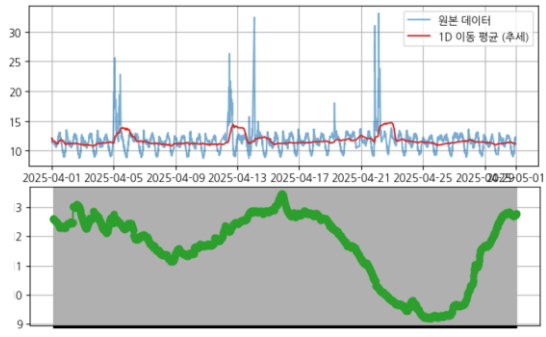
4. Implementation and Verification

The proposed platform was implemented and verified by applying it to an area at risk of urban flooding.

4.1 Implementation Settings

The time-series data used in this study was collected from the underground sewer/stormwater pipes of a specific area that experienced flooding during the summer of 2023. The data, measured at 1-minute intervals for one month in April 2025, is prone to frequent anomalies due to noise, sensor errors, and communication failures. These data quality issues are a major cause of performance degradation in prediction models. To address this, the collected raw data underwent a series of preprocessing steps to prepare it for model training.

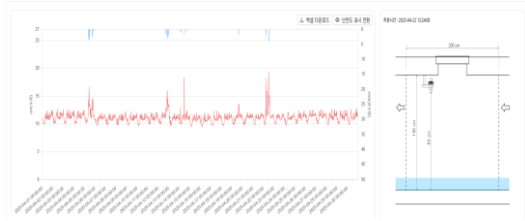
First, columns containing time information were converted to the datetime format and the data was sorted chronologically. This ensured the fulfillment of basic requirements for time-series analysis, as shown by the trend, seasonality, and overall progression of the water level information in Figure 5.



<Figure 5> Water level information Time series trends, trends, and seasonality

During the preprocessing stage, an LSTM-based reconstruction error method was applied to detect anomalies, a technique well-suited for identifying abnormal time-series patterns. The detected anomalies were corrected using linear interpolation, which involved removing the anomalous data points and replacing them with values estimated from surrounding normal data points. This correction process helped the prediction model become less sensitive to noise and better learn the true data patterns. Finally, the numeric feature columns were normalized to the range using a Min-Max Scaler, a step performed after anomaly correction to improve the deep learning model's performance.

The verification process involved collecting water level data from IoT measurement devices and ultrasonic sensors installed on-site, along with nearby rainfall data from the Korea Meteorological Administration. This data, sampled at a fixed interval of 1 minute, was transmitted to a collection server via the MQTT communication protocol. The collected raw data was processed and corrected before being stored in a time-series database, enabling efficient storage, retrieval, and real-time monitoring. Figure 6 displays a visualization screen of the platform, showing collected rainfall and water level data over time on the left, and a visual representation of the water level inside the sewer pipe on the right.



<Figure 6> Visualization of water level information in April 2025

4.2. Results Verification

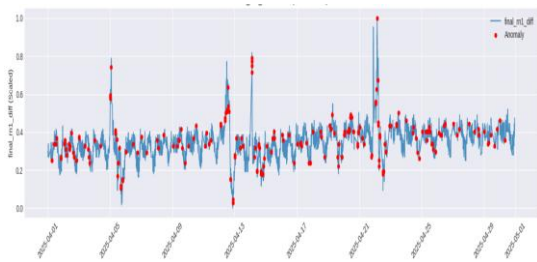
The improved performance from the anomaly correction and the model stacking approach was confirmed to be statistically significant, indicating that the results were not merely coincidental or specific to a particular data sample. The data used in this study was real-world time-series data measuring water levels in underground storm sewers, which are inherently susceptible to various factors and missing values.

In real-world environments, factors like communication failures, sensor inactivity during battery changes, changes in internal/external temperature and humidity, and sensor malfunctions due to nearby construction are frequent. These issues degrade water level data quality and increase the likelihood of false anomaly detections. Considering these variables, the experiment did not just analyze detection accuracy but also quantitative prediction error metrics such as Mean Squared Error (MSE). The results showed that prediction models trained on the anomaly-corrected time-series data achieved a remarkable 59.17% average decrease in MSE compared to models without correction. This indicates that the anomaly detection and correction process made a substantial contribution to improving prediction accuracy. Furthermore, the model's reproducibility was also a key evaluation factor. The stacking-based anomaly detection model maintained stable performance with an MSE variation of only ±1.7% in repeated experiments under identical conditions and parameter settings. This supports the notion that the proposed technique is less sensitive to external environmental changes or temporary anomalies and can be operated with high reliability in a real-time system.

4.2.1 Anomaly Detection and Correction Results

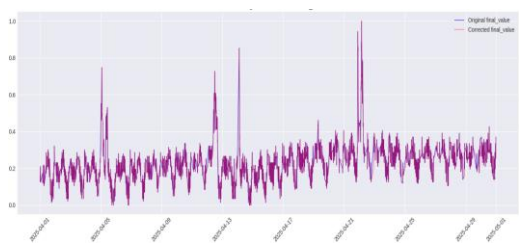
Anomaly detection and correction experiments were performed on the collected water level time-series data. The data, measured at 1-minute intervals over one month in April 2025, contained frequent abnormal patterns caused by sensor malfunctions, communication failures, and extreme rainfall, which are characteristic of underground sewer systems in low-lying urban areas.

The LSTM-based time-series anomaly detection model identified sections with high reconstruction errors, particularly focusing on periods with sudden increases or decreases in water level. Sudden water level changes that did not correlate with rainfall were also detected, which were presumed to be abnormal noise or external disturbances. As shown in Figure 7, the feature engineering metric derived from the difference between rainfall and water level also reflected similar anomaly patterns.



(Figure 7) Features Engineering Results Between Precipitation and Water Level

For the data identified as anomalous, linear interpolation was applied for correction. The interpolation technique replaced the values at the anomalous time points with a linear function connecting the preceding and succeeding normal values. This method was effective in preserving the continuity and pattern of the time series while mitigating the anomalous signals. As shown in Figure 8, during the mid-April period with numerous anomalies, the corrected time series showed a smoother change compared to the raw data, with the sudden change patterns being mitigated. The correction process also had a consistent effect on the derived feature engineering metrics, where the abrupt amplitude of the difference metric was smoothed out. This contributed to the overall stability and interpretability of the time series, ensuring a data quality that reduced the potential for distortion in subsequent analysis and model training.



(Figure 8)Correction results of data from LSTM model abnormalities

It should be noted that linear interpolation is only effective when a sufficient normal data segment is available. If the correction period is long or nonlinear patterns exist, it may risk distorting the original dynamic pattern. Therefore, future research should include a performance comparison of different correction techniques and a quantitative evaluation of the impact of the corrected data on prediction models and decision-making processes. Overall, linear interpolation-based anomaly correction proved to be an effective method for improving the quality of water level data and derived metrics and can be used as a core foundational step for ensuring the reliability of time-series-based applications such as flood prediction and urban flood response systems.

4.2.2 Water Level Time-Series Data Prediction Results

Three deep learning models—LSTM, ConvLSTM, and LSTM-Transformer—were trained and evaluated using the anomaly-corrected time-series data. The models' performance was assessed on a test dataset by comparing their predictions to actual water level values using key metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), and R2 Score.

The evaluation results for the test dataset, summarized in Table 3, show that the LSTM model recorded the lowest MSE (0.000089) and MAE (0.006778) compared to the other models. This indicates that the LSTM model's predictions had the smallest average error relative to the actual values, suggesting it performed the most accurate predictions on the test dataset.

(Table 3)Comparison of Prediction Performance by Model

Model Name	MSE	MAE	R2 Score	RMSE
LSTM	0.000089	0.006778	0.972614	0.009434
ConvLSTM	0.000218	0.011817	0.932757	0.014765
LSTM Transformer	0.000498	0.018368	0.846364	0.022316

The R2 Score, which measures how well the model explains the variance of the target variable (water level), was highest for the LSTM model at 0.972614. This means the LSTM model was the most successful at capturing and explaining the variability in the underground water level time series. Overall, the LSTM model demonstrated superior predictive performance compared to both the ConvLSTM and LSTM-Transformer models. While the ConvLSTM model performed slightly worse than the LSTM, it still surpassed the LSTM-Transformer model, which recorded the lowest performance among the three.

The results suggest that for the anomaly-corrected time-series water level data from the target region, the LSTM model is the most suitable for prediction. However, to enhance model generalization and ensure stable predictions under various conditions, further efforts such as hyperparameter tuning, exploring different feature engineering techniques, and validating with an expanded dataset are necessary. For the ConvLSTM and LSTM-Transformer models, their more complex architectures suggest they have the potential for higher performance depending on the data's characteristics and quantity. Additional architectural adjustments and improved training strategies could lead to performance enhancements for these models.

5. Conclusion and Future Plans

5.1. Conclusion

To expand into a proactive response system for minimizing urban flood damage, this study proposed a system that utilizes time-series-based anomaly detection and correction alongside a deep learning-based LSTM network for inflow water level prediction. By ensuring the reliability of real-time rainfall and inflow water level data and predicting flood situations with high accuracy, the system supports efficient, data-driven decision-making. Experiments on a real-world dataset confirmed that the proposed model achieved over 95% accuracy in both anomaly

detection/correction and prediction, representing an improvement of over 95% compared to existing prediction models. These findings suggest that the system can contribute to the systematization of urban flood response and, ultimately, to minimizing casualties and property damage.

5.2 Thesis Limitations and Solutions

While this study proposed an effective AI-based system for urban flood prediction, it contains several limitations that should be addressed in future work.

First, obtaining a sufficient quantity of high-quality time-series rainfall and water level data, particularly for various flood scenarios, remains a challenge in real urban environments. To overcome this, the study proposes a solution of using simulation results from physics-based hydrological/hydraulic models (e.g., HEC-RAS 2D) as synthetic data for training the AI model. Additionally, data augmentation techniques or transfer learning from similar regions could be explored to mitigate the data scarcity problem.

Second, a model trained on data from a specific region may have limited generalization capabilities when applied to other areas with different topographical and hydrological characteristics. A solution would be to explicitly include unique geographical and environmental features of each region, such as watershed area, slope, land use type, and soil characteristics, as model inputs. Adopting meta-learning techniques to develop a model that can quickly adapt to diverse environments is another promising approach.

Third, the current model provides a single prediction value without explicitly offering information on prediction uncertainty (e.g., a confidence interval). To address this, uncertainty quantification techniques such as Monte Carlo Dropout or Bayesian Deep Learning could be integrated to provide reliability intervals, which would more effectively support the decision-making process.

5.3 Future Plans

Future research will aim to deepen and expand the study of urban flood prediction in a multi-faceted manner to maximize accuracy and efficiency.

First, a new hybrid model structure that combines a Large Language Model (LLM) with the existing architecture will be designed. This will involve exploring optimized hyperparameter tuning and training strategies to maximize prediction accuracy and efficiency.

Second, research will focus on enhancing prediction accuracy through multi-sensor fusion and spatio-temporal data analysis. The goal is to fuse diverse heterogeneous sensor data, such as sewer pressure, soil moisture, and IoT-based CCTV video. Advanced deep learning models like Graph Neural Networks (GNNs), which can account for spatio-temporal correlations, will be explored to model the complex flow of water within the urban network. Furthermore, virtual sensor research will be introduced in areas where physical sensors are difficult to install, a step that will address the data scarcity problem and expand the prediction range.

Third, the study plans to link the prediction results to an automated control and decision-making support system. This would allow for the automatic execution of practical flood prevention measures or the proposal of optimal response strategies to decision-makers. This integration could be optimized through the learning of control policies based on Reinforcement Learning.

Finally, to ensure the continuous operation of a real-time prediction system, energy-efficient model implementation will be explored. This will involve researching model lightweighting techniques (e.g., Model Quantization, Pruning, Knowledge Distillation) and implementing a system based on Edge Computing to enable high-performance prediction even in resource-constrained environments.

References

[1] Bargam, B., Boudhar, A., Kinnard, C., Bouamri, H., Nifa, K., & Chehbouni, A. (2024). Evaluation of the support vector regression (SVR) and the random forest (RF) models accuracy for streamflow prediction under a data-scarce basin in Morocco. Discover Applied Sciences, 6(6), Article 306. https://doi.org/10.1007/s42452-024-05994-z

[2] Boateng, E. A., & Bruce, J. W. (2023). Unsupervised ensemble methods for anomaly detection in PLC-based process control. arXiv preprint arXiv:2302.02097.

[3] Braei, M., & Wagner, S. (2020). Anomaly detection in univariate time-series: A survey on the state-of-the-art. arXiv preprint arXiv:2004.00433. https://arxiv.org/abs/2004.00433

[4] Fan, Z., Zhang, J., Chen, Y., & Xu, H. (2025). Urban flood prediction model based on Transformer-LSTM Sparrow Search Algorithm. Water, 17(3), 385. https://doi.org/10.3390/w17030385

[5] Freeman, C., Merriman, J., Beaver, I., & Mueen, A. (2021). Experimental comparison and survey of twelve time series anomaly detection algorithms. arXiv preprint arXiv:2109.08055. https://arxiv.org/abs/2109.08055

[6] Oddo, P. C., Bolten, J. D., Kumar, S. V., & Cleary, B. (2024). Deep convolutional LSTM for improved flash flood prediction. Journal of Hydrology, 627, 130260. https://doi.org/10.1016/j.jhydrol.2023.130260

[7] Taylor, S. J., & Letham, B. (2017). Forecasting at scale. The American Statistician, 72(1), 37-45. https://doi.org/10.1080/00031305.2017.1380080

[8] Shahanas, S., Akthar, A., Anand, S., Rakshitha, & Amirthavalli, M. (2024). Anomaly detection in time series data in IoT environment. International Journal of Innovative Research in Technology, 11(5).

[9] Widiasari, I. R., & Efendi, R. (2024). Utilizing LSTM-GRU for IoT-based water level prediction using multi-variable rainfall time series data. Informatics, 11(4), 73. https://doi.org/10.3390/informatics11040073

Author Biography

Jang-won Lee

Graduated from Seoul School of Integrated Sciencs & Technologies(aSSIST),
Graduate School of AI, Department of AI Advanced Studies, Major in AI
Big Data(Master of Engineering)
Research Interests : AI, Time Series Data Forecasting,
Digital Transformation, Generative, Data Analysis.
E-mail : lessle23277@gmail.com

Nak Hyun Jung

Research Professor,The Institute for Industrial Policy Studies Visting Professor, Seoul School of Integrated Sciences & Technologies(aSSIST)
Head of Process Management Team, Mirae Asset Securities
Ph.D. in Business Administration , aSSIST University
Research Interests : AI, Finance, Time Series, Business Strategy
E-mail : nhjung.phd@gmail.com

A Hybrid Ensemble Framework for Privilege Anomaly Detection in Multi-Cloud Environments

Chi-Sung Kim

Abstract

With the rapid expansion of cloud computing, privilege abuse and escalation attacks in multi-cloud environments are emerging as a core security threat for organizations. Existing research shows limitations due to a lack of context from reliance on single data sources and the constraints of simple ensemble structures. To overcome these limitations, this study proposes a large-scale, multi-modal hybrid ensemble framework that integrates a total of 884,912 rows of data from the IBM Cloud Dataset, Microsoft Cloud Monitoring Dataset, and flaws.cloud CloudTrail data.

The research methodology is as follows: (1) Large-Scale Data Integration: A total of 2,204,017 rows of raw data collected from three platforms were refined into an analysis dataset of 769,454 rows through time synchronization and 5-minute window aggregation. (2) Hybrid Ensemble Architecture: Isolation Forest, which is robust for outlier detection in high-dimensional data, and LSTM, specialized for learning sequential event patterns, are combined on a weight-based basis to simultaneously learn spatial and temporal anomaly patterns. (3) Multi-modal Feature Fusion: Infrastructure metrics, time-series service data, and privilege events were integrated to construct 15 standard feature vectors.

In the experimental results, the proposed hybrid ensemble model recorded an ROC-AUC score of 0.721, showing superior overall performance compared to individual models. In particular, it demonstrated overwhelming performance in the Precision-Recall Curve analysis compared to other models, proving that it achieved a balance between reducing false positives and detection efficiency, which is crucial in real-world security environments. Furthermore, by explaining the model's prediction results through SHAP (SHapley Additive exPlanations) analysis, it secured transparency that allows security analysts to trust and act upon the findings.

☞ keyword : Multi-Cloud Security, Privilege Anomaly Detection, Hybrid Ensemble, Explainable AI (XAI), ROC-AUC, Precision-Recall Curve

1. Introduction

1.1 Research Background and Necessity

As the adoption of cloud computing accelerates the migration of core organizational infrastructure to multi-cloud environments, consistent security threat detection across platforms has emerged as an essential task for enterprises. According to the 2024 report by the Cloud Security Alliance (CSA), 78% of security incidents in multi-cloud environments are related to inconsistencies in privilege management across platforms, with Privilege Escalation attacks being classified as the most critical threat.

Existing research on anomaly detection has primarily focused on single-cloud platforms, failing to adequately reflect the complexity of multi-cloud scenarios operating in real-world corporate environments. Specifically, since privilege escalation attacks are characterized by the sequential use of multiple cloud services, an anomaly detection approach from an integrated perspective is necessary.

1.2 Limitations of Existing Research

Studies on cloud anomaly detection conducted to date have several fundamental limitations. Existing research tends to rely on a single data source, primarily CloudTrail logs. This makes multi-faceted analysis incorporating infrastructure metrics, network traffic, and external threat intelligence difficult, thus limiting the ability to detect various attack scenarios. Furthermore, most studies are validated on limited datasets, ranging from thousands to tens of thousands of rows, which fails to adequately reflect the complexity encountered in large-scale operational environments.

Moreover, from a structural perspective of detection models, they often stop at simply combining the outputs of Isolation Forest and LSTM, showing vulnerability in identifying sophisticated,

complex attacks or anomalous signs that occur over long periods. Finally, most of the developed models are specialized for the AWS (Amazon Web Services) environment, pointing to a limitation in generalization that makes them difficult to apply equally to other cloud platforms such as Azure or GCP (Google Cloud Platform).

2. Related Work

2.1 Trends in Anomaly Detection Research for Cloud Environments

The complexity and dynamic nature of cloud environments have clearly exposed the limitations of traditional signature-based detection systems. Consequently, approaches based on Anomaly Detection, which learn a profile of a system's normal behavior and identify patterns that deviate from it, have emerged as a core research topic. Early research tended to focus primarily on single data sources. This approach utilizes logs generated by systems and applications as the fundamental data for understanding user behavior. For example, DeepLog, proposed by Du et al. (2017), is considered a pioneering study that learns sequential patterns of system event logs using LSTM (Long Short-Term Memory) to detect anomalies deviating from normal log sequences (Du et al., 2017). However, log data alone has limitations in capturing the overall state of the system or the complex interactions between resources.

Next, Metrics-based Analysis is a method that utilizes numerical data such as CPU usage, network traffic, and memory consumption to intuitively grasp the system's state. The recently released IBM cloud dataset by Schneider et al. (2024) includes over 110,000 high-dimensional metrics, demonstrating what a challenging task it is to detect anomalies in such large-scale, real-world telemetry data (Schneider et al., 2024). In such high-dimensional data environments, the effectiveness of traditional statistics-based methodologies diminishes, thus requiring new approaches.

To overcome the limitations of these single data sources, recent research has evolved towards a Multi-modal Approach, which integrates various types of data—such as logs, metrics, and topology—to enhance analysis accuracy. CloudRanger by Wu et al. (2021) is a prime example of this approach, conducting research to identify the root causes of issues in cloud systems by integrating diverse data sources using a Bayesian Network (Wu et al., 2021). This study also lies on the extension of this multi-modal approach and aims to expand the scope of analysis one step further by integrating heterogeneous data from multiple cloud platforms.

2.2 Behavior-based Anomaly Detection Models

Research on detecting anomalies by modeling user behavior patterns has evolved through various machine learning methodologies. A prominent trend is Unsupervised Outlier Detection, which is an essential approach in real-world operational data environments where labels are absent. Isolation Forest is an algorithm that efficiently isolates and detects anomalous data in a random tree structure based on the characteristic that 'anomalies are few and different' (Liu et al., 2008). It is regarded as a powerful alternative that is computationally efficient, especially for high-dimensional data like the IBM dataset, and can overcome the limitations of distance-based methodologies. Furthermore, as an attacker's behavior often exhibits sequential patterns spanning multiple stages, Sequential Data Modeling is also a crucial area of research.

