International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(23s) 2026 pp. 315-326
https://doi.org/10.70917/ijcisim-2026-5576
Article

NATURAL LANGUAGE PROCESSING: RECENT DEVELOPMENTS, TRENDS, AND DIFFICULTIES

Dr. J. Jeyshri[1], Kartikeya Tiwari[2], Srinivas V.[3], Dr. G. Saravanan[4], M. Guru Vimal Kumar[5], Abhimanyu[6]

[1]Assistant Professor, Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai – 600119, Tamil Nadu, India. Email: jeyshri.j.cse@sathyabama.ac.in
[2]Assistant Professor, Department of Computer Science and Engineering, Gautam Buddha University, Greater Noida – 201312, Uttar Pradesh, India.
[3]Assistant Professor, Department of MCA, Bangalore Institute of Technology, Bengaluru, Karnataka, India. Email: srinivas.jaya666@gmail.com
[4]Professor, Department of Artificial Intelligence and Data Science, Erode Sengunthar Engineering College, Tamil Nadu, India. Email: gsaravanan.pacet@gmail.com
[5]Associate Professor, Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai – 600062, Tamil Nadu, India. Email: guruvimal09@gmail.com
[6]Vijayanagara Sri Krishnadevaraya University, Karnataka, India. Email: iamabhimanyusharma12@gmail.com
Corresponding Author: Dr. J. Jeyshri

Abstract: Natural Language Processing (NLP) is a rapidly advancing domain within Artificial Intelligence (AI) that allows computers to comprehend, manipulate, and produce human language. This document provides a summary of the latest advancements, burgeoning trends, and significant obstacles in natural language processing. It examines the progression of NLP from rule-based frameworks to transformer architectures and Large Language Models (LLMs), as well as the uses of Natural Language Understanding (NLU) and Natural Language Generation (NLG). The document moreover examines prominent NLP development frameworks, including as Transformer, BERT, and GPT models, and underscores their contemporary uses in diverse fields. Furthermore, it explores significant obstacles like linguistic ambiguity, multilingual processing, computational intricacies, and ethical dilemmas, while delineating prospective study avenues for the advancement of more effective and dependable NLP systems.
Keywords: Artificial Intelligence, Natural Language Generation, Transformer, BERT, GPT, Large Language Models

1. INTRODUCTION

Artificial Intelligence (AI) is one of the most impactful technologies of the 21st century, shaping the way people interact with computers and intelligent systems. The goal of AI is to build robots capable of doing things that ordinarily need human intellect. These things include learning, reasoning, solving problems, sensing and decision making. The rapid growth in processing power, data and machine learning algorithms led to the widespread of use of AI in many areas such as healthcare, education, finance, manufacturing, transportation and customer services. Natural Language Processing (NLP) is one of the disciplines of AI that has attracted much interest due to its capacity to make computers comprehend, interpret, analyse and synthesise human language.

Natural Language Processing (NLP) is an interdisciplinary discipline that integrates artificial intelligence, computer science, linguistics and machine learning to allow communication between people and computers using natural language. Human language is far more complicated than formal computer languages. Ambiguity, contextdependent meanings, different grammatical forms, the variety of languages, etc. NLP seeks to solve these problems by creating computational methods that enable robots to understand written and spoken language appropriately. It consists of essentially two primary components: Natural Language Understanding (NLU) which involves understanding and extracting meaning from text or voice, and Natural Language Generation (NLG) which involves computers being capable of generating reasonable and meaningful language answers.

The relevance of NLP has expanded dramatically with the fast expansion of digital communication and unstructured textual data created by websites, social media platforms, emails, online documents and conversational systems. Modern NLP technologies are extensively utilised for machine translation, sentiment analysis, chatbots, virtual assistants, voice recognition, question-answering systems, text summarisation, spam detection, information retrieval and content development. In addition, NLP has become an important technology in many fields such as healthcare for clinical text analysis, finance for fraud detection and customer service, education for intelligent tutoring systems, e-commerce for recommendation systems, legal document analysis, and cybersecurity for threat intelligence. These applications have greatly enhanced automation, decision making and human-computer interaction.

The development of NLP has gone through numerous important stages. Most early NLP systems relied on rule-based systems, which processed language using hand-crafted linguistic rules and dictionaries. These systems worked well in narrow areas, but they lacked flexibility and scalability. Statistical methodologies and machine learning techniques allow models to understand linguistic patterns from big datasets, leading to better performance on a variety of NLP tasks. At that point, deep learning methods advanced language representations by using neural networks. The real breakthrough came with the introduction of the Transformer architecture, which dramatically increased the ability to perceive context and process in parallel. Transformer-based models, e.g., Bidirectional Encoder Representations from Transformers (BERT) and Generative Pre-trained Transformer (GPT), have attained state-of-the-art performance in several NLP applications. More recently, the advent of Large Language Models (LLMs) has revolutionised the discipline, allowing more accurate, context-aware and human-like language processing and creation.

