ISSN 2409-2665
Journal of Logistics, Informatics and Service Science
Vol. 11 (2024) No. 2, pp. 324-345
DOI:10.33168/JLISS.2024.0220

AI Ads: Practicability of Text Generation for F&B Marketing

Amos Chow Li Kuang[1], Tong Ming Lim[2], Chi Wee Tan[1], Chuk Fong Ho[1],

Noor Aida Husaini[1]

[1] Faculty of Computing and Information Technology,Tunku Abdul Rahman University of Management and Technology, Kuala Lumpur, Malaysia

[2] Centre for Business Incubation and Entrepreneurial Ventures,Tunku Abdul Rahman University of Management and Technology, Kuala Lumpur, Malaysia

amosclk-wm20@student.tarc.edu.my

Abstract. Despite numerous studies and implementations on text generation in the field of Artificial Intelligence (AI), the purpose of this research is to study the challenges and practicality of using Natural Language Generation (NLG) on Food and Beverages (F&B) marketing text generation using generative AI techniques. Since there are a number of different researchers that claim they are able to generate realistic text that is acceptable, we would like to investigate the degree to which AI models can effectively generate marketing text for advertising purpose. The goal of this research is to investigate several existing techniques for text generation where implementation of Long Short-Term Memory (LSTM), Open Pretrained Transformer (OPT) and Keyword to Text Generation (K2T) will be experimented. Locally collected marketing text samples will be applied to train and validate based on how realistic the generated texts are from the perspective of human expert based on realistic and practical attributes. The text-based advertisement will be developed and reviewed in the current research. A discussion based on the models' theoretical and algorithmic fundamentals has been presented from the perspective of models' performance.

Keywords: Natural Language Generation, Keyword to Text Generation, F&B Marketing Advertisement, Long Short-Term Memory, Open Pretrained Transformers

1. Introduction

Marketing plays an important role in building customer loyalty and relationships with them, it is also critical for a business's success because it helps generate revenue, which is essential for the organizations to continue to operate profitably (Hill, 2010). In the field of advertising, it is a process to attract customers by capturing their attention on businesses such as system integrators (Kiselicki, 2017; Petrevski et al., 2017), database marketers (Khan et al., 2023; Leary, 2023), telemarketers (Ghatasheh et al., 2020; Tékouabou et al., 2022) and internet firm products (Chatzoglou and Chatzoudes, 2016; Wang and Vaughan, 2014) to be adopted in the field of marketing to increase profit in order to compete with one another (Mergent, 2022).

An advertisement for an organization and its products that is paid for and distributed to a targeted audience may be marketed through mass media channels (Kotler and Armstrong, 2021). Advertising serves the purpose of promoting the sale of goods and services through various media outlets, such as television, radio, print, and social media. Its goal is to reach a wide audience and convince them to purchase the advertised products or services, while also raising brand recognition and providing information about new offerings (Sama, 2019).

Natural Language Processing (NLP) is also one of the fields of AI that focuses on the interaction between computers and humans using natural language. It involves development of algorithms used to build models with training data that can understand, interpret, and generate human language. Some common applications of NLP include language translation (Ghasemi and Hashemian, 2016; Kolhar and Alameen, 2021), text classification (Fucci et al., 2022; Yue et al., 2022), sentiment analysis (Norambuena, Lettura and Villegas, 2019; Zhou, 2022), and chatbots (Liu, Lin and Sun, 2020; Shum, He and Li, 2018).

There are challenges when it comes to the generalizability of NLG tools. They may not perform well when they are used in a different domain from the one they were originally trained on. To maintain good performance in new settings or domains, it may be necessary to be retrained or perform domain adaptation (Groot et al., 2021).

This means that different fields of knowledge may not apply well to another field since every NLG algorithm is learned from the training data. For instance, medical data used for model training cannot be applied to marketing. It requires data sources related to marketing and alternative data pre-processing and fine-tuning procedures prior to model training. Designing a model that can generate effective marketing advertisement text is a time consuming and complex process that requires careful planning and attention to detail (Beltis, 2021).

For this investigation, the following objectives have been set:

- To study different approaches to generate text by using three generative NLP models.
- To develop a text-based marketing-related advertisement application for testing and validationpurposes.

In this research, three generative AI algorithms in NLG have been implemented to test their practicality and level of realistic outcome produced by each model. These models have been validated by human experts. Specifically, this article focuses on three algorithms: Long Short-Term Memory (LSTM), (Hochreiter, 1997), pretrained transformers such as Open Pretrained Transformers (OPT) (Zhang et al., 2022), and KeytoText (K2T) (Gagan, 2021), powered by Google T5 (Raffel et al., 2020). These algorithms will be studied and conducted as experiments to generate advertising text for the (F&B) industry.

2. Literature Review

Past research papers on generative AI models will be reviewed. Section below will focus on LSTM, OPT, K2T, GAN, and GPT-3 in depth. According to previous researchers (Fatima et al., 2022), there are two approaches to deep learning models in the context of text generation. These approaches include traditional and advanced methods. Under the traditional model category, researchers have explored recurrent neural networks (RNN), long short-term memory (LSTM), gated recurrent units (GRU), and convolutional neural networks (CNN).On the other hand, advanced models typically involve large language models (LLM) or pre-trained models, such as attention-based models, Transformers, and BERT.

According to (Tighe et al. 2020), researchers have conducted a deep learning-based text generation by adopting Word2Vec, a technique that converts words from each abstract into vectors, considering the words that are in proximity to them. This approach contributes to implementing and conducting surveillance measures of NLP methodology and analyzing output. In another study, Strobelt et al. (2022) developed a prototype and framework called Generation Negotiation Interface (GenNI). GenNI aims to facilitate collaboration between humans and AI in the complex field of data-backed text generation using machine learning systems. It assists users in creating actionable constraints that are compatible with systems intended for user controllable text generation.

According to research (Devlin et al., 2019), few researchers from Google AI team have introduceda Bidirectional Encoder Representations from Transformers (BERT), it is a Large Language Model (LLM) as well as a transformer provides the usability in terms of training unsupervised text data. It provides clarity regarding the search engine in Google, receiving better understanding of user intent and return a more relevant content (Roy, 2022). BERT contributed to involves extending the applicability of these findings to deep bidirectional architectures, thereby enabling a single pre-trained model to effectively address a wide range of NLP tasks.

2.1. Long Short-Term Memory (LSTM)

An LSTM network is a type of Recurrent Neural Network (RNN). A RNN is a type of neural network that is designed to process sequential data, unlike traditional neural networks that process fixed-sized inputs. Therefore, LSTM has been designed to be able to learn and retain long-term dependencies between input and output data. This is achieved using a special type of memory cell called a "long short-term memory" cell, which is able to hold on to information for extended periods of time, and gates that allow the network to selectively retain or forget information as needed. Figure 1 shows the design of LSTM and how the model processes the data, where (f_t) is the ForgetGate, (i_t) is the Input Gate and (o_t) is the Output Gate.

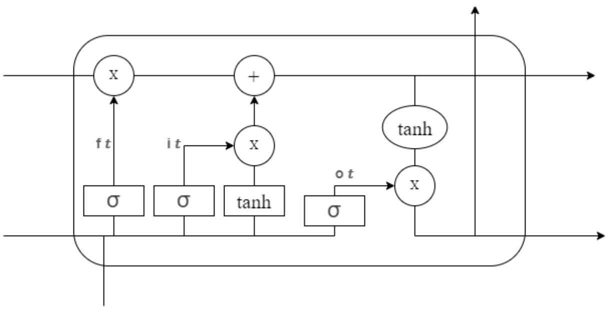

Fig. 1: Architecture of LSTM

The forget gate layer (f_t), which is the first sigmoid layer in an LSTM network, is responsible for determining what information should be removed from the cell state based on the input and short-term memory from the previous time step. The input gate layer (i_t), which is the second sigmoid layer, determines what new information should be added to the cell state to replace the forgotten information. The output gate layer in an LSTM (o_t) is responsible for controlling the flow of information from the memory cell to the output layer, by determining what information is going to be output at each time step. This new information is based on the result of the input gate layer multiplied by the output of a tanh layer, which is similar to the role of the tanh layer in a traditional RNN (Ryan, 2020).