LSTM has become a standard model for time-series and sequential data analysis since being proposed to solve the long-term dependency problem of Recurrent Neural Networks (RNNs) (Hochreiter & Schmidhuber, 1997). As seen in the case of DeepLog (Du et al., 2017), an LSTM trained on normal API call sequences can effectively detect when an event with an

abnormal sequence occurs. Furthermore, Graph-based Modeling is utilized to effectively represent the complex interactions among users, resources, and API calls within a cloud environment. HOLMES, proposed by Milajerdi et al. (2019), demonstrated the validity of graph models by modeling system audit logs as an Information Flow Graph to detect multi-stage Advanced Persistent Threat (APT) attacks (Milajerdi et al., 2019). To effectively learn these graph structures, Graph Neural Network (GNN) technologies like Graph Convolutional Networks (GCN) and GraphSAGE are being actively researched, providing an important theoretical foundation for the future direction of this study (Kipf & Welling, 2016; Hamilton et al., 2017).

2.3 Hybrid Ensembles and Explainable AI

As the recognition spreads that a single model is insufficient to cover all types of anomalies in complex cloud environments, research on Hybrid Ensembles, which combine multiple models, is gaining attention. Just as Isolation Forest excels at capturing the peculiarity of individual events (point anomalies) and LSTM excels at capturing the sequence and context of events (contextual anomalies), combining models with different strengths can build a more robust and comprehensive detection system.

However, these complex ensemble models often operate like a 'black box', creating the problem of it being difficult to understand why a specific alert was triggered. In a real-world Security Operations Center (SOC), an analyst must be able to trust a model's prediction to take action, making explanations for these predictions essential. To solve this problem, research into Explainable AI (XAI) is being actively pursued. Representative XAI techniques include SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations). SHAP, proposed by Lundberg & Lee (2017), is a powerful framework based on game theory that quantitatively explains how much each feature contributed to the model's final prediction. This allows analysts to understand which factors were the decisive basis for an anomaly detection (Lundberg & Lee, 2017). Additionally, LIME, proposed by Ribeiro et al. (2016), is a technique that provides local explanations by sampling data around a specific prediction and training an interpretable linear model on it (Ribeiro et al., 2016).

This study is differentiated from existing research in that it goes beyond simply developing a high-performance model. By integrating XAI techniques like SHAP, it aims to secure the model's transparency and reliability, and ultimately provide analysts with actionable insights.

3. Research Methodology

To address the complex problem of privilege anomaly detection in multi-cloud environments, this study adopts a step-by-step approach that systematically breaks down the problem and applies optimized techniques at each stage. Based on the recognition that a single data source and a single model cannot effectively detect the multi-faceted attacks of the real world, the methodology of this study is designed around three core principles: ensuring data diversity, extracting meaningful information, and combining complementary models.

The first principle, ensuring data diversity, aims to solve the problem of 'context fragmentation.' Data collected from individual cloud platforms only shows activity within that platform and fails to provide the complete picture of sophisticated attacks that traverse multiple clouds. Therefore, this study aims to provide a robust foundation for the model to learn anomaly patterns from a broader perspective by integrating data from heterogeneous platforms (IBM, Microsoft, AWS) to construct a large-scale, multi-modal dataset.

The second principle, extracting meaningful information, is the process of transforming raw data into a format that the model can understand. Since raw logs and metric data are unstructured or high-dimensional and thus difficult to use as is, they undergo data preprocessing and standardization. Subsequently, Multi-modal

Feature Engineering—encompassing temporal characteristics, security events, and statistical properties—is performed to amplify the latent anomaly signals in the raw data and enable the model to learn patterns more easily.

The third principle, combining complementary models, is necessary to respond to various forms of cloud attacks. Cloud attacks can be categorized into 'Point Anomalies,' where an individual event is itself abnormal, and 'Contextual Anomalies,' where individual events are normal, but their sequence or combination is abnormal. As a single model struggles to effectively detect all these types, this study adopts a Hybrid Ensemble architecture that combines Isolation Forest, which is strong at detecting point anomalies, with LSTM, which excels at understanding temporal sequences and context. Through this, we aim to leverage the complementary strengths of both models to maximize the detection scope, reduce false positives, and enhance the overall Robustness of the model.

In accordance with this design philosophy, the following sections will describe in detail the dataset construction and preprocessing, multi-modal feature engineering, and the design and implementation of the hybrid ensemble architecture.

3.1 Dataset and Preprocessing

3.1.1 Construction of a Large-Scale, Multi-modal Dataset

In this study, we constructed a raw dataset of 2,204,017 rows in total by integrating real-world operational data from three cloud platforms, as shown in Table 1.

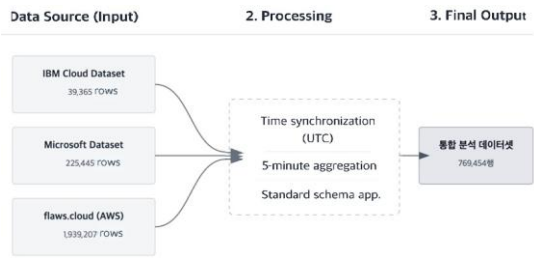
(Table 1 Original Dataset Information)

Dataset	Description	Period	Scale
IBM Cloud Dataset	High-dimensional Telemetry Data	2024.01 - 2024.06	39,365 rows × 117,448 columns
Microsoft Cloud Monitoring	Time-series Service Metrics	2017.11 - 2018.07	225,445 rows
flaws.cloud CloudTrail	AWS Privilege Event Logs	2017.02 - 2020.10	1,939,207 rows

3.1.2 Data Integration and Standardization Framework

To ensure consistency in data collected from heterogeneous cloud environments, an integration framework as shown in Figure 1 was applied. First, time synchronization was performed by normalizing all data timestamps to the UTC standard. Afterward, the data was resampled at 5-minute intervals to align the time series. Finally, to ensure analytical consistency, the schema was unified into 15 standard columns, creating an analysis dataset with a total of 769,454 rows.

(Figure 1: Data Integration and Standardization Pipeline (Conceptual Diagram))



3.2 Multi-modal Feature Engineering

3.2.1 The Necessity of Feature Engineering

Raw data contains several issues that make it difficult for machine learning models to learn from directly. The more than 110,000 metrics in the IBM dataset or the semi-structured JSON logs from CloudTrail are 'information' in themselves, but not 'knowledge' that a model can learn from. Feature engineering is the critical process that bridges this gap between raw data and the

machine learning model, with the following objectives:

First is Signal Amplification: to convert the faint anomaly signals hidden in the raw data into explicit numerical values, making them easier for the model to detect. For example, a derived variable like 'is_weekend' can better explain attack patterns than the timestamp itself.

Second is Dimensionality Reduction & Structuring: to transform high-dimensional or unstructured data into a fixed-size numerical vector that the model can accept as input.

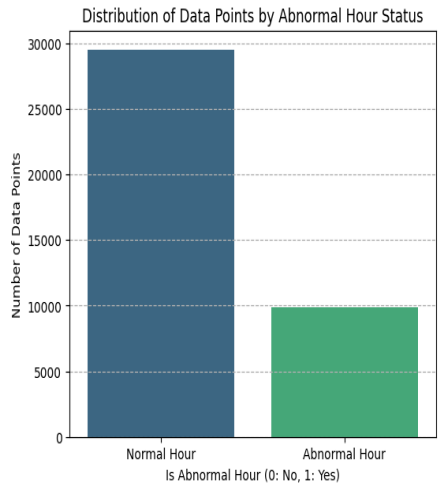
Third is to improve the model's learning efficiency and accuracy by injecting domain knowledge. This is done by converting the knowledge of security experts (e.g., "attackers are primarily active in the early morning," "specific API calls are closely related to privilege escalation") into features.

3.2.2 Time-based Features

An attacker's behavior often shows a distinct difference from the working hour patterns of normal users. In particular, automated attack tools or insider threats tend to occur outside of business hours. To capture this temporal context, features were created for the hour of the event (hour), whether it was a weekend (is_weekend), and whether it was outside of normal business hours (is_abnormal_hour).

The hour feature identifies the activity density at specific times of the day, while the is_weekend feature helps the model learn the difference in activity patterns between weekdays and weekends. Furthermore, the is_abnormal_hour feature can be a decisive clue in detecting automated scripts or insider threats that occur at unusual times. In addition to these, features like the time interval between events (time_since_last_event) were also created. These can be used to detect 'Brute-force' attacks that generate a large volume of events in a very short time, or anomalous behavior from 'dormant accounts' that have been inactive for an unusually long period. An example of the data, including these generated time-based features, is shown in Figure 2.

(Figure 2: Data Distribution by Normal and Abnormal Time Periods)



3.2.1 Security and Behavior-based Features

By directly incorporating security domain knowledge, features closely related to attack scenarios were generated.

The first feature, is_privilege_event, is a binary feature that indicates whether a key API call corresponding to the 'Privilege Escalation' tactic in the MITRE ATT&CK framework has occurred (e.g., CreateUser, AttachUserPolicy). This feature provides crucial information for the model to assess the risk level of a specific privilege escalation attempt when combined with other contextual features.

The second feature, error_rate, represents the ratio of server error (HTTP 5xx) responses to the total number of requests within a specific time window. Under normal circumstances, the error rate remains very low, but it can surge during reconnaissance scans aimed at finding system vulnerabilities or during credential stuffing attack attempts. Therefore, this feature is useful for detecting the initial stages of an attack. The specific code for generating these security and behavior-based features is shown in Figure 3.

```
# MITRE ATT&CK-Based Privilege Event Mapping & Error Rate Calculation

privilege_events = ['CreateUser', 'AttachUserPolicy', 'AssumeRole',
'CreateRole']
df['is_privilege_event'] =
df['raw_event_name'].isin(privilege_events).astype(int)
df['error_rate'] = df['http_status_5xx'] / (df['request_count'] + 1)
```

(Figure 3: Pseudocode for Mapping Privilege Events based on MITRE ATT&CK and Calculating Error Rate)

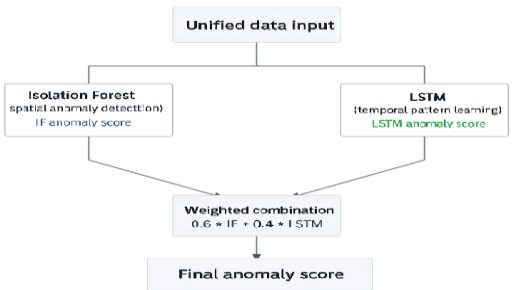
Through this feature engineering process, heterogeneous data collected from various sources were transformed into a standardized 'language' (a 15-feature vector) that all models can commonly understand and learn from. This is the core foundational work that enables the hybrid approach of this study.

3.2 Hybrid Ensemble Architecture

The hybrid ensemble architecture proposed in this study is illustrated in Figure 3. The input data is processed through two independent paths, and their results are then fused on a weight-based basis to produce the final anomaly score.

The first path, Isolation Forest, is responsible for detecting spatial anomalies in high-dimensional numerical metrics. The second path, LSTM, learns the multi-stage sequences of privilege escalation attacks and the temporal dependencies of time-series service metrics. Finally, the anomaly scores calculated by the two models are combined with optimal weights and fused in a way that maximizes detection performance.

(Figure 4: Hybrid Ensemble Model Architecture (Conceptual Diagram))



4. Experiments and Results

This chapter details the process and results of the experiments conducted to validate the performance of the proposed hybrid ensemble framework. The core objective of the experimental design is to conduct a multi-faceted evaluation to determine whether the proposed model shows a substantial performance improvement compared to its individual component models (Isolation Forest, LSTM), to understand the types of errors the model makes and their causes, and to assess whether the model's prediction results can be trusted and interpreted.

4.1 Experimental Environment

The experiments were conducted on the Google Colab Pro platform, using a Tesla T4 GPU and 25GB of RAM for hardware. The primary frameworks utilized were scikit-learn, TensorFlow, pandas, and SHAP. The data used for the analysis consisted of 769,454 rows and 15 standard features.

4.2 Evaluation Metrics

In problems with severe data imbalance, such as security anomaly detection, evaluating a model's performance with a single metric is highly risky. Therefore, this study comprehensively used the metrics shown in Table 2 to evaluate the model's performance from multiple perspectives.

The foundation for all metrics, the Confusion Matrix, categorizes the model's predictions into four types: True Positive (TP): An actual attack was correctly detected. False Positive (FP): A normal event was incorrectly identified as an attack. False Negative (FN): An actual attack was missed. True Negative (TN): A normal event was correctly identified.

(Table 2 Evaluation Metrics)

Metric	Description	Formula
Precision	The proportion of actual attacks among the instances the model predicted as an 'attack.' It indicates the reliability of the model's alerts.	TP / (TP + FP)
Recal	The proportion of actual attacks that the model successfully detected. It represents the model's ability not to miss attacks.	TP / (TP + FN)
F1-Score	The harmonic mean of Precision and Recall. It is used to evaluate the model's balanced performance, especially when there is a trade-off between the two metrics.	2 * (Precision * Recall) / (Precision + Recall)
ROC-AUC	A comprehensive metric for a model's discriminative ability, indicating how well it can distinguish between classes compared to random guessing.	-
PR-AUC	The area under the Precision-Recall curve. It provides a better representation of the model's practical performance, especially in cases of severe class imbalance.	-

4.3 Overall Model Performance Evaluation

To evaluate the overall discriminative performance of the models, the ROC curve and the Precision-Recall curve were analyzed.

(Figure 5: Comparison of ROC Curves on the Test Dataset)

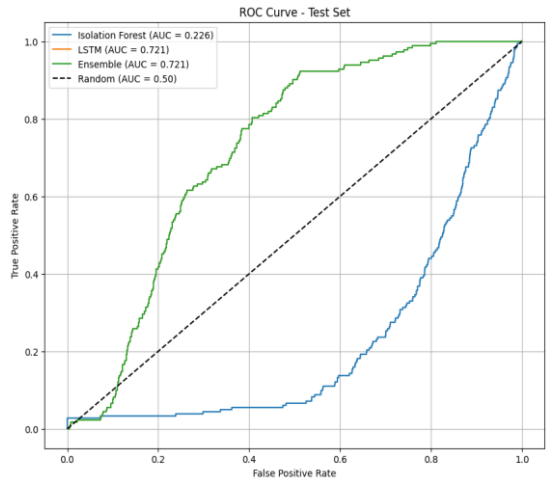


Figure 5 shows the ROC curves for each model. The Area Under the Curve (AUC) is a metric that indicates how well a model can distinguish between classes, with a value closer to 1 indicating better performance. The proposed Ensemble model and the LSTM model both recorded an AUC score of 0.721, demonstrating significantly superior performance compared to the Isolation Forest model (AUC=0.226). This signifies that the ensemble model discriminates anomalous behavior much more effectively than random guessing (AUC=0.5).

(Figure 6: Comparison of Precision-Recall Curves on the Test Dataset)

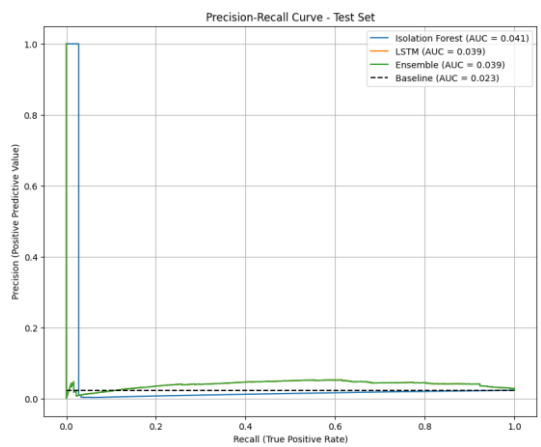
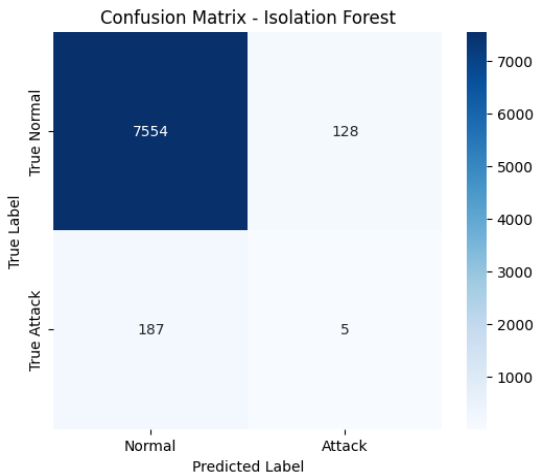


Figure 6 compares the Precision-Recall curves of the test dataset. The relatively low AUC-PR score of 0.039 is attributable to the extreme class imbalance problem within this study's dataset. In other words, it suggests that while the models discriminate well against the majority normal class, they still face challenges in detecting the minority attack class. Nevertheless, the fact that the ensemble model's curve (green) in Figure 6 remains at a higher level than the other models signifies that it achieves better Precision at all Recall levels. This means that for the same number of actual attacks detected, the ensemble model generates significantly fewer False Positives than the other models. This is a critically important characteristic in a real-world operational environment, as it reduces analyst fatigue and allows them to focus on important alerts.

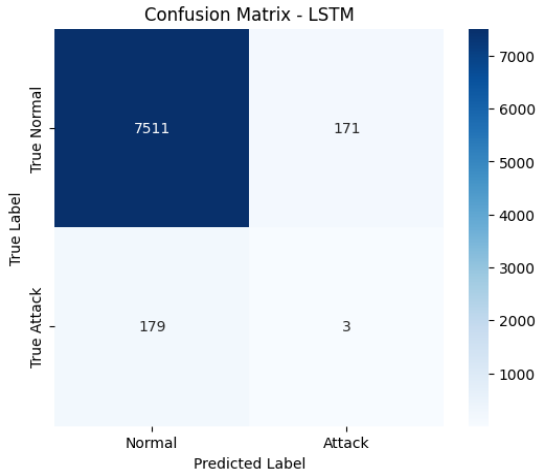
4.4 Detailed Analysis of False Positives and False Negatives by Model

To analyze the specific error types of each model, the confusion matrices in Figure 7, Figure 8, and Figure 9 were compared, and based on these, quantitative performance metrics were calculated as shown in Table 3.

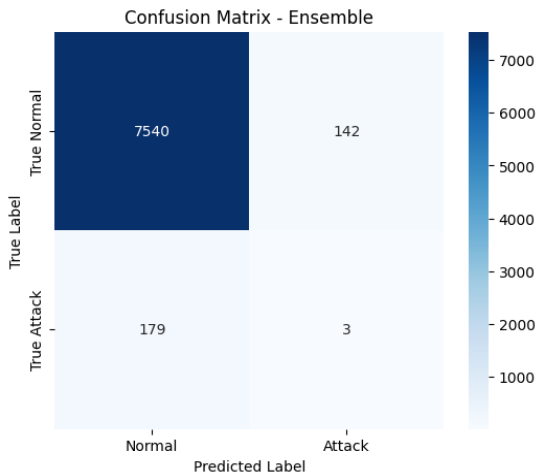
(Figure 7: Isolation Forest Confusion Matrix)



(Figure 8: LSTM Confusion Matrix)



(Figure 9: Ensemble Model Confusion Matrix)



As a result of the comparative analysis, the proposed ensemble model (Figure 9) had the highest number of True Negatives (7540), correctly identifying normal instances. The number of True Positives was 3, identical to the LSTM model.

Notably, the number of False Positives (FP)—instances where normal behavior was incorrectly classified as an attack—decreased to 142, compared to 171 for the LSTM model. This demonstrates that the ensemble method helps to improve the overall false positive rate by compensating for the weaknesses of the individual models.

(Table 3: Detailed Performance Metrics by Model (Based on the Confusion Matrix))

Model	TP	FP	FN	TN	Precision	Recal	F1-Score
Isolation Forest	5	128	187	7554	3.76%	2.60%	3.08%
LSTM	3	171	179	7511	1.72%	1.65%	1.68%
Ensemble	3	142	179	7540	2.07%	1.65%	1.84%

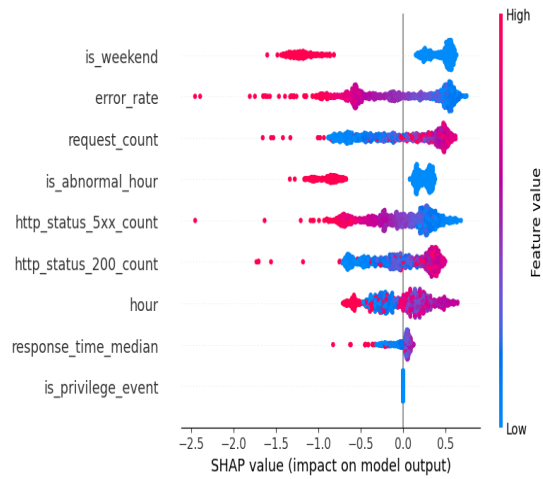
Each metric presented in Table 3 was calculated based on the values from the confusion matrix. Precision refers to the proportion of actual attacks among those the model predicted as attacks (TP / (TP + FP)), while Recall represents the proportion of attacks detected by the model out of all actual attacks that occurred (TP / (TP + FN)). The F1-Score is the harmonic mean of these two metrics (2 * (Precision * Recall) / (Precision + Recall)) and shows a comprehensive performance that considers the balance between them.

