Product Copy Generation for Fashion Footwear Data

Sudesna Baruah[1] *and Bagya Lakshmi V.*[2]

Abstract

Natural Language Generation is a specialized field of Artificial Intelligence (AI) that deals with generation of text, given a set of inputs, like text, image or both. It is also deals with training of a given algorithm or machine to learn about a particular information and build relevant set of information about it in the form of natural language or text, given an input, while validating the same. This technology has stirred up a storm in the field of AI and has a great impact on automation science.

Keywords : **Bi-directional and Autoregressive Transformer (BART), Convolutional Neural Network (CNN), Generative Pre-Trained (GPT), Long Short Term Memory (LSTM), Natural Language Generation (NLG), Product Copy Generation (PCG), Recurrent Neural Network (RNN)**

I. INTRODUCTION

Natural language generation (NLG) is a field of Artificial Intelligence that interprets data and presents it in such a way that it is easily understood by humans. NLG tools are used for processing large structured or unstructured data sets, for generating useful text, which relates or describes the content of the data. NLG solutions tools are so powerful that they have the capability of replacing a human assigned for a task. We utilize this technology, through this work in such a way that it minimizes human efforts and assists humans in generating text.

II. NATURAL LANGUAGE GENERATION (NLG)

We might not realize the applications of NLG in our day-to-day life, nowadays, but when we do all these things, we are actually using NLG offerings :

- **Online Shopping :** Product descriptions writing can be automated using NLG.
- **Talking to Voice Assistant :** When we interact with our voice assistant, like Alexa or Google, NLG technology is used to respond to our requests.
- **Playing Video Games :** *Call of Duty* generates weekly game summaries, thanks to natural language generation. So, it is likely that we may have had already interacted with Natural Language Generation without knowing it.

Most importantly, it is noted that NLG is a boon in the field of copywriting for a product and especially the copy writers themselves. The augmentation of this specialized field of AI covers the following aspects :

- It speeds up the writing process. The time spent in writing a copy by a human copy writer is saved.
- It shoulders repetitive tasks. Any newly added similar product can get its copy generated in no time.

Manuscript Received : August 10, 2021 ; Revised : September 3, 2021 ; Accepted : September 8, 2021. Date of Publication : October 5, 2021.

[1]S. Baruah, *Data Scientist,* Retail Strategic Initiatives, Retail Cluster, Tata Consultancy Services Pvt. Ltd., SIPCOT IT Park, Siruseri, Chennai - 603 103, Tamil Nadu.
Email : baruahsudesna@gmail.com ; ORCID iD : https://orcid.org/0000-0001-7298-1344
[2] B. Lakshmi V., *Principal Scientist,* Retail Strategic Initiatives, Retail Cluster, Tata Consultancy Services Pvt. Ltd., Magnum SEZ IT Park, Pallikaranai Marshland, Kotivakkam, Chennai - 600 096, Tamil Nadu.
Email : bagyalakshmi_vasudevan@yahoo.com ; ORCID iD : https://orcid.org/0000-0002-3969-771X

DOI : https://dx.doi.org/10.17010/ijcs/2021/v6/i5/166513

➷ It turns non-writers into passable writers. The AI generated copy enables a non-writer to add the necessary features associated with the product.

➷ Reduces research time to an extent. A lot of time is spent in understanding the product for writing a good description. The AI generated copy helps in overcoming this.

The amount of data that can be generated in one minute is mind blowing. Companies need to analyze and interpret this huge amount of data in the most efficient and cost-effective way possible. Artificial Intelligence through Natural Language Generation can provide a solution for this by extracting ideas and communicating them in natural language.

Ultimately, this solution serves to relieve the workload, increasing productivity, and efficiency without involving employees in tasks that can be easily automated.

A. Statement of the Problem

Retailers nowadays are looking for options to automate the process of writing product information. Hence, minimizing manual efforts as much as possible is the target for all retailers. At this stage NLG plays a vital role in the field of generating product titles, descriptions, related information etc.

