iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** XII **Month of publication:** December 2025

DOI: https://doi.org/10.22214/ijraset.2025.76052

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

A Study on Natural Language Processing Techniques for Intelligent Text Generation

Archana Sharma[1], Shaurya Giri[2], Srishti Pandey[3], Pragati Pandey[4], Mr. Shrawan Kumar Pandey[5]
[1, 2, 3, 4]*Students,*[5] *Assistant Professor, Department of Computer Science and Engineering, Buddha Institute of Technology, Gorakhpur, India*

Abstract: ***Natural Language Processing (NLP) is crucial for machines to understand, interpret, and generate human language. This paper presents an overview ofNLP techniquesincluding traditional techniques and its modern advancements. Traditional NLP techniques are the foundation of language understanding which is further enhanced by the modern NLP techniques. The paper also gives the applications of NLP across various fields. In addition, an experiment using Google's Gemini AIillustratesthe generation of insights, coherent, and context-based text for career growth. The study shows the evolution of NLP and its growing potential in developing intelligent and human-centric AI application.***

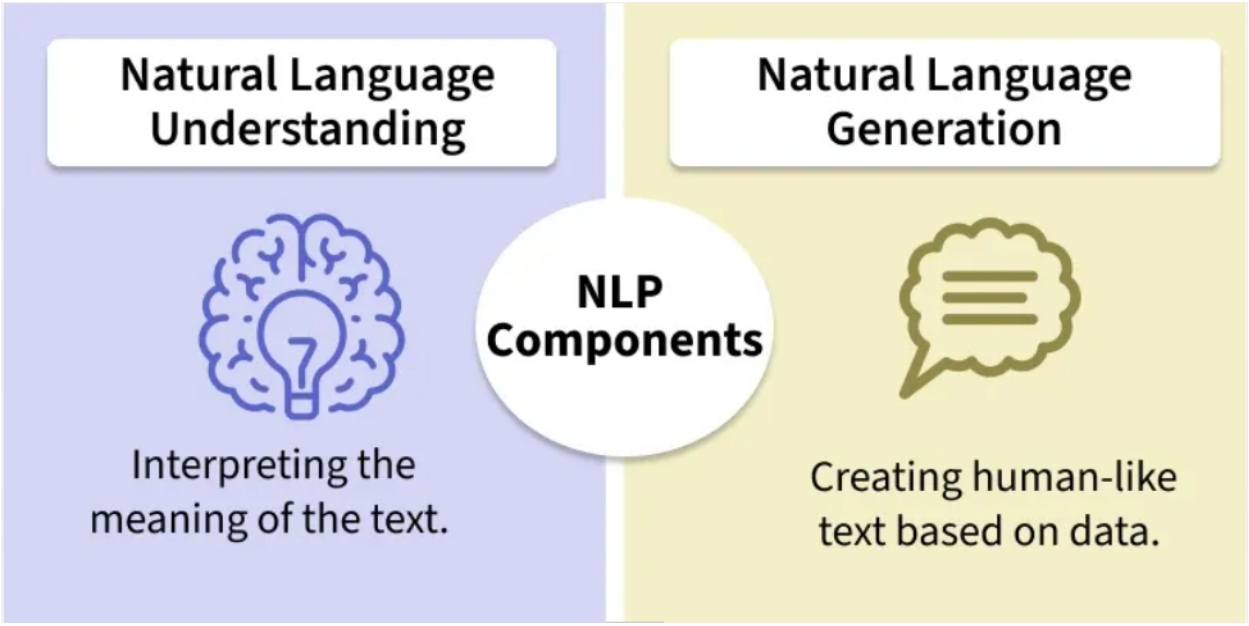
Keywords: Natural Language Processing, Text Generation, Gemini AI, Automated Systems, Transformer Models, Contextual Embedding.

I. INTRODUCTION

Natural Language Processing is core domain of Artificial Intelligence that enables computers to understand, generate and manipulate human language. Text generation, which is also referred as natural language generation has become one of the most important tasks of NLP. It has become very popular in recent years for its application across various fields such as machine translation, content creation and automated systems. NLP systems can now analyze sentiments, summarize documents, extract information, answer questions and generate human like texts.

Despite the impressive progress of neural models for text generation, their effectiveness is often restricted by large size of available datasets. The existing datasets are considered to be very small for many supervised text generation tasks, causing deep models to frequently overfit and struggle to generalize well when trained on limited data.