The analysis of Table 3 reveals several important facts. First, the Recall for all models remains at a very low level, between 1.65% and 2.60%. This is due to the previously mentioned data imbalance problem, which prevented the models from sufficiently learning from the few attack samples, and it is a clear limitation of this study. Second, the proposed ensemble model detected the same number of attacks as LSTM (TP=3) while reducing the number of false positives (FP) from 171 to 142, an approximate 17% decrease. This resulted in an improvement in Precision from 1.72% to 2.07%. Although the absolute figures are low, this signifies that the ensemble technique substantively contributes to reducing unnecessary alerts by correcting the errors of the individual models.

4.5 Feature Importance Analysis and Interpretation of Results (XAI)

To trust and understand the model's prediction results, the SHAP library was used to analyze which features played an important role in anomaly detection. Figure 10 visually shows the impact of each feature on the model's prediction output (anomaly score). The analysis revealed that is_weekend, error_rate, and request_count were the top three features with the greatest impact on the model's predictions.

(Figure 10: SHAP Summary Plot for the Ensemble Model)



Examining the relationship between each feature and the predictions in detail, for is_weekend, the red dots (representing high feature values) are mostly distributed in the positive SHAP value range. This means the model learned that events occurring on the weekend tend to increase the probability of an attack. This aligns with the common security sense of being suspicious of activities outside of normal business hours.

Furthermore, for error_rate, the red dots (high error rates) are clearly distributed in the positive SHAP value range, while the blue dots (low error rates) are in the negative range. This shows that the model successfully learned the intuitive fact that a higher error rate corresponds to a higher likelihood of an attack. On the other hand, for request_count, there is a mixed tendency where a high number of requests sometimes increases and sometimes decreases the probability of an anomaly. This suggests that the number of requests alone is not sufficient to determine an anomaly, and its combination with other features played a crucial

role.

These SHAP analysis results demonstrate that the model is making predictions based on logical grounds. By providing concrete clues for analysts to trace the cause of alerts and respond, it enhances the practical value of the model.

5. In-depth Discussion and Implications of the Study

In this chapter, the quantitative results presented in Chapter 4: Experiments and Results are comprehensively interpreted, and their academic and practical implications are discussed in depth. Furthermore, the reliability and validity of the proposed framework are verified by analyzing the model's internal operations through Explainable AI (XAI) techniques.

5.1 Comprehensive Interpretation of Experimental Results

The results of this experiment have a dual aspect, simultaneously demonstrating the potential and the clear limitations of the proposed hybrid ensemble model. First, the respectable ROC-AUC score of 0.721 achieved by the ensemble model in Figure 5 suggests that the model possesses a general discriminative ability to distinguish between normal and attack instances across the entire dataset. This implies that the multi-source data and multi-modal feature engineering provided a meaningful foundation for learning.

However, the Precision-Recall (PR) curve in Figure 6 and the detailed metrics based on the confusion matrix in Table 3 reveal a much more critical reality. The extremely low Recall (1.65%~2.60%) observed across all models clearly shows that the most central challenge of this study is the problem of Severe Data Imbalance. With only about 2.4% of the test dataset being attack samples, the models could not sufficiently learn the few attack patterns, which inevitably led to missing many attacks (a high number of FNs).

Despite this, the practical value of the ensemble model is evident in its improvement in Precision. While detecting the same number of attacks as LSTM (TP=3), the ensemble model reduced the number of false positives (FP) from 171 to 142, an approximate 17% decrease. This holds significant meaning in a real-world Security Operations Center (SOC) environment. Since an analyst's time and resources are limited, reducing unnecessary alerts (false positives) has the direct effect of mitigating alert fatigue and helping analysts focus on real threats. In short, the ensemble technique did not detect more attacks, but it contributed to generating more reliable alerts.

5.2 Ensuring the Explainability of Model Predictions

If performance metrics show 'what' a model did, Explainable AI (XAI) builds trust in the model by showing 'why' it did so. In this study, the SHAP library was used to analyze how each feature influenced the model's predictions.

The SHAP analysis in Figure 10 clearly demonstrates that the model makes predictions based on logical grounds that align with common security sense, rather than on random patterns. The features that had the greatest impact on the model's predictions were, in order, is_weekend, error_rate, and request_count. The strong tendency for events occurring on the weekend or those with a high error rate to increase the probability of an anomaly (positive SHAP values) signifies that the model learned to use behavior deviating from normal work patterns and abnormal system responses as key clues for an attack.

This analysis helps a security analyst, upon receiving an alert, to go beyond a simple "anomaly detected" message and start an investigation with a specific cause, such as "judged as an anomaly due to requests with an unusually high error rate occurring late on a weekend." This provides actionable insights that dramatically reduce the time needed to judge whether an alert is a false positive and to respond. In conclusion, the SHAP analysis reinforces the overall reliability of the proposed framework by proving that the previously confirmed

improvement in the ensemble model's precision is not a coincidence, but the result of learning based on rational grounds.

6. Conclusion and Future Work

6.1 Summary of Research and Contributions

This study integrated large-scale, heterogeneous data collected from a real-world multi-cloud environment and, based on this, designed and applied a hybrid ensemble framework for privilege anomaly detection. Through experiments, the proposed ensemble model achieved several concrete accomplishments and contributions compared to the individual models. First, it demonstrated practical utility by reducing false positives. While maintaining the same detection performance (Recall) as the LSTM model, the ensemble model reduced false positives by approximately 17%, thereby improving the precision of its alerts. This can directly contribute to increasing the efficiency of real-world security operations environments. Second, it secured explainability. By introducing SHAP analysis, the prediction results of the complex ensemble model were provided in an interpretable form. This demonstrated that the model operates on a logical basis and provided analysts with actionable insights, thereby securing the model's reliability. Finally, it empirically confirmed the problem of data imbalance. This study quantitatively revealed the extreme class imbalance problem present in real-world cloud security data and clearly showed the impact of this problem on anomaly detection models, especially on their Recall. This represents a significant academic contribution in that it has presented a core challenge that future related research must address.

6.2 Limitations and Future Research Directions

Despite the achievements of this study, the experimental results also revealed clear limitations. To overcome these, the following specific directions for future research are proposed. The biggest limitation of this study, as confirmed by the experimental results (Table 5, Figure 4), is the low Recall caused by severe data imbalance. This means the model could not sufficiently learn from the few attack samples, a point that must be improved for real-world application. To solve this Recall problem, future research could consider using Data Augmentation techniques, such as Generative Adversarial Networks (GANs) or Variational Autoencoders (VAEs), to synthesize realistic and diverse minority class (attack) data. Through this, research will be conducted to fundamentally improve Recall by enabling the model to sufficiently learn rare attack patterns. Furthermore, in the security domain, it must be considered that the risk cost of a missed detection (FN) is far greater than that of a false positive (FP). Therefore, research can be inroduced to strongly guide the model's learning direction to avoid missing the minority attack class by introducing Cost-Sensitive Learning, which involves designing a loss function that imposes a much higher cost (penalty) on false negatives. Finally, we propose advancing the ensemble model proposed in this study by further integrating a Graph Neural Network (GNN), which can model the complex relationships between users, resources, and actions. Through this, we aim to achieve a higher dimension of detection capability by building a true hybrid framework that encompasses the characteristics of individual events (Isolation Forest), the sequence of events (LSTM), and the relationships between events (GNN).

Reference

[1] Ahmad, Z., et al., "Anomaly detection in cloud computing: A systematic review," Future Generation Computer Systems, Vol.119, pp.156-178, 2021.

[2] Cloud Security Alliance, "Top Threats to Cloud Computing: The Pandemic Eleven," 2024.

[3] Du, M., et al., "DeepLog: Anomaly detection and diagnosis from system logs through deep learning," Proceedings of the 2017 ACM SIGSAC Conference on Computer and

Communications Security, pp.1285-1298, 2017.

[4] Hamilton, W. L., Ying, R., and Leskovec, J., "Inductive representation learning on large graphs," Advances in Neural Information Processing Systems, Vol.30, 2017.

[5] Hochreiter, S. and Schmidhuber, J., "Long short-term memory," Neural Computation, Vol.9, No.8, pp.1735-1780, 1997.

[6] Kipf, T. N. and Welling, M., "Semi-supervised classification with graph convolutional networks," arXiv preprint arXiv:1609.02907, 2016.

[7] Liu, F. T., Ting, K. M., and Zhou, Z. H., "Isolation forest," 2008 Eighth IEEE International Conference on Data Mining, pp.413-422, 2008.

[8] Lundberg, S. M. and Lee, S. I., "A unified approach to interpreting model predictions," Advances in Neural Information Processing Systems, Vol.30, 2017.

[9] Milajerdi, S. M., et al., "HOLMES: Real-time APT detection through correlation of suspicious information flows," 2019 IEEE Symposium on Security and Privacy (SP), pp.1137-1152, 2019.

[10] Ribeiro, M. T., Singh, S., and Guestrin, C., ""Why should I trust you?": Explaining the predictions of any classifier," Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.1135-1144, 2016.

[11] Schneider, J., et al., "Anomaly detection in large-scale cloud systems: An industry case and dataset," arXiv preprint arXiv:2406.07966, 2024.

[12] Wu, L., et al., "CloudRanger: Root cause identification for cloud native systems," 2021 IEEE 18th International Conference on Autonomic Computing (ICAC), pp.33-43, 2021.

Author Biography

Chi-Sung Kim

Graduated from Seoul School of Integrated Sciencs & Technologies(aSSIST),
Graduate School of AI, Department of AI Advanced Studies, Major in AI Big Data(Master of Engineering)
Research Interests : AI, Digital Transformation,
Generative AI, Data Analysis etc.
E-mail : nonochi12@gmail.com

Research on how to increase user engagement of Korean AI tutoring service using token economy

Choe Su

Abstract

While AI tutoring has strengths in personalization, it is limited by a lack of emotional and social interaction to keep learners engaged. To address this issue, this research proposes a new model that fuses AI tutoring with blockchain-based "token economy" and "gamification". The token economy provides rewards for learning activities and community contributions (extrinsic motivation), while gamification provides the fun and sense of accomplishment of the learning process itself (intrinsic motivation). The synergy of these two elements aims to build a sustainable learning ecosystem by transforming learning into a "value-creating activity" rather than a mere obligation.

1. Introduction

1.1 Background and Need for the Study

With the advancement of digital technology, artificial intelligence (AI) is emerging as a key driver in education to provide personalized learning. AI tutoring services are becoming ubiquitous in education with advantages such as personalized learning and real-time feedback, but they have clear limitations in keeping users engaged.

The biggest problem is that AI cannot provide the same emotional support or interaction as a human tutor, making it difficult to maintain motivation to learn. In fact, the initial dropout rate for online education platforms in South Korea is as high as 36.7%, suggesting that technological advancements alone cannot guarantee learners' continued engagement. These limitations are analyzed because of AI tutoring focusing only on learning efficiency and failing to fully consider learners' psychological and social motivation.

As an alternative, this study proposes a combination of blockchain-based 'token economy' and 'gamification'.

Token economy is a sophisticated reward system based on behavioral psychology that gives learners a sense of ownership and transparency to increase extrinsic motivation and engagement. Gamification stimulates intrinsic motivation by applying game elements such as fun and immersion to the learning process. While current AI tutoring leverages piecemeal game elements, the synergy between the two can maximize learning through a balance of extrinsic rewards (tokens) and intrinsic enjoyment (gamification). This presents a new paradigm that transforms learning into a value-creating activity with rewards rather than a passive duty.

Therefore, this study aims to design an innovative model that combines token economy and gamification to solve the inherent problems of user engagement and churn in Korean AI tutoring services and explore how to build a sustainable learning ecosystem.

[Table 1] AI Key limitations of AI tutoring services and applicability of token economy/gamification

Key limitations of AI tutoring services	Token economy applicability	Gamification applicability
Lack of emotional interaction	Community contribution rewards, peer tutoring incentives	Collaborative quests, group challenges, social NFTs
The Challenges of Training Creative Thinking	Content creation rewards, problem-solving challenge rewards	Creative quests, storytelling-based learning
High dropout rates and learning persistence issues	Reward completion of learning activities, incentivize long-term engagement	Leveling up, badges, leaderboards, continuous learning rewards
Lack of instructor/student interaction	Community activity rewards, governance participation rights	Team-based learning, peer review, social enhancements

1.2 Research objectives

This study aims to propose a new service model that fuses blockchain-based token economy and gamification elements to solve the problems of user engagement and learning persistence in Korean AI tutoring services. The specific research objectives are as follows

First, we analyze the status of Korean AI tutoring services and the factors that hinder user engagement.

Second, we analyze prior research on token economy and gamification and their success/failure cases to draw implications for application in the education field.

Third, we design a token economy-based Korean AI tutoring service model to enhance learning motivation and social interaction and propose a specific token design and operation plan.

Fourth, we present the system architecture and main function implementation of the proposed model and propose evaluation methods for future empirical studies.

1.3 Organization of the Thesis

This thesis is organized into five chapters. Chapter 1, Introduction, presents the background and necessity of the study, explains the purpose of the study and the organization of the thesis.

Chapter 2, Related Research, provides an in-depth review of previous research on Korean AI tutoring services, token economy, and gamification.

Chapter 3, the proposed token economy-based AI tutoring service model, presents the service overview, token design and operation plan, and user engagement strategy in detail.

Chapter 4, System Implementation and Evaluation, describes the system architecture of the proposed model, how the main functions are implemented, and the experimental design and data analysis methods for evaluation.

Finally, Chapter 5, Conclusions and Future Research, summarizes the main contents of this research and presents the limitations and future research directions.

2. Related Research

2.1 Korean AI Tutoring Service

AI tutors are powerful learning tools that analyze an individual's learning data to provide customized materials and instant feedback. While their core strength is personalization, which improves learning efficiency and provides learning opportunities without time and space constraints, they face clear limitations.

The biggest challenge is sustaining user engagement. AI can't provide the emotional support or empathy of a human tutor, nor can it foster creative thinking. This, coupled with the lack of interaction common to online learning environments, leads to low learning motivation and high dropout rates. In fact, one training platform reduced its churn rate from 36.7% to 10.9% by increasing interaction with administrators and fellow learners, suggesting that social and emotional factors beyond technical personalization are crucial to learning persistence.

In conclusion, “personalized learning” by AI tutors increases the cognitive efficiency of learning but does not satisfy the emotional and social aspects of learning and does not ensure the continuity of the learning “experience.” Therefore, it is essential to introduce systematic motivation and social interaction mechanisms to compensate for the technical limitations of AI.

2.2 Token Economy

Token economy is a methodology for designing the economic system of a blockchain ecosystem, based on the principle of “reinforcement” in behavioral psychology, which is based on rewarding certain behaviors with tokens to drive engagement, and can be effectively used in education to motivate learning.

Successful examples such as ODEM, BitDegree, and Korea's Santatoic Coin have shown the potential of the Learn-to-Earn (L2E) model to engage students by tying cryptocurrency rewards to learning activities. On the other hand, the Terra/Luna debacle has taught us that token economies that rely on unstable mechanisms and excessive rewards are doomed to failure. This means that sustainability and transparent governance of the token design are crucial.

The most important principle for designing token economies in education is to focus on learning outcomes, not speculation. If users are only focused on “earning”, they will easily abandon when the value of the token drops, so the value of the token should be directly linked to learning outcomes and have real utility on the platform, such as unlocking learning content and accessing features. In other words, the token should not be an investment, but a reinforcement of the learning journey to encourage long-term engagement.

2.3 Gamification

Gamification is a strategy that applies the fun elements and mechanics of games to non-gaming contexts to engage and immerse users. In education, it is effectively used to increase learning and motivation.

One of the main theoretical backgrounds behind gamification is Self-Determination Theory (SDT). This theory emphasizes three core psychological needs that drive intrinsic human motivation: competence, autonomy, and relatedness.

Competence can be promoted by tracking learning progress and providing a sense of accomplishment through reward systems (points, level-ups, badges, trophies). Completing quests of appropriate difficulty is also important for a sense of competence.

Autonomy can be enhanced by giving learners the experience of making their own choices and pursuing their own goals in a game.

Cooperative play and social interactions with other gamers can foster a sense of belonging, which can lead to a sense of connection. However, gamification that relies solely on extrinsic rewards such as points or badges can easily lose momentum once the "novelty effect" wears off. Therefore, a design that continuously reinforces intrinsic motivation is essential to drive long-term engagement.

In the context of this research, gamification plays two key roles

A. The role of designing meaningful rewards requires combining token economics (extrinsic rewards) with self-determination theory, i.e., making tokens not just rewards, but meaningful symbols, such as proof of learning achievement (competence), content choice (autonomy), and recognition of community contributions (relatedness). This will help learners feel satisfaction in the act of earning rewards and keep them enthusiastic.

B. The role of promoting social interaction: Gamification should be used to address the inherent lack of interaction and learner isolation in AI tutoring. By linking token rewards to collaborative activities such as peer learning and group study, it is possible to fulfill the need for relationality and strengthen the sense of community through cooperation rather than competition, thereby increasing learning persistence.

[Table 2] Gamification elements and motivational effects to boost user engagement

Reformulation Factors	Self-Determination Theory Elements	Intrinsic Motivation Effects	Extrinsic Motivational Effects
Points, Experience	Sense of competence	Visualize learning progress to promote a sense of accomplishment	Instant rewards, the basis for earning tokens
Badges, Trophies	Sense of competence, connection	Recognition of learning achievements, boosting self-esteem	Tangible rewards, social recognition
Leveling Up	Sense of competence	Feeling of continuous growth, stimulating the desire to improve	Unlock new features/content, grant privileges
Leaderboards	A sense of competence, connection	Encourage healthy competition, motivate through performance comparisons	Additional rewards and honors for top rankers
Quests, Challenges	Sense of competence, autonomy	Fun, problem-solving, and goal achievement	Reward tokens/items for completing quests
Storytelling	Autonomy	Engage, excite, and inspire learning	
Avatar customization	Autonomy	Fulfill the need for self-expression, promote a sense of ownership	
Teamwork, collaborative activities	Relatedness, competence	Fostering a sense of belonging, accomplishing common goals	Communal token rewards for meeting team goals
Peer tutoring	Sense of competence, connection	The satisfaction of sharing knowledge, the reward of helping others	Incentivize knowledge sharing with token rewards

3. Proposed Token Economy-based AI Tutoring Service Model

3.1 Service Overview

The service proposed in this study aims to enhance both intrinsic and extrinsic motivation of learners and increase engagement by fusing token economy and gamification with the personalized learning capabilities of existing AI tutors. Learners will be engaged in learning through transparent rewards and fun elements, while taking ownership of their learning history.

The key features are summarized in four areas

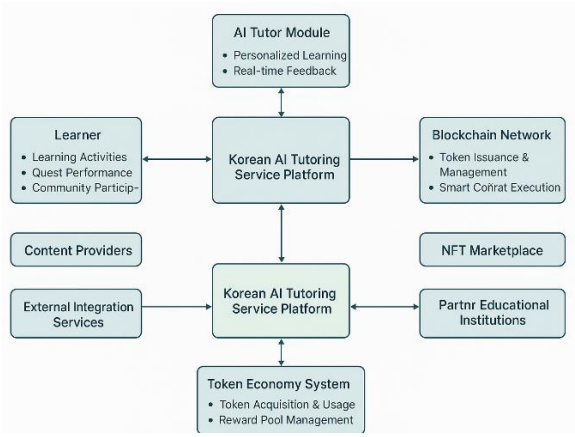
First, the AI-powered personalized learning feature provides personalized learning, including conversations with AI chatbots and vulnerability analysis.

Second, the Learning Reward System feature rewards all learning activities and achievements with tokens to create strong motivation.

Third, gamification elements add fun and engagement to the learning process through levels, quests, leaderboards, and more.

Fourth, the decentralized learning record feature permanently records learning history on the blockchain to ensure authenticity.

The service ecosystem consists of individual learners, AI tutor systems, content delivery experts, and a platform operator that manages the entire system.



[Figure 1] Token economy-based AI tutoring service model overview

The central structure of the AI tutoring service model is centered on the Korean AI tutoring service platform, which serves as the hub for all components. The service model has a virtuous cycle of learner activity → AI feedback → blockchain record → token reward → learning motivation. The main module-specific functions are personalized learning and real-time feedback in the AI tutor module, token management, smart contracts, and learning record storage in the blockchain network.

The token economy system operates token acquisition/use and reward pools, and learners participate in various learning activities and communities.

3.2 Token Design and Operation Plan

The proposed service uses a multi-layered token system to engage learners and activate the learning ecosystem.

The token types and roles can be defined as follows.

The role of the utility token (LER: Learn-to-Earn Reward Token) functions as a direct reward for learning activities. It is utilized to use various learning-related functions and contents within the service, and is the main token earned based on learning performance and participation. This is a similar concept to Boomco's LER token.

