Sentimental Symphony: Maximizing GPT Capabilities for Text Generation and Emotional Insight

Thesis submitted in partial fulfillment of the requirements for the degree of

Master of Engineering
in
Computer Science and Engineering

Submitted by

Soumik Maity
Regn. No. - 163561 of 2022 - 2023
Exam Roll No. - M4CSE24007
Class Roll No. - 002210502010
Session - 2022 - 2024

Under the supervision of

Prof. Diganta Saha

Department of Computer Science and Engineering
Jadavpur University
188, Raja S.C. Mallick Rd,
Kolkata - 700032, West Bengal, India

Department of Computer Science and Engineering
Faculty of Engineering and Technology
Jadavpur University, Kolkata - 700032

CERTIFICATE OF RECOMMENDATION

This is to certify that the work embodied in this thesis entitled **"Sentimental Symphony: Maximizing GPT Capabilities for Text Generation and Emotional Insight"** has been satisfactorily completed by **Soumik Maity** (Registration Number 163561 of 2022 - 2023; Class Roll No. 002210502010; Examination Roll No. M4CSE24007). It is a bonafide piece of work carried out under my supervision and guidance at Jadavpur University, Kolkata, for partial fulfilment of the requirements for the awarding of the **Master of Engineering** in **Computer Science and Engineering** degree of the Department of Computer Science and Engineering, Faculty of Engineering and Technology, Jadavpur University, during the academic year 2022 - 2024.

Prof. Diganta Saha,
Department of Computer Science and Engineering,
Jadavpur University.
(Supervisor)

Forwarded By:

Prof. Nandini Mukherjee,
HOD, Department of Computer Science and Engineering,
Jadavpur University.

Prof. Dipak Laha,
DEAN, Faculty of Engineering and Technology,
Jadavpur University.

Department of Computer Science and Engineering
Faculty of Engineering and Technology
Jadavpur University, Kolkata - 700032

CERTIFICATE OF APPROVAL

This is to certify that the thesis entitled **"Sentimental Symphony: Maximizing GPT Capabilities for Text Generation and Emotional Insigh"** is a bonafide record of work carried out by **Soumik Maity** (Registration Number 163561 of 2022 - 2023; Class Roll No. 002210502010; Examination Roll No. M4CSE24007) in partial fulfilment of the requirements for the award of the degree of **Master of Engineering** in **Computer Science and Engineering** in the Department of Computer Science and Engineering, during the period of September 2022 to May 2024. It is understood that by this approval, the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein, but approve the thesis only for the purpose of which it has been submitted.

Signature of Examiner 1
Date:

Signature of Examiner 2
Date:

Department of Computer Science and Engineering
Faculty of Engineering and Technology
Jadavpur University, Kolkata - 700032

DECLARATION OF ORIGINALITY AND COMPLIANCE OF ACADEMIC ETHICS

I hereby declaree that the thesis entitled **"Sentimental Symphony: Maximizing GPT Capabilities for Text Generation and Emotional Insight"** contains literature survey and original research work by the undersigned candidate, as a part of his degree of **Master of Engineering** in **Computer Science and Engineering, Jadavpur University**. All the information have been obtained and presented in accordance with academic rules and ethical conduct.
I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name: Soumik Maity

Examination Roll No.: M4CSE24007

Registration No.: 163561 of 2022 - 2023

Thesis Title: Sentimental Symphony: Maximizing GPT Capabilities for Text Generation and Emotional Insight

Signature of the Candidate:

Acknowledgements

I am deeply grateful for the journey that has led to the completion of this thesis, and I wish to express my heartfelt thanks to all those who have supported me along the way.

First and foremost, my profound gratitude goes to my esteemed supervisor, Prof. Diganta Saha. His exceptional guidance, insightful feedback, and unwavering support have been the cornerstone of this research. His passion for Natural Language Processing and dedication to academic excellence have been a constant source of inspiration.

I also express my sincere appreciation to the Head of the Department, Prof. Nandini Mukherjee, for her continuous encouragement and leadership. Her dedication to fostering an environment of innovation and research has been invaluable to my work.

My deepest thanks go to the Dean of the Faculty, Prof. Dipak Laha. His steadfast support and commitment to providing the necessary resources and facilities have played a crucial role in the successful completion of this thesis.

I am immensely grateful to the faculty and staff of the Department of Computer Science Engineering at Jadavpur University. Their support and assistance have been instrumental throughout my academic journey. Special thanks to my fellow research colleagues and friends for their camaraderie and collaboration.

I also express my heartfelt appreciation to Jadavpur University. The vibrant academic community, state-of-the-art resources, and nurturing environment provided by this esteemed institution have been integral to my growth as a researcher. I am proud to be a part of such a distinguished university.

Finally, I owe my deepest gratitude to my family for their unwavering love, patience, and encouragement. Their constant support has been my foundation and source of strength throughout this academic pursuit.

Thank you all for being an integral part of this journey and for making this achievement possible.

Soumik Maity

Examination Roll No.: M4CSE24007

Registration No.: 163561 of 2022 - 2023

Department of Computer Science and Engineering,

Jadavpur University.

Abstract

In recent years, Natural Language Processing (NLP) models, particularly Generative Pre-trained Transformers (GPT), have demonstrated remarkable capabilities in generating coherent and contextually relevant text. However, while these models excel in linguistic proficiency, they often fall short in capturing the nuanced emotional undertones present in human communication. This paper introduces "ECGText" (Emotionally Controlled Generation of Text) a novel model designed to enhance text generation by integrating emotional intelligence into GPT-based systems. By leveraging advances in sentiment analysis and emotion recognition, Sentimental Symphony empowers GPT to imbue generated text with appropriate emotional nuances, thereby fostering more engaging and relatable communication. We present a comprehensive overview of the framework, discussing its architecture, training methodology, and evaluation metrics. Additionally, we conduct experiments on diverse text generation tasks, showcasing the efficacy of Sentimental Symphony in producing emotionally resonant outputs across various domains. Our results demonstrate significant improvements in generating text with nuanced sentiment, paving the way for more emotionally intelligent AI systems that better mirror human expression and understanding.