The evolution of large language models has further transformed NLP by enabling greater adaptability and reasoning across various tasks. Hence NLP has become a core technology in many applications such as automated text generation, information retrieval, conversational agents and decision-making systems. This paper explores the techniques, principles, applications, and advancements in NLP in making intelligent and automated AI systems.



II. LITERATURE REVIEW

A. History and Evolution of NLP

In recent decades has seen a remarkable progress, moving from traditional rule-based systems to statistical modelling approaches, and now to highly advanced deep learning and transformer-based models. Initially, NLP relied heavily upon handcrafted linguistic rules. Because of this reliance, it is difficult for systems built using handcrafted linguistic rules to scale, grow, or adapt. Following the use of handcrafted rules were statistical methods, such as n-gram and probabilistic algorithms. While statistical methods improved the ability to create language models, they also generated a number of challenges with regard to the use of long-range dependencies within language. The introduction of neural networks (namely, the Recurrent Neural Networks and Long Short-Term Memory) led to a paradigm shift in the way NLP systems were created; as a result, neural networks improved an NLP model's ability to interpret the context of a sentence. The creation of transformer architectures created an even greater opportunity for researchers in NLP to create large scale language generation models such as BERT, GPT, and Gemini, and have them create very realistic, contextually-relevant sentences.

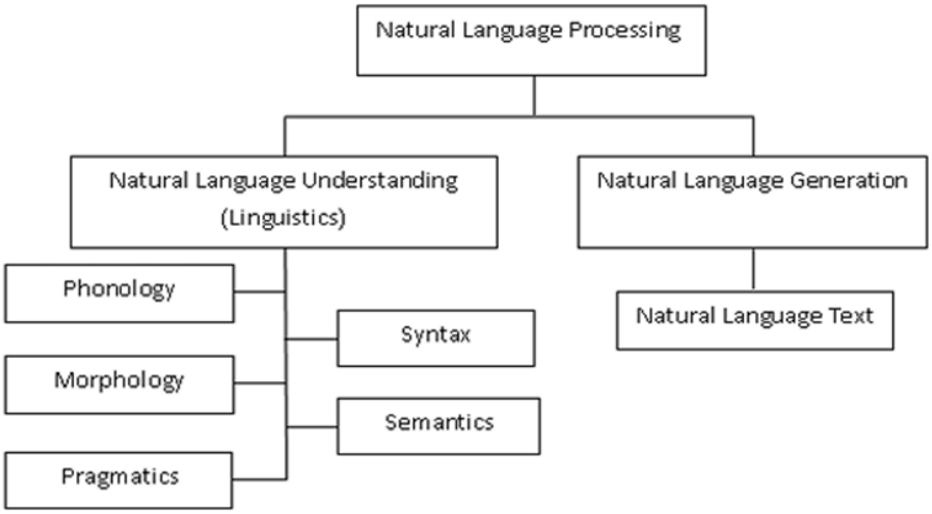
Numerous studies have examined how to improve existing NLP models for the purpose of text generation. A review of the various models used to create pre-trained text generation methodologies illustrates that large language models have effectively replaced traditional feature engineering with contextual embeddings. Studies have been conducted on how to generate controlled output by developing set prompts for each of the individual components of a model; the results of these studies have demonstrated that by creating transformer-based models and then conditioning their output by creating distinct structures to guide the development of text generated using transformer-based architectures, models received better coherence and improved relevance when creating their outputs.

In addition to these advancements, though a lot of current studies (e.g., research studies) indicate that various issues remain, for example deep models may suffer from overfitting because most supervised-produced text generation datasets are relatively small. There are also problems with language ambiguity, bias and lack of domain adaptation for NLP models created from pre-trained transformer models. Due to these issues, transformer-based and large language model (LLM) methods that are more effective at addressing many of the most complex text generation problems across multiple domains are becoming increasingly popular within the NLP community.

Overall, existing literature indicates a clear shift in NLP architectures from rule-based systems to transformer-based and data-driven approaches. This paper will provide a foundational knowledge base that can be used to analyze advanced models like Gemini within current use cases, as well as highlight how today's advances in NLP technologies facilitate intelligent, contextual text generation.

B. Classification of NLP

NLP can be classified into two main parts i.e. Natural Language Understanding and Natural Language Generation which is used to understand and generate human language.



