Fake News Detection

Multimodal NLP Model for Text and Image Classification

1. Team

- Surya Putrevu - sputre2@pdx.edu
- Kajal Patil - kajal@pdx.edu
- Rohan Singh - rohsingh@pdx.edu
- Dan Shitkar - dshitkar@pdx.edu

2. Motivation

In today's digital world, the spread of fake news has become a concerning issue that undermines informed decision-making and societal stability. This report explores the creation and assessment of a multimodal NLP model that utilizes Bert/ GPT /ERNIE models to detect fake news effectively. By leveraging the capabilities of a robust natural language processing framework, the proposed model is capable of extracting meaningful information from both textual and visual components, thereby improving the accuracy and comprehensiveness of fake news identification. Also towards the end our aim is also to compare the results of all 3 models and find out which one suits the most as per our requirement.

3. Introduction

The prevalence of fake news, which involves intentionally fabricated and spread misinformation, poses a significant challenge in today's digital world. Its harmful impact lies in its ability to sway public opinion, undermine trust in institutions, and even incite violence. As a result, countering fake news has become a crucial task in protecting the accuracy of information and promoting societal well-being. Traditional methods for detecting fake news often rely on analyzing only text or images separately. However, such approaches overlook the crucial interdependencies between visual and textual cues, which can be vital in identifying fake news. In contrast, multimodal NLP models address this limitation by integrating information from various modalities such as text, images, and audio. This comprehensive approach allows for a deeper understanding of the content, resulting in improved detection of fake news.

4. Methodology

4.1. NLP Task Addressed

- Binary classification of a News based on Title and images associated with it.

4.2. Dataset

- We utilized the ['multimodalfake']dataset by Hariharan RL, consisting of textual features (author, clean_title, domain) and image URLs and other fields, which we sourced from Kaggle.
- Dataset size: approximately 59000 samples.
- Dataset includes 3 labels: 2_way_label (which identifies news as fake or true), 3_way_label (which identifies news as fake, neutral, true) and 6_way_label (which identifies news as fake, may be fake, neutral, may be neutral, true, maybe true)

4.3. Preprocessing

- Limited dataset size to 10000 rows for faster experimentation.
- Selected relevant columns for analysis.
- Dataset was split into 80:20 ratio to form train data and test data respectively.

4.4. Models and Techniques

- Implemented ERNIE-based text tokenization and vectorization functions.
- Implemented GPT-based text tokenization and vectorization functions.
- Implemented BERT-based text tokenization and vectorization functions.
- Implemented a common image vectorization function using OPENCV library.All images were resized to 224*224 pixels. Then transformed that image into tensor form , which is in the NumPy array format.
- Developed a neural network architecture which can consume 2 input layers as text and image input. Used Relu as an activation function to enhance the learning process.
- Used TensorFlow and Hugging Face Transformers libraries to train and evaluate the model.

