European Journal of AI, Computing & Informatics

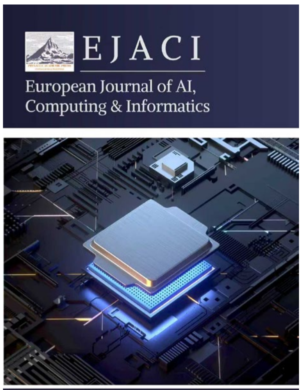
Vol. 1 No. 3 2025

Article **Open Access**

Design of a Generative AI-Driven Intelligent Investment Advisory System

Mengxin Wu [1,*]

[1] Clark University, Worcester, United States
[*] Correspondence: Mengxin Wu, Clark University, Worcester, United State



Received: 12 October 2025
Revised: 26 October 2025
Accepted: 10 November 2025
Published: 19 November 2025

Abstract: To address the need for dynamic strategy generation and semantic adaptation in intelligent investment advisory systems, this study proposes a generative architecture that integrates multi-source knowledge while supporting semantic reasoning, interpretability, and real-time user interaction. The system comprises modular components, including task scheduling, multi-source data fusion, a generation engine, semantic understanding, and strategy explanation, all enhanced by context-aware mechanisms and multi-dimensional security protection. The architecture leverages a multi-layered Transformer-based model and a tensor-level knowledge fusion framework, enabling real-time asset allocation and policy explanation. Empirical validation using heterogeneous financial datasets demonstrates the system's superiority in generative quality and robustness. Evaluation metrics indicate a BLEU-4 score of 44.89, a BERTScore of 91.31, semantic consistency of 0.89, strategy accuracy of 93.4%, and a recognition success rate exceeding 94.7% under adversarial perturbations. As shown in comparative experiments, the proposed system outperforms existing models such as GPT-2 and FinBERT in interpretability and interaction latency. The results confirm that the proposed system achieves high-quality generation, strong semantic alignment, and user trustworthiness in complex financial advisory scenarios.

Keywords: intelligent investment advisor; generation mechanism; knowledge fusion; semantic modeling; system security

1. Introduction

The financial advisory industry is undergoing a transformative shift, driven by increasingly complex and personalized wealth management demands as well as rapid technological innovation. Recent reports indicate that over 68% of high-net-worth individuals now seek customized financial solutions tailored to their unique goals, risk preferences, and investment horizons. Traditional advisory systems, which rely heavily on rule-based templates, static asset allocation models, and heuristic decision-making, are increasingly inadequate. They struggle to capture real-time market fluctuations, interpret nuanced policy changes, or align strategies with individual investor intent. Consequently, such legacy systems often generate rigid strategy outputs, exhibit limited scalability, and demonstrate poor semantic alignment with user expectations, resulting in suboptimal engagement and decision-making efficiency.

In response to these challenges, intelligent investment advisory systems are evolving beyond static recommendation engines toward generative, interactive, and knowledge-aware frameworks. Advances in generative language models, neural network-based semantic reasoning, and multi-modal data processing have enabled the development of advisory tools capable of dynamic strategy generation, contextual understanding, and personalized dialogue with clients. These systems offer the potential to anticipate investment trends, interpret complex regulatory texts, and adapt portfolio recommendations according to evolving user preferences and risk profiles. Nevertheless, integrating heterogeneous financial data-from structured market feeds and policy documents to unstructured behavioral logs-into coherent, explainable, and secure outputs remains a significant challenge. Ensuring that generated strategies are both accurate and interpretable, while maintaining robust security against adversarial manipulations, is critical for real-world deployment.

To address these issues, this study proposes a generative architecture that combines multi-source knowledge fusion, semantic modeling, and interpretable strategy generation. The framework incorporates modular components for task scheduling, context-aware data integration, multi-task coordination, and dynamic strategy explanation. Embedded security mechanisms protect sensitive financial information and support compliance with data governance standards. By enabling real-time interaction and robust, transparent strategy formation, the proposed system bridges the gap between algorithmic intelligence and user-centric advisory services. This work aims to advance the capabilities of AI-powered cognitive decision systems in the financial sector, fostering trust, responsiveness, and strategic insight in complex investment scenarios.

