International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 05, IIRA 5.0 (2026)

Exploring Applications and Challenges in NLP: From Pre-trained Models to Real-world Use Cases

G.Pradeepa
Apex Institute of Technology
Chandigarh University
Punjab, India
g.pradeepa1309@gmail.com

Ashmeet Kaur Deol
Apex Institute of Technology
Chandigarh University
Punjab, India
ashmeetdeol17@gmail.com

Mr. Prabjot Singh
Apex Institute of Technology
Chandigarh University
Punjab, India
prabjot.pb.bali@gmail.com

***Abstract*— Natural Language Processing (NLP) has developed greatly in accuracy and efficiency, transforming modern AI applications with continuous improvements in deep learning and pre-trained models. This paper provides a comprehensive review of recent advancements in NLP, areas of its applications, and challenges being encountered from training to deployment. The paper focuses on the main principles of NLP such as computational linguistics, Natural Language Understanding (NLU) and Natural Language Generation (NLG). Current multi-domain applications and areas that can benefit from leveraging NLP techniques are discussed, along with a survey of key challenges and methods proposed to mitigate them so far. The work also touches upon the role of Pre-trained Language Models (PLMs) such as GPT, BERT and their variants which have enhanced NLP tasks. Comparative analyses of various NLP models and methodologies are presented on the basis of performance metrics and benchmarking results. This paper aims to offer researchers and practitioners a clear perspective on NLP's current state, its potential, and future directions.**

Keywords—NLP, BERT, GPT, transformer-based models, ambiguity, ethical AI

I. INTRODUCTION

The combination of linguistic principles with computer science has initiated the concept of Natural Language Processing (NLP). It is a fast-moving field and a subset of the broader concepts of Artificial intelligence and Machine Learning [1]. NLP enables computers to interpret, analyze, and generate human language, thus playing a major role as the core technology of many applications ranging from voice assistants and chatbots to translation tools such as Siri, ChatGPT, Google Translate, etc. Significant advancements in this area have made human-computer interactions smoother and have reduced language barriers. However, ongoing research continues to work on overcoming the limitations of NLP and improving its efficiency.

The initial step in NLP while interpreting human language involves the conversion of input to machine-readable formats and performing a series of NLU (Natural Language Understanding) tasks. The response is generated by the machine after applying statistics and deep learning to the processed data, with the utilization of knowledge bases and pre-trained models. These functions of NLP power today's real-world applications like spam detection, automated customer service, sentiment analysis, search engines, document summarization, coding tasks, and many more in various sectors including healthcare, finance, legal analytics, and social media monitoring [2].

In recent years, deep learning has revolutionized NLP by powering pre-trained models that serve as the focal point of modern NLP by improving contextual understanding and text generation capabilities that are more human-like [3]. BERT and GPT have rapidly gained popularity due to their use in conversational AI applications that ingest vast datasets from multiple sources and consist of complex neural architectures designed to perform language-related tasks [4].

Despite reducing manual efforts and human errors in our everyday lives by automating linguistic tasks, even some of the highly advanced NLP models pose major challenges like bias during training, high computation costs, multiple types of ambiguity, and lack of support for less-known languages, among other issues [5].

The paper is structured into sections that explore the fundamentals of NLP, its applications, and key challenges, followed by a comprehensive review of recent advancements in the field, focusing on pre-trained models.

Natural Language Understanding (NLU) and Natural Language Generation (NLG) are the two primary aspects of the technique of Natural Language Processing (NLP). These are the steps that are crucial to effectively process the linguistic data that has been received and provide the desired outcomes. Getting the computer to understand the complex human language i.e., NLU, is a more challenging task, when compared to enabling it to generate a response in natural language i.e., NLG.

A. Natural Language Understanding (NLU)

The first step to comprehension is the reception and recognition of a language given as input in any form such as text or speech and converting it to machine language for further processing. Such encoding for voice recognition is nowadays commonly done using Hidden Markov Models (HMM) that can convert speech into text by fragmenting the speech in smaller sections, applying mathematical calculations based on the phenome data fed for training the data and comparing each fragment to it [7]. The Machine Learning model predicts the words uttered through this method and keeps the record.