Keywords: Generative AI; Text Generation; Robot with Sentiment; Enhancing GPT; ECGText

Contents

Abstract

List of Tables

List of Figures & Images

List of Abbreviations

1. Introduction 1 - 3

    1.1. Motivation 1

    1.2. Objective 1

    1.3. Novelty 2

    1.4. Contribution 3

2. Literature Review 4 - 9

    - Research Gap 9

3. Methodology 10 - 17

    3.1. Problem Statement 10

    3.2. Working of the Models 11

    3.3. Datasets 13

    3.4. Algorithm 16

4. Result and Analysis 18 - 22

    4.1. Output 18

    4.2. Metrics Evaluation 19

    4.3. Analysis 22

5. Conclusion 23 - 24

    - Future Work 24

List of Tables

1. Average Sentence Length Metrics 19
2. POS Tagging Metrics 19
3. Winnowing Score Metrics 21
4. Final comparison of results 21

List of Figures

1. ECGText Architecture 12

2. WordCloud of Obama Dataset 14

3. WordCloud of Movie Dataset 15

4. WordCloud of Bengali Dataset 15

5. Cosinie Similarity of Movie Dataset 20

6. Cosinie Similarity of Obama Dataset 20

1. Introduction

In recent years, the field of natural language processing (NLP) has witnessed remarkable progress, particularly with the advent of advanced language models such as Generative Pre-trained Transformers (GPT).

These models have demonstrated impressive capabilities in text generation tasks, yet they often lack the ability to capture the subtle emotional nuances inherent in human communication. Emotions play a crucial role in conveying meaning, context, and intent, making them a vital aspect of effective communication.

1.1 Motivation

When we started analyzing various text generation models, we found most of the models pick text from various training sets and rephrase them in a way that it feels customized for the end users. There was not much work available on generating completely new text that the machine would generate based on its Natural Language Understanding. Initially we tried out LSTM and Bi-LSTM. We found they were either copying from the corpus or generating gibberish and repeat words. Various strategies were employed subsequently in order to eliminate these hallucinations. Our motive was not to make the model generate factually accurate generation but to generate long, coherent text that is brand new and not copied from any corpus.

1.2 Objective

In this paper, we propose a novel model for integrating emotional intelligence into text generation models. Our approach leverages the capabilities of GPT-2, a state-of-the-art language model, and the RoBERTa-based GoEmotions model, an advanced emotion recognition system which we call the ECGText modelEmotional Controlled Generation of Text . The primary objective is to enable AI-generated text to convey appropriate emotional nuances, thereby enhancing its relevance and impact. Integrating emotional intelligence into text generation poses several challenges, including algorithm design,

model architecture, and evaluation methodologies. However, it also presents exciting opportunities for advancing the field of NLP and exploring new applications at the intersection of affective computing and artificial intelligence.

1.3 Novelty

The major novelty of the proposed work is as follows:

- Learn to Feel- Our model can understand and express emotion. Till now researchers who have worked on Generative AI have been able to generate text but not with emotional intelligence. Even ChatGPT-based chatbots can generate text on some specific static sentiment classes. Generating text with emotional intelligence comparable to human intelligence is the first of its kind and novel.
- Authorial Style Preservation: A significant aspect of our work is preserving the author's writing style across different emotional contexts. Our model can accurately predict what an ancient author would say or write in the latest topic or event. Ensuring that generated text maintains the author's unique linguistic characteristics, we enable more authentic and engaging expression of the author's emotions in textual content.
- Surpassing the performance of baseline: Through rigorous iteration and parameter tuning, we surpassed the metrics established by our baseline paper. Notably, our generated texts not only accurately conveyed the intended emotions but also preserved the original author's writing style, a significant achievement in mimicking emotional expression through text. Also our model is language independent. When we picked out the LLM with similar architecture but trained in Bengali language we found out that it was able to generate flawless coherent text imitating the style of the author. The fine tuning weights representing the strategy of generation were unchanged. So our text generation strategy works perfectly across different languages in zero shot.

1.4 Contribution

The main contributions during our research go as follows:

- We've performed strong literature survey covering all in the last 5 years within existing theoretical frameworks. This involved synthesizing relevant literature, identifying gaps, and developing a theoretical framework to guide the analysis.
- Delving into the selection of datasets, considerable effort was dedicated to analyzing and understanding various state-of-the-art options. We meticulously crafted pre-processing rules tailored to our specific task, laying the groundwork for subsequent stages.
- Our approach didn't end with text generation alone. Recognizing the importance of emotional context, we integrated RoBERTa, a variant of BERT specialized in text classification, trained on the Go-Emotions dataset for emotion classification. An iterative algorithmic refinement process ensued to optimize performance, exploring various parameters for both text generation and emotion classification.
- To evaluate the efficacy of our model, we developed and applied diverse metrics, including cosine similarity and winnowing, alongside fundamental measures such as part-of-speech (POS) distribution and average sentence length.
- Through meticulous documentation and articulation, we aim to convey the essence of our findings, methodologies, and insights to the scholarly community. This process involves distilling complex concepts and research outcomes into clear and coherent prose, supported by rigorous empirical evidence and logical reasoning. By adhering to the conventions and standards of academic writing, we strive to ensure the integrity, clarity, and accessibility of our manuscript. Through this endeavour, we aspire to contribute meaningfully to the discourse and advancement of knowledge in the domain of text generation, fostering dialogue, collaboration, and innovation within the research community.

2. Literature Review

In the past decade also we have seen a lot of advancements, catalyzing the research endeavours. The proliferation of Transformer based models showcased the capability of producing a next degree of relevant texts. Starting from a seminal work “Attention is All You Need” by Vaswani el.al(2017) which introduced the attention mechanism that led the model to capture intricate language patterns. It achieves success by guiding the model to assess the significance of individual words. Rashford et al.(2019) made an influential disruption with the paper “Language Models are Unsupervised Multitask Learners”. Generative Pre-trained(GPT) model was introduced that managed to become the state-of-the-art in many NLP based tasks like summarization, question-answering, translation and also on conditional text generation and prompt based generation that we focus on in this paper. Keskar et al.(2019) [6] presented “CTRL: A Con-ditional Transformer Language Model for Controllable Generation” facilitated granular manipulation of style, content and personalization of content generation. With the target of achieving more real text data, Generative Adversarial Networks (GANs) were introduced in NLP tasks. SeqGAN, proposed by Yu et al,(2017) [7] consists of a CNN based discriminator and an LSTM generator. These two units challenged each other and thrived by in turn beating each other to facilitate the overall improvement of the generated text.

Several papers have paved their way for leveraging text generating models for addressing real world challenges like data annotations, citations and many more. (Veniamet al. , Manoel et al. , Robert et al.) [9] in 2023 showed how annotating and evaluating is effectively carried out by Large Language Models (LLMs). Also (Chiang et al. and Lee et al.) [10] in 2023 took a step towards incorporating AI as an evaluation mainstream where it showed overwhelmingly similar evaluations scores as high graded evaluators. This shows how much understanding capabilities these models possess as

they are able to identify how high quality human written text looks like. (Tiyanyu et al and Howard et al,) [11] in 2023 were able to generate text based on citations. They specifically trained their LLM to produce text on given premises and facts. They also had put major emphasis on eliminating model's hallucination. (Daniel, Sara, Ashley and Mutlu) [12] in their work in 2023 have shown how LLM generated content can be used to design courses that are very helpful or educational for adults. Surely there must be areas that can help teenagers too. (Allen and Sanjay, 2022) [13] in their work also showed how LLMs can generate content that is engaging and poetic. Various interesting strategies were employed by them such as imposing various constraints of generation using filters. The idea was to restrict the models' normal generation and enforce them to be a little more casual similar to what a poet would do. While tackling real life problems like annotation and reviews posting using LLMs, the authenticity of crowdsourcing is challenged as surveys show a hike in the use of LLMs by crowdsource workers to enhance their income and effectiveness.