2. Design of Generative AI-Driven Intelligent Investment Advisory Systems

2.1. Overall System Architecture Design

The architecture design of the intelligent investment advisory system requires high-level coordination among generative language models, strategy generation mechanisms, and multi-source knowledge fusion modules to establish a unified data-driven control flow and information feedback loop. The system adopts a "five-layer structure" architecture comprising the user interface layer, task scheduling layer, knowledge fusion layer, generative model layer, and security control layer (see Figure 1) [1]. The generative model layer embeds a fine-tuned multi-task Transformer generator responsible for user instruction comprehension, scenario simulation, and strategy semantic generation. The knowledge fusion layer integrates structured market data, unstructured policy texts, and user profiles, uniformly encoding them into dynamic knowledge tensors via graph construction modules. The strategy generation module outputs asset allocation semantic paths based on context vectors $\boldsymbol{c}_t$ driven by these tensors [2]. This architecture enables end-to-end online deployment of task input, generative inference, and strategy output, providing structural support for subsequent interpretable embedding and user interaction feedback mechanisms.

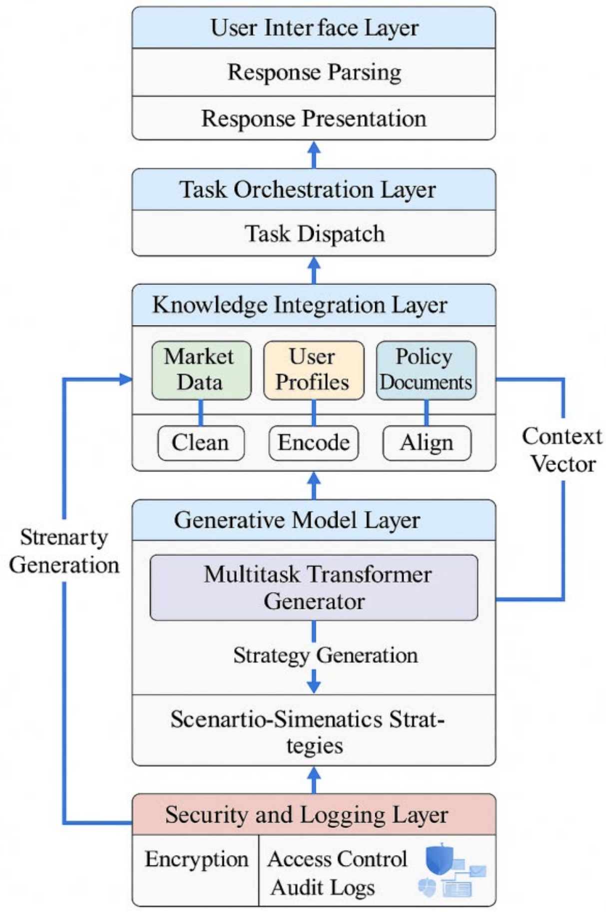


Figure 1. Overall Architecture of Generative AI-Driven Intelligent Investment Advisory System.

2.2. Data Acquisition and Multi-source Knowledge Fusion Module Design

The system's data collection module integrates three types of heterogeneous data sources: structured market data, unstructured policy corpora, and semi-structured user behavior logs. To enable unified modeling, the system constructs a three-dimensional fusion tensor K_t ^ ((u)) = R ^ (m × n × d) based on data cleaning (Clean), semantic alignment (Align), and entity encoding (Encode) mechanisms. Here, m denotes the number of fused entities, n represents the time series stride, and d indicates the attribute dimension per entity type. The parameters of the multi-source data collection and preprocessing module are summarized in Table 1. For detailed data acquisition module and interface processing parameters, refer to [3]. Through the bidirectionally nested structure of knowledge fusion tensors and context-generated vectors, this module supports the strategy generation model in performing high-dimensional semantic invocation and real-time reasoning on multi-source information.