The ML model for NLU is basically composed of various types of classifiers designed to work on various segments of linguistic data and produce filtered outcomes at each stage to achieve end results. At the minimum level, the entire vocabulary corpus of a natural language is used as primary data for model training. Neural Networks such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are mostly well-suited for NLP sequential data processing [8]. Although a majority of the NLP models are trained with supervised learning methodology, there have been recent

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 05, IIRA 5.0 (2026)

advancements in the unsupervised learning sector for them, as well [8].

Natural Language Understanding (NLU) follows a structured sequence for processing text. It begins with POS Tagging, where each word is classified under grammatical categories (noun, verb, adjective, etc.) [9]. Unidentified words are re-labeled using n-grams, and bidirectional inference is commonly applied. The "Guided Learning" algorithm (Shen et al., 2007) achieved 97.33% per-word accuracy [9]. Next is Shallow Parsing (Chunking), which groups tagged words into phrases for contextual understanding. Shen & Sarkar (2005) attained 95.23% tagging accuracy using a voting classifier with a Viterbi decoder [10]. Named Entity Recognition (NER) then identifies entities like names, dates, and locations using ML, CRF, or deep learning models [6]. The most efficient method is deep learning, with Spacy and NLTK being key frameworks. Ando & Zhang (2005) achieved 89.31% accuracy using semi-supervised training with an unlabeled Reuters corpus [11]. Finally, Semantic Role Labeling (SRL) determines sentence structure by analyzing subjects and predicates, generating parse trees for syntax analysis [12].

B. Natural Language Generation (NLG)

Natural Language Generation (NLG) converts structured computer data into human-like language with the use of sequential processing [13]. It starts with Content Determination in which relevant words, sentences, and components are compiled to define rough syntax and semantics. In the next step, Document Organization and Aggregation elements structure the content logically in order to ensure correct sequencing and clarity by grouping related sentences together [13]. This step is followed by Linguistic and Grammatical Structuring part where the most suitable words from a vocabulary corpus are selected while applying predefined grammatical rules to refine composition [1]. In the final step of Realization and Presentation, the processed output is delivered in the form of text or speech according to user requirements and technological support [13]. NLG is the reverse of NLU, applying understood language structures to generate meaningful responses [10]. Bidirectional encoding is preferred [9], often using RNN-based models. While supervised learning is commonly used, unsupervised methods such as seq2seq Neural Machine Translation (Bahdanau et al., 2015) [14] are increasingly deployed to handle linguistic ambiguities more naturally [15].

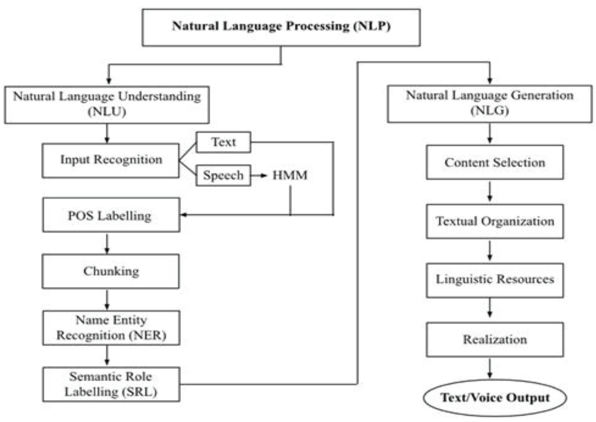

Fig. 1. NLP Procedure Flow

II. Applications of NLP

- Sentiment Analysis- NLP techniques are used to analyse, understand and interpret the meaning of sentiments expressed through text. It is used by many businesses to identify the consumer's views and sentiments via online mediums to enhance their products. It plays a crucial role in monitoring social media. NLP can recognize and gather data on the general feelings of a large number of people by examining various posts, comments, and responses about a certain topic when it is challenging to go through these manually. Sentiment analysis is a machine learning tool that analyses text for polarity, from positive to negative. The range of polarity lies between -1 to 1, where -1 denotes extremely negative and 1 is extremely positive. Support Vector Machine (SVM) is a learning technique that performs well on sentiment classification. The performance of SVM depends on the used kernel function. Polarity feedback can be used to monitor social media sentiments about a product or brand, identifying the latest interests of people. A polarity feedback system can also be used to analyse political speeches, tweets, and other forms of communication to identify the behaviour of people toward a certain topic or issue [16].