LSTM networks are a specialized type of Recurrent Neural Network (RNN) that are designed to be able to effectively learn and retain long-term dependencies between input and output data. Traditional RNNs tend to struggle with this task due to the vanishing gradients problem, which makes it difficult for the network to propagate error gradients back through time (Kratzert et al., 2018; Lipton, 2015). However, LSTM networks address this problem through the use of special memory cells and gates that allow them to selectively retain or forget information, allowing them to effectively learn and model long-term dependencies (Bengio, Simard and Frasconi, 1994).

2.2. Open Pretrained Transformers (OPT)

The Meta AI team introduced Open Pretrained Transformer (OPT) in June 2022 which provides an open source of decoder-only pre-trained transformers ranging from 125 million to 175 billion parameters (Zhang et al., 2022). The OPT models are trained to achieve similar performance andsize to the GPT-3 models, using the most recent best practices in data collection and efficient training (Cameron, 2022). The Meta team developed the suite of OPT models with the goal of enabling large-scale, reproducible, and responsible research and including more diverse perspectives in the examinationof the effects of these Large Language Models (LLMs).

Since OPT models are open source, developers who choose to use them can save the energy-intensive initial learning process of creating their own from scratch (Davies, 2022). HuggingFace also provides the OPT-350m pipeline that is used for text generation, it shows that it is also useful in text generation by implementing the pipeline, or fine-tuning by user dataset (HuggingFace, 2022). It has been successfully nominated as one of the milestones that helps develop NLP programs by providing such useful pretrained parameters. Another research found that regarding to the Zero-shot NLP Evaluation Averages across a variety of tasks and model sizes, the performance of OPT has approximately matched the GPT-3 (Zhang et al., 2022).

2.3. Keyword to Text Generation (K2T)

Although ordinal sentence generation input is functional by using according to input words as a seed word to start generate sentence, there are still limitations especially the model does not provide user convenience in their desired field of text generation. In current research, a method that uses keyword text generation, that is categorized as a text-to-text generation based on Google T5 (Gagan, 2021) will be experimented.

T5 is a Text-to-Text Transfer Transformer and is a LLM developed by Google's AI research team. It is based on the Transformer architecture and is trained using a "text-to-text" approach, where the model is trained to map an input text to an output text. The T5 model is a pre-trained LLM on a massive dataset of text from the internet, including books, articles, and websites. It is trained to perform various natural language processing tasks, including language translation, question answering, summarization, and more. The text-to-text approach means that T5 can be fine-tuned for a specific task by simply providing a set of input- output examples in the form of text (Raffel et al., 2020).

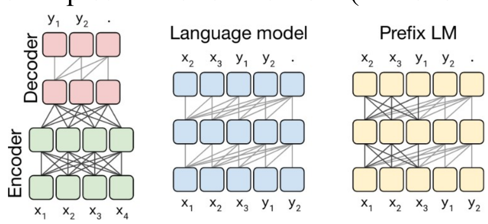

Fig. 2: Architecture of Google T5

Figure 2 illustrate the encoder-decoder architecture used in this language model is a standard transformer. The encoder component utilizes a fully visible attention mask, similar to the masking technique employed in BERT as a Pre-trained Transformer model (Devlin et al., 2019; Kharazi, n.d.). This allows each input token to contribute to the attention computation of every other input token in the sequence. The decoder component, on the other hand, employs causal attention during training, which means that only the previous output tokens contribute to the attention computation of the current output token in the output sequence. This attention mechanism is a fundamental aspect of the auto-regressive modelling approach utilized in this architecture, which is a combination of the BERT architecture and traditional language modelling approaches. T5 has achieved state-of-the-art results on several NLP benchmarks, the most used cases of Google T5 include machine translation, classification, question answering, and summarization. It is open-source and available for researchers and developers to use. This approach is different from LSTM and OPT where LSTM and OPT model construct sentencesby generating text based on user input to start with, for e.g., if "chicken delicious" is provided, LSTMand OPT will generate a sentence such as "chicken delicious". However, K2T generates based on several keywords provided.

2.4. Generative Adversarial Network (GAN)

Recently, researchers have achieved promising results in text generation by using Generative Adversarial Nets (GANs) and combining them with policy gradient methods. GANs use a discriminative model to guide the training of a generative model, treating it as a reinforcement learning policy.

A previous study found that one traditional method for text generation is to train a RNN to predict the next word in a sequence, given the previous words. The RNN is typically trained to maximize the log-likelihood of the ground-truth words, given the previous words (Graves, 2013). However, this approach suffers from a problem called exposure bias, which arises from the difference between the training and inference stages. During training, the model is presented with ground-truth words and must generate the next word based on them. During inference, the model must generate the next word based on the terms it has previously generated itself. However, this discrepancy can cause problems and there is inconsistency in using this RNN (Huszár, 2015).

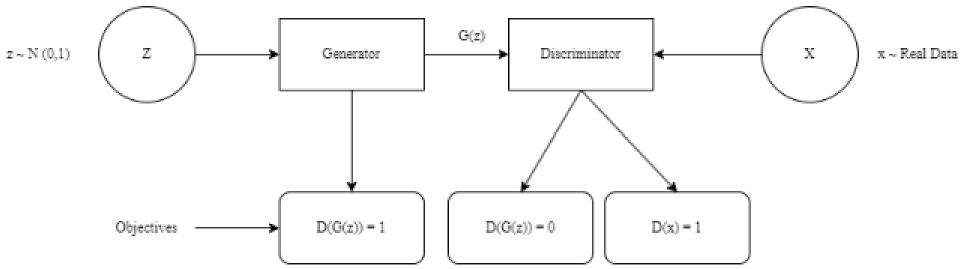

Fig. 3: Generative Adversarial Network Architecture

Figure 3 as designed by Goodfellow illustrates that GAN are specialized for image-generating models, which primarily deal with continuous data. Text generation can be viewed as a series of sequential decisions, where the state represent the previously generated words, the action corresponds to the next word to be generated, and the Generative Model (G) acts as a policy that maps the current state to a distribution over the possible actions, determining the next word. The generated text is then given to a Discriminative Model (D), which is trained to distinguish real and generated text samples and provide reward signals for updating the generative mode (Goodfellow et al., 2014) There is another approach of GAN which is LeakGAN proposed by Guo, which is a technique thatallows the generator in a GAN (Generative Adversarial Network) to receive more detailed feedback from the discriminator, enabling the generation of more nuanced and accurate output (Guo et al., 2018).

The main difference between LeakGAN and GAN is LeakGAN consist of a high-level Manager using an LSTM model, and a Low-level worker in the layer of generator (G). The generator in GAN is responsible for producing the next word in a sentence, while the discriminator evaluates the quality of the generated sentence. This process is adversarial, with the generator trying to produce realistic sentences and the discriminator trying to identify which sentences are generated and which are real. The key innovation of this approach is that the discriminator shares its internal state of feature with the generator during the training process, providing the generator with more detailed and frequent feedback to improve the quality of the generated sentences. Therefore, LeakGAN outperforms previous solutions and demonstrates significant improvements in performance in natural language processing applications.

2.5. Generative Pretrained Transformers (GPT)

Recently, the Generative Pretrained Transformers (GPT) is offered by OpenAI which specializes in AI research and development. The popular Chat-GPT model (Nerdynav, 2022) is originally in a previous version of GPT-2 launched in the year of 2019, GPT-3 in the year 2020 and GPT-4 in March of 2023 (OpenAI, 2023). GPT-2, a language model developed by OpenAI, is an improvement on the original GPT model. Itcan generate text in a manner similar to human language and has a larger capacity of up to 1.5 billion parameters, making it more powerful than the original GPT model. (OpenAI, 2019).