However, NLP still has many technical and practical obstacles such linguistic ambiguity, multilingual processing, domain adaption, computational complexity, data privacy, model bias, explainability, and ethical issues. With the fast pace of the development of NLP technologies, there is an increasing demand to analyse the recent developments, future trends, current development frameworks, transformer-based models, and the challenges that affect future research. Therefore, this article provides a complete review of recent advances, trends, applications and challenges in Natural Language Processing with insights into existing technology and promising areas for future study.

2. LITERATURE REVIEW

Khurana et al. (2023) have noted that natural language processing (NLP) has lately attracted significant interest for its ability to represent and analyse human language using computer means. Its applications have proliferated across many domains like machine translation, email spam detection, information extraction, summarisation, medical areas, and question answering, among others. This study first delineates four stages by examining several tiers of NLP and elements of Natural Language Generation, thereafter outlining the historical progression and development of NLP. We subsequently examine comprehensively the cutting-edge developments, showcasing the many applications of NLP, prevailing trends, and obstacles. Ultimately, we provide an analysis of several accessible datasets, models, and assessment criteria in natural language processing.

Hirschberg et al. (2015) assert that natural language processing utilises computer methods to acquire, comprehend, and generate human linguistic material. Initial computational methodologies in linguistic study concentrated on automating the examination of language's structural components and creating fundamental technologies like machine translation, voice recognition, and speech synthesis. Contemporary scholars enhance and use these instruments in practical scenarios, including spoken conversation systems and speech-to-speech translation technologies, extracting data from social media about health or finance, and discerning mood and emotion related to goods and services. We outline achievements and obstacles in this rapidly evolving field.

Kalyanathaya et al. (2019) assert that Natural Language Processing (NLP) is a subdivision of Artificial Intelligence, now receiving significant attention in research and development owing to the rise of its applications. The primary research domains are conversational systems, language processing, machine translation, and deep learning. Investigations in these domains result in the creation of several instruments for constructing industrial applications. The integration of Deep Learning methodologies with Natural Language Processing is discovering several applications across sectors like Healthcare, Finance, Manufacturing, Education, Retail, and customer service. This document offers a comprehensive overview of progress in the research, development, and application domains of Natural Language Processing. This document delineates 21 research focal points, 22 development instruments, and 6 sectors in which Natural Language Processing is seeing fast progress. Essential terminology referenced in this document includes: NLP, Natural Language Processing, Deep Learning, Sentiment Analysis, Question Answering, Dialogue Systems, Parsing, Named-Entity Recognition, POS Tagging, Chatbots, Human-Computer Interface.

Young et al. (2018) Deep learning techniques use several processing layers to acquire hierarchical data representations, achieving cutting-edge outcomes across various fields. Recently, an array of model architectures and techniques has proliferated within the realm of natural language processing (NLP). This article examines notable deep learning models and techniques used for various NLP applications and discusses their progression. We also encapsulate, evaluate, and differentiate the many models while presenting a comprehensive insight into the historical, contemporary, and prospective aspects of deep learning in NLP.

Ghazizadeh et,al (2020) This study article conducted an extensive literature analysis to examine the applicability of Natural Language Processing (NLP) across many disciplines. Furthermore, using qualitative research, we want to examine the evolution of the present condition and the challenges of NLP technology as a pivotal component of Artificial Intelligence (AI) technology, highlighting certain limits, dangers, and potential. In this study, we use primary data derived from relevant laws and secondary public domain data sources that provide pertinent information from case analyses. Through an examination of the structure and substance of existing literature, NLP-based applications have been distinctly categorised into various domains, including natural language understanding, natural language generation, speech recognition, machine translation, spell correction, and grammar checking. The trajectory of progress, unresolved issues, and constraints have also been examined.

- **Research gap**

The literature review indicates the considerable advancement of Natural Language Processing (NLP) including its development, applications, deep learning approaches, and transformer-based models. While there has been significant discussion in previous research about various NLP tasks, development tools, and domain-specific applications, the issues of linguistic ambiguity, multilingual processing, and computational complexity have been identified. However, most of the studies are limited to particular areas of NLP such as deep learning approaches, industrial applications, or individual language models, rather than a consolidated view of recent developments, emerging trends, modern development frameworks, transformer-based architectures (Transformer, BERT, and GPT), and current research challenges in one study. Hence, a complete analysis that incorporates recent advances, development tools, applications, and existing obstacles is essential to provide a holistic picture of the present status and future prospects of Natural Language Processing.

3. OBJECTIVES OF THE STUDY

- To learn about the latest advancements in Natural Language Processing (NLP).
- To explore the latest trends and use cases of NLP in different fields.
- To recognise the key challenges and issues of NLP techniques.
- To discuss the areas of future scope and emerging directions in research of Natural Language Processing.

