NLP Project Report: Predicting Stock Trends with News Sentiment

Anonymous Author(s)
Affiliation
Address
`email`

Abstract

This project aims to investigate how combining technical indicators with sentiment analysis can improve the prediction of stock price trends. Focusing on forecasting whether the intraday opening price of the S&P 500 will rise or fall, we propose a hybrid deep learning framework that integrates Transformer models with sentiment signals. Specifically, we use FinBERT to score the sentiment of daily financial news headlines and combine these scores with traditional indicators like RSI and MACD as inputs to our models. We compare two architectures: the Transformer and Bidirectional LSTM (BiLSTM).The experimental results suggest that incorporating sentiment data can enhance predictive performance. However, while the Transformer model excels at capturing long-term dependencies, it struggles without high-quality sentiment inputs. In contrast, the BiLSTM model performs more reliably in small-sample settings, though it sometimes exhibits overfitting. The result highlight the critical role of high-quality financial text data and demonstrate a practical approach for market prediction.

1 Introduction

In financial markets, information asymmetry and investor sentiment play a significant role in driving asset price movements. With the advancement of NLP technologies, an increasing number of studies have begun to extract market sentiment from textual sources, such as financial news, social media posts, and official announcements to support quantitative models in trend prediction.

However, building a predictive system that can effectively integrate textual understanding with time-series modeling remains challenging due to the highly nonlinear and dynamic nature of market behavior. In recent years, deep learning framework such as Transformer models and Bidirectional Long Short-Term Memory (BiLSTM) networks have shown strong potential for financial prediction tasks. At the same time, pretrained language models like FinBERT, which are specifically tuned for financial contexts, offer powerful tools for extracting high-quality sentiment signals.

In this project, we focus on predicting the direction of the next day's opening price of the S&P 500 Index, aiming to assess how incorporating sentiment information alongside technical indicators can enhance trend forecasting. We use daily sentiment scores generated by FinBERT, based on the "Combined News DJIA" dataset from Kaggle, and combine these with traditional technical indicators—RSI, MACD, and SMA—to construct the time-series input. These inputs are then fed into two deep learning models: a Transformer and a BiLSTM, for comparative analysis.

Through a series of experiments using real market data, we demonstrate the feasibility of integrating structured and unstructured data for stock trend forecasting. Our findings not only provide insights into the performance of different modeling approaches but also offer practical guidance for future research in financial sentiment analysis and predictive modeling.

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

2 Related Work and Task Definition

2.1 Related Work

In recent years, the application of sentiment analysis in finance has gained increasing attention. Early studies relied on sentiment dictionaries or lexicons to assign scores to financial texts such as news articles and corporate announcements. For example, the Loughran-McDonald financial dictionary has been widely used in analyzing reports. However, such methods often fail to capture contextual semantics and generally exhibit poor generalization.

To address these limitations, pretrained language models (PLMs) have been introduced into financial text analysis. Among them, FinBERT, a BERT-based model has demonstrated superior performance in extracting positive, neutral, and negative sentiment from financial news. It has outperformed traditional methods in several financial text classification tasks and serves as the sentiment extractor in our project.

On the other hand, traditional financial trend prediction has long relied on technical indicators (e.g., RSI, MACD) and time-series models (e.g., ARIMA, GARCH). BiLSTM networks, which model bidirectional temporal dependencies, have shown strong performance in tasks such as price prediction and trend classification. Meanwhile, Transformer models, known for their parallelism and ability to capture long-range dependencies, are gradually being adopted in financial time series analysis. For example, the StockFormer series encodes multi-asset information as embedded sequences to perform trend classification and asset ranking.

Compared with existing literature, our work focuses on integrating sentiment scores extracted from financial text with structured technical indicators, and comparing the performance of Transformer and BiLSTM architectures in the context of market trend prediction.

2.2 Task Definition

The objective of this project is to predict the direction of the opening price on day $T + 1$, based on technical indicators and financial news sentiment from the previous T days.

We formulate this as a binary classification problem, where the target label is defined as:

$$y_{t+1} = \begin{cases} 1 & \text{if } \text{Open}_{t+1} > \text{Open}_t \\ 0 & \text{otherwise} \end{cases}$$

Each input sample is constructed using a sliding window of length T, containing the features:

- **Technical indicators:** Open, High, Low, Close, RSI, SMA, EMA, MACD
- **Text-based sentiment score:** Daily sentiment scores extracted via FinBERT

The resulting input is a time-series sequence of shape (T, d), where d is the number of features.

We evaluate model performance using two key indicators:

- **Mean Squared Error (MSE):** Measures the accuracy of the regression output on price predictions.
- **Directional Accuracy (DA):** Evaluates the consistency between the predicted and actual movement directions.

The formula for directional accuracy is:

$$DA = \frac{1}{N-1} \sum_{t=1}^{N-1} \mathbb{1}\left[(\hat{y}_{t+1} - \hat{y}_t)(y_{t+1} - y_t) > 0\right]$$

where $\hat{y}_{t+1}$ and $\hat{y}_t$ are the predicted values, and y_{t+1} and y_t are the true values. This indicator reflects whether the model correctly predicts the direction of price change rather than the exact value.

3 Data and Feature

3.1 Data Sources

We use two publicly available Kaggle datasets:

- **Market Data (2008–2016)**: SPY ETF price history representing the S&P 500 index trends.
- **News Data (2008–2016)**: The Combined News DJIA dataset contains 25 top headlines per trading day.

After aligning the dates between the two sources, we obtain 1985 aligned trading-day observations.

3.2 Technical Indicator

In order to give the representation of the price time series, we extract the following technical indicators from each day's data:

- **Relative Strength Index (RSI)**:
$$\mathrm{RSI}_t = 100 - \frac{100}{1 + RS_t}, \quad \text{where } RS_t = \frac{\text{avg gain}}{\text{avg loss}}$$
- **Simple Moving Average (SMA)**:
$$\mathrm{SMA}_t = \frac{1}{n} \sum_{i=t-n+1}^{t} P_i$$
where $n = 14$ and P_i is the closing price.
- **Exponential Moving Average (EMA)**:
$$\mathrm{EMA}_t = \alpha \cdot P_t + (1 - \alpha) \cdot \mathrm{EMA}_{t-1}, \quad \alpha = \frac{2}{n+1}$$
This recursive formula emphasizes recent prices more heavily.
- **MACD (Moving Average Convergence Divergence)**:
$$\mathrm{MACD}_t = \mathrm{EMA}_{12,t} - \mathrm{EMA}_{26,t}$$

3.3 Daily Sentiment Score with FinBERT

We use the FinBERT model to compute daily sentiment scores:

1. Merge all 25 headlines from the same day into a single input string.
2. Use the FinBERT classifier to obtain probabilities for [Positive, Neutral, Negative] classes.
3. Define the sentiment score as:
$$\mathrm{Sentiment}_t = P(\mathrm{Positive}_t) - P(\mathrm{Negative}_t)$$

The resulting score ranges in $[-1, 1]$, representing the direction and intensity of sentiment.

3.4 Sliding Window Construction

We construct input samples using a sliding window approach:

- Set the window length $L = 10$;
- Each input sample contains the previous 10-day feature sequence;
- The prediction target is the next day's open price.

Formally, each sample is defined as:

$$\mathbf{x}_t = [\mathbf{z}_{t-L+1}, \ldots, \mathbf{z}_t], \quad y_t = \mathrm{Open}_{t+1}$$

This preprocessing yields approximately 8000 time-aligned samples for model training and evaluation.

4 Model Architectures

4.1 Transformer-Based Regressor: StockFormer

Inspired by paper Support for Stock Trend Prediction Using Transformers and Sentiment Analysis, we design a lightweight sequential regression model named **StockFormer**. The model includes:

- A linear projection layer to embed input features into a d-dimensional space:

$$\mathbf{h}_t = \mathbf{W}_{\text{in}}\mathbf{z}_t + \mathbf{b}_{\text{in}}$$

- Multi-head self-attention and positional encoding are applied to capture cross-day dependencies.
- The last token's representation is passed to a final feed-forward layer:

$$\hat{y}_t = \mathbf{w}_{\text{out}}^\top \mathbf{h}_t + b_{\text{out}}$$

4.2 BiLSTM Baseline

We use a BiLSTM-based as a benchmark model:

- A bidirectional LSTM with two layers and hidden size 64.
- The hidden states from both directions at the last time step are concatenated.
- A linear layer outputs the predicted next-day open price:

Motivation: BiLSTM is capable of capturing short-term memory patterns in a time series and is widely used as a benchmark in financial forecasting tasks.

5 Empirical Results

5.1 Performance Comparison

We compare the performance of the Transformer-based model (StockFormer) and the BiLSTM baseline using two evaluation metrics:

- **Mean Squared Error (MSE)**: Measures the average squared difference between predicted and actual open prices.
- **Directional Accuracy (DA)**: Measures the percentage of times the predicted direction of price change matches the actual direction.

Model	MSE	Directional Accuracy
StockFormer (Transformer)	2944.57	0.5071
BiLSTM	**1031.02**	**0.9584**

Table 1: Performance comparison of StockFormer vs. BiLSTM on validation set

5.2 Dynamic Comparision

Figure 1 shows the training evolution of both models. While StockFormer quickly converges to a plateau in directional accuracy (50%), BiLSTM continues to improve over epochs, achieving over 95% accuracy. This highlights the latter's ability to exploit short-term temporal dependencies more effectively in this setting.

The Transformer model exhibits a marked decline in predictive accuracy around periods of extreme market movement, such as the March 2020 COVID-19 crash. This shows its limited capacity to model abrupt volatility shifts.

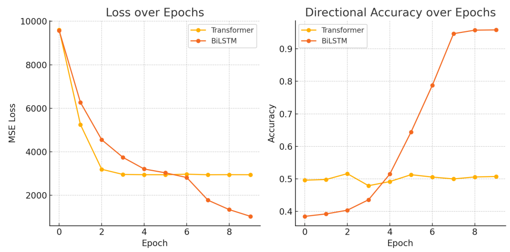

Figure 1: Training loss and directional accuracy over epochs for both models

5.3 Critical Analysis

Despite the theoretical strengths of Transformers in capturing long-range dependencies, StockFormer underperforms in this case due to several factors:

- **Small dataset size**: Transformers are data-hungry; the limited number of training samples (8000) hinders their ability to generalize.
- **Lack of pretraining**: Unlike NLP applications where Transformers benefit from large-scale pretrained models, our model is trained from scratch.
- **Overfitting risk**: The higher model capacity of Transformers leads to quicker overfitting on small datasets.

In contrast, BiLSTM, with a smaller parameter count and strong inductive bias toward sequential patterns, learns meaningful representations even with limited data. This aligns with findings in [1], which emphasize the importance of inductive bias in time-series financial forecasting.

6 Conclusion

By adapting and extending the work of Garcia et al. (2023), we built a full modeling pipeline that merges structured market data with unstructured text, and provided a practical comparison between two deep learning approaches.Our experiments demonstrate that while Transformers offer high modeling capacity, they underperform in low-data regimes, where BiLSTM models with stronger sequential inductive bias show superior generalization. The findings highlight the value of incorporating sentiment into financial forecasting and point to future directions such as using alternative sentiment sources, applying causal or reinforcement learning methods, or exploring time resolutions like hourly predictions.

References

[1] Garcia, A., Zhang, Y., Kumar, V. (2023). Support for Stock Trend Prediction Using Transformers and Sentiment Analysis. *arXiv preprint arXiv:2305.14368*.

[2] Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL-HLT*.

[3] Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., Le, Q. V. (2019). XLNet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems*, 32.

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes] , [No] , or [NA] .
- [NA] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes] " is generally preferable to "[No] ", it is perfectly acceptable to answer "[No] " provided a proper justification is given (e.g., "error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering "[No] " or "[NA] " is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "NeurIPS paper checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly state that the goal is to compare Transformer and BiLSTM models for financial forecasting using structured and unstructured data. These claims are supported by the methodology and results discussed in Sections 3, 4, and 5.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]

Justification: Conclusion includes a discussion of limitations, such as the data scarcity problem affecting Transformer performance and potential improvements using other sources of sentiment, finer time resolutions, or reinforcement learning.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: This project is empirical and does not involve any theoretical results or formal proofs. The paper focuses on experimental comparison of models rather than deriving new theoretical guarantees.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: Section 3 and Section 4 clearly explain the data sources, feature engineering steps, model architectures (Transformer and BiLSTM) and evaluation metrics (MSE and directional accuracy). The code is mentioned as available via GitHub.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: Both datasets used (SPY and DJIA News) are publicly available on Kaggle, and the code is planned to be hosted on GitHub. The preprocessing steps and modeling pipeline are clearly written in the report.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.

- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [NA]

   Justification: The report describes model hyperparameters, loss function (MSE), training epochs, and batch size. It also details the train-validation split (80/20), window size for time series modeling, and normalization procedures.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification:

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [No]

Justification:

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: The research uses publicly available datasets from Kaggle and open-source models FinBERT under permitted licenses.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [No]

Justification:

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [No]

Justification:

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification:

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: This project does not introduce any new datasets or pretrained models, nor does it develop any new open-source tools or frameworks. Therefore, there are no new assets or documentation that require separate publication.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.

- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification:

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification:

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.