Table 2 shows the area of deploying GPT-2 in several fields. It provides assistance for humans especially in the field of text language understanding and generation. Apart from that, GPT-3 which is a new generation of GPT-2 contains up to 175 billion parameters,it also provides few-shot learning, one short learning and also zero-shot learning. Previous studies found that GPT-3 has demonstrated the ability to rapidly adapt to and perform tasks that require quickreasoning, such as rearranging scrambled words, performing arithmetic, and using new words in a sentence after only seeing them defined once. It also has the capacity to produce synthetic news articles that are difficult for human evaluators to differentiate from those written by humans in a few-shot setting (Brown et al., 2020).

GPT-3 has shown strong performance on tasks that are defined on-the-fly, sometimes nearly matching the performance of state-of-the-art fine-tuned systems, and producing high-quality samples. These results suggest that extremely large language models may play a crucial role in creating adaptable, general language systems. GPT-3 represents a significant advancement over GPT-2 in terms of size, capacity, and performance, but it comes at a higher cost since it has more parameters to be trained which means it is more resource intensive. Previous research also shows that the energy usage for training the GPT-3 175B is considered as a large-scale parameter that requires immense amounts of computational resources (Brown et al., 2020).However, GPT-3 has more accurate results while performing most tasks regarding accuracy evaluation, including language translation, text summarization, and language generation for chatbots and virtual assistants.

In addition, GPT-4 has the objective to development of infrastructure and optimization methods that demonstrate consistent behavior across a broad range of scales was a fundamental aspect. GPT-4 is a multimodal model capable of accepting both image and text inputs and generating text outputs. Although it may not be as competent as humans in certain real-world situations, it has shown human-level performance on numerous professional and academic benchmarks (OpenAI, 2023b). It exhibits superior English-language performance compared to GPT-3.5 as shown in Figure 5. Furthermore, it also outperforms them in low-resource languages like Latvian, Welsh, and Swahili. (OpenAI, 2023a). GPT-4 can incorporate Reinforcement Learning from Human Feedback (RLHF), then improve the model's adaptability and performance significantly, resulting in a highly efficient natural language processing system (Liu et al., 2023).

2.6. Summary of Review

As new generations of technology are developed, they tend to represent improvements over their predecessors. For instance, LSTM is an improved version of RNN, GAN uses LSTM to implement as a foundation, GPT provides a large number of parameters enabling the hybrid of several models including LSTM. Meanwhile. OPT is a pretrained transformer that provide open-source parameter based on the accuracy of GPT. Table 1 justifies the reviewed of past researcher findings and potential methods to developing a text generation technique.

Table 1. Summary of Recent Researcher Works of Potentials Text Generation Techniques

Authors	Objectives	Techniques	Future works
(Cho et al.,2014)	Phrase Representations Statistical Machine Translation	RNN Encoder and Decoder	Improve performance of RNN Encoder Decoder with Neural Net Language model
(Bauer, Hoedoro and Shuk, 2015)	Produce text in various natural languages	Rule-based text generation Automated Text Markup Language(ATML3)	-
(Gero, Karamanolakisand Chilton, 2018)	Generate text with literary style	Transfer Learning AWD-LSTM utilize pre-trained weights	Optimal training embedding to a real-time interface
(Guo et al.,2018)	Improving LeakGAN. Generating long text	LeakGAN	Apply adoption of LeakGAN to NLP application
(Kratzert et al.,2018)	Explore the potential of the LSTM Learn long term dependencies	LSTM Calibration	Interpret LSTM network
(Devlin et al.,2019)	Generate deep bidirectional representations from unannotated text	BERT	-
(Tighe et al.,2020)	Quantify topics related to pain.	Word2Vec	Examine the connections betweentopics related to pain
(Brown et al.,2020).	Few-shot learners Building GPT-3	GPT-2 Training Fine tune GPT-2	Precisely proceed with few-shotlearning works Improving pre-training sample
(Raffel et al.,2020)	Discover Transfer Learning techniques Convert text-based language difficulties into text-to-text format	T5	Utilize for model training Well-organized knowledge extraction
(Kumar, Gangadharaiahand Roth,2022)	Text generation by Transfer Learning	Transfer Learning Combination of BERT Encoder and GPT-2Decoder Training Adapters	-
(Strobelt et al., 2022)	Collaborative development of data backed text generation	GenNI	Develop common encoding that can shared with wide range of NLP models
(Upadhyay, Sudhakar and Maheswaran, 2022)	Text Generation by using Reinforcement Learning (RL)	Dense Reinforcement Learning (DRL)DRL as decoder Transformer RL training	-
(Zhang et al.,2022)	Contribute Open Source of Pre-trained Transformers to researcher	OPT	Improvements on writing prompt
(Google, 2023)	To develop a more powerful language model that can be used to solve a wider range of problems	Transformer based LLM Masked Language Modelling (MLM)	Scaling models Improving datasets and refining architectures

3. Research Methodology and Data Preparation

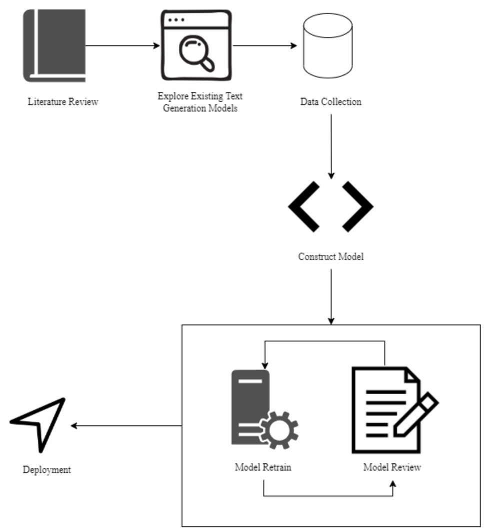

Fig. 4: Flowchart of Research Project

Figure 4 shows that literature review was conducted in this research initially. Based on the past researchers' findings, techniques and unresolved problems were studied and identified. This process helps to justify and rationalize the research to be undertaken and potential contribution that can be obtained at the end of the work. Based on literature review, we are able to shortlist several potential text generation models to be studied in order to consider appropriate algorithms for the purpose of F&B marketing text generation. As part of the research activities, data collection is carried out to gather text data from several F&B Facebook sites. Due to limited resources and time, this research only focuses on several locally popular F&B retailers in Malaysia such as KFC, Starbucks and Tealive.

	message		message
0	calling all deaf hard of hearing talents starb...	0	call all deaf hard of hear talent starbucks ma...
1	the flowers are blooming and the weather is wa...	1	the flower be bloom and the weather be warm up...
2	spring is finally here now its the moment to b...	2	spring be finally here now it the moment to be...
3	let the countdown begin cherry blossom season ...	3	let the countdown begin cherry blossom season ...
4	what comes in rich nutty comfort with every si...	4	what come in rich nutty comfort with every sip...
...	...	...	...
35575	penang rojak anyone	35575	penang rojak anyone
35576	need a drink after your shopping spree youll b...	35576	need a drink after your shopping spree youll b...
35577	hungry	35577	hungry
35578	what makes a perfect sunday\nwell our kerabu m...	35578	what make a perfect sunday well our kerabu man...
35579	its time for a break after all the shopping\nc...	35579	it time for a break after all the shopping com...

35580 rows × 1 columns | 35580 rows × 1 columns

Fig. 5: Collected Unnormalized Dataset Compare to Normalized (Lemmatized) Dataset

Figure 5 shows the sample output of the normalized and unnormalized. The purpose of preparing three different normalization types of data (unnormalized clean dataset, stemmed dataset, lemmatized dataset) is to investigate different type of data on different models and the results obtained on their realistic and practical levels. This is then followed by model building, testing and validation using collected sample data. Since we focus at the Malaysian F&B industries, advertising texts compiled are made of English combine with some Bahasa Malaysia and other slangs. After these data were collected and pre- processed (Refer to Section 7.1), there are a total 35580 rows of marketing text messages overall that will be used as sample for the experiments.