Our goal is to write product descriptions through language generation techniques.

B. Objective

The aim of this project is to generate automated product description for fashion footwear, given some textual information, like title, brand, color, fabric etc.

The project objective will be useful for retailers across retail domain, for generating descriptions for any product with basic information available. The project targets automation of product copy writing process. It also aims at minimizing manual efforts of copywriters. These are achieved through natural language generation techniques.

C. Project Goal and Scope

The project aims at generating quality product descriptions that attract customers for purchase of products without the intervention of human efforts. The project takes only textual information into account for writing product copy.

The scope of such a project is automating the copywriting process by copywriters, thereby, minimizing manual efforts.

This leads to efficiency of a retailer selling products online.

D. System Overview

This project involves the implementation of Open AI's GPT-2 model for text generation. The data for text generation has been extracted from public retail websites that sell fashion footwear for both men and women. In our case study, we have chosen women's footwear only as it has more features to describe, compared to men's footwear. This consideration elucidates the efficiency of the model under consideration.

Also, the same can be used not only for copy generation of footwear products, but is applicable across all fashion products under retail domain like apparels, accessories, etc.

E. NLG Technique(s) Involved

The NLG model involved in this project is *Open AI's GPT-2.*

III. LITERATURE REVIEW

Here are some relevant papers and articles that have been read before working this technology out for generating copies for a given product.

A. Relevant Works

(a) ***Plug and Play Language Models :*** *A Simple Approach to Controlled Text Generation by Sumanth Dathathri, Andrea Madotto, Janice Lan, Jane Hung, Eric Frank, Piero Molino, Jason Yosinski, and Rosanne Li [1].*

Although large transformer based language models have shown incredible performance when trained on larger datasets, controlling attributes of the generated text has always been a problem. A simple alternative Plug & Play Language Model (PPLM) has been proposed for controllable language generation. It comprises of a pre-trained language model (LM) that does not require further training. This model is flexible in terms of differentiable attributes combination. The model is guided by sampling which entails forward and backward pass in which the model gradients push the hidden activation functions of the LM.

(b) *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks by Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela [2].*

Pre-trained models trained on large data have proven to produce state-of-the-art results but their ability to access and manipulate the knowledge precisely is limited. We produce a general purpose fine tuned recipe generator (retrieval augmented generation – RAG) which combines pre-trained parametric and non-parametric memory for language generation. The model generates more specific, diverse and factual language than state-of-the-art seq-2-seq model.

(c) ***Hugging Face's Transformers :*** *State-of-the-art Natural Language Processing by Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, JulienPlu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush [3].*

In recent times, Transformer based architectures have enabled higher capacity models that are capable of being trained on a large data. We present the 'textit' transformer, \textit{Transformers} is an open-source library that consists of state-of-the-art engineered transformers under a unified API. It is fast, robust, and has been designed for researchers, practitioners, and for industrial deployments.

(d) ***BART :*** *Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension by Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdel-rahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer [4].*

We present the BART model, which is an auto-encoder and has been trained through the following : 1) corrupting the text data with an arbitrary noise function. 2) Learning to reconstruct the original data. Despite being a simple structure, its performance is far better than that of models like BERT or GPT-n. Evaluation is purely on the basis of various noising techniques, where spans of text are replaced by single mask tokens. It is capable of achieving state-of-the-art results on a range of abstractive dialogues, question-answers, and text summarization which scores 6 ROUGE and 1.1 BLEU scores.

(e) ***Google Colaboratory :*** *Tool for Deep Learning and Machine Learning Applications [5].*

The given paper presents a tool for executing Deep learning models and also discusses about transfer learning models xlm and MobileNetV2. The tool is Google Colaboratory which enables developers and researchers to utilize the platform for processing large datasets. Google Colaboratory provides a free graphical processing unit (GPU) which is essential for disseminating deep learning models.