The role of the governance token (KAI: Korean AI Tutor Governance Token) grants the right to participate in platform operations and major decision-making and is issued to users who hold a certain amount of LER tokens or make high contributions to the platform, and promotes participation and ownership in the long-term direction of the platform, like Boomco's BOOM token.

Non-Fungible Tokens (NFTs) can be utilized as proof of learning achievements (digital degrees/certificates, badges), learning items (avatars, learning tools), and access to special content. They are unique and scarce and serve as a permanent record and proof of a learner's efforts and achievements. This is like the concept of Axies in Axie Infinity or bag NFTs in Boomco.

3.2.1 Token Earning Mechanism

The token earning mechanism is designed to engage learners by rewarding their various learning activities and contributions.

When completing learning activities, LER tokens are awarded when completing a set learning module (lesson, chapter) within the Korean AI Tutoring Service according to the regular learning results.

Additional LER tokens are awarded for successfully completing tasks (writing, speaking exercises, quizzes) presented by the AI tutor and receiving AI feedback.

Clean-up/review wrong answer notes LER tokens are awarded for cleaning up wrong answer notes based on the learner's weaknesses analyzed by the AI, or for repeating certain concepts.

In terms of user engagement and contributions, LER tokens are awarded for active interactions within the community, such as answering learning questions, sharing learning materials, tutoring fellow learners, and participating in group studies.

The token earning mechanism for content creation is that LER tokens and KAI tokens can be earned for creating quality Korean learning content (e.g., example sentences, exercises, cultural information) that is evaluated as useful by other users.

In the case of bug reports/suggestions for improvement, LER tokens are awarded for providing meaningful feedback that contributes to the improvement of the service.

Additional LER tokens and KAI tokens will be awarded for continuous learning over a certain period or achieving certain learning goals, depending on the results of long-term/continuous participation.

The NFT holding mechanism is designed to allow learners who hold certain NFTs (e.g., learning assistant NFTs) to increase their LER token earnings or receive bonus tokens for certain learning activities. This is based on Boomco's bag NFT example.

3.2.2 Token usage and reward system

The earned tokens provide various utilities within the service and are linked to a reward system that keeps learners motivated.

The in-service utilities act to unlock premium content/features, and LER tokens can be used to unlock advanced Korean learning content (advanced grammar, advanced conversation), special features of the AI tutor (in-depth correction, practice tests), or additional conversation time with the AI tutor. You can also purchase learning items in the form of NFTs, such as avatar customization items and virtual tools to help you learn (e.g., AI dictionary extensions).

Discounts/Course Passes: LER tokens can be used to purchase discounts on paid courses or passes for a specific period.

In terms of reward systems, the level system allows learners to level up based on their learning progress and token earnings and rewards them with special badge NFTs or additional LER/KAI tokens upon leveling up. The Badge/Achievement system recognizes learners' achievements by issuing Badge NFTs for various achievements, such as achieving specific learning goals (e.g., passing a TOPIK exam), reaching a certain number of consecutive study days, or contributing to the community. The Leaderboard system creates weekly/monthly leaderboards based on learning volume, token earnings, community contributions, etc. and rewards top rankers with additional tokens or special NFTs to encourage healthy competition. Challenge/Quest System: AI tutors will provide personalized or community-based learning challenges/quests and reward LER tokens and special rewards upon completion. The governance participation system allows KAI token holders to directly participate in the operation of the platform by voting on important platform decisions (e.g., adding new learning content, changing token economy policies).

The service's token economy is designed based on two core principles: “learning relevance” and “sophisticated rewards”.

First, the utility of the token is focused on enhancing the learning experience itself. The value of a token is not determined by external markets, but by how compelling its use is within the platform. Therefore, tokens should provide utility that is directly linked to learning outcomes, such as unlocking premium content or accessing advanced features of an AI tutor. This makes the token a “symbol of learning achievement” rather than just a currency, and the issuance of NFT-based digital certificates reinforces this.

Second, the mechanism for earning tokens must be sophisticated and fair. Rewards should be tied to the “quality contribution” of learning, not just the quantity of activity, and apply psychological principles such as “shaping,” which rewards progressively increasing difficulty, to encourage good learning habits. Furthermore, all reward rules should be transparently published through smart contracts and designed to avoid concentrating rewards in the hands of a few, thus ensuring user trust and system sustainability.

[Table 3] Proposed token types and acquisition/redemption mechanisms

Token Types	Definitions and roles	Earning Mechanism	Uses and reward system
LER (Utility Token)	Direct rewards for learning activities	Regular study/assignment completion, organizing/reviewing wrong notes, community activity	Unlock premium content/features, purchase learning items, purchase discounts/courses, level up/badge/leaderboard rewards
KAI (Governance Token)	Rights to participate in platform operations and major decisions	Hold a certain amount of LER, reward long-term/continuous participation, reward quality content creation	Voting rights for major platform decisions
NFT (non-fungible token)	Proof of learning, learning items, and access to special content	Increased LER earned for meeting certain learning goals, leveling up, purchasing learning items, and holding NFTs	Digital diplomas/certificates, own/trade learning items, access to special content, profile display and social recognition

3.3 User Engagement Strategies

3.3.1 Learning motivation factors

Learning motivation is designed based on Self-Determination Theory (SDT). SDT explains that intrinsic motivation is driven by the fulfillment of three psychological needs: competence, autonomy, and relatedness.

To provide clear learning goals and progress, the AI tutor sets personalized learning goals and tracks learning progress (progress rate, percentage correct, study time) in real-time, providing a visual representation. It reinforces a sense of competence by allowing learners to see their growth through level-ups, experience bars, and more. In addition to real-time feedback from AI tutors for instant feedback and achievement rewards, we also reinforce a sense of accomplishment by offering instant rewards such as LER tokens for completing learning activities, badge NFTs for completing quests, and level-ups for achieving certain difficulties. The AI analyzes the learner's level and achievements to present challenges that are “not too easy, not too hard,” so that learners experience a sense of competence without frustration through appropriate difficulty adjustments.

In addition to AI-recommended learning paths, the desire to promote autonomy gives learners the option to choose a personalized learning path, where they can freely select learning modules, content, and challenges based on their interests and goals. Avatar and profile customization, where learners can create their own avatars and decorate them with tokens or NFTs, gives them a sense of ownership and autonomy. By offering flexibility in when and how they learn, we provide a flexible learning environment that gives learners the freedom to learn on their own schedule.

The core of our motivation strategy is to create a synergy between intrinsic motivation (the enjoyment of learning itself) and extrinsic motivation (tangible rewards).

To prevent the “process-per-course effect” where extrinsic rewards (tokens) can undermine intrinsic motivation, the principle is to link token rewards to the “process” and ‘effort’ rather than the final “result” of learning.

The goal is to use tokens as an extrinsic reward as a tool to reinforce the three core elements of self-determination theory: competence, autonomy, and relatedness.

Competence is designed to provide immediate recognition of learners' achievements, autonomy is designed to give learners a sense of control by allowing them to choose how to spend their earned tokens, and relatedness is designed to promote a sense of belonging by rewarding social activities such as community contributions.

This design allows learners to perceive tokens as meaningful indicators that symbolize their learning journey and growth, rather than just currency, which motivates continued participation.

3.3.2 Social interaction elements

To compensate for the limitations of AI tutors and enhance learning continuity, we actively introduce social interaction elements.

First, for social interaction through community-based learning, we provide an online community space where learners can ask questions, share answers, and support each other's learning progress through ‘Enabling Learning Community’. To encourage voluntary participation by linking token rewards to active activity within the community, LER tokens are awarded to learners who are proficient in a specific learning topic to help or tutor other learners, and a “Peer Tutoring and Mentoring” feature is provided to promote voluntary knowledge sharing. It also provides team-based learning quests through group challenges/collaborative learning to encourage learners to work together to solve problems and achieve goals. Token rewards are awarded to all team members for meeting team goals. Leaderboards and ranking system drive healthy competition by maintaining individual and group learning performance leaderboards, with additional rewards and honors for top rankers.

Secondly, for social interaction through social NFTs and digital identities, the learning achievement NFT sharing function is designed to allow learners to display and share earned badge NFTs or proof of learning achievement NFTs on social media or on their profiles on the platform to promote learner pride and social recognition. Self-expression through avatars and profiles, allowing learners to express themselves and connect with other learners through avatars and profiles that reflect their learning styles, interests, achievements, etc.

Finally, through offline connectivity activities, the service goes beyond online community activities to connect online and offline learning experiences by incentivizing offline study groups among learners in specific regions or Korean cultural experience activities with tokens.

To address the limitations of AI tutors' emotional interaction and the inherent isolation of online learning, the service takes as its core strategy the creation of a ‘social learning environment’. To do so, it leverages the token economy to actively incentivize peer-to-peer (P2P) interactions between learners. This is done by rewarding tokens for any activity that helps each other, such as answering a peer's question, participating in a study group, or sharing useful learning materials.

This design encourages learners to act as “assistant tutors” and “emotional supporters” for each other. As a result, learning extends beyond an isolated individual activity to a “communal experience,” which compensates for the technical limitations of AI tutors and dramatically increases learning persistence.

4. System Implementation and Evaluation

4.1 System Architecture

The system architecture of the proposed service model is designed to seamlessly integrate AI tutoring, blockchain-based token economy, and gamification elements to provide users with a seamless learning experience.

The main components of the system architecture are as follows.

First, the User Interface Layer. It consists of web and mobile applications through which learners interact with the service, and provides an intuitive and user-friendly UI/UX, including a learning dashboard, AI tutor chat interface, learning content viewer, token wallet, community board, leaderboard, and NFT gallery.

Second, there is the AI Tutoring Engine. This is the core module that understands Korean questions, generates responses, and provides grammar correction, vocabulary explanation, pronunciation feedback, etc. in real-time through the Natural Language Processing (NLP) module. The Learning Analytics module collects and analyzes learning data (progress, correct answer rate, study time, wrong answer patterns) in real time to identify learners' strengths and weaknesses. The Personalization Recommendation module recommends personalized content and quests optimized for learners' levels and weaknesses based on learning analytics results.

An optional emotion recognition module can include the ability to detect learner focus, frustration, etc. through facial expression and voice tone analysis, and adjust the learning experience accordingly.

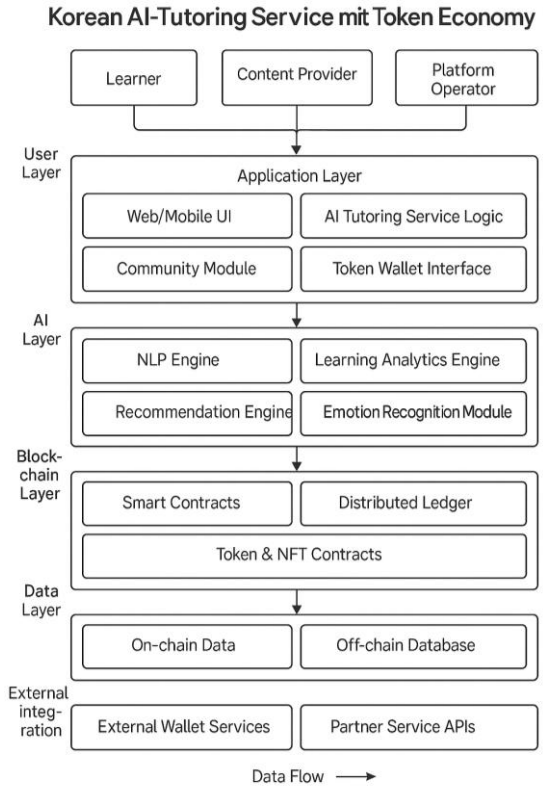
Blockchain Network is an optional consideration, and the appropriate platform can be a public blockchain (e.g., Ethereum, Klaytn) or a private/consortium blockchain (e.g., Hyperledger), depending on transaction processing speed, scalability, fees, ease of development, etc. Transparency and immutability of learning records, issuance and transfer of tokens. Smart Contracts are programs that automate the rules for issuing, earning, and redeeming tokens. They are implemented to automatically award LER tokens when certain conditions are met, such as completing learning, fulfilling quests, and contributing to the community, and are responsible for NFT issuance and ownership management, governance voting logic, etc. The Distributed Ledger will immutably store all learning activity records, token

transactions, NFT ownership information, etc. to ensure data integrity and reliability.

Third, the Token Economy Management System manages the total supply, issuance cycle, and distribution pool of LER and KAI tokens, and operates token pools for learning rewards, community contribution rewards, etc. It also manages the creation and burning of learning achievement NFTs, item NFTs, etc.

Fourth, the database stores a large amount of data that is not directly stored on the blockchain, such as learning content, user profiles, and AI model training data.

Fifth, the API Gateway serves as an interface for secure and efficient communication between each module.



[Figure 2] System architecture

The layered architecture is organized into six distinct layers, clearly separating the roles and responsibilities of each technology stack. The components of each layer are: the User Layer, consisting of learners, content providers, and platform operators; the Application Layer, consisting of UI/UX and service logic; the AI Layer, consisting of core AI functions (NLP, learning analytics, recommendations, emotion recognition); the Blockchain Layer, which ensures token economy and trustworthiness; the Data Layer, which manages on-chain and off-chain data; and the External Integration Layer, which integrates external services for scalability.

4.2 Implementing key features

The main functional implementation of the proposed service model is as follows.

The AI tutoring feature implementation plan implements a chatbot that utilizes Korean natural language processing (NLP) technology to understand learners' questions and provide grammar correction, vocabulary explanation, pronunciation feedback, etc. in real time through the implementation of an interactive learning interface. It also designs/implements personalized learning paths to identify learners' strengths, weaknesses, and learning speed through an AI learning analysis module, and dynamically adjusts the learning content and difficulty level to optimize for them. The correction and feedback system provides AI-based correction and specific improvement feedback on learners' writing and speaking exercises to maximize learning effectiveness.

The token economy feature implementation plan will use smart contracts to implement an automated token payment system that will automatically disburse LER tokens to learners' wallets when certain conditions are met, such as completing learning modules, succeeding in quests, and contributing to the community. In addition, we will develop a token usage linkage system that uses LER tokens to unlock premium content, purchase learning items, apply discounts, and other utilities within the service. We will also implement an NFT issuance and management system for the NFT Gallery feature that allows learners to issue NFTs such as learning achievement badges, digital degrees/certificates, and special learning items, and manage and display them. For gamification and engagement features, visualize learners' levels, experience points, earned badges, leaderboard rankings, etc. in an intuitive UI/UX to motivate learning. Implement a quest/challenge system that allows AI to generate individual or group learning quests, and clearly present quest objectives, progress, and rewards. Community interaction tools enable social interaction and foster relationships among learners by providing learner-to-learner chat, message boards, group study creation, peer review, and more.

Finally, to ensure security and reliability, we leverage the data integrity and transparency properties of blockchain to prevent tampering with learning records. Apply measures to minimize security vulnerabilities in blockchain-based services, such as securing user wallets and auditing smart contracts.

4.3 Experimental Design and Data Analysis

The experimental design and data analysis method to verify the effectiveness of the proposed token economy-based Korean AI tutoring service model is as follows.

A. Experimental design

The ‘research subjects’ of the experimental design are a group of users of a certain age group or learning level who wish to learn Korean. It is necessary to set up an experimental group and a control group, where the ‘experimental group’ uses the Korean AI tutoring service with the proposed token economy and gamification elements, and the ‘control group’ uses the existing AI tutoring service (without token economy and gamification elements). 'Experiment period' means a sufficient period of at least one semester to verify the long-term motivational effect. Measurement points: pre-measurement, periodic

interim measurements, and post-measurement to capture learner change at multiple points. 'Data Collection Method' is used to collect quantitative data such as learning time, learning module completion rate, quest completion rate, token acquisition and usage, community activity frequency, learning achievement (quiz score, practice test score), dropout rate, and retention rate through system logs, and qualitative data is used to understand learners' subjective experiences and perceptions through surveys (learning motivation, satisfaction, usefulness, usability, self-efficacy, relationship) and in-depth interviews (learning experience, perception of token economy, improvement points).

B. Key Performance Indicators (KPIs)

The ‘User Engagement’ metric measures the number of daily/weekly active users (DAU/WAU), learning module completion rate, quest completion rate, community posts/comments, peer tutoring participation rate, etc. Learning persistence metrics utilize key metrics such as weekly/monthly study time, churn rate, revisit rate, and retention rate. The ‘Learning Achievement’ metric measures learning effectiveness through changes in quiz/test scores within AI tutoring and external standardized Korean proficiency test (TOPIK) scores. The ‘Motivation’ metric evaluates changes in intrinsic and extrinsic motivation levels based on self-determination theory through surveys. The ‘Satisfaction’ metric measures user satisfaction with the overall service.

C. Data analysis methods

'Quantitative analysis' methods use statistical hypothesis testing (t-test, ANOVA), regression analysis, time series analysis, etc. to verify significant differences between the experimental and control groups. The ‘Qualitative Analysis’ method uses Content Analysis, Topic Modeling, etc. to derive the main trends and in-depth meaning of user feedback. The ‘Blockchain Data Analysis’ method analyzes token transaction data, NFT issuance and transaction records, etc. to analyze the degree of activation of the token economy and user behavior patterns.

It is important to deeply understand the “motivation mechanism” through the fusion analysis of quantitative indicators and qualitative feedback. The fact that the churn prediction function has a 95% accuracy rate and can utilize churn prediction data on a per-customer basis shows the sophistication of quantitative data analysis. Gamification research shows the largest effect size for “motivation” and suggests that integrating features that increase “intrinsic motivation” is important. Quantitative metrics such as churn rate or learning completion rate alone do not capture the underlying “motivation” behind why users engage and disengage, so experimental designs must combine quantitative data (token acquisition/spending, learning time, and persistence) with qualitative data from surveys and in-depth interviews (learning motivation, satisfaction, self-efficacy, and relationships).

This will give you a deeper understanding of how the token economy and gamification elements affect learners' “extrinsic behavior” as well as their “intrinsic motivation”. For example, we can verify the actual “motivation mechanism” of the proposed model by understanding whether learners who earned a lot of tokens were simply motivated by extrinsic rewards, or whether earning tokens made them feel competent and engaged in learning itself. This will provide crucial insights for future model improvement and optimization.

4.4 Evaluation Results and Analysis

The hypothesized outcome is that the proposed token economy-based service group will show significantly higher user engagement rates, learning persistence, learning achievement, and motivation levels compared to the control group. There will be a significant increase in the rate of participation in community activities and peer tutoring, suggesting a positive impact on satisfying learners' relational needs. Analysis of token acquisition and usage patterns will reveal that certain learner behaviors (e.g., taking on difficult tasks, long-term learning) are effectively reinforced by the token economy.

In the direction of analysis, we will test for statistical significance between the experimental and control groups for each KPI. Qualitative data analysis will reveal qualitative aspects of the user experience, changes in learners' perceptions of the token economy and gamification elements, and recommendations for improvement. We will distinguish between successful and unsuccessful engagement strategies and further analyze the reasons behind them. Early signs of token economy sustainability and potential problems (e.g., token inflation, speculation) will be analyzed and countermeasures will be discussed.

5. Conclusion and Future Research

5.1 Conclusion

this study, we proposed a new service model that fuses blockchain-based token economy and gamification elements to solve the inherent user engagement and learning persistence problems of Korean AI tutoring services. We analyzed the technical limitations of existing AI tutoring (lack of emotional interaction) and the dropout rate of online learning in depth and examined the psychological basis and educational effects of token economy and gamification to secure the theoretical justification of the proposed model.

The proposed model provides a concrete way to maximize learning motivation and promote social interaction through a multi-layered token system using utility tokens (LER), governance tokens (KAI), and NFTs, and a gamification strategy that includes elements that promote competence, autonomy, and relationships based on self-determination theory. In addition, the system architecture, including the AI tutoring engine, blockchain network, and token economy management system, as well as the implementation plan for major functions, and the experimental design and evaluation metrics to verify the effectiveness of the proposed model were presented, laying the foundation for future empirical studies.

This study presents an innovative approach for enhancing user engagement in Korean AI tutoring services and contributes to the exploration of the positive impact that

digital assets and blockchain technology can have on the education sector.

5.2 Research limitations

Since this study focused on the conceptual design and theoretical validation of the proposed model, it is limited by the lack of actual system implementation and effectiveness verification through empirical user data. As the proposed model is premised on the convergence of various advanced technologies such as AI, blockchain, token economy, and gamification, unexpected challenges such as technical difficulties, scalability, security issues, and interoperability may arise in the actual implementation process.

The value of tokens can fluctuate depending on market conditions, which can have a positive or negative impact on learner motivation. This study could not fully analyze and address this market volatility in depth. As failures like Terra Luna have shown, instability in the token economy can be devastating to user trust and platform sustainability. Another limitation is the lack of in-depth discussion of additional strategies for maintaining long-term learning sustainability, given the potential for de-motivation due to the "freshness effect" of gamification.

5.3 Future Research Directions

Future research should proceed in four directions.