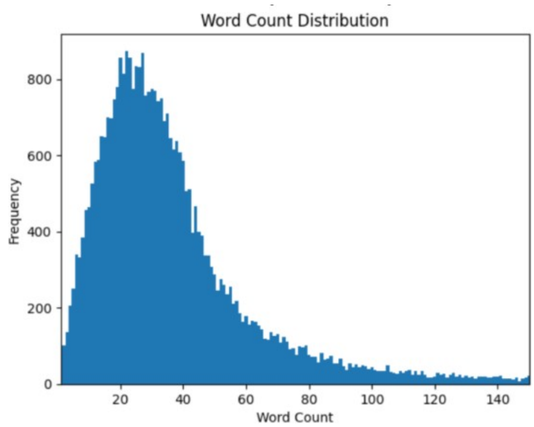

Fig. 6: Word Count Distribution of Dataset Collected

In Figure 6, it shows the word count from each data row as a distribution graph. As observed, the minimum word from the dataset would be 1 word, and the maximum of would be more than 150 words. Each line of text in the dataset contains the most frequent words, with a range of 10 to 50 occurrences. It is important to identify these frequently occurring word count to selectively choose the frequent text pattern for the model to train. Following the completion of data collection, the processed data will be loaded in batches—startingwith a small sample of 5,000 rows and increasing to the entire batch of data fit into the training model. This is to monitor performance and avoid overloading the system. Model construction, testing, and validation using the sample data gathered come next.

When these models are being trained, we will evaluate the experiments' outcomes by the project team of four members to classify outputs are 'good', 'moderate' or 'poor' based on majority votesfrom team member and fine tune based on the normalization of data such as normal data, stemmed data and lemmatized data. The team member consists of 4 ML and NLP related field expert to evaluate the generated text based on context and grammatical criteria. The number with different epochs used for training are also applied when training these models to generate more accurate text outcome. The best model from each proposed model will be chosen for analysis and deployment when the comparison is completed.

4. AI Generative Models and Text Generating Web App

The following few subsections discuss the pseudocode of the three Deep Learning models to be implemented including OPT, LSTM and K2T.

4.1. The design of Text Generating Web Application

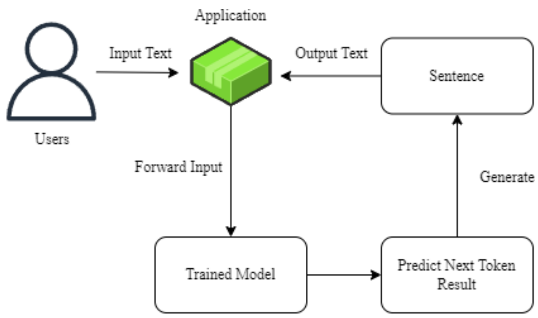

Fig. 7: Flowchart of Text Generating Application

For testing and validation of the developed algorithms, we deploying using Flask and Figure 7 has show the conceptual flowchart of the web application. In following Section 4.2, 4.3 and 4.4, pseudocode for the three (3) chosen generative AI algorithm: LSTM, OPT and K2T, are presented.

4.2. Long Short-Term Memory

The equation of LSTM as follows:

$$f_t = \sigma(W_f[x_t; z; c] + U_f h_{t-1} + b_f)$$
$$i_t = \sigma(W_i[x_t; z; c] + U_i h_{t-1} + b_i)$$
$$o_t = \sigma(W_o[x_t; z; c] + U_o h_{t-1} + b_o)$$
$$c_t = f_t \otimes c_{t-1} + i_t \otimes \sigma(W_c[x_t; z; c] + U_c h_{t-1} + b_c) h_t = o_t \otimes relu(c_t)$$

Where,

f_t: Forget Gate, i_t: Input Gate, o_t: Output Gate, c_t: Current Cell State, h_t: Current Hidden State, σ: Sigmoid Function, W: Weight Matrix, x_t: Input Sequence, U: Weight Associated, h_{t-1}: Hidden State of Previous Timestamp, b: Bias, $\otimes$: Pointwise Multiplication Operation

The equations presented above are applicable only for a single time step, implying that they need to be recalculated for each subsequent time step. Therefore, if there are 10-time steps in a sequence, the equations will have to be computed 10 times for each of those time steps. This indicates that the weight matrices remain constant across multiple time steps. In simpler terms,the same weight matrices are employed to compute the outputs for various time steps. Figure 8 demonstrate the idea using LSTM to generate the sentence is using word embedding feature while learning to predicting the next possible word.

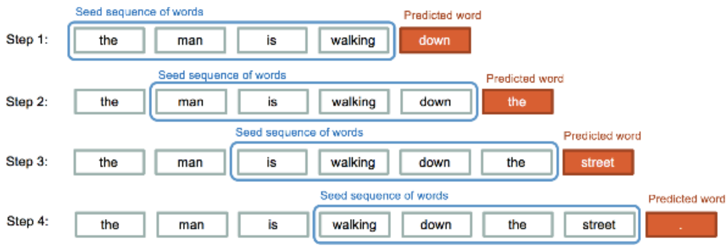

Fig. 8: Text generation concept using LSTM

Proposed LSTM Implement Algorithm

Pre-process with word count distribution
FOR the words in each row
DO split where >=10 or <=50 append split result
Function tokenizing each of the words
Fit tokenizer on input dataframe
Calculate unique words in tokenizer
Convert data into tokens
Generate n-gram sequence and add to input sequence list
Return input sequence and total words
Create function generate padded sequence
Determine the maximum length of input sequence
Pad the same input sequences as the maximum sequence length
Split input sequence to predictors and labels
Convert labels to one-hot encoded
Return predictors, label and max sequence
Generate padded sequence from input sequence
Create Keras model
Add Embedding feature
Add LSTM layer
Add Dropout layer
Add Output layer
Create model with max sequence and total words
Training
epochs = 100
verbose = 1
Save model
Deployment
END

The reason considers to implementing LSTM into advertisement text generation is because it provides the capability to effectively memorize large amounts of previous data. Phrases in text are sequences of words, and LSTM can be utilized to predict the next word by taking a sequence of words as input and generating a probability matrix for each word in the dictionary to be the next word in the sequence. The model also learns the similarity between words or characters and calculates the probability of each.

4.3. Open Pretrained Transformers

Proposed OPT Implemented Algorithm

Load Transformers OPT-350M tokenizer and pretrained model
Train Test Split

```
        train ratio = 0.7
        test ratio = 0.3
Load to HuggingFace Datasets
Tokenize
Create function group text and mapping tokens with fixed blocks of length
        block_size = 128
        concate input to list
        concate and round to nearest multiple of blocksplit
        concate into chunks of block
        copy token id to the dictionary by label
        return dictionary
Training
        verbose = 20
        batch_size = 8
        Save Model
Deployment
END
```

OPT is a LLM model trained from a large corpus of text data, that is also ready to be fine-tuned into a specific task. It provides the ability to perform text generation specifically in the F&B field with marketing advertisement.

4.4. Keyword to Text Generation

Proposed K2T Implemented Algorithm

```
Use Keybert to extract keywords
Append keyword column
Train test split
        train ratio = 0.7
        test ratio = 0.3
Create T5 model
Training
        verbose = 20
        batch_size = 2
Save Model
Deployment
END
```

The strengths of KeytoText are it has the ability perform keywords to sentence generation, it is different compared to previous model in a form of sentence making. Also, it can identify the input keywords and predict the idea and the sentence from the user. Therefore, it might provide more accurate and ideal result since user can control the variables by using keywords.

5. Implementations of Models, Results and Analysis

5.1. Data Preprocessing

At this stage, stop words were reserved, this is because generating text sentence should include stop words otherwise the sentence would be strange, another fact is we wanted to ensure model able to capture and learn the pattern of text generation. Others include symbols, emoji, hashtags, email, Facebook Tags and Uniform Resource Location (URL) were removed. The reason that digits removed was because the Pandas data frame previously detected non-string values while pre-processing the data that contain digits. Furthermore, if the dataset contains digits, thealgorithm cannot determine the contextual subjectivity and objectivity of the digits play in the text which may affect the usefulness of the generated results.