4. OVERVIEW OF NATURAL LANGUAGE PROCESSING

Work on NLP started around 1950 with the development of what is called as “Turing Test” and in 1957 a rule based system of syntactic structures. The development was gradual until 1990 since processing power was restricted and the systems were built on complicated sets of hand written rules and limited vocabularies. The interest in research and applications is expanding lately with the advent of machine learning and continual development in computer capacity. Recently, the primary breakthrough areas of NLP include voice recognition, language processing and dialogue systems by using deep learning methods. Though NLP is still facing lot of obstacles (such human computer interfaces) yet lot of research interests have been generated and it has led up numerous prospects for employing the approaches in robotics, automation and digital transformation.

NLP is a cross-disciplinary area that focuses on the interaction between people and computers in natural language. Its goal is to enable computers to understand, decode and synthesise human-like written or spoken statements. NLP employs sophisticated computational methods and linguistic principles to allow computers to analyse, interpret, and generate human language in a contextually appropriate and semantically rich manner. NLP is the capacity to interpret and modify human language using different computing models that analyse linguistic patterns automatically. In NLP, atomic phrases like "bad," "old," or "fantastic" are often used as the fundamental building blocks of language. These atomic phrases may be concatenated to compound terms, for example “very good movie” or “young man”. An atomic term is just a single word . A compound term is a sentence composed of many words . Words are the fundamental building blocks of human language and understanding them is crucial for any NLP work. Now, NLP has been used in various industries, such as: voice recognition, question answering system, online translation, text categorisation and so on NLP plays a significant role.

- **Recent Developments and Difficulties in Natural Language Processing**

Facilitating computer comprehension of human language presents significant difficulties. The computer comprehends not just the explicit meaning of the statement but also the implicit information it conveys. The subsequent challenges encountered by machines in natural language processing.

Initially, a significant quantity of lexemes has many interpretations in natural language processing. For instance, in the phrase "I desire an apple," the "apple" intended by an individual signifies the "Apple phone," yet the computer can read it as the fruit, illustrating that the meaning of a term can shift based on context. This exemplifies the uncertainty intrinsic to linguistic expression. To tackle this problem, artificial intelligence methodologies, including machine learning and deep learning frameworks, are used to discern and rectify textual discrepancies. Furthermore, the model's parameters may be modified to more effectively suit various linguistic settings and circumstances, hence enhancing the precision and efficacy of ambiguity identification.

Secondly, natural language processing depends on contextual comprehension, although the computer models' processing capabilities are constrained. One factor contributing to this is the volume and calibre of data. Despite the substantial size of the data sets used by contemporary NLP models, their efficacy may diminish when addressing infrequent or specialised activities. A emotion analysis model could excel with typical movie critiques but struggle with uncommon, specialised opinions, such those pertaining to professional cinematography. The long-tail problem arises when several words and phrases occur seldom in real language, and this infrequent data may be overlooked in the training set, leading to suboptimal performance of NLP models in addressing these uncommon instances. A further rationale is constrained computational resources. The NLP model requires substantial memory and computational resources to analyse millions of text segments. Consequently, using a machine with inadequate processing capabilities results in prolonged context processing times. To address this issue, augmenting the quantity of GPUs may decrease processing duration; nonetheless, this strategy may result in elevated training expenses.

Thirdly, there are linguistic variances in natural languages. In the contemporary digital landscape, a resilient NLP model capable of managing linguistic variety is essential. For several minor languages, the available data set for training is very limited, posing significant hurdles to model training. Due to the limited availability of extensive training data, models often demonstrate worse performance in lesser-spoken languages relative to more widely spoken ones. Furthermore, the fundamental disparities across languages are evident in their grammatical variations. In French, adjectives often follow nouns, but in English, they precede them. In text creation, if the computer fails to accurately comprehend certain grammatical principles, the produced text will be nonsensical or challenging to interpret.

- **The Applied Branch of Natural Language Processing**

Natural Language Processing fundamentally comprises two main divisions. One aspect of this is Natural Language Understanding (NLU), which denotes the capability of computers to comprehend and interpret human natural language. This is an essential research pathway in the field of artificial intelligence, aimed at equipping computers with the capacity to comprehend and manipulate human language, thereby enabling fluid and effortless communication with people. Another domain is Natural Language Generation (NLG), which primarily focuses on producing human-readable text from both structured and unstructured data to provide feedback in an easily comprehensible manner. As seen in Figure 1, this constitutes the primary branch module for NLP applications.

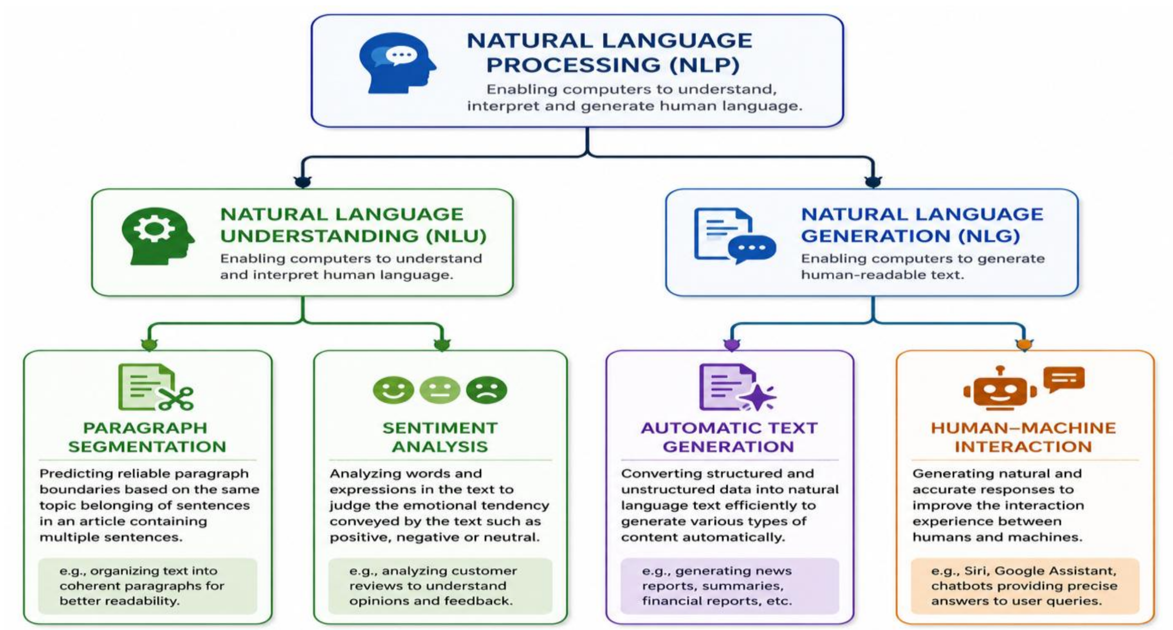

Figure 1: The main branch module of NLP applications

Natural Language Understanding is primarily employed in text segmentation and sentiment evaluation. Text segmentation pertains to the forecasting of dependable paragraph limits predicated on the thematic coherence of phrases inside a singular paragraph of a multi-sentence piece. This segmentation procedure aids in structuring material into logically consistent units and is crucial for enhancing text readability. Emotion analysis is examining the language and phrases inside a text to assess the emotional inclination it communicates, whether good, negative, or neutral. Companies use sentiment analysis to interpret what consumer feedback reveals about their products or services.

NLG is mostly used for the automation of text creation and facilitating human-computer communication. In the realm of automated text creation, NLG assists users in effectively producing several forms of textual material by transforming structured data into natural language.

In the domain of journalism, NLG can swiftly produce precise and current news articles, alleviating the burden on human editors. In the realm of human-machine interaction, NLG technology facilitates the machine's ability to provide natural and precise responses, hence improving the user engagement experience. For instance, artificial intelligence platforms like Siri and Google Assistant employ NLG technology to deliver precise interactive enquiries and responses.

- **Development Frameworks And Tools For Nlp**

The frameworks and tools for development will assist in constructing the industrial applications mentioned before. A plethora of development tools are currently accessible, owing to the considerable enthusiasm exhibited by open source communities globally. These frameworks and tools provide integrated libraries and are also modifiable to meet the particular requirements of the sector.

