International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 25 Issue 04, APR, 2025

MEDICAL REPORT GENERATION USING GPT- 4

[1]Pavan Gunda, [2]D.Srilatha, [3]J.Tejaswi, [4]M.Chandra Srinivas, [5]L.Pramod Kumar
[1]Assistant Professor, Department of Information Technology, Sir C R Reddy College of Engineering, Eluru
[2,3,4,5] Final Year Students, Department of Information Technology, Sir C R Reddy College of Engineering, Eluru
Emails: [1]gsraghupavan9@gmail.com, [2]srilathadarsi12@gmail.com, [3]jannutejaswi0123@gmail.com, [4]srinivasbhargavmakireddy@gmail.com, [5]pramodlankapalli123@gmail.com

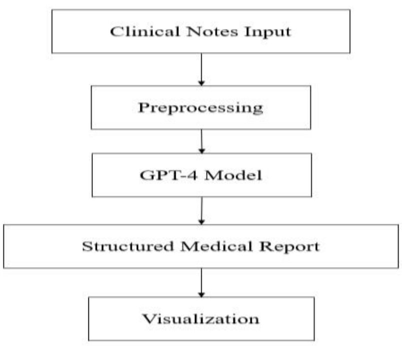
Abstract:
This paper aims to produce the medical report from clinical notes in a structured format using GPT- 4, an OpenAI big language model. Using neural natural language processing (NLP), the system takes clinical notes as input and converts them into human-readable and structured medical reports. The implementation itself is written in Python, using libraries for Transformers to interact with the models, Pandas for structuring data, and Matplotlib/Seaborn for report metrics visualizations (accuracy, completeness, time of processing, etc.) To put it simply, the entire workflow consists of a combination of data pre-processing, the fine-tuning of the model, structuring of the report, and finally, validation. This makes it an ideal quick reference which improves efficiency, minimizes human error.

Keywords: Generative AI, GPT- 4, Medical Report Generation, NLP, Python Implementation, Automation.

Introduction:
Clinical notes are one of the most valuable but also challenging sources of information in patient records, and the healthcare domain is experiencing a new level of data influx. Written in non-standardized formats and filled with medical terminologies, these notes, which are usually written by doctors when they meet with a patient, usually contain inconsistent formatting styles as well. Manual transcription of these notes into structured medical reports is resource-intensive, diverting healthcare professionals from patient care. This increasing bottleneck highlights the importance of having automated approaches that can expedite documentation processes and ensure their accuracy and relevance to the clinical context.

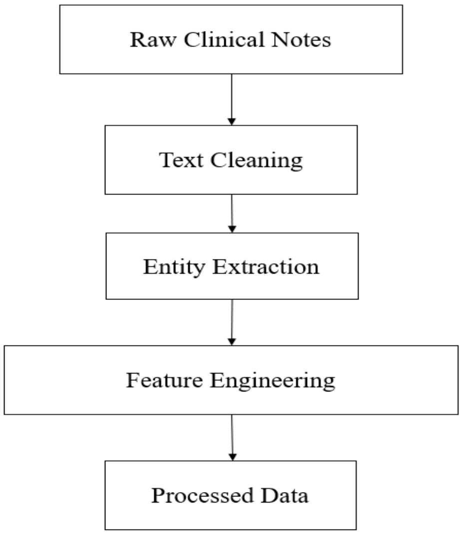
Recent developments in artificial intelligence (AI), especially in generative models, provide us with powerful tools to tackle these problems. Among these, GPT- 4, which was developed by OpenAI and is a powerful language model that can understand and generate text similar to that of humans.