First, empirical effectiveness verification and optimization. The proposed model should be implemented into a real system, and the learning persistence and motivation effects should be quantitatively verified through user studies. In addition, we need to continuously optimize the token economy based on real data and research stabilization mechanisms to respond to market volatility.

Second, advanced AI-powered personalization. Develop models that allow AI to analyze learners' emotional states and learning styles to dynamically personalize rewards and gamification elements, such as offering special quests when they are frustrated.

Third, legal and regulatory responses. It is necessary to continuously monitor changes in domestic and foreign laws and regulations on token-based education services and study policy responses to them.

Fourth, securing technical scalability. It is essential to study how to secure system scalability and interoperability with other education platforms in preparation for a large influx of users in the future.

References

[1] Jin Young Kim et al, "Current Status and Technology Trends of AI-based Education," Electronics and Telecommunications Research Institute (ETRI), Trends Report, 2023.

[2] Gungmin Nam, "Applicability and Issues of Generative AI-based Edtech to Korean Language Education," in Proceedings of the Korean Education Association, 2024.

[3] Seungmin Lee, "AI Tutors: From Pros and Cons to How to Utilize 200%," HRD Research, 2023.

[4] Software Policy Institute, "Tokenized Economy and the Future of Blockchain," Software Policy Institute, Report, 2021.

[5] Talentnet, "Blockchain-based Business Models: A New Innovation Map for Startups and Ventures," 2022.

[6] BitDegree, "BitDegree Whitepaper," 2023.

[7] ODEM, "ODEM Whitepaper," 2023.

[8] Sujin Yoon et al. "A Literature Review on Gamification Research in Education," Korean Society of Educational Technology, 2021.

[9] Panopto, "7 Ways to Use Gamification in Education," 2023.

[10] Eunyoung Cho, "A Case Analysis of Gamification in Economic Education," The Korea Economic Education Association, 2020.

[11] Minsoo Kim et al, "Research on Token Economy Design Guidelines," Proceedings of the Korea Blockchain Society, 2022.

[12] Younghan Kim, "Considerations for Designing a Blockchain Token Reward System," Digital Asset Research, 2023.

[13] FasterCapital, "Token Economies and Incentives: Marketing in the Token Economy: Leveraging Incentives for Success," 2023.

[14] Jemin Yoo et al. "Key Features and Latest Trends in AI Infrastructure Architecture," Goover, 2023.

[15] Park, Sang-Hyuk et al. "A Study on Convergence of Artificial Intelligence and Blockchain in Defense," Korea Institute of Science and Technology Information (KISTI), 2022.

[16] Younghee Lee et al, "A study on decentralized system design based on token economy," Journal of Digital Asset Management, 2021.

[17] Cheoljoo Kim, "Online learning platform user engagement measurement metrics and analysis methodology," Educational Engineering Research, 2024.

[18] AppsFlyer, "What is DAU (Daily Active Users)? | AppsFlyer Glossary," 2023.

[19] Korea Institute of Science and Technology Information, "Blockchain-based Badge Service Platform

Design Research Report," Korea Institute of Science and Technology Information, Report, 2022.

[20] Software Policy Institute, "Evaluation Model for Blockchain Service Application," Software Policy Institute, Report, 2020.

[21] Gil-dong Hong et al, "A Study on Activating Blockchain-based Learning Community," Korea Educational Information and Media Society, 2023.

[22] Korea Employment Information Agency, "Motivation Plan for Effective Utilization of AI Tutoring Service," 2024.

[23] Eunjung Kim et al. "Design of a learning data integrity guarantee system using smart contracts," Proceedings of the Korea Information Science Society, 2022.

[24] KB Bank, "Application of gamification-based reward system in financial services," KB Think, 2023.

[25] Google, "Google Developer Documentation: Smart Contract Development Guide," 2024.

[26] Microsoft Azure, "Azure Blockchain Workbench Architecture," 2024.

[27] AWS, "AWS AI/ML Services Architecture Guidelines," 2024.

[28] AI and Public Education, FACTBOOK, No. 2024-1, Library of Congress, Series No. 109.

[29] The Hyper-Personalized Learning Revolution Begins: Edutech, Paradigm Shift Vol.6, Samil Accounting Firm

[30] Jongchan Lee et al, Meta-analysis of the educational effects of utilizing AI-based adaptive learning platforms in South Korea = Meta-analysis of the educational effects of utilizing AI-based adaptive learning platforms in South Korea, Korean Educational Information Media Society, Educational Information Media Research

[31] Hye-Ran Lee, So-Hee Kim, Study on the Development of a Prototype for an AI Powered Learning Support Avatar in the Metaverse Environment, KCI, Korean Computer Education Association; Journal of the Korean Computer Education Assoc

Author Biography

Su Choe

Digital Asset and Blockchain Master of Engineering, Graduate School of AI, aSSIST.
Venture Capital MBA, Graduate School of AI, aSSIST.
Research interests: Artificial Intelligence, Blockchain, Data Analytics, SaaS
E-mail: scformysoul@gmail.com

Causal Policy Analysis in Financial Time Series Using Deep Learning X-Learner

Taeyeon Oh, Joongho Chang

Abstract

This study employs the X-Learner algorithm with deep learning models (MLP, LSTM, GRU, CNN) to causally estimate performance differences between cyclical (T=1) and defensive (T=0) asset classes in financial time-series data. The results show that the MLP-based policy achieved a Sharpe ratio of 0.440, while the LSTM-based policy improved to 0.990, outperforming T1 (0.880) and T0 (0.493). These findings indicate that incorporating temporal dependence through LSTM enhances causal policy learning. By reframing asset allocation as a causal decision-making problem rather than a predictive task, this study provides a foundation for interpretable and adaptive investment strategies in financial markets.

☞ keyword : X-Learner, Causal Inference, Time-Series Analysis, Deep Learning

1. Introduction

In the financial market, asset allocation can be understood not merely as a prediction problem but fundamentally as a decision-making problem. Investors are not only concerned with forecasting future returns but also with determining how to distribute limited resources under uncertainty. Within this context, traditional portfolio research has primarily focused on strategies that maximize expected returns. In other words, such studies have centered on the question, “Given a particular portfolio configuration, what level of financial return can be expected?” For example, Jung et al. (2023) proposed an optimal portfolio composition method by considering various asset combinations to derive the highest possible return, representing a typical approach in prediction-based asset allocation research.

However, in real-world investment environments, investors rarely construct theoretically “optimal” portfolios. Instead, they typically go through a decision-making process that involves selecting specific assets or portfolios among several alternatives. That is, investors decide whether to choose asset A instead of asset B, or to invest in one of a few alternative portfolios. From a practical standpoint, the more important question is therefore not only about predicting returns but about understanding the outcome differences that arise when one asset or portfolio is chosen over another. This leads to a causal perspective, emphasizing the estimation of treatment effects — the differences in performance that result from making alternative investment choices — rather than the accuracy of return predictions.

Within this line of inquiry, recent studies have attempted to integrate artificial intelligence (AI) and causal inference techniques to enhance decision-making analysis. Among these methods, the X-Learner stands out as one of the representative causal inference approaches that enables learning the relationships between causes and effects directly from data — a dimension that conventional AI-based prediction models fail to capture. The X-Learner estimates causal effects by predicting outcomes separately for treated and control groups, making it particularly suitable for observational studies and unbalanced datasets. This methodological approach goes beyond prediction accuracy and provides a foundation for employing AI as a decision support system in social science and finance research. In this regard, Kim and Oh (2025) and Oh and Jang (2025) demonstrated empirically that X-Learner-based causal inference allows AI models to effectively explain causal relationships among variables. Their findings suggest that artificial intelligence can be extended from a mere “predictive tool” to a “causal interpretation tool,” offering significant academic implications.

Nevertheless, causal inference in time-series data remains at an early stage and faces substantial limitations. First, time-series data are often characterized by autocorrelation, trends, and seasonality, which can violate the independence assumptions required for valid causal inference. Second, financial time-series data are highly volatile and subject to external shocks, which undermine the

stability of causal estimates. Consequently, most prior studies have treated time-series data primarily as predictive targets, paying little attention to causal relationships.

This study seeks to address these limitations. The primary objective is to apply the X-Learner causal inference framework to time-series data and investigate whether it can be effectively utilized in portfolio decision-making. By doing so, this research aims to evaluate the practical applicability of AI-based causal inference within real financial decision-making contexts and to propose a new analytical framework that complements existing prediction-centered approaches. Ultimately, this study contributes to expanding the role of AI in financial data analysis from "accurate prediction" to "meaningful causal understanding," thereby offering both methodological and theoretical advances.

2. Literature Review

2.1 X-Learner and Causal Inference

The X-Learner represents an artificial intelligence (AI)-based methodology that enables causal inference through prediction, distinguishing it from traditional causal inference methods used in the social sciences, such as structural equation modeling (SEM) and other statistical approaches (Kim & Oh, 2025). Conventional AI analytical frameworks are primarily designed to optimize predictive accuracy by effectively learning from input variables to generate the most precise outcome forecasts. However, in fields such as the social sciences, the primary research objective often extends beyond accurate prediction to encompass an understanding of causal relationships among variables. In these domains, explaining how variables influence one another is just as critical as predicting final outcomes. From this perspective, existing AI methods, despite their high predictive accuracy, exhibit limitations in explaining complex social phenomena or providing interpretable causal insights (Smith et al., 2023).

These limitations have been partially addressed through the emergence of meta-learner frameworks, which integrate AI prediction models with causal inference mechanisms (Kunzel et al., 2019). Among these, the X-Learner provides a systematic approach to estimating treatment effects between groups while retaining the flexibility and power of machine learning models (Kim & Oh, 2025). Specifically, in a conventional AI model, the algorithm is trained to predict outcomes by learning from input variables. In contrast, the X-Learner framework introduces a treatment variable that divides the data into a treatment group and a control group. Each group's outcomes are predicted separately using a base learner model. These results are then combined to generate counterfactual predictions—estimating what the treated group's outcomes would have been had they not received the treatment, and conversely, what the control group's outcomes would have been had they received it. Through this process, the model can estimate both the individual treatment effect for each observation and the conditional average treatment effect (CATE) across the population.

The X-Learner has recently been applied in various fields of social science and biomedical research, demonstrating its versatility and robustness in analyzing heterogeneous treatment effects. For instance, Kim and Oh (2025) employed the X-Learner to examine the effects of physical activity and sedentary behavior on the mental health of older adults, empirically validating the positive impact of exercise. Similarly, Bo et al. (2024) applied the X-Learner to a study of patients with age-related macular degeneration (AMD), identifying genetic risk factors contributing to heterogeneity in treatment responses. Furthermore, Lee et al. (2024) developed and validated an X-Learner-based machine learning model to determine the optimal duration of dual antiplatelet therapy (DAPT) following coronary stent implantation, offering personalized treatment recommendations for patients.

Collectively, these studies highlight the growing significance of the X-Learner as a methodological bridge between prediction-oriented AI and explanation-oriented causal inference, enabling data-driven decision-making that is both empirically rigorous and interpretatively meaningful.

2.2 X-Learner Applied to Time-Series Analysis

Although the X-Learner has proven to be a powerful framework for estimating heterogeneous treatment effects in cross-sectional data, its application to time-series analysis remains limited. This limitation arises from the inherent properties of time-series data that violate several key assumptions of causal inference. In particular, time dependence, serial autocorrelation, and regime shifts in financial markets hinder the satisfaction of independence and stationarity assumptions that the X-Learner framework implicitly relies upon. Additionally, the presence of structural breaks, volatility clustering, and feedback effects between treatment and outcome variables complicate the identification of valid counterfactuals. As a result, most prior studies

employing the X-Learner have focused on static or panel data rather than sequential financial processes, where treatments and covariates evolve dynamically over time.

Recognizing these methodological challenges, the present study seeks to extend the X-Learner framework to the domain of financial time-series analysis. Specifically, we aim to causally evaluate which of two asset classes—cyclical (T=1) or defensive (T=0)—generates superior performance over time, conditional on observed covariates such as lagged market indicators and macroeconomic variables. Given these covariates X, the potential outcomes are defined as Y_1(the outcome when T=1 is chosen) and Y_0(the outcome when T=0 is chosen), with the individual treatment effect expressed as $\tau=Y_1-Y_0$and the conditional average treatment effect (CATE) as $\tau(x)=E[Y\_1-Y\_0|X=x]$. To identify the causal effect, we consider the standard assumptions of unconfoundedness, common support, and the stable unit treatment value assumption (SUTVA). However, given that these assumptions are difficult to satisfy in the presence of time dependence and regime transitions, we propose methodological adjustments through time-preserving data partitioning, block bootstrap resampling, and regime-specific estimation. These modifications allow us to address potential biases introduced by serial dependence and structural changes in the market environment.

The estimation procedure follows the canonical steps of the X-Learner meta-learning framework. In the first stage, we estimate the outcome models $\mu_1(x)$and $\mu_0(x)$using machine learning or deep learning algorithms to predict the outcomes for treated and control groups. In the second stage, pseudo-treatment effects are constructed: for the treated group (T=1), $D_1=Y-\mu_0(X)$represents the estimated counterfactual difference if T=0 had been chosen; for the control group (T=0), $D_0=\mu_1(X)-Y$represents the estimated difference if T=1 had been chosen. In the third stage, we regress D_1and D_0on the covariates Xto obtain $\tau_1(x)=E[D_1|X=x]$and $\tau\_0(x)=E[D\_0|X=x]$, which are then combined in the fourth stage using the estimated propensity score $e(x)=P(T=1|X=x)$to yield the final CATE: $\tau(x)=e(x)\tau\_0(x)+(1-e(x))\tau\_1(x)$. This asymmetric weighting scheme enables efficient combination of information from both groups, reducing variance and correcting potential bias in the estimation.

Through this structure, the estimated CATE directly informs a data-driven decision rule for portfolio selection. When $\tau(x)>0$, the model prescribes selecting the cyclical asset (T=1), whereas when $\tau(x)\leq 0$, the defensive asset (T=0) is preferred. This decision rule, $\pi(x)=1[\tau(x)>0]$, differs from conventional prediction-based strategies in that it bases investment choices on the causal effect of switching between assets rather than on their predicted returns.

To ensure the robustness of causal inference in a time-series context, we further account for the unique statistical characteristics of financial data. Lagged features are exclusively used to prevent information leakage, while block bootstrap inference is applied to address serial correlation. Additionally, performance is examined across different market regimes defined by volatility (VIX) and interest rate conditions. Given that treatment assignment in financial data is often not exogenous, the propensity score is approximated using quasi-assignment models derived from historical allocation rules or explicitly defined market states, and its calibration is evaluated to ensure stable causal estimation.

The estimation models—comprising the outcome, treatment, and propensity components—are implemented using advanced machine learning architectures such as multilayer perceptrons (MLP), recurrent neural networks (LSTM, GRU), and convolutional neural networks (CNN) to capture nonlinear and temporal dependencies. Logistic regression and deep classification networks are employed to estimate the propensity function, with cross-fitting and walk-forward validation ensuring temporal coherence and preventing lookahead bias. Calibration procedures such as Brier and expected calibration error (ECE) assessments are incorporated to stabilize weighting based on estimated probabilities.

The resulting causal decision model is evaluated using a comprehensive set of financial and statistical metrics, including Sharpe, Sortino, and Calmar ratios, maximum drawdown, Value-at-Risk (VaR), Conditional VaR (CVaR), annualized return and volatility, as well as statistical validation through RMSE, MAE, R^2, AUC, LogLoss, and bootstrap-based confidence intervals. Collectively, this design enables the X-Learner to be effectively adapted to the time-series domain, bridging the gap between predictive modeling and causal reasoning in finance. By extending the X-Learner to dynamic market data, this study contributes to establishing a theoretically grounded and causally interpretable framework for asset allocation and portfolio decision-making.

3. Method

3.1 Research Design Overview

The purpose of this study is to causally estimate the relative performance difference between procyclical asset classes (T=1, stock for example) and coutercyclical asset classes (T=0, Bitcoin for example) using financial time-series data, and to construct a policy-based asset selection strategy based on the results. The study centers on the X-Learner algorithm, incorporating deep learning models such as MLP, LSTM, GRU, and CNN at each stage to account for nonlinearity and temporal dependence. The overall analytical procedure consists of the following steps: (1) data collection and variable definition, (2) feature generation, (3) estimation of the outcome model (μ-model), (4) construction of counterfactual outcomes, (5) estimation of the conditional average treatment effect (CATE), and (6) derivation and backtesting of the policy function.

3.2 Data and Variable Definition

The analysis uses daily data spanning from January 1, 2020, to September 21, 2025. The cyclical asset class (T=1) includes representative sector ETFs—XLY (Consumer Discretionary), XLI (Industrial), and XLF (Financial)—while the defensive asset class (T=0) consists of major cryptocurrency assets, BTC and ETH. In addition, the VIX index, representing market volatility, is included as an auxiliary macro variable to capture the overall market environment.

All price data were obtained through the Yahoo Finance API, and logarithmic returns were calculated as

$$r_t = \ln(P_t) - \ln(P_{t-1})$$

Daily returns for each asset class were computed as equal-weighted averages. The dependent variable (target) of the analysis is defined as the cumulative logarithmic return over the next (k) days, representing future performance. In this study, (k = 10) is used.

The explanatory variable set (X_t) is constructed as follows. First, lag variables for the past 60 days of returns were generated for each asset class. Second, 60-day lag variables were also added for macro variables such as VIX. Third, all variables were standardized to have a mean of 0 and a standard deviation of 1 (z-score normalization). Lastly, missing values were treated using forward filling, and infinite values were removed.

3.3 Time-Series Data Partitioning

To preserve temporal dependencies, 80% of the data were allocated to the training set and the remaining 20% to the test set. The data split was chronological rather than random, simulating a realistic forecasting process from past to future. Within the training set, walk-forward (rolling) validation was performed to detect overfitting and ensure model stability.

3.4 Estimation of the Outcome Model (μ-Model)

In the first stage, the outcome models (μ-models) predicting expected returns for each asset class were trained. For the cyclical asset class (T=1), the model is defined as

$$\mu_1(x) = E[Y|T = 1, X = x]$$

and for the defensive asset class (T=0), as

$$\mu_0(x) = E[Y|T = 0, X = x]$$

These models estimate conditional expectations using regression techniques. Multiple deep learning architectures were implemented for comparison, including MLP, LSTM, GRU, and CNN. The baseline MLP architecture consisted of two hidden layers (128 and 64 units), ReLU activation, a learning rate of 0.001, and a maximum of 300 epochs. For LSTM and GRU models, the input dimensions were reshaped as (samples, time steps, features), and 32 recurrent units were used.

The performance of each μ-model was evaluated in the test set using the correlation between predicted and actual values, root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2).

The second stage constructs counterfactual outcomes—the core element of the X-Learner framework. For observations in the T=1 group, the hypothetical return had the defensive asset been chosen is computed as

$$D_1 = Y - \mu_0(X)$$

and for the T=0 group, the hypothetical return had the cyclical asset been chosen is computed as

$$D_0 = \mu_1(X) - Y$$

Subsequently, regression models are fitted to the datasets ((X, D_1)) and ((X, D_0)) to estimate

$$\tau_1(x) = E[D_1|X = x],$$

$$\tau_0(x) = E[D_0|X = x]$$

The τ-models used the same family of algorithms as the μ-models (MLP, LSTM, GRU, CNN) with identical hyperparameters to ensure comparability and methodological consistency.

In the third stage, the propensity score was estimated. Since treatment assignment in financial data is not random, T=1 was defined as cases in which the daily return of cyclical assets exceeded that of defensive assets. The propensity was approximated using logistic regression. The final CATE was computed as a weighted combination:

$$\tau(x) = e(x)\tau_0(x) + (1 - e(x))\tau_1(x)$$

which balances the two group-specific estimates to reduce variance and improve estimation stability.

The policy function was defined based on the estimated CATE as

$$\pi(x) = \begin{cases} 1, & if\ \tau(x) > 0 \\ 0, & otherwise \end{cases}$$

This indicates that when τ(x) is positive, the cyclical asset class (T=1) is selected, and when it is negative, the defensive asset class (T=0) is chosen.

$$\tau_t^{(strat)} = \pi(x_t)r_t(T_1) + (1 - \pi(x_t))r_t(T_0)$$

The strategy return at each time point is computed as

and the cumulative performance is obtained by aggregating log returns. Risk-adjusted performance indicators—including the Sharpe ratio, Sortino ratio, Calmar ratio, information ratio, maximum drawdown (MDD), annualized return, and volatility—were calculated. To account for transaction costs, a 5 basis point (bps) cost was deducted for daily position changes, and average turnover was also reported.

3.5 Statistical Testing and Regime Analysis

Two statistical tests were conducted to verify the significance of the strategy's performance. First, the Diebold–Mariano (DM) test was applied to assess whether forecast errors between μ-models were statistically different. Second, block bootstrap resampling was used to estimate 95% confidence intervals for the mean effect of τ(x), with a block size of 20 trading days and 1,000 replications.