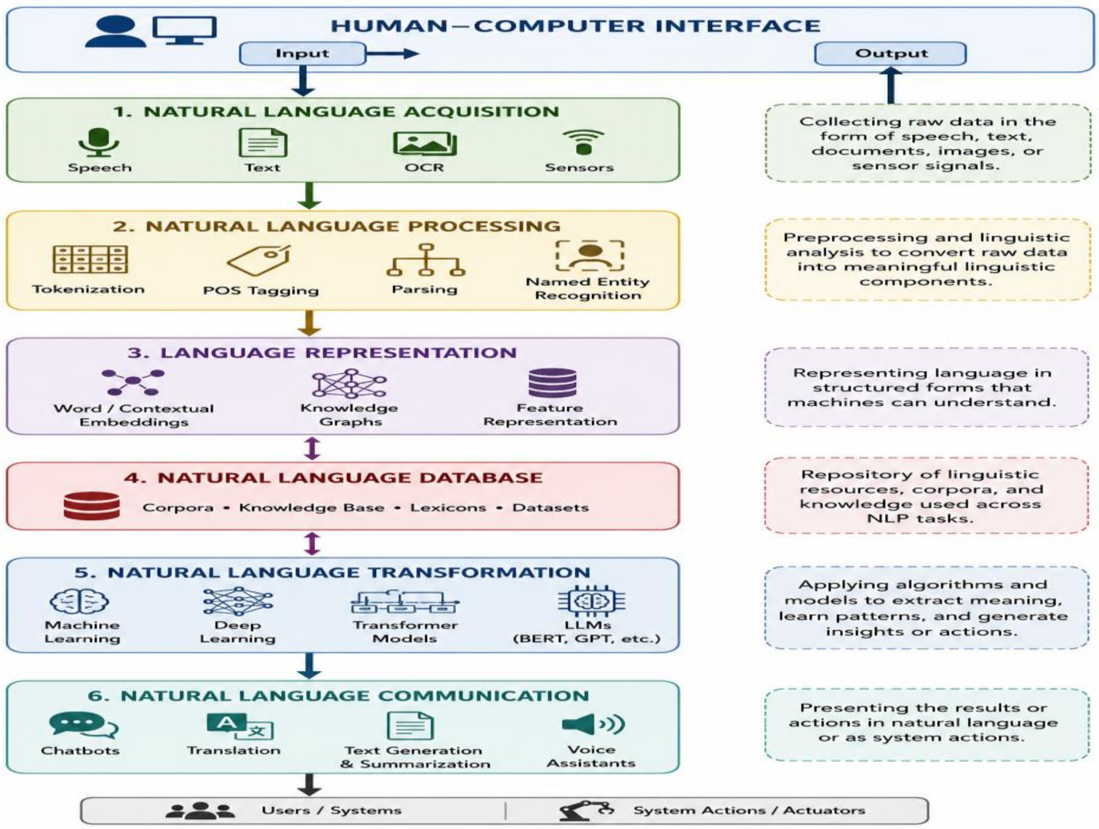

Fig 2: Block representation of stages in the development of NLP tools

Figure 2 illustrates the block diagram of several phases in the development of NLP applications. The Natural Language acquisition module constructed using voice processing, computer vision, or any data collection instruments to integrate Natural language text into the system.

Table 1: Natural Language Processing Tools

Tool	**Developer / Supported by**	**Native Language Support**	**Usage**	**Application Areas**
CoreNLP	Stanford Group	Java	Open source (GPL)	Common NLP tasks and analytics
NLTK	Steven Bird and Edward Loper	Python	Apache License	Common NLP tasks and analytics; supported by Apache Software Foundation
TextBlob	Steven Loria	Python	Open source	Specialized library for text analytics on top of NLTK

Gensim	RaRe Technologies	Python	GNU Lesser GPL	Specialized library for topic modelling
spaCy	Explosion.ai	Python	MIT License	Common NLP tasks and seamless integration with Python AI ecosystem
OpenNLP	Apache Software Foundation	Python	Apache License	Supports common NLP tasks and useful for building text analytics applications
UIMA	IBM / Apache Software Foundation	Java and C++	Apache License	Supports text analytics and multimodal analytics
GATE	GATE Research Team, University of Sheffield (UK)	Java	GNU Lesser GPL	GUI-based development tool for text mining and language processing
Chainer	Preferred Networks Inc.	Python	Open source	Supports feed-forward networks, convolutional networks, recurrent networks, and recursive networks
Deeplearning4j	Open Community	Java and Scala	Apache License	Supports topic modelling, parallel GPUs and RNN, Word2Vec, Doc2Vec, and GloVe algorithms
DeepNL	Giuseppe Attardi	Python	Open source (GPL)	Supports common NLP tasks
DyNet	Carnegie Mellon University	C++ with wrappers for Python	Open source	Supports common NLP tasks
Nlpnet	MIT	Python	MIT License	Supports common NLP tasks
OpenNMT	MIT	Python	MIT License	Specialized library for machine translation, summarization, image-to-text, speech recognition, language modelling, and sequence tagging
PyTorch	Facebook and Uber AI Team	C++, Python	Open source	Supports development of

				computer vision and NLP tasks
TensorFlow	Google Brain Team	C++ and Python	Apache License	Supports many deep learning algorithms and API interfaces for Python, C, C++, Java, and Scala
TFLearn	TFLearn Contributors	C++ and Python	MIT License	Built on top of TensorFlow; easy-to-use deep neural network API with helper functions
Theano	Montreal Institute for Learning Algorithms	Python	Open source (BSD License)	Supports mathematical and matrix operations
Keras	François Chollet	Python	MIT License	Supports easy prototyping and various deep neural networks
LingPipe	Alias-i Inc.	Java	Alias-i License	Specialized in text processing; provides Java APIs
MALLET	Andrew McCallum	Java	Common Public License	Supports common NLP tasks
LUIS	Microsoft	Microsoft SDK	Pay per use	Cloud-based APIs supporting rapid development of bots and conversational systems using pre-built domain-specific models

The Natural Language representation module employs structured, tree, or graph frameworks to depict Natural Language comprehension. A Natural Language database serves as a collection of Natural Language data, such as MNIST or analogous databases, which are then used by machine learning algorithms for various NLP applications. This database is used by representation and transformation modules to execute their functions. Natural Language transformation will include a collection of diverse learning and extraction algorithms designed to derive significant actions from NLP jobs. Natural Language communication represents the intended behaviours resulting from NLP tasks. The produced result may manifest as either Natural Language or as a computational action, such as the movement of a robotic arm. The subsequent table provides a summary of the prevalent instruments presently accessible in the marketplace.

5. APPLICATION MODEL OF NATURAL LANGUAGE PROCESSING

The advancement of NLP is intrinsically linked to many significant models. In 2017, Google unveiled the Transformer model, which utilises the self-attention mechanism, establishing a crucial basis for advancements in NLP and having a significant impact. In 2018, Google unveiled the Bert model, while OpenAI concurrently produced the GPT-1 model. Both frameworks integrate pre-training components built upon the Transformer architecture.

