

Text Generation with the help of natural language processing: A Review

Sanjay Kumar¹, Ms. Kusum Sharma²

Student, Computer Science & Engineering, RSR-Rungta College of Engineering and Technology, Bhilai, India¹

Assistant Professor, Computer Science & Engineering, RSR-Rungta College of Engineering and Technology,
Bhilai, India²

Abstract: The creation of the alphabet can be understood from a prehistoric era. The early man would write on caves as wall painting and then would take to clay tablets, later to parchment, paper and so on. In the modern world, there was printed text on a typewriter and in the current era there is written word on the computing devices which can then be printed out via a printer on paper. Artificial Intelligence, Machine Learning and Data Science have led to the development of Natural Language Processing which can help people in the output of text. Natural Language Processing involves constructs that require careful assistance by people. Automated Generation of sentences has been achieved utilizing Data Science and Artificial Intelligence. In this research article, we would like to present a survey of the present state of the art in sentence output. Also an algorithm called TAVERN which stands for Text Applying Vernacular is presented which helps in the output of sentences.

Keywords: Text, Artificial Intelligence, Machine Learning, Natural Language Processing, TAVERN.

I. INTRODUCTION

A language has developed over history and has required the input of scholars. Eminent researchers and scientists have marvelled at the ability of language to convey information and knowledge. English seems to be the common way of communicating information and knowledge to people across the world. The foundation of English starts in Kindergarten where students are made to learn the alphabet. Later they learn words and finally a scholar is able to write a sentence. Artificial Intelligence is a novel field which has involved the practice of Data Science in order to help people in sentence generation.

Natural Language Processing contains constructs and systems that can help people in creating sentences with the help of computers. The principle of sentence generation might require people that are able to utilize technology in order to produce output. In this research work a basic review of the literature on sentence output is presented. Also an algorithm called TAVERN (Text Applying Vernacular) is also presented.

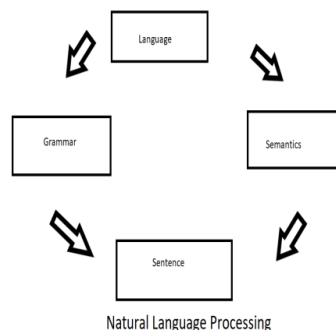


Fig1. Natural Language Processing

II. NATURAL LANGUAGE PROCESSING

A Natural Language Processing system involves the following phases

1. Language: Language consists of units that can convey meaning to people.

2. Grammar: Grammar defines the structure of a language.
3. Semantics: Semantics helps in giving meaning to language.
4. Sentence Output: The output of a sentence requires people and theory.

III. RELATED RESEARCH WORK

At present, we would like to review the state of the art in Natural Language Processing.

In [1] (Bas, A., Topal, M. Onat., Duman, C., Heerden, I. V. (2023)) discuss a glimpse of text generation by applying deep learning. A particular subject in Artificial Intelligence theory and in Natural Language Processing focuses on the mechanistic creation of realistic content. In order to course through the enormous and fastly developing content, this research work furnishes a brief review of phases in the past of text creation. To achieve these milestones, the research article elaborates deep learning projects for a large audience, concentrating on contemporary, linear, modern and abstract constructs as well as logical paradigm.

In [2] (Rani, A. (2023)) explore constructing a deep learning model for creation of text from input text data. The creation of text applying deep learning models has acquired a significant amount of interest in the recent past. This literature survey describe the various methodologies and ways utilized to construct deep learning models that can create modern text by taking into account an input set of text. The survey concentrates on various paradigms, knowledge preprocessing techniques, training strategemes, testing metrics and the usage of these concepts in the subject of Natural Language Processing (NLP) and creative writing. By analyzing of current research, this survey intends to further knowledge into the development done, issues faced, and pertinent areas for further researh in the creation of text applying deep learning models. In this recommended model, the authors demonstrate a new deep learning scheme for the creation of text applying coherent and relevant text. The recommended model, the authors display a new deep learning approach for the production of text applying concepts. The model intends to better existing frameworks and improve the production fo text. The recommended model banks on the power of self-attention processes to sustain logic in the input data and produce high quality output. The authors also develop novel structures to help the model comprehend context while keeping sanctity. By evaluating, the authors demonstrate the usefulness of their model in a vareity of tasks related to the production of text.

In [3] (Shi, Z., Chen, X., Qiu, X., Huang, X. (2018)) elaborate a variety of sentence production with Q Learning. The production of text is an important task in Natural Language Processing. Currently, various graph theoretical models have been recommended to better alleviate issues in the creation of text. Even though these models have achieved good rates of success, they might need sustenance. In order that these issues be employed in this research article, the authors utilize Q Learning for sentence production. Particularly, the Q Learning framework understands the concepts involved. The policy and gain are optimized in an alternate manner. The method recommended by the authors has two benefits: 1) The gain function is disciplined. 2) The creation policy is focused on creating a variety of sentences. Observational data are evaluated and it can be understood that the recommended scheme can produce content.

In [4] (Qader, R., Portet, F., Labbe, C. (2019)) understand partially supervised content creation by team based learning of basic language creation and models that comprehend natural language. In the creation of basic language, models that have been nurtured are now in fashion. Basic models require a good amount of data annotated in order to achieve fruition. Moreover, relating to such data is an enormous task. In this research work, the authors recommend a semi-detailed machine learning scheme that can look at un-annotated data and semi-annotated data if these is a need for it. It utilizes an NLG and an NLU end-to-end models which focus on history. Evaluation on two standard datasets display the fact that with limited content, the process can achieve small results while not applying any pre-processing. These results show us the way to the management of knowledge bases which is the major issue in E2E NLG schemes for new utilities.

In [5] (Guerrero, J. A. (2023)) comprehend isolating Artificial Intelligence created text applying machine learning. For human beings separating text from various sources is not easy. Natural Language Processing models have been produced in order to do language processing along with human being. This research work explores text detection models. The primary focus of this research work is to explain the present issue and text detection models in general. In the second chapter there is a sufficiently detailed literature review of text detectors. The third chapter explains an analysis where RoBERTa was utilized to evaluate linear models as compared to basic models.

In [6] (Lancaster, T. (2023)) describe machine learning and processes for text generation and ChatGPT which offer a digital watermarking solution. Text creation processes are very commonly displayed as a kind of data science. They can often be thought of as a recommendation system. The creation of tools such as ChatGPT appeared to have offered solutions to industry. This research article furnishes an offering to text creation which is meant for a novice explaining the kinds of evaluation that can be offered displaying the kinds of prompts that can be applied to create text while illustrating various watermarking techniques that may help in detecting generated text. A small evaluation of watermarking offers that this scheme is viable and that it can be sustained. Other possible solutions exist which can be then produced. This research article ends with a conclusion that discusses seven possible areas for exploration in the future.

In [7] (Ghatzia, D., Oliver Lemon and Verena Rieser (2017)) explain sentence creation that can help industry. Industry relies on sources of information for eg. data gathered from people. This research article furnishes an analysis of various subjects for unstructured data and pioneers the measurement of undertakings by people. The authors apply a theoretical process. It is shown that the application of Natural Language Generation (NLG) betters industry under various scenarios. In a study with 440 adults, the authors discovered that presentations applying NLG led to a betterment of industry. At last, the authors present an analysis of industry data and explore the future implicants for NLG system.

In [8] (Li, J., Li, Z., Mou, L., Jiang, X., Lyu, M. R., King, I (2020)) demonstrate text creation by applying fashion. In this study, the authors propose TGIF, a new framework for text creation by applying fashion. The authors begin by utilizing a process that can look at the sentences. Then a fashion based model applies industrialists to produce output. The dynamics between input and examples can be then understood in order to better industry. The offered model outperforms legacy systems. It can also achieve comparable performance to other systems.

In [9] (Liu, B., Yin, G (2023)) recognize graphutils for sentence generation. In sentence generation, LLMs (Large Language Models) tries to emit sentences utilizing the softmax function. Nonetheless, the node probabilistic knowledge of paired words based on a graphic-specific dataset is important in selecting a sentence which help in ensuring that the subject of the selected text is then suitable. In order to help digest the co-occurring knowledge, the authors propose a softmax function for process-particular sentence selection. Utilizing a static mathematical function helps in softmax to select the appropriate sentence by looking at both the local dataset and the graphic-particular knowledge base. The contemporary softmax procedure is helped by a term which leads to incorporation of local knowledge meaning that a model can then do probabilistic analysis amongst sentences in a graphic-particular knowledge base. The recommended softmax is static and can be offered into certain industries for sentence selection. By trying out the system, the authors demonstrate that the softmax function can help in industrial work by offering utilities for NLP processes. Also the text generated by softmax can be distinguished.

In [10] (Kumar, M., Singh, A., Kumar, A., Kumar, A. (2021)) discuss the evaluation of text creation by discussing data science. A chatbot is a computer program that can help in conversation with people applying processes in society. The objective of this work is to utilize and better data science in order to prepare a chat bot. In fashion chatbots are prepared applying people, machine learning or new techniques which can help humanity. There are three techniques like LSTM (Long and Short Term Memory), RNN (Recurrent Neural Network) and GRU (Gated Recurrent Unit) which have been evaluated. The prepared chatbots are then of use in industry.

In [11] (Bayer, M., Kaufhold, M, A., Buchhold, B., Keller, M., Dallmeyer, J., Reuter, C. (2023)) notice information addition in NLP: a new text creation method for short and long text categorizers. In various instances of Artificial Intelligence, development recommends that the management of training data might have a greater importance than the categorizers themselves. Thus information addition have been developed to better categorizers by manually creating training data. In this research article, we give and evaluate a content creation process meant for bettering the performance of categorizers for short and long texts. The authors were able to achieve good betterment when testing long as well as short text jobs with the betterment by our content creation method. Particularly with respect to data, augmentation can achieve gains of up to 15% and 4% inside a created content, with respect to a no addition baseline and another information addition methodology. As the present track of these created content can show good betterment in the present world data jobs. As the authors are evaluating the process from many viewpoints, they also get to know scenarios where the process may not be appropriate. The authors evaluate and look for a successful application of their scheme on various kinds of information.

In [12] (Mou, L. (2022)) maintain search and learning for manageable text generation. With the progress in deep learning methods, text creation is leading to a growing interest in machine learning (ML) community, leading to a wider applications and because it a pertinent part of sociology. Contemporary text creation methods are trained in a

supervised manner, that needs greatly labeled parallel corpora. In this research work, I will document the current work on learning and search methodologies to unsupervised text creation, where an algorithmic output evaluates the performance of a candidate text and discrete search algorithm to level out the noise and improve efficiency. The method is pertinent to the industry for creating viable products for a novel job, it also has a great social impact for helping languages.

In [13] (Patil, G. S., Thube, R. S., Patil, H., Dongradive, J. (2024)) give details on the comprehensive literature review on text creation applying deep neural networks. Text creation applying deep neural networks has led to good work in the fields of research. Deep neural networks such as recurrent neural networks (RNNs) and LSTMs have shown considerable capacities in creating relevant and pertinent content. By trying out these models on big content, they try out gathering the basic mechanism of the concept. This leads them to create novel content that is like the presentation and content of the data for training. Content creation applying deep neural networks has a good a good range of utilities, like chatbots, language transliteration, and prose generation.

In [14] (Hassani, H., Silva, E. S. (2023)) give a brief on the role played by ChatGPT in Machine Learning as to how AI-assisted chat interfaces are new fashion in the subject. ChatGPT, a chat AI interface that requires NLP and AI algorithms, is new fashion and is of good help in many verticles of the industry. Given the good in this project throught this relevant article, the authors intend to furnish a review of the potential gains and issues related to applying ChatGPT in machine learning, while offering reviewers with a overview of the benefits and lead to interest of its applicaiton for machine learning projects. The article explores how ChatGPT can help data scientists in assisting the workflow, also in data preprocessing and lceaning, training of the model and interpretation of the results. It also brings to attention how ChatGPT has the capability to lead to novel insights and better processes by looking at data. The authors then evaluate the benefits of ChatGPT's design, while looking at the capability to be attuned for a broad range of language related jobs and create synthetic data. Issues and limitations are also assessed, typically those about content when applying ChatGPT. Also, the research aritlce summarizes that the gains are good and ChatGPT has the required capacity to lead to good output. ChatGPT can help with a broad range of NLP jobs in machine learning, like translation, analysis of sentiment and text categorization. Moreover, while ChatGTP can help in saving time and resource. Also, the output of ChatGPT can be interpreted and could help in applications of Data Science.

In [15] (Li, C., Xing, W. (2021)) look at NLG applying deep learning to help Massive Open Online Course (MOOC). In all the educational resources amongst MOOCs such as movie lectures and coursework, the conversational platform is a good forum. Moreover, team work on MOOC forums are seen. The presence of MOOC learners can lead to good learning. This project intends to evaluate the deep learning based natural language generation (NLG) models that are good. Particularly, as per the framework of social scinces theory, the research has evaluated the application of latest deep learning models like Recurrent Neural Networks (RNN) and Generative Pretrained Transformer 2 (GPT-2) to provide furnish students with help for NLG and it may require further discussion. We first train a Recurrent Neural Network GPT-2 model with 13,850 entries. Metric based evaluation is also necessary. The output displayed the fact that GPT-2 outscored RNN in certain areas. The authors then summarized the results as gathered by people and GPT-2 which showed that GPT-2 is effective but it also requires technical guidance. The authors reviewed participants to assess whethered the gathered information would be of help. The outcome of the GPT-2 model was that it is quite supportive and that it can work well with people.

In [16] (Yang, S., Liu, Y. (2021)) prescribe input to sentence creation by the application of planning. Sentence creation based on planning from basic information is a good scheme that can help people in society. The contemporary methods can help in planning as they assist in the processes that can lead to fashion. In the development of a new sentence, the model at first does analysis and then looks at basic things, finally emitting output. Empirical analysis show that the process leads to sentence output. The model performs well on the E2E dataset.

In [17] (Zhang, Q., Guo, B., Wang, H., Liang, Y., Hao, S., Yu, Z) work on Machine Learning applied sentence output for harmonious people-machine association: present sate and future scope. In the past twenty years, the scenario of sentence output has gone through some changes and needs to be understood by analysis. Novel methodologies for sentence output varies from matching to new fashion. Also, the scientific goals have been guiding from output sentences to blending customized attributes so that variety can be achieved. Along with the fast development of sentence output applications, a thorough review is required in order to assist in the future. In this review paper, we furnish the basic structured scheme by fashioning the commonly applied knowledge and then explain the basic model of sentence output.

In [18] (Zhou, S., (2020)) speak about utilization of machine learning in sentence output. In the current past, new fashion and sentence output based on it requires analysis. In the shape of a various kinds of information, sentence output is a mainstream type of data and can be applied to certain areas. When someone needs to find and effectively look at information from large knowledge bases sentence output can provide some help. The application of fashion can be of assistance in the development of contemporary NLG processes. It can try to understand the relation between historical corpora. The sentence output system can also be applied in historical verticals and can produce fashion. This research article tries to find the sentence output process that applies fashion and then creates a cluster based approach. Empirical analysis shows that the recommended system achieves performance.

In [19] (Hu, Z., Chan, H. P., Huang, L. (2022)) give ideas on MOCHA which stands for a Multiple Organized Clustered and Hierarchical Application which can lead to a novel approach for sentence output from the perspective of analysis. Explaining applications to produce sentences is a new domain. In the recent past, there have been results from supervised models. Also there is significant differences between human output and machine generated sentences. In this research project, the authors recommended a new multi-faceted teaching apparatus for relevant sentence output based on the fashionable applications. The authors have evaluated the model on three traditional patterns like sentence output, sentence creation and sentence reading. Empirical analysis display that the model achieves output on supervised settings. The evaluation by people has shown that the model can output sentences.

IV. PROPOSED METHODOLOGY

The TAVERN algorithm stands for Text Applying Vernacular.

A Vernacular is a kind of language where words from local populace are derived along with standard vocabulary. Text can be written on a medium and consists of alphabets. A String can be defined as a sequence of alphabets A Word is a string that can be found in a dictionary. The English alphabet consists of 26 alphabets starting from A to Z.

The speciality in the TAVERN algorithm is the application of a vernacular. In the English language, there is vernacular from Ireland and America.

For instance, along with standard english, there are words like pint, eclair, asparagus etc., The TAVERN algorithm follows the following steps.

- 1. T - Templatize**
- 2. A - Approximate**
- 3. V - Visit**
- 4. E - Elicit**
- 5. R - Revise**
- 6. N - Nice**

Templatize phase consists of looking at templates of sentences.

Approximate phase consists of approximations of letters.

In the Visit phase, there is a visitor pattern that can combine letters.

In the Elicit phase, there is eliciting of vowels where they are needed after the Visit phase.

In the Revise phase, there is a revision of words so that it appears in a dictionary.

In the Nice phase, a nice sentence is emitted after taking recommendations from people.

V. CONCLUSION

We can arrive at the conclusion that there is a considerable body of work in natural language processing that can help in sentence generation. Also, sentence generation will need the help of people who are skilled in language.

An algorithm called TAVERN (Text Applying Vernacular) has been presented which can be of good help to the literary community.

REFERENCES

- [1] Bas, A., Topal, M. Onat., Duman, C., Heerden, I. V. (2023). A Brief History of Deep Learning-Based Text Generation, International Conference on Computer and Applications.
- [2] Rani, A. (2023). Developing a Deep Learning Model for Text Generation from a Given Set of Text, International Journal of Science and Research, Vol. 12, Issue 8.
- [3] Shi, Z., Chen, X., Qiu, X., Huang, X. (2018). Toward Diverse Text Generation with Inverse Reinforcement Learning, Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence.
- [4] Qader, R., Portet, F., Labbe, C. (2019). Semi-Supervised Neural Text Generation by Joint Learning of Natural Language Generation and Natural Language Understanding Models, Proceedings of the 12th International Conference on Natural Language Generation.
- [5] Guerrero, J. A. (2023). Detecting AI Generated Text Using Neural Networks, Office of Graduate Studies Texas A & M University-San Antonio.
- [6] Lancaster, T. (2023). Artificial intelligence, text generation tools and Chat GPT-does digital watermarking offer a solution, International Journal for Educational Integrity.
- [7] Ghatzia, D., Oliver Lemon and Verena Rieser, (2017). Data-to-Text Generation Improves Decision-Making Under Uncertainty, IEEE Computational Intelligence Magazine.
- [8] Li, J., Li, Z., Mou, L., Jiang, X., Lyu, M. R., King, I (2020). Unsupervised Text Generation by Learning from Search, 34th Conference on Neural Information Processing Systems.
- [9] Liu, B., Yin, G. (2023). Graphmax for Text Generation.
- [10] Kumar, M., Singh, A., Kumar, A., Kumar, A. (2021). Fourth International Conference on Computational Intelligence and Communication Technologies.
- [11] Bayer, M., Kaufhold, M. A., Buchhold, B., Keller, M., Dallmeyer, J., Reuter, C. (2023). Data augmentation in natural language processing: a novel text generation approach for long and short text classifiers, International Journal of Machine Learning and Cybernetics.
- [12] Mou, L. (2022). Search and Learning for Unsupervised Text Generation, AI Managize.
- [13] Patil, G. S., Thube, R. S., Patil, H., Dongradive, J. (2024). A Systematic Literature Review on Text Generation Using Deep Neural Networks.
- [14] Hassani, H., Silva, E. S. (2023). The Role of ChatGPT in Data Science: How AI-Assisted Conversational Interfaces are Revolutionizing the Field, Big Data and Cognitive Computing.
- [15] Li, C., Xing, W. (2021). Natural Language Generation Using Deep Learning to Support MOOC Learners. International Journal of Artificial Intelligence in Education.
- [16] Yang, S., Liu, Y. (2021). Data-to-text Generation via Planning, Journal of Physics.
- [17] Zhang, Q., Guo, B., Wang, H., Liang, Y., Hao, S., Yu, Z. AI-Powered Text Generation for Harmonious Human-Machine Interaction: Current State and Future Directions.
- [18] Zhou, S., (2020). Research on the Application of Deep Learning in Text Generation, CISAI 2020, Journal of Physics: Conference Series.
- [19] Hu, Z., Chan, H. P., Huang, L. (2022). MOCHA: A Multi-Task Training Approach for Coherent Text Generation from Cognitive Perspective, arXiv.