In their work, [17] propose a hybrid architecture combining RoBERTa with CNN/LSTM layers for sentiment analysis that achieves 96.28% accuracy on IMDB reviews and 94.2% on Twitter datasets using SMOTE for class imbalance mitigation.

Companies employ natural language processing (NLP) to examine social media posts, reviews, and online comments. Sentiment analysis is used by businesses such as Netflix and Amazon to measure client opinion and enhance offerings [18]. Customer Data Extraction in E-Commerce improves product recommendations and customer engagement, and NLP-based algorithms examine user preferences. In industries like retail and finance, AI-powered chatbots help with answering frequently asked questions, enhancing customer service, and lowering manual labour.

- Machine Translation – NLP helps to translate data from one language to another using different tools like Google Translate, which allows us to convert data in real time while tackling challenges such as idiomatic expressions and contextual ambiguity [19]. It also involves summarization of news, reports and documents. It helps to retrieve useful information from large datasets which is useful for businesses and researchers. [2]. Sign language translation, converting gestures into text/audio and speech to text conversion for users improves accessibility for disabled persons. Such tools of assistive technology help bridge the gap and empower people to interact more confidently and with ease with technology and society. [20]

- Healthcare and medical applications - NLP structures the Electronic Health Records (EHRs) to help the healthcare industry, it also helps to convert the unstructured and not understandable doctor notes into a structured way and make them usable for predictive analysis. NLP facilitates by making chatbots that help to check symptoms. Other than this, it analyses patients' symptoms through emotion-based medication systems to help recommend personalized treatments according to the requirements of the patients individually. It also retrieves

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 05, IIRA 5.0 (2026)

critical details from prescriptions and research papers which optimizes healthcare decision making [21].

- Education- NLP can be used to develop intelligent teaching systems that offer students their personalised instruction and support. These systems assess students' answers, give feedback, and modify the result according to the requirements of the individual [5]. It can be used to detect plagiarism in assignments and worksheets by comparing the data with a vast database. It helps students by providing interactive tools for improving vocabulary, correcting grammar and polishing language. Students can receive feedback based on their performance using NLP to evaluate themselves and know the areas where improvement is required.

- Agriculture – The economy of world relies significantly on agriculture. The term digital agriculture is the integration of technology with crops and animal management. Data intensive techniques are applied to maximise the agricultural productivity and reducing environmental effects. [2] NLP enhances digital agriculture by improving communication, data analysis and decision making. To identify various diseases in crops it uses various data such as farmer feedback and agricultural publications. NLP algorithms are used to identify different patterns and trends and presence of crop illness by analysing the collected data which enables farmers to take preventive measures. It also identifies the presence of weeds and pests in the crop. It can be used for analysing weather-related information from textual data sources, such as weather reports. By analysing this information, NLP algorithms can provide insights into climate patterns, forecast changes, and help farmers make data-driven decisions regarding irrigation, planting schedules, and crop selection. By analysing new articles, market trends, and supply-demand dynamics, algorithms in NLP can help farmers and traders plan for crop selling, storage, and further crop growing. By reviewing textual data from sensor data, historical records, and farm management systems, NLP algorithms may help farmers by forecasting harvests, and more efficiently managing resources like watering and fertilisation.

- Finance, Law and Compliances – NLP automates financial reporting, it generates summaries and detects inconsistencies in compliance documents [22]. It detects unusual transaction patterns and spots fake reviews to enhance fraud detection. Through Legal Question Answering (LQA) to help lawyers with research and legal interpretations, Similar Case Matching (SCM) to locate pertinent prior rulings, and Legal Judgement Prediction (LJP), where AI predicts case outcomes, the legal profession benefits from natural language processing (NLP).

- Cybersecurity and Fraud Prevention – NLP monitors security logs, detect phishing attempts and analyse threat patterns. Anomaly detection in banking and online transactions and secure financial operations, this benefits the fraud detection [23]. Fake product reviews that manipulate consumer perception can also be identified using NLP which benefits the E-commerce. Explainable AI is integrated with NLP to make decision making process of NLP more transparent. Models like Local Interpretable Model-Agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP) clarify how Ai arrives at it conclusion, ensuring accountability and trust which is mostly used in areas like text classification , sentiment analysis and fraud detection [19].

III. Challenges in NLP

- Ambiguity – Ambiguity in NLP refers to the sentence or phrase that may have two or more possible interpretations or the words that have different meanings but are spelled same. It is one of the most persistent challenges that occurs at different linguistic levels of NLP such as lexical, syntactic and semantic. Natural human language is ambiguous due to which NLP models faces the difficulty to interpret the data accurately, which becomes a barrier while performing certain tasks such as translating language, analysing sentiments, question answers, where understanding the context and language is very important. [2]

Pre-trained language models such as BERT, GPT, and XLNet are trained on large datasets before they are fine-tuned for specific tasks, improving contextual understanding [3]. These models help to reduce ambiguity and also improve sentiment analysis, question answering and text classification.

- Syntax and Grammar Variation—Syntactic differences across languages make it challenging for models to distinguish and analyze them accurately [24]. It is because in some languages there are strict rules that are needed to be followed, while some languages have free word order.

Transformer-based hybrid approaches like BERT + GPT in which the masked language modelling of BERT is combined with the casual language modelling of GPT to improve syntactic understanding and text generation[4, 29].

- Idioms and Sarcasm Detection—NLP models are context-dependent models, and they often struggle with non-literal language. They misinterpret the words or phrases or sentences based on the words written rather than focusing on the actual meaning of the idiom [25]. A sentence may seem like praise, but its true meaning could be entirely different, making it difficult for the model to understand.

To detect sarcasm in text, deep learning models such as RoBERTa, XLNet, BERT are contextually embedded to analyze non-literal meaning in text [25].

- Data noise and errors—Noisy data includes misspelled words, slangs, abbreviations, or mistakes done during the speech-to-text translation. [26] The main reasons for mistakes in speech-to-text can occur due to mispronunciation of words, different accents, or having some speech issues. All these factors are responsible for the performance of the machine and it effects the accuracy of the model. Since the models are only trained on data it may not be able to interpret the actual intentions of the writer.

Data filtering, active learning and curriculum learning techniques are used to overcome the issue of data noise and errors which improves the quality of data and makes the model more equitable [26].

- Bias in NLP models [27]

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 05, IIRA 5.0 (2026)

Bias from Data: The data on which the NLP models are trained is derived from various sources, such as newspaper articles, which may be written historically by a standardized group. Due to this, the data contains selection bias, where the model performs better in certain cases, leading to overrepresentation or underrepresentation of certain groups or people.

Bias from Annotations: Annotation occurs when the labeled data is based on the background, experiences, or perspective of the annotator. NLP models are trained on such labeled datasets and any inconsistency in the data can cause the prediction to be biased. In tasks where subjectivity is involved such as emotion recognition and syntactic parsing, this type of bias becomes challenging.

Bias from Input Representation: Gender, racial, and societal biases are introduced even if the datasets are well proportioned through word vectors and contextualized word representation. For instance, an example is provided in this paper that says that embedding has been shown to relate “man” with “programmer” and “woman” as "homemaker,” which is known as semantic bias that can result in discrimination in NLP model output.

Bias from Models: Machine learning models have the potential to magnify biases in data, even when the data is better. This happens because models may take advantage of erroneous associations (such as linking gender to occupation) in order to maximize predicted accuracy. Certain sentiment analysis algorithms, for instance, assign varying ratings to the same statement depending on the gendered pronouns.

Bias from Research Design: The majority of research work is done in English or Indo-European languages; availability bias—a self-reinforcing cycle in which English receives disproportionate attention and more tools are developed for it while other languages remain underrepresented—occurs. Furthermore, model creation may be influenced by the inherent biases of researchers, which could have unanticipated negative consequences.