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

Natural Language Understanding (NLU) is the ability of machine to understand what users say. NLUinvolves extracting unstructured text and converting it into structural informationwhich can be easily processed.It supports various application such asspeech recognition systems, sentiment analysis, spam classification etc.

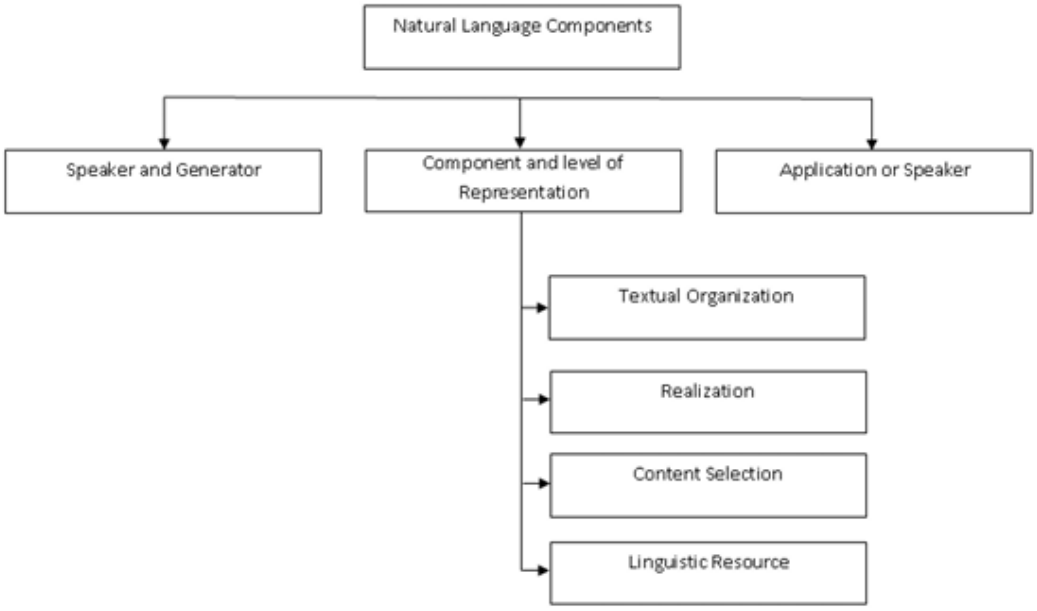
Natural Language Generation is the process of generating language the way human speak or write. It transforms data such as statistics or database entries, into coherent, contextually appropriate text or speech. Applications of NLG include Chatbots, Voice assistants etc.

III. TEXT GENERATION IN NLP

Text generation is the main task of Natural Language Generation (NLG). It is a domain in Artificial Intelligence and NLP which is used to generate contextually relevant text from the various types of data. It allowscomputers to express thoughts or observations in natural language.

NLP includes the wide range of interpreting and generating language, whereas NLG mainly focuses on the generation aspect such as turning raw data into readable text. It is widely used in automated news articles, personalized emails, financial report systems and intelligent chatbots. Text generation has several real-world applications such as machine translations, chatbots and virtual assistants and summarization systems.

There are several approaches for text generation such as rule-based text generation (traditional NLP), statistical methods, deep learning approaches, and transformer-based models.



IV. METHODOLOGY

A. *Text Processing*

Text Processing is very first step inNatural Language Processing which converts the raw data into clean and analyzable form. It consists of four steps i.e. tokenization, stop-word removal, stemming, and lemmatization.

- Tokenization is the process of dividing words into smaller words or subwords.
- Stop-words removal refers to the elimination of words which containlow information. It is used to reduce noise.
- Stemming and lemmatization are used to convert the word into their base or root forms, which helps models treat various forms of the same word consistently.

B. *Syntactic Analysis*

Syntactic analysisrefers to parsing the structure of a sentence. It helps in understanding the grammatical structure of the sentence.It involves several techniques such as POS(Parts-Of-Speech) Tagging, Dependency Parsing,and Chunking.

- POS tagging is used to assign the grammatical categories to each word which enables the structural interpretation of the words.
- Dependency parsing refers to the analysis that recognize the grammatical relationship between words in a sentence. It directly builds the links between words rather than grouping them into phrases.
- Chunking refers to the grouping of phrases i.e. noun phrases or verb phrases.

