Beyond Neural Networks: Enriching ChatGPT with Rule-Based Approaches

PRINCY PAPPACHAN[1], MASSOUD MOSLEHPOUR[2], MOSIUR RAHAMAN[3],SOURASIS CHATTOPADHYAY[3], SREERAKUVANDANA[4]
[1]Center for the Development of Language Teaching and Research, Asia University, Taiwan
[2]Department of Business Management, Asia University, Taiwan
[3]Department of Computer Science and Information Engineering, Asia University, Taiwan
[4]Department of AI-ML and Cybersecurity, Jain University, India

ABSTRACT Over the years, Generative AI has revolutionised and transformed many aspects of life, leading to its latest development, Chat Generative Pretrained Transformer (ChatGPT). Trained on a large corpus of datasets, ChatGPT can generate relevant responses by identifying the statistical patterns and semantic connections. However, this reliance on statistical learning often limits its performance. This paper hopes to address this limitation by proposing an approach to enrich ChatGPT by integrating a rule-based approach over the neural network. Our proposed approach involves designing a rule-based system incorporating domain knowledge, explicit reasoning, and linguistic rules to improve the robustness of the model. This integration will also help ChatGPT to handle complex linguistic structures and generate consistent and deterministic responses.

KEYWORDS ChatGPT, neural-network approach, rule-based approach, GPT architecture

I. INTRODUCTION

The advent of Open AI's generative model that utilises statistical learning to make predictions based on patterns in the data has significantly revolutionised and transformed Natural Language Processing (NLP) [1]. Before GPT, Open AI algorithms included techniques like rule-based inference, maps for semantic linguistic analysis, Bayesian networks, similarity measure approaches like the k-means approach and support vector machines, genetic algorithms, and artificial neural networks (Duan et al., 2019). Newer AI algorithms adopted deep learning, advanced neural networks, and transfer learning that could run on extensive unstructured data. Integrating deep learning and neural network transformer architecture created the popular ChatGPT.

The success of ChatGPT is evident in various NLP tasks such as question-answering, dialogue generation, and sentiment analysis. However, ChatGPT has also demonstrated its inability to reason explicitly over the input, lack of interpretability of its predictions and sometimes incorrect and non-existent responses. This limitation is also apparent when dealing with complex sentence structures and linguistic ambiguity. The present article addresses this limitation by designing a rule-based system that operates in conjunction with the neural network of ChatGPT to augment its predictions and provide a more comprehensive understanding of the input text. This proposed approach has significant implications for enhancing the performance and interpretability of neural network-based NLP systems.

In the following section, this article reviews the literature on the different approaches and discusses how the rule-based approach can be integrated with the neural network approach to enrich ChatGPT's generated responses with a simple Python algorithm illustration. In the end, we conclude the article by highlighting the potential for future research in this area.

II. LITERATURE REVIEW

A. RULE-BASED APPROACH

The rule-based AI approach is a specific type of AI that employs rules to solve a problem. Essentially, human experts write these rules instead of AI learning from datasets. Many early rule-based systems were built on manually programmed rules that subject matter experts or linguists developed. They were also found to work well in straightforward tasks like parts-of-speech labelling. However, initially, it needed a lot of help to efficiently handle the complexity and variety found in natural language [6]. Since the early research in the discipline, rule-based approaches have been the fundamental element of language processing [4]. SHRDLU developed by Terry Winograd, is regarded as one of the oldest rule-based natural language processing systems [7]. The program could understand and manipulate the virtual world objects using commands based on a clear set of rules. The rule-based method made analysing and producing natural language possible, which helps encode linguistic information [5].

In the 1900s and 2000s, however, rule-based approaches were abandoned in favour of data-driven techniques, which came with the advent of machine learning and other statistical techniques [9]. Subsequently, rule-based techniques, when combined with models based on neural networks, garnered significant attention and interest [10].