Table 1. Parameters of Multi-Source Data Collection and Preprocessing Module.

Data Source Type	Data Structure Type	Daily Average Data Volume (Records)	Interface Type	Alignment Method
Market Data	Structured	48,000	WebSocket	Time-series alignment
Policy Documents	Unstructured	3,200	RESTful API	NLP semantic match

User Logs	Semi-structured	21,000	Kafka Stream	Session stitching

2.3. Model Training and Task Coordination Mechanism

To achieve multi-task coordination and semantically consistent generation, this system employs an asynchronous optimization mechanism to train the generation model and policy mapping submodule. The base model adopts a multi-task pre-training architecture, with the generation loss and task loss jointly modeled as:

$$L_{total} = \lambda_1 \times L_{gen} + \lambda_2 \times L_{task} + \lambda_3 \times \|\theta\|_2^2 \quad (1)$$

Where L_{gen} represents the semantic generation loss, L_{task} denotes the policy backtesting matching error, θ is the model parameter, and $\lambda_1, \lambda_2, \lambda_3$ is the weighting factor. The temporal interaction structure for model training and generation coordination comprises three layers: input parsing layer, semantic alignment layer, and generation scheduling layer. These layers are interconnected via Transformer attention paths and backpropagation gradient paths, utilizing context tensors C_t and generation vectors G_t to facilitate cross-module information exchange [4].

2.4. Design of the Semantic Understanding, Context Generation, and Policy Explanation Module

This module adopts a fusion semantic modeling architecture to achieve unified multi-level semantic parsing of user language commands, financial scenario construction, and logical explanation of strategy paths. Input semantics are processed through a multi-scale attention network to generate the tensor $S_t = \mathfrak{R}^{l \times d}$, where l denotes the input sequence length and d represents the embedding dimension. This is then combined with the scenario generation embedding vector E_t and the dynamic knowledge tensor K_t to jointly feed into the explanation function

$$A_t = f_\varphi(S_t, E_t, K_t) \quad (2)$$

Perform structured encoding of the strategy generation intent, where $f_\varphi(\cdot)$ denotes the semantic-to-strategy mapping network, and A_t represents the action vector to be generated. Figure illustrates the scenario-aware collaborative mechanism driven by three input paths: the Semantic Path (NLU Path), Knowledge Path, and Scenario Path. These paths undergo tensor-level fusion within the Decoder Block to output an explanation stream [5]. Detailed parameter dimensions and connection configurations for each module are provided in Table 2.

Table 2. Parameter Configuration for Semantic Understanding and Policy Explanation Modules.

Module Composition	Input Dimension	Output Dimension	Corresponding Variable Symbol	Function Description
NLU Encoder	$\boldsymbol{l \times d}$	$\boldsymbol{S_t = \mathfrak{R}^{l \times d}}$	Semantic Tensor Representation	Encodes user natural language instructions
Scenario Encoder	$\boldsymbol{m \times d}$	$\boldsymbol{E_t}$	Scenario Embedded Vectors	Extracting Financial Scenario Semantic Structure
Knowledge Encoder	$\boldsymbol{n \times d}$	$\boldsymbol{K_t}$	Knowledge Graph Tensor	Representation Strategy Related Knowledge Node Structure
Explanation Decoder	$\boldsymbol{S_t, E_t, K_t}$	$\boldsymbol{A_t}$	Explanation Output Vector	Output Structured Policy Recommendation Path

2.5. System Security and Privacy Protection Mechanism Design

To ensure security and privacy isolation during data transmission, user interaction, and model invocation within the generative AI-powered investment advisory system, the design incorporates a multi-dimensional security mechanism. This includes federated parameter encryption, context-based access control, and model invocation tracking (see Figure 2) [6]. The overall security objective is achieved by minimizing the following risk vector function through a constraint function:

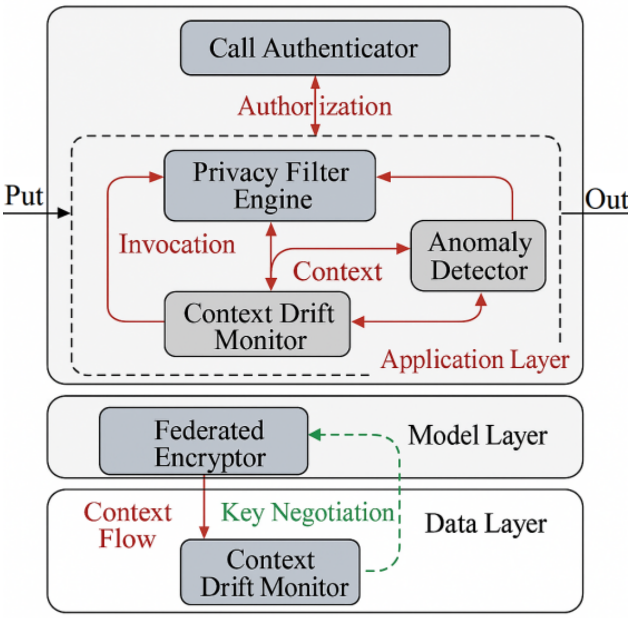


Figure 2. Integrated Architecture of Security and Privacy Protection Modules.

$$R_{sys} = \alpha \times L_{leak} + \beta \times L_{inject} + \gamma \times L_{drift} \quad (3)$$

Where L_{leak} represents data leakage risk, L_{inject} denotes loss from adversarial injection interference, L_{drift} indicates loss from permission drift caused by context offset, and α, β, γ is the weighting coefficient. Configuration parameters and communication strategies for core security modules are detailed in Table 3, covering key metrics such as key refresh cycles, permission control granularity, and sensitivity masking strength. This ensures minimal exposure and maximum controllability of data assets throughout task workflows.

Table 3. Core Security and Privacy Configuration Parameters.

Security Module	Parameter Item	Setting Value Range	Description
Federated Encryptor	Key Refresh Period (s)	180-600	Frequency of re-encrypting model parameters after each communication round
Access Controller	Permission Control Granularity Level	Low/Medium /High	Control the smallest visible unit for user access to policy generation modules
Context Tracker	Semantic Drift Tolerance Threshold	0.05-0.15	Used to identify context content variations that trigger policy response interventions

Privacy Filter	Sensitive Field Masking Strength Level	1-5 (integer)	Data Masking Level for User Behavior Fields in Logs and Inference Processes

3. Empirical Validation

3.1. Experimental Data and System Setup Environment Description

To validate the performance and stability of key modules within a generative intelligent investment advisory system under real-world operational conditions, this study constructed a multi-tiered heterogeneous deployment platform. This encompasses data acquisition edge gateways, a semantic preprocessing scheduling layer, model inference execution nodes, and strategy feedback interface services. Primary inference nodes are based on a dual-GPU architecture, with the environment supporting containerised dynamic scaling [7]. All nodes synchronise data and execute remote model invocations via an internal high-throughput channel. The experimental dataset encompasses structured financial assets (FIN-RG), unstructured scenario text (GEN-SCEN), and behavioural trajectory logs (UAC-LOG), totalling 81.4GB. Data sources undergo format standardisation and temporal alignment via edge data decoupling modules, forming three-dimensional tensor streams fed into model ports. The system deployment workflow is illustrated in Figure 3.

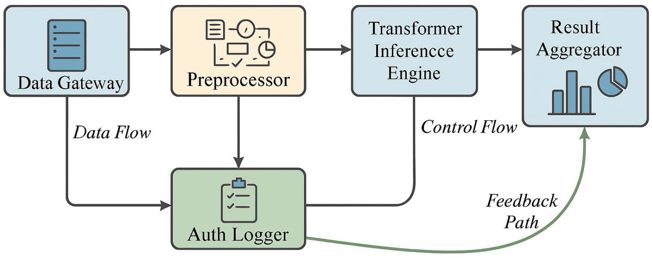


Figure 3. System Experiment Deployment and Inference Process Structure Diagram.