5. Experiments and Results Evaluation

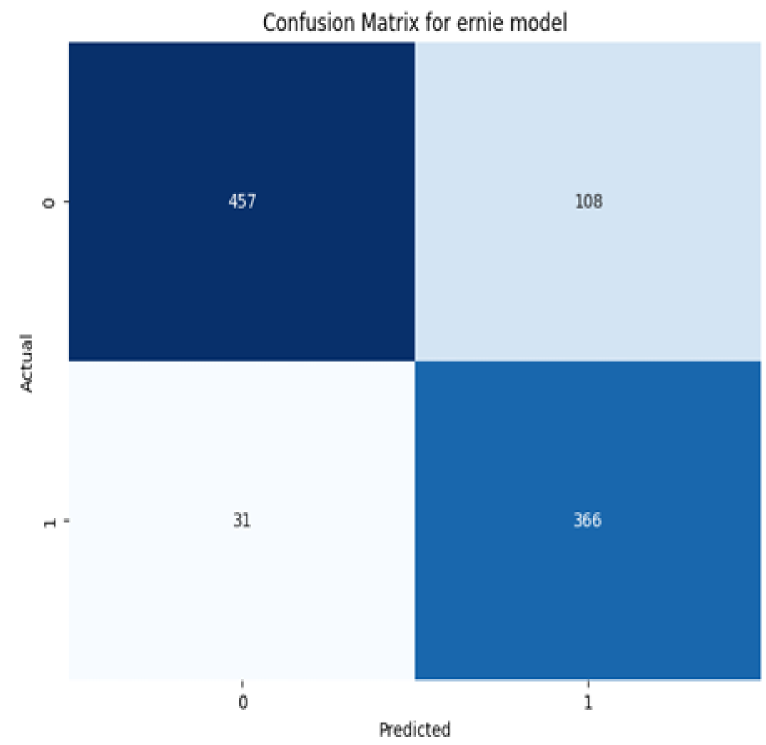
5.1. ERNIE model - Surya Putrevu

As we can see in the graph training loss is dropped as we proceed through epochs and training accuracy is improved along the side. On the contrary, validation loss has improved as we proceed and accuracy has dropped . We can see alternative high and lows in validation accuracy.

Despite this, the Ernie Model has given maximum accuracy among all 3 models, which was very surprising. I was able to achieve 86% accuracy. But there is a significant difference between precision for class 0 and 1 which is fake news and real news.

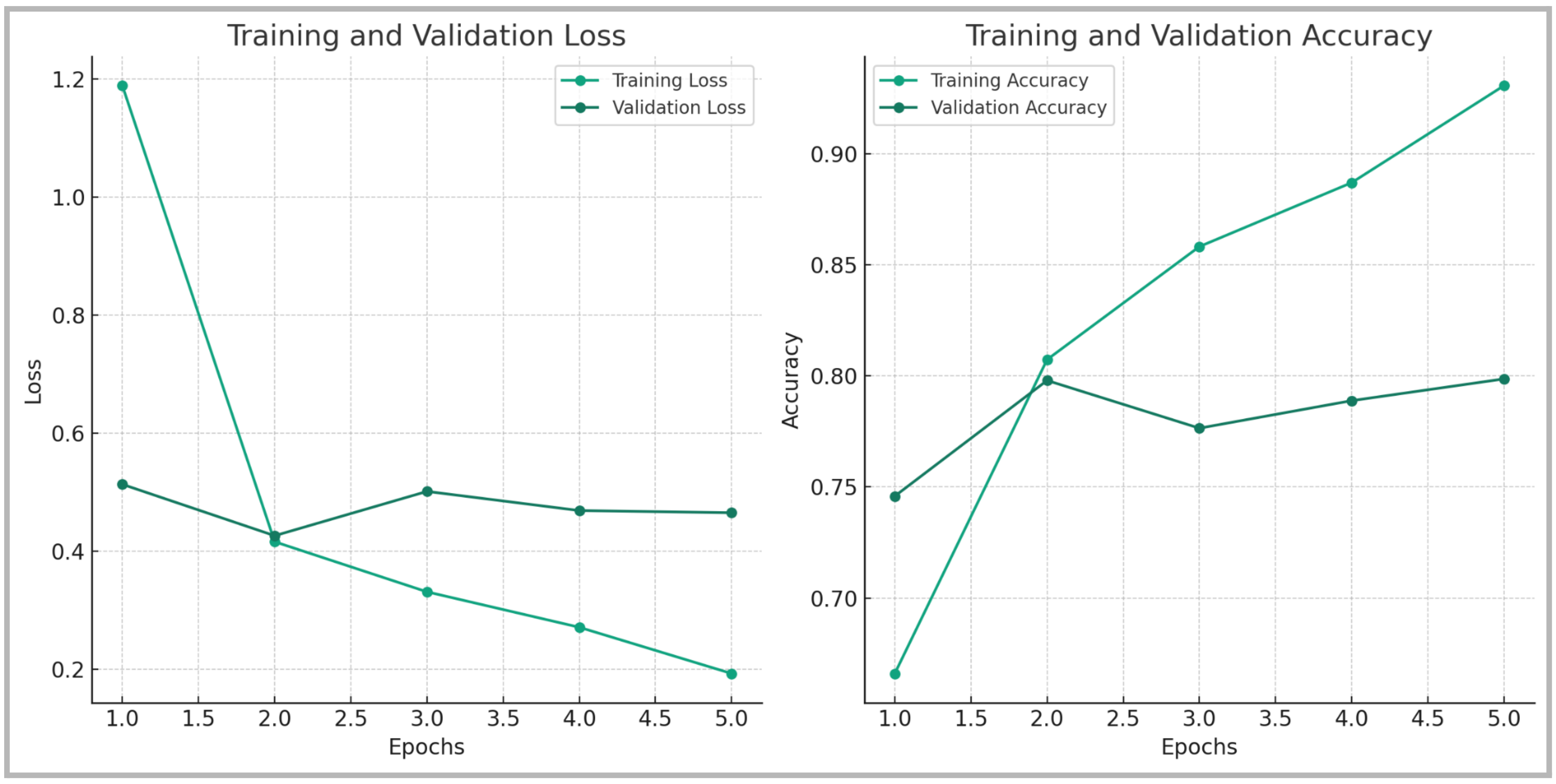
Classification Report

	Precision	Recall	F1 Score	Support
0	94%	81%	87%	565
1	77%	92%	84%	397
Accuracy			86%	962
Macro Avg	85%	87%	85%	962
Weighted Avg	87%	86%	86%	962



5.2. GPT Model - Rohan Singh

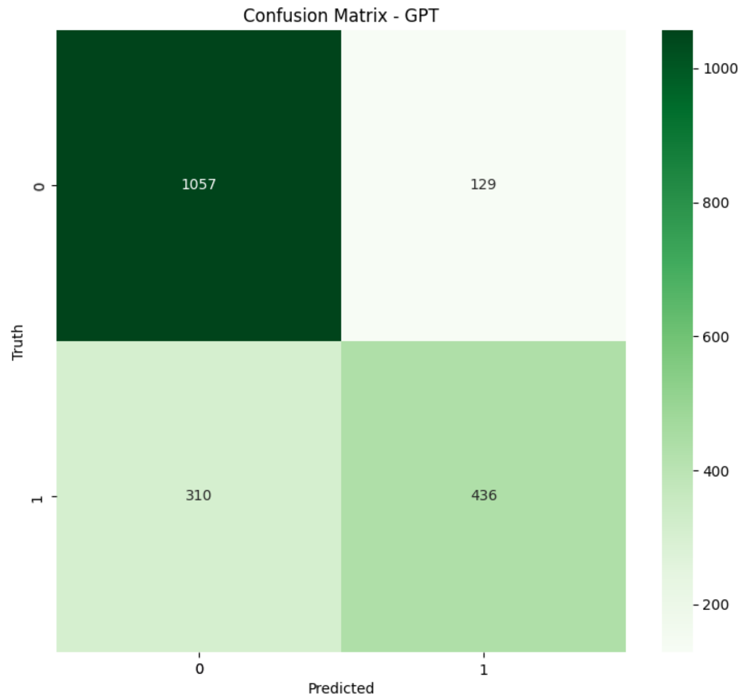
This model only considers the actual line of text that features the news and ignores the author and domain to highlight any measurable changes in the quality of prediction in contrast to the other models which also consider the author and domain.. I have presented my observations here:



As shown in the above figures, the validation loss, although starting out lower than training loss, ended up higher than training loss after epoch two. Also, the validation accuracy dropped after epoch 2 before improving eventually.

Classification Report

	Precision	Recall	F1 Score	Support
0	0.77	0.89	0.83	1186
1	0.77	0.58	0.67	746
Accuracy			0.77	1932
Macro Average	0.77	0.74	0.75	1932
Weighted Average	0.77	0.77	0.77	1932



I got an accuracy of 77%, and as depicted in the confusion matrix above, in terms of wrong predictions it is evident that the occurrences of false negatives are somewhat higher than those of false positives. As I had hamstrung my model on purpose by discarding the author and domain fields it can be inferred from observing the exhibited data and the results of other models that it did negatively affect the accuracy of predictions, although not by a very significant margin.

