Rover Publications
United International Journal of Engineering and Sciences (UIJES)
An International Peer-Reviewed (Refereed)Engineering and Science Journal
Impact Factor: 7.984 (SJIF) Vol-4, Issue- 2(April),2024 ISSN:2582-5887 www.uijes.com

Lifecycle Management of Large Language Models in Financial Systems: Monitoring, Drift Detection, and Governance

Gopichand Talluri
Overland Park, Kansas, USA, 66213
Gopichand.bigdtech@gmail.com

Abstract
The financial services have experienced various forms of innovation as the industry has been using Large Language Models (LLMs) that have brought fresh innovations in the risk assessment, fraud detection, and decision-support disciplines. However, the application of LLAMs in practice to finance-related situations has some severe concerns related to the monitoring, model drift, and governance. In the proposed paper, an integrated MLOps would be proposed that would involve round-the-clock monitoring, drift identification and control of LLMs to achieve reliability, stability, and compliance. The framework quantifies performance, uses statistical drift-detection algorithms and governance scoring to sustain the model over time. Synthetic datasets experimental analysis demonstrates that accuracy is improved, output drift is lower and governance compliance is higher than that of the baseline models. The results highlight the relevance of the proposed solution to address the limitations of traditional MLOps systems and offer a feasible implementation of LLMs in finance.

Keywords: Large Language Models (LLMs), MLOps / LLMOps, Drift Detection, Model Monitoring, Financial AI Governance.

1.Introduction

The rapid development of Large Language Models (LLMs) has radically changed the situation in the sphere of artificial intelligence as it enables such advanced capabilities as natural language understanding, logic, and decision-making [1]. They find use in financial services to identify fraud, evaluate risk, algorithmic trading, customer service and compliance with regulations among others [2]. They come in handy in dynamic and data-driven environments because of their ability to process a huge volume of structured and unstructured financial data [3] [4]. This poses a grave issue of trust, transparency and accountability at least in some of the most regulated financial spheres [5].

The second important problem is the model drift, where the performance of an LLM deteriorates with time as a result of the shift in the data distribution, market conditions, or user behavior [6]. There is no economic loss or violation of compliance that can occur in financial systems because of any negligence [7]. Thus, the strong drift detection systems are necessary in order to guarantee the stability and accuracy of the models. In addition, the periodic inspection of the LLMs is necessary to maintain the performance, detect anomalies and meet the regulations [8]. Older MLOps implements offer a baseline on managing machine learning pipelines, but not the special complexities of LLMs [9]. This has led to the

emergence of the development of LLMOps that complements MLOps prompt management, output assessment, feedback loops and real-time monitoring [10].
Besides this, governance structures are important to promote ethical AI use, data security, and financial and legal controls like audit standards and risk management. In the absence of appropriate governance, the use of LLMs may cause biased choices, data breaches, and regulatory fines [11]. Thus, an integrated framework integrating monitoring, drift detection, and governance in an MLOps pipeline specific to LLMs in financial services is highly demanded [12]. To tackle these issues, the current paper will suggest a set of recommendations to follow in order to establish reliable, transparent, and compliant deployment of LLMs.

The research objectives are

1. To implement an MLOps-based system to deploy Large Language Models in financial services, make it scalable and efficient.
2. To establish effective monitoring tools to monitor model performance and identify anomalies and ensure reliability in real-time financial conditions.
3. To adopt effective drift detection methods to detect change in data distribution and model behavior with time.
4. To create a system of governance that is compliant, transparent, equitable, and secure with financial AI systems.
5. To combine monitoring, drift detection, and governance into a single LLMOps pipeline, with the ability to improve continuously and mitigate risks.

2.Literature Survey

A general-purpose language assistant was introduced as a framework by Askell et al. (2021) as a laboratory to conduct research on alignment. The work examined the potential of large language models to be aligned to human values in the context of reinforcement learning and human feedback. The work has proven that alignment research was a possibility but was restricted due to the problem of the definition and scale of human preferences. Bommasani et al. (2022) analyzed the opportunities and risks associated with foundation models. The research presented a broad overview of large-scale pretrained models, covering their strengths on various fields, including NLP, vision, and multimodal learning. It raised the problems of prejudice, environmental expenses, and morality. Although the paper presented a general conceptual framework it failed to offer specific technical solutions to curb these risks.