Variability and linguistic ambiguity were handled exceptionally well by rule-based approaches. The key advantage of a rule-based approach is its ability to be interpretable. By encoding expert knowledge and domain-specific rules, these rule-based systems greatly aid in disambiguating the meaning of the words and phrases, making them especially useful for sentiment analysis, matching reading comprehension, and dialogue generation [11]. Rule-based approaches can also be designed to demonstrate insight behind their reasoning and prediction [12]. This gives rule-based approaches a significant edge over black-box machine learning models. This also makes the rule-based approach advantageous for applications that require transparency and accountability, such as legal or ethical decision-making [13]. Several successful rule-based systems have emerged over the years. For example, IBM's Watson famously defeated human champions in the quiz show Jeopardy, which relies heavily on handling complex language tasks [14]. Rule-based approaches, therefore, remain a crucial area of research in natural language processing, especially in combination with machine learning and statistical techniques [15]. They offer valuable inputs and insights that can help improve the performance and the interpretability of natural language processing systems.

B. NEURAL NETWORK APPROACH

Neural networks have been the next major focus in the domain of artificial intelligence and machine learning since its origin in the 1940s [16]. The earliest known neural network was built on a rather straightforward feedforward network for text classification and sentiment analysis [17]. In 1986, Rumelhart and McClelland proposed one of the earliest neural network models in NLP, the feedforward neural network [18]. This model was used to learn statistical patterns and dependencies in language data, particularly the mapping between orthography and phonology in English. Post this, a range of neutral network architectures has come about for handling different kinds of NLP tasks. For example, the Recurrent Neural Network (RNN) have been proven efficient in language modelling and sequence-to-sequence tasks such as machine translation [19]. On the other hand, CNN or Convolutional Neural Networks was used for text classification and sentiment analysis [20].

With time, more complex models were developed for handling sequence modelling and language generation tasks like the RNN and Long Short-Term Memory (LSTM) [21]. In the early 2010s, the introduction of word embeddings, such as Word2Vec and GloVe, helped to revolutionise the field of NLP [22]. These embeddings enabled neural networks to learn distributed representations of words, which enabled them to capture even more nuanced and complex relationships between words and phrases [23]. Ultimately, CNNs, RNNs, and transformers became some of the most often used neural network designs for natural language processing [24].

Later in 2017, the domain of NLP witnessed a significant development: the introduction of the Transformer architecture, which serves as the base for the GPT model (Generative Pre-trained Transformer) [25]. These models achieved state-of-the-art performance on various NLP tasks by relying on self-attention processes to learn contextual links between words and phrases [26]. Undoubtedly, neural networks transformed the NLP discipline and continued to be a prominent area of study. Despite their need for a large amount of training data, which remains a challenge, their propensity to discover erroneous correlations and their difficulty in recognising linguistic subtleties, NLP remains a robust tool for natural language understanding and generation.

C. HYBRID APPROACH

The next of the kind within the NLP model is the Hybrid approach, which combines the rule-based and neural network methods for NLP tasks [27]. The hybrid approach aims to leverage the strengths of rule-based and neural network methods while efficiently mitigating their limitations [28]. Rule-based systems can handle domain-specific knowledge and linguistic structures, while neural networks generally generalise learning from large amounts of data [29]. The hybrid approach offers a promising avenue for fusing the strengths of these approaches and improving performance [30].

Many studies have explored the potential benefits of integrating rule-based and neural network approaches in various NLP tasks. For example, machine translation uses rule-based systems to handle language-specific idiosyncrasies and improves the quality of translations generated by neural machine translation models [31]. Rule-based systems have been employed in sentiment analysis to enhance the interpretability and explainability of deep learning models [32]. Another case where rule-based systems have been employed is in question-answering to improve the accuracy of neural models by handling complex linguistic structures and domain-specific knowledge [33].

Comprehensively speaking, combining rule-based and neural network approaches has strengths and weaknesses. The choice of approach typically depends on the task and the context. There are also downsides to this integration; the integration of rules can restrict the model's adaptability, thus making it more challenging to accommodate new linguistic rules, types of language use or even the vase stipulation. Additionally, integrating rule-based and neural network-based approaches can be complex and time-consuming, and it can cost a lot of significant resources to design and maintain.

III. BOOSTING CHATGPT

The synergy of rule-based and neural network-based approaches depends on the particular NLP problem. One typical strategy is to use rules to direct the learning process of a neural network model [34]. In this approach, the neural network

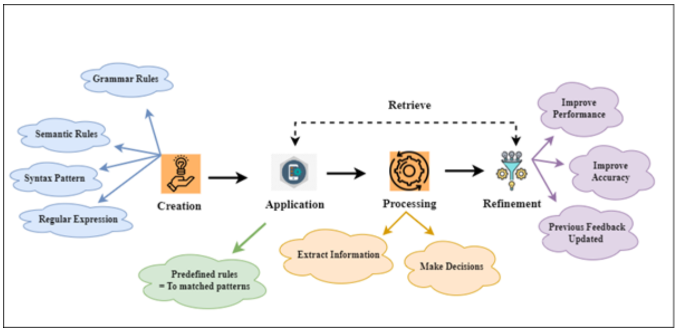

FIGURE 1: Rule based approach in Natural Language Processing

```
# Pseudocode algorithm for implementation Rule based and neural network approach
C_i is the Combined text (Rule based, Neural network)
R_i, N_i Rule based text and neural network text. C_i ∈ R_i, N_i
function init (generates_text),
Then
R_i rule generated_text = {Regular_Expression, grammar_rules,syntax_pattern,Semantic_rules}
N_i neural_generated_text= {Regular_Expression, grammar_rules,syntax_pattern,Semantic_rules}
C_i Combined_text= {R_i, N_i}
Return=C_i
Applied Rule based approach          //Neural_network generated numerical
                                     representation by converting seed
Generated_text=request_rules Return generetad_text
Then
C_i = {R_i, N_i}
Return C_i
Function processing (Combined text);
Processing_combined text=predefined (combined text)
Transformation process_text =R_i+N_i (predefined_combined_text)
Return processed_text
Then
After processing.                    // Improve Accuracy and performance
Refined _C_i = Request_refinement_rules (C_i)
Return refined_C_i.
```

is trained to learn from a set of explicit rules that capture a language's fundamental patterns and structures. This ensures that the accuracy and efficiency of the neural network can be improved and perfected by using rules that will direct its attention to the pertinent features of the input text [35]. For instance, in a sentiment analysis task where the goal is to detect whether a given input has a positive or negative sentiment, the hybrid approach can produce some significant results. With this approach, the rule-based approach can create a list of keywords defining what terms are associated with positive and negative sentiment and guide the learning process of the neural network model as it learns to correlate the text with the proper sentiment [36].

Another instance where we can employ the hybrid approach is a medical diagnosis task, where rules can be identified for particular diseases by incorporating fundamental symptoms and diagnostic standards. At the same time, the neural network can be taught to employ these guidelines to generate a diagnosis by learning from a dataset that involves medical records. Ultimately, the neural network, in conjunction with the rule-based approach, can produce diagnoses that are interpretable by medical experts and congruent with medical knowledge [40]. In the domain of NLP, an example is language generation tasks like text summarisation or question answering. In both tasks, the rule-based approach may provide a template or structure for the output text, alongside, a neural network used to generate the specific words and phrases [41]. In this scenario, the rules can guarantee that the output text is coherent and consistent, certifying that the neural network can capture the nuances of language [42].

A. ILLUSTRATION INTEGRATING RULE-BASED AND NEURAL NETWORK-BASED APPROACHES:

One of the first NLP techniques is the rule-based approach, which analyses and processes information in the text according to specified language standards. Applying a specific set of rules to identify frameworks, extract data, or carry out activities like text classification and other similar ones is known as a rule-based approach. Pattern matching and recurrent expressions are two standard rule-based methods.

This execution example demonstrates how a neural network, and a rule-based approach can produce a response in language production tasks. The rule-based technique sets the summary's structure and format, while the neural network generates the precise words and phrases. Combining these two methods can produce and sound more like the original text. Rules of language, such as grammar rules, syntactic patterns, and semantic conventions, are developed based on the objectives that are needed [45]. For obtaining information, making choices, or doing other duties, the text information is processed in line with the outcomes of the associated limitations. The regulations are occasionally revised and amended based on prior comments.

Algorithm for implementing pseudocode Methods for natural language processing is used in a rule-based and neural network approach to unstructured data analysis and automatic content generation. Language models like Chat-GPT, which can evaluate unorganized content and produce trustworthy articles according to the text, are one example.

IV. CONCLUSION

Undoubtedly, ChatGPT has become the most significant transformative technological development of all time in generative Open AI. Ideologically, the next question is what can be done to take ChatGPT further ahead in the game. This article focuses on this one element by advocating the integration of a rule-based approach over the neural network.

This application of hybrid approaches also opens up new research directions in NLP, as it can be used in several other technological tools. For instance, extending this hybrid approach to multi-modal NLP tasks like image captioning or video description, where visual and linguistic inputs need to be processed to generate an output, the rule-based approach can provide a structure for the output aiding the neural network to capture visual and linguistic features. Another direction that can be taken in future research is exploring other AI algorithm techniques like reinforced learning, knowledge graph, multi-modal learning, and active learning with rule-based and neural-network approaches. Although extensive research and further investigations are required to optimise this integration, the article proposes an avenue that can be pursued wholeheartedly to enrich ChatGPT's performance.

REFERENCES

[1] M. Abdullah, A. Madain, and Y. Jararweh, "ChatGPT: Fundamentals, Applications and Social Impacts," in 2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS), Nov. 2022, pp. 1–8. doi: 10.1109/SNAMS58071.2022.10062688.

[2] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable AI: A Review of Machine Learning Interpretability Methods," Entropy, vol. 23, no. 1, Art. no. 1, Jan. 2021, doi: 10.3390/e23010018.

[3] M. T. Mutiwokuziva, M. W. Chanda, P. Kadebu, A. Mukwazvure, and T. T. Gotora, "A neural-network based chat bot," in 2017 2nd International Conference on Communication and Electronics Systems (ICCES), Oct. 2017, pp. 212–217. doi: 10.1109/CESYS.2017.8321268.

[4] D. Richards, "Two decades of Ripple Down Rules research," Knowl. Eng. Rev., vol. 24, no. 2, pp. 159–184, Jun. 2009, doi: 10.1017/S0269888909000241.

[5] E. M. Bender, "Linguistic typology in natural language processing," Linguist. Typology, vol. 20, no. 3, pp. 645–660, Dec. 2016, doi: 10.1515/lingty-2016-0035.

[6] "Implementation of interpretable methods for paraphrasing and text disambiguation." http://odr.chalmers.se/items/d162b84a-7852-4c35-9c37-987666c466f9 (accessed May 14, 2023).

[7] M. Pietikäinen and O. Silven, "Challenges of Artificial Intelligence – From Machine Learning and Computer Vision to Emotional Intelligence." arXiv, Jan. 05, 2022. doi: 10.48550/arXiv.2201.01466.

[8] "Full article: Survey on frontiers of language and robotics." https://www.tandfonline.com/doi/full/10.1080/01691864.2019.1632223 (accessed May 14, 2023).

[9] P. Ntylgeris, "Unsupervised information extraction and linking to open data," Apr. 2023, Accessed: May 14, 2023. [Online]. Available: https://repository.ihu.edu.gr//xmlui/handle/11544/30193

[10] "Intrusion Detection in Industrial Internet of Things Network-Based on Deep Learning Model with Rule-Based Feature Selection." https://www.hindawi.com/journals/wcmc/2021/7154587/ (accessed May 14, 2023).

[11] L. Bambaci, "Critical Apparatus as Domain-Specific Languages: A Rule-based Parser for Encoding an Eighteenth-Century Collation of Hebrew Manuscripts," Int. J. Inf. Sci. Technol., pp. 23-34 Pages, Sep. 2021, doi: 10.57675/IMIST.PRSM/IJIST-V5I1.194.