Biases generated from imbalanced datasets and societal prejudices can be addressed using fairness-aware training, bias-aware embeddings, and subgroup performance metrics [19, 27]. Moreover, human oversight, fairness testing and structured ethical review process, all these factors ensure ethical deployment of NLP models, which prevents from biases and unintended harm in AI-driven decision-making [28, 31].

- Lack of Explainability and Interpretability – Many NLP models are based on a function called “Black Box”, which means the processing of data is not visible to the user and it is difficult to interpret their decision-making process. It is a significant issue in the fields like healthcare and judiciary purposes [19].

Explainability tools like SHAP, LIME, and Grad-CAM [19] are used to overcome the black box nature of NLP models and it promotes transparency through ethical review board [31].

- Limited Domain and Data Dependency - NLP models are based on specific domains and perform well in that certain domain. Generalising their performance with unfamiliar domains is a challenge. The reliability of NLP models on a large amount of training data can cause hindrance in performance as the availability of domain-specific or language-specific labelled data may be comparatively less. Many languages don’t have enough amount of training data which lead to the underfitted model and affects the accuracy of the model [2]

Multilingual PLMs such as Mbert, XLM-R with cross-lingual transfer learning allow the models that are trained on high-resource language to generalize to low-resource language to give significantly accurate results. Retrieval augmented generation (RAG) and continual learning integrate external information sources to keep the models updated to overcome the issue of outdated knowledge and static learning [30].

- Privacy and security risks- NLP models used in healthcare and finance sector store personal data about their patients and customers which poses a privacy risk [28]. Confidential information can be exposed as the data will remain stored in the model even when not required.

Data anonymization, encryption and federated learning methods are used to address issues related to privacy and security risks like data leaks, unauthorized data retention [23, 28], ensuring compliance with data protection standards like GDPR and HIPAA.

- Computational Costs and Resource Inefficiency- NLP models like GPT-4 require massive computational resources and large datasets to train, leading to high training and inference costs. [5].

Techniques like model pruning, quantisation and parameter efficient tuning are used to reduce the high computational costs of large-scale models that help lower the memory consumption and interference time while maintaining the accuracy.[26]

IV. Pre-trained Language Models (PLMs)

The problem of data scarcity in training conversation systems is highlighted by Zaib et al.(2020) in their survey [3],which reduces the capacity of system to learn syntax, grammar, decision making and reasoning effectively. As a solution PLMs are proposed, which are used to pre-train models on large amount of unlabelled data and then the data is fine tune on smaller and task-specific datasets. To correctly identify the context of word in sentences and overcome the limitations of traditional word embeddings like Word2Vec and GloVe, PLMs such as ELMo, GPT, BERT and XLNet have been designed.

Earlier models like RNNs and LSTMs did not perform well with long sequences due to the vanishing gradient problem. This limited their ability to maintain context and parallel processing became a difficult task. In contrast to GPT’s unidirectional approach, BERT improved upon this with a bidirectionality, analyzing both forward and backward context. This helped in tasks like text classification and question answering. XLNet builds on both GPT and BERT by using Transformer-XL to model bidirectional contexts. It overcomes BERT's reliance on masked tokens and achieves state-of-the-art results by integrating autoregressive and autoencoding methods [4].

GPT-BERT [29], a hybrid language model that combines CLM (GPT-like) and MLM (BERT-like) approaches was introduced by Charpentier & Samuel (2024). By merging the two training objectives, the model benefits from both

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 05, IIRA 5.0 (2026)

efficient text generation (CLM) and deep language understanding (MLM). GPT-BERT outperforms both pure MLM and pure CLM approaches on various BabyLM benchmarks.

Currently, some of the most advanced models include GPT-4, and BERT [4, 5, 29] among more notable models like RoBERTa, XLNet, ALBERT, and StructBERT [4,17,32,33]. The foundational works like the original BERT [34] and Transformer-XL [35] have been modified by multiple researchers to improve efficiency and overcome their limitations.

In a recent study [36], Pookduang et al. (2025) demonstrate RoBERTa's superiority over traditional deep learning models like Naive Bayes, KNN, CART, and LSTM with 96.30% accuracy and 98.11% F1-score on Amazon reviews and highlight advantages in e-commerce sentiment analysis. In their paper [37], Yang et al. (2019) explain XLNet's permutation-based training which gets 87.9% EM score on SQuAD 2.0 and outperforms BERT on GLUE/RACE benchmarks and finds applications in QA systems and text summarization. Parameter-reduction techniques of factorized embedding, cross-layer parameter sharing and advance SOTA on 12 tasks, including SQuAD v2.0 and RACE was introduced by Lan et al. (2020) in their work [32]. The StructBERT [33] model proposed by Wang et al. (2019) enhances BERT with word/sentence-level structural objectives and achieves 89.0 GLUE score and 93.0 F1 on SQuAD v1.1.

Table 1. Comparison of NLP Models

Model	Key Innovation	Benchmark Performance
BERT (2018)	Bidirectional context via masked language modeling (MLM)	Foundation for modern NLP; 93.5% accuracy on SQuAD v1.1
RoBERTa (2019)	Optimized training (dynamic masking, larger batches)	96.3% accuracy on Amazon reviews
ALBERT (2019)	Parameter efficiency via cross-layer sharing	SOTA on RACE (89.4%) and SQuAD v2.0 (92.2%)
DeBERTa (2020)	Disentangled attention (separate positional/token encoding)	90.8% on MNLI, 93.1% on SQuAD v2.0
GPT-3 (2020)	Autoregressive training at scale (175B params)	75.1% accuracy on SuperGLUE
GPT-4 (2023)	Multimodal architecture optimization	86.4% on MMLU (5-shot), 88.8% HellaSwag
GPT-4 Turbo (2024)	Enhanced inference efficiency	4x faster than GPT-4 with comparable accuracy
GPT-4o (2025)	Real-time multimodal processing	1.8x speed gain over Turbo, 92.1% on MMLU
Transformer-XL (2019)	Segment recurrence + relative positional encoding	0.99 bpc on enwiki8, 18.3 PPL on WikiText-103

Feature	BERT (Bidirectional Encoder Representations from Transformers)	GPT (Generative Pre-training Transformer)	Hybrid (GPT-BERT)	Advantages	Limitations	Reference
Architecture	Encoder (Bidirectional)	Decoder (Autoregressive)	Combines both	Balances text understanding and generation	Requires hybrid fine-tuning	[29]
Training Objective	Masked Language Model (MLM), Next Sentence Prediction (NSP)	Causal Language Model (CLM)	Masked Next-Token Prediction (MNTP)	Better for multitasking NLP applications	Computationally heavy	[29]
Best for	Understanding & classification tasks	Text generation, chatbots	Both comprehension & generation	Covers broader NLP applications	Increased model complexity	[3, 29]
Limitations	Weak at long text generation	Prone to biases & hallucinations	Requires higher computational cost	May require large-scale datasets	Limits adaptation to highly specific tasks	[29, 30]
Popular Applications	Search engines, question answering (Google BERT, SQuAD)	Chatbots (ChatGPT), Text Generation (GPT-3, GPT-4)	Hybrid NLP systems	Improves efficiency in multiple domains	Training requires significant fine-tuning	[3, 29]

Table 2. BERT vs. GPT vs. Hybrid: Feature Analysis

V. Conclusion And Future Directions

AI – driven communication has evolved as a result of NLP, which has enabled advancements across variety of fields. Language understanding has been enhanced by pre-trained models like BERT, GPT, and XLNet. Although it has made extraordinary strides in recent years. Inspite of the advancements, some issues such as prejudice, vagueness, high technology expenses, and absence of clarity remain problems that need solutions.

Future research should address the problems of inclusivity and the mitigation of bias, interpretation of the model, and multilingual NLP. Promote sustainable development through improved efficiency by applying model pruning and quantization techniques. Combining artificial intelligence with traditional techniques improves accuracy and model explainability. Enhancing efficiency while maintaining social responsibility in natural language processing will be achieved through ethical AI.

References

[1] Khurana, D., Koli, A., Khatter, K., & Singh, S