5.3. Bert Model - Kajal Patil

Our dataset has 3 label columns which are 2_way_label, 3_way_label and 6_way_label. After comparing the results we went ahead with 2_way_label as it gave us the most accurate results among the three. Also I experimented with data size of 100, 1000 and 10000 records. Experiments showed a significant change from 100 to 1000, where accuracy went from 73% to 81%. Whereas with 10000 dataset I got an accuracy of 82.56 %.

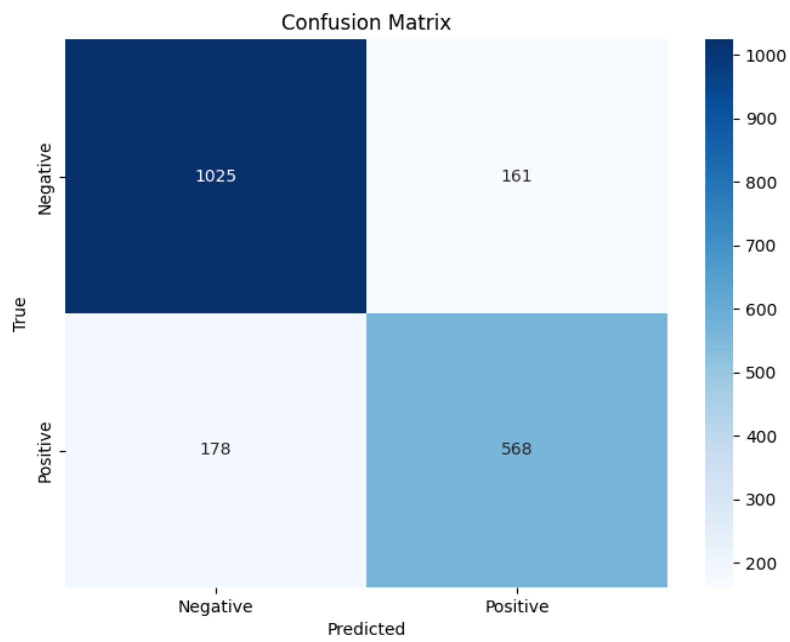
Outcome for 1000 dataset:

```
Accuracy: 0.81
F1 Score: 0.81
```

Outcome for 10000 dataset:

Classification Report

	Precision	Recall	F1 Score	Support
0	0.85	0.86	0.86	1186
1	0.78	0.76	0.77	746
Accuracy			0.82	1932
Macro Average	0.82	0.81	0.81	1932
Weighted Average	0.82	0.82	0.82	1932

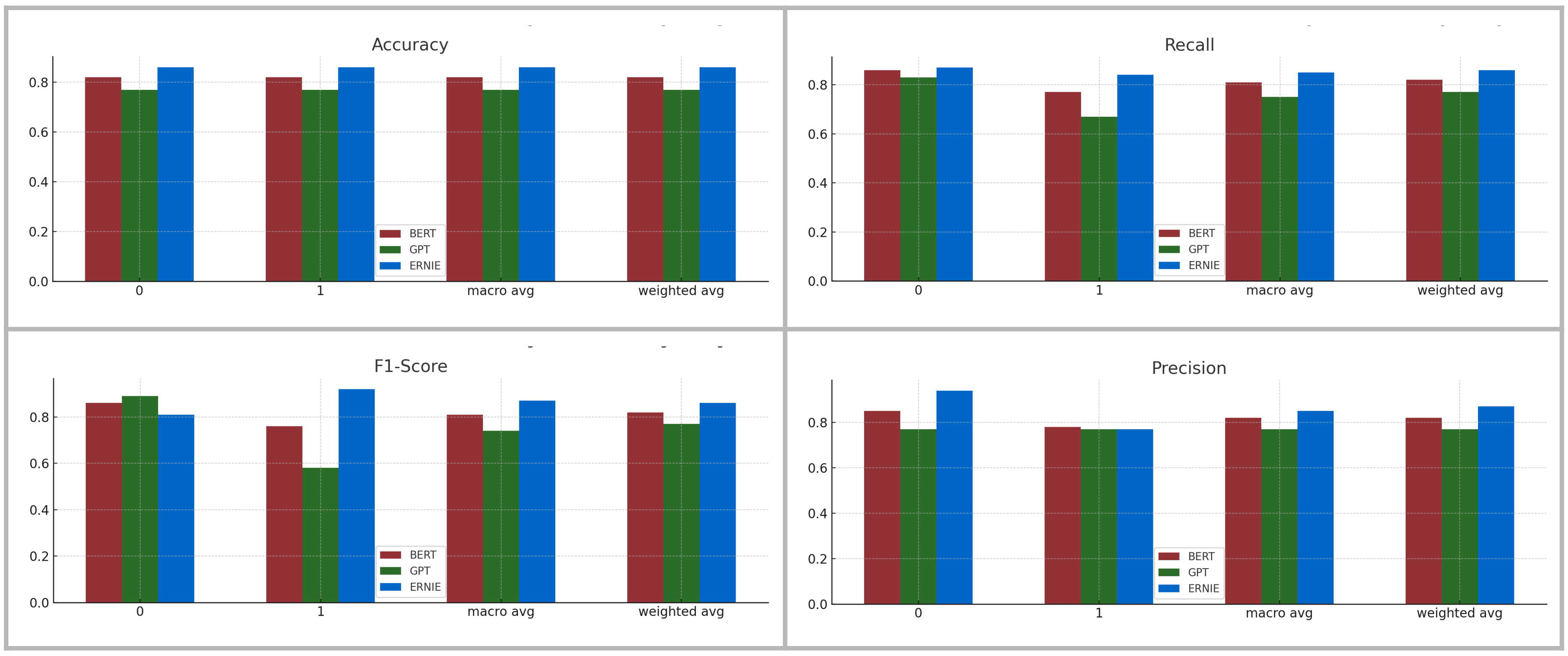


6. Related Work

In the past, researchers have delved into the concept of multimodal models with the aim of enhancing the accuracy of classification tasks that involve both text and images. Our method takes inspiration from this approach but takes it a step further by combining the advanced capabilities of Transformer models such as Bert /GPT /ERNIE with established image processing techniques. This powerful fusion creates a robust hybrid approach that can identify even the most nuanced connections between text and images. With this approach, we strive to achieve more refined and precise classifications, which can have a significant impact in crucial fields like natural language processing and computer vision. Our method offers the potential to bring about breakthroughs in these fields, and we are optimistic about the possibilities that it can offer.

7. Interesting Insights

- Successful integration of text and image data.
- Challenges in handling diverse image sources and quality.
- Experimentation revealed the need for further data augmentation.
- By contrasting the results of BERT and GPT models, it was observed that considering the author and domain for a particular line of text increases the overall accuracy of the model and a possible explanation for that is that the model learns to recognize repeat offenders posting fake news and the domain they operate from, so it gets more efficient in flagging them.
- Comparison between Bert, GPT and ERNIE models:



8. Future Work

As of now each time we execute the code we end up downloading the image, which is a very memory heavy operation. Going forward we want to save these images in a folder and then consume it from the folder for further processing. This can reduce the work for the second time onwards. From what we observed, our project is biased towards predicting real data. According to our research about this topic, we can implement the CCD(Casual intervention and Counterfactual reasoning based Debiasing) framework to remove this bias and make our prediction more accurate. We believe that CCD benefits from the removal of psycholinguistic bias with causal intervention as well as the mitigation of the image-only bias via counterfactual reasoning.

9. Group Contribution

Kajal Patil : I worked on setting up the initial flow for the project. Did a research to find out what all models can be used for our project as it involves images as well. Also looked into multiple datasets to see which dataset can be useful for us . Found out a way to vectorize images and feed it to the neural network model. I did set the outline of a code so that we can compare model performance at the end.
While doing so I faced issues while vectorizing an image and managing its size. As we need to pass text and image to the same neural network, managing its size was a major challenge. Since text input is 1D vector I had to flatten out the image vector to 1D.

Rohan Singh : I worked on the GPT aspect of our project, examining how different data selections impact predictions and by extension, the accuracy of the model I worked on. I also consolidated the results so that they could be compared to arrive at a definitive conclusion regarding the efficacy of the different models, which is essentially the primary objective of this project. I also handled the styling and formatting of the presentation, by removing bulky text and replacing it with succinct bullet points for better clarity. Apart from that, I presented my section of the project and I also concluded our project presentation.

Surya Putrevu : Throughout the project, I played a pivotal role in achieving our desired outcomes. Specifically, my expertise and attention to detail were crucial in carefully selecting and working with the Ernie model. To ensure project success, I maintained regular communication with the team, scheduling meetings and reviewing progress. Proactively, I took the initiative to create project slides for our presentation and final report, ensuring the project's timely completion at a high standard. While working with the Ernie model, I encountered computational and session timeout challenges, which required quick thinking and problem-solving skills. Skilfully, I utilized the dataset controller to manually manage the dataset count, optimizing the model's performance for timely and efficient results. Overall, my contribution was critical to the project's success, and I am proud of our team's accomplishments.

Dan Shitkar

10. Code Repository

Code repository: https://github.com/kajal8879/MultiModal-Fake-News-Detection

11. Conclusions

Throughout our project, we conducted a thorough investigation into the effectiveness of a multimodal natural language processing (NLP) model for text and image classification. Our findings were largely positive, as we were able to achieve promising results in terms of accurately classifying the data we were working with. However, we also encountered several challenges and identified certain areas for improvement.

One of the key issues we identified was the need for responsible data collection practices. Specifically, we recognized that the data we were using could potentially be biased in certain ways, and that this could have a negative impact on the accuracy of our model. To mitigate this risk, we took great care to ensure that our data was collected in a way that was as representative and unbiased as possible.

Another important ethical consideration that we discovered was the need for energy-efficient computing practices. Given the large amounts of data that our model was processing, we recognized that there was a significant environmental impact associated with our work. To address this issue, we explored various strategies for reducing our energy consumption, such as optimizing our code and using more energy-efficient hardware.

Overall, our project has highlighted the complex ethical considerations that must be taken into account when working with advanced NLP models. By carefully examining these issues and working to address them in a responsible and sustainable manner, we believe that we can help to ensure that these technologies have a positive impact on society and are used in a way that is both ethical and responsible.

12. Future Directions

Our ongoing efforts to improve the model's performance will involve a range of activities that aim to expand its capabilities and accuracy. Specifically, we will focus on expanding the dataset to include more diverse and representative samples. This will allow the model to better generalize to new and unseen data, and improve its overall performance.

In addition, we will explore the integration of more advanced image processing techniques, such as feature extraction and image segmentation, to obtain more detailed information from the images and improve the model's ability to detect patterns and features.

Moreover, we plan to investigate the use of transfer learning, which involves leveraging pre-trained models to extract useful features and weights that can be used to improve the performance of our model. By doing so, we can reduce the amount of training data required and improve the model's ability to learn from smaller datasets.

Overall, these efforts will help us to develop a more robust and reliable model that can be applied to a wide range of image recognition tasks, and ultimately improve our ability to extract meaningful insights and knowledge from visual data.

13. References

- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.
- Brownlee, J. (2021). How to Develop a Multimodal Deep Learning Model for Time Series Forecasting. Machine Learning Mastery. [Online]. Available:

https://machinelearningmastery.com/how-to-develop-a-multimodal-deep-learning-model-for-time-series-forecasting/.

- Hugging Face Transformers. (n.d.). Transformers Library. [Online]. Available: https://huggingface.co/transformers/.
- TensorFlow. (n.d.). TensorFlow Documentation. [Online]. Available: https://www.tensorflow.org/guide.
- Impact Statements Guide. (2022). Writing Impact Statements. [Online]. Available: https://www.nsf.gov/pubs/2019/nsf19084/nsf19084.jsp.
- Experiment Impact Tracker. (2023). Tracking the Impact of AI Experiments. [Online]. Available: https://github.com/AI-RRI-Toolkit/experiment-impact-tracker.
- multimodalfake. Available: https://www.kaggle.com/datasets/hariharanrl/multimodalfake/