- **Transformer Model**

The Transformer architecture was first used for translating natural languages. In contrast to earlier predominant models like as LSTM and GRU, the Transformer exhibits two notable advantages: Initially, the Transformer employs a multi-head attention mechanism and may use distributed CPUs for simultaneous training, enhancing both training efficiency and accuracy. Secondly, conventional RNNs and LSTMs sometimes exhibit subpar performance in capturing long-term relationships because to their sequential processing of input, which results in issues of gradient vanishing or gradient explosion. The Transformer model employs a self-attention mechanism that enables it to immediately assess the relationships between any two positions in the sequence, hence enhancing its efficiency in capturing long-term dependencies. The Transformer architecture primarily consists of an input component, an N-layer encoder segment, an N-layer decoder segment, and an output component. The whole process of the Transformer is shown in Figure 3.

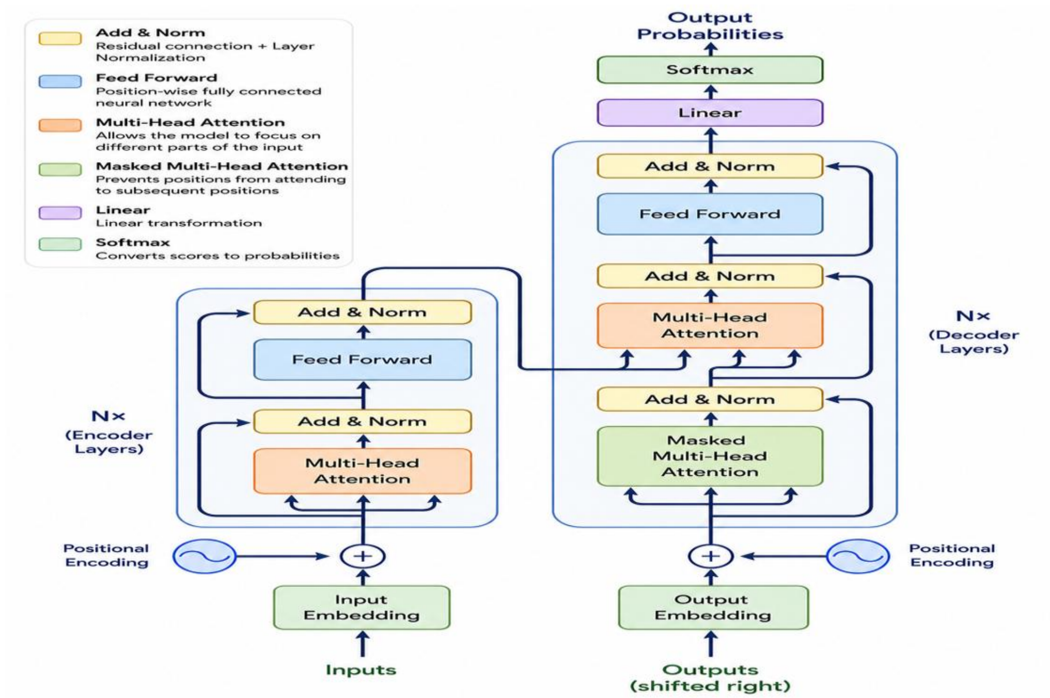

Figure 3: Complete process of Transformer model

The multi-head attention mechanism inside the decoder serves as the foundation of the Transformer model. The attention mechanism computes the final attention result by multiplying the attention weight with the QK matrix and then applying this weight to the V matrix. Figure 3 illustrates the methodology for computing attention outcomes. Multi-head denotes the simultaneous feature extraction from several individual single-head self-attention methods. Exploring parameters across several parameter spaces will inherently enhance the model's accuracy as the number of retrieved parameters rises. Nonetheless, the limitation of the Transformer model is in its applicability just to the study of brief phrases, since it fails to discern the positional relationships among words for evaluation.

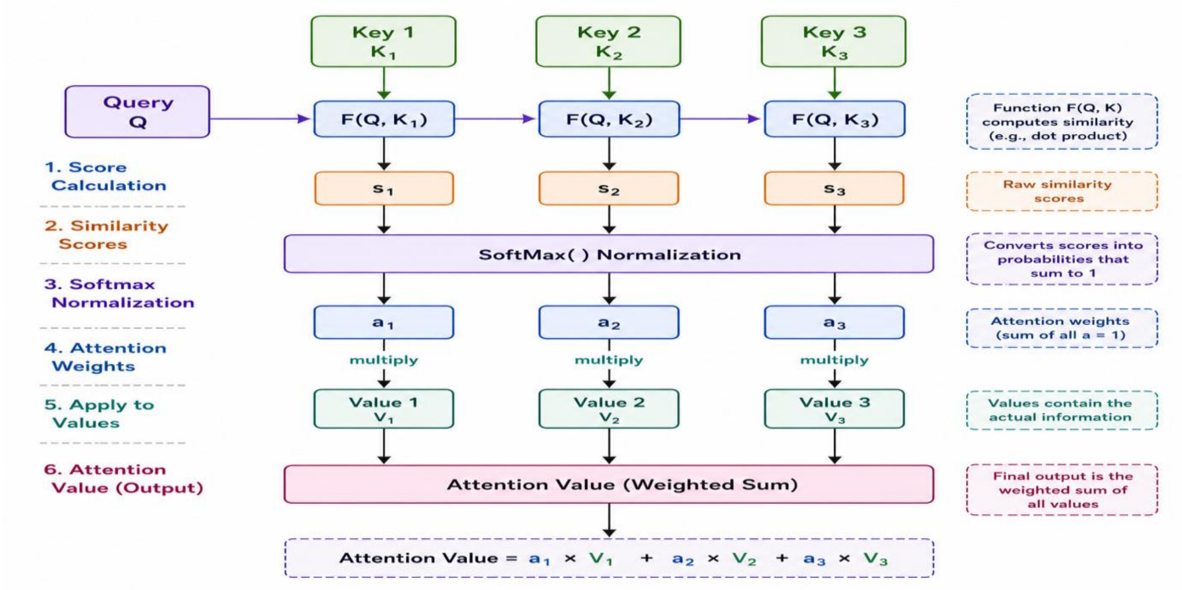

Figure 4: The process of calculating the outcome of attention

The Transformer architecture employs a training dataset: The WMT 2014 English-German dataset comprises 45,000 pairs of sentences. The sentence preprocessing uses Byte-pair encoding to reduce the size of the training set's vocabulary, resulting in a lexicon of 37,000 tokens. The apparatus used consists of 8 NVIDIA P100 GPUs. The criteria for assessment used are: BLEU (Bilingual Evaluation Understudy) Understudy is a technique used to assess the calibre of machine translation, specifically to gauge the resemblance between machine-generated translations and human-produced translations. It specifically assesses the intersection of n-grams, which are consecutive sequences of n elements from a specified text sample, between the translated result and the reference translation. Elevated BLEU scores indicate a greater degree of similarity, implying that the machine-generated translation aligns more closely with human translation quality. This measure is especially significant in assessing the fluency and precision of translation models, establishing it as a fundamental element in the domain of machine translation research and development. It specifically assesses the intersection of n-grams (consecutive sequences of n elements from a particular text sample) between the translated result and the reference translation. The alternative training set is the WMT 2014 English-French dataset, a substantial corpus that contains 36 million words and categorises the tokens into 32,000 unique phrases. The equipment used consists of eight NVIDIA P100 GPUs.

- **BERT Model Based on Transformer**

The Bert model employs the Encoder component of the Transformer architecture. The primary distinction from the Transformer Model is the incorporation of two pre-training tasks: Mask Language Model (MLM) and Next Sentence Prediction (NSP) are two essential tasks in the pre-training of language models. MLM entails the arbitrary omission of certain words within the input text, then training the model to deduce these concealed words from the contextual clues. The NSP job is on instructing the model to ascertain if two phrases are sequentially related within a cohesive text. A further distinction is the incorporation of bidirectional processing during the pre-training phase. The Bert model is capable of assessing the contextual information that precedes and follows each word. The Bert model is extensively used in natural language understanding tasks due to its robust contextual comprehension capabilities.

- **GPT Model Based on Transformer**

GPT employs the encoder architecture of the Transformer as its main function is text production, while the decoder is tailored for generative tasks. GPT has the ability to forecast the subsequent word based on the preceding word, indicating that the input phrase is linear. The unidirectional self-attention mechanism is used, concentrating just on preceding words in the sequence, so exhibiting a robust capacity for generation and interactive question responding.

Moreover, the most significant characteristic of GPT is its extensive number of parameters, which guarantees the quality and precision of interactive conversations. The pre-trained GPT architecture consists of 12 Transformer layers, and its dimensional specifications are.

$$d_{model} = 768 \quad (1)$$

The aggregate count of parameters surpassed 110 million. As the number of model parameters steadily rises, Google has persistently refined the GPT model. Introduced in 2020, GPT-3 (also referred to as chatGPT) is the most formidable and comprehensive language model to date, producing output that is remarkably consistent and contextually relevant. The capability of GPT-3 to assimilate knowledge from a limited number of examples is a significant breakthrough, swiftly comprehending fresh data from minimal samples. This enhancement is attributable to the augmented quantity of parameters used in GPT-3 relative to GPT-2 and GPT-1. GPT comprises 175 billion parameters, while its forerunners, GPT-1 and GPT-2, contain 117 million and 1.5 billion parameters, respectively.

6. DISCUSSION

This document emphasises the notable progress in Natural Language Processing (NLP), especially the shift from rule-based frameworks to transformer structures and Large Language Models (LLMs). The research illustrates that architectures like Transformer, BERT, and GPT have significantly enhanced the efficacy of several NLP applications, including machine translation, sentiment analysis, text summarisation, and conversational AI. It also highlights the increasing use of NLP in healthcare, education, banking, e-commerce, and several other sectors. Notwithstanding these accomplishments, obstacles including linguistic ambiguity, multilingual processing, computing expenses, data confidentiality, prejudice, and ethical issues persist in shaping the advancement of dependable and effective NLP systems.

7. CONCLUSIONS

Natural Language Processing has emerged as a critical domain within Artificial Intelligence, allowing robots to comprehend and produce human language with enhanced precision. This outlined the progression of NLP, its primary divisions, development instruments, transformer-oriented models, contemporary trends, and real-world applications. It also addressed the principal obstacles that persist in creating resilient and reliable NLP systems. Owing to ongoing progress in deep learning, Large Language Models, and multilingual language technologies, NLP is anticipated to have an increasingly significant role in sophisticated human–computer interaction and digital transformation across diverse sectors. Subsequent investigations need to concentrate on enhancing model efficiency, interpretability, equity, and confidentiality to facilitate the development of more dependable and accountable NLP programs.

REFERENCES

1. Ghazizadeh, E., & Zhu, P. (2020, October). A systematic literature review of natural language processing: Current state, challenges and risks. In *Proceedings of the future technologies conference* (pp. 634-647). Cham: Springer International Publishing.
2. Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing. *ieee Computational intelligenCe magazine*, *13*(3), 55-75.
3. Kalyanathaya, K. P., Akila, D., & Rajesh, P. (2019). Advances in natural language processing–a survey of current research trends, development tools and industry applications. *International Journal of Recent Technology and Engineering*, *7*(5C), 199-202.
4. Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, *349*(6245), 261-266.
5. Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. *Multimedia tools and applications*, *82*(3), 3713-3744.
6. Ardkhani, P., Vahedi, A., & Aghababa, H. (2023). *Challenges in natural language processing and natural language understanding by considering both technical and natural domains*. In *2023 6th International Conference on Pattern Recognition and Image Analysis (IPRIA)* (pp. 1–5). IEEE.
7. Azhar, U., & Nazir, A. (2024). *Exploring the natural language generation: Current trends and research challenges*. In *2024 International Conference on Engineering & Computing Technologies (ICECT)* (pp. 1–6). IEEE.

8. Chang, M. W., Devlin, J., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of deep bidirectional transformers for language understanding*. OpenAI Blog.
9. Das, S., & Das, D. (2024). *Natural language processing (NLP) techniques: Usability in human-computer interactions*. In *2024 6th International Conference on Natural Language Processing (ICNLP)* (pp. 783–787). IEEE.
10. Ganesh, A., Strubell, E., & McCallum, A. (2019). *Energy and policy considerations for deep learning in NLP*. *arXiv preprint arXiv:1906.02243*.
11. Gupta, N. K., Chaudhary, A., Singh, R., & Singh, R. (2023). *ChatGPT: Exploring the capabilities and limitations of a large language model for conversational AI*. In *2023 International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT)* (pp. 139–142). IEEE.
12. Guo, D. C. (2024). *Research on natural language understanding problems in open-world scenarios* (Master's thesis). Beijing University of Posts and Telecommunications.
13. Jiang, L., Tang, H. L., & Chen, Y. J. (2024). *A review of natural language processing based on the Transformer model*. *Modern Computer*, (14), 31–35.
14. Kurt, U., & Çayir, A. (2023). *A modern Turkish poet: Fine-tuned GPT-2*. In *2023 8th International Conference on Computer Science and Engineering (UBMK)* (pp. 1–5). IEEE.
15. Li, X., Wang, S., Wang, Z., & Zhu, J. (2021). *A review of natural language generation*. *Computer Applications*, *41*(5), 1227–1235.
16. Mitra, A. (2020). *Sentiment analysis using machine learning approaches (Lexicon based on movie review dataset)*. *Journal of Ubiquitous Computing and Communication Technologies*, *2*(3), 145–152.
17. Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). *Natural language processing: An introduction*. *Journal of the American Medical Informatics Association*, *18*(5), 544–551.
18. Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving language understanding by generative pre-training*. OpenAI.
19. Satpute, R. S., & Agrawal, A. (2023). *Machine learning approach for ambiguity detection in social media context*. In *2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI)* (pp. 516–522). IEEE.
20. Shazeer, N., Vaswani, A., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). *Attention is all you need*. *Advances in Neural Information Processing Systems*, *30*, 5998–6008.
21. Xia, Z., Zhang, C., & Woodland, P. C. (2021). *Adapting GPT, GPT-2 and BERT language models for speech recognition*. In *2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)* (pp. 162–168). IEEE.
22. Zhao, Y. B., Jiang, F., & Li, P. F. (2024). *A multi-level coherent text segmentation method based on BERT*. *Computer Applications and Software*, (10), 262–268, 324.
23. Zheng, X., Zhang, C., & Woodland, P. C. (2021). *Adapting GPT, GPT-2 and BERT language models for speech recognition*. In *2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)* (pp. 162–168). IEEE.