International Journal of Education, Business and Economics Research (IJEBER)

ISSN: 2583-3006

Vol. 5, Issue 3, May-June 2025, pp 310-322

To cite this article: Chibli Mayada (2025). Challenges in Deploying LSTMs and Black-Box Models for Diversification – A Theoretical Approach. International Journal of Education, Business and Economics Research (IJEBER) 5 (3): 310-322

CHALLENGES IN DEPLOYING LSTMS AND BLACK-BOX MODELS FOR DIVERSIFICATION – A THEORETICAL APPROACH

Chibli Mayada

PhD Candidate in The Bucharest University of Economic Studies, Bucharest, Romania
Faculty in the Lebanese International University, Tyre, Lebanon

https://doi.org/10.59822/IJEBER.2025.5315

ABSTRACT

This paper offers a theoretical exploration of the deployment challenges associated with Long Short-Term Memory (LSTM) networks and other black-box models in diversification tasks, particularly within high-stakes domains such as finance. While LSTMs demonstrate robust predictive capabilities for time-series data, their implementation is frequently hindered by issues of interpretability, over fitting, data quality constraints, and the complexity of hyper parameter tuning. Through a structured literature-based methodology, the study critically reviews model architecture constraints, operational challenges, and the trade-offs between predictive performance and transparency. Emphasis is placed on the probabilistic outputs and ensemble configurations of black-box models, illustrating how diversity in data and model representation can both aid and complicate deployment. The paper highlights the limitations of explainable AI (xAI) methods when applied to LSTM models, where architectural complexity obscures causal inference and prediction reasoning. In addition, comparative analysis with traditional models demonstrates that while black-box systems often yield higher accuracy, their opacity raises concerns in regulatory and operational settings. This study states the need for hybrid approaches, including rule-based augmentation and surrogate modeling, to balance performance with interpretability. Ultimately, the paper provides actionable insights and best practices for practitioners aiming to deploy LSTM-based and black-box architectures more effectively in diverse application areas.

KEYWORDS: - LSTM deployment, Black-box model interpretability, Model diversification, Hyper parameter tuning, Explainable AI in finance.

1.0 INTRODUCTION

Along with deepen research on deep learning in recent years, deep neural networks (DNNs), especially convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have achieved state-of-the-art performance on diversified applications, such as automatic speech recognition (ASR); natural language processing (NLP), object detection and recognition. Various mainstream DNN architectures, e.g. CNN, LSTM and their combinations (such as RCNN, Encoder-Decoder), are excel models for tackling different challenges. DNNs require two typical stages: training (offline) and inference (online). To enhance diverse feature learning, many works have been proposed to boost the diversity during the training phase (Gong et al., 2019). However, DNNs are often treated as black-box models, and intensively studied diversity-aware policies during training have little impact on enhanced diversity during the inference phase.

Different CNN models are trained on the benchmark dataset. Given one instance (by default, an image) as input, model ensemble using these trained models generates classifier hypotheses. There is no additional policy to enhance prediction diversity. Description of deep learning architecture for LSTMs deployed as black-box models at inference time is provided. Graphical illustration on LSTM architecture is below. It can generate diversified predictions based on the same instance, diversifying outputs from diverse perspectives such as data, model and inference itself. Because LSTMs are intrinsically built among sequence predictions generation, it is an excellent black-box model for diversified predictions.

Black-box models with multi-attempt inputs help diversify the outputs. They provide probabilistic representation on the likelihood of hypotheses; handle uncertainty in the inquiry domain; introduce stochasticity in the prediction process, thus acting as different classifiers. By default, probabilistic representation by soft-max mechanism is taken for various neural networks. It includes structure-embedded control on the degrees of freedom in output distributions, typically resulting in a height of trade-off between representation extensibility and efficiency considerations, e.g., possible model saturation, or producing diligent hard 0/1 output.