(f) *Night Time Headlight Detection using CNN Based Object Tracking [6].*

This paper explores a technique that changes the beam of a car headlight from high to low when another vehicle/car is detected approaching from the opposite side. This prevents causing of glare at the opposite driver's end and reduces the risk of accidents caused in the dark due to low visibility. The model uses a CNN structure for training on various images of car headlights, which helps in detection of the same and reducing the beam. In absence of other vehicles, the beam is set to high.

IV. FEASIBILITY STUDY AND REQUIREMENT ANALYSIS

A. Feasibility Study

The field of NLP is ever expanding and is rapidly progressing in various techniques used but the field of NLG is still a major challenge as there is involvement of a lot of trainings with huge data set for performance. On top of this, the text generated is uncontrollable and a lot of tuning is essential.

You may be able to clearly understand what is being spoken in a language but you may not be able to speak it yourself, or atleast not as proficiently. That is because the process of generating words and sentences in any language is much more complex than that of understanding it. Generating language uses more of our brain than understanding it does. It is not a surprise then that the process of natural language generation by artificially intelligent agents is more complicated and challenging than natural language processing. While natural language processing (NLP) enables computers to understand what humans say or type, natural language generation (NLG) gives computers the ability to generate output in a way that is easily understood by humans. This could be in the form of written text or speech. There may be many other things that are considered while generating text or speech in any language. Although humans get these steps mostly right, training machines to do the same can be extremely difficult. An artificially intelligent machine can memorize all the words and grammatical rules, but there are a lot of rules that it must follow while attempting to generate natural language or text.

B. Requirement Analysis

After extensive analysis of the problems, we are familiarized with the requirement that the current system needs. The requirements that the system needs are listed below :

- The system should be able to identify the product under consideration by virtue of keywords and phrases like 'title', 'brand', 'style', 'fabric', 'print' etc.
- The system should be able to generate a relevant description for the product based on the above keywords or phrases provided and recall appropriate text information fed to it during training phase.

V. DATASET UNDERSTANDING

The dataset is a collection of Fashion footwear : ID, Title, Brand, Price, and Description. The columns “Title + Brand” & “Description” forms the part of the training set for the model. For our ease, we combine the “Title” and the “Brand” columns and for testing/validation purpose, we shall be having the combination of “Title” + “Brand” columns only. The output of which will be the Product Description or the Product Copy of a product.

VI. SYSTEM DESIGN AND ARCHITECTURE

A. Solution Approach

Below are a set of approaches that have been taken into consideration for implementing the technology mentioned here for generation of text :

- **Business Understanding :** Understand business problems for business perspectives.
- **Data Understanding :** Collect data, perform NLP/NLG techniques and all necessary actions to meet business requirements from the data.
- **Data Preparation :** Perform various activities to derive insights from the raw data and prepare a final dataset. The final dataset is used for modelling. This step includes data cleaning and data transformation.
- **Modelling :** Select and apply various modelling techniques. Focus is on choosing a model that can best predict subscription to term deposit based on campaign.
- **Evaluation :** Thoroughly evaluate the model using statistical techniques ensures that business objectives are met.
- **Deployment :** Generate insights and recommendations from the data and model that can be presented to the business.

B. Architecture

Generative Pre-trained Transformer series-2 (GPT-2) is an open-source Artificial Intelligence model by Open-AI. It was released in February 2019. The highly intelligent model is capable of translating text, answering questions, summarizing passages, and generating text output. It is a general-purpose learner that has been pre-trained on billions of parameters. It can be re-trained with a specific data to perform text generation task with respect to data.

The GPT architecture comprises of a transformer model combined with an attention network. This combination makes it possible to learn and generate text. This model has outperformed many other traditional AI based models like CNN, and RNN based LSTM structures.

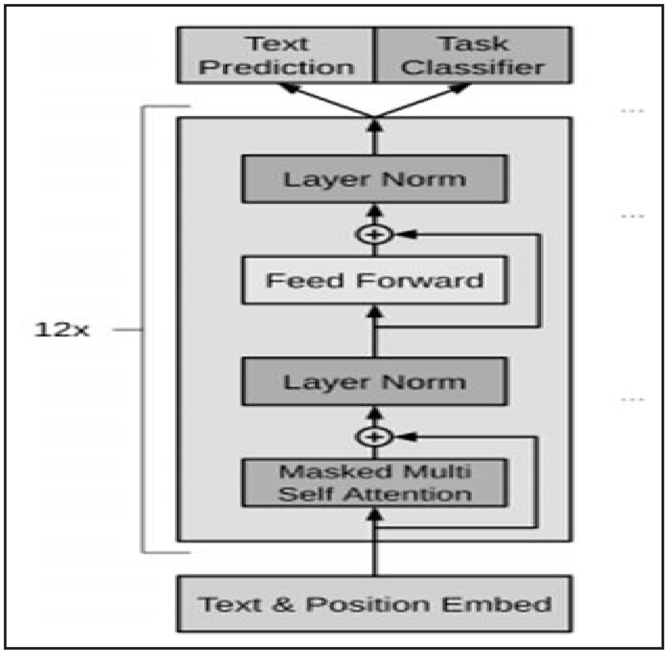

Fig. 1. General GPT-n Architecture

The Attention network forms the most significant component of the model. In fact, Attention is what makes the transformer model work on text generation.

The architecture is a twelve-layer decoder-based transformer which uses twelve masked attention heads. The Adam Optimization algorithm has been incorporated. The learning rate can be manually adjusted as per requirement.

C. Modelling

It is inevitable that Artificial Intelligence (AI) has already made its way into the copy and content creation world. Afterall, words are data that can be parsed like any other piece of information.

Platforms like GPT are pre-trained on a massive amount of information; some 45 terabytes of text data run through 175 billion parameters. GPT-2 is a powerful technology that is transforming the way we produce content intended for human consumption. One of the biggest challenges that any content marketer or media organization will tell you they have in the content game is simply keeping up. This is where AI-driven content platforms offer a tremendous advantage. They can spin up content virtually on demand.

D. Implementations on Dataset

The dataset considered for product copy generation task is of Fashion Footwear data.

The data comprises of 'Product ID', 'Product Title', 'Product Description', 'Image URL', 'Price' and 'Brand'.

We combine the *product title* and *brand* columns to include the brand name in the product title itself.

The next step is to load the medium version of GPT-2 model-*355M (*which is ideally pre-trained with 355 million parameters). Depending on the parameters

```
[ ]  prompt = 'Women''s Response Super Sneakers by Adidas'

     result_text_one = gpt2.generate(sess,
                   length=100,
                   temperature=0.85,
                   prefix=prompt,
                   nsamples=2,
                   batch_size=2,
                   truncate = '<|endoftext>|',
                   include_prefix=True,
                   return_as_list=True
                   )

result_text_one

['Isnt the Womens Response Super Sneakers by Adidas, something that you are looking for?  If you are looking for a pair of shoes that is compatible with your sr
 "Isnt the Womens Response Super Sneakers by Adidas, something that you are looking for? You will be pleased to know that there is no press release or anything
```

Fig. 2. Input and Output statements

used for pre-training, GPT-2 has four versions 'small', 'medium' 'large', and 'XL'. For our purpose of generating product copies for fashion footwear, we have chosen 355M medium GPT-2 model. We import the data and perform the necessary pre-processing steps like dropping nulls and lowercasing the text. We train-test split the data. We consider the 'Product Title', 'Brand', and 'Product Description' columns. We combine the title and brand columns to one and consider description column as the other. We add <|startoftext|> & <|endoftext|> tags at the beginning and end of each set of title and description respectively to mark the start and end of each set of text. This information is used for training the model. On a GPU run, the training roughly takes 20-25 minutes. On completion of training, we test the model performance using 'Product Title + Brand' information. When a product title + brand is provided in the prompt, the result is in the form of a caption describing the product. We have set the number of sample outputs to 2. Hence, there will be two copies for a given product title + brand. The samples can be set to any number '*n*', as per choice, but *n* samples=2 yields relatable copies. In certain cases, there might be a requirement to consider more samples of output for getting suitable descriptions or copies.