This is followed with blank spaces, null rows and escape characters removal, followed by performing normalization on text to ensure the training process is able to accept string or bytes like object. The purpose for generating different normalization types of data (Refer to Section 3), by testing various normalization types of data, it would be easier to determine which types of models would be most appropriate to utilize moving forward based on the results.

5.2. Long Short-Term Memory

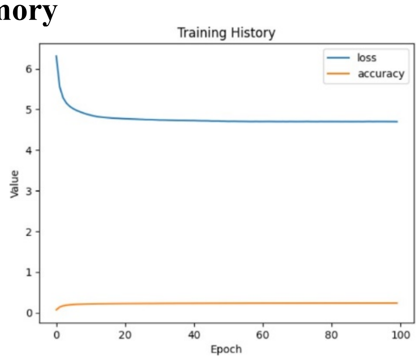

Fig. 9: Training Loss and Accuracy of Full LSTM Training

Figure 9 shows the training loss and accuracy by using Keras "categorical cross entropy" and evaluation metrics "accuracy" after training the LSTM model for 100 epochs. As observed, the training loss is capped at a minimum value of 4.8 and the training accuracy is capped at 0.2. The reason of beencapped might be overfitting after 5 epochs, further adjustment in the parameters such as activation's function, layers of LSTM and accuracy need to be enhanced due to time constraint. The LSTM modelwas tested using several sample inputs. The generated results are shown in Table 4.

Table 2. LSTM Full Trained Model Generated Result

Input Words as seed	Generated Words	Overall Results
Get Extra	Get Extra Free Boost For A Limited Time Only Get Your Favourite	Good
Chicken is	Chicken Is Burger Be A Perfect Combination Of A Soft Shell Crab	Moderate
KFC	Kfc Be Always A Good Way To Enjoy A Free Boost	Moderate
Spicy	Spicy Chicken Burger With Our New Spicy Chicken Burger Be Back	Poor
Grab your	Grab Your Favourite Pizza At Rm With A Minimum Spend Of Rm	Good
Offer only	Offer Only Rm Be A Member Day To Enjoy A Free Boost	Good
Meals offer	Meals Offer Be A Day To Enjoy A Free Boost For You	Moderate

Table 2 shows the result generated for eight different seed input words. Out of seven tests, we can see three 'good', three 'moderate', and two 'poor'. overall results are considered average with a mix of good, moderate, and poor acceptance.

There are certain criteria to review the generated messages. For good results, we can identify the message provides meaningful and contained marketing elements with the least grammatical error. Moderate results show grammatical error and some contexts of phrase are not related and make sense. For poor results, it shows messages that are not meaningful and produced grammatical error such as repeating phrase.

Although LSTM is an advanced RNN model, it does not perform that well when it comes to comparing with other large language model (LLM). There is also an obvious disadvantage on LSTM as it is trained once on the dataset but OPT and K2T are pre-trained models that are fine-tuned base on the collected dataset. This is because LLM such as OPT and K2T are specialize dealing with unsupervised text data and providing effective and stable approach in NLG (Chen et al., 2023). Moreover, LSTM is a traditional model that required more effort to deploy, this is because LSTM are resource consumption in terms of data source, number of parameters and LSTM layers needed to be proceeded with feature tuning during the model training. Also, we do not apply syntactic and semantics rules for LSTM to offer the proper form of generating human likely sentences. Hence, the generated results occasionally not ideal. To review the concept and implementation, we have successfully applied the algorithm discussed above in (4.2) by predicting the next possible word to generate the sentence in sequence.

5.3. Open Pretrained Transformers

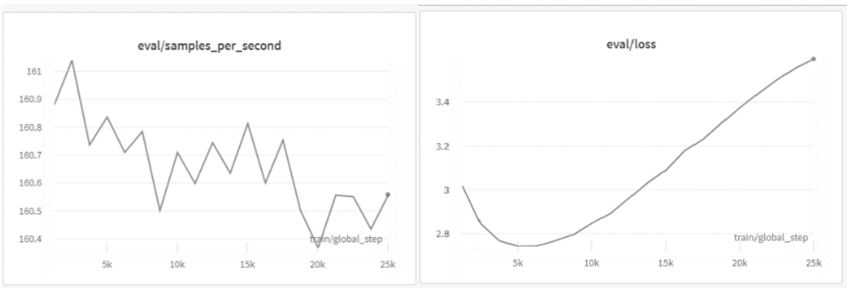

Fig. 10: Evaluation Samples per Second and Evaluation Loss

Global steps refer to the number of batches seen by the graph. Every time a batch is provided, the weights are updated in the direction that minimizes the loss, the global steps is used to track the batches during the training. Figure 10 shows that, as the OPT training global steps increase, the evaluation of samples per secondstarts to decrease, it might be due to this model is overfitting and the trainer does not process much from the new data. Most importantly, the validation loss started to increase after 6000 global steps of training and reached a peak value of 3.5990. It means that the model is performing worse on the validation dataset. This can be an indication that the model is overfitted on the training data, may be due to the dataset istoo small to be learned and generalized by a complex model with such huge number of parameters and causing overfitting.

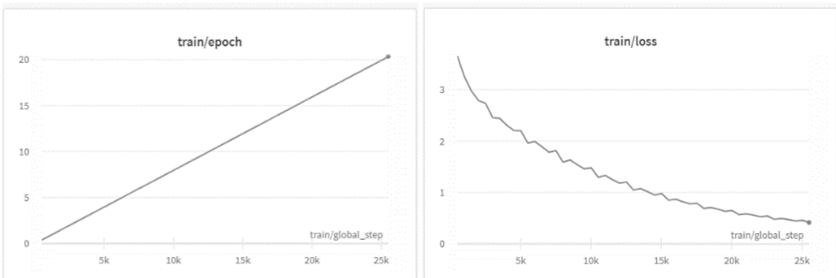

Fig. 11: OPT Epoch with Global Steps and Training Loss

As observe in Figure 11, after 20 epochs of OPT training, the learning rate decreased with a fixed gradient value. The lower the training loss, the better the model is performing on the training data. As the model learns, the training loss should decrease, this justified the learning of OPT model able to generalize new data. Meanwhile, the training loss started to be capped at the minimum value of 0.4616. Another observation is that OPT is case sensitive, meaning that the capitalization of input words will result in different generated outputs.

Table 3. OPT Capital Compare Non-Capital Generated Result

Input Words	Generated Words (Capitalised First Alphabet)	Generated Words (Uncapitalised Text)
Get extra	Get extra savings when you order with grabfood with ourbuy free deal today	get extra satisfaction with our new spicy chicken burger order now viamcdelivery or drivethru
Chicken is	Chicken is the best food to eat when youre feeling hungry	chicken is the best food to share with your loved ones
KFC	KFC breakfast is back with a new look and taste get yourhands on our new breakfast menu today	kfc delivery is now available in peninsular malaysia order now via kfcapp or kfccommy
Grab your	Grab your friend and come over to our store to enjoy ourbuy free promot today	grab your favorite cuppa and head over to your nearest starbucks storeto enjoy this exclusive promotion
Offer only	Offer only valid dinein takeaway mcdelivery drivethru	offer only valid on during local store hours valid photo id must be shown at store to redeem free boost only first name is eligible for this promotion only one redemption per customer
Meals offer	Meals offer a great way to spend time with your famili andfriend enjoy a meal with your	meals offer valid during local store hours valid photo id must be shownat store to redeem free boost only first name is eligible for this promotiononly one redemption per customer

Table 3 shows that OPT is a case-sensitive model where generative outputs are different based on the different inputs. This is because OPT is one of the LLM models. LLM are trained on vast amounts of natural language data, such as text or speech, using deep learning techniques. These models are typically based on neural networks, which allow them to learn and understand the patterns and structures of language in a way that mimics human language processing. LLM have been trained on massive amounts of text data, they also capable of generating human- like text, answering questions, and completing tasks that involve natural language processing. Some examples of large language models include OpenAI's GPT-3, Google's BERT, and Facebook's RoBERTa.

Although there is no formulae or algorithm provided by the developers of OPT. However, there is research indicates OPT has the ability to generate text and respond in a manner that closely resembles human-produced content (Tan et al., 2023). In review, we have successfully fine-tuned the OPT-350 million parameters model into a F&B marketing text generation model.

5.4. Keyword to Text Generation)

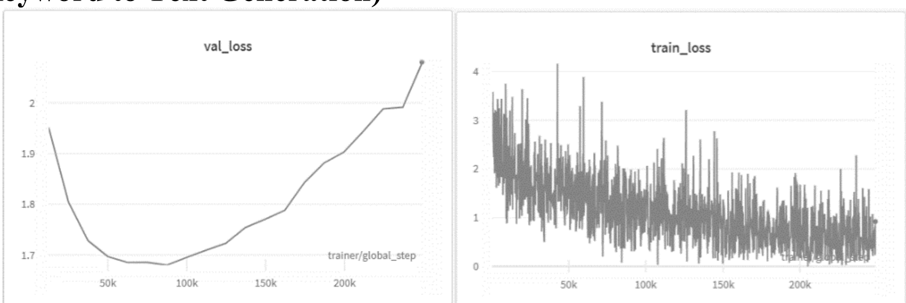

Fig. 12: KeytoText Validation Loss and Training Loss

Figure 12 shows that when the K2T model's global steps reached 90k, the validation loss reached the lowest value which means it is a good fit. However, the overfitting will occur after 90k of global steps. In the K2T model building, with 20 epochs, the global steps will reach 249k; beyond 249k, the validation loss started to overfit and reach a maximum validation loss value of 2.081. Besides that, the training loss starts at 3.432, after 20 epochs of training and 249k global steps, it reaches the minimum value of 0.932, this indicates the model is learning from the data. When the training loss is decreasing, it means that the model fitness has slowly been obtained. This is accomplished as the model's parameters were adjusted during the training stage. Well-fitted model delivers predictions that are able to obtain high predicted outcomes with the training loss decreases.

Table 4. KeytoText Generated Result Based on N-gram

Input Wordsas Seed	Generated Words (Input 1-gram)	Generated Words(Input 2-gram)	Generated Words(Input 3-gram)	Generated Words(Input 4-gram)
chicken, offer, delicious ,low price	the chicken riceshop	when you have chickenwe offer you	our chicken be back offer you delicioussside and a whole lot more	delicious at a low price we have something for everyone from our chicken to offer youonly the
noodle, tasty, hot, limited	noodle	noodle and everythingtasty	noodle it tasty and hot	noodle or tasty even hot and limited timeonly noodle available in all outlet today
KFC, celebrate, perfect, sides	KFC	at KFC and we celebrate	what we do at the KFC and we celebrate it with our perfect chicken patty	we get it right here at the KFC we celebratewith our perfect two sides a they
drinks, deliver, discount, mango	get your drinksnow	get your drinks deliver toyou	get your drinks deliver to you withdiscount from rm for select drink only |	discount for one of our mango drinks get itdeliver to you for free psst
coffee, best,relax, aroma	cant get enoughof coffee	coffee be the bestmedicine	coffee be always the best thing to do sosit back and relax with u	coffee be always the best thing to relax withand the aroma of your choice add it to your

Table 4 shows the K2T generation based on different number of gram of keywords affect the generated results. KeytoText used Google T5 model to carry out keyword text generation. Unfortunately, the generated results are not ideal due to it only produces human likely result by input 3-grams. Although the keyword input and generated text, from the model cannot learn the pattern of the text advertisement generation. Compared to OPT, KeytoText rarely produce human likely text message. In addition, with a different gram of words, the KeytoText model also generates worse results when using 1-gram and 2-gram as observed. This is because Google T5 is a text-to-text LLM, the more of the input keywords, the model able to catch the idea of generating more accurate and a more human likely sentence.

5.5. Summary of Models Performance

Table 5. Deployed Model Specifications

Methods	Epochs	Dataset Used	Training Time(Hours. Minutes)
Open Pretrained Transformers	20	Clean dataset	1.36

(OPT)			
Long Short-Term Memory (LSTM)	100	Lemmatised dataset	7.22
KeytoText (K2T)	20	Lemmatised dataset	13.33

Table 5 specifies the epochs and datasets used for training the three models. It is observed that OPT is best suited for training using an unnormalized clean dataset (as referred to in Section 3). This is because OPT has the ability to learn the patterns of Part of Speech (POS). For LSTM and KeytoText models, a lemmatized dataset is applied to generate normalized results for the experiment. Nevertheless, we do not proceed to use stemmed data because the stemmed result is not ideal for current experimentations.

6. Discussion and Conclusion

6.1. Discussion of Deployed Model

Table 6. Evaluation of Proposed Text Generation Model

Methods	Advantages	Disadvantages
Open Pretrained Transformers(OPT)	Produce accurate and relevant text advertisement. Able to identify where to sentence should stop. Sometimes produce attractive results.	Unable to use keywords to generate text.
Long Short-Term Memory(LSTM)	Able to let the user input generated number of words.	Not smart enough to identify where the sentence should end properly with input number. Produce poor results, rarely generating accurate phrases and relevant marketing messages.
KeytoText (K2T)	Generate text by keywords instead of traditional sentence generation. Sometimes generate attractive results	The generated results depend on the amount of input keywords. The fewer the keywords, the worse the generated output will be.

Table 6 stated the advantages and disadvantages of OPT, LSTM and KeytoText validated by human NLP experts in terms of sentence accuracy, flexibility of use and marketing effectivity. In contrast to these three successfully deployed text generation models, the most suitable to adopt in the field of industry is the OPT, which produces quite relevant and accurate marketing messages as shown. It also rarely occurs generating invalid types of phrases or sentences. The developer of OPT Meta team also stated that belongs to the same family of decoder-only models like GPT-3. As such, it was pretrained using the self-supervised causal language modelling objective. These models are typically trained using large amounts of text data, which enables them to learn patterns and relationships between words that can be used to make accurate predictions.

LSTM proceeded with enhancement such as word count distribution and increasing numbers of data training. Nevertheless, LSTM is the worst to use compared to the other LLM models. LSTM as a traditional model would lose much of the priority in terms of training parameters adjustment, training resource consumption and generated accuracy. The reason LSTM is incapable of generating high accuracy and human-like sentences because it required a lot of adjustment, especially rules for the feature to control the next word predicted.

Unfortunately, there is no research paper exists for the KeytoText library, as it was developed by a programmer named Gagan and has not gained widespread popularity. Nonetheless, the results of the experiment demonstrate that it is possible to use keywords to generate text. This approach is particularly powerful and easy to use, as it allows users to input their desired idea by keyword, then the generator can produce corresponding advertisements based on those input keywords generate sentence which are more intelligent and smarter in terms of understanding what the user wants.

6.2. Discussion and Future Improvements of Research

There are improvements can be made in the future. The data selection process was restricted to certain fields within the F&B industry, specifically focusing on fast food companies and those with retail shops in Malaysia. Consequently, while the data used may be representative of popular food items, it is possible that relevant food data beyond this scope is not included in the analysis. Meanwhile, the data collection process does not filter based on certain conditions where the collected marketing text messages are suitable for the training model learn to achieve accurate and attractive context. Moreover, the pre-processing of the lemmatized data also can be improved, this is because some ofthe text data are not lemmatized although using POS-tag then lemmatize by NLTK wordnet.

After evaluating those findings, we have found that the data may not enough for training purposes, at the same time we need to consider the computational resource are able to handle the training, therefore the increment of data will put in the future consideration. Additionally, testing the perfect fitting and overfitting results is also one of the future improvements that can be made, as identifying the difference generated results. Besides that, the model enhancementis also one of the considerations because the proposed model not yet carried out deeper experimentation such as manipulating the learning rate, and batch size to discover the difference to produce a better- generated result. A larger parameter such as parameters up to billion from OPT may be used and re-train the same data with perfect fitting to observe the generated results compare to the current applied 350 million parameters. Lastly, the KeytoText model should proceed with word count distribution, this is becausethe generated results contain short sentences between 1 to 4 words only, this might be the reason duringtraining and fail to learn the appropriate pattern of sentences.

6.3. Contributions

The studies contribute to the idea of using NLP and NLG approaches to generate F&B field text advertisement. It also justifies few deployed text generation models including LSTM, KeytoText and OPT proving these models can generate F&B marketing advertisement. Some of the models does help in generating F&B advertisements especially for OPT, it generates quite accurate results for promoting the specialized field of F&B.

However, the well-known and popular text generator Chat-GPT still has yet revealed the techniques used to deal with NLP and NLG because it is confidential for business purposes. Besides that, there are many ways of text generating, transformers are well popular, Keras word embedding, text to text generation by using summarization of keywords. This research would be a milestone in terms of using LSTM and pretrained transformers as LLM to train a model for generating text in a specific field.

6.4. Conclusion

After conducting through research, the objectives set out at the beginning of the study have been successfully achieved. The investigation involved delving into the various approaches to generating text using NLG models. Through this process, a range of effective techniques was discovered and OPT, LSTM, KeytoText are implemented to create several text-based marketing-related advertisements. Another objective also has been accomplished, those deployed models have successfully developedfew text-based marketing-related advertisement applications for testing and validation purposes. We have discussed those deployed models in terms of their implementation, training validation and generated results that are useful for future enhancement.

Despite the success of the study, it is worth that the OPT model can generate advertisements with the least error and higher accuracy rate of marketing text advertisement. Nevertheless, it should be acknowledged that all the deployed model occasionally produced structures that were not entirely correct or absolute in their form, which could potentially impact the reliability of the generated results. In conclusion, the research conducted provides valuable insights into the use of NLP models for generating text-based F&B marketing materials. The findings could be used to inform the development of future models and applications, ultimately leading to more effective and accurate text-generation techniques.

Acknowledgements

We would like to extend our heartfelt gratitude to Tunku Abdul Rahman University of Management and Technology (TARUMT) providing the computational resources for the training purposes and supporting the academic research with professional advisor. In addition, we would like to appreciate Centre for Business Incubation and Entrepreneurial Ventures (CBIEV) for inspiring the research idea and playing a significant role in the development of Natural Language Generation in the field of marketing advertising. Their contributions have been instrumental in making this research a milestone achievement

References

Bauer, A., Hoedoro, N. and Schneider, A. (2015). *Rule-based approach to text generation in natural language - Automated Text Markup Language (ATML3)*. [online] Available at: https://ceur-ws.org/Vol-1417/paper20.pdf Accessed 14 May 2023

Beltis, A.J. (2021). *How to Make an Ad: A 10-Step Guide*. [online] blog.hubspot.com. Available at: https://blog.hubspot.com/marketing/how-to-make-an-ad. Accessed on Jan 2, 2023

Bengio, Y., Simard, P., and Frasconi, P. (1994). "Learning long-term dependencies with gradient descent is difficult," in IEEE Transactions on Neural Networks, vol. 5, no. 2, pp. 157-166 doi: 10.1109/72.279181 Retrieved from https://ieeexplore.ieee.org/document/279181 . Accessed on Jan 5, 2023

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. Accessed on Jan 6, 2023

Cameron R.Wolfe. (2022). Understanding the Open Pre-Trained Transformers (OPT) Library. Retrieved from https://towardsdatascience.com/understanding-the-open-pre-trained-transformers-opt-library-193a29c14a15 Accessed on Apr 21, 2023

Chatzoglou, P. and Chatzoudes, D. (2016). Factors affecting e-business adoption in SMEs: an empirical research. *Journal of Enterprise Information Management*, 29(3), pp.327–358. doi:https://doi.org/10.1108/jeim-03-2014-0033.

Chen, Y., Wang, R., Jiang, H., Shi, S. and Xu, R. (2023). *Exploring the Use of Large Language Models for Reference-Free Text Quality Evaluation: A Preliminary Empirical Study*. [online] Available at: https://arxiv.org/pdf/2304.00723.pdf Accessed on 13 May ,2023.

Chen, Y.-C., Gan, Z., Cheng, Y., Liu, J. and Liu, J. (2020). *Distilling Knowledge Learned in BERT for Text Generation*. [online] Association for Computational Linguistics, pp.7893–7905. Available at: https://aclanthology.org/2020.acl-main.705.pdf Accessed on 14 May 2023.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bougares, F., Schwenk, H. and Bengio, Y. (2014). *Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation*. [online] Available at: https://arxiv.org/pdf/1406.1078.pdf.

Davies, T. (2022). *Meta AI Releases OPT-175B, Set Of Free-To-Use Pretrained Language Models*. [online] W&B. Available at: https://wandb.ai/telidavies/ml-news/reports/Meta-AI-Releases-OPT-175B-Set-Of-Free-To-Use-Pretrained-Language-Models--VmlldzoxOTQwOTU1 Accessed on Jan 6, 2023

Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. [online] Available at: https://arxiv.org/pdf/1810.04805.pdf. Accessed on May 11, 2023

Fatima, N., Imran, A.S., Kastrati, Z., Daudpota, S.M. and Soomro, A. (2022). A Systematic Literature Review on Text Generation Using Deep Neural Network Models. *IEEE Access*, 10(2169-3536), pp.53490–53503. doi:https://doi.org/10.1109/access.2022.3174108. Accessed on 11 May, 2023

Fucci, D., Romano, S., Baldassarre, M., Caivano, D., Scanniello, G., Thuran, B. and Juristo, N. (2022). A Longitudinal Cohort Study on the Retainment of Test-Driven Development. *Deep Learning Based*

Text Classification: A Comprehensive Review. [online] https://arxiv.org/pdf/2004.03705.pdf Accessed on 14 May, 2023.

Gagan. (2021). keytotext. Retrieved from https://github.com/gagan3012/keytotext Accessed on Apr 1, 2023

Gero, K.I., Karamanolakis, G. and Chilton, L. (2018). *Transfer Learning for Style-Specific Text Generation*. [online] Available at: https://nips2018creativity.github.io/doc/Transfer%20Learning%20for%20Style-Specific%20Text%20Generation.pdf Accessed on 14 May, 2023.

Ghasemi, H. and Hashemian, M. (2016). A Comparative Study of Google Translate Translations: An Error Analysis of English-to-Persian and Persian-to-English Translations. *English Language Teaching*, 9(3), p.13. doi:https://doi.org/10.5539/elt.v9n3p13.

Ghatasheh, N., Faris, H., AlTaharwa, I., Harb, Y. and Harb, A. (2020). Business Analytics in Telemarketing: Cost-Sensitive Analysis of Bank Campaigns Using Artificial Neural Networks. *Applied Sciences*, 10(7), p.2581. doi:https://doi.org/10.3390/app10072581.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2014). *Generative Adversarial Nets*. [online] Available at: https://arxiv.org/pdf/1406.2661.pdf.

Google (2023). *PaLM 2 Technical Report*. [online] Available at: https://ai.google/static/documents/palm2techreport.pdf [Accessed 18 May 2023].

Graves, A. (2013). Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850. Accessed on Jan 4, 2023

Groot, O.Q., Ogink, P.T., Oosterhoff, J.H. and Beam, A.L. (2021). Natural language processing and its role in spine surgery: A narrative review of potentials and challenges. *Seminars in Spine Surgery*, 33(2), p.100877. doi:https://doi.org/10.1016/j.semss.2021.100877.

Guo, J., Lu, S., Cai, H., Zhang, W., Yu, Y. and Wang, J. (2018). Long Text Generation via Adversarial

Training with Leaked Information. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1). doi:https://doi.org/10.1609/aaai.v32i1.11957. Accessed on Jan 4, 2023

Hill, M.E. (2010). *Marketing strategy in play : questioning to create difference*. New York, N.Y. (222 East 46Th Street, New York, Ny 10017): Business Expert Press.