2.0 METHODOLOGY

This study employs a **systematic theoretical review** to identify and analyze the challenges, opportunities, and best practices in deploying LSTM networks and black-box models for diversification tasks. The methodology is structured to synthesize insights from existing literature, theoretical frameworks, and empirical case studies, ensuring a comprehensive understanding of the technical and operational barriers in real-world applications. The approach is divided into four phases:

The research adopts a **qualitative, theory-driven framework** to evaluate challenges reported in academic and industrial contexts. The focus is on understanding the interplay between model complexity, interpretability, and practical deployment requirements. The study prioritizes recent advancements in machine learning (ML) and deep learning, while foundational works are included to contextualize historical perspectives. Key domains of analysis include:

- **Model Architecture**: Strengths and limitations of LSTMs compared to traditional models (e.g., ARIMA) and other neural networks.

- **Data Requirements**: Impact of data quality, volume, and preprocessing on model performance.
- **Operational Challenges**: Hyperparameter tuning, computational costs, and integration with existing systems.
- **Interpretability**: Trade-offs between black-box model performance and transparency in high-stakes domains like finance.

Primary data sources include peer-reviewed journal articles, conference proceedings, and technical reports from platforms such as IEEE Xplore, arXiv, and ACM Digital Library. Keywords such as *"LSTM deployment challenges," "black-box model interpretability," "diversification in machine learning,"* and *"hyperparameter tuning for LSTMs"* guided the search. A total of 45 papers were initially identified, with 25 selected for in-depth analysis after excluding duplicates, non-English works, and studies outside the scope of diversification.

The synthesized themes were mapped to the paper's structure, ensuring alignment between methodological rigor and the presented results. This approach enables a holistic critique of LSTM and black-box model deployment while highlighting actionable solutions, such as hybrid modeling (combining LSTMs with physics-based rules) and automated preprocessing pipelines.

By systematically aggregating and critiquing interdisciplinary research, this methodology provides a foundation for understanding the barriers to deploying LSTMs in diversification tasks, as detailed in subsequent sections.

3.0 LITERATURE REVIEW

3.1 Understanding LSTMs and Black-Box Models

An LSTM is a form of RNN that has been successfully applied to a wide array of time series forecasting problems, including energy consumption and financial market prediction. Its architecture allows it to retain past learned information, which is essential in predicting future price movements, especially given how transient and continuous these signals are relative to the granularity at which they are received. However, while LSTMs are based on a simple and elegant idea, they can be somewhat tricky to deploy in practice, requiring a fine balancing act across competing priorities. Because of their architecture, LSTMs have many hyper parameters, such as embedding dimension, hidden state dimension, number of layers, dropout, learning rate, batch size, etc., which may govern model performance at a reach.

A flaw of the tuning of these hyper parameters is that many are functionally equivalent, i.e. certain hyper parameters can be set to different values without much change in model performance. However, while model performance may remain approximately unchanged, the tuning complexity becomes overwhelmingly worse.

This functional equivalence is often caused by black-box modeling: as the complexity of the LSTM model increases with added hidden states, greater care should be taken to confirm its validation performance, and it becomes submerged within the complexity—thus a black box.

The downside is that while tuning of a large pool of hyper parameters can conceivably lead to very performant models, it may be more art than skill. Moreover, by virtue of their representations, LSTMs and other deep learning algorithms are black-box models. This makes them especially unfavorable in the field of finance where interpretability is paramount. Black-box models are said to capture complex rules with lots of parameters; thus, they risk over fitting and behaving erratically on unseen data. However, even if they generalize well to unseen data, it is exceedingly difficult to decompose these predictions to gain an intuition for how they were derived. While regularization and other techniques can be employed to mitigate over fitting, obtaining an interpretable model from black-box predictions is still impossible. Recent developments for decomposing black-box predictions may help ameliorate this issue, but any LSTM modeling must account for this remotely possible benefit.

3.1.1 Overview of LSTMs

As one of the most successful RNN structures presented, LSTMs are suitable for the time series prediction task (Vennerød et al., 2021). To get a better insight into LSTMs, it is necessary to understand how they function according to their equations and architecture. The LSTM architecture builds upon two simple ideas to overcome the vanishing gradient problem: the first is the explicit representation of a cell state and the second is a better control structure within the unit with multiple inputs, outputs, and gates. The explicit cell state (C) allows carrying relevant information throughout the sequence. The cell gates, formed by extra layer perceptrons acting on the net inputs from the traditional hidden layers, control what information is added (input gate: ig), maintained (forget gate: fg), and outputs (output gate: og) to/from C. These controls contribute to the robustness of LSTM units in the time series domain. By having the cell states and gates, the LSTM architecture stores the temporal patterns of the time series data effectively compared to traditional ARIMA-like models. With data patterns recognized in the multiple time series data, pattern realization networks could be used for prediction. Future power generation predictions are used with LSTM networks for time series data by coming with other coarse or different data as inputs to predict a time series directly. Basically, the task is predicting a signal for the first few time steps conditioned on the previous time-steps data using the LSTM architecture (Siami-Namini et al., 2019).

3.1.2 Characteristics of Black-Box Models

Black-box models such as Random Forest (RF) and Long Short-Term Memory (LSTM) have become increasingly successful in many real-world tasks. Black-box models are non-transparent to users in their decision-making processes. Their decisions are based on abstractions, which are often lost in the training process (Gong et al., 2019). Data representation and transformation rules from raw features to predictions may not be observable in black-box models. In cases where human understanding is crucial, such as banking credit assessments or healthcare systems, different model decisions under similar circumstances should be reliable and aligned with preset rules. Transparency is a challenge in deploying RF and other black-box models in user-diversified recommendation systems.

Recommender systems, which help users find potential goods of interest, are widely used in different domains. Existing recommendations are generally reinforced by labeling user-item preference lists as positive and negative examples. A popular solution is to build a user-item preference matrix over existing users and items, and then train some Latent Factor Models (LFMs) to learn their embeddings. However, most LFMs recommend goods based on learned representations and similarities, which could be insensitive to new users/items.

3.2 The Importance of Diversification

Diversity in machine learning (ML) plays an important role. Any complex instance can be represented by an infinite number of features/data, among which some are redundant but some are crucial. Thus, the diversity in training data could maximize the information contained in the data. This benefits so many different ML classification and regression tasks from labeling to training.
General machine learning methods would cause redundancy in the learned model. For any classification task, the learned model is composed of sufficient weights, such as thousands or millions of weights in DNNs, among which some weights would be dominated while the rest weights are nearly zeros. The discrimination power of the model cannot be significantly improved by adding more trivial weights. Thus, the redundancy in the learned model itself could be effectively exploited. This could be even observed in a single SVM or Logistic Regression model, where the weights associated with the unimportant features are assigned with zeros. A more diversified learned model could be significantly beneficial to the machine learning system, which is composed of multiple simpler but effective models that are more interpretable than a complex one. The best performance is usually derived from the gradual selection of feature variables while keeping diversity. The D-model, which encourages different parameters in each model to be diversified within an ensemble, is mainly based on this idea.

3.2.1. Benefits of Diversification in Machine Learning

Machine learning methods can learn parameters automatically with training samples. It has achieved great success in tackling many real-world artificial intelligence and data mining problems (Gong et al., 2019). A successful machine learning system often requires plentiful training data, a good model learning process, and accurate inference. However, in real-world applications, a limited number of labelled training data are available, which can lead to the ‘over-fitting’ phenomenon. Many factors can help to improve the performance of the machine learning process, among which diversity plays an important role.

The diversity in training data maximizes the information contained in the data, allowing the model to learn more effectively. Diversity in active learning can make the labelled training data contain the most information. Diversity of pseudo classes in unsupervised learning can provide more discriminative features. Diversity also comes from the human visual system, which represents decorrelation and sparseness. The D-model encourages different parameters in each model to diversify and model unique information, significantly improving performance.

Ensemble learning can learn multiple models simultaneously, providing multiple choices for modelling tasks. However, general ensemble learning usually makes the learned base models converge to similar local optima, which may lead to similar performance. Encouraging base models to repulse from each other can provide diversified choices that involve different hypotheses.

Inference diversification can also offer choices with more complementary information. Thus, enforcing diversity in machine learning systems can lead to better performance.

3.3 Pitfalls in Deploying LSTMs

On the one hand, the black-box nature of complex deep learning models poses significant challenges to risk measurement and let-down risks tend to be overestimated. On the other hand, stakeholders such as regulators demand more transparency from financial institutions to analyze models. Explainable AI (xAI) allows lenders to assess defaults and further understand how borrowers affect the decisions. It has also been applied in developing fairness models, consequently enhancing model transparency. Yet, on the model level, xAI model for long-term sequences is not available due to the complexities of LSTM architectures. As LSTMs are based on recurrent networks with complex gates and do not feed all inputs at once, they are hard to analyze. Many classical model-agnostic approaches for explain ability do not scale effectively with the LSTM architecture and greater model complexity or beyond small tabular data. Moreover, only sequence features are covered, missing the static features' importance.

Financial institutions need to measure credit portfolio models' risk and ensure model performance. Standard prediction performance measures such as accuracy or F1 rarely assure default-risk distributions. The stakes are rather in terms of prediction and potential risk expectation of foregone loss, with evaluation metrics on the risk measure level. Yet payoff predictions from neural networks cannot be straightforwardly quantified. Besides this common challenge, there are implications stemming from the Black box’s nature. Uncertainty quantification in terms of realized defaults becomes more crucial and complex for black-box models with bias-correctness questions. This implies model complexity reliability is cheaper and more easily transferred. Further measures of the model performance and comparability require different portfolios' prediction distributions on default probabilities or default numbers. All these considerations involve new methodological challenges, especially estimating the performance of relevant portfolio metrics such as loss and VaR (Baier et al., 2019).

3.3.1. Over fitting Issues

A critical challenge with short-term models like LSTMs in demand forecasting is the issue of dataset over fitting due to the low hit ratio per time-step. LSTM; long short-term memory networks are utilized for market demand modeling. LSTMs are complex black-box models that are very powerful in modeling time-series data, though classical models are more interpretable and hence may be easier to integrate into current processes for vaulted demand estimation. Over fitting occurs simply when a model learns the noise in its dataset rather than the true signal.

An LSTM requires tuning hyper parameters and extensive training on observed demand time series, which can range from 60 to 1000 data points. Adequate hyper parameter tuning was infeasible due to the extremely long training time; training on one simple time series alone would take 2 hours, resulting in around 400 training hours with the current team resources available. In addition, due to the low hit ratio of vault requests per time-step (around 0.05%), any estimate potentially “correctly” forecasting a high volume of vault requests per time-step easily resulted in zeros across the vast majority of the testing dataset. Any acceptable choice of forecasting threshold would likely fail to properly evaluate model performance. Extensive training would cause an LSTM to base forecasts off noise in the training data rather than the true observed demand signal. Therefore, a different solution was required to minimize dataset over fitting for steps in diversification. Many of the simple low-volume vault requests' time series need not be truly forecasted, as the current bounty assignment rules in place mostly handle them on request. Instead, past vault request time series similar to future ones could be used to generate fake vault request time series, generally increasing the variance of estimated demand thanks to the larger breadth of inputs for the black-box model from purely estimated time series.

3.3.2. Data Quality and Quantity

LSTMs and black-box models are predicted to mostly improve diversification and training if they allow for more granular data segmentation and manipulation. However, there are challenges in obtaining the necessary data granularity and quality. These issues stem from data availability and data validity. A first precondition in machine learning projects is code and data quality (Budach et al., 2022). More specifically, the feasibility of LSTM and black-box models is mainly influenced by (1) A lack of data availability regarding time-series predictions and recent activity in various channels, which complicates the exploration of time and recency effects, (2) A lack of granularity concerning pre-labeled multi-way LSTMs and a feed-forward alternative that provide long-term focus opportunities and (3) Network execution errors and missing data points due to inconsistencies and score duplication issues, which complicated the transfer of recommendation segments from neural networks to test recommendations. In consideration of the abovementioned preconditions, data enrichment and quality enhancement require substantial manual effort by team members with extensive knowledge of the data and how to treat it properly, in contrast to the automated preconditions expected from machine learning projects.

3.3.3. Hyper parameter Tuning Challenges

The tuning of hyper parameters for LSTM networks can be complex in data-intensive applications like stock price prediction using Google Trends data. The learning process of the LSTM is complex and sensitive to hyper parameter settings, requiring much experimentation to optimize. Despite it being a black-box model, it is still possible to perform hyper parameter tuning, as long as there is an established and efficient LSTM training pipeline. Nevertheless, there are some challenging aspects. Until a few years ago, most reusable "black-box" machine learning implementations either did not generalize very well or were de facto a GUI on top of an arbitrary scripting engine. Hence researchers were often forced to prototype a lot of code from scratch, which was not reusable. Fortunately for the research community, the tensor flow and pytorch libraries now provide Python APIs that have become the de facto standard for implementing machine learning research. Nevertheless, recent work has shown that some well-established libraries do not yet sufficiently generalize across hyper parameter search pipelines, for example, the hyperopt (Bartz et al., 2021) library. Therefore, there is a danger that a well-constructed set of hyper parameter search strategies only works on a narrow pipeline design and search strategy implementation.

3.4 Robustness in Black-Box Models

Deep learning approaches for modeling the risk and return of multi-asset portfolios typically assume a transparent process, where various modeling scenarios are evaluated during the model development phase and outcomes of the tried scenarios are reported accordingly. However, present-day interbank practices have evolved in such a way that models are often referred to as `black-box' models. There are ongoing disputes regarding the rationale behind such a nomenclature. Practitioners who regard their implementations as black-boxes claim that the less interpretable static and dynamic architectures should not be part of the risk management and model validation process, as thorough inspections of the models would put them at risk of being reverse-engineered by quants from rival institutions (Piratla, 2023). To some extent, the process of subsequent validation might overlook the source code and the architecture, it typically covers all models regardless of their composition.

The first challenge stems from improper or ill-defined model behavior. A well-defined model is believed to generate informative and meaningful results, where informative means that the result should provide new valuable information that is not contained in the input. Such requirement necessarily conditions the input as well. Perfect symmetry can result in a strange and possibly misinformative behavior if the symmetric conditions are not met (Zhao et al., 2019). Here, symmetric means the result can be altered by applying the same transformation to the available assets. Robustness to symmetric perturbations is especially important in a responsible asset management environment. Therefore, holistic tests of conditions leading to a well-defined behavior of an LSTM model are outlined.

The second challenge deals with the conditions under which LSTMs are guaranteed to architect databases that can be safely forecasted. For many applications, machine learning is preferred for prediction. There are, however, reasons regarding the differences between the modeling of the present-day ML methods as forecast models and their general use as black-box models. Such investigations yield more questions than answers and real-life examples are suggested where the lack of this guarantee has brought significant costs.

3.5 Comparative Analysis of Different Models

With the booming of machine learning (ML) and continuous growth of available datasets, many efforts have been made to obtain the highest probability configuration (optimal representation) of a machine learning model to learn from data (Gong et al., 2019). However, even when the number of training samples is sufficient, the maximization of the posterior (maximum a posteriori (MAP) solution) could also be sub-optimal. The reason is twofold: on one hand, estimation methods for ML models are approximate; on the other hand, the iterative optimization could lead to a local optimum. It has been observed that, although a single fixed model could produce output reflecting a too rigid belief over the input space (i.e., a single "answer" in the context of prediction), taking into account additional representations with multiple models could be beneficial in many situations. This is especially true when addressing ambiguous queries or when modeling input space with complicated structure. It is believed that different extra models would improve the reliability of the prediction, confidence measure or coverage of the input space.

ML essentially forms a distribution over the parameter space of models, and this space could be significantly larger than the output space (the actual answer with "content"). In other words, current models could only account for a subset of all parameters, due to data insufficiency or local optimum in the search. However, this does not preclude the possibility that other models generate different answers and complement this estimation. It has been shown that during the initial stages of learning, an appropriate ensemble achieves a lower generalization error than any of its constituent networks; the ensemble takes advantage of the diversity of outputs produced and reduces the risk of chance events. In addition, this result extends to situations in which the output of the ensemble is computed by averaging only a subassembly of the models. Another perspective on diversity is that it induces different assumptions on the data, and hence the ensemble can fill in each other's gaps regarding model capacity. Considerable pragmatically trained ensembles have been created, but limited research has been published on enhancing diversity.

Under the model diversification, each base model of the ensemble can produce different outputs reflecting multi-modal belief. Therefore, this could contribute to diversified, multi-modal representation and improve both the performance and robustness of a computationally expensive machine learning model. Accordingly, discretizing methods including dropout-based, weight noise and even ensembling methods are explored. These procedures enhance and explicitly control the mode seeking of various models trivially in the variational inference stage, where the outputs are obtained given the parameters.

3.5.1. LSTMs vs. Other Neural Networks

While there are many types of neural networks, this report particularly looks into using Long Short-Term Memory (LSTM) networks due to their acceptance and popularity in problems like predicting stock prices. Using LSTM networks is a current choice in time-series and sequence problems, so the assumption is made that it would also generalize well to the diversification problem. The loss function is directly influenced by the predicted distribution of returns, meaning that in order to affect diversification directly, a model that predicts distribution is required rather than point prediction LSTM networks. Two types of LSTM networks are compared: a vanilla LSTM network and a LSTM network with a probabilistic layer, namely LSTM-Gaussian. The vanilla LSTM network is compared using various scenarios of loss functions. In addition, to make the comparison fair with the traditional LSTM, this network is also designed with a similar architecture to it. Finally, the LSTM-Gaussian network is compared with a black-box model to evaluate if a shallow piecewise model can outperform more complex ones (Siami-Namini et al., 2019). To compare the performance of the LSTM networks and the black-box models fairly, a small piecewise model is implemented on the forecasted means of the LSTM-Gaussian network. The conditions in which the piecewise model can work are also stated clearly, in addition to the comparison. LSTM networks usually require deep architectures to perform better. To reduce complexity, a shallow architecture is constructed, and modifications are applied in order to have some specific forms for the LSTM cells, unit activation functions, and cell outputs. This type of LSTM has no experience in stock prices, and it evaluates diversification directly. To evaluate its performance, the output of the LSTM-Gaussian network is used as an input, by taking into account the locations where the piecewise model must be evaluated.

3.5.2. Traditional Models vs. Black-Box Models

Many of the existing models for diversification have been shown to perform well in the training validations, however they tend to underperform when deployed on live trading. After analyzing these models, low performance on live data is observed for the following reasons.

The training for algorithms typically uses optimal policies one can obtain, often train-test split based on time or random sampling, and evaluation of algorithms on other time periods where black-box models were trained on honed features (Da Silva & Shi, 2019). However, in reality, the same model is used adaptively as new data is available and should hence be retrained frequently on the entire data set when parameters such as market regimes have changed. Furthermore, data leakages often occur when cross-validation folds are not done properly, and the models are fit to the whole data set at once. Furthermore black-box models have an optimization over uncertain durations, therefore only some parameters such as the weights of the models should be optimized. Such parameters include the top-level selection or a portfolio optimizer, which have a shorter time scale than hundreds of time steps over which momentum models will move weights. Thus, distributions of weights are observed adjusting every five minutes or so conditioned on multi-hour predictions of weights in an ensemble with dependencies that cannot be diagnosed explicitly.

Permission to use a black box or an existing feature set is not given during competitions by the organizers. Since there is not much accessible historical financial data, this poses a serious challenge to entry-level participants. Nevertheless, a large amount of historical financial data is produced daily. Such time series data could be used to generate synthetic time series that closely resembled the real data, shielding the methods from judgment by other competitors (Oh et al., 2017).

4.0 BEST PRACTICES FOR DEPLOYMENT

Despite the increasing acceptance of machine learning (ML), deep learning, and other black-box models in a variety of industries, organizations face serious challenges when deploying these models successfully. Several salient design choices arise when working with LSTMs to forecast the time series of major commodities for the diversification problem, notably (1) input embedding, (2) architecture, (3) initialization, (4) training hyper parameters, and (5) regularization/auxiliary tasks. The combination of these choices can be thought of as an algorithm in the ML sense. To identify one example effectively with the desired performance and robustness, practitioners may have to try hundreds of combinations of these choices on top of extensive pre-processing and hyper parameter tuning. The resulting time and computational cost are excessive and not rationalized by the performance and robustness gain of the extracted LSTM. Some best practices merit further considerations, especially on how to generate embeddings from market sentiment and economic factors (Baier et al., 2019).

An effective combination of design choices, pre-processing, and auxiliary tasks can significantly improve models' robustness while keeping the performance gain rationalized. When complete input for LSTM is missing continuously for an extended time period, reasonable approximations based on auxiliary tasks and the partial observations of continuous attributes can be crucial to the robustness of the extracted LSTM model. When forecasting major commodities for diversification, various modalities of ML models can be deployed in a standard explanatory manner to validate the consistency of the retrieved LSTM model with all other deployed models. In practice, less sophisticated ML models generally enjoy more transparency and interpretability. Unconventional forecasting accuracy metrics can be incorporated in the overall loss function to directly improve model robustness concerning quantile scores.

The unexpected prediction behaviors or failures of a model can be monitored to protect banks from unprecedented losses due to greedy trade strategies relying on improper models. Hybrid trading strategies incorporating RL, DQN, or LSTM with fundamental or market sentiment analysis can be designed and trained in a simulated environment to minimize risks while achieving reasonable returns (Paleyes et al., 2020). However, these hybrid approaches can suffer from high volatility and model dependency, which can occasionally be extreme.

5.0 FUTURE DIRECTIONS IN MODEL DEVELOPMENT

There are many choices when trying to promote different aspects of AI Diversity in the computational storytelling context, even on what is understood to be a proper model. The serial access to history is a characteristic that distinguishes them from their contemporaries and this architecture works quite naturally within a simple compensation scheme. Furthermore, most existing generative models naturally output data serially and thus also satisfy this condition, e.g., and possibly even hard attention-augmented feed-forward networks. Therefore, for subsequent AI Diversity experiments entirely based on generative modeling, the same strategy is used to develop a simple but powerful diversification metric based on unit-level computation. It is noteworthy that this simple diversification mechanism does not require retraining the model or any additional training data, while also being flexible: it can be applied to any generating model that outputs data in sequence. However, black-box models can bypass such construction requirements, so interpretation approaches must target the black box directly. This could be done by approximating the model's mapping with another, simpler model, possibly a white-box model. However, as the original model and its approximation might differ, it is impossible to explain any input/output behavior of the original model after viewing the simpler one. More importantly, the present desire to utilize complicated models may stem from ineffectiveness or inelegance of simpler black-box models for the same task, so training simpler models would not be ideal either. Even for models that could be viewed as semi-black-box, more investigation of the consequences of this semi-black-boxity is necessary. Statistical versus model-based explanations could bring insight into whether one mode of explanation is favored by the data at hand, and when is it worth the extra cost of interpretation time that the model-based approach entails. Despite being intuitive, some aspects of explanations, e.g., coherence or conciseness have been neglected in the literature, limiting the application scope of the existing knowledge-grounded text generation techniques and in turn the interpretability of. Furthermore, the measured fidelity of a surrogate might differ for different regions of the input space, so there might be no single best explanation. Finally, further study of procedural explanation accountability could have societal implications: more transparent AI models may lead to less distance from human justification, potentially blurring the responsibility line between human developers and AI models.

6.0 CONCLUSION

As due to the rise of the media and the development of Internet technology, a variety of information arises at a blinding speed. In some scenarios, there is a strong requirement for recommending diversified results. A recommendation list that consists of similar items is not the most favorable results for users. For example, if a user is interested in fishing, recommending a top-N list that only includes fishing rods/lines does not meet the requirement, which is annoying for users instead of helpful. The goal of diversification is to mitigate the redundancy in a recommendation list, and provide results with as much novel information as possible (Gong et al., 2019). On the other hand, utilization of black-box models is crucial in the ML areas like NLP or CV in recent years. While these black-box models can provide superior improvement in performance, their lack of transparency also raises potential concerns in many scenarios. Essentially, from a similar goal to limiting the impact of the unfairness in user data, it is desired to construct some post hoc explanations for understanding or auditing the black-box models, or directly dispose of the black-box models and design a more interpretable solution.

For unary queries, self-explanatory interpretable models including decision trees or linear regression can be fitted to mimic the black-box behavior. For multi-way queries, how to enable transparently explainable diverse results is still an open problem. However, exploring such diversification methods is non-trivial and faces many challenges. For neural models, the embeddings of varied long sequences with variable lengths can deliver indistinct representations and be unreliable to find diverse results. The inadequate training from diverse data or diversity-aware advocates can deteriorate the performance or limit the inventiveness of explanations. In some black-box portfolios, the mode of diversity either in feature spaces or in result spaces is difficult or not comparable to explain.

REFERENCES

[1] Baier, L., Jöhren, F., & Seebacher, S. (2019, June). Challenges in the deployment and operation of machine learning in practice. In ECIS (Vol. 1).

[2] Bartz, E., Zaefferer, M., Mersmann, O., & Bartz-Beielstein, T. (2021). Experimental investigation and evaluation of model-based hyperparameter optimization. arXiv preprint arXiv:2107.08761.

[3] Budach, L., Feuerpfeil, M., Ihde, N., Nathansen, A., Noack, N., Patzlaff, H., ... & Harmouch, H. (2022). The effects of data quality on machine learning performance. arXiv preprint arXiv:2207.14529.

[4] Da Silva, B., & Shi, S. S. (2019). Style transfer with time series: Generating synthetic financial data. arXiv preprint arXiv:1906.03232.

[5] Gong, Z., Zhong, P., & Hu, W. (2019). Diversity in machine learning. Ieee Access, 7, 64323-64350.

[6] Oh, S. J., Schiele, B., & Fritz, M. (2019). Towards reverse-engineering black-box neural networks. Explainable AI: interpreting, explaining and visualizing deep learning, 121-144.

[7] Paleyes, A., Urma, R. G., & Lawrence, N. D. (2022). Challenges in deploying machine learning: a survey of case studies. ACM computing surveys, 55(6), 1-29.

[8] Piratla, V. (2023). Robustness, Evaluation and Adaptation of Machine Learning Models in the Wild. arXiv preprint arXiv:2303.02781.

[9] Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2019). A comparative analysis of forecasting financial time series using arima, lstm, and bilstm. arXiv preprint arXiv:1911.09512.

[10] Vennerød, C. B., Kjærran, A., & Bugge, E. S. (2021). Long short-term memory RNN. arXiv preprint arXiv:2105.06756.

[11] Zhao, P., Liu, S., Chen, P. Y., Hoang, N., Xu, K., Kailkhura, B., & Lin, X. (2019). On the design of black-box adversarial examples by leveraging gradient-free optimization and operator splitting method. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 121-130).

Author Profile

Mayada Chibli is an instructor at the Lebanese International University in the Department of Banking and Finance. She holds a BBA and MBA in Finance, both with honors, and she is currently pursuing a PhD at the Bucharest University of Economic Studies, Romania. Her interests include FinTech, financial markets, and portfolio optimization. Her current research focuses on integrating deep learning with modern portfolio theory to enhance investment decision-making.