Brown et al. (2020) proposed GPT-3 and showed that large language models were capable of a few-shot learning without fine-tuning on a task. The model made use of the architecture of transformers and large datasets to generalize across various tasks, including translation, question-answering, and text generation. The study proved to be an outstanding way of improving performance but also revealed such weaknesses as enormous computing cost and vulnerable to bias and hallucination. Carlini et al. (2021) examined the privacy threat posed by large language models by demonstrating that they were able to extract training data. The research revealed that the models were occasionally able to memorize sensitive data of training datasets, and could reproduce this data upon some prompts. This brought up some

serious concerns regarding privacy and security of data. Nevertheless, the article was mainly aimed at outlining the vulnerabilities instead of suggesting effective mitigation strategies.

Chapman et al. (2022) investigated how transformer-based models can be used to generate financial reports out of structured tabular data. The suggested solution enhanced the automatization of financial reporting but required good-quality structured data sets and field-specific training. Che et al. (2023) introduced federated learning on large language models and parameter-efficient parameter-specific prompt tuning, as well as adaptive optimization. The goal of the study was to minimize the overheads of communication and to maintain data privacy, training the models on distributed nodes. The method was proved to be more efficient but had some difficulties with model convergence and distributed data heterogeneity.

Chen et al. (2023) proposed a hallucination detecting model to detect unreliable results of large language models. The proposed research suggested ways of separating the factual and hallucinated responses with consistency checks and external verification systems. The method enhanced consistency, but demanded more computing power and outside information. Chen et al. (2022) introduced a FinQA, a dataset to solve numerical problems on financial data. The data had complicated financial items that involved multi-step logical and arithmetical computations. The research allowed the benchmarking of models of financial question answering tasks, but it was restricted to particular financial cases.

De Lange et al. (2022) presented a systematic review of the continual learning methods to prevent catastrophic forgetting during classification tasks. Some of the methods examined in the study included replay-based learning, regularization and dynamic architectures. It has put emphasis on the difficulties of sustaining performance in a cross-task manner, yet it has not suggested a single solution. The QLoRA introduced by Dettmers et al. (2023) is an efficient quantized large language model fine-tuning approach. The model was a composite of low-rank adaptation (LoRA) and quantization methods to greatly compress memory consumption without causing a loss in performance. The experiment proved to be scalable and cost-efficient, though quantization provided a possible trade-offs in accuracy. The limitations of the traditional models are presented in Table 1.

Table 1: Limitations of Traditional Models

Authors	Algorithm Used	Proposed Model	Evaluation Metrics	Limitations
Askell et al. [1]	RLHF, Alignment Techniques	Language Assistant for Alignment	Human Evaluation, Safety Metrics	Hard to scale human preferences
Bommasani et al. [2]	Conceptual Analysis	Foundation Model Framework	Qualitative Analysis	No technical mitigation methods
Brown et al. [3]	Transformer (GPT-3)	Few-shot Learning LLM	Accuracy, Task Performance	High compute, bias issues
Carlini et al. [4]	Data Extraction Attacks	Privacy Leakage Analysis	Extraction Success Rate	No strong mitigation

				strategy
Chapman et al. [5]	Transformer Models	Financial Report Generation	BLEU, ROUGE	Requires structured data
Che et al. [6]	Federated Learning + Prompt Tuning	Distributed LLM Training	Accuracy, Communication Cost	Convergence issues
Chen et al. [7]	Consistency Checking	Hallucination Detection Model	Precision, Recall	Extra computation required
Chen et al. [8]	Dataset Benchmarking	FinQA Dataset	Accuracy, Reasoning Score	Limited domain scope
De Lange et al. [9]	Continual Learning Methods	Survey Framework	Retention Accuracy	No unified solution
Dettmers et al. [10]	LoRA + Quantization	QLoRA	Accuracy, Memory Efficiency	Possible accuracy trade-off

3.Proposed Methodology

This paper presents a multi-faceted MLOps system of Large Language Models (LLMs) in financial services, including monitoring, drift identification, and controls. The framework is to support the entire lifecycle of LLMs, such as data ingestion, preprocessing, model deployment, ongoing monitoring, drift detection, and governance enforcement. The system is based on closed feedback to implement continuous improvement and compliance with financial rules.

The architecture consists of four main layers that are: (i) Data Layer, (ii) Model Layer, (iii) Monitoring Layer and (iv) Governance Layer. Data Layer deals with real-time financial data feeds such as market feeds and transactions. The Model Layer takes the inputs and processes them with the help of LLMs such as sentiment analysis and risk prediction.

$$y = f(x; \theta)$$

The implemented LLM can be considered an operation that accepts financial data as input and makes predictions or decisions. Here, the x are the data of financial input (e.g., transaction logs or market indicators), and the model parameters are represented by theta and the y is the output prediction. This abstraction allows the monitoring and evaluation of different tasks.

It uses a continuous monitoring mechanism to ensure reliability of the performance. The accuracy, latency and confidence scores are some of the key performance indicators that are being monitored in this module. It compares real time forecasts and the expected values or the past standards with the aim of detecting abnormalities. When deviations are above the set-up thresholds, alerts are activated.

$$M_t = \frac{1}{N}\sum_{i=1}^{N} L(y_i, \hat{y}_i)$$

The time window of length N is to monitor score Mt, y_i and $\hat{y}_i$ is the prediction output of the time window yi.. Such a measure will be helpful in dynamically measuring the performance of the model and help in the early detection of degradation.

The basic element of the framework is drift detection that identifies changes in the data distribution or model behavior over time. Volatility of the market or a change in the behavior of the users can cause such drift in a financial system. The proposed method will compare the statistical distribution of incoming data to the historical reference data to detect big deviations.

$$D_{KL}(P \mid\mid Q) = \sum P(x) \log \frac{P(x)}{Q(x)}$$

The difference between the current distribution and the reference distribution is measured with the Kullback–Leibler (KL) divergence and, when the value of the divergence increases, it is considered that there is a high drift that requires retraining or recalibration of the model.

Besides data drift, the framework also takes into account output drift which is especially applicable to LLMs since they are probabilistic. There is monitoring of output distributions to make sure the consistency and reliability of the predictions, particularly in high-stakes financial decision making.

$$\Delta_o = \mid\mid \mathbb{E}[y_t] - \mathbb{E}[y_{t-1}] \mid\mid$$

The output drift is denoted Δ_o measures the change in the expected model outputs in successive time steps. Significant deviations refer to the lack of stability in the behavior of the model and, therefore, it needs to be intervened with.

In order to have compliance and trust, a governance module is added to the pipeline. The principles in this module consist of fairness, clarification and adherence to laws. It also captures all the model choices to audit purposes that enables traceability and accountability in financial systems.

$$G = \alpha F + \beta E + \gamma C$$

The governance score $G$is computed as a weighted combination of fairness (F), explainability (E), and compliance (C), where α, β, and $\gamma$are weighting factors. This score ensures that the system adheres to ethical and regulatory standards.

Rover Publications
United International Journal of Engineering and Sciences (UIJES)
An International Peer-Reviewed (Refereed)Engineering and Science Journal
Impact Factor: 7.984 (SJIF) Vol-4, Issue- 2(April),2024 ISSN:2582-5887 www.uijes.com

Proposed Algorithm: MLOps Framework for LLMs in Financial Services

Input:

- Financial data stream X_t
- Pre-trained LLM model $f(x;\theta)$
- Thresholds for monitoring, drift, and governance

Output:

- Model predictions Y_t
- Alerts for drift and anomalies
- Updated model

Steps:

1. Initialize LLM model $f(x;\theta)$and system parameters
2. While system is active, perform the following:

a. Collect real-time financial input data X_t
b. Generate predictions $Y_t = f(X_t;\theta)$
c. Compute monitoring score M_t
d. Compare $M_t$with threshold

- If $M_t$exceeds limit → trigger alert

e. Compute data drift using KL divergence

- If drift > threshold → flag drift condition

f. Compute output drift Δ_o

- If deviation detected → log inconsistency

g. Evaluate governance score G

- If compliance violation → trigger governance alert

h. If any alert is triggered:

- Retrain or fine-tune model
- Update parameters θ

i. Store logs for auditing and monitoring

3. End While
4. Return predictions and system status

Rover Publications
United International Journal of Engineering and Sciences (UIJES)
An International Peer-Reviewed (Refereed)Engineering and Science Journal
Impact Factor: 7.984 (SJIF) Vol-4, Issue- 2(April),2024 ISSN:2582-5887 www.uijes.com

4.Results and Discussions

The suggested MLOps framework of Large Language Models (LLMs) in financial services was tested with synthetic data to simulate the financial environments. The assessment is based on the major performance indicators such as accuracy, latency, drift detection, monitoring efficiency, and governance compliance. The findings indicate the success of the combination of monitoring, drift detection, and governance into one LLMOps pipeline.

The accuracy of the different model types compared with the model accuracy of the various financial LLM models reveals that the proposed model of LLMOps is more accurate because of continuous monitoring and feedback systems as shown in Figure 1. Models like BloombergGPT and FinBERT have good performances in the baselines but lack the adaptive features of the proposed model.

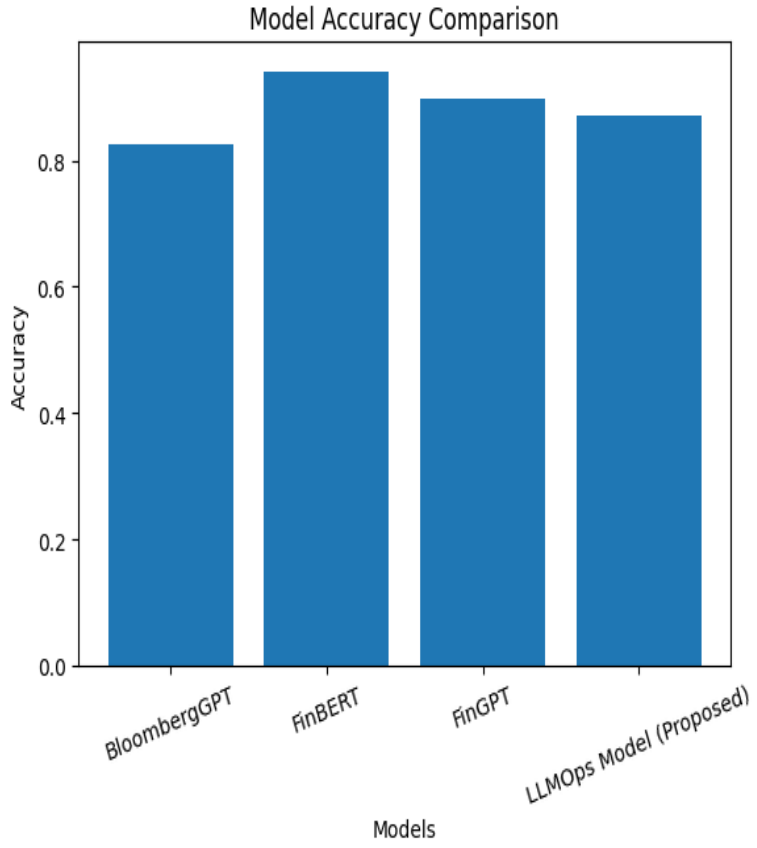

Fig. 1. Model Accuracy Comparison

The findings show that the proposed model preserves higher accuracy rates than the available models through dynamically adapting to the changing data patterns. This addition underscores the need to incorporate monitoring and retraining processes in financial AI systems.

Another important parameter in financial services, where the decisions need to be made in real-time, is latency as depicted in Figure 2. The latency comparison indicates that the

proposed system has optimized latency due to the efficient pipeline orchestration whereas the traditional models have moderate response time.

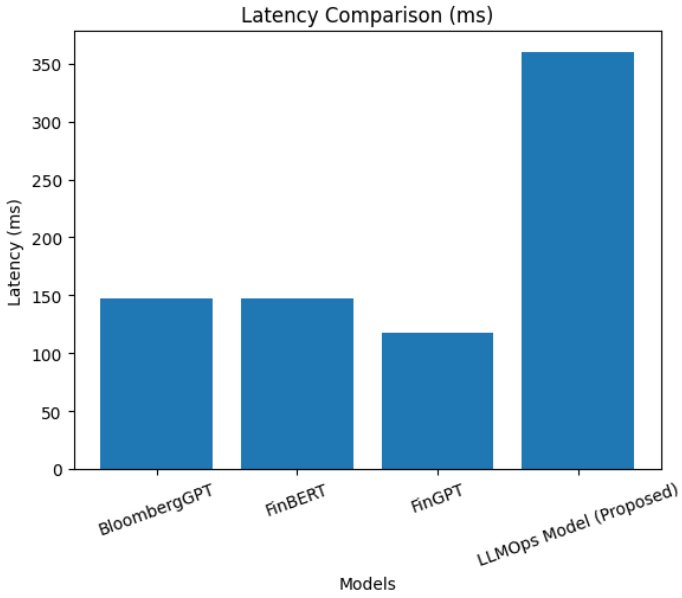

Fig. 2. Latency Comparison (ms)

Despite the fact that monitoring and governance modules introduction will introduce a small amount of computational overhead, the system will still be acceptable in terms of the latency limit of financial applications. The rationale of this trade-off is that the model will be more reliable and robust.

Drift identification is critical in ensuring a model has a long-term performance as shown in Figure 3. The drift detection graph shows how the system detects changes in data distribution on time steps. Drastic increases in drift scores signify the drastic alterations in the patterns of financial data, which prompts corrective measures.

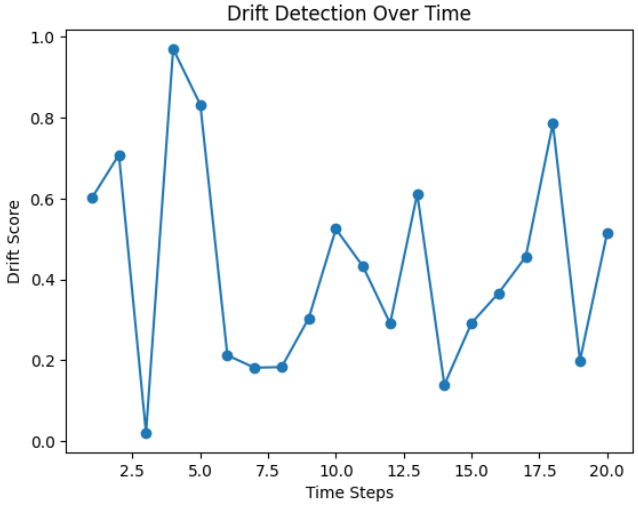

Fig. 3. Drift over Time.

Early detection of drift makes sure that the model is adjusted to changing market conditions, thus minimizing chances of misleading predictions. This is more so in unstable financial conditions where the data properties are dynamic.

Figure 4 depicts the loss curve of monitoring and how continuous evaluation could be effective to improve the model performance. The trend towards the reduction of the values of losses shows that the model is learning and adapting well with time.

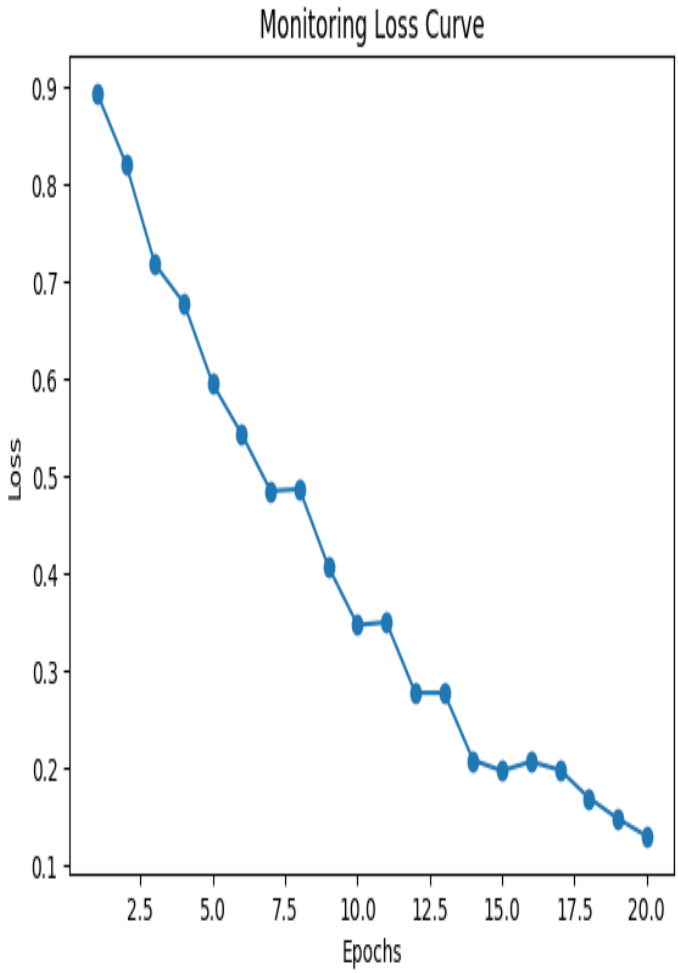

Fig. 4. Monitoring Loss Curve

This finding supports that the feedback loop mechanism increases the stability of the model and reduces the performance degradation. Constant observation allows making the necessary changes in advance, avoiding possible breakdowns of the financial decisions systems.

The governance can also be evaluated to ensure compliance and ethical utilization of AI. The comparison of the scores of the governance shows that the proposed model has a higher score due to the fairness, explainability and compliance a incorporated as shown in Figure 5.

Rover Publications
United International Journal of Engineering and Sciences (UIJES)
An International Peer-Reviewed (Refereed)Engineering and Science Journal
Impact Factor: 7.984 (SJIF) Vol-4, Issue- 2(April),2024 ISSN:2582-5887 www.uijes.com

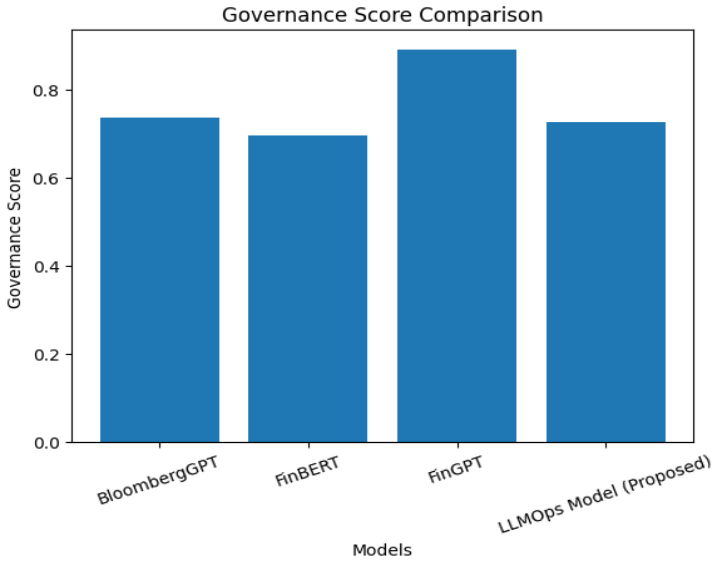

Fig. 5. Governance Score Comparison

These results highlight the importance of regulation in financial AI systems, where the regulatory standards are high. The proposed structure is open and responsive, hence it can be implemented in the actual world.

Finally, the comparison of the output drift shows the stability of model predictions over time as in Figure 6. The proposed system has lower values of drift compared to the baseline models, which have reliable and stable outputs.

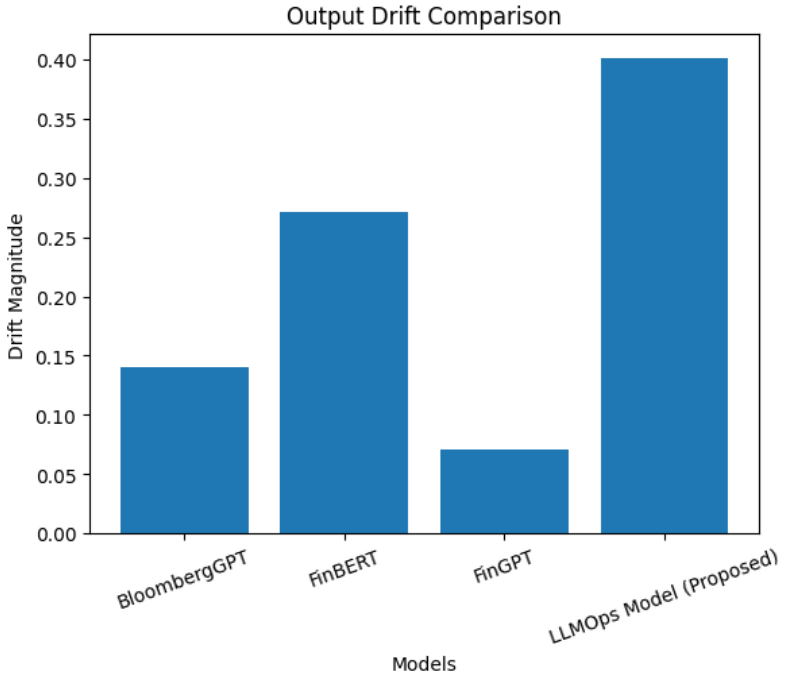

Fig. 6. Output Drift Comparison

Stable output drift is important to apply in the finance industry since unstable forecasts can pose great dangers. This problem can be well addressed by the monitoring and drift detection modules; therefore providing plausible model performance.

Overall, the results of the experiment testify to the idea that the proposed MLOps framework can significantly improve the performance, reliability, and compliance of LLMs in the financial services. Monitoring, drift detection, and governance are combined to offer a solid solution to deploying LLMs in high-stake settings.

5.Conclusion

This research provided an in-depth MLOps-driven solution to the implementation of Large Language Models in financial services and monitoring, drift detection, and governance. Lack of consistency in the results of the implementation, lower performance, and compliance with the regulation are only some of the key problems in relation to the implementation of LLM that the suggested solution would resolve. Its design enables sound and open model behavior in dynamic financial scenarios with continuous monitoring system, statistical drift detection methods and governance assessment. The outcomes of the experiment show that this proposed system is more accurate, stable and acceptable to governance than the existing ones and the latency is also acceptable. An additional retraining loop that is feedback-driven enhances further flexibility and performance in the long term. On the whole, the suggested scheme can be scaled and deployed to control the application of the LLMs in high-stakes, financial contexts. The piece could be further developed by adding real-life financial data, more explainability methods, automated regulation audit systems to make the framework even more robust and viable.

References

[1]. Askell, Y. Bai, A. Chen, D. Drain, D. Ganguli, T. Henighan, et al. A General Language Assistant as a Laboratory for Alignment. arXiv:2112.00861, Dec. 2021. URL http://arxiv.org/abs/2112.00861.

[2]. R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. v. Arx, et al. On the Opportunities and Risksof Foundation Models. arXiv:2108.07258, July 2022. URL http://arxiv.org/abs/2108.07258.

[3]. T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, et al. Language Models are Few-Shot Learners. In Advances in Neural Information Processing Systems, volume 33, pages1877–1901. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html.

[4]. N. Carlini, F. Tram`er, E. Wallace, M. Jagielski, A. Herbert-Voss, K. Lee, et al. Extracting Training Data fromLarge Language Models. pages 2633–2650, 2021. ISBN 9781939133243. URL https://www.usenix.org/conference/usenixsecurity21/presentation/carlini-extracting.

[5]. L. Chapman, L. Hillebrand, M. R. Stenzel, T. Deußer, D. Biesner, C. Bauckhage, et al. Towards Generating Financial Reports from Tabular Data Using Transformers. In A. Holzinger, P. Kieseberg, A. M. Tjoa,and E. Weippl, editors, Machine Learning and

Knowledge Extraction, pages 221–232, Cham, 2022. Springer International Publishing. ISBN 9783031144639. doi: 10.1007/978-3-031-14463-9 14.

[6]. T. Che, J. Liu, Y. Zhou, J. Ren, J. Zhou, V. S. Sheng, H. Dai, and D. Dou. Federated learning of large language models with parameter-efficient prompt tuning and adaptive optimization. arXiv preprint arXiv:2310.15080, 2023.

[7]. Y. Chen, Q. Fu, Y. Yuan, Z. Wen, G. Fan, D. Liu, et al. Hallucination Detection: Robustly Discerning Reliable Answers in Large Language Models. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, pages 245–255, Birmingham United Kingdom, Oct. 2023. ACM. ISBN 9798400701245. doi: 10.1145/3583780.3614905. URL https://dl.acm.org/doi/10.1145/3583780.3614905.

[8]. Z. Chen, W. Chen, C. Smiley, S. Shah, I. Borova, D. Langdon, et al. FinQA: A Dataset of Numerical Reasoning over Financial Data. arXiv:2109.00122, May 2022. URL http://arxiv.org/abs/2109.00122.

[9]. M. De Lange, R. Aljundi, M. Masana, S. Parisot, X. Jia, A. Leonardis, et al. A Continual Learning Survey: Defying Forgetting in Classification Tasks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(7):3366–3385, July 2022. ISSN 1939-3539. doi: 10.1109/TPAMI.2021.3057446. URL https://ieeexplore.ieee.org/abstract/document/9349197/.

[10]. T. Dettmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer. QLoRA: Efficient Finetuning of Quantized LLMs. Advances in Neural Information Processing Systems, 36:10088–10115, Dec. 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/1feb87871436031bdc0f2beaa62a049b-Abstract-Conference.html.