Details of the example shown in Fig.2 are given next.

Input : Women's Response Sugar Sneakers by Adidas

➷ **Output Result 1 :** Isn't the Womens Response Super Sneakers by Adidas something that you are looking for? If you are looking for a pair of shoes that is compatible with your dress, they are the shoes that can be worn. If you are looking for a pair that is not made of synthetic material, you can buy these shoes that are made of comfort material. They have a softer texture and a high-quality product that is compatible with any dress. If you are looking for a pair of shoes that can be worn as regular sneakers you have them.

➷ **Output Result 2 :** "Isn't the Womens Response Super Sneakers by Adidas, something that you are looking for? The Womens Rubber Sole Sneakers and Dress Sneakers are utilitarian type of shoes that you can wear all day and take for sure. You can use them to go to work or to go out with friends. You can even wear them to your student's dance recital. You can also wear them to daily life. You can wear them to party with friends"]

Fig.3 shows a slide displaying another example of footwear with its product information with image and the generated sequences or outputs.

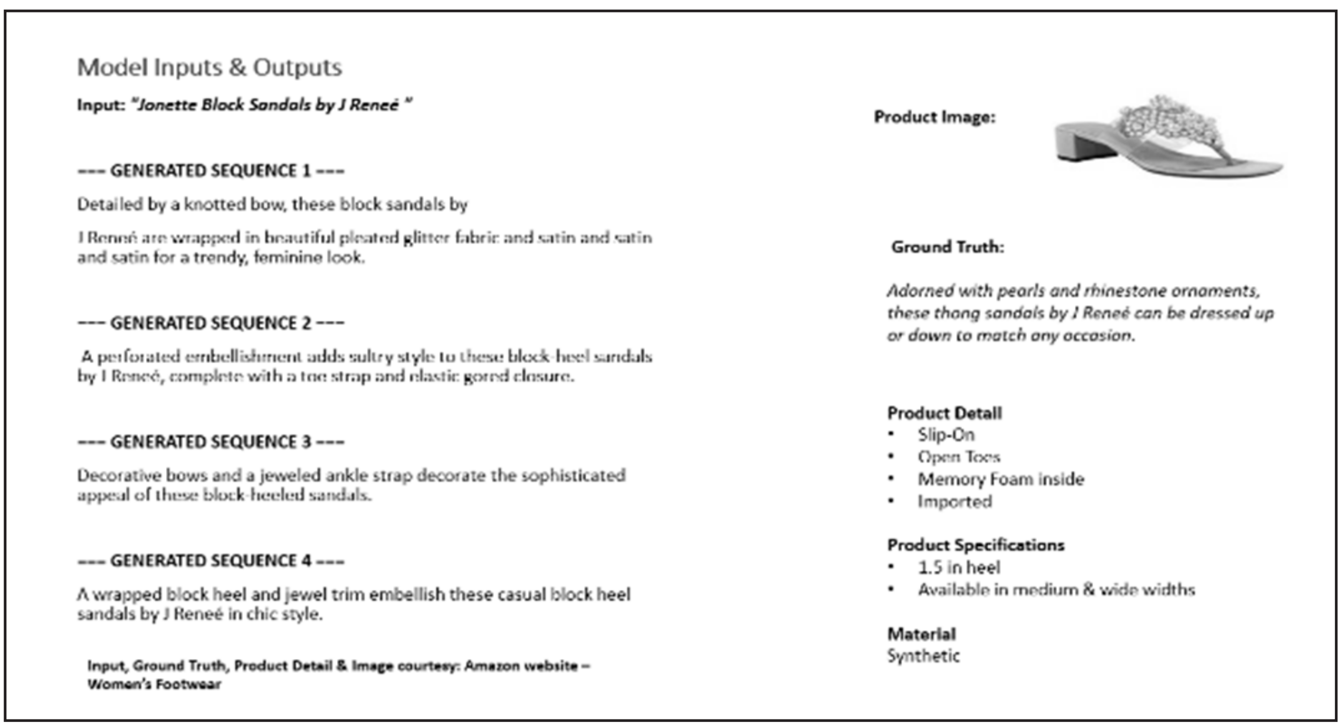

Fig. 3. Generated Sequences with Product Details

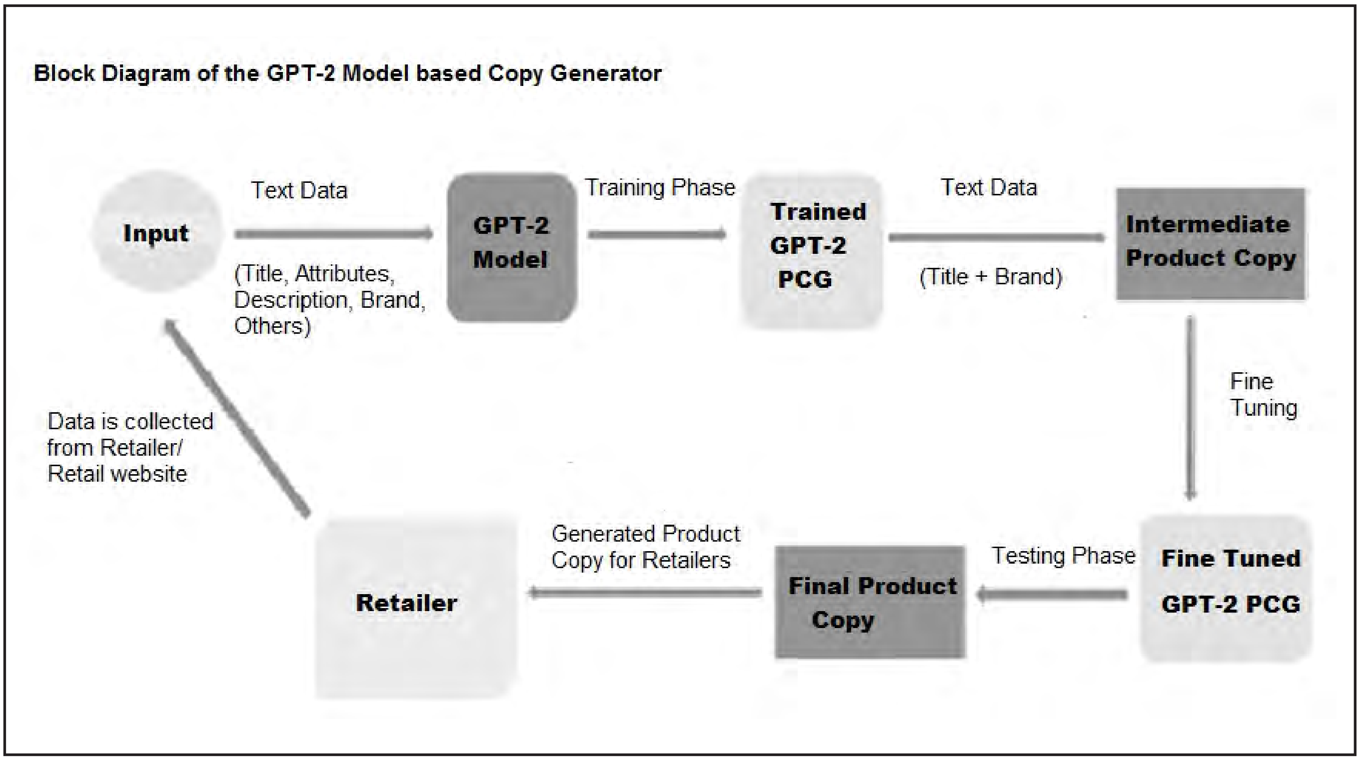

Fig. 4. Block Diagram Representation of GPT-2 Copy Generator

Fig. 4 highlights the whole process in a block diagrammatic representation.

VII. LIMITATIONS AND FUTURE ENHANCEMENTS

A. Limitations

While Generative Pre-Trained Transformers is an impactful technology in the field of artificial intelligence, it is not equipped to handle complex language formations. In absence of sufficient training beforehand, it is incapable of performing the text generation task of any specialized fields like generating copies for products like garment, footwear, grocery etc. or fields like literature, finance or medicine.

It is not a feasible solution currently due to the significant computing resources required.

Even choosing the correct value for 'nsamples' or number of sample outputs is another limitation of this model. The same value might not be suitable for all the elements of a data.

Moreover, in absence of an image, it is almost impossible to capture all the features of a product or entity with given text data alone. For example, pattern or design in a pair of shoes. This might or might not be mentioned as a part of text data. In absence of such information, it is somewhat difficult to generate a well describing copy. Thus, image plays a crucial role along with text in generating product copy.

B. Future Enhancements

Open AI has recently released GPT-3, which is an enhanced version of GPT-2. Trials are still going on, while the version is now available for working on.

GPT-3 is a major improvement upon GPT-2 and features far greater accuracy for better use cases. This is a significant step forward for AI development, impressively accomplished in just a two-year time frame.

Early tools that have been built on GPT-3 show great promise for commercial usability such as : no-code platforms that allow you to build apps by describing then ; advanced search platforms using plain English ; and better data analytics tools that make data gathering and processing much faster.

Open-AI announced plans to release a commercial API which will enable organizations to build products powered by GPT-3 at scale. However, many questions remain about how exactly this will be executed — pricing, SLA, model latency, etc.

Open-AI is openly committed to creating AI for the benefit of humanity, but still monitoring for misuse at scale will be difficult to achieve. This raises a broader question about the necessity of government involvement to protect the rights of individuals.

Also, use of text and image data processing models is encouraged in generating product copy. This makes use of an encoder-decoder based model which is a combination of CNN and RNN structures.

The encoder comprises of a CNN and RNN structure of which the CNN processes the image data and the RNN captures the text data. The decoder comprises of an RNN alone which generates the final text output.

The encoder trains on the image and text data. While validating, an image or image and text are given as input to the model for which a text is generated as output. This produces copies or captions, which literally captures all the product features.

VIII. CONCLUSION

This work aims at generating product copies/descriptions for a product, using text information only. The footwear dataset used in this work comprises of both text and image information for each product. We have chosen to generate copies using the text information alone. The more textual information is available of a product, for training, the better the generated copies.

The copies generated are purely based on machine understanding, which is derived through training on large data. The model used is a pre-trained model but needs specific data training to be able to generate suitable copies for a given entity. GPT-n series has brought in a revolution in the field of language generation. The latest version available is GPT-3 which is more robust than the GPT-2. This work can be reproduced using the latest version GPT-3 for deriving copies which are even better in terms of describing a product or any entity. Nevertheless, GPT-2 is a powerful tool.

GPT-2 itself has proved to be very powerful and can be utilized for minimizing human efforts by industries, especially retailers. It proves to be a boon in the field of generative text.

The copy generation technique can be further enhanced to capture more product features by including image data with text data. This is done by using an encoder-decoder based model, which is a combination of CNN and RNN structures. The encoder comprises of a CNN and RNN structure, of which the CNN processes the image data and the RNN captures the text data. The decoder comprises of an RNN alone which generates the final text output.

This work is an example of how text data can be utilized in generating text as output, as an automated process in order to reduce human efforts of copywriting.

APPENDIX

Fig. 1 : General GPT Architecture used for generating text based outputs.

Fig. 2 : Input & Output. Input provided to the model & the output description for it.

Fig. 3 : Generated descriptions obtained for a given product, having certain title, attributes/features and a gold standard product description (written by human).

Fig. 4 : Block diagram representation of the flow of steps of a GPT-2, text based product copy generator. It represents the end-to-end flow of the entire process of copy generation for a given product.

ACKNOWLEDGMENT

The author Sudesna Baruah is immensely grateful to her guide Bagya Lakshmi for actively mentoring throughout the entire research process and for synergistically helping in producing the desired outcome of the presented work.

AUTHORS' CONTRIBUTION

Both the authors had been actively involved in the presented work. Sudesna Baruah worked on creating the whole product copy generation of fashion footwear. She read the mentioned references, worked on the data and also built the model, fine tuned & processed the outputs generated from the model. Bagya Lakshmi closely monitored each step in this work, synergistically helped as a mentor in bringing the desired outcome.

CONFLICT OF INTEREST

The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in the manuscript.

FUNDING ACKNOWLEDGEMENT

The authors received no financial support for the research, authorship, and/or for the publication of the article.

REFERENCES

[1] S. Dathathri, A. Madotto, J. Lan, J. Hung, E. Frank, P. Molino, J. Yosinski, and R. Liu, "Plug and play language models: A simple approach to controlled text generation," 2019. [Online]. Available: https://arxiv.org/abs/1912.02164

[2] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W. Yih, T. Rocktäschel, S.Riedel, and D. Kiela, "Retrieval-augmented generation for knowledge-intensive NLP tasks," 2020. [Online]. Available: https://arxiv.org/abs/2005.11401

[3] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, and P. Cistac et al., "HuggingFace's Transformers : State-of-the-art Natural Language Processing," 2019. [Online]. Available: https://arxiv.org/abs/1910.03771

[4] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O.Levy, V. Stoyanov, and L. Zettlemoyer, "BART: Denoising sequence-to-sequence pre-training for Natural Language Generation, translation, and comprehension," in *Proc. of the 58th Annual Meeting of the Assoc. for Computational Linguistics,* July 5-10, 2020, pp. 7871-7880. [Online]. Available: https://aclanthology.org/2020.acl-main.703.pdf

[5] P. Gujjar J. and N. Kumar V., "Google Colaboratory: Tool for Deep Learning and Machine Learning Applications," 2021. [Online]. Available: http://www.indianjournalofcomputerscience.com/index.php/tcsj/article/view/165408

[6] S. Badri and V. R. Mounika, "Night time headlight detection using CNN based object tracking," 2019. [Online]. Available: http://www.indianjournalofcomputerscience.com/index.php/tcsj/article/view/150425

About the Authors

Sudesna Baruah has been working as a Data Scientist with Tata Consultancy Services, Bengaluru and then Chennai for over five years. She completed B.Tech. degree in Electronics & Communication Engineering from Manipal Institute of Technology, Majhitar Campus, Sikkim. She also has a Post Graduate Diploma in Data Science from Manipal Academy of Higher Education, Bengaluru Campus, Karnataka. She is a Data Science enthusiast with many research activities and publications in the field of Artificial Intelligence to her credit.

Bagya Lakshmi V. has been working as a Principal Scientist at Tata Consultancy Services, Chennai for over 15 years. She completed her B.Tech. degree in Electrical & Electronics Engineering from Government College of Technology, Anna University, Coimbatore. She is also a Postgraduate in Robotics. She is a pronounced Data Science enthusiast with many research activities and publications in the field of Artificial Intelligence to her credit.