We have also looked deep into the papers that employ GANs as their main premises in generating text. It helped us to also gain insights of the faults in generation by GANs. This in turn was also a key factor in beating the generation quality of our baseline GAN model. Wu and Wang (2020)[19] introduced a Text Generation Service Model based on Truth-Guided SeqGAN. Their work focuses on enhancing the quality and coherence of text generated by generative adversarial networks (GANs) by incorporating truth signals as guidance. Lippi et al. (2019) [21] study investigates the statistical properties and linguistic characteristics of text generated by Long Short-Term Memory (LSTM) models. By examining patterns such as word frequency distributions, n- gram probabilities, and syntactic structures, the authors provide valuable insights into the generative capabilities of LSTM-based text generation models. Zhao et al. (2020) proposed FGGAN [22] introduced a novel approach to guiding the text generation process using feature representations. By leveraging feature guidance, the proposed model achieves enhanced performance in text generation tasks, allowing for more precise control over the style, content, and coherence of generated text. Chen et al. (2019) [24] proposed Customizable Text Generation via Conditional Text Generative Adversarial Network, enabling users to customize the content and style of generated text using conditional

generation techniques. By incorporating user-defined constraints and preferences, the proposed model offers flexibility and versatility in text generation, catering to diverse user needs and preferences.

There are some papers that helped us in identifying various research gaps and problems with the existing generating models. (Neeraj and Wenlin, 2023) [8] through their work introduced evidence on how LLMs hallucinate. By hallucination they meant that LLMs often produce undistinguishable free flowing text that would convincingly not be distinguished but ultimately the text generated is false and away from the ground truth. (Fryer et al., 2022) [16] discusses the limitations of existing counterfactual generation methods, which typically rely on wordlists or templates and produce simple counterfactuals that do not take into account grammar, context, or subtle sensitive attribute references. They introduce a task for generating counterfactuals that overcomes these shortcomings and demonstrate how large language models (LLMs) can be leveraged to make progress on this task. Another major point of focus was the improvement of coherence. In their study, (Sun, Yang, and Yousefzadeh, 2020) [17] focused on enhancing the coherence and consistency of sentences within a language generation model. They employed a two step approach to do so and achieved near human level coherence at certain fields of generation. In their work, (Son et al., 2022) [18] introduced the Writing Relevance (WR) training method designed for neural encoder-decoder natural language generation models. The primary goal of this approach is to enhance the coherence of generated text by incorporating negative examples. The WR loss function utilized in this method guides the regression of vector representations of both the context and generated sentences towards positive continuations, while simultaneously creating distance from negative examples. There are also researches that focus on adversarial attacks to Natural Language Generation modulus. One such example is Alsmadi et al. (2022) [26] conducted a comprehensive survey on Adversarial Machine Learning in Text Processing, providing insights into the use of adversarial techniques for enhancing the robustness and security of text processing systems. Their survey covers various adversarial attacks and defenses in the context of text processing tasks, offering valuable guidance for researchers and practitioners in the field.

(Wang, J., Zhang, C., Zhang, D., Tong, H., Yan, C. and Jiang, C. 2024) [30] in their recent journal points out the lack of casualty in the field on NLG even though they have excelled in the field of Translation, dialogue generation and many more. They suggest the incorporation of real world knowledge and external constraints should not be ignored. The paper identifies four solid problems in CTG and proposes different approaches to solve those CTG problems. (Hu, Z. and Li, 2021) [31] highlights that in the field of Casual Language Modeling involves two main fundamental tasks - attribute conditioned generation and text attribute transfer. They observed that historically these two problems were addressed separately. Their paper shows the disadvantages of such an approach and how their novel unified model can take care of all those drawbacks with appropriate measures. Shu, L., Papangelis, A., Wang, Y.C., Tur, G., Xu, H., Feizollahi, Z., Liu, (2020) [34] B. and Molino, P. presents their novel FVN which solves two leading issues of Text Generation which is diversity as well as according to the needs of the generation.

In recent times, with the advancement of quantum computers, QNLP has emerged to be a possibility and we have included studies about it while studying various strategies of generation. The core premise of Quantum Natural Language Processing (QNLP) lies in uniting language semantics and grammatical structure through the lens of categorical quantum mechanics. This framework holds promise for a deeper comprehension of linguistic meanings. Utilizing a visual language enables the representation of word meanings and their syntactic relationships, surpassing traditional dependency-based models. A Foundational model common to most QNLP approaches is the Categorical Distributional Compositional (DisCoCat) model[35]. DisCoCat facilitates the encoding of word and phrase meanings as quantum states, which can be implemented through quantum circuits. However, due to current hardware limitations,[36] QNLP primarily employs a hybrid methodology[37], combining classical and quantum operations. Despite theoretical advantages, practical constraints necessitate hybrid solutions, given the restricted scale of available quantum hardware, notably NISQ computers.

Also we have an amazing collection of latest survey papers that describes in detail about different approaches to text generation using different approaches. Fatima et al.

(2022) [27] conducted a systematic literature review on Text Generation Using Deep Neural Network Models. De Rosa and Papa (2021)[28] conducted a survey on Text Generation Using Generative Adversarial Networks, examining the use of GANs for text generation across different domains and applications. Iqbal and Qureshi (2020)[29] surveyed Text Generation Models in Deep Learning, providing an overview of state-of-the-art text generation models and techniques in the context of deep learning. Zhang, H., Song, H., Li, S., Zhou, M. and Song, D(2023) [32] in their work named “A Survey of Controllable Text Generation using Transformer-based Pre-trained Language Models; A Survey of Controllable Text Generation using Transformer-based Pre-trained Language Models” highlights how PLMs have helped researchers quickly able to manage to generate diverse and fluent texts. But there is some level of interpretability issues with the Neural nets. They mention how in the emerging years various subfields have emerged aiming to solve a specific problem with controlled constraints. Their work helps to bring all CTG techniques under a single umbrella.

