IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, June 2026
Impact Factor: 8.2

Salary Consumption Using Large Language Models

Dr. Sanjaykumar P. Pingat[1], Mohit Sunil Warke[2], Sai Sharad Ghoalp[3], Harshwardhan Ramesh Kolekar[4], Siddhesh Rajendra Marne[5]
Department of Computer Engineering[1-5]
Shrimati Kashibai Navale College of Engineering, Pune, Maharashtra, India
sanjaypingat@gmail.com, mohitwarke70@gmail.com, saighoalp2004@gmail.com,
Kolekarharsh78@gmail.com, siddheshmarne02@gmail.com

Abstract: *The proliferation of Large Language Models (LLMs) has created new opportunities for intelligent financial analysis and decision-support systems. This paper presents a novel LLM-based framework for salary consumption analysis, designed to assist individuals and organizations in understanding, predicting, and optimizing salary expenditure patterns. The proposed system combines transformer-based language models with Retrieval-Augmented Generation (RAG), enabling contextually aware financial guidance, anomaly detection in spending behavior, and personalized budgeting recommendations. The framework integrates structured financial data preprocessing, semantic embedding, and real-time retrieval to ensure accurate and reliable insights. Experimental evaluation demonstrates that the system significantly improves prediction accuracy for expenditure categories and delivers actionable financial guidance while reducing dependency on static financial advisors. The results affirm that LLMs integrated with financial domain knowledge provide an effective and scalable solution for salary consumption management.*

Keywords: large language models, salary consumption, financial analytics, retrieval-augmented generation, expenditure prediction, personalized budgeting, transformer models

I. INTRODUCTION

The management of personal and organizational finances has become increasingly complex in modern economies. Salary consumption—the manner in which earned income is allocated, spent, and saved directly influences individual financial well-being, organizational payroll efficiency, and macroeconomic stability. Despite the availability of financial tools and advisory services, a significant portion of individuals and small enterprises continue to face challenges related to overspending, poor budgetary planning, and inadequate financial forecasting.

The emergence of Artificial Intelligence (AI) and Natural Language Processing (NLP) technologies, particularly Large Language Models (LLMs), has introduced transformative possibilities for intelligent financial assistance. LLMs such as GPT-4, LLAMA, and similar transformer-based architectures possess remarkable capabilities for contextual understanding, reasoning, and language generation, enabling them to interpret complex financial narratives, answer domain-specific queries, and generate actionable recommendations.

However, deploying LLMs directly for salary consumption analysis presents challenges including hallucination risks, lack of domain-specific grounding, and inability to process structured financial datasets effectively. These limitations necessitate the integration of LLMs with complementary techniques such as Retrieval-Augmented Generation (RAG), which anchors generative outputs in verified financial knowledge, and structured data processing pipelines capable of handling transactional records.

This paper proposes an intelligent Salary Consumption Analysis System (SCAS) that leverages LLMs integrated with RAG to deliver personalized, accurate, and context-aware financial insights. The system is designed to analyze income and expenditure data, identify consumption patterns, predict future spending behaviors, and provide tailored budgetary

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, June 2026
Impact Factor: 8.2

recommendations. By combining transformer-based models with financial knowledge retrieval, the proposed system addresses the key limitations of standalone LLM applications in financial contexts.

The primary contributions of this work are: (i) the design of an LLM-based framework specifically tailored for salary consumption analysis, (ii) the integration of RAG to ensure factual accuracy and domain relevance in financial guidance, (iii) the implementation of an adaptive user profiling module to enable personalized recommendations, and (iv) an experimental evaluation demonstrating improved performance over conventional financial analysis tools.

II. LITERATURE REVIEW

The intersection of artificial intelligence and financial analytics has attracted considerable research attention in recent years. Transformer-based architectures have emerged as foundational models enabling state-of-the-art performance across NLP tasks, providing the computational basis for modern LLMs [1].

Several studies have explored the application of LLMs in financial domains. Araci introduced FinBERT, a domain-specific BERT model fine-tuned on financial corpora, demonstrating superior performance in sentiment analysis of financial news and earnings call transcripts [2]. This work established the viability of domain-adapted language models for specialized financial tasks. Yang et al. presented a comprehensive study on the application of pre-trained language models for financial question answering and information extraction. Their findings highlight that LLMs fine-tuned on financial datasets substantially outperform general-purpose models in tasks requiring domain-specific reasoning [3].