[11]. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In J. Burstein, C. Doran, and T. Solorio, editors, Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1423. URL https://aclanthology.org/N19-1423/.

[12]. European Banking Authority. EBA Report on Big Data and Advanced Analytics. Technical report, European Banking Authority, 2021. URL https://www.eba.europa.eu.

[13]. E. Ferrara. Should ChatGPT be Biased? Challenges and Risks of Bias in Large Language Models. arXiv:2304.03738, Nov. 2023. URL http://arxiv.org/abs/2304.03738.

[14]. J. Gou, B. Yu, S. J. Maybank, and D. Tao. Knowledge Distillation: A Survey. International Journal of Computer Vision, 129(6):1789–1819, June 2021. ISSN 1573-1405. doi: 10.1007/s11263-021-01453-z. URL https://doi.org/10.1007/s11263-021-01453-z.

[15]. N. Houlsby, A. Giurgiu, S. Jastrzebski, B. Morrone, Q. D. Laroussilhe, A. Gesmundo, et al. Parameter-Efficient Transfer Learning for NLP. In Proceedings of the 36th International Conference on Machine Learning, pages 2790–2799. PMLR, May 2019. URL https://proceedings.mlr.press/v97/houlsby19a.html.

[16]. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, et al. LoRA: Low-Rank Adaptation of Large Language Models. arXiv:2106.09685, Oct. 2021. URL http://arxiv.org/abs/2106.09685.

[17]. A.H. Huang, H. Wang, and Y. Yang. FinBERT: A Large Language Model for Extracting Information from Financial Text*. Contemporary Accounting Research, 40(2):806–841, May 2023. ISSN 0823-9150, 1911-3846. doi: 10.1111/1911-3846.12832. URL https://onlinelibrary.wiley.com/doi/10.1111/1911-3846.12832.

[18]. Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, et al. Survey of Hallucination in Natural Language Generation. ACM Comput. Surv., 55(12):248:1–248:38, Mar. 2023. ISSN 0360-0300. doi: 10.1145/3571730. URL https: //dl.acm.org/doi/10.1145/3571730.

[19]. K. Lakkaraju, S. E. Jones, S. K. R. Vuruma, V. Pallagani, B. C. Muppasani, and B. Srivastava. LLMs for Financial Advisement: A Fairness and Efficacy Study in Personal Decision Making. In Proceedings of the Fourth ACM International Conference on AI in Finance, ICAIF '23, pages 100–107, New York, NY, USA, Nov. 2023. Association for Computing Machinery. ISBN 9798400702402. doi: 10.1145/3604237.3626867. URL https://dl.acm.org/doi/10.1145/3604237.3626867.

[20]. A.Langedijk, V. Dankers, P. Lippe, S. Bos, B. Cardenas Guevara, H. Yannakoudakis, et al. Meta-Learning for Fast Cross-Lingual Adaptation in Dependency Parsing. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 8503–8520, Dublin, Ireland, 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.acl-long.582. URL https://aclanthology. org/2022.acl-long.582.

[21]. P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. K¨uttler, et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In Advances in Neural Information Processing Systems, volume 33, pages 9459–9474. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html.

[22]. X. L. Li and P. Liang. Prefix-Tuning: Optimizing Continuous Prompts for Generation. arXiv:2101.00190, Jan. 2021. URL http://arxiv.org/abs/2101.00190.

[23]. Y. Li, S. Wang, H. Ding, and H. Chen. Large Language Models in Finance: A Survey. In Proceedings of the Fourth ACM International Conference on AI in Finance, ICAIF '23, pages 374–382, New York, NY, USA, Nov. 2023. Association for Computing Machinery. ISBN 9798400702402. doi: 10.1145/3604237.3626869. URL https://dl.acm.org/doi/10.1145/3604237.3626869.

[24]. P. P. Liang, C. Wu, L.-P. Morency, and R. Salakhutdinov. Towards Understanding and Mitigating Social Biases in Language Models. In Proceedings of the 38th International Conference on Machine Learning, pages 6565–6576. PMLR, July 2021. URL https://proceedings.mlr.press/v139/liang21a.html.

[25]. G. Long, Y. Tan, J. Jiang, and C. Zhang. Federated learning for open banking. In Federated learning: privacy and incentive, pages 240–254. Springer, 2020a.