Finally as we move towards our transformer based models where we need to be able to leverage the superiority of LLMs by using their pre-trained model and finetuning them. (Kale, 2020) [14] studied the pre-train + fine-tune strategy for data-to-text tasks. The experiments indicate that text-to-text pre-training in the form of T5, enables simple, end-to-end transformer based models to outperform pipelined neural architectures tailored for data-to-text generation, as well as alternative language model based pre-training techniques such as BERT and GPT-2. They rely on a significant amount of labelled data for each task, which is costly to acquire and thus limits their application to new tasks and domains.(Maruyama and Yamamoto, 2019) [15] discuss the use of pre-trained transformer language models for extremely low resource text simplification. They highlight the success of transformer-based translation models and the challenges of working with small parallel corpora for text simplification. The authors propose fine-tuning pre-trained language models for text simplification with a small parallel corpus, and they experiment with transformer-based encoder-decoder models and a language model called TransformerLM. They demonstrate that TransformerLM outperforms a strong baseline and achieves comparable performance with only 3,000 supervised examples. In a related study, Yang et al. (2020) [23] investigated Emotional Text

Generation Based on Cross-Domain Sentiment Transfer, exploring techniques for transferring sentiment across different domains to generate emotionally expressive text. Their study demonstrates the efficacy of sentiment transfer methods in imbuing generated text with desired emotional tones, facilitating more engaging and expressive communication.

Research Gap: Iterating through various papers, documentations and model descriptions, we identified some interesting issues faced by researchers working with language generating models. The most common of those are incomplete generation, lack of coherence and meaningful flow, deviation from initial prompt which is also known as long range dependency problem. Also there are problems of generating gibberish and repeat words. To improve the above mentioned problems we employed various tactical solutions. Here we are going to give an overview of how we manipulated the models generation so they are more humanly and free flowing. We optimized the attention mechanisms within the transformer architecture to better capture long-range dependencies. We began by refining the preprocessing steps and curating the training data to ensure high-quality input. By tweaking the attention layers, we enabled the model to maintain context over longer spans of text, reducing instances of deviation from the initial prompt and enhancing coherence. The repeat word problem was solved simply by adding a penalty if the model generates two exact same words consecutively. Also there were strategies involved that increases the casualness of the text that was generated which made the text more humanly.

3. Methodology

3.1. Problem Statement

In this study, the focus lies on the task of text generation employing a fine-tuned GPT-2 model. The process entails initializing the model with specific start words and an emotional label attributed to the author. The overarching objective is to produce text that exhibits characteristics akin to the author's writing style, vocabulary, emotional tone, and syntactic structure, thereby yielding coherent and thematically relevant output. This endeavor aims to mitigate the occurrence of nonsensical content and word repetition while ensuring adherence to the provided prompt.

Formally, the problem can be defined as follows

Let:

- *ECGText* represents the proposed model.
- *StartWords* denote the initial starting sequence.
- *EmotionalLabel* signifies the emotional attribute associated with the author.

The task involves generating a sequence of words, *GeneratedText*, which aligns closely with the provided prompt and embodies the desired authorial characteristics. The generated text is subject to a maximum length constraint, *MaxLength.*

The transformer-based language model is parameterized by Θ, encapsulating its learnable parameters. The scoring function $f\Theta(X, Yi)$ evaluates the likelihood of token y_i being the next word in the sequence, given the input X and the model parameters Θ. This function is defined as

$f\Theta(X, Yi) = g\Theta(h\Theta(X), y_i)$

Where *hΘ(X)* represents the hidden representations derived from the transformer model for the input sequence X and gΘ maps these representations and the output token to a score.

The probability of observing token yi as the next word, conditioned on the input sequence and the model parameter Θ, is computed using the softmax function:

$$P(y_i|X,\Theta) = \frac{e^{f_\Theta(X,y_i)}}{\sum_{j=1}^{m} e^{f_\Theta(X,y_j)}}$$

3.2 Working of the models

Bi LSTM and LSTM Models: We tried to train our own models based on CNN discriminator and LSTM and BI-LSTM generators. The problem with those generators was that they often ran into overfitting, required a lot of data for training, repeat words leading to incomplete generation. Due to this problem of overfitting, often when a prompt is given from the corpus, it tends to copy exactly the same text present in the corpus. Thus, the model did not make into evaluation phase as without nobel text, we cannot compute cosine similarity or winnowing score. The datasets despite being enormous were insufficient to train the model to generate quality text. Also there were problems of repeating and at times printing gibberish.

Indexed based zero-shot model: Next we started thinking about generation strategies from scratch which led to the development of one of the zero shot learning generative model. Our aim at this point was to generate noble uncopied text. We worked on generating noble paragraphs. Noble paragraph was made of sentences that might be present as it is in the corpus. But those sentences together would make a paragraph which is completely noble. Now it order to make the generation meaningful, we indexed the whole corpus of training data by sentences, assigned a cosine similarity function to score the sentences. Now the first sentence would be given as a prompt. The model will calculate cosine similarity score of all the sentences wrt the prompt sentence. Now the model starts picking the most relevant next sentences. Inorder to restrict the model from picking sentences consecutively, we attached a min distance condition where a

sentence would only be picked if it is at least 100 lines apart from the last picked sentences. Some of the generations were shockingly too good. Also as they picked sentences from the corpus, the model imitated the author's vocabulary as well as writing style like similar average sentence length. Having seen such versatile generations given the simplicity of the model pushed us further to take this sentence level generation to word level generation which led to the formation of nobel sentences as well.

ECG Text: Our approach begins with fine-tuning the GPT-2. Fine-tuning involves adapting the pre-trained GPT-2 model to a specific dataset or task by further training it on domain-specific data. During fine-tuning, we employ various training arguments to optimize the model's performance. Specifically, we set the `num_train_epochs` parameter to 3, indicating the number of training epochs, and `per_device_train_batch_size` to 4, which determines the batch size per device during training. Additionally, we configure the `save_steps` parameter to 10000, specifying the interval at which model checkpoints are saved during training. To manage storage resources efficiently, we limit the total number of saved checkpoints using the `save_total_limit` parameter, which is set to 2.