In the area of personal finance management, Khasawneh et al. proposed a machine-learning-based system for personal expenditure classification and prediction, leveraging transaction history to forecast spending categories [4]. While effective, their approach lacked natural language interaction capabilities and contextual adaptability.

Retrieval-Augmented Generation (RAG) was formally introduced by Lewis et al. as a hybrid model that combines parametric memory from pre-trained LLMs with non-parametric retrieval from external document stores, demonstrating significant improvements in knowledge-intensive NLP tasks [5]. This paradigm has since been applied across healthcare, legal, and educational domains. Shah et al. investigated the use of conversational AI agents for financial advisory services, reporting improved user engagement and satisfaction compared to traditional rule-based advisory systems [6]. However, their system operated without grounding in personalized financial data, limiting its utility for individual salary consumption analysis.

Recent surveys on generative AI highlight the potential of LLMs for multi-domain decision support while identifying persistent challenges including hallucination, bias, and ethical governance [7]. In the financial domain, these challenges are particularly consequential, necessitating robust validation and retrieval-based grounding strategies. Despite advances, existing literature reveals a gap in end-to-end LLM-based frameworks specifically designed for salary consumption analysis that integrate user profiling, structured financial data processing, and real-time retrieval. The proposed system addresses this gap by combining these components into a unified, scalable architecture. III.

III. METHODOLOGY

A. System Architecture

The proposed Salary Consumption Analysis System (SCAS) is built on a modular architecture comprising five core components: the Data Ingestion and Preprocessing Module, the Financial Knowledge Base, the Retrieval Engine, the User Profiling Module, and the LLM-based Response Generator. Together, these components enable end-to-end salary consumption analysis, from raw financial data input to actionable budgetary recommendations.

The Data Ingestion and Preprocessing Module accepts structured financial data in formats such as CSV, JSON, and PDF bank statements. Transaction records are parsed, normalized, and classified into expenditure categories using rule-based and ML-assisted classifiers. Preprocessed data is subsequently stored in a relational database and partially vectorized for retrieval.

The Financial Knowledge Base consists of curated domain resources including financial planning guidelines, tax regulations, budgeting frameworks (such as the 50/30/20 rule), and domain-specific FAQs. These documents are chunked and converted into dense vector embeddings using a pre-trained Sentence-BERT model, stored in a vector database for semantic retrieval.

The Retrieval Engine performs similarity-based search over the knowledge base upon receiving a user query or analysis request. Using cosine similarity and approximate nearest neighbour algorithms (HNSW), the engine retrieves the top-k most relevant documents to ground the LLM response in verified financial knowledge.

The User Profiling Module maintains individual financial profiles, including income brackets, recurring expenses, savings history, and stated financial goals. This module enables the system to personalize responses and recommendations based on each user's unique financial context.

B. Workflow

The operational workflow begins when a user submits a natural language query or uploads financial data. The system preprocesses the input, retrieves relevant financial context from the knowledge base, and constructs an enriched prompt combining the user query, retrieved context, and user profile. This prompt is processed by the LLM, which generates a contextually grounded and personalized response. Outputs are delivered through an interactive interface, and interaction logs update the user profile for continuous personalization.

C. Algorithmic Techniques

Expenditure Classification: Transaction data is classified into categories (housing, food, transport, entertainment, savings) using a fine-tuned BERT-based classifier trained on labeled financial transaction datasets, achieving improved accuracy over keyword-based methods.

Semantic Retrieval: User queries are embedded using Sentence-BERT and matched against the financial knowledge base via HNSW-based ANN search, ensuring sub-millisecond retrieval latency at scale.

Prompt Engineering: Structured prompts are designed to incorporate retrieved context, user profile data, and financial query, guiding the LLM to produce domain-accurate, personalized guidance while minimizing hallucination.

D. Prediction Model

Future expenditure prediction is achieved through a hybrid approach combining time-series analysis (ARIMA and LSTM networks) on historical transaction data with LLM-generated narrative insights. The LSTM model captures long-term spending trends, while the LLM provides interpretable explanations and actionable recommendations aligned with the predicted patterns.

IV. RESULTS AND DISCUSSION

The proposed SCAS was evaluated on a dataset comprising 2,400 anonymized monthly salary and expenditure records collected from individuals across diverse income brackets. The system was assessed on four dimensions: expenditure classification accuracy, prediction performance, retrieval relevance, and user satisfaction with generated recommendations.

A. Expenditure Classification

The fine-tuned BERT-based classifier achieved an overall accuracy of 92.4% across seven expenditure categories, outperforming a baseline keyword-matching classifier (74.3%) and a traditional SVM-based model (84.6%). The improvement is attributed to the contextual encoding capabilities of transformer models, which correctly disambiguate transactions based on merchant context and description semantics.

Table 1: Expenditure Classification Accuracy Comparison

Model	Accuracy (%)	F1-Score
Keyword Matching	74.3	0.71

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, June 2026
Impact Factor: 8.2

SVM Classifier	84.6	0.83
FinBERT (Fine-tuned)	89.1	0.87
Proposed (BERT+RAG)	92.4	0.91

B. Expenditure Prediction

The hybrid LSTM-LLM prediction model achieved a Mean Absolute Percentage Error (MAPE) of 6.8% for 30-day expenditure forecasting, compared to 11.4% for standalone ARIMA and 9.2% for a baseline LSTM model. The LLM component contributed interpretable narrative explanations for predicted trends, which user feedback indicated significantly improved comprehension and trust in the recommendations.

C. Retrieval and Response Quality

Retrieval relevance was assessed using precision@5 and Mean Reciprocal Rank (MRR). The HNSW-based retrieval system achieved a precision@5 of 0.89 and MRR of 0.93 on a curated financial query benchmark. RAG-augmented responses scored an average of 4.3/5.0 for factual accuracy as rated by domain experts, compared to 3.1/5.0 for non-augmented LLM responses, confirming the effectiveness of knowledge grounding.

D. User Satisfaction

A user study with 45 participants over a four-week evaluation period found that 87% rated the personalized budgeting recommendations as useful or highly useful. Average response comprehension score improved by 34% compared to static financial advisory materials. System response latency averaged 1.4 seconds per query, satisfying real-time interaction requirements.

V. FUTURE SCOPE

The proposed system can be extended in several directions. First, multi-modal input support can be incorporated to accept images of receipts, scanned invoices, and voice-based financial queries, broadening accessibility. Second, reinforcement learning from human feedback (RLHF) can be applied to continuously improve the quality and personalization of recommendations based on user interactions and outcomes. Third, integration with open banking APIs and real-time transaction streams would enable continuous monitoring and proactive financial alerts. Fourth, the system can be adapted for organizational payroll analytics, enabling HR and finance departments to identify compensation inefficiencies and optimize salary structures. Finally, multilingual support and adherence to diverse regional financial regulations will be critical for global deployment and inclusive financial literacy.

VI. CONCLUSION

This paper presented an LLM-based Salary Consumption Analysis System (SCAS) that integrates transformer-based language models with Retrieval-Augmented Generation and adaptive user profiling to deliver accurate, personalized, and contextually grounded financial insights. By addressing the limitations of standalone LLMs through domain-specific knowledge retrieval and structured financial data processing, the proposed system achieves superior performance in expenditure classification, spending prediction, and recommendation quality. Experimental results demonstrate that the RAG-augmented LLM framework outperforms conventional approaches across all evaluation metrics, while user studies confirm high satisfaction and practical utility. The modular architecture ensures scalability and adaptability for both personal and organizational financial contexts.

REFERENCES

[1] A. Vaswani et al., "Attention Is All You Need," in Proc. 31st Int. Conf. Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 2017, pp. 5998-6008.

[2] D. Araci, "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models," arXiv preprint arXiv:1908.10063, 2019.
[3] Y. Yang, M. Uy, and A. Huang, "FinBERT: A Pretrained Language Model for Financial Communications," arXiv preprint arXiv:2006.08097, 2020.
[4] T. Khasawneh, S. Alsubhi, A. Alzahrani, and W. H. Alharbi, "Personal Finance Management Using Machine Learning," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 3, pp. 122-130, 2020.
[5] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," Advances in Neural Information Processing Systems, vol. 33, pp. 9459-9474, 2020.
[6] S. Shah, Z. Shaikh, T. Khan, and R. Kaur, "Conversational AI for Financial Advisory Services," IEEE Access, vol. 10, pp. 41230-41245, 2022.
[7] S. Bengesi et al., "Advancements in Generative AI: A Comprehensive Review of GANs, GPT, Autoencoders, Diffusion Model, and Transformers," IEEE Access, vol. 12, pp. 69812-69840, 2024.