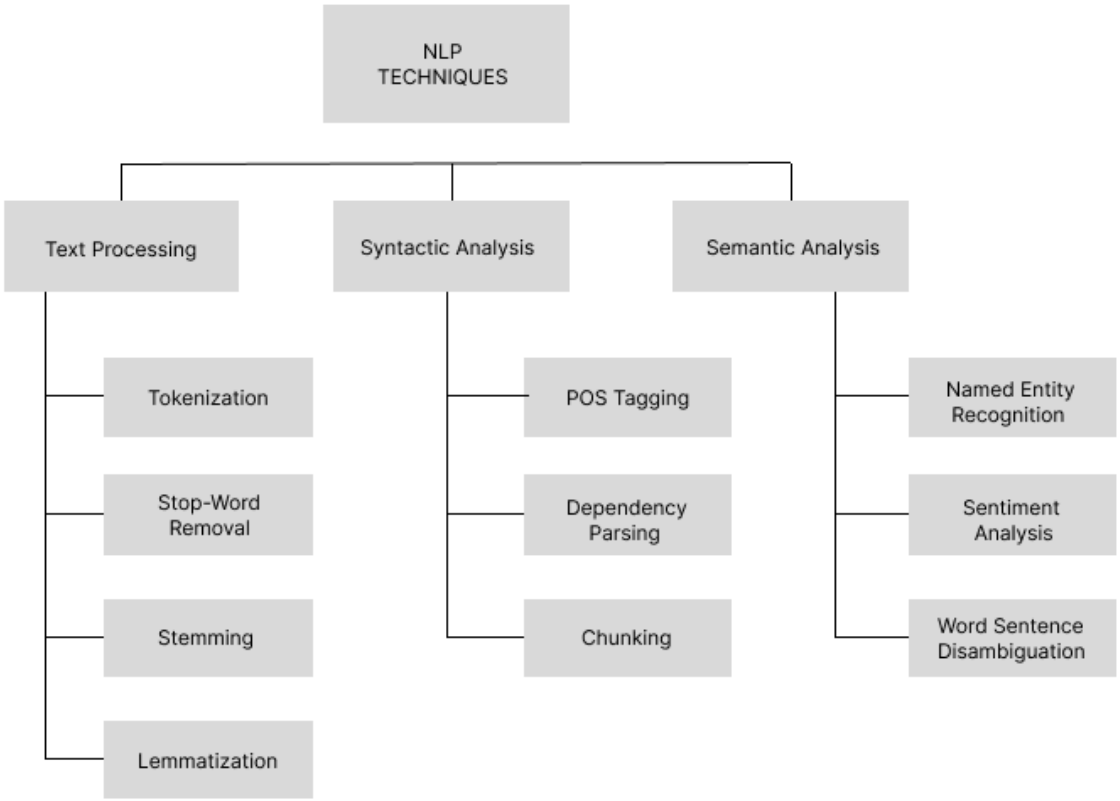
Syntactic analysis isuseful for NLP systems to understand the sentence creation, syntactic relationships and the structural dependency of texts or data.

C. Semantic Analysis

Semantic analysis is a stage in NLP that includes understanding the meaning of words in a sentence. It is used to check sense and meaning. Improvement in the question answering, machine translation, summarization, conversational chatbots and information extraction are the result of semantic analysis. It helps machines understand context rather than just focusing on the text.

Components of semantic analysis are Named Entity Recognition (NER), Sentiment Analysis and Word Sense Disambiguation (WSD).

- NER is the process of identifyingreal-world entities and classifying them into predefined categories such as person, organizations, locations etc.
- Sentiment analysis is a process in machine learning and NLP that looks for the positive or negative polarity in the text. It classifies the text according to the emotions they convey.
- WSD determine the correct meaning of a text based on its context. Most of the words have several meanings, and WSD helps identify which meaning is correct based on the context.



V. ADVANCE TECHNIQUES IN NLP

Modern NLP focuses on the advanced model for the improved accuracy and generate text based on the context. It includes more deeper and nuanced processing. Advance NLP leverage deep learning, transformers and knowledge of language to achieve human-level understanding and language generation. It enables the machines to interpret, generate and reason with human language at the human-level or near human-level.

A. Transformer Models

Transformer models use self-attention mechanism for effectively handling context of the words. It includes several NLP tasks such as translation, summarization and question answering. Some of the transformer models are BERT, T5,GPT (1-5), Gemini and Claude models. GPT generates meaningful and contextually relevant text based of the prompt provided by the user.