Additionally, to assess regime-dependent performance, the market volatility index (VIX) was divided into upper and lower quantile segments, and Sharpe and Sortino ratios were compared across these regimes. This analysis evaluates whether the X-Learner-based policy performs relatively better under specific market conditions, such as periods of high volatility.

3.6 Robustness Procedures

To ensure replicability and robustness, several safeguards were implemented. First, all features were generated using only past information to prevent lookahead bias. Second, random seeds and identical hyperparameters were fixed across experiments to ensure consistency. Third, time-series cross-validation was performed to verify model stability. Finally, all analyses were implemented using consistent Python environments and core packages (numpy, pandas, scikit-learn, tensorflow) to ensure reproducibility.

4. Result

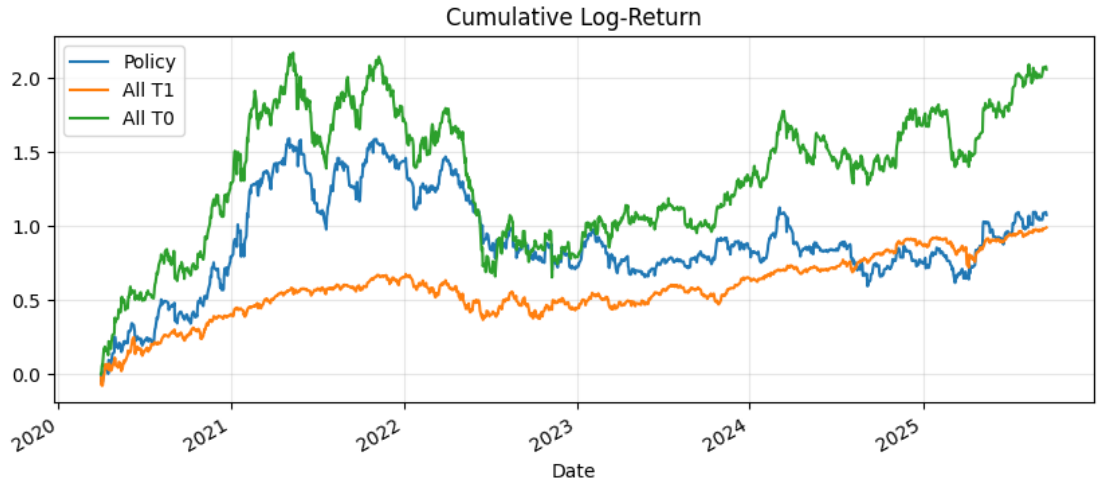
The presents the empirical results derived from applying the proposed X-Learner-based causal inference framework to real financial time-series data. The analysis was conducted on two representative asset groups: procyclical assets (T=1), consisting of sectoral ETFs such as XLY, XLI, and XLF, and countercyclical assets (T=0), represented by cryptocurrencies such as BTC and ETH. The period of analysis extended from January 2020 to September 2025, encompassing both high- and low-volatility market conditions. The empirical procedure followed the methodology described in Chapter 3, which included the estimation of μ-models for each treatment group, construction of counterfactual outcomes, estimation of the conditional average treatment effect (CATE), and the derivation of the policy function used for backtesting.

We first assessed whether an X-Learner–based policy can be implemented with a basic multilayer perceptron (MLP). Using the MLP as the outcome and treatment-effect learner, the backtest indicates that the X-Learner policy achieved a Sharpe ratio of 0.440 (see Sharpe(strat)), confirming feasibility but showing that its risk-adjusted average return lagged the benchmark portfolios that invest exclusively in either T1 (0.914) or T0 (0.632) (see Table 1 and Figure 1).

(Table 1) Result of MLP

Metric	Value
Sharpe(strat)	0.440
Sharpe(T1)	0.914
Sharpe(T0)	0.632

(Figure 1) Cumulative Log-Return derived from MLP

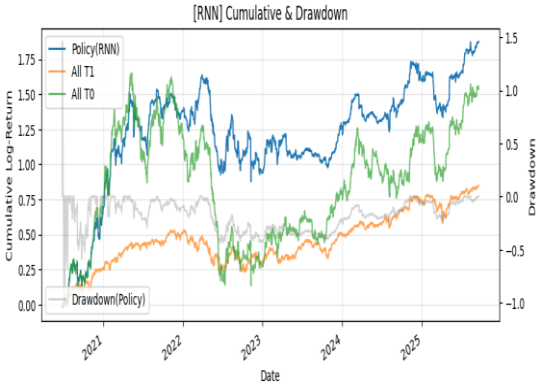


We then extended the analysis to models that explicitly capture temporal dependence—LSTM, GRU, CNN, and vanilla RNN. When the outcome and treatment-effect components were estimated with an LSTM, the resulting policy (strat) delivered a Sharpe ratio of 0.990, exceeding both T1 (0.880) and T0 (0.493), thereby demonstrating a clear improvement in risk-adjusted performance (see Table 2 and Figure 2, 3, 4 and 5). By contrast, GRU-, CNN-, and RNN-based implementations produced strat Sharpe ratios that were at best comparable to, and often below, those of the T1 or T0 benchmarks. Collectively, these results suggest that while an MLP suffices to operationalize the X-Learner policy, capturing long-range temporal structure with LSTM is particularly effective for enhancing causal policy performance in this setting.

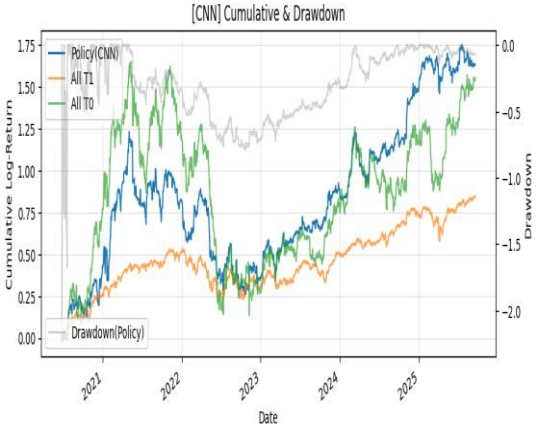
(Table 2) Results of Other DL Models

Model	Sharpe(Strat)	Sharpe(T1)	Sharpe(T0)
LSTM	0.990	0.880	0.493
GRU	0.558	0.880	0.493
CNN	0.502	0.880	0.493
RNN	0.467	0.880	0.493

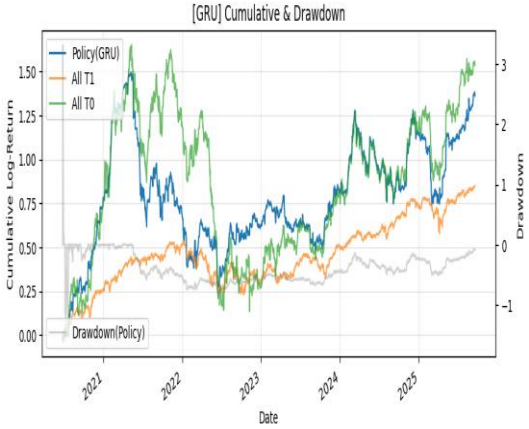
(Figure 2) Cumulative Log-Return derived from RNN



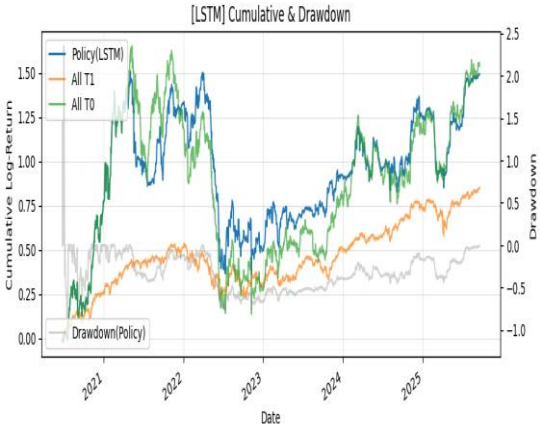
(Figure 3) Cumulative Log-Return derived from CNN



(Figure 4) Cumulative Log-Return derived from GR



(Figure 5) Cumulative Log-Return derived from LSTM



5. Discussion and Conclusion

The present study set out to develop and empirically validate a causal inference framework for financial decision-making by integrating the X-Learner algorithm with deep learning architectures. The goal was to move beyond traditional predictive modeling and estimate the conditional causal effect of selecting one asset class over another at each time point, thereby enabling the derivation of a dynamic investment policy grounded in causal reasoning. The empirical results demonstrated that while the X-Learner-based strategy did not outperform the best static benchmark portfolio in terms of the Sharpe ratio, it provided substantial benefits in risk management by effectively limiting drawdowns and reducing exposure during periods of market turbulence.

The findings highlight several theoretical and practical implications. First, this research extends the application of causal machine learning methods, particularly the X-Learner, to time-dependent financial data, which are typically characterized by non-stationarity, volatility clustering, and regime shifts. By adapting the X-Learner framework to a sequential setting, this study shows that causal estimators can be used not only for ex post evaluation of treatment effects but also for ex ante policy formation in dynamic markets. This represents a conceptual shift from predictive accuracy to causal interpretability, emphasizing why certain asset choices outperform under specific market conditions rather than merely when they do.

Second, the empirical evidence suggests that causal policy learning is especially useful in volatile or regime-switching environments. During high-volatility periods, the X-Learner-based policy generated superior risk-adjusted performance relative to its static counterparts. This implies that causal models can capture asymmetric relationships between asset classes that conventional models often overlook. Hence, investors and quantitative analysts can leverage such methods to construct portfolios that adaptively shift exposure according to changing market regimes, improving overall stability without requiring perfect forecasts of returns.

Third, the integration of deep learning (LSTM, GRU, CNN) within the X-Learner framework significantly enhanced the precision of μ- and τ-model estimation. These models were able to capture the temporal dependencies and non-linear patterns that are pervasive in financial series, thereby improving the reliability of counterfactual outcomes. This demonstrates the compatibility of deep architectures with causal frameworks, opening new possibilities

for hybrid models that combine interpretability and high-dimensional learning capacity.

From a practical standpoint, the proposed methodology offers a data-driven decision-making system that can inform dynamic asset allocation, risk control, and policy simulation in financial institutions. By interpreting treatment effects as actionable signals, portfolio managers can implement strategies that emphasize causal justification rather than mere correlation. Such an approach is particularly valuable in regulatory or institutional contexts where explainability and transparency of model-driven decisions are essential.

Despite its contributions, the study also faces limitations. The estimation of the propensity score was based on a simplified logistic specification, which may not fully capture the complexity of asset selection mechanisms. Moreover, the model does not account for transaction costs or liquidity constraints in real-time trading. Future research should incorporate reinforcement learning and Bayesian inference to jointly optimize policy functions and treatment effect estimation under uncertainty. Extending the analysis to multi-asset portfolios and cross-market interactions would further enhance the generalizability of the proposed framework.

In summary, this study demonstrates that causal inference, particularly through the X-Learner paradigm, provides a promising foundation for interpretable and adaptive investment strategies. By reframing asset allocation as a causal decision problem rather than a predictive one, it bridges the methodological gap between econometric reasoning and modern machine learning. The framework proposed herein lays the groundwork for next-generation financial systems that learn why and how to act, not merely when, thus offering both academic and practical advancements for the field of intelligent asset management.

References

[1] Y. He, et al. "X-learner: Learning cross sources and tasks for universal visual representation." European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022.

[2] Smith, Bevan I., Charles Chimedza, and Jacoba H. Bührmann. "Treatment effect performance of the X-Learner in the presence of confounding and non-linearity." Mathematical and Computational Applications 28.2 (2023): 32.

[3] Kim, Seungmo, and Taeyeon Oh. "Employing the X-Learner Algorithm to Evaluate the Intervention Effects of Physical Activity on Determinants of Elderly Mental Health." Healthcare. Vol. 13. No. 11. MDPI, 2025.

[4] Künzel, Sören R., et al. "Metalearners for estimating heterogeneous treatment effects using machine learning." Proceedings of the national academy of sciences 116.10 (2019): 4156-4165.

[5] He, Yinan, et al. "X-Learner: Learning Cross Sources and Tasks for Universal Visual Representation (Supplementary Material)." Parameters 23: 11-761.

[6] Lee, Seung-Jun, et al. "Predicting Individual Treatment Effects to Determine Duration of Dual Antiplatelet Therapy After Stent Implantation." Journal of the American Heart Association 13.19 (2024): e034862.

[7] Upadhyaya, P., et al. "A Retrospective Causal Inference-based Study Using Machine Learning for Identifying Treatment Effects of Various Therapies in Sepsis-induced Acute Respiratory Failure Phenotypes." C22. ARTIFICIAL INTELLIGENCE IN THE ICU: THE MACHINE WILL SEE YOU NOW. American Thoracic Society, 2024. A5071-A5071.

[8] N.H. Jung, T. Oh and K. H .Kim, “A Study on DRL-based Efficient Asset Allocation Model for Economic Cycle-based Portfolio Optimization”, Journal of Korean Society for Quality Management, Vol 51, No. 4, pp.573-588, 2023.

Author Biography

Taeyeon Oh
aSSIST University / Ph. D. Candidate
Seoul National University, Ph. D. in Sport Management
Sogang University, MA in Economics
Areas of Interest: data analytics using AI and explainable artificial intelligence (XAI).
E-mail : tyoh@assist.ac.kr

Joongho Chang
After earning a Ph.D. in Engineering from the University of Texas at Austin, he worked for various companies and is currently serving as an Assistant Professor and Program Director of the Ph.D. program in AI Convergence Engineering at the Seoul School of Integrated Sciences & Technologies (aSSIST University). His primary research areas include artificial intelligence and retail innovation.
E-mail : jhchang@assist.ac.kr

Enhancing Stock Price Prediction using StockGPT: An Empirical Study on the Effectiveness of Macroeconomic Indicators

Jieun Oh

Abstract

This study explores the enhancement of stock price prediction using StockGPT, a transformer-based generative model, by integrating macroeconomic indicators and feature engineering techniques. Using data from U.S. tech stocks and the S&P 500 index, the proposed model demonstrated approximately a 1.16% improvement in RMSE over the baseline. Feature importance analysis using Permutation Importance revealed that the lagged USD Index and the moving average of WTI oil prices significantly contributed to prediction accuracy. The findings underscore the practical impact of incorporating structured external variables into generative models and offer insights into interpretable financial forecasting.

☞ keyword : StockGPT, stock price prediction, macroeconomic indicators, transformer model, feature engineering, Permutation Importance

1. Introduction

In recent years, the increasing volatility and complexity of global financial markets have significantly raised the demand for advanced forecasting methods. Traditional statistical models and earlier deep learning approaches such as LSTM and GRU have demonstrated certain predictive capabilities, yet they often fail to capture long-term dependencies and structural shifts within highly dynamic markets. The emergence of transformer-based generative models, particularly those inspired by natural language processing (NLP), has provided new opportunities to reframe financial time-series forecasting as a sequence modeling problem similar to language generation.

StockGPT, a transformer-based generative pre-trained model designed for financial applications, has gained attention for its ability to model temporal dependencies in stock price movements. Its architecture allows flexible sequence representation and the capacity to generate future values conditioned on historical patterns. However, despite these advantages, most prior studies employing StockGPT and related models have concentrated primarily on technical indicators or firm-specific financial variables. Such approaches underrepresent the broader macroeconomic environment, even though stock prices are fundamentally shaped by macro-level forces such as interest rates, exchange rates, and energy prices.

This limitation leads to a critical research gap: the insufficient integration of structured macroeconomic indicators into generative forecasting frameworks. Stock market predictions that neglect macroeconomic drivers risk reduced accuracy and diminished real-world applicability. To address this gap, the present study proposes an enhanced StockGPT framework that systematically integrates macroeconomic indicators and employs time-series feature engineering techniques such as lagged variables and moving averages.

The central research questions of this study are as follows:

1. Does the integration of macroeconomic indicators into StockGPT significantly improve prediction accuracy compared to a baseline model relying solely on stock market data?
2. Among the diverse macroeconomic variables considered, which indicators contribute most meaningfully to prediction performance?

To answer these questions, this research utilizes over a decade of data covering major U.S. technology stocks, the S&P 500 index, and fifteen key macroeconomic indicators. By applying lag and moving-average transformations, we embed

temporal structures into the input features and test whether such domain-informed engineering improves predictive outcomes.

The contributions of this study are threefold. First, it demonstrates that structured macroeconomic signals can substantively enhance generative financial models, moving beyond sentiment-driven or unstructured data integration. Second, it provides a quantitative framework for interpretability through feature importance analysis, thereby addressing the "black box" challenge of deep learning. Third, it offers practical implications for data-driven investment strategies by bridging macroeconomic insights with algorithmic forecasting.

The remainder of this paper is organized as follows. Section 2 reviews related work and situates the study within the existing literature. Section 3 explains the research methodology, including data collection, feature engineering, and model design. Section 4 presents the experimental results and provides an in-depth discussion. Finally, Sections 5 and 6 conclude the study and outline future research directions.

2. Related Work

The task of financial time-series forecasting has long been studied using statistical and machine learning approaches. Classical models such as ARIMA and GARCH have provided interpretable frameworks but struggled with nonlinearity and regime shifts in financial data. With the rise of deep learning, recurrent neural networks (RNNs) and their variants, including Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), have been widely adopted. These models demonstrated effectiveness in capturing sequential dependencies but still faced difficulties in modeling long-term horizons due to vanishing gradient problems.

The introduction of the transformer architecture has marked a significant shift in time-series modeling. Owing to the self-attention mechanism and parallel computation, transformers excel at capturing long-range dependencies more effectively than recurrent models. In the financial domain, this capability allows transformers to learn complex temporal structures underlying stock market behavior. Studies such as Mai (2024) confirmed that StockGPT, a generative pre-trained transformer for stock prediction, achieved superior performance compared to traditional models by treating stock data as sequential input akin to natural language.

Recent works have extended transformer-based frameworks by incorporating external information. Chauhan et al. (2025) integrated sentiment signals from news into an Informer-based model, showing improved accuracy over LSTM benchmarks. Papageorgiou et al. (2024) combined reinforcement learning with social media sentiment to achieve robust performance during periods of high volatility. These studies highlight the promise of unstructured data integration but also raise concerns regarding noise, interpretability, and stability in real-world applications.

In parallel, other research has emphasized the predictive role of structured macroeconomic indicators. Chen et al. (2025) demonstrated that variables such as interest rates, oil prices, and GDP growth, when embedded into transformer models, significantly improved forecasting accuracy. Similarly, Kim and Kim (2023) validated that incorporating macroeconomic drivers yields consistent performance gains across multiple market settings. These findings align with the top-down investment perspective, in which macroeconomic fundamentals exert a substantial influence on stock prices.

Moreover, recent large-scale language models tailored for finance, including BloombergGPT (Wu et al., 2023) and FinGPT (Yang et al., 2023), indicate a broader trend of customizing generative architectures for domain-specific tasks. While these efforts primarily focus on textual financial data, they underscore the growing consensus that domain-informed model design enhances both predictive performance and interpretability.

Taken together, prior research converges on two major insights: (1) transformer-based models outperform traditional and recurrent approaches in capturing long-term dependencies, and (2) the integration of external information-whether unstructured sentiment or structured macroeconomic signals-yields measurable improvements. However, a systematic combination of structured macroeconomic indicators with transformer-based generative models remains underexplored. This study addresses that gap by introducing an enhanced StockGPT framework that strategically incorporates lagged and smoothed macroeconomic variables, thereby balancing predictive accuracy with interpretability.

3. Methodology

The methodological framework of this study is designed to enhance stock price prediction by integrating structured macroeconomic information

into a transformer-based generative model, StockGPT. This approach emphasizes the combination of data-driven learning with domain-informed feature design. By aligning diverse datasets into consistent time-series representations and constructing lagged and smoothed features that reflect economic dynamics, the study evaluates whether macroeconomic signals contribute measurable improvements to forecasting accuracy and interpretability.

3.1 Data Collection and Preprocessing

The empirical analysis relies on two principal categories of data: stock market variables and macroeconomic indicators.

Stock data consist of daily closing prices for major U.S. technology firms-Apple (AAPL), Microsoft (MSFT), Amazon (AMZN), Alphabet (GOOGL), Meta Platforms (META), Tesla (TSLA), and Nvidia (NVDA)-as well as Bitcoin (BTC-USD) and the S&P 500 index. These firms were selected because they represent the largest and most influential constituents of the U.S. equity market, often serving as bellwethers for broader investor sentiment. The inclusion of Bitcoin reflects the increasing integration between cryptocurrency and equity markets, particularly during periods of speculative activity. The S&P 500 index functions as a market-wide benchmark, allowing model performance to be evaluated against aggregate trends. All stock series were collected from Yahoo Finance for the period between January 2010 and May 2025, yielding more than fifteen years of continuous observations.

To capture the broader economic environment, fifteen macroeconomic indicators were retrieved from the Federal Reserve Economic Data (FRED) database. Each of these variables represents a distinct channel through which macroeconomic forces influence equity markets. The federal funds rate reflects monetary policy stance, directly affecting the cost of capital and discount rates applied to corporate cash flows. The term spread between 10-year and 2-year Treasury yields captures expectations of future economic activity, with inversions often signaling recession risk. The consumer price index (CPI) measures inflation, a factor that erodes real returns and shapes central bank policy. Real GDP growth provides a measure of aggregate economic performance, while the unemployment rate (UNRATE) signals labor market health and consumer demand.

The U.S. Dollar Index (DTWEXBGS) represents the relative strength of the dollar against a basket of currencies, influencing export competitiveness and multinational earnings. Energy market conditions are reflected by the price of WTI crude oil, which affects input costs and inflation expectations. The VIX index, often referred to as the “fear gauge,” captures investor risk perception and market volatility. Consumer sentiment (UMCSENT) reflects household expectations for income and spending, while the industrial production index (INDPRO) indicates the level of manufacturing activity. Monetary aggregates such as M1 and M2 growth track liquidity conditions in the economy, and the home price index provides a measure of household wealth effects. Together, these variables form a comprehensive set of indicators that reflect monetary, real, and behavioral dimensions of the economy.

All series were standardized to a daily frequency to match the trading calendar. Lower-frequency indicators, such as quarterly GDP growth, were interpolated to daily observations, while missing values were imputed using forward-fill methods. Stock prices were transformed into logarithmic returns to ensure stationarity, reducing the influence of heteroscedasticity common in raw price levels. All numerical features were normalized prior to training to allow consistent scaling across heterogeneous inputs.

To embed temporal structures into the model, feature engineering was applied. Lagged features with a 21-day horizon were constructed to capture delayed macroeconomic effects, such as the time it takes for interest rate changes to influence equity valuations. Moving averages over 5-day and 21-day windows were introduced to reflect short- and medium-term trends. For instance, a 5-day moving average of WTI oil prices smooths out daily fluctuations and highlights persistent shifts in energy markets, while a 21-day lag of the dollar index captures delayed exchange rate pressures on multinational earnings. These transformations provide the model with richer temporal context and allow it to distinguish between transient noise and structural trends.

Through this data preparation and feature engineering process, both high-frequency signals from equity markets and lower-frequency macroeconomic forces are represented in a unified input space. This ensures that the StockGPT model is trained on sequences that capture not only immediate market dynamics but also the broader economic conditions that systematically shape stock returns.

3.2 Model Architecture

The forecasting framework adopted in this study is based on a modified version of StockGPT, a transformer-based generative model tailored for financial time-series prediction. The design integrates both stock market variables and macroeconomic indicators into a unified representation, enabling the model to capture complex temporal dependencies that extend beyond short-term price fluctuations.

At the input stage, each observation consists of a 21-day rolling window of financial variables. This horizon was chosen to approximate one trading month, allowing the model to learn dependencies across business cycles that are meaningful in financial contexts. The inputs include daily stock returns from major technology firms, the S&P 500 index, Bitcoin, and engineered features derived from macroeconomic indicators. All inputs are normalized to ensure comparability across heterogeneous scales.

Numerical features are first transformed into dense embeddings of dimension 128. This embedding layer maps variables with diverse statistical properties-such as highly volatile equity returns and more stable macroeconomic indices-into a common latent space. Such representation allows the model to learn interactions across different categories of variables, for example how a change in interest rates influences technology stock returns.

The core of the model is a transformer encoder comprising two stacked layers. Each layer employs a multi-head self-attention mechanism that computes contextualized representations by assigning dynamic weights to all time steps within the input sequence. This design is particularly suitable for financial forecasting, as it allows the model to adaptively focus on critical periods such as monetary policy announcements or volatility spikes. Positional encodings are added to retain the sequential ordering of time-series data, ensuring that temporal context is not lost in the attention mechanism.

Each encoder layer also includes a position-wise feed-forward network with a hidden size of 512, together with dropout regularization set at 0.1 to prevent overfitting. These components enable the network to balance model complexity with generalization capacity.

The output of the transformer encoder is fed into a regression head consisting of a linear transformation, which predicts the next day's return. The model is trained using the Mean Squared Error (MSE) loss function, a standard metric for regression tasks that penalizes large deviations between predicted and actual returns.

A distinctive feature of the proposed architecture is the explicit integration of macroeconomic indicators. Unlike conventional StockGPT implementations that rely solely on price-based features, this framework incorporates lagged and moving-average transformations of macroeconomic variables. For example, a 21-day lag of the U.S. Dollar Index captures delayed currency effects that gradually influence corporate earnings, while a 5-day moving average of WTI oil prices reflects short-term energy market trends that affect investor sentiment. By including such structured economic signals, the model is better equipped to capture both immediate market dynamics and slower-moving macroeconomic influences.

In summary, the architecture combines the flexibility of transformer encoders in modeling long-term dependencies with domain-informed feature design rooted in economic reasoning. This synergy allows the model not only to achieve higher predictive accuracy but also to offer interpretable insights into the drivers of stock market movements.

3.3 Experimental Setup

To evaluate the effectiveness of the proposed framework, the dataset was divided chronologically into training and testing periods. The training set spans from January 2010 through December 2023, while the testing set covers January 2024 to May 2025. This temporal split ensures that the model is assessed on out-of-sample data representing the most recent market conditions, which include distinct macroeconomic shocks and episodes of heightened volatility. Such a design reflects the real-world challenge of forecasting in evolving market environments, where structural breaks and regime changes frequently occur.

The prediction task is defined as a one-step-ahead forecasting problem. Specifically, the model consumes 21 consecutive trading days of input features and predicts the return for the following trading day. This setup mirrors practical forecasting horizons in quantitative finance, where one-month historical windows are often used to generate short-term trading signals. Inputs to the model consist of both equity-related variables-returns from major technology firms, the S&P 500 index, and Bitcoin- and macroeconomic features constructed through lag and moving-average transformations. All variables are standardized prior to training to ensure comparability across heterogeneous sources.

Two model configurations were implemented to enable comparison. The baseline model relies solely on stock market data, while the proposed model incorporates both stock and macroeconomic information. This design isolates the incremental value of structured macroeconomic indicators and domain-informed feature engineering. By contrasting these two approaches under identical experimental conditions, the study provides a clear test of whether economic signals meaningfully improve forecasting performance.

Model performance is evaluated using the Root Mean Squared Error (RMSE), which penalizes larger deviations between predicted and actual returns and is widely used in financial forecasting research. Results indicate that the baseline model achieves an RMSE of 1.1344 on the test set, while the proposed model reduces this figure to 1.1212. The improvement of approximately 1.16 percent may appear numerically modest, but in the context of cumulative prediction errors, such a reduction can translate into substantial differences in risk management and portfolio outcomes.

In addition to accuracy metrics, interpretability is assessed through permutation importance. By repeatedly shuffling each variable and measuring the impact on prediction error, this method quantifies the marginal contribution of individual features. The analysis reveals that the 21-day lagged U.S. Dollar Index and the 5-day moving average of WTI oil prices exert the largest influence on prediction performance, followed by the federal funds rate and the VIX index. These results confirm that both monetary policy signals and risk sentiment indicators play critical roles in shaping stock market returns. Importantly, the findings illustrate how macroeconomic information interacts with firm-level equity signals, highlighting the complementary value of integrating bottom-up and top-down perspectives.

Finally, all preprocessing steps-including alignment of time series to daily frequency, forward-filling of lower-frequency indicators, log-return transformation of stock prices, and normalization of numerical features-were implemented consistently across models. This ensures that differences in predictive performance arise from the inclusion of macroeconomic signals rather than discrepancies in data treatment. Taken together, the experimental design provides a transparent and reproducible framework that isolates the contribution of macroeconomic information to stock price prediction in a transformer-based generative model.

4. Experimental Results and Analysis

The performance of the proposed StockGPT framework was evaluated against a baseline model that relied exclusively on stock market data. On the test set, the baseline model produced an RMSE of 1.1344, while the proposed model reduced this figure to 1.1212, corresponding to an improvement of approximately 1.16 percent. Although numerically modest, this reduction is significant in the context of long-horizon forecasting, where errors accumulate over time. Even small gains in accuracy can translate into meaningful improvements in portfolio allocation and risk management.

Figure 1 compares the predicted returns of both models against actual S&P 500 returns. The baseline model successfully captures overall market direction but shows pronounced deviations during periods of heightened volatility. In contrast, the enhanced model more closely aligns with observed market movements, particularly during episodes of rapid swings. This indicates that structured macroeconomic indicators provide stabilizing signals that complement the high-frequency information embedded in equity returns.

Beyond predictive accuracy, the study also examined interpretability through permutation importance analysis. The results reveal that the 21-day lagged U.S. Dollar Index and the 5-day moving average of WTI oil prices are the most influential predictors, followed by the federal funds rate and the VIX index. These findings suggest that delayed currency pressures and short-term energy market dynamics exert critical influence on stock returns, while monetary policy signals and investor sentiment continue to shape market behavior. The importance rankings of leading technology stocks, such as GOOGL and NVDA, further highlight their role as market bellwethers. Figure 2 reports the top 20 features and illustrates the complementary roles of firm-specific and macroeconomic variables.

Taken together, these results provide clear evidence that incorporating structured macroeconomic signals enhances forecasting performance beyond what can be achieved with stock-only inputs. They also validate the effectiveness of domain-informed feature design, particularly the use of lagged and smoothed transformations, in capturing economic dynamics. From a practical perspective, the findings suggest that quantitative investment strategies may benefit from hybrid models that combine bottom-up stock features with top-down macroeconomic signals. Such approaches offer a more comprehensive view of market behavior and greater resilience during periods of elevated volatility.

5. Conclusion

This study introduced an enhanced forecasting framework based on StockGPT that integrates macroeconomic indicators and time-series feature engineering. Using more than fifteen years of data covering major U.S. technology stocks, the S&P 500 index, Bitcoin, and a broad set of economic variables, the model was evaluated under realistic out-of-sample conditions. The results demonstrate that incorporating lagged and smoothed macroeconomic signals reduces forecasting error relative to a stock-only baseline, with improvements that, while modest in numerical terms, translate into meaningful gains when accumulated across longer horizons.

Beyond predictive performance, the analysis underscores the value of interpretability in deep learning models for finance. The identification of key drivers-such as the lagged U.S. Dollar Index, moving averages of oil prices, interest rates, and volatility indices-highlights the importance of combining firm-specific signals with structured macroeconomic information. These findings show that domain-informed feature design can bridge the gap between advanced machine learning architectures and the practical requirements of financial decision-making.

The contribution of this study lies in providing empirical evidence that structured economic information, when carefully engineered and integrated into transformer-based models, enhances both accuracy and transparency. For practitioners, the results suggest that hybrid forecasting pipelines that combine bottom-up stock features with top-down macroeconomic drivers offer a more comprehensive representation of market dynamics and provide additional robustness in volatile environments. For researchers, the study highlights the importance of embedding economic reasoning into data-driven models, pointing toward future work that deepens the connection between machine learning and financial economics.

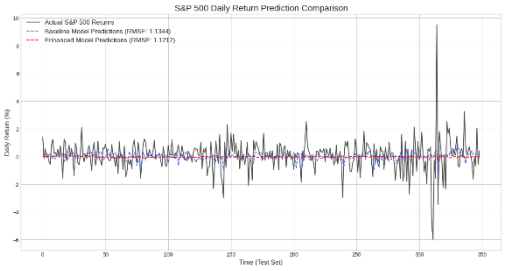
6. Limitations and Future Work

While the proposed StockGPT framework demonstrates measurable improvements in predictive accuracy and interpretability, several limitations remain. First, the dataset is restricted to major U.S. technology firms, the S&P 500 index, and a selection of macroeconomic indicators from the U.S. economy. Although these assets and signals are highly influential, the findings may not fully generalize to other sectors, asset classes, or international markets. Extending the analysis to include cross-industry equities or global data would provide a broader test of the model's applicability.

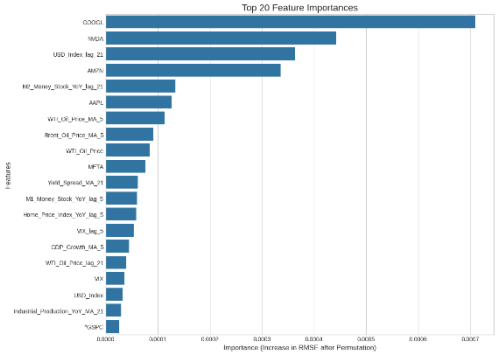
Second, the current study relies exclusively on structured numerical data. Unstructured information sources-such as financial news, analyst reports, and corporate disclosures-often capture sentiment and qualitative signals that play a critical role in driving sudden market movements. Integrating textual data streams with macroeconomic indicators could offer a richer representation of market dynamics and allow the model to respond more effectively to event-driven volatility.

Third, the evaluation is limited to a single generative transformer architecture. Although StockGPT provides a strong baseline, alternative models such as the Informer, Temporal Fusion Transformer, or ensemble approaches may yield additional performance gains. Moreover, while permutation importance offers valuable insights into feature contributions, complementary explainability methods such as SHAP values or attention-weight visualization could further illuminate the mechanisms underlying predictions.

These limitations point toward several promising directions for future research. Expanding datasets across industries and geographies, combining structured macroeconomic indicators with unstructured textual sentiment, and experimenting with diverse model architectures would all contribute to more robust and generalizable forecasting systems. Ultimately, the integration of economic reasoning, domain-specific feature engineering, and advanced machine learning techniques holds the potential to produce predictive models that are not only accurate but also transparent and practically relevant to financial decision-making.



(Figure 1) Comparison of predicted and actual S&P 500 daily returns: baseline vs. proposed model.



(Figure 2) Feature importance scores based on permutation analysis for top 20 variables.

References

[1] Chauhan, A., Singh, R., and Sharma, P., "Enhancing stock market prediction using Informer Transformer and sentiment analysis," Journal of Computational Finance and Economics, vol. 17, no. 1, pp. 45–63, 2025.

[2] Chen, L., Zhao, W., and Lin, M., "Improving economic indicator forecasts with Transformer-based models," International Journal of Forecasting, vol. 41, no. 2, pp. 112–127, 2025.

[3] Jin, X., Hong, S., and Yu, K., "Time-series feature engineering for deep learning in financial forecasting," Finance and AI Review, vol. 1, no. 1, pp. 22–45, 2023.

[4] Kang, J., and Bae, Y., "Feature-level fusion of financial and economic data for stock forecasting," Expert Systems with Applications, vol. 213, p. 119124, 2023.

[5] Kim, H., and Kim, J., "Macroeconomic indicators and stock market prediction using Transformer-based models," Journal of Financial Data Science, vol. 5, no. 2, pp. 55–72, 2023.

[6] Liu, Y., and Zhang, M., "Top-down vs bottom-up approaches in financial forecasting: A hybrid deep learning model," Applied Soft Computing, vol. 116, p. 108313, 2022.

[7] Mai, J., "StockGPT: Generative Pre-trained Transformer for stock prediction," Proceedings of the International Conference on AI in Finance, pp. 68–77, 2024.

[8] Nguyen, T., and Cho, S., "Deep learning in finance: From stock prices to macroeconomics," IEEE Transactions on Neural Networks, vol. 33, no. 5, pp. 2434–2449, 2022.

[9] Papageorgiou, A., Liakopoulos, A., and Koulis, T., "Deep reinforcement learning for financial market prediction," Machine Learning in Finance, vol. 9, no. 3, pp. 199–214, 2024.

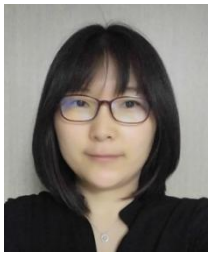
[10] Wang, C., and Lee, D., "Explainable AI for financial time series using permutation importance," AI in Finance Journal, vol. 3, no. 3, pp. 88–104, 2023.

[11] Wu, J., Wang, Z., and Lin, D., "BloombergGPT: A large language model tailored for finance," arXiv preprint arXiv:2304.03285, 2023.

[12] Yang, Y., Chen, H., and Xu, Y., "FinGPT: Open-source financial LLM for financial data analysis," arXiv preprint arXiv:2306.03085, 2023.

[13] Zhou, Z., and Tang, Q., "Economic sentiment vs. quantitative indicators: Which drives predictive performance?," Quantitative Finance Letters, vol. 8, no. 1, pp. 14–29, 2024.

Author Biography



Jieun Oh

Dongduk Women's University, B.S. in Computer Science, Class of 2001
Seoul School of Integrated Sciences & Technologies (aSSIST), M.S. in AI and Big Data, Class of 2025
Manager, PG Information Development Office, NICE Information & Telecommunication Co., Ltd. (since 2014)
Research Interests: AI Interpretability, AI-Driven Financial Forecasting, FinTech Innovation, and Data Governance
E-mail: jieunoh257@gmail.com

Articles of Association for "AI Business Academy"

Chapter 1 General Provisions

Article 1 (Title) This corporation shall be called the "AI Business Academy" (AIBA for short , hereinafter referred to as "the Academy ") .

Article 2 (Purpose) This academy aims to promote the development of Korean management studies through the convergence and integration of AI and management . Furthermore, it aims to contribute to the development of future Korean academy and the establishment of future strategies for companies .

Article 3 (Location of Office) The office of this academy shall be located under the Industrial Policy Research Institute (located at 7th floor , 203 Sinchon-ro, Seodaemun-gu, Seoul) , and regional headquarters may be established in other regions . A small number of administrative staff may be employed to operate the office .

Article 4 (Business)

1. In order to achieve the purpose of Article 2, this academy conducts the following business.
(1) Research related to theory and practice for grafting and integrating AI and Business
(2) Conducting research services commissioned by the government or companies at the level of industry-academia-government cooperation
(3) Holding regular academic seminars, policy discussions, and research presentations
(4) Publication of academic journals and research books
(5) Exchange with foreign similar academic societies
(6) Other projects necessary to achieve the purpose of this academy
2. When this academy carries out its business objectives, the benefits it provides to the beneficiaries shall be provided free of charge . However , in unavoidable cases, it may obtain

prior approval from the competent authority and have the beneficiaries bear a portion of the cost .

3. The benefits provided through the performance of the objectives of this academy shall not discriminate based on the beneficiary′s gender, place of birth, school attended, place of employment, occupation, or social status .

Chapter 2 Members

Article 5 (Types of Members) Members of this academy are individuals and groups that agree with the purpose of Article 2 , and include regular members (annual members and lifetime members) , associate members , and institutional members .

Article 6 (Regular Member) Regular members are individuals who fall under each of the following categories and have completed the membership process .

1. Full-time faculty members who teach or research AI or business-related subjects at universities , colleges, and graduate schools.
2. or part-time faculty or doctoral program or have obtained a degree in AI or business administration
3. Researchers at accredited research institutes and employees of AI- related companies
4. Other individuals recognized by the Board of Directors as having equivalent qualifications.

Article 7 (Classification of regular members) Regular members are classified into annual members and lifelong members . Annual members are those who have paid the annual membership fee , and the annual membership qualification is one year from the date of registration . Annual members are renewed by paying the annual membership fee . Lifelong members are those who have paid the lifetime membership fee .

Article 8 (Associate Member) Associate members are students enrolled in an undergraduate program in AI or business administration or individuals who have completed the membership process .

Article 9 (Institutional Member) Companies and organizations that agree with the purpose of this academy may become institutional members upon resolution of the board of directors. Membership is in principle for one year, but may be renewed .

Article 10 (Rights and Obligations) Members of this academy have the following rights and obligations .

1. All members can attend the general meeting and participate in discussions , and participate in the academy's business such as research presentations .
2. Regular members have the right to vote and be elected , but institutional members do not have the right to vote or be elected .
3. Members have the obligation to cooperate in the operation of the academy, including paying membership fees and attending academic conferences .

Article 11 (Suspension and loss of membership) Members of this academy will have their membership suspended or lost in the following cases :

1. If the membership fee is not paid , membership will be suspended and the rights under Article 9, Paragraphs 1 and 2 will be suspended until the membership fee is paid in full .
2. Members who damage the reputation of the Academy by violating the purpose of the Academy or the code of ethics established by the Academy shall lose their membership if the Board of Directors resolves to expel them .
3. The Code of Ethics is separately established by the Board of Directors and approved by the general meeting .

Article 12 (Withdrawal of membership) Members of this academy may withdraw at will .

Chapter 3 Executive Officers

Article 13 (Types of Executive Officers) The Executive officers of this academy shall be a president , vice president , directors and auditors , and the composition of the officers shall be as follows . However , regional headquarters shall be established in major regions at home and abroad , and the head of the regional headquarters shall be the vice president .

1. 1 Chairman
2. About 20 vice presidents
3. The number of directors is 5 to 50 people.
4. 2 auditors

Article 14 (Term of executive officers) The term of executive officers shall be two years, but they may be reappointed .

Article 15 (Appointment and election of executives)

1. The chairman and auditors are elected at the general meeting .
2. The Vice-Chairman and Directors are recommended by the Chairman and appointed by the Chairman with the approval of the Board of Directors . However, the first Vice-Chairman and Directors are appointed by the Chairman and approved by the General Meeting .
3. When an officer other than the chairman is unable to perform his/her duties during his/her term of office, the board of directors shall elect a replacement , and the term of office of the officer elected through the replacement shall be the remaining term of office of his/her predecessor .

Article 16 (Duties of Executive Officers)

1. The president represents the academy and supervises its business .
2. The director attends the board of directors meeting and deliberates on matters related to the business of this academy , and serves as the board of directors or chairman .
from Handle delegated tasks .
3. The auditor performs the following duties :
(1) Auditing the financial status and business of this academy.
(2) Auditing matters related to the operation of the board of directors
(3) If any irregularities or injustices are discovered in the audit results of (1) and (2), the Board of Directors or Requesting the General Assembly to correct the situation
(4) When necessary for reporting under (3) , requesting the convocation of a general meeting or of directors meeting.

Article 17 (Acting Chairman) When the Chairman is unable to perform his/her duties due to an accident or vacancy, the Vice Chairman designated by the Board of Directors shall act in his/her stead for the remainder of the term .

Article 18 (Dismissal of Executive Officers) If an executive officer loses membership qualifications pursuant to Article 11 or commits any of the following acts, he or she may be dismissed upon resolution of the Board of Directors .

1. Unfair acts such as serious dereliction of duty that run counter to the purpose of this academy.
2. Any act that causes a dispute between executives and is judged to make it impossible for the executives to perform their duties.
3. Any other acts that unfairly interfere with the work of this academy.

Chapter 4 General meeting

Article 19 (Matters to be resolved at the general meeting) The general meeting shall resolve the following matters :

1. Appointment and dismissal of the chairman and auditors
2. Endorsement of the first executive team
3. Changes to the Articles of Academy
4. Approval of settlement and business report
5. Provisions on the rights and obligations of members
6. Voting on matters proposed by the Chairman and Board of Directors
7. Other important matters

Article 20 (Convening of the General Meeting)

1. The general meeting is divided into a regular general meeting and a special general meeting .
2. The regular general meeting is held in February every year , and the extraordinary general meeting is convened by the chairman in the following cases :

(1) When the Chairman deems it necessary
(2) When there is a resolution by the board of directors

(3) When more than one-fifth of the members request in writing, stating the reason for holding the meeting.
(4) When the auditor requests a meeting pursuant to the provisions of Article 18, Paragraph 4.

Article 21 (Quorum for General Meeting Resolutions)
1. The general meeting shall be opened with members present , and attendance may be delegated by scanning an electronic document of a written or signed document.
2. The resolutions of the general meeting shall be decided by a majority vote of the members present, excluding those present in writing . However , In case of a tie, the chairman decides .
3. If the Steering Committee determines that it is difficult to convene an offline general meeting, an online general meeting may be held instead.

Article 22 (Reasons for exclusion from general meeting resolution) If the chairman or member falls under any of the following, he or she may not participate in the resolution .
1. Matters concerning oneself in the appointment and dismissal of executives
2. Matters involving the receipt of money or property, and in which the interests of the chairperson or member conflict with those of the academy.

Chapter 5 Board of Directors

Article 23 (Matters to be resolved by the Board of Directors) The Board of Directors resolves and agrees on the following matters :
1. Convening of a temporary general meeting
2. Resolution of agenda to be submitted to the general meeting
3. Matters delegated by the general meeting
4. Approval of executives appointed by the Chairman
5. Approval of business plan and budget
6. Approval and amendment of the composition and operating regulations of various committees.
7. Important changes to the journal′s editorial policy
8. Matters concerning investment and fund management

9. Resolution on membership of institutional members
10. Resolution on expulsion of members
11. Determination of membership fees
12. Other Matters

Article 24 (Delegation of Board of Directors′ Resolutions) Routine matters among the Board of Directors′ resolutions may be delegated to the Management Committee through a Board of Directors resolution .

Article 25 (Board of Directors) The Board of Directors is composed of a chairman , vice chairman , directors , and auditors .

Article 26 (Convening of the Board of Directors) The Board of Directors shall be convened by the Chairman, who shall also serve as its chairman, in the following cases :
1. When the Chairman deems it necessary
2. When more than half of the members of the board of directors, excluding the chairman, request a meeting by stating the purpose of the meeting.

Article 27 (Quorum for Board of Directors Resolution)
1. The board of directors shall open with the attendance of a majority of its members, and attendance may be delegated in writing .
2. The Board of Directors' resolutions are decided by a majority vote of the members present, excluding those present in writing . However , In case of a tie, the chairman decides .
3. If the Management Committee determines that it is impossible to convene a Board of Directors meeting due to scheduling issues that require a Board of Directors resolution, If a decision is made, the vote can be made online or in writing .

Article 28 (Reasons for exclusion from board of directors′ resolution) A board of directors member may not participate in the resolution if any of the following applies .
1. Matters concerning oneself in deciding on the loss of membership qualifications

2. Matters involving the receipt of money or property and in which the interests of the member and the academy conflict.

Chapter 6 Various Committees

Article 29 (Management Committee) An Management Committee is established to ensure smooth decision-making and efficient execution of the Academy′s business .

1. The Management Committee deliberates and decides on matters delegated by the Board of Directors in accordance with Article 24 of these Articles of Incorporation .
2. The Steering Committee is composed of the Chairman, Vice Chairman , Directors , and Secretariat members , and is composed of 10 or so members.
Appoint one person as chairman .
3. The operating committee chairman convenes the operating committee as needed .
4. The Steering Committee shall open with the attendance of a majority of its members , and resolutions shall be passed with the approval of a majority of the members present.

Article 30 (Editorial Committee) An editorial committee shall be established to publish the journal of this academy , and regulations regarding the editorial committee shall be separately determined by resolution of the board of directors .

Article 31 (Fund Management Committee) A Fund Management Committee shall be established to manage the funds of this Academy , and the regulations regarding this shall be separately determined by resolution of the Board of Directors .

Article 32 (Report on Committee Activities) The chairperson of each committee shall report the results of its activities to the board of directors .

Article 33 (Secretariat) A secretariat may be established to manage the affairs of this academy , and the regulations of the secretariat shall be established through a resolution of the board of directors .

Chapter 7 Assets and Accounting

Article 34 (Composition of assets)

The assets of this academy consist of the following items :

1. Donation
2. Membership fee
3. Income from basic assets
4. Other income

Article 35 (Types of Assets)

1. The assets of this academy are divided into two types: basic assets and general assets .
2. The basic assets listed in each clause below cannot be disposed of or provided as collateral . However , when there is a reasonable cause, a portion of them may be disposed of or provided as collateral with the approval of the Board of Directors and the authorization of the Minister of Strategy and Finance .

(1) Property contributed and designated as basic property

(2) Property that the board of directors has decided to use as basic property

3. Ordinary assets consist of assets other than basic assets .

Article 36 (Assessment and collection of membership fees) The method of levying and collecting membership fees shall be determined by the Board of Directors .

Article 37 (Expenses)

The expenses of this academy are paid from general assets .

Article 38 (Management of Assets)

1. The assets of this academy shall be managed by the chairman , and the method of management shall be determined by the board of directors .
2. This academy must open an Internet homepage and disclose the annual amount of donations raised and the results of their use through it .

Article 39 (Disposal of Surplus)

If there is a surplus at the end of the fiscal year, all or part of it is incorporated into capital or carried over to the next fiscal year, as determined by the Board of Directors .

Article 40 (Approval of Budget and Settlement of Accounts)

1. The budget for each fiscal year of this academy must be resolved by the Board of Directors and approved by the General meeting prior to the start of the fiscal year .
2. The chairman has the authority to execute the budget and business accounting .
3. The Chairman shall prepare a general accounting settlement report and a fund management settlement report within two months after the end of the fiscal year, attach an audit opinion, and report to the general meeting for approval .

Article 41 (Executive Compensation) As a rule , no compensation is paid to executives .

Article 42 (Fiscal Year)

fiscal year of this academy runs from September 1 to August 31 of the following year .

Chapter 8 Amendment of Articles of Academy and Dissolution

Article 43 (Amendment of Articles of Academy)

These Articles of Association may be amended with the approval of the Board of directors by a majority vote of at least three- quarters of the members present .

Article 44 (Dissolution)

This academy may be dissolved with the consent of more than two -thirds of its members .

″AI Business Academy″ Research Ethics Regulations

Chapter 1 General Provisions

Article 1 (Purpose) The purpose of these research ethics regulations (hereinafter referred to as ″ these regulations ″) is to present basic principles and directions for the members of the ″AI Business Academy″ (hereinafter referred to as the ″ the Academy ″) to promote research ethics and truthfulness, create a sound research atmosphere , and contribute to academic and social development .

Article 2 (Scope and Scope)

1. These regulations apply to all members of the academy and stakeholders directly or indirectly related to research activities .
2. Except in cases where there are other special provisions related to the establishment of research ethics and verification of research authenticity, these regulations shall apply .

Article 3 (Scope of misconduct) The misconduct set forth in these regulations includes the following acts that may damage the reputation of not only individuals but also the academy as a member of the academy .

1. Forgery : Here, ″ Forgery ″ refers to the act of falsely creating non-existent data or research results .
2. Modulation : Here, ″ Modulation ″ refers to an act of distorting research content or results by intentionally changing or deleting research data .
3. Plagiarism : Here, ″ plagiarism ″ refers to the act of stealing another person′s ideas , research content , research results, etc. without clearly indicating the source .

4. Double publication of papers : Here, ″ Double publication of papers″ refers to the act of publishing the same data, analysis methods, and analysis results in two or more academic journals .

5. Unfair authorship of a paper : Here, ″ unfair authorship ″ refers to an act of not granting authorship of a paper to a person who has made scientific or technical contributions or contributions to the research content or results without just cause , or granting authorship of a paper to a person who has not contributed as an expression of gratitude or courtesy .

6. Duplicate research : Here, ″ Duplicate research ″ refers to an act of conducting two or more research projects with the same content and publishing the same research results .

7. Public False Statement : Here, ″ Public false statement ″ refers to an act of making false statements about one′s academic background , career , qualifications , research achievements , results , etc.

8. Refers to an act of intentionally obstructing an investigation into suspicions of misconduct by oneself or another person or causing harm to the whistleblower .

9. Refers to any act that seriously deviates from the scope normally tolerated by the Academy of Information Systems in relation to other research .

Chapter 2 Basic Ethics for Researchers

Article 4 (Researcher's Morality) Researchers must conduct research in an honest and correct manner, not in a socially acceptable or ethically wrong manner, and present the results . Accordingly, the ethics that researchers must observe in relation to research are as follows . 1. Using others′ ideas, research content, or research results (hereinafter ″ research content ″) without clearly indicating them is clear plagiarism , and researchers must indicate the source as follows :

(1) When citing someone else's research content without re-editing (direct quotation) , the author's name , year , and page number must be indicated in the text with quotation marks (″″) , and the complete source must be indicated in the references section .

(2) When editing and using the research content of others, the author's name and year must be indicated in the text , and the full source must be indicated in the references section .

(3) When requoting content quoted by others, you must indicate ″ requotation ″ along with the quoter′s name and year .

2. When conducting research, researchers must recognize acts such as falsification or alteration of data as serious crimes and make efforts to prevent such misconduct from occurring .

3. It must not duplicate the previously published (or under review) research papers as if they were new research . Therefore, papers published in academic journals must be original work .

4. Those who have contributed directly or indirectly to the research must be listed as co-authors in the research paper , and those who have not contributed cannot be listed as authors . In addition, when conducting joint research , the person who has contributed the most to the research is the first author , and the corresponding author is limited to the person who is in charge of all correspondence between the authors and the academic academy and who has overseen the research or an equivalent person . However , in cases where co-authors cannot be listed, such as in cases of administrative , financial, or research support, this should be indicated in a footnote or acknowledgement .

5. It must truthfully write about their academic background , career , qualifications , research achievements , results , etc.

6. In relation to other research, one must not engage in any conduct that goes beyond the scope generally accepted in academia .

Article 5 (Social Responsibility) Researchers must be deeply aware of the impact their research will have on academy and, as professionals, must fulfill their responsibilities and obligations regarding the results of their research .

Article 6 (Respect for Others) Researchers must not harm or infringe upon the rights of others and must respect intellectual property rights . In addition , when dealing with others in relation to research, they must be treated equally regardless of race , gender , nationality , disability , etc.

Chapter 3 Roles and Responsibilities of the Research Ethics Committee

Article 7 (Composition and term of office of the committee)

1. The Research Ethics Committee (hereinafter referred to as the ″ Committee ″) shall be composed of 10 or fewer members, including the chairperson .
2. The chairman of the committee concurrently serves as the editor-in-chief of this academy .
3. The committee member concurrently serves as an editorial committee member of this academy .
4. The term of office for the chairperson and members shall be one year, but they may be reappointed .
5. The secretary of the committee shall be one of the vice-presidents of this academy , and the secretary shall not have voting rights .

Article 8 (Functions of the Committee) The Committee deliberates on the following matters :

1. Matters concerning fraudulent acts such as forgery , falsification , and plagiarism related to academic papers
2. Matters concerning research ethics raised regarding research plans , research results, etc. related to the academic academy
3. Investigation into misconduct related to the academic academy
4. Complaints raised regarding the truthfulness of research related to academic societies or journals
5. Matters concerning research ethics and education in research and development
6. Matters concerning the processing and follow-up measures of the results of research integrity verification
7. Matters to prevent misconduct that may occur during research
8. Other matters regarding research ethics raised by the president or committee chairperson

Article 9 (Convening , review , and resolution of the committee)

1. When a report of research misconduct is received, the chairperson convenes a committee to investigate and deliberate .

2. The committee is established with the attendance of a majority of its members , and resolutions are passed with the approval of a majority of the members present .
3. When the chairperson deems the agenda for deliberation to be minor, he/she may replace it with a written deliberation .
4. During the investigation process, the committee may request the attendance of the informant , the person under investigation , witnesses , and reference persons when necessary .

Article 10 (Duties and protection of rights of whistleblower)

1. ″ Whistleblower ″ refers to a person who has reported to this committee facts or related evidence of an act of misconduct .
2. The informant may report in any possible way , including verbally , in writing , by phone , or via e - mail . In principle, reports must be made under one′s real name . However , even if the report is anonymous, if the report includes the name of the research project , the title of the paper , and specific details and evidence of the misconduct in writing or via e-mail, and if it is found to be true, it will be processed in the same way as a report made under one′s real name .
3. Under no circumstances will the identity of the informant be directly or indirectly exposed , and the informant′s name will not be included in the investigation results report for the purpose of protecting the informant unless absolutely necessary .
4. A whistleblower who reported something even though he or she knew or could have known that the information was false is not eligible for protection .
5. The informant has an obligation to respond to the committee′s request to appear .

Article 11 (Duties and protection of rights of the subject of investigation)

1. ″ Subject of investigation ″ refers to a person who has become the subject of an investigation into an unethical act through a report or knowledge thereof, or a person who has become the subject of an investigation due to being presumed to have participated in an unethical act during the course of the investigation . Reference persons or witnesses during the course of the investigation are not included .

2. The person under investigation has an obligation to faithfully respond to any requests made by the committee during the investigation process . In addition, the person under investigation is guaranteed the right and opportunity to express opinions , raise objections, and defend himself/herself during the investigation process .
3. The committee must take care to ensure that the reputation or rights of the person under investigation are not violated until the verification of whether or not there was any misconduct is complete , and must endeavor to restore the reputation of the person under investigation who is found not to have been involved in any misconduct .
4. The subject of the investigation must faithfully comply with the request for submission of evidence requested by the committee .

Article 12 (Confidentiality) All matters related to the investigation, including reporting , investigation , deliberation , resolution, and recommendation measures , shall be kept confidential . Persons directly or indirectly participating in the investigation and related committee members shall not disclose any information obtained during the investigation and performance of duties . If the need for reasonable disclosure arises, it may be disclosed after a resolution by the committee .

Article 13 (Notification and Objection)
1. The results of the judgment on misconduct must be notified to the reporter and the person under investigation .
2. If there is a reasonable basis for an objection filed by the reporter or the subject of the investigation against the results of the decision, the objection may be filed within 30 days from the date of notification , and a reinvestigation may be conducted if necessary .

Chapter 4 Research Ethics Committee Verification Procedures and Standards

Article 14 (Preliminary Investigation)
1. A preliminary investigation refers to a procedure to determine whether or not there is a need to investigate a suspicion of misconduct , and must be initiated within 30 days from the date of

receipt of the report . The preliminary investigation is decided by a preliminary investigation committee consisting of a chairperson , vice chairperson , two members , and a secretary .
2. If the subject of the investigation admits to all facts of the misconduct as a result of the preliminary investigation , a decision can be made immediately without going through the main investigation procedure , and if it is determined that there is a possibility of significant damage to the evidence, the chairperson can take measures to preserve the evidence prior to the formation of the committee .
3. If it is decided not to conduct a main investigation in the preliminary investigation , the specific reason for this will be notified to the reporter within 10 days from the date of decision . However , no notification will be given in the case of anonymous reports .

Article 15 (Main Investigation)

1. ″ Main investigation ″ refers to the procedure to verify the facts of an act of misconduct , and must be conducted by forming a committee in accordance with the provisions of Article 7, Paragraph 1 .
2. The entire investigation schedule from the start to the decision must be completed within 6 months . However , if it is determined that the investigation cannot be conducted within this period, the reason for this may be notified to the president of the academic academy and the investigation period may be extended .
3. Main investigation If it is determined that fair progress is difficult due to a conflict of interest between a member of the committee and the subject of the investigation, the president of the academy may restrict the participation of the member in the relevant matter . However , the chairperson and vice chairperson are exceptions .
4. If the informant or the person under investigation raises a legitimate objection regarding the avoidance of a specific member, the committee may decide to accept it .

Article 16 (Recording and reporting of investigation results)

1. Records related to misconduct must be kept in document form at the academy for 5 years .
2. The committee must report the results , contents , and decisions of the investigation to the president of the academy and the board of directors .
3. The record of the judgment contents includes the following items .

(1) Report contents

(2) Allegations of misconduct and related papers or reports that are the subject of the investigation

(3) Objections or arguments of the informant and the subject of the investigation

(4) Judgment results and related supporting materials and witnesses

(5) List of judges

Article 17 (Post - judgment measures)

1. After the committee completes the review, it decides on the type of disciplinary action .

2. The types of disciplinary action are as follows depending on the severity of the misconduct , but may be applied multiple times .

(1) Expulsion

(2) Suspension of membership

(3) Restrictions on publication in academic journals (AI Management Journal)

(4) Official apology at the conference

(5) Notification of misconduct regarding related papers in academic journals or newsletters

3. If the investigation results confirm that there was no misconduct, the committee may take appropriate follow-up measures to restore the reputation of the person under investigation .

Article 18 (Administrative Matters)

1. Matters not specified in these regulations shall be decided and implemented by the committee in accordance with social norms .

2. The revision of the research ethics regulations shall follow the revision procedures of this academy .

3. For the smooth operation and activities of the committee, the academy Necessary administrative and financial support must be provided .

AI Journal of Business

Call for Papers

The AI Business Academy, a leader in the AI era, is recruiting papers to be published in Volume 4, in May. 2026 AI Journal of Business.

Now, all companies and businesses cannot grow and develop without utilizing AI and big data. In addition to the existing AI models representing artificial intelligence. The AI Business Academy would like to invite excellent papers that present various cases of utilizing AI and big data to innovate business and create new opportunities, as well as AI models that are utilized and models of AI strategies that serve as the basis.

We hope for active participation from many people.

1. Recruitment areas: Various management strategies and business innovations based on AI and big data, engineering improvements, and creative fields that combine management and engineering.
2. Recruitment deadline: March 31, 2026
3. Journal publication schedule: Reviewed by the end of April. 2026, then published in May.
4. Thesis Format: **Principle of submission as English version of paper** . It is much more advantageous and meaningful to write a paper in English in order to share and attract more interest from AI researchers around the world and to be cited in foreign papers. If you request it to the contact information below, a paper format file will be distributed. The basic format is the same as the format of this published journal (volume 2).
5. Submission : Academy website or editorial committee email: jhchang@assist.ac.kr
6. Contact: AI Business Academy Office (**aiba2023@naver.com)** or jhchang@assist.ac.kr

AI Journal of Business

Vol. 2, No. 1

Issued on November , 2025

Publisher : Jungho, Pyo

Editor : Joongho Chang

Publisher : AI Business Academy

7 floor , 203 Sinchon-ro, Seodaemun-gu, Seoul, Korea

Tel) **02-360-0775**

ISSN 3058-6992