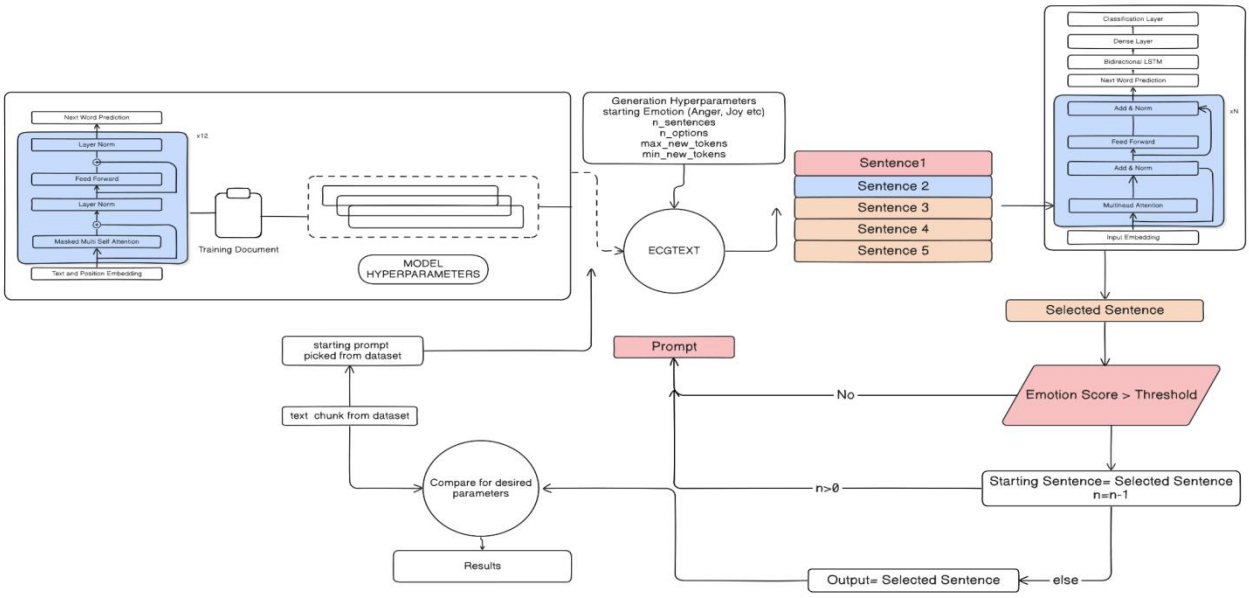


Figure 1: Architecture of ECGText

We fine tune the model on the relevant dataset which is Obama RNN or Movie Review in this case.Once the ECGText model is fine-tuned for text generation, we proceed with the synthetic text generation algorithm, which integrates emotion prediction into the process. The algorithm operates as follows:

Initialization:

- Set hyperparameters including the `starting_emotion`, which specifies the initial emotion for text generation, `n_sentences` representing the desired number of sentences to generate, n_options indicating the number of options to consider for each sentence, `max_new_tokens` defining the maximum number of tokens to generate per sentence, `min_new_tokens` specifying the minimum number of tokens to generate per sentence, and n determining the number of sequences to generate for each sentence.

Text Generation Loop:

- Start with a given starting sentence and initialize it with the specified `starting_emotion`.
- For each sentence to generate:
    - Generate n candidate sequences using the fine-tuned GPT-2 model.
    - Utilize a sentiment analysis pipeline to predict emotions for each generated sequence.
    - Select the sequence with the highest emotion score.
    - If sufficient emotion score is not achieved then discard the generation and start regenerating again from first step, otherwise
    - Append the selected sequence to the previous sentence.

3.3 Datasets

3.3.1 *Obama Text Data:*

i. The dataset comprises 1000 texts of Barack Obama's speeches in .txt format.

ii. Collected from reliable sources, covering diverse themes like firearms, Middle East issues, economics, and religion.

iii. Unlabelled dataset termed the Obama Speech Dataset, devoid of explicit labeling.

iv. Size: 1000 texts, providing a substantial corpus for analysis of Obama's oratory patterns.



Figure 2: Wordcloud of Obama Dataset

3.3.2 *Film Review Data:*

i. "Film Review Dataset" contains 1000 positive and 1000 negative reviews, labeled as P (positive) or N (negative) for sentiment.

ii. Size: 2000 reviews in total evenly split between positive and negative sentiments.

iii. Reviews encompass opinions, critiques, and evaluations of various films.



Figure 3: : Wordcloud of Film Dataset

3.3.3 *Rabindranath Tagore Data:*

i. Addressed the need for a comprehensive dataset with granular access to Tagore's works across genres.

ii. Includes 8 CSV files under /CSV directory and 7 txt files under the /txt directory containing the complete works of Rabindranath Tagore.

iii. Each CSV file includes all the literary works found in that genre in CSV format



Figure 4: : Wordcloud of Bengali Dataset

3.4 Algorithm

Algorithm 1: Proposed Algorithm

Data: Preprocessed dataset containing N text samples annotated with emotion labels, an emotional intelligence module, GPT-2 architecture, an Input Sequence X, an Output Sequence Y

Result: Generated Text Y_{gen}

$p_{EI} \leftarrow softmax(f_{EI}(X))$;
$i \leftarrow 0$;
while $i < N_{sentences}$ **do**
    $Y_i \leftarrow M_{GPT-2}(X, p_{EI}) = argmax_Y P(Y'|X, p_{EI})$
    $Y \leftarrow Y_i$
    i=i+1;
end
$L \leftarrow L_{LM}(X, Y) + \lambda * L_{EI}(X, p_{EI})$;
$e \leftarrow 0$;
while $e < Epochs_{max}$ **do**
    Compute gradients of the joint loss L with respect to model parameters;
    Update model parameters using gradients and optimizer;
    e=e+1;
end
Adjust necessary parameters;
while *not_converged* **do**
    Evaluate model performance on validation set;
    Update parameters based on performance;
    Check convergence criteria;
end
$Y_{gen} = argmax_Y P(Y|X_{prompt}, E_{desired})$;

Emotional Intelligence Module Integration: In this step, the emotional intelligence module is integrated into the text generation process. It computes emotional probabilities denoted by *p(*EI) using the `softmax` function applied to the output of the emotional intelligence module's classification function. The classification function outputs scores denoted by *f*EI(X), representing the likelihood of various emotional states given an input text X. The `softmax` function converts these scores into a probability distribution, ensuring that they sum up to 1.

Adaptation of transformer Architecture: Here, the transformer architecture is adapted for text generation while considering emotional probabilities. The algorithm initializes the generated text sequence denoted by Y as an empty sequence and enters a loop for text generation. Within this loop, it generates text iteratively by leveraging the modified GPT-

2 model denoted by MGPT-2 . This model takes the input text X and the emotional probabilities *p*EI as input to generate the next token in the sequence. The generated token is then appended to the existing text sequence Y.

Training Objective Training the model involves defining a joint loss function that incorporates both language modeling and emotional intelligence objectives. The language model loss LLM measures how well the model predicts the next token in the sequence based on the input text X and the generated text sequence Y. The emotional intelligence loss LEI evaluates how accurately the model captures the desired emotional tone based on the emotional probabilities *p*EI. The overall loss L is a combination of LLM and LEI, with a trade-off parameter λ controlling the balance between the two objectives.

Fine-tuning and Optimization: This step focuses on fine-tuning the model parameters to minimize the joint loss function L. It initializes an epoch counter denoted by e and enters a loop for training epochs. Within each epoch, the algorithm computes the gradients of the joint loss function with respect to the model parameters. These gradients are denoted by ∇L. The model parameters are then updated using these gradients and an optimizer denoted by optimizer, such as stochastic gradient descent (SGD) or Adam. This process continues until the maximum number of epochs denoted by Epochsmax is reached.

Necessary parameter Optimization: Parameters, such as the learning rate denoted by learning rate and regularization strength, play a crucial role in model performance. This step involves tuning these hyperparameters to optimize model performance further. It initializes a loop for hyperparameter optimization and evaluates the model's performance on a validation set. Based on this evaluation, it updates the hyperparameters iteratively until convergence, where the model achieves satisfactory performance.

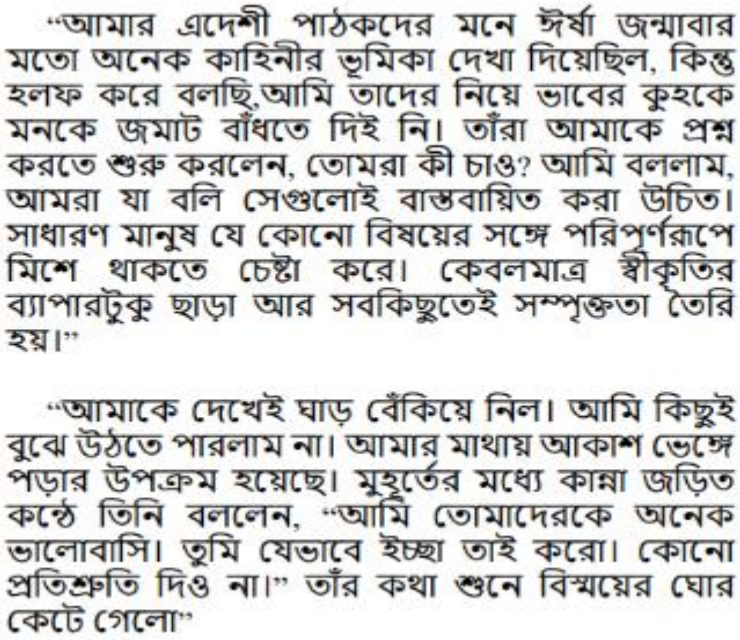
Text Generation: Finally, the algorithm outlines the process for generating text with a specified emotional tone. Given an input text prompt `Xtextprompt` and a desired emotion category `Edesired`, it generates text that embodies the desired emotion. This

involves leveraging the trained model and using `argmaxY` to find the most probable sequence of tokens given the input prompt and desired emotion. The generated text is denoted by `Ygen`.

4. Result and Analysis

4.1 Output:

Through our work we were able to generate good quality text that cannot be found at any document, i.e., the text generated is absolutely novel and is not lifted from any corpus. Even for the Bengali model, the text generated is coherent, matches the prompt and tries to conserve meaning for a long generation. Some of the examples are:



"আমার এদেশী পাঠকদের মনে ঈর্ষা জন্মাবার মতো অনেক কাহিনীর ভূমিকা দেখা দিয়েছিল, কিন্তু হলফ করে বলছি,আমি তাদের নিয়ে ভাবের কুহকে মনকে জমাট বাঁধতে দিই নি। তাঁরা আমাকে প্রশ্ন করতে শুরু করলেন, তোমরা কী চাও? আমি বললাম, আমরা যা বলি সেগুলোই বাস্তবায়িত করা উচিত। সাধারণ মানুষ যে কোনো বিষয়ের সঙ্গে পরিপূর্ণরূপে মিশে থাকতে চেষ্টা করে। কেবলমাত্র স্বীকৃতির ব্যাপারটুকু ছাড়া আর সবকিছুতেই সম্পৃক্ততা তৈরি হয়।"

"আমাকে দেখেই ঘাড় বেঁকিয়ে নিল। আমি কিছুই বুঝে উঠতে পারলাম না। আমার মাথায় আকাশ ভেঙ্গে পড়ার উপক্রম হয়েছে। মুহূর্তের মধ্যে কান্না জড়িত কণ্ঠে তিনি বললেন, "আমি তোমাদেরকে অনেক ভালোবাসি। তুমি যেভাবে ইচ্ছা তাই করো। কোনো প্রতিশ্রুতি দিও না।" তাঁর কথা শুনে বিস্ময়ের ঘোর কেটে গেলো"

Only the first sentence in each case was provided as a prompt. As it is evident from the above text, once we adjust just the parameter weights, it works exceptionally well for all languages. A few changes here and there to adjust language dependency is enough for the model to produce author-like generation.

4.2 Metrics Evaluation:

Here, we introduce some metrics to analyze generated texts from various aspects, such as plagiarism percentage, text similarity with trained dataset, word frequency and emotional tendency. We use Winnowing to calculate the similarity between generated texts and training texts, analyze text diversity by cosine similarity, analyze the word frequency by POS tagging, and analyze the subjective and sentimental of generated texts by SentiWordNet.

In Table 1, we also analyze the average sentence length of the generated texts. The average sentence length is expressed as the average of the number of words in each sentence. Often, longer texts and comments give people a more real feel, so longer average length of generated texts mean better performance of text generation method.

Dataset	Markov	Seq2Seq	RNN	CTGAN	ECGText	Ground truth
Movie	12.56	9.91	13.5	14.42	23	14.21
Obama	13.64	10.88	14.2	14.95	20	15.14
Bengali					12.67	13.67

Table1: Average Sentence Length

Then, we use POS tagging to analyze the frequency of special keywords (e.g. noun, verb and adjective) of all generated texts and real texts in Table 2. From Table 2, we can find that the word frequency of the texts generated by ECGText is more similar with the ground truth texts in most cases, especially with noun frequency.

Datasets	Markov			Seq2Seq			RNN			CTGAN			ECGText			Ground		
	Noun %	Verb %	Adj %	Noun %	Verb %	Adj %	Noun %	Verb %	Adj %	Noun %	Verb %	Adj %	Noun %	Verb %	Adj %	Noun %	Verb %	Adj %
Movie	27.15	20.39	9.22	27.35	22.04	10.07	25.2	17.26	11.65	25.86	17.89	10.44	23.08	14.98	9.25	25.71	18.15	10.97
Obama	28.55	18.9	9.97	28.74	19.75	10.29	26.1	17.31	11.85	27.03	17.2	11.63	26.05	18.97	7.85	26.68	17.99	11.36
Bengali	-	-	-	-	-	-	-	-	-	-	-	-	42.12	14.2	7.69			

Table 2: POS Tagging

Then, we use the Cosine Similarity score to calculate the similarity between generated texts and real texts. In Figures above, the similarity of GPT2 model is around 0.1–0.2, while the similarity of Markov Chain and Seq2Seq model is around 0.03–0.06. So, in terms of grammar, we can find the texts generated by GPT2 are more real, and texts generated by Markov Chain and Seq2Seq are different from human written text and preserve less semantic meaning. The similarity score generated by Bengali-GPT2 is more than 0.5, as trained on a different dataset, implying a very semantic generated text. Next, we evaluate the diversity of generated texts of four text generation methods via winnowing score on all datasets. Winnowing score is calculated based on the Jaccard Similarity between the generated text and the real text.