B. Contextual Word Embedding

Contextual Word Embeddings are the advancements in the word representation which show the meaning of words based on their context. It is useful in summarization, question answering, chatbots, machine translations and search engines. Several contextual word embedding models are ELMo, BERT, RoBERTa, ALBERT, GPT embedding, and sentence BERT (SBERT).

C. Advanced Text Generation

It is the advanced and improved version of the classical text generation which uses models such as LLaMA, Gemini, GPT models and diffusion-based text generation. Various applications of advanced text generation are story generation, email writing, chatbots and code generation.

D. Sentiment & Emotion Analysis

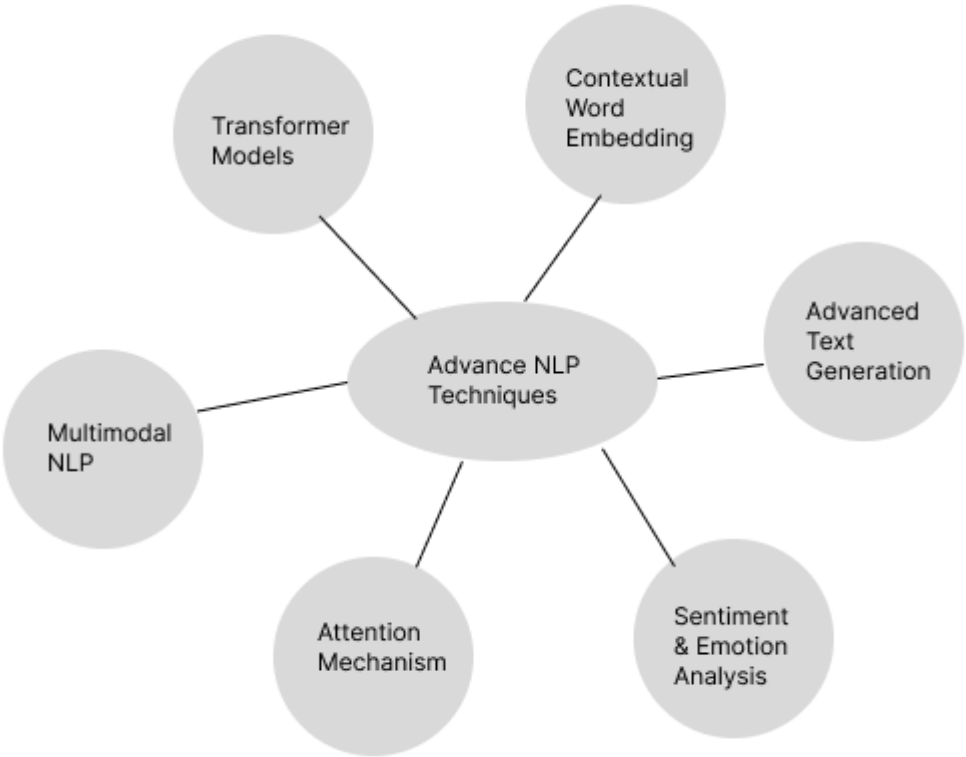
This type of analysis is the extended version of theclassical sentiment analysis technique in NLP. Rather than just analyzing negative or positive sentence, it also detects the emotions, intent, sarcasm and opinion targets.

E. Attention Mechanism

Attention Mechanism is a technique in deep learning thatenables the model to decide which input part is more suitable for a given output. Attention solves the problem oftraditional models like RNN and LSTM by allowing models to assign different and important weights to the different words. There are several types of attention mechanism such as soft attention, hard attention, self-attention and multi-head attention. Self-attention is the core part of the transformers.

F. Multimodal NLP

MultimodalNLP refers to advanced techniques or models in NLP. It includes the understanding of text, image,speech, video and audio together. It is used for video understanding, image captioning, and Visual Question Answering (VQA). Some examples of multimodal are GPT-4o, Gemini multimodal, CLIP and LLaVA.



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

VI. APPLICATIONS OF NLP

Natural Language Processinghas its applications across various areas including Machine Translation, Email Spam Detection, Summarization, Question Answering, Information Extraction etc.

A. Text Classification

Text classification is the process of categorizing text data intopredefined labels based on its context. It is also known as Text Categorization. It takes a large amount of data and assigns them to predefined categories.It helps machine learn pattern from labelled data which is then used to classify new and unseen text.Intent classification, sentiment classification and topic classification also fall in the category of text classification.