[12] "Explaining black box models by means of local rules | Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing." https://dl.acm.org/doi/abs/10.1145/3297280.3297328 (accessed May 14, 2023).

[13] "AI-Assisted Decision-making in Healthcare | SpringerLink." https://link.springer.com/article/10.1007/s41649-019-00096-0 (accessed May 14, 2023).

[14] "Beauty Is in the A.I. of the Beholder: Artificial and Superintelligence | SpringerLink." https://link.springer.com/chapter/10.1007/978-3-030-16920-6_6 (accessed May 14, 2023).

[15] "Full article: Natural language processing and machine learning to enable automatic extraction and classification of patients' smoking status from electronic medical records." https://www.tandfonline.com/doi/full/10.1080/03009734.2020.1792010 (accessed May 14, 2023).

[16] "A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence - Michael Haenlein, Andreas Kaplan, 2019." https://journals.sagepub.com/doi/abs/10.1177/0008125619864925?journalCode=cmra (accessed May 14, 2023).

[17] G. Alexandridis, I. Varlamis, K. Korovesis, G. Caridakis, and P. Tsantilas, "A Survey on Sentiment Analysis and Opinion Mining in Greek Social Media," Information, vol. 12, no. 8, Art. no. 8, Aug. 2021, doi: 10.3390/info12080331.

[18] P. Smolensky, R. McCoy, R. Fernandez, M. Goldrick, and J. Gao, "Neurocompositional Computing: From the Central Paradox of Cognition to a New Generation of AI Systems," AI Mag., vol. 43, no. 3, Art. no. 3, Sep. 2022, doi: 10.1002/aaai.12065.

[19] Jang, M., Seo, S., Kang, P. (2019). Recurrent neural network-based semantic variational autoencoder for sequence-to-sequence learning. Information Sciences, 490, 59-73..

[20] J. Mutinda, W. Mwangi, and G. Okeyo, "Sentiment Analysis of Text Reviews Using Lexicon-Enhanced Bert Embedding (LeBERT) Model with Convolutional Neural Network," Appl. Sci., vol. 13, no. 3, Art. no. 3, Jan. 2023, doi: 10.3390/app13031445.

[21] Eskandari, H., Imani, M., Moghaddam, M. P. (2021). Convolutional and recurrent neural network based model for short-term load forecasting. Electric Power Systems Research, 195, 107173.

[22] S. Baek, W. Jung, and S. H. Han, "A critical review of text-based research in construction: Data source, analysis method, and implications," Autom. Constr., vol. 132, p. 103915, Dec. 2021, doi: 10.1016/j.autcon.2021.103915.

[23] "Placing language in an integrated understanding system: Next steps toward human-level performance in neural language models | PNAS." https://www.pnas.org/doi/abs/10.1073/pnas.1910416117 (accessed May 14, 2023).

[24] V. Sorin, Y. Barash, E. Konen, and E. Klang, "Deep Learning for Natural Language Processing in Radiology—Fundamentals and a Systematic Review," J. Am. Coll. Radiol., vol. 17, no. 5, pp. 639–648, May 2020, doi: 10.1016/j.jacr.2019.12.026.

[25] C. Alt, M. Hübner, and L. Hennig, "Fine-tuning Pre-Trained Transformer Language Models to Distantly Supervised Relation Extraction." arXiv, Jun. 19, 2019. doi: 10.48550/arXiv.1906.08646.

[26] Liu, G., & Guo, J. (2019). Bidirectional LSTM with attention mechanism and convolutional layer for text classification. Neurocomputing, 337, 325-338..

[27] "JMIR Medical Informatics - Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review." https://medinform.jmir.org/2019/2/e12239/ (accessed May 15, 2023).

[28] "Recent advances and applications of deep learning methods in materials science | npj Computational Materials." https://www.nature.com/articles/s41524-022-00734-6 (accessed May 15, 2023).