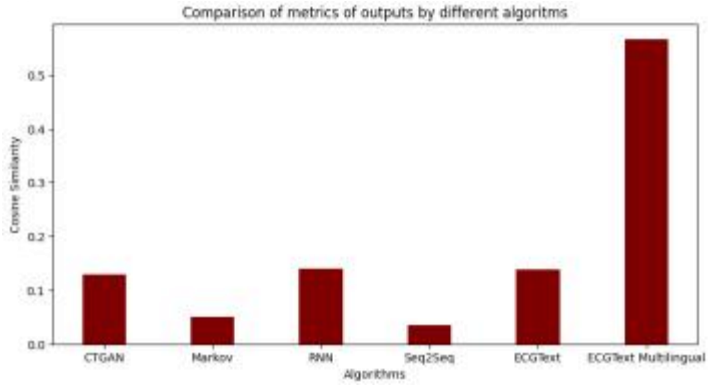


Figure 1: Cosine Similarity of Film Dataset

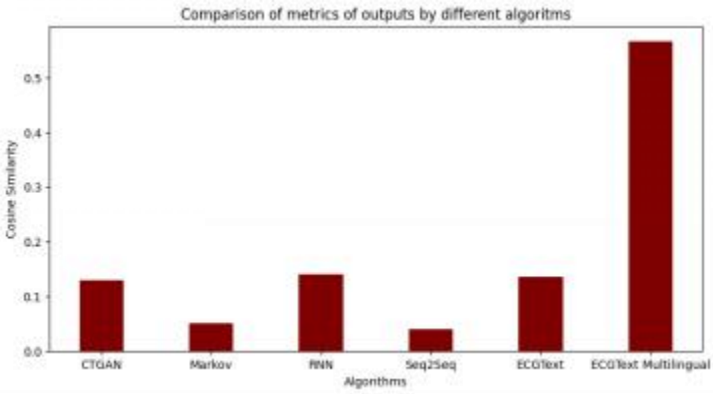


Figure 2: Cosine Similarity of Movie Dataset

As employed by baseline, the diversity of the text generated by the model and ground truth is calculated by winnowing score. Better and non copied text should have lower

winnowing score. A lower winnowing score represents that the model can generated properly without depending upon the content of the ground truth.

Dataset	Markov	Seq2Seq	RNN	CTGAN	ECGtext
Movie	0.32	0.3	0.33	0.34	0.21
Obama	0.33	0.32	0.35	0.36	0.2
Bengali	-	-	-	-	0.17

Table 3: Winnowing Score

Metrics	CTGAN			ECGText			Ground Truth		
Avg Sentence Length	14.685			20.75			18.5		
POS Tagging	N	A	V	N	A	V	N	A	V
	26.45	11.04	17.55	24.57	8.6	16.98	26.2	11.17	18.07
Cosine Similarity	0.13			0.144			-		
Winnowing	0.36			0.2			-		

Table 4: Comparison of results

Finally in Table 4 we clearly see the difference in performance on various metrics between our model,baseline model and the ground truth. The list of metrics tried is: (a) Average Sentence Length (b) POS Tagging (c) Cosine Similarity (d) Winnowing Score. In order to unify the score across different datasets we multiply each factor with 0.5 and deduce a score. Eg: If there are two readings of data X1 and X2, we do 0.5X1 * 0.5X2 to fill the scores. Metrics c and d don't apply to the ground truth. For calculating those scores we need the ground truth readily as we calculate the similarity and diversity of the generated text in comparison to the ground reality.

4.3 Analysis

Based on the four metrics, we can tell that ECGText can generate longer sentences than CTGAN which is also more close towards the ground text's average sentence length. Also transformers as a whole are better than GAN in generation due to which we have the desired higher Cosine Similarity score and lower Winnowing score than

CTGAN. The main differentiating factor would be the transformers ability of Natural Language Understanding which incorporates the writing style of the author resulting in higher cosine similarity score. Next the decoder part of the transformer responsible for generation is also better due to which more diversity is found in the text generated by ECGText resulting in lower winnowing score. The other two metrices: POS Tagging and Average Sentence Length was a measure of how much writing style of the author is preserved. Here we observe that Average Sentence Length of our model is more similar to the ground truth. Initially when we employed only the generation module of ECG Text, our POS Tagging results were also as similar to the ground as CTGAN. But later when we started taking emotions under consideration, there were several restriction that was applied to the model's generation due to which the text was a bit more casual, humanly and free flowing. This led to the shift of POS Tags and it was a bit off from the ground truth. Also overtime POS Taggers have improved a lot so it is not very logical to just conclude from numbers that CTGAN's generation maintained more resonance with ground.

5. Conclusion

The integration of emotion prediction into fine-tuned language models, as outlined in this research, presents a significant advancement in text generation technology. Here, we delve into the implications, challenges, and potential applications of this novel approach.

Integrating emotion prediction into text generation enables the production of content with heightened emotional nuance. By considering emotional context, the generated text becomes more resonant and relatable to users, fostering deeper engagement and connection.

Emotionally-aware text generation holds promise for enhancing user experience across various applications. Whether in chat bots, virtual assistants, or storytelling platforms, the ability to tailor content to users' emotional states can create more personalized and immersive interactions.

Despite its potential, accurately predicting emotions from text remains a challenging task. The complexity and subjectivity of human emotions pose hurdles for existing models, leading to potential inaccuracies or misinterpretations. Addressing these challenges requires refining emotion prediction algorithms and considering broader contextual cues.

Integrating emotion prediction into text generation raises ethical concerns regarding privacy, consent, and emotional manipulation. Ensuring transparency, fairness, and user autonomy is paramount to responsibly deploying emotionally-aware systems. Robust safeguards and ethical guidelines must accompany technological advancements in this area.

Scaling emotionally-aware text generation systems to handle real-time or large-scale applications presents technical challenges. Optimizing computational resources, model

efficiency, and deployment infrastructure is essential for practical feasibility and widespread adoption.

The success of emotionally-aware text generation hinges on interdisciplinary collaboration. Bringing together expertise from natural language processing, affective computing, psychology, and ethics fosters a holistic understanding of emotional intelligence in AI systems. Collaborative efforts drive innovation and address complex challenges more effectively. Future research should focus on refining emotion prediction models, developing adaptive text generation strategies, and exploring multi-modal integration. Additionally, conducting user studies and field experiments will provide valuable insights into the usability and impact of emotionally-aware text generation systems in real-world settings.