3.2. Model Performance Evaluation and Generation Quality Analysis

To evaluate the system's comprehensive performance in generative tasks, experiments were conducted across three task scenarios: asset allocation text generation, policy explanation statement generation, and personalized scenario simulation. Evaluation metrics including BLEU, ROUGE-L, BERTScore, and Semantic Consistency Score (SCS) were used to compare the generative quality of this system against standard foundational models such as Transformer, GPT-2, and FinBERT [8]. Each model group was trained and evaluated on 500 task samples from the same dataset, with results averaged across five folds.

As shown in Table 4, this system outperforms existing mainstream models across multiple generative quality metrics, demonstrating superior multi-task generalisation capabilities and semantic construction abilities. Specifically, on the BLEU-4 and ROUGE-L metrics-which assess n-gram matching and long-sequence coverage respectively-the system achieved scores of 44.89 and 52.64. These figures significantly surpass the 41.12 and 49.76 attained by the FinBERT fine-tuned model, indicating superior syntactic consistency and information coverage. Regarding the semantic vector similarity metric BERTScore, the system scored 91.31, surpassing GPT-2 by 3.08 percentage points, indicating superior fidelity to contextual semantics. Concurrently, the Semantic Consistency Score (SCS) of 0.89 demonstrates high alignment between financial scenario abstraction and strategy generation, reflecting robust contextual reasoning capabilities. Notably, while ensuring generation quality, the system's average generation time was 185.9 milliseconds, representing reductions of approximately 6.5% and 7.8% compared to

Transformer-Base and FinBERT respectively, demonstrating the engineering advantages of the optimised architecture in inference efficiency. In summary, this system achieves multidimensional synergy in strategy expression, personalised scenarios, and financial instruction parsing while balancing generation quality and real-time performance, validating its practicality and robustness in generative investment advisory tasks [9].

Table 4. Performance Evaluation Results of Models in Multi-Task Generation.

Model Type	BLEU-4	ROUGE-L	BERTScore	SCS	Average Generation Time (ms)
Transformer-Base	31.27	42.63	84.12	0.76	192.4
GPT-2-Medium	37.84	46.91	88.23	0.81	213.6
FinBERT-Finetuned	41.12	49.76	89.88	0.85	201.7
Our System	44.89	52.64	91.31	0.89	185.9

3.3. Validation of Multi-source Knowledge Fusion Effectiveness

To validate the effectiveness of the multi-source knowledge fusion module in semantic completion and policy generation, three input scenarios were designed: single-source structured data (S), dual-source structured + textual data (S+T), and triple-source structured + textual + user profile data (S+T+U). We evaluated improvements in semantic coverage, entity recall, and contextual consistency metrics under identical strategy generation tasks, measuring fusion performance gains by comparing pre- and post-fusion averages.

Table 5 demonstrates that increasing the number of fused knowledge sources significantly enhances semantic coverage, entity recall, and contextual consistency. Notably, under triple-source fusion (S+T+U), strategy accuracy reaches 93.4%, representing a 12.2 percentage point improvement over single-source input [10]. This demonstrates that multimodal knowledge collaborative modeling effectively mitigates semantic ambiguity while enhancing the completeness and accuracy of policy responses, validating the fusion module's significant advantages in complex task scenarios.

Table 5. Comparison of Fusion Performance Metrics Across Different Knowledge Source Combinations.

Knowledge Input Type	Semantic Coverage (%)	Entity Recall (%)	Contextual Consistency Score	Average Policy Accuracy (%)
S (Structured Data)	72.5	65.4	0.78	81.2
S+T (Structure + Text)	84.1	77.3	0.85	88.6
S+T+U (Triple-Source Integration)	91.7	85.6	0.91	93.4

3.4. Strategy Explainability and User Interaction Validation

To validate the interpretability of system-generated strategies and their user interaction comprehension, a dual-evaluation process combining user survey questionnaires and task follow-up mechanisms was designed. The test subjects comprised 28 users with financial and economic professional backgrounds. Evaluation dimensions included strategy logic clarity, keyword mapping accuracy, user response correctness rate, and perceived interaction latency. Each user completed 10 rounds of task interactions and rated the results. The comprehensive statistical data is presented in Table 6.