B. Machine Translation

There are many languages with different grammar and sentence structure, and this has become language barrier.Machine translation refers to the automatic conversion of text from one language to another. The challenge in machine translation is not with the translation text from one language to another but keeping meaning and context intact along with grammar and tenses. In past few years, several methods have come in picture for the automatic evaluation of machine translation quality by the comparison of hypothesis translation with the reference translation. Machine translation is useful for multi-language content creation, global communication and localization.

C. Information Extraction

Information extractionis the process of extraction or identifying structured and useful information such as places, dates names and relationships. It involves various taskssuch as NER, resume parsing, relation extraction and it has become very important over past years. Knowledge discovery is the important area of research and uses various techniques to gain or extract meaningful information from the source document. These techniques involve Parts of Speech (POS) tagging, stemming, chunking, and stop-word removal.

The extracted information can be used across various fields and is applied on several purposesuch as identify keywords,prepare summary, build databases, classify text into predefined categories etc.

D. Chatbots & Virtual Assistants

Chatbots and virtual assistants are the most important applications of NLP. These systems use NLP so that they can better understand user's query and generatecontext-based and human-like response. Some examples of conversational systems that uses NLP are Siri, Alexa, Google Assistant and customer support bots. From answering questions to helping with daily activities, they can perform various tasks.

E. Healthcare Application

NLP is widely used in medical field as well and is useful in analysing clinical reports, lab reports, doctor notes and other medical records.In addition, NLP also helps in disease identification, symptoms detectionand supporting medical decision-making from the medical reports.NLP improves accuracy, enhance care and reduce the manual workload by transforming raw clinical text into meaningful insights.

F. Career Growth

NLP is very useful for the career growth of students and working professionalsby helping them analyze skills, choosing career path and improving the professional profiles. NLP is useful for resume screening, generating resume & cover letters, interview preparation and career recommendation. By providing industry insights, it enables the user or students to understand industry trends, trending skills, areas for improvement and the salary ranges by job roles.

VII. RESULTS AND DISCUSSION

For the evaluation of practical application and effectiveness of advanced NLP techniques, a prototype was built and tested through Google's Gemini AI model. The model includes the functionality such as resume and cover-letter generation, interview question answers, based on user input. The outputs were context-based and according to the user requirements, which shows the effectiveness of LLM-based text generation. When the user provides a basic information about their background such as their skills and domain, the model generates a structured industry insight that gives the information about the industry trends, the trending and in-demand skills in that industry or domain and the salary ranges by role. This shows the ability of advanced NLP techniques to extractinformation and convert it into the meaningful insights.

These outcomes supportthe findings discussed in this paper, highlighting the significance of transformer-based models andtheir advancement over traditional NLPin accuracy, relevance, context-based text generation and contextual reasoning.