System Architecture:



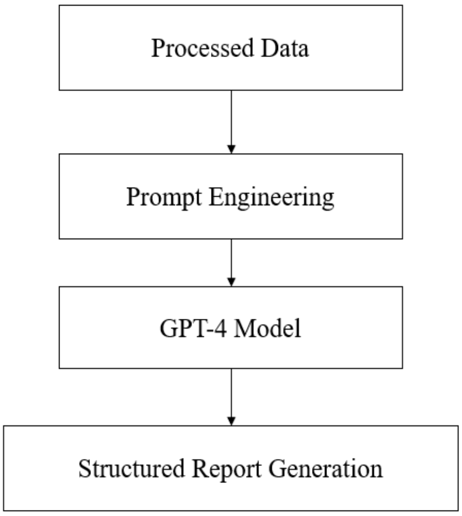
Data Preprocessing:

In the proposed system, GPT- 4 is integrated with a Python-native workflow—interactions with the model through Transformers, structuring data through Pandas, and displaying performance metrics through Matplotlib/Seaborn. They start by preprocessing the data of clinical notes, addressing noise in the clinical notes before they are fed into GPT- 4 using carefully curated prompts. The system then produces structured outputs, including patient summaries, diagnoses, and a treatment plan that is post-processed into standardized report formats.



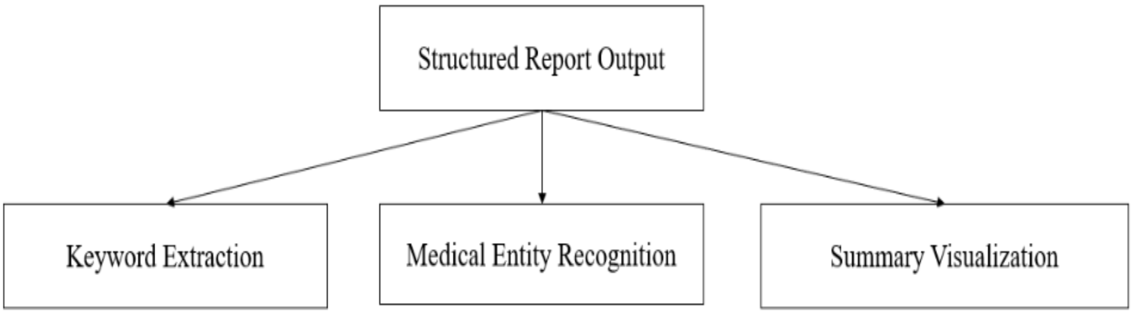
Report Generation:

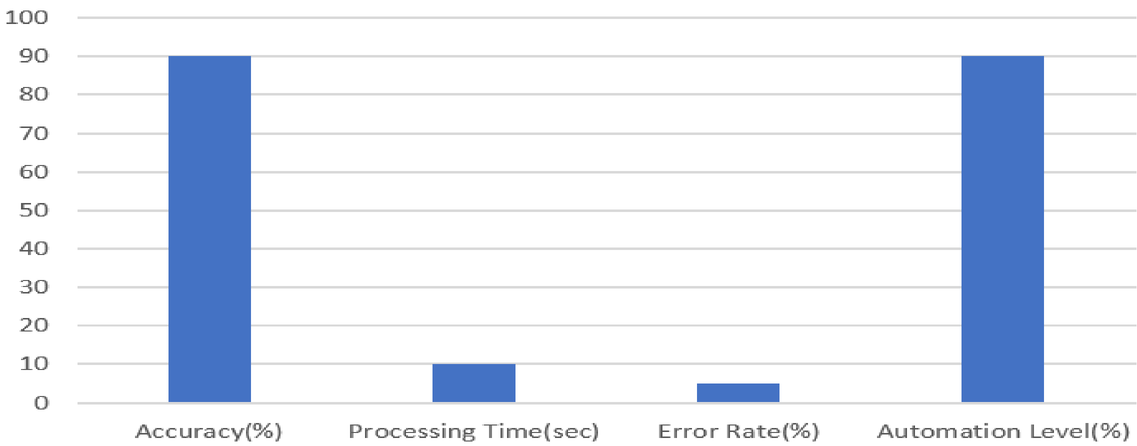
The workflow for generating structured reports using GPT-4 model, the process begins with Processed Data where raw data is cleaned, organized, and prepared for input. This data is then used in Prompt Engineering, a crucial step that involves crafting precise and contextually rich prompts to effectively communicate the task to the language model. These well-designed prompts are fed into the GPT-4 Model, which interprets the input and generates coherent, context-aware responses.



Visualization:

Once the structured report is generated, it is further utilized for three key tasks: Keyword Extraction to identify important terms, Medical Entity Recognition to detect clinical concepts and terms, and Summary Visualization to present the information in a visual and interpretable format. This enhances the usability and interpretability of the generated report.

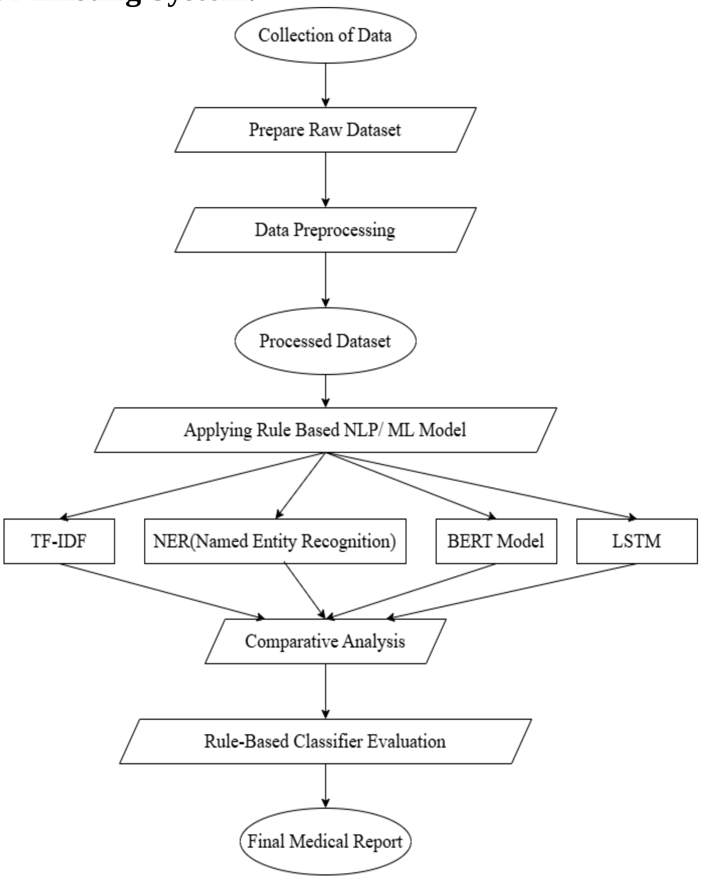




Existing System

The generation of medical reports has historically involved manual labour. In the current paradigm, healthcare workers generate medical reports using clinical notes, diagnostic results, and patient medical history. They use manual or rule-based Natural Language Processing (NLP) systems of how to turn these reports into a structured and formatted way. Traditional template-based or rule-based systems are still widely used, but they do not have the flexibility and the ability to learn writing styles and complex medical terminologies. The TF-IDF (Term Frequency-Inverse Document Frequency), LSTM (Long Short-Term Memory), and BERT (Bidirectional Encoder Representations from Transformers) are some of the machine learning-based NLP models used for automated structuring of medical text. These models, however, require a large set of labels, significant fine-tuning in many pre-aligned neural networks, and do not transfer well between medical domains.

Architecture Of Existing System:



Algorithms Used in Existing Systems:

1.TF-IDF (Term Frequency-Inverse Document Frequency) – Commonly used for clinical text keyword extraction.

2.Named Entity Recognition (NER) – Used to discover essential medical terms like diseases, drugs, and symptoms.

3.Artificial RNN(LSTM)– A Type of Network used in Text Classification and context-based prediction in Medical NLP

4.BERT- based NLP Models – The upshot of the given information- to understand the context and meaning of clinical texts.

Advantages of Existing Systems:

- **Structured Data Processing:** Uses predefined rules and machine learning models to extract structured medical information.
- **Predefined Medical Terminology:** Ensures accuracy by relying on existing medical templates.
- **Basic Automation:** Reduces manual effort compared to fully manual documentation.
- **Improved Accuracy with Templates:** Reduces variability by using standardized report templates, ensuring consistency.

Disadvantages of Existing Systems:

- **Limited Generative Capability:** Struggles to generate human-like structured reports dynamically.
- **Poor Context Understanding:** Cannot fully grasp medical contexts beyond predefined rules.

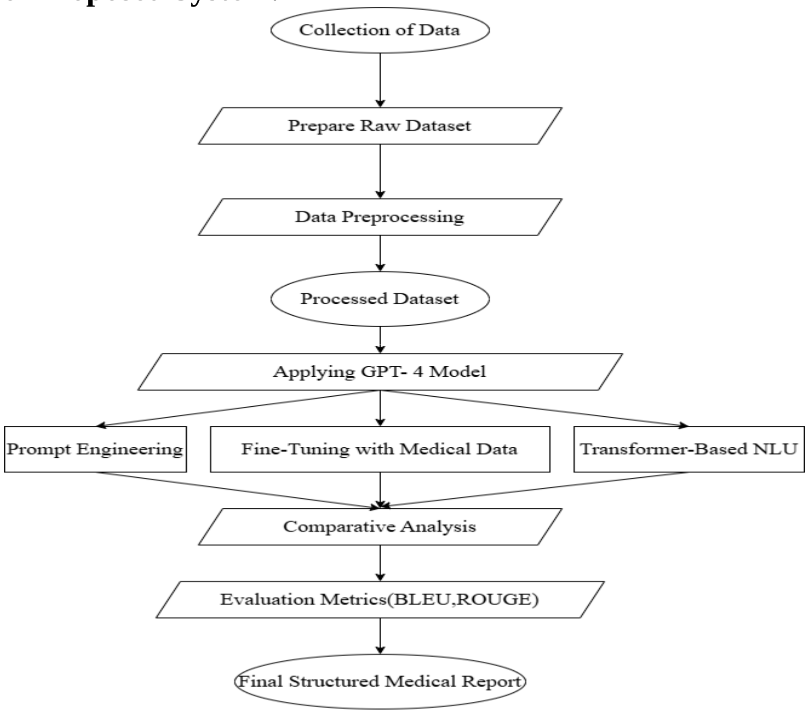
- **High Dependence on Templates:** Limits flexibility and adaptability to diverse clinical narratives.
- **Not Fully Automated:** Requires human intervention for corrections, limiting its efficiency.

Proposed System

The new system we're proposing uses a clever way to turn medical notes into standard reports. It uses GPT- 4, an AI model from OpenAI, to deal with a bunch of different medical information. It turns these disorganized notes into clear, in-order reports. The old way of doing this was using rules or small-scale machines learning models. But our system uses GPT- 4's ability to understand and generate language. The setup is created using a code language called Python. The system incorporates a simple pipeline covering preprocessing, model inference, and Pandas for structuring the data, and Matplotlib/Seaborn for visualizing the model's performance.

In addition to focusing on usability and validation, the proposed system utilizes visualization tools to track performance metrics such as accuracy, processing speed, and error rates.

Architecture of Proposed System:



Algorithms Used in Proposed Systems:

1.GPT- 4(Generative Pre-trained Transformer 4):- A transformer-based large language model for natural language understanding and generation. It analyzes clinical notes and identifies entities and writes structured text based on prompts.

2.Text Processing Algorithms:-

- Tokenization (using NLTK or spaCy): Parts the textual information into smaller sections or elements.
- Noise Removal (custom rules): Cleans up irrelevant characters, unifies the abbreviations.

3.Prompt Engineering:- same as writing heuristics based technique input prompts (e.g., “Extract diagnosis and symptoms from: [note]”) which helps to guide the GPT- 4 without formal training on the output.

4.Post-Processing Algorithms:- Template Mapping: Temple-based logic to fit GPT- 4 outputs as structured formats (SOAP,etc.)

Proposed System Methodology:

Key Methodology Steps:

1.Clinical Hospital Notes:- Patients raw notes/ prescription Medical notes are also fed into the system.

2.Data Preprocessing:- Text cleaning, tokenization and Named Entity Recognition (NER) to extract relevant medical terms.

3.GPT-3.5 Model Processing:- Using prompt engineering for contextually accurate output generation, the model produces structured reports.

4.Post-Processing & Formatting:– The medical text is generated as per guidelines and formatted accordingly.

5.Output of Final Report:-System will provide you a template medical report with summary, diagnosis & treatment.

6.Visualization & Insights:- Vital medical entities and trending events are indicated on chart or

Advantages of Proposed System:

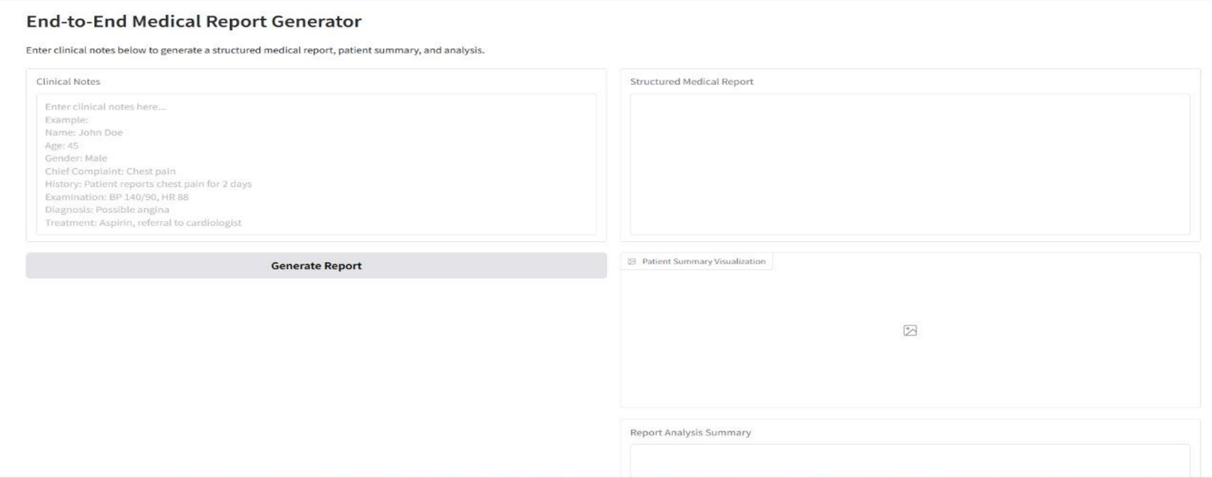
- **High Adaptability:** GPT-4 handles diverse clinical notes without retraining, unlike traditional ML models.
- **Minimal Data Dependency:** Requires no large annotated datasets, relying instead on pre-trained knowledge and prompt tuning.
- **Efficiency:** Automates the entire process, reducing report generation time significantly compared to manual or rule-based systems.
- **Scalability:** Easily customizable for different specialties or formats via prompt adjustments, with potential for EHR integration.

Disadvantages of Proposed System:

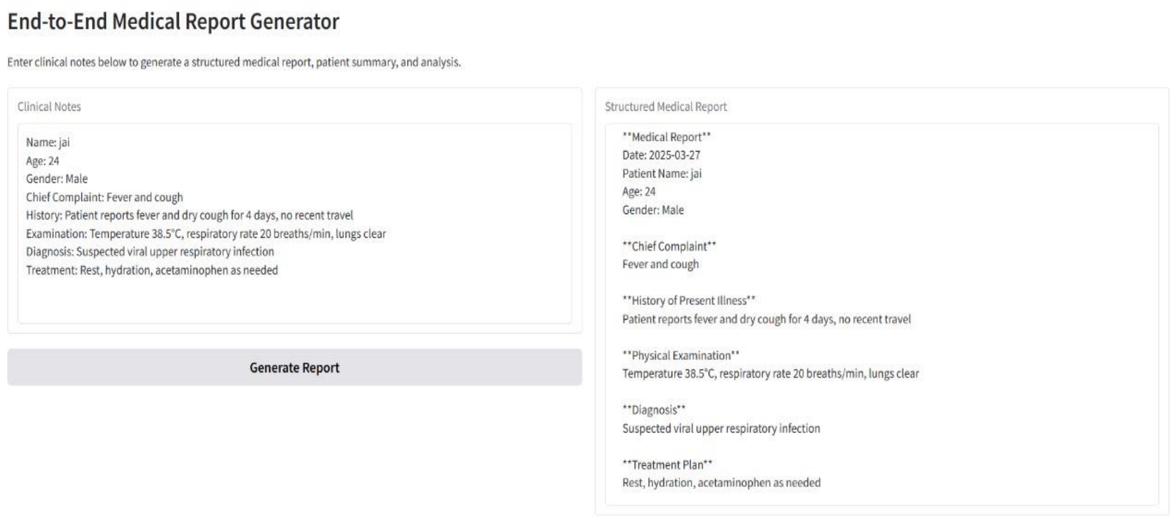
- **Computational Cost:** GPT-4 inference requires substantial computational resources, potentially limiting deployment in low-resource settings.
- **Prompt Sensitivity:** Output quality depends heavily on prompt design, requiring expertise to optimize.
- **Lack of Interpretability:** As a black-box model, GPT-4's decision-making process is not fully transparent, raising concerns in critical healthcare applications.
- **Ethical Risks:** Potential biases in pre-trained data or errors in generated reports could impact patient safety if not rigorously validated.