[29] I. Banerjee et al., "Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification," Artif. Intell. Med., vol. 97, pp. 79–88, Jun. 2019, doi: 10.1016/j.artmed.2018.11.004.

[30] "A survey of state-of-the-art approaches for emotion recognition in text | SpringerLink." https://link.springer.com/article/10.1007/s10115-020-01449-0 (accessed May 15, 2023).

[31] K. Kann et al., "AmericasNLI: Machine translation and natural language inference systems for Indigenous languages of the Americas," Front. Artif. Intell., vol. 5, 2022, Accessed: May 15, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frai.2022.995667

[32] "Editable machine learning models? A rule-based framework for user studies of explainability | SpringerLink." https://link.springer.com/article/10.1007/s11634-020-00419-2 (accessed May 15, 2023).

[33] R. Xie, Y. Lu, F. Lin, and L. Lin, "FAQ-Based Question Answering via Knowledge Anchors," in Natural Language Processing and Chinese Computing, X. Zhu, M. Zhang, Y. Hong, and R. He, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020, pp. 3–15. doi: 10.1007/978-3-030-60450-9_1.

[34] I. H. Sarker, "Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions," SN Comput. Sci., vol. 2, no. 6, p. 420, Aug. 2021, doi: 10.1007/s42979-021-00815-1.

[35] "Graph Neural Networks in Natural Language Processing | SpringerLink." https://link.springer.com/chapter/10.1007/978-981-16-6054-2_21 (accessed May 15, 2023).

[36] M. Arora and V. Kansal, "Character level embedding with deep convolutional neural network for text normalisation of unstructured data for Twitter sentiment analysis," Soc. Netw. Anal. Min., vol. 9, no. 1, p. 12, Mar. 2019, doi: 10.1007/s13278-019-0557-y.

[37] R. Sharma, S. Morwal, B. Agarwal, R. Chandra, and M. S. Khan, "A deep neural network-based model for named entity recognition for Hindi language," Neural Comput. Appl., vol. 32, no. 20, pp. 16191–16203, Oct. 2020, doi: 10.1007/s00521-020-04881-z.

[38] "Natural language understanding approaches based on joint task of intent detection and slot filling for IoT voice interaction | SpringerLink." https://link.springer.com/article/10.1007/s00521-020-04805-x (accessed May 15, 2023).

[39] R. Dias and A. Torkamani, "Artificial intelligence in clinical and genomic diagnostics," Genome Med., vol. 11, no. 1, p. 70, Nov. 2019, doi: 10.1186/s13073-019-0689-8.

[40] "IEEE Xplore - Page not Found." https://s3-us-west-2.amazonaws.com/ieeeshutpages/xplore/xplore-shut-page.html (accessed May 15, 2023).

[41] "Natural language processing: state of the art, current trends and challenges | SpringerLink." https://link.springer.com/article/10.1007/s11042-022-13428-4 (accessed May 15, 2023).

[42] "From theories on styles to their transfer in text: Bridging the gap with a hierarchical survey | Natural Language Engineering | Cambridge Core." https://www.cambridge.org/core/journals/natural-language-engineering/article/from-theories-on-styles-to-their-transfer-in-text-bridging-the-gap-with-a-hierarchical-survey/C2C386DEEB83B6281280ACD899AA1A3F (accessed May 15, 2023).

[43] W. Kryściński, B. McCann, C. Xiong, and R. Socher, "Evaluating the Factual Consistency of Abstractive Text Summarization." arXiv, Oct. 28, 2019. doi: 10.48550/arXiv.1910.12840.

[44] B. Zhang and R. Sennrich, "Root Mean Square Layer Normalization," in Advances in Neural Information Processing Systems, Curran Associates, Inc., 2019. Accessed: May 15, 2023. [Online]. Available: https://proceedings.neurips.cc/paper/2019/hash/1e8a19426224ca89e83cef47f1e7f53b-Abstract.html

[45] "Rule Based Approach in NLP - GeeksforGeeks." https://www.geeksforgeeks.org/rule-based-approach-in-nlp/ (accessed Jun. 23, 2023).