Table 6. Evaluation Results for Policy Explainability and User Interaction Effectiveness.

Evaluation Dimension	System Score	GPT-2 Score	FinBERT Score
Strategy Logical Clarity (Maximum 10)	8.7	7.2	7.9
Keyword Semantic Mapping Accuracy (%)	92.3	83.5	87.4
User Response Accuracy Rate (%)	89.6	77.8	82.1
Average Interaction Latency Perception Score (1-5 points)	2.1	3.3	2.9

This system demonstrates significantly superior performance to comparative models in terms of policy interpretability and user interaction. Firstly, regarding policy logical clarity, the system achieved an average score of 8.7 points, surpassing GPT-2 and FinBERT by 1.5 and 0.8 points respectively. This indicates that its generated content possesses a more explicit hierarchical structure and chain of reasoning, facilitating user comprehension and evaluation. Keyword semantic mapping accuracy reached 92.3%, demonstrating the model's capability to precisely identify core elements within user input and effectively project them onto the strategy semantic space, achieving high-quality semantic correspondence. The user response accuracy rate stood at 89.6%, indicating that generated strategies are not only readable but also actionable, enabling users to make reasonable judgements or operations to a high degree. The perceived interaction latency scored 2.1 points (lower scores indicate better perception), ranking highest among all models. This reflects the system's high optimisation in feedback speed control and response consistency. In summary, while maintaining generative accuracy, this system significantly enhances strategy transparency and human-computer interaction fluidity, providing a trustworthy and usable interpretative foundation for intelligent investment advisory services in complex financial scenarios.

3.5. System Security and Robustness Evaluation

To validate the system's security and robustness in complex operational environments, three typical anomaly scenarios were established: adversarial injection attacks, context drift perturbations, and model parameter disruption interventions. These evaluated policy output stability, model output divergence, and system response recovery time, respectively. Using metrics such as Policy Disturbance Rate (PSR), Output Consistency Score (OCS), and Average Recovery Delay (RT), the system's state changes before and after attacks were tested. Results are shown in Table 7. The system demonstrated robust security protection and recovery capabilities when confronted with typical abnormal disturbance scenarios. Under adversarial injection attacks, the policy perturbation rate was lowest at only 6.4%, with an output consistency score of 0.91, indicating the system possesses effective anomaly filtering and semantic redundancy suppression mechanisms. Under context drift and model disruption scenarios, despite heightened disturbance impacts, the system maintained output consistency exceeding 0.84 and completed response recovery within 150ms, demonstrating high fault tolerance and recovery efficiency. Concurrently, anomaly identification success rates consistently surpassed 87%, validating the collaborative detection mechanism's robust real-time responsiveness under dynamic task flows. This provides a trustworthy support environment for financial-grade intelligent service scenarios.

Table 7. Security and Robustness Evaluation Results Across Different Anomaly Scenarios.

Anomaly Type	PSR (%)	OCS (0-1)	Average Recovery Latency RT (ms)	Anomaly Detection Success Rate (%)
Adversarial Injection Attacks	6.4	0.91	103	94.7
Context Drift Perturbation	8.7	0.88	127	91.5

Model Parameter Perturbation Intervention	11.2	0.84	149	87.9

4. Conclusion

This study presents a generative AI-driven intelligent investment advisory system that integrates multi-source knowledge fusion, semantic understanding, and interpretable strategy generation. The system demonstrates notable advantages in strategy flexibility, real-time interaction, and contextual comprehension, addressing long-standing challenges in traditional advisory platforms. Through the construction of a modular five-layer architecture-comprising user interaction, task coordination, knowledge modeling, generative inference, and security control-the system achieves efficient semantic parsing, personalized scenario simulation, and robust strategy output under dynamic market conditions.