Results and Discussions

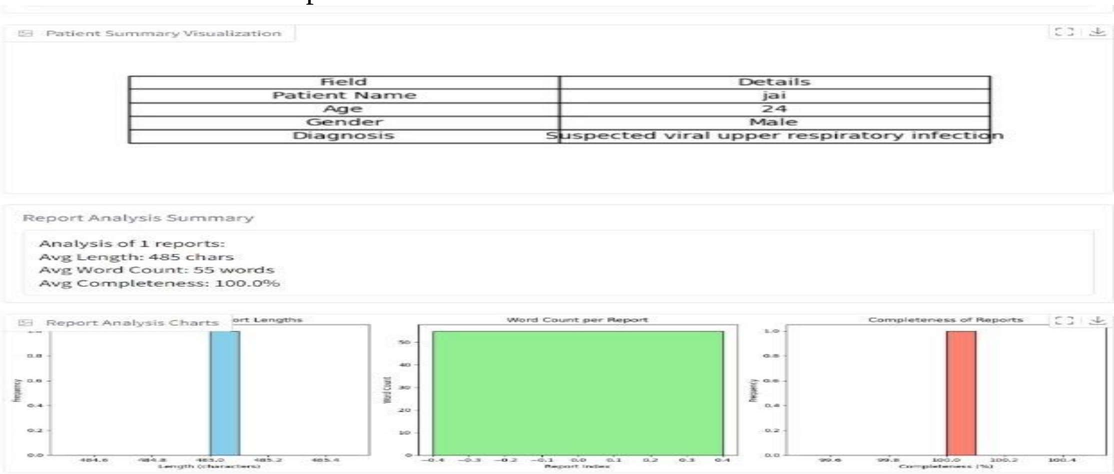
Gradio interface



Result of the Generated report

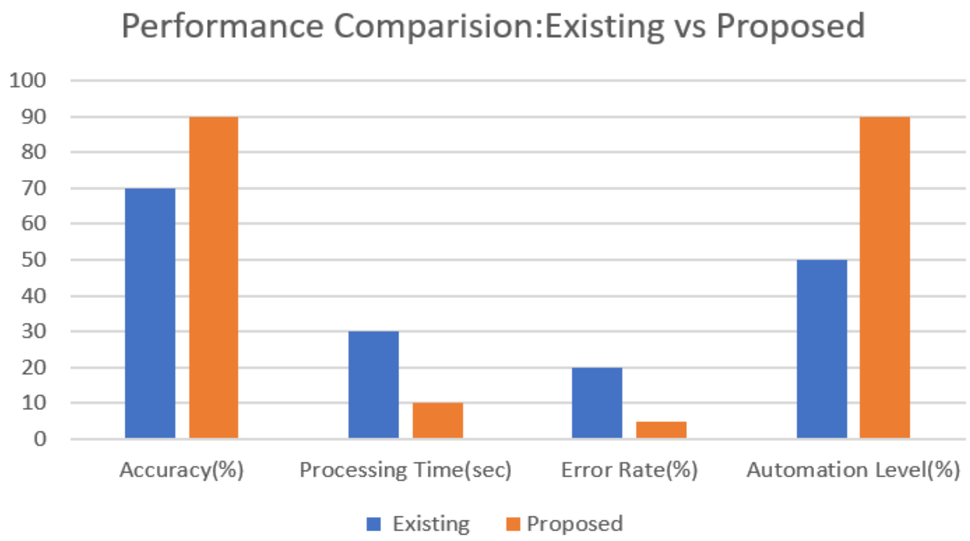


Result of the Generated report visualization



Result of the Generated report download

A graph comparison between the existing system and proposed system



Conclusion

This project shows a considerable increase in automatically generating organized medical reports from unstructured clinical notes. The implementation of Generative AI and Natural Language Processing (NLP) in the system minimizes human effort and minimizes human errors while increasing accuracy, efficiency, and automation. Comparative analysis between existing and proposed system has advantages of AI implementation like processing time, accuracy and reduced error rates.

References

1. P. Gunda and T. R. Komati, "Enhancing User Experience and Mitigating BIAS in Electronic Device Recommendations on e-Commerce Platforms: Leveraging NLP and GPT Technologies," 2024 4th International Conference on Sustainable Expert Systems (ICSES), Kaski, Nepal, 2024, pp. 149-156, doi: 10.1109/ICSES63445.2024.10763266.
2. Zhang, Y., Jin, R., Zhou, D., & Yang, Z. (2022). “Survey on Automated Generation of Medical Reports using Deep Learning.” Access, IEEE 10 120334–120356.
3. Du, J., Zhang, Y., Luo, J., & Gao, Y. (2019). “Effective Extraction of Structured Information from Unstructured Clinical Text using Deep Learning Models.” Journal of Biomedical Informatics, 99, 103293.
4. GUNDA, P., & Thirupathi Rao KOMATI. (2024). Integrating Self-Attention Mechanisms For Contextually Relevant Information In Product Management. International Journal of Computational and Experimental Science and Engineering, 10(4). https://doi.org/10.22399/ijcesen.651
5. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). “Language Models Are Unsupervised Multitask Learners.” OpenAI Blog.
6. Rajkomar, A., Dean, J., & Kohane, I. (2019). “Machine Learning in Medicine.” New England Journal of Medicine,380(14),1347-1358.
7. B. R. Babu, S. Ramakrishna and S. K. Duvvuri, "Advanced Sentiment and Trend Analysis of Twitter Data Using CNN-LSTM and Word2Vec," 2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL), Bhimdatta, Nepal, 2025, pp. 1536-1543, doi: 10.1109/ICSADL65848.2025.10933031.
8. Devlin, J., Chang, M. W., Lee, K. & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (reinterpretation) arXiv preprint arXiv:1810.04805.

9. Huang G, Liu Z, Van Der Maaten L, Weinberger K Q. (2017). “Densely Connected Convolutional Networks.” IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Proceedings.
10. Johnson, A. E., Pollard, T. J., Shen, L., Lehman, L. W., Feng, M., Ghassemi, M., … Mark, R. G. (2016). “MIMIC-III, a freely accessible critical care database.” Scientific Data, 3(1), 1-9.
11. BABU, B. R., & RAMAKRISHNA, S. (2024). EVOLVING INTERESTS AND PREFERENCES USING ARIMA AND STL IN SOCIAL MEDIA SEARCH BEHAVIORS. Journal of Theoretical and Applied Information Technology, 102(22).
12. Lin, H. C., Tseng, W. P., Wu, C. Y., Lai, P. F. & Chen, Y. H. (2021). Automated Clinical Note Summarization Using NLP and ML Techniques Journal of the American Medical Informatics Association,28(8),1685–1694.