Hochreiter, Sepp & Schmidhuber, Jürgen. (1997). Long Short-term Memory. Neural computation. 9. 1735-80. 10.1162/neco.1997.9.8.1735. Accessed on Apr 21, 2023

https://doi.org/10.1016/j.semss.2021.100877 Accessed on Jan 2, 2023

HuggingFace. (2023). facebook/opt-350m. Retrieved from https://huggingface.co/facebook/opt-350m Accessed on Apr 21, 2023

Huszár, F. (2015). How (not) to Train your Generative Model: Scheduled Sampling, Likelihood, Adversary? 1–9. http://arxiv.org/abs/1511.05101 Accessed on Jan 4, 2023

Khan, S.O., Hasan, R., Hussain, S. and Malik, M.H. (2023). Inventory Management Optimization with Data Analytics for a Trading Company. [online] Available at: https://www.researchgate.net/publication/368454977_Inventory_Management_Optimization_with_Data_Analytics_for_a_Trading_Company Accessed on 13 May, 2023.

Kharazi, D. (n.d.). *https://dkharazi.github.io/notes/ml/nlp/t5/*. [online] dkharazi.github.io. Available at: https://dkharazi.github.io/notes/ml/nlp/t5 Accessed 14 May 2023.

Kiselicki, M. (2017). Business process improvement of the system integrator companies through a case study analysis. p.7(17):149.

Kolhar, M. and Alameen, A. (2021). Artificial Intelligence Based Language Translation Platform. *Intelligent Automation & Soft Computing*, 28(1), pp.1–9. doi:https://doi.org/10.32604/iasc.2021.014995.

Kotler, P. & Armstrong, G. (2021). Principles of Marketing, 18th Edition. Accessed on Jan 2, 2023

Kratzert, F., Klotz, D., Brenner, C., Schulz, K. and Herrnegger, M. (2018). Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences*, 22(11), pp.6005–6022. doi:https://doi.org/10.5194/hess-22-6005-2018

Kumar, V., Gangadharaiah, R. and Roth, D. (2022). *Unsupervised Neural Stylistic Text Generation using Transfer learning and Adapters*. [online] Available at: https://arxiv.org/pdf/2210.03264.pdf Accessed 14 May 2023.

Leary, C.O. (2023). Using E-Business Information in Database Marketing Processes. [online] Available at: https://www.researchgate.net/publication/255617622_USING_E-BUSINESS_INFORMATION_IN_DATABASE_MARKETING_PROCESSES Accessed on 13 May 2023.

Lipton, Z.C. (2015). A Critical Review of Recurrent Neural Networks for Sequence Learning. [online] Available at:https://www.researchgate.net/publication/277603865_A_Critical_Review_of_Recurrent_Neural_Networks_for_Sequence_Learning Accessed on 14 May, 2023.

Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., Wu, Z., Zhu, D.,Li, X., Qiang, N., Shen, D., Liu, T. and Ge, B. (2023). *Summary of ChatGPT/GPT-4 Research and Perspective Towards the Future of Large Language Models*. [online] Available at: https://arxiv.org/pdf/2304.01852.pdf Accessed 14 May 2023

Liu, Z., Lin, Y. and Sun, M. (2020). *Representation Learning for Natural Language Processing*.

Singapore: Springer Singapore. doi:https://doi.org/10.1007/978-981-15-5573-2.

Mergent (2022). Advertising & Marketing Services - Quarterly - ProQuest. [online] Available at: https://www.proquest.com/docview/2723081134?accountid=38945 Accessed on Jan 1, 2023

Nerdynav (2022). *73 Important ChatGPT Statistics & Facts For Mid Feb 2023 + An Infographic - Nerdy Nav*. [online] Nerdynav. Available at: https://nerdynav.com/chatgpt-statistics/ Accessed 14 May, 2023.

Norambuena, B.K., Lettura, E.F. and Villegas, C.M. (2019). Sentiment analysis and opinion mining applied to scientific paper reviews. *Intelligent Data Analysis*, 23(1), pp.191–214. doi:https://doi.org/10.3233/ida-173807.

OpenAI (2023a). *GPT-4*. [online] openai.com. Available at: https://openai.com/research/gpt-4 Accessed 14 May, 2023.

OpenAI (2023b). *GPT-4 Technical Report*. [online] Available at: https://cdn.openai.com/papers/gpt-4.pdf. Accessed 14 May, 2023.

Petrevski, V., Josimovski, S. & Kiselicki, M. 2017, "Business Process Improvement of The System Integrator Companies Through A Case Study Aanalysis", *Journal of Sustainable Development (Skopje),* vol. 7, no. 17, pp. 149-174.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W. and Liu, P. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research*, [online] 21, pp.1–67. Available at: https://jmlr.org/papers/volume21/20-074/20-074.pdf. Accessed on Apr 21, 2023

Roy, J. (2022). *Everything you need to know about Google BERT*. [online] seositecheckup.com. Available at: https://seositecheckup.com/articles/everything-you-need-to-know-about-google-bert. Accessed on 11 May 2023

Ryan, T. (2020). LSTMs Explained: A Complete, Technically Accurate, Conceptual Guide with Keras. Retrieved from https://medium.com/analytics-vidhya/lstms-explained-a-complete-technically-accurate-conceptual-guide-with-keras-2a650327e8f2. Accessed on Jan 6, 2023

Sama, R. (2019). Impact of Media Advertisements on Consumer Behaviour. *Journal of Creative Communications*, [online] 14(1), pp.54–68. Available at: https://journals.sagepub.com/doi/10.1177/0973258618822624.

Strobelt, H., Kinley, J., Krueger, R., Beyer, J., Pfister, H. and Rush, A.M. (2022). GenNI: Human-AI Collaboration for Data-Backed Text Generation. *IEEE Transactions on Visualization and Computer Graphics*, 28(1), pp.1106–1116. doi:https://doi.org/10.1109/tvcg.2021.3114845.

Tan, C.W., Khor, Y.K., Tan, J.H. and Tan, G.J. (2023). *Meta AI's Open Pretrained Transformer (OPT): The Future of Text Generation?* [online] Research Gate. Available at: https://www.researchgate.net/publication/370582714_Meta_AI's_Open_Pretrained_Transformer_OPT_The_Future_of_Text_Generation Accessed on 13 May, 2023.

Tékouabou, S.C.K., Gherghina, Ş.C., Toulni, H., Neves Mata, P., Mata, M.N. and Martins, J.M. (2022). A Machine Learning Framework towards Bank Telemarketing Prediction. *Journal of Risk and Financial Management*, 15(6), p.269. doi:https://doi.org/10.3390/jrfm15060269.

Tighe, P.J., Sannapaneni, B., Fillingim, R.B., Doyle, C., Kent, M., Shickel, B. and Rashidi, P. (2020). Forty-two Million Ways to Describe Pain: Topic Modeling of 200,000 PubMed Pain-Related Abstracts Using Natural Language Processing and Deep Learning–Based Text Generation. Pain Medicine, 21(11), pp.3133–3160. doi:https://doi.org/10.1093/pm/pnaa061. Accessed on May 11, 2023

Wang, F. & Vaughan, L. 2014, "Firm web visibility and its business value", *Internet research,* vol. 24, no. 3, pp. 292-312.

Yue, X., Zhou, T., He, L. and Li, Y. (2022). Research on Long Text Classification Model Based on Multi-Feature Weighted Fusion. *Applied Sciences*, 12(13), p.6556. doi:https://doi.org/10.3390/app12136556.

Zhang, S., Roller, S., Goyal, N., Artetxe, M., Chen, M., Chen, S., Dewan, C., Diab, M., Li, X., Lin, X.V., Mihaylov, T., Ott, M., Shleifer, S., Shuster, K., Simig, D., Koura, P.S., Sridhar, A., Wang, T. and Zettlemoyer, L. (2022). OPT: Open Pre-trained Transformer Language Models. *arXiv:2205.01068 [cs]*. [online] Available at: https://arxiv.org/abs/2205.01068.

Zhou, Z.G. (2022). Research on Sentiment Analysis Model of Short Text Based on Deep Learning.*Scientific Programming*, 2022, pp.1–7. doi:https://doi.org/10.1155/2022/2681533.