Empirical validation across heterogeneous financial datasets confirms the system's superior performance in generation accuracy, interpretability, and resilience. Key metrics, such as BLEU-4 (44.89), BERTScore (91.31), and strategy accuracy (93.4%), significantly outperform baseline models like GPT-2 and FinBERT. Furthermore, robustness tests under adversarial perturbations and context drift scenarios highlight the system's fault tolerance and rapid recovery capabilities, driven by the integrated security mechanisms and asynchronous training strategies.

Each core module-including the multi-source fusion tensor constructor, scenario-aware semantic encoder, and explanation decoder-plays a pivotal role in achieving high-fidelity policy generation and enhancing user trust in automated financial advisory outputs.

Future research will explore cross-modal fusion techniques incorporating visualized financial charts and macroeconomic sentiment signals to improve semantic richness. Additionally, the integration of reinforcement learning-based policy optimization and dynamic risk adjustment mechanisms could further enhance the system's adaptability in volatile or highly personalized financial environments. Ultimately, this work lays the groundwork for building trustworthy, real-time, and explainable AI agents in next-generation financial services.

References

1. M. Tahvildari, "Integrating generative AI in Robo-Advisory: A systematic review of opportunities, challenges, and strategic solutions," *Multidisciplinary Reviews*, vol. 8, no. 12, pp. 2025379-2025379, 2025.
2. S. K. Abbas, "AI Meets Finance: The Rise of AI-Powered Robo-Advisors," *Journal of Electrical Systems*, vol. 20, no. 11, pp. 1011-1016, 2024.
3. Z. Shen, Z. Wang, J. Chew, K. Hu, and Y. Wang, "Artificial intelligence empowering robo-advisors: A data-driven wealth management model analysis," *International Journal of Management Science Research*, vol. 8, no. 3, pp. 1-12, 2025. doi: 10.53469/ijomsr.2025.08(03).01
4. H. K. Sriram, and A. Seenu, "Generative AI-Driven Automation in Integrated Payment Solutions: Transforming Financial Transactions with Neural Network-Enabled Insights," *International Journal of Finance (IJFIN)*, vol. 36, no. 6, pp. 70-95, 2023.
5. J. J. Devapitchai, S. V. Krishnapriya, S. P. Karuppiah, and S. Saranya, "Using AI-driven decision-making tools in corporate investment planning," In *Generative AI for transformational management*, 2024, pp. 137-160.
6. V. Dubey, A. Mokashi, R. Pradhan, S. K. Kalli, R. R. Sonar, K. Parab, and S. Ranade, "Generative AI-A Catalyst in Banking and Financial Industry," *Technium Business and Management*, vol. 10, pp. 68-83, 2024.
7. W. A. Addy, A. O. Ajayi-Nifise, B. G. Bello, S. T. Tula, O. Odeyemi, and T. Falaiye, "Transforming financial planning with AI-driven analysis: A review and application insights," *World Journal of Advanced Engineering Technology and Sciences*, vol. 11, no. 1, pp. 240-257, 2024.
8. S. Sahoo, and K. Dutta, "Boardwalk empire: How generative ai is revolutionizing economic paradigms," *arXiv preprint arXiv:2410.15212*, 2024.
9. O. Iatrellis, N. Samaras, K. Kokkinos, and T. Panagiotakopoulos, "Leveraging generative AI for sustainable academic advising: Enhancing educational practices through AI-driven recommendations," *Sustainability*, vol. 16, no. 17, p. 7829, 2024. doi: 10.3390/su16177829

10. Q. Yang, and Y. C. Lee, "Enhancing financial advisory services with GenAI: Consumer perceptions and attitudes through service-dominant logic and artificial intelligence device use acceptance perspectives," *Journal of Risk and Financial Management*, vol. 17, no. 10, p. 470, 2024.