

Statistical Data Analysis using GPT3:

An Overview

ASHWIN SHARMA

Computer Engineering

Thakur College of Engineering and
Technology

Mumbai, India

sharmaashwin4000@gmail.com

DISHA DEVALIA

Computer Engineering

Thakur College of Engineering and
Technology

Mumbai, India

devaliadisha06@gmail.com

WILFRED ALMEIDA

Computer Engineering

Thakur College of Engineering and
Technology

Mumbai, India

almeidawilfred642@gmail.com

DR. HARSHALI PATIL

HoD Computer Engineering
Department

Thakur College of Engineering and
Technology

Mumbai, India

harshali.patil@thakureducation.org

ANIKET MISHRA

Professor Computer Engineering
Department

Thakur College of Engineering and
Technology

Mumbai, India

aniket.mishra@tcetmumbai.in

Abstract— Though automated statistics has started gaining some momentum in the field of data analysis, it is not unified and very slow with large datasets. Due to computing limitations or lack of specific domain knowledge, general statistics have been used most commonly. But now research advisors are attracted towards a machine learning-based approach for statistical analysis of Data Sets which may help bridge gaps between traditional approaches like correlation matrices, p-values, etc., and new models like GPT3. This paper proposes a novel approach for the analysis of large datasets which uses GPT3 to predict insights from calculated statistics of data. The research addresses the limitations of existing methods and proposes a novel framework to analyze large statistical data sets, which solves many computationally challenging problems in efficient ways. Our proposed method works on top of GPT3's features, where it learns to predict individual words from particular parts of the dataset you pass as prompts (cumulative sums/means etc.) enabling us to analyze extremely large datasets such as telecom churn or census data. A comparison of traditional methods, statistical analysis, and machine learning approaches with GPT3 will be made. Furthermore, a discussion on the pros and cons of using GPT3 for this research is also discussed from the point of view of performance, accuracy, and reliability concerns.

Keywords— GPT3, prompts, Data Insights, Data Analysis using GPT-3

I. INTRODUCTION

GPT-3 (Generative Pre-trained Transformer) is a third-generation, autoregressive language model that uses deep learning to provide human-like text. Or to place it more simply, it's a computational system designed to generate sequences of words, code, or other data, starting from a source input, called the prompt[1]. It's used, for instance, in computational linguistics to predict word sequences statistically. Using GPT-3 is elementary, no harder than searching for information through a search engine[2]. In the same way as search engines "reads" our queries without understanding them, and offers relevant answers, likewise, GPT-3 writes a text continuing the sequence of our words (the prompt), without any understanding.

GPT-3 is not just incredibly powerful but large as well. Having 175 billion parameters and the whole internet as its training corpus, GPT-3 describes itself to be the most efficient and perfect NLP model to work with complex data relations. In this paper, we focus on how we can use GPT-3 for analyzing datasets using prompts. We plan to analyze three different datasets using GPT-3 and compare the results to traditional analysis methods to provide benchmarks on the accuracy of GPT-3 using a parametrical statistics approach. We conduct experiments on Ecommerce Sales, Heart attack, and telecom churn rate datasets. For each dataset, we use GPT-3 to analyze the dataset using prompts and compare its accuracy with traditional methods of analysis. We find that GPT-3 is able to provide reasonable insights on datasets in a very small amount of time but misses out on some information due to which it might not be as accurate as expected for every use case.

Despite GPT-3 being trained on only language modeling objectives, these models can perform relatively well at new tasks that they need not been explicitly trained to perform. the reason being large language models generalize to new tasks because of an implicit process of multitask learning[1]. As a by-product of learning to predict the subsequent word, a language model is forced to learn from a mixture of implicit tasks included in their pretraining corpus[3].

The main target of this paper is on looking at ways we can use GPT-3 for performing analysis on a dataset using prompts. We decide to perform analysis on three different datasets stated in table 1 and compare the results with results from traditional analysis methods to benchmark the accuracy of analysis done by GPT-3. [2]

TABLE I. DATASETS USED

Set. No.	Details of Dataset
D1	Ecommerce Sales
D2	Heart Attack Prediction
D3	Telecom churn dataset

Our proposed method will help analyze extremely large datasets using GPT-3 and provide insights into them. We also plan to compare traditional methods for statistical analysis with our proposed method to see the pros and cons of using GPT-3 as a tool for analysis based on the performance, accuracy, and reliability of the results.

II. LITERATURE REVIEW

[3] Proposed a method to do what the user means instead of doing whatever he/she says. In this study, they discussed that natural language interaction features can be designed in Tableau and how it will help the user to minimize their effort while creating data visualizations using the NLI tool thereby improving UX [3]. They first describe a pre-design Wizard of Oz study that offered insight into this question and narrowed the space of designs under consideration. The work provides insights into the design of natural language interaction in visual analytics as well as a reflection on the value of pre-design empirical studies in the development of visual analytic systems. Our research is similar to theirs where we propose GPT3 based Statistical Data Analysis Tool which takes input as dataset processed statistics than outputting the insights extracted from the data, which would otherwise require complex coding by both coders and non-coders alike saving time and efforts required for same.

[4] Proposed an Analytics-as-a-Service (AaaS) Tool for Unstructured Data Mining. They have introduced a framework that is capable of extracting and organizing terms and topics from unstructured data sources such as NoSQL databases etc. [4]. Moreover, they have also justified the architectural design of their proposed framework in detail which includes algorithms like concurrency search, linear search, etc. Also, its performance has been observed by conducting macro tasks like filtering and tagging to find out the most important features of an unstructured data set stored in a NoSQL database or any other source. Our research will be based on similar grounds but our primary focus will be on using GPT3 to perform statistical analysis on structured data sets present publicly at different websites/data repositories like Kaggle or UCI Machine Learning repository available online, which can extract various features from it without much effort. Compared to traditional methods used earlier to analyze statistical datasets because GPT3 has shown tremendous performance when it comes to natural language processing thus making way for a new era where AI's are able to predict future outcomes with high accuracy thus making business predictions more efficient than before especially when there is no clear pattern between independent variables X_1, X_2, \dots, X_n

[5] Proposed a broader definition of big data that captures its defining characteristics [5]. Their study also reinforces the need to devise new tools for predictive analytics using machine learning which is a subset of artificial intelligence in the field of computer science that often uses statistical techniques to give computers the ability to "learn" (i.e., progressively improve performance on a specific task) with data, without being explicitly programmed. Our research will be based on similar grounds where we analyze whether GPT can successfully solve complex statistical problems like

hypothesis testing/parametric approach and make useful predictions from textual datasets or not. Similarly, by using datasets like Census Data/Academic Performance Scores, etc., we may find out whether GPT3 can be useful in solving real statistics problems or not. Moreover, our study may also help us evaluate how well does it perform compared to other existing models?

III. METHODOLOGY

In this paper, we propose a practical framework for the statistical analysis of huge datasets using GPT-3 API. This requires a unique approach to prompt the zero-shot model and feed it with adequate data in order to extract useful insights from it[8][10]. Our proposed system follows the below-mentioned steps to perform its operations.

1) Data Pre-processing

GPT-3 API has the ability to extract information from raw data with the help of its zero-shot model [8] [10]. But there is an existing limit of 16kb or 4000 tokens per request. So firstly, we need to pre-process the dataset in order to calculate the required statistics of the dataset. These statistics are provided as prompts to GPT-3 in order to generate insights from that data. As this method can be used to shrink millions or probably billions of lines of data into a statistically reasonable and easily understandable output, hence it saves us a lot of time and effort.

This pre-processing involves sorting the data in ascending order, taking their cumulative sum/means, median values, etc.

Parameters calculation:

- a) Count b) Mean c) Median d) Max e) Min f) Skewness
- g) Standard Deviation h) Correlation Matrix

2) Data Transformation

The next step involves processing data into the format that is required by GPT-3. This can be achieved after performing some mathematical operations on the raw dataset [11]. In our proposed system all these transformations are done in Python. We do not directly use the whole dataset for analysis but rather transform it into a more generic form like cumulative sums, means, correlations, etc. There is no specific limit of lines of data or number of columns since the GPT-3 API has zero-shot learning capabilities and doesn't require the data to be labelled for analysis, however, the total length of input for statistical data should not exceed 16kb or 4000 tokens per request [6]. Also, it is imperative that we pass the dataset prompts in a comma-separated format.

3) Analysis of Dataset

Next, the dataset into the processed data is fed into GPT-3 for analysis. Primarily we decided to process our dataset in two different ways in which GPT-3 can perform analysis on it for the best possible outcomes. Firstly, we provide the GPT-3 API with the Correlation matrix of the dataset as one prompt and other statistical data like median, mean, standard deviation, etc. as another prompt in order to generate useful insights from it[12][13]. We also performed analysis on raw data by processing and labelling it before

passing it into the model as prompts in order to see if the accuracy increases or not compared to processing the transformed data statistically.

IV. RESULTS OF ANALYSIS USING GPT3

Code for processing the raw data into statistical data like mean, median, skewness, etc. has been written in python. The prompts are pre-defined questions that the GPT3 API analyzes and answers. The prompts can be any natural language query or question that you would like to ask about the dataset. Figure 4 & 5 shows an example of how the queries or prompts are provided to GPT3 API. The output obtained by GPT3 is compared with results from traditional analysis methods (Manual data plotting and pattern analysis) to benchmark the accuracy of analysis done by GPT-3. For each dataset, we conduct two processes. First to transform the data into statistical information and correlation matrix; then analyze the data using GPT3.

STATISTICAL METHODOLOGY

The paper demonstrates the use of GPT-3 to automatically generate insights from statistical data. Specifically, the paper shows how the combination of processed statistical information (e.g., mean and standard deviation of individual columns, correlation values, etc.), attribute labels (e.g., "category" information), and prompts can be used to predict short-form phrases in the language of the author's choosing.

The framework used for this analysis is based on GPT-3, which enables the use of prompts to extract insights from statistical data rather than manually creating a list of insights for prediction by the model. Different variations of the prompts were used to influence the outputs of predicted insights, allowing for the generation of multiple sets of predictions based on statistical data. Overall, GPT-3 provides a sophisticated way to represent conclusions or insights as short-form phrases by taking advantage of its unique prompt/response feature.

PARAMETERS GPT-3

GPT-3 is a language model that can generate natural outputs from inputs. It uses different types of inputs and combines them in its proposed approach, including continuous features like cumulative sums or counts along with Boolean flags indicating non-continuous information such as column data type (nominal/numeric), etc.

Key parameters that affect prompt outputs:

Temperature: It controls creativity of model. Values < 0.6 will not be creative and > 1.0 will have no context sensitivity. Hence, We used values in 0.7 to 0.9 range.

Top-p: Controls diversity via nucleus sampling approach. We used value 1.

Max tokens: We use top-p to prune the outputs so that only meaningful insights are generated. We used values in 400 to 500 range.

Frequency penalty: It controls for rare words and hence, reduces repetition by fixing the top-k value. We used 0.44

Presence penalty: It controls for repetitive words and hence, reduces repetition by fixing the top-p value in function. We used 1.5, to let the model draw conclusion based on it's wide domain of training data, a high presence penalty was required while tuning.

Many different input parameters were tried out to understand what affects the performance of this research study on a sample dataset "Telecom churn rate" obtained from kaggle.com (the final experiment was carried out on real-world datasets like census and telecom churn).

GPT-3 is a machine learning model that takes in descriptive statistics like mean and standard deviation, as well as nominal/numeric flags, to make predictions about data. In addition to these, GPT-3 can also draw conclusions based on its knowledge of whether a particular attribute value is continuous or not, along with other information such as the maximum values of each column, etc.

The following are the statistical attributes used by GPT:

Mean: Higher values imply a greater proportion of high-value records and lower mean implies more low-valued records in that column.

Standard deviation: Standard curve distribution(bell shape) measures the variation of individual results. Lower values imply less deviation in results and higher values imply more deviation in results.

nth percentile: Outliers will tell us about the presence of extremely high or low values that fall 3 times more than the standard deviation from the mean.

Skewness and Kurtosis: Kurtosis is used to measure how peaked a distribution is, while Skewness tells us if there are outliers in the data. Skewness can be used to describe the direction of skew (i.e., left-skewed vs right-skewed) in order to draw conclusions regarding the correlation between different values or deviation from mean allowed by upper bound computations like $1.5 * \text{std_dev}$ above mean or below. Skew & kurtosis connote information similar to that given by standard deviation but with more details on outliers

The above data briefs us about the statistical data this paper uses for insights extraction from raw data.

Descriptive statistics are used to summarize the characteristics of a dataset by describing or presenting features using arrays, graphs, charts, and tables of values (mean, median, etc.). This approach is used to understand the dataset with minimal use of statistical procedures or methods.

DATA TRANSFORMATION

Statistical data analysis fascinates most of the existing approaches with high accuracy and reliability. We generated the correlation matrix and used it in prompt input.

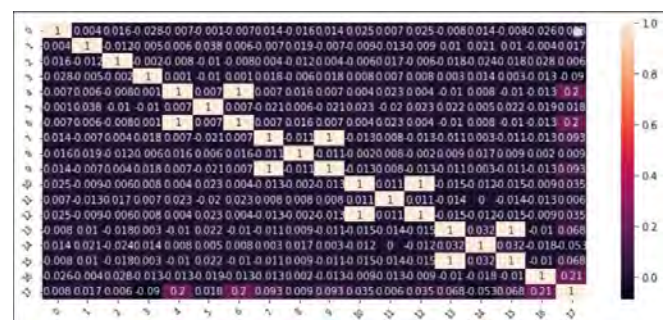


Fig. 1. Correlation matrix for providing more information

There are parameters to be adjusted in GPT-3 API model. In our results, we decided to use code-davinci-002 and text-davinci-002.

```
temperature=0.72, max_tokens=527, top_p=1,
frequency_penalty=0.44, presence_penalty=1.49,
stop= ["-> Insight 11:"]
```

A. Analysis of dataset D1

a) Dataset

The dataset consists of a historical sales record of 3 branches of a supermarket company over 3 years.

b) Prompt

The model was given the labeled correlation matrix and some statistics about the data and was asked to generate 10 insights on the data.

c) Generated Output

1. Product line: Health and beauty have the highest mean price. So, customers may prefer to buy the most expensive products in the health and beauty category than any other product category. This may be because of the fact that people pay more attention to their body looks or even they are using costly perfumes or cosmetics which are from this product category only.
2. Payment mode: Electronic is used maximum times but its average amount is low as compared to cash payment mode. The reason behind that can be electronic payments like debit card, credit card has lower transaction fees for processing payments thus stores provide a discount on these modes of payment which reduce the cost price and hence the net transaction value by increasing the quantity sold in case of e-payments.

d) Discussions

The prompt was executed 10 times and each time distinct results were generated. The generated insights were all in human-readable sentences. Often the results were simply explaining a data point like two parameters being the same due to their correlation being 1. For the sales dataset, the model successfully pointed out the branches with the highest and lowest sales. It was able to rank store branches according to their ratings from highest to lowest.

```
# STATISTICS
['longitude', 'latitude', 'housing_median_age', 'total_rooms', 'total_bedrooms', 'populat
'median_income', 'median_house_value']
['count'][3000.0, 3000.0, 3000.0, 3000.0, 3000.0, 3000.0, 3000.0, 3000.0]
['mean'][-119.589, 35.635, 28.845, 2599.579, 529.951, 1482.799, 489.912, 3.807, 285846.27]
['std'][1.995, 2.13, 12.555, 2155.593, 415.654, 1030.543, 365.423, 1.855, 113119.687]
['min'][-124.18, 32.56, 1.0, 6.0, 2.0, 5.0, 2.0, 0.5, 22500.0]
['25%'][-121.81, 33.93, 18.0, 1401.0, 291.0, 780.0, 273.0, 2.544, 121200.0]
['50%'][-118.485, 34.27, 29.0, 2106.0, 437.0, 1155.0, 409.5, 3.487, 177650.0]
['75%'][-118.02, 37.69, 37.0, 3129.0, 636.0, 1742.75, 597.25, 4.656, 263975.0]
['max'][-114.49, 41.92, 52.0, 30450.0, 5419.0, 11935.0, 4930.0, 15.0, 500001.0]
['skew'][-0.298, 0.46, 0.019, 4.168, 3.863, 2.95, 3.56, 1.699, 0.99]
['kurtosis'][-1.363, -1.124, -0.804, 32.044, 28.488, 16.414, 26.184, 5.615, 0.393]
```

Fig. 2. Correlation matrix parameters calculated and defined as prompt

Some insights were generated on the distribution and skewness of the data. Often times the generated insights were sentences explaining data points like the average custom spending and transaction amount. These insights however save the effort of visualizing the data to determine some information like understanding the most popular mode of payment.

The model was able to determine the relationship between various entities of the data and successfully interpreted how a variance in the value of one entity can affect other values. It also concluded that the data was not clean and had some missing values. However, this observation is incorrect because the prompt had preprocessed data which was cleaned and null values were dropped.

It can be concluded that the quality of results generated depends on the quality of the data. The prompt used had two only two parameters, the correlation matrix, and the statistical data. Providing a prompt with more data points can yield in more accurate results.

B. Analysis of Dataset D2

a) Dataset

The dataset consists of pasts records of patients who have had a heart attack.

b) Prompt

The model was given the labeled correlation matrix and some statistics about the data and was asked to generate 10 insights on the data.

c) Generated Output

10 insights from the above data are as follows:

1. Insight 1:

Out of 303 observations, 204 contain chest pain-type information. Of these only 143 mentions slope parameters and 138 have a number of major vessels colored by fluoroscopy information. This means that for almost 52% and 46% data we don't have full information respectively.

2. Insight 2:

There are 72% males and 28 % females in our dataset which is approximately close to real-world scenarios where males suffer from heart diseases more than females.

3. Insight 3:

58% people don't have coronary artery disease while 42 % do have it which hints to us that this might be an imbalanced dataset where our target class is outnumbered by other classes.

4. Insight 4:

The age group between 60-70 seems to be suffering most from heart diseases as per the given data followed by the age group 70-80. Also, it is seen that the maximum number of people having CAD belongs to the age group 50 - 60 while the minimum belongs to the 30 - 40 age group.

From the above analysis, we can see that the patients included in this dataset are relatively older, with a cholesterol

level that is slightly high but within normal limits. On average, they have a higher systolic blood pressure and a lower diastolic blood pressure, which suggests that they may be suffering from hypertension. Their resting heart rate is also on the higher side, although their maximum heart rate is within normal limits

d) Discussions

Most of the time results were simply the average values from the given dataset but this may however save the time and effort of visualizing the data.

Although the model points out that the maximum heart rate is within normal limits it is to be noted that this is not actually the case.

Just to test whether the generated insights were random or meaningful insights, we provided the model with the imbalanced dataset. To our surprise, the model did point out that the dataset might be imbalanced in one of the insights. The model was also able to determine relations between data points and linked them to a probable cause of heart attack (hypertension)

C. Analysis of Dataset D3

a) Dataset

Dataset of telecom churn data consisting of calls, minutes, and charges for different users.

b) Prompt

The model was given the labeled correlation matrix and some statistics about the data and was asked to generate 10 insights on the data.

c) Generated Output

1. Insight 1:

The parameters which have the highest correlation with Churn are Total eve charge followed by Customer service calls and Total intl minutes. State code has no effect on Churn since it has a very low Pearson Correlation Coefficient of 0.004 with Churn.

2. Insight 2:

From the statistics of the data, we can see that the Total eve charge has a mean of 17.084 and a standard deviation of 4.311 with a 50% value of 17.12 which means that 50% of the total eve charges lie between 12.81 and 21.43 where most people tend to churn if their Total eve charge goes higher than 21.43

3. Insight 3:

From the correlation matrix, Customer service calls are also very highly correlated with Churn at 0.209 so it will be very easy for us to predict churn using this parameter because of its high coefficient values due to insight 1 we have come up with, I believe Total eve charge would give us more accurate results in predicting churn compared to customer service calls since they both are highly correlated with Churn

d) Discussions

The prompt was executed 10 times and each time distinct results were generated. The model generated insights in human-readable sentences which successfully explained the data from the provided set of parameters. It also stated which entities have more relation with churn, total eve charge, and customer service calls since they have a higher correlation with churn as compared to other entities.

This is a very practical problem as it will help telecom companies to figure out why users are likely to leave their services. Such insight can reduce expenditure for such companies because churn can cost millions when considered at giant scales like call centers having millions of users.

With the first few executions, this model produced some erroneous outputs concluding that the correlation matrix contains negative values while it only contained positive values. This points out how precision depends on how clean your data is because the model suggests correlations might be negatively correlated due to seeing some negative values.

Since the prompt had preprocessed data consisting of only two parameters, correlation matrix and statistics, it is possible that providing more data points as given in the actual dataset might yield better result as then we will be able to see how churn depends on various factors other than correlating correlated factor with churn. Though this model was not very precise but still managed to provide insightful information about the data which can be used by telecom companies to reduce their revenue losses due to customer churning.

CONCLUSION

In this paper, we analyze the ability of the GPT-3 model to provide useful insights from raw and preprocessed datasets. We evaluate the model on three different types of datasets and find that it is accurate most of the time, providing insights on the rates of heart patients per 1000 individuals. While interpreting the telecom churn data, it has been found that customers are more likely to leave a company if their total evening charges are higher than other parameters in the dataset. Super Market Company dataset provided us with insights into how customers are more likely to choose products from expensive product lines like health and beauty over other categories. This implies that customers prefer companies that provide them with cheaper rates at nighttime. Human-readable sentences. However, GPT-3 fails at some points; for example, in telecom churn data analysis, it used statistics as insights without actually incorporating any domain knowledge.

The language model GPT-3 from OpenAI has great potential to interpret tabular data, but its performance depends on the quality of the data. If the model performs well on a given set of parameters, it will likely perform much better on a larger dataset with multiple parameters. However, since the model has not been trained on extracting insights, it may still have some errors in predictions. Despite this, the GPT-3 API could be used by many companies to reduce expenditure, for example by predicting the churn rate of telecom customers. The API could also serve as a great analyst tool, helping analysts to figure out why an event happened by feeding data into the prompt - something which was previously done manually.

Limitations of using GPT3 for Analysis

This proposed model is not perfect and has some limitations too which are discussed below:

1. Training data used for GPT-3 is limited to the whole internet, so if your data does not match with it then GPT-3 might fail to extract correct insights from it. Using real-world datasets like census or telecom churn rate etc. might help in finding more accurate results than examples taken from Kaggle or UCI ML repository because those datasets can have bias while generated them [7].
2. The total length of input for statistical data should not exceed 16kb or 4000 tokens per request thus it may be hard to find patterns between parameters having very high skew values.
3. It also misses correlation among different parameters as our experiments show that it performed poorly when we took raw data sets as prompts where we need to perform labeling before classification.
4. Results generated by GPT3 may be dependent on the quality of data provided to it [9]. Like if your data is not clean and has null values, then GPT3 might fail to predict some insights accurately.

REFERENCES

- [1] SearchEnterpriseAI. (n.d.). *What is GPT-3? Everything You Need to Know*. [online] Available at: <https://www.techtarget.com/searchenterpriseai/definition/GPT-3>.
- [2] V. Sanh, A. Webson, C. Raffel, S. H. Bach, L. Sutawika, Z. Alyafeai, A. Chaffin, A. Stiegler, T. L. Scao, A. Raja, M. Dey, M. S. Bari, C. Xu, U. Thakker, S. S. Sharma, E. Szczechla, T. Kim, G. Chhablani, N. Nayak, D. Datta, J. Chang, M. T.-J. Jiang, H. Wang, M. Manica, S. Shen, Z. X. Yong, H. Pandey, R. Bawden, T. Wang, T. Neeraj, J. Rozen, A. Sharma, A. Santilli, T. Fevry, J. A. Fries, R. Teehan, T. Bers, S. Biderman, L. Gao, T. Wolf, and A. M. Rush, "Multitask prompted training enables zero-shot task generalization," *arXiv.org*, 17-Mar-2022. [Online]. Available: <https://arxiv.org/abs/2110.08207v3>.
- [3] Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., Smith, N. and Allen (2020). *Don't Stop Pretraining: Adapt Language Models to Domains and Tasks*. [online] pp.8342–8360. Available at: <https://aclanthology.org/2020.acl-main.740.pdf>.
- [4] Lomotey, R.K., and Deters, R. (2014). Analytics-as-a-Service (AaaS) Tool for Unstructured Data Mining. 2014 IEEE International Conference on Cloud Engineering. pp.319-324. doi:10.1109/IC2E.2014.15.
- [5] Ogsulee, P., Chotchuaung, V., Bamrunsi, E., and Rodcheewit, T. (2018). Big Data, Predictive Analytics and Machine Learning. 2018 16th International Conference on ICT and Knowledge Engineering (ICT&KE). pp.1-6. doi:10.1109/ICTKE.2018.8612393.
- [6] Radford, A. (2019). *Better Language Models and Their Implications*. [online] OpenAI. Available at: <https://openai.com/blog/better-language-models/>.
- [7] www.kaggle.com. (n.d.). *How to find your model is bias?? | Data Science and Machine Learning*. [online] Available at: <https://www.kaggle.com/general/195939>.
- [8] Cooper, K. (2021). *OpenAI GPT-3: Everything You Need To Know*. [online] Springboard Blog. Available at: <https://www.springboard.com/blog/data-science/machine-learning-gpt-3-open-ai/>.
- [9] Romero, A. (2022). *The New Version of GPT-3 Is Much, Much Better*. [online] Medium. Available at: <https://towardsdatascience.com/the-new-version-of-gpt-3-is-much-much-better-53ac95f21cfb>.
- [10] Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C. and Hesse, C. (2020). Language Models are Few-Shot Learners. *arxiv.org*. [online] Available at: <https://arxiv.org/abs/2005.14165>.
- [11] Tyagi, N. (n.d.). *What is the OpenAI GPT-3? | Analytics Steps*. [online] www.analyticssteps.com. Available at: <https://www.analyticssteps.com/blogs/what-openai-gpt-3>
- [12] Pezzotti, N. and Hall, C. (2021) *GPT-3 for few-shot dialogue state tracking*, *Cam.ac.uk*. Available at: https://www.mlmi.eng.cam.ac.uk/files/2020-2021_dissertations/gpt_3_for_few_shot_dialogue_state_tracking.pdf.
- [13] Wang, L. et al. (2021) "An evaluation of generative pre-training model-based therapy chatbot for caregivers," *arXiv [cs.HC]*. doi: 10.48550/ARXIV.2107.13115.