- *Future Work*

Extend the algorithm to incorporate multi-modal inputs, such as images, audio, or video, alongside textual context. Explore methods for effectively combining and processing multi-modal data to generate emotionally coherent and contextually relevant text.

Investigate approaches to personalize the text generation process based on individual user preferences, demographics, or interaction history. Develop adaptive algorithms capable of learning and adjusting to user feedback over time, providing tailored and engaging content.

Address ethical concerns related to AI-generated content, including biases, fairness, and potential societal impacts. Implement mechanisms for bias detection and mitigation, transparency, and accountability to ensure the responsible deployment of emotionally-aware text generation systems.

Conduct user studies and real-world deployments to evaluate the effectiveness and usability of emotionally-aware text generation systems in practical settings. Collaborate with domain experts in fields such as mental health, education, or entertainment to explore specific applications and user requirements.

6. References

[1] Chen, J., Wu, Y., Jia, C., Zheng, H. and Huang, G., 2020. Customizable text generation via conditional text generative adversarial network. *Neurocomputing*, *416*, pp.125-135.

[2] Keskar, N.S., McCann, B., Varshney, L.R., Xiong, C. and Socher, R., 2019. Ctrl: A conditional transformer language model for controllable generation. *arXiv preprint arXiv:1909.05858*.

[3] Yu, L., Zhang, W., Wang, J. and Yu, Y., 2017, February. Seqgan: Sequence generative adversarial nets with policy gradient. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 31, No. 1).

[4] Veselovsky, V., Ribeiro, M.H. and West, R., 2023. Artificial artificial artificial intelligence: Crowd workers widely use large language models for text production tasks. *arXiv preprint arXiv:2306.07899*.

[5] Chiang, C.H. and Lee, H.Y., 2023. Can large language models be an alternative to human evaluations?. *arXiv preprint arXiv:2305.01937*.

[6] Gao, T., Yen, H., Yu, J. and Chen, D., 2023. Enabling large language models to generate text with citations. *arXiv preprint arXiv:2305.14627*.

[7] Leiker, D., Finnigan, S., Gyllen, A.R. and Cukurova, M., 2023. Prototyping the use of Large Language Models (LLMs) for adult learning content creation at scale. *arXiv preprint arXiv:2306.01815*.

[8] Roush, A., Basu, S., Moorthy, A. and Dubovoy, D., 2023. Most language models can be poets too: An AI writing assistant and constrained text generation studio. *arXiv preprint arXiv:2306.15926*.

[9] Wu, Y. and Wang, J., 2020. Text generation service model based on truth-guided SeqGAN. *IEEE Access*, *8*, pp.11880-11886.

[10] Lippi, M., Montemurro, M.A., Degli Esposti, M. and Cristadoro, G., 2019. Natural language statistical features of LSTM-generated texts. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(11), pp.3326-3337.

[11] Yang, Y., Dan, X., Qiu, X. and Gao, Z., 2020. FGGAN: Feature-guiding generative adversarial networks for text generation. *IEEE Access*, *8*, pp.105217-105225.

[12] Chen, J., Wu, Y., Jia, C., Zheng, H. and Huang, G., 2020. Customizable text generation via conditional text generative adversarial network. *Neurocomputing*, *416*, pp.125-135.

[13] Yu, L., Zhang, W., Wang, J. and Yu, Y.S., 2017. Sequence generative adversarial nets with policy gradient. 492 In. In *AAAI conference on artificial intelligence* (Vol. 493).

[14] Huang, W., Hew, K.F. and Fryer, L.K., 2022. Chatbots for language learning—Are they really useful? A systematic review of chatbot‐supported language learning. *Journal of Computer Assisted Learning*, *38*(1), pp.237-257.

[15] Sun, R., Yang, J. and Yousefzadeh, M., 2020. Improving language generation with sentence coherence objective. *arXiv preprint arXiv:2009.06358*.

[16] Son, S., Lim, J., Jang, Y., Lee, J. and Zhang, B.T., 2022. Learning to write with coherence from negative examples. *arXiv preprint arXiv:2209.10922*.

[17] Alsmadi, I., Aljaafari, N., Nazzal, M., Alhamed, S., Sawalmeh, A.H., Vizcarra, C.P., Khreishah, A., Anan, M., Algosaibi, A., Al-Naeem, M.A. and Aldalbahi, A., 2022. Adversarial machine learning in text processing: a literature survey. *IEEE Access*, *10*, pp.17043-17077.

[18] Wang, J., Zhang, C., Zhang, D., Tong, H., Yan, C. and Jiang, C., 2024. A Recent Survey on Controllable Text Generation: a Causal Perspective. *Fundamental Research*.

[19] Hu, Z. and Li, L.E., 2021. A causal lens for controllable text generation. *Advances in Neural Information Processing Systems*, *34*, pp.24941-24955.

[20] Shu, L., Papangelis, A., Wang, Y.C., Tur, G., Xu, H., Feizollahi, Z., Liu, B. and Molino, P., 2020. Controllable text generation with focused variation. *arXiv preprint arXiv:2009.12046*.

[21] Grover, L.K., 1996, July. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing* (pp. 212-219).

[22] Coecke, B. and Meichanetzidis, K., 2020. Meaning updating of density matrices. *arXiv preprint arXiv:2001.00862*.

[23] Borsley, R. and Börjars, K. eds., 2011. *Non-transformational syntax: Formal and explicit models of grammar*. John Wiley & Sons.

[24] 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*, pp.53490-53503.

[25] De Rosa, G.H. and Papa, J.P., 2021. A survey on text generation using generative adversarial networks. *Pattern Recognition*, *119*, p.108098.

[26] Iqbal, T. and Qureshi, S., 2022. The survey: Text generation models in deep learning. *Journal of King Saud University-Computer and Information Sciences*, *34*(6), pp.2515-2528.

[27] Zhang, H., Song, H., Li, S., Zhou, M. and Song, D., 2023. A survey of controllable text generation using transformer-based pre-trained language models. *ACM Computing Surveys*, *56*(3), pp.1-37.

[28] Kale, M. and Rastogi, A., 2020. Text-to-text pre-training for data-to-text tasks. *arXiv preprint arXiv:2005.10433*.

[29] Maruyama, T. and Yamamoto, K., 2019, November. Extremely low resource text simplification with pre-trained transformer language model. In *2019 International Conference on Asian Language Processing (IALP)* (pp. 53-58). IEEE.

[30] Zhang, R., Wang, Z., Yin, K. and Huang, Z., 2019. Emotional text generation based on cross-domain sentiment transfer. *IEEE Access*, *7*, pp.100081-100089.