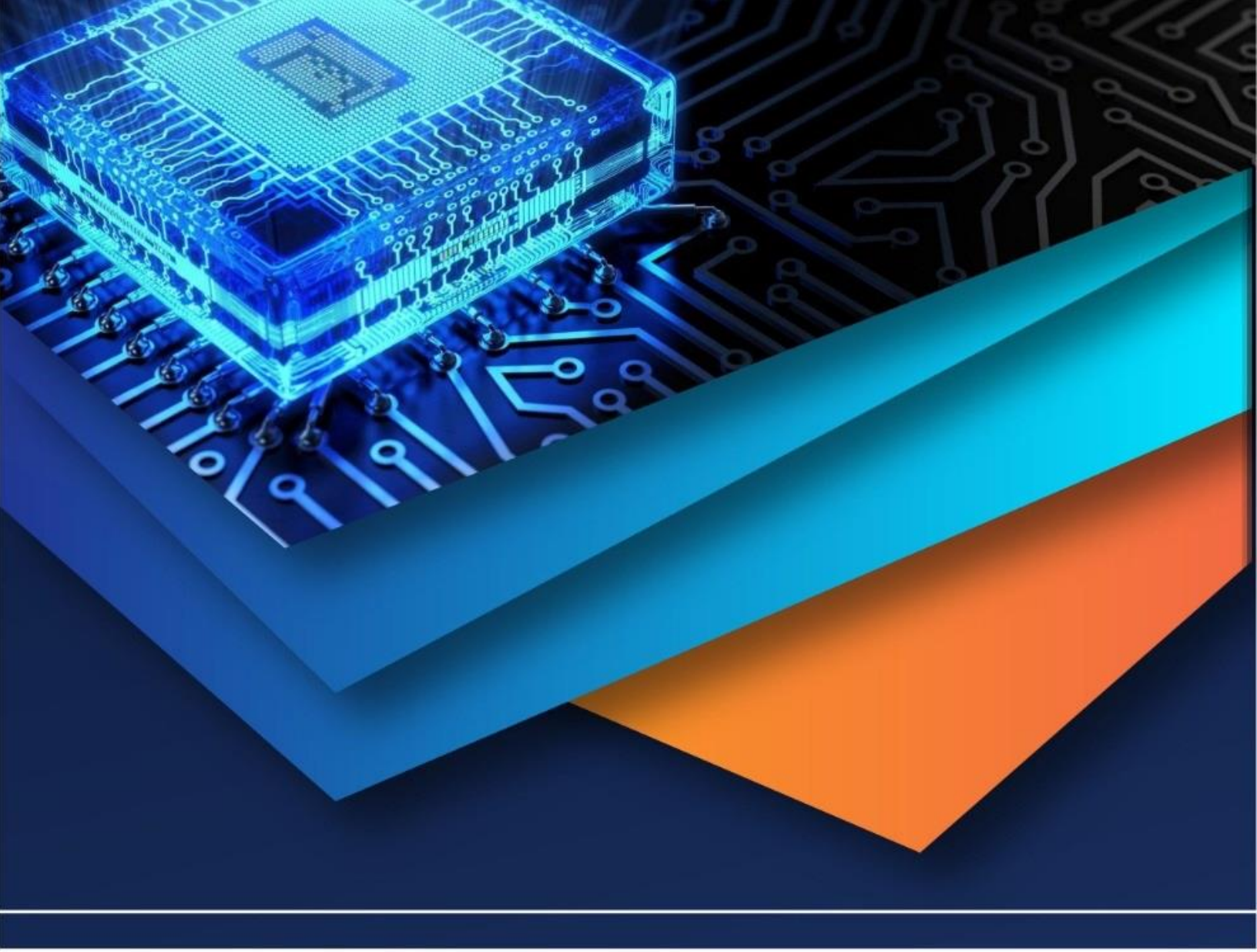
VIII. CONCLUSION

Natural Language Processing has emerged as a powerful fieldand has made tremendous stridesin enabling machinesto understand, interpret, analyse and generate human-like language with improved accuracy. This paper reviewed and provided thetraditional and foundational NLP techniques such as text processing, syntactic and semantic analysis, and their interpretation into the advancedNLP techniques such as transformer models, contextualembeddings, attention mechanism, multimodal and emotion analysis. These advancements enable the advance text generation which is coherent, contextual and more human-like.

The outcomes of the experimental observation demonstrate the effectiveness of modern NLP in handling contextual reasoning and generating meaningful outputs including resume generation and career related insights. Even though NLP has ongoing challenges like data scarcity, bias and ambiguity, the current trends show the continuous progress in creating more robust and scalable NLP systems. In summary, NLP has a crucial role in shaping intelligent systems and serves asthegroundwork for the development of more advanced human-AI interaction models.

REFERENCES

[1] D. Khurana, A. Koli, K. Khatter, and S. Singh, “Natural Language Processing: State of the Art, Current Trends and Challenges,” Proceedings of the 2017 International Conference on Machine Learning and Data Science, pp. 262–267, 2017.
[2] J. Li, T. Tang, W. X. Zhao, and J.-R. Wen, “Pretrained Language Models for Text Generation: A Survey,” arXiv preprint arXiv:2201.05273, 2022.
[3] A. Samyuktha, M. Pallavi, L. Jagan, D. Y. Srinivasulu, and M. Rakesh, “Sentiment Analysis Using Machine Learning,” International Journal for Research in Applied Science & Engineering Technology (IJRASET), vol. 11, no. 1, pp. 317–324, 2023.
[4] GeeksforGeeks, “Introduction to Natural Language Processing,” Available: https://www.geeksforgeeks.org.Accessed: Nov. 2025
[5] IBM, “What is Natural Language Processing (NLP)?” Available: https://www.ibm.com/topics/natural-language-processing.Accessed: Nov. 2025
[6] A. Vaswani et al., “Attention Is All You Need,” Advances in Neural Information Processing Systems (NeurIPS), 2017.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY