Wildlife: Exploring Wildlife Trafficking through Animal-Related Ads

Jin Zhou
New York University
New York, USA
jz3928@nyu.edu

Junzhe Zhou
New York University
New York, USA
jz3709@nyu.edu

Chhatrapathi Sivaji Lakkimsetty
New York University
New York, USA
cl7203@nyu.edu

I. Introduction

The illegal wildlife trade (IWT) poses significant threats to biodiversity, ecosystems, and human health. To combat this illegal activity, the recent researches propose diverse set of methods ranging from X-ray imaging to genetic analysis.

In ”Detecting illegal wildlife trafficking via real time tomography 3D X-ray imaging and automated algorithms,” 3D X-ray CT technology is used together with machine learning to develop an algorithm that is capable of identifying concealed wildlife within luggage and cargo with high accuracy. The researchers utilized 294 scans from 13 rare species to train the model, and they are able to achieve a detection rate of 82% with a 1.6%false positive rate. [1], [2]

The focus shifts to the impact of IWT on spreading infectious diseases in both nature and human habitats in research paper ”Illegal Wildlife Trade and Emerging Infectious Diseases:Pervasive Impacts to Species, Ecosystems and Human Health”. The authors conduct a comprehensive literature review on 82 papers from 1990 to 2020 in order to fully understand the link between IWT and certain types of pathogens. [3]

The study ”A Survey on Identification of Illegal Wildlife Trade” highlights the unawareness of the general public about illegal animal trafficking by explaining several ways of trade that is happening on the web. The author Nalluri first explains how trades are conducted on the ”clear web”. On legitimate platforms like Ebay, advertisements regarding animal trade are being sent to potential interested buyers via recommendation systems on a weekly basis. Since these ads exists for only a short period of time, it is difficult to track them back to their source. [4] The other lesser known space is dark web. A P2P network and a Tor browser are tools for cybercriminals to mask their identities and conduct illegal businesses. This is why investigating crime on the dark web is even more difficult. Later in the paper, author recommends the use of deep learning to identify both sellers and buyers. More specifically, researchers can use APIs to collect large dataset from social media platform like twitter. Then, a deep learning model can help identify visual, verbal, and audiovisual content regarding IWT. [4], [5]

II. Problem Formulation

The main issue we are addressing in this project is identifying online advertisements for wild animal trading to potentially reduce the problems mentioned above. To achieve this, Machine Learning will be utilized to classify information from images and texts. This problem can be divided into three smaller tasks: data gathering, image analysis, and text analysis.

- For data gathering, advertisements need to be collected from multiple trading sites such as eBay. Then, the collected data will be cleaned with weak supervision, for example, remove the ads that did not provide an image or text description, to avoid bad data and reduce the accuracy of the model.
- Then, the images provided need to be processed into relative text or numerical information that can be analyzed. This step can be completed either separately to provide more inputs for a model with a simpler structure, or passed in along with other data but have to be evaluated differently.
- Finally, the text data such as description and location needs to be analyzed alongside either the modified image data or the original image to form a final classification of the advertisement, with the final output being either whether it is wild animal trading or the type of product the advertisement is trying to promote.

Using the final result from this three-step processing, we will be able to distinguish whether an advertisement is attempting to trade wild life.

III. Related Work

There have been multiple efforts in image-to-text classification using machine learning. The paper ”Deep Learning for Image-to-Text Generation: A Technical Overview” by Xiaodong He highlights the difficulty of detecting and understanding the relationships between various elements within an image. He et al. proposes two solutions that enhance video captioning. [6] First one is an end-to-end framework, which is also called vector sequence learning. Basically, the encoder processes the image to generate a global visual feature vector using Convolutional Neural Networks(CNNs) and the decoder will use that output to generate a caption using Recurrent Neural Networks(RNNs). The second method is the attention mechanism. This allows the model to focus on specific parts of an image rather than the entire image. By dynamically focusing on different subregions of an image, the caption generated will evolve over time. To evaluate their results, they used both automatic metrics and human studies. A quantifiable metric would be the fraction of n-grams between the hypothesis and the inference, which we can also consider using in our research. [6]

Another novel approach uses Maximum Mean Discrepancy (MMD) to help in the image-to-text synthesis process, rather than relying solely on the traditional Generative Adversarial Networks (GANs). Das et al. created two separate autoencoders, where one is for texts and the other is for images. [7]For the image autoencoder, ResNet50 is used to first transform images into high-dimentional vectors. For the text autoencoder, the authors employ a pre-trained LSTM on the One Billion Word Benchmark dataset. For the translation between embedding spaces of the two modalities, the authors compare the performance of a MMD-based mapping network and a GAN-base mapping network. This study offers an advanced multi-modal learning technique that outperforms the traditional image-to-text frameworks. [7]

More recently, GPT-4Vision has been known to benchmarking image-to-text taks using Large Language Models(LLMs). It is capable of accurately extracting both text and visual features from user input images. In further details, similar with aforementioned work, the features are extracted and mapped to a semantic space. [8] Then, the data is fed into an LLM like GPT so that text generation can begin. It is potentially useful to include the use of LLM-based models to compare the performance of different classifiers that we build.

In addition to the above methods, the transformer architecture is also powerful in understanding human language. Unlike CNNs or RNNs that rely heavily on self-attention mechanism, transformers use a multi-head attention framework. This not only makes dynamic weighting of the input sequence possible, it also allows multiple components in the training process to be processed in parallel [9]. Specific to our project, Bidirectional Encoder Representation from Transformers(BERT) is a branch of transformers that specializes in context understanding [10]. As we aim to extract useful information from wild animal product advertisements, this pre-trained model can significantly increase the efficiency of our research.

IV. Data Insights

Numerous sources offer advertisements related to wildlife, spanning a broad spectrum of wildlife products, both directly and indirectly connected. However, sifting through the vast internet landscape to gather this information is not computationally efficient for this project. Therefore, our research narrows its focus to a single prominent online platform, eBay.com, as the primary source for data collection. The ads found on eBay.com provide a rich dataset, notably abundant in listings directly linked to animal products. Our dataset comprises various elements such as image URLs, descriptions, titles, prices, domains, locations, and more from these advertisements. To effectively comprehend and utilize this data, it's essential to analyze and process both the image and textual information. This approach transforms raw data into structured, meaningful insights that accurately reflect the characteristics of the advertisements.

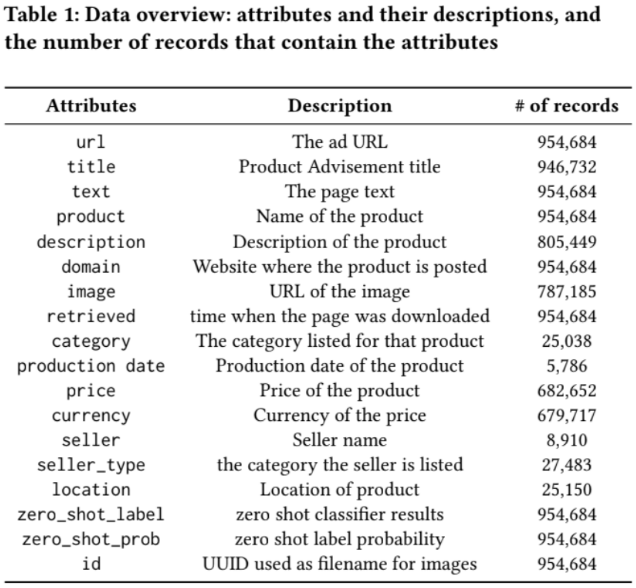

Table 1: Data overview: attributes and their descriptions, and the number of records that contain the attributes

Attributes	Description	# of records
`url`	The ad URL	954,684
`title`	Product Advisement title	946,732
`text`	The page text	954,684
`product`	Name of the product	954,684
`description`	Description of the product	805,449
`domain`	Website where the product is posted	954,684
`image`	URL of the image	787,185
`retrieved`	time when the page was downloaded	954,684
`category`	The category listed for that product	25,038
`production date`	Production date of the product	5,786
`price`	Price of the product	682,652
`currency`	Currency of the price	679,717
`seller`	Seller name	8,910
`seller_type`	the category the seller is listed	27,483
`location`	Location of product	25,150
`zero_shot_label`	zero shot classifier results	954,684
`zero_shot_prob`	zero shot label probability	954,684
`id`	UUID used as filename for images	954,684

Fig. 1. Data Composition

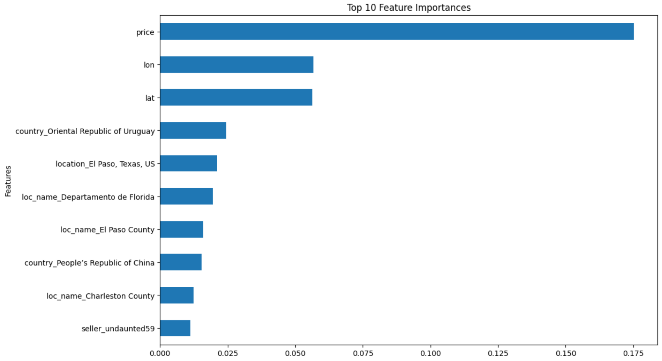

Fig. 2. Feature Importance

We also performed a feature importance analysis, which identifies features that significantly influence the classification result of the advertisements. We discovered that price is the most impactful feature of all features. Following price, we have longitude and latitude, which help us infer the location of the transaction. As we see in Figure 2, the rest important features are specific values of a feature. For example, Florida, the United States is shown to be highly influential. This could mean that the state of Florida is an even better indicator of illegal wild animal product than the location feature in general.

V. Methods, Architecture, and Design

A. Workflow

Each pipeline is intricately constructed to maximize the utility of the collected data, ensuring its precise processing and analysis to fulfill the project's goals.

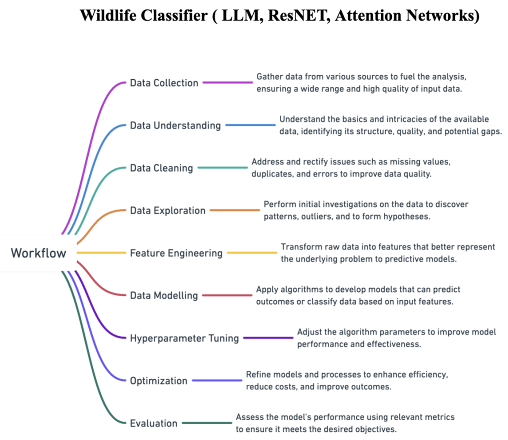
Wildlife Classifier (LLM, ResNET, Attention Networks)
Workflow
Data Collection
Gather data from various sources to fuel the analysis, ensuring a wide range and high quality of input data.
Data Understanding
Understand the basics and intricacies of the available data, identifying its structure, quality, and potential gaps.
Data Cleaning
Address and rectify issues such as missing values, duplicates, and errors to improve data quality.
Data Exploration
Perform initial investigations on the data to discover patterns, outliers, and to form hypotheses.
Feature Engineering
Transform raw data into features that better represent the underlying problem to predictive models.
Data Modelling
Apply algorithms to develop models that can predict outcomes or classify data based on input features.
Hyperparameter Tuning
Adjust the algorithm parameters to improve model performance and effectiveness.
Optimization
Refine models and processes to enhance efficiency, reduce costs, and improve outcomes.
Evaluation
Assess the model's performance using relevant metrics to ensure it meets the desired objectives.

Fig. 3. Workflow

This project is architected around two critical pipelines, pivotal to its comprehensive development and eventual success. Below we elucidate the organization and operational specifics of these pipelines.

B. Data Collection Pipeline

In the extensive digital ecosystem, advertisements for wildlife products are prevalent, with both direct and indirect relevance. However, the sheer scale of online data poses computational challenges for our project. Consequently, we concentrate our data gathering efforts on eBay.com, a principal source. This pipeline aims to methodically harvest product listings from eBay and analogous platforms, thereby generating an ‘animal_products‘ dataset. This dataset includes essential attributes such as Image URL, Product Sales, Seller Information, Location, Text Descriptions, and Titles, all of which are instrumental in furnishing deep insights. Utilizing a Jupyter notebook service, the ‘Minio_data‘ pipeline provides a shared data storage solution, facilitating the dissemination of sizable datasets to users as a read-only resource within their notebook environments.

C. Data Inference Pipeline

Crucial to the project, this pipeline focuses on the processing, cleansing, transforming, and analytical examination of the raw data to extract valuable insights. It commences with the purification of raw data, a prerequisite for detailed analysis. The Image URL field, rich in visual data, necessitates transformation to unlock its information potential. This pipeline incorporates several key procedures, like data wrangling, identification of outliers, Exploratory Data Analysis along with feature engineering, in order to make a better classifier.

D. Data Cleaning

The raw dataset is not fit for direct use as it has many flaws that can negatively affect the performance of the classifier. Firstly, there are duplicate or similar features. For example, feature loc_name, $location$, and $country$ all contain information about the location of the transaction. Some are more specific, having zip code and city or county names, while some only have country name.

Secondly, same values in the same column can have very different format. For instance, the country column has many formatting issues. Some values are abbreviations like ’US’; some are in Russian or Mandarin; the rest are full names of the countries. To have unified values in the same column, we created a dictionary that maps unwanted format of specific values to the desired values.

Furthermore, we simplified the data by using only the country column as the location indicator. During the process of doing so, we discovered that almost 60% of the rows of country has null values. We then extracted country information from $location$ column and loc_name column and fill the country column with that extracted information.

After performing the above data cleaning, we noticed that there are still plenty of rows of null values. We further minimized null values by utilizing the longitude and latitude columns. With Python’s geopy library, we implemented a function that is able to infer country from the geographical coordinate, and then fill the country column.

E. Outlier Identification

Price feature in the dataset has too many outliers as the price distributions seems to be too randomly distributed. It leads to bias in the model predictions and outlier resolution is truly essential. Outliers in the price are due to the different currency types, which make the price ranges too high. To normalize the prices, currency converters are used, where API’s convert all the prices to USD and it maintains the uniformity in the price.

- Skewness and Range: Both distributions appear to be highly right-skewed, indicating that the majority of prices are low, with a few ads having very high prices. This is typical for product datasets where a large number of inexpensive items and a few premium or high-priced items exist.
- Similar Shape: The shapes of the distributions in both plots are similar, which suggests that the conversion from the original currency to USD did not alter the relative distribution of prices. This is expected as a currency conversion should uniformly scale the prices.
- Outliers: The presence of extreme values or outliers is noticeable in both plots. These outliers could significantly affect the mean price, making the median a more reliable measure of central tendency for this data.

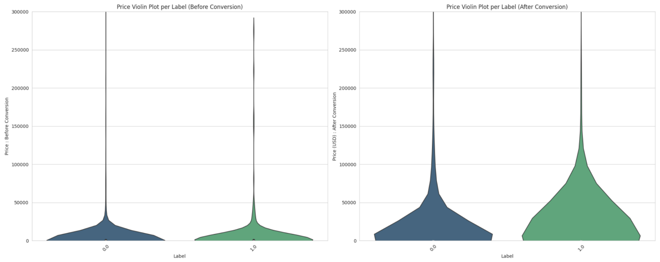

Fig. 4. Price Normalization

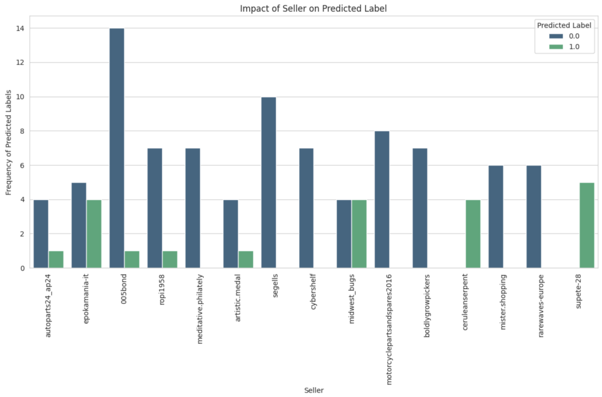

Fig. 5. Top 15 Sellers Distribution

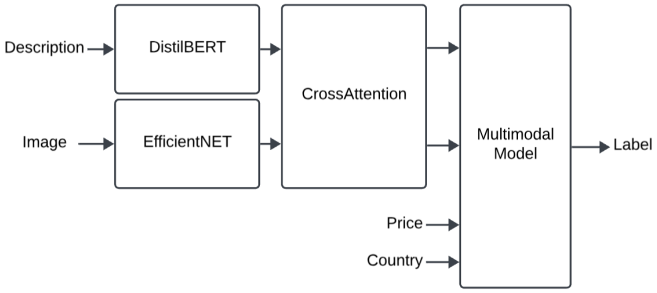

Fig. 6. Multimodal Model

F. Advanced Feature Transformations

In this project, we employed SparkMLib to enhance feature engineering processes by deriving new attributes from existing data points such as the product's location, price, category, and seller details, aiming to boost the accuracy of predictions for specific variables. Utilizing techniques such as Frequent Pattern Matching, we generate itemsets to identify potential features including Price, Country, Seller, and Product Category. These itemsets facilitate the determination of confidence levels and support metrics within the dataset.

Significantly, we implemented the Quantile Discretizer to segment price into 10 distinct bins, alongside categorization of seller names, domains, and countries. This segmentation reveals that attributes such as price range, seller identity, and country of origin exhibit high confidence in predicting outcomes, underscoring their potential as robust predictors in our analytical models

G. Multimodal Model

With the classification model, both image and text have to be appropriately transformed into numerical data that the model can be trained on.

1) ***Image Transformation (EfficientNet):*** A significant correlation exists between raw image data, necessitating preprocessing raw images into numerical feature maps. To achieve this, a pre-trained EfficientNet model, a Convolutional Neural Network that uniformly scales images of all dimensions, will be applied to the image data for converting images of different sizes into consistent-sized feature maps.

Additionally, since CNNs also serve as a classification method, the final layer will be removed from the model so that only the information from the images are preserved and later applied to the Multimodal Model. [11]

2) ***Text Transformation (DistilBERT):*** Description from the advertisements also serves a major role in providing its underlying information. Like Images, it's essential to convert them into numerical data. For this task, DistillBERT will be applied to the text. DistillBERT is a smaller and faster version of BERT, which is a Deep Bidirectional Transformer for Language Understanding. The model is utilized to tokenize information within descriptions which will then be feed into the Multimodal Model for classification.

3) ***CrossAttention:*** It should be clear that not all information from the image and description are helpful in the classification and there are even different importance levels across valuable data, thus Attention layers need to be applied to both image and text data for the model to learn about this information. Therefore, the CrossAttention layer is created to map attention across images and text.

4) ***Multimodal Model:*** Finally, the data from EfficientNet and DistilBERT redistributed through the CrossAttention will be passed in alongside normalized price and country into two fully connected Perceptron layers to formulate the final output. Since country also needs to be converted into numerical data, a country list is created for all existing countries in the dataset and using index as the country value for the model. To make the model fit more data, the country list can be expanded to include all countries and more data including those countries will be needed.

VI. Results

After data cleaning, the dataset size has decreased. To improve the model's performance on this limited data, random transformations such as horizontal flips are applied to the images during each epoch, increasing the diversity of the dataset.

Next, the dataset is split into training and testing datasets with an 80% to 20% distribution, respectively. The model is then trained for 20 epochs, yielding the following results:

A. Model Performance

The table below shows the performance of the model on both the training and testing datasets for precision, recall, and accuracy:

Dataset	Precision	Recall	Accuracy
Training	99.07%	99.53%	99.58%
Testing	85.11%	93.02%	94.08%

TABLE I
Model performance metrics for training and testing datasets

The performance metrics indicate that the model achieved near-perfect results on the training data and also performed strongly on the testing data, though with a slight decrease in precision and recall.

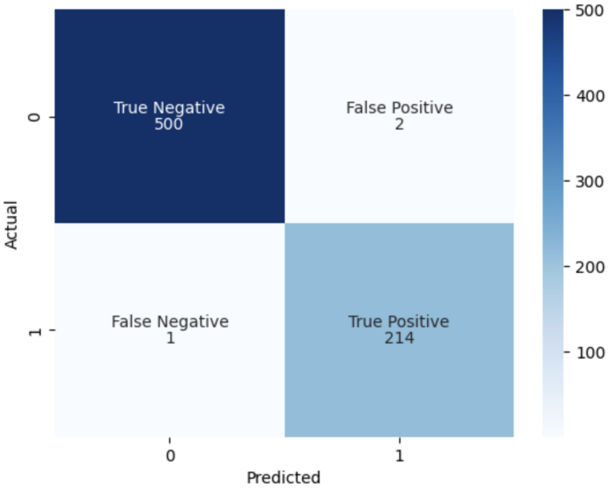

Fig. 7. Training Confusion Matrix

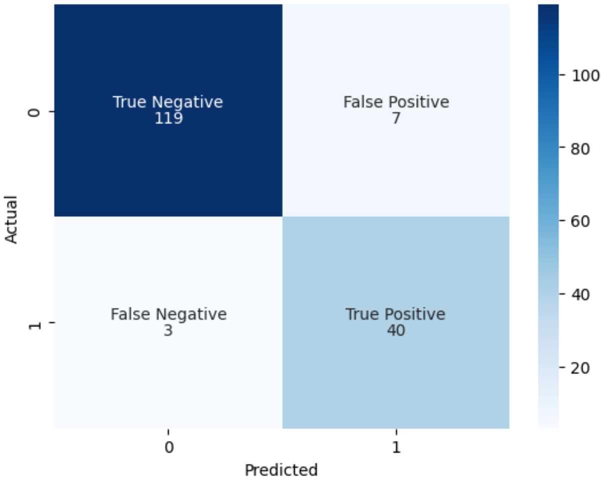

Fig. 8. Testing Confusion Matrix

B. Discussion of Results

The model demonstrates excellent performance on the training dataset with near-perfect precision, recall, and accuracy. On the testing dataset, the model still achieves strong results, with precision at 85.11%, recall at 93.02%, and accuracy at 94.08%. The slight decrease in performance on the testing data compared to the training data is typical in machine learning models, as the model is exposed to previously unseen data.

The relatively high recall on the testing data indicates that the model is successful at identifying positive instances, though there is a trade-off with precision, meaning that some false positives were likely classified. Further fine-tuning may help improve precision while maintaining strong recall and accuracy.

C. Future Work

To further improve the model's performance, additional training data could be added to enhance generalization. Additionally, further hyperparameter tuning and exploration of more advanced architectures could help narrow the performance gap between training and testing data, especially in terms of precision.

Acknowledgment

We thank Professor Juliana Freire for giving us access to the existing dataset and codebase on NYU JupyterHub.

References

[1] V. Pirotta, K. Shen, S. Liu, H. T. H. Phan, J. K. O'Brien, P. Meagher, J. Mitchell, J. Willis, and E. Morton, "Detecting illegal wildlife trafficking via real time tomography 3d x-ray imaging and automated algorithms," *Frontiers in Conservation Science*, vol. 3, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fcosc.2022.757950

[2] G. K. Moloney and A.-L. Chaber, "Where are you hiding the pangolins? screening tools to detect illicit contraband at international borders and their adaptability for illegal wildlife trafficking," *PLOS ONE*, vol. 19, pp. 1–25, 04 2024. [Online]. Available: https://doi.org/10.1371/journal.pone.0299152

[3] E. R. Rush, E. Dale, and A. A. Aguirre, "Illegal wildlife trade and emerging infectious diseases: Pervasive impacts to species, ecosystems and human health," *Animals*, vol. 11, no. 6, 2021. [Online]. Available: https://www.mdpi.com/2076-2615/11/6/1821

[4] S. Nalluri, S. J. R. Kumar, M. Soni, S. Moin, and K. Nikhil, "A survey on identification of illegal wildlife trade," in *Proceedings of International Conference on Advances in Computer Engineering and Communication Systems*, C. Kiran Mai, B. V. Kiranmayee, M. N. Favorskaya, S. Chandra Satapathy, and K. S. Raju, Eds. Singapore: Springer Singapore, 2021, pp. 127–135.

[5] E. Di Minin, C. Fink, H. Tenkanen, and T. Hiippala, "Machine learning for tracking illegal wildlife trade on social media," *Nature Ecology & Evolution*, vol. 2, no. 3, pp. 406–407, 2018. [Online]. Available: https://doi.org/10.1038/s41559-018-0466-x

[6] X. He and L. Deng, "Deep learning for image-to-text generation: A technical overview," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 109–116, 2017.

[7] A. S. Das and S. Saha, "Self-supervised image-to-text and text-to-image synthesis," *CoRR*, vol. abs/2112.04928, 2021. [Online]. Available: https://arxiv.org/abs/2112.04928

[8] X. Zhang, Y. Lu, W. Wang, A. Yan, J. Yan, L. Qin, H. Wang, X. Yan, W. Y. Wang, and L. R. Petzold, "Gpt-4v(ision) as a generalist evaluator for vision-language tasks," 2023.

[9] T. Xiao and J. Zhu, "Introduction to transformers: an nlp perspective," 2023.

[10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2019.

[11] F. Phe, "Paying attention to text and images for visual question answering," https://blog.dataiku.com/paying-attention-to-text-and-images-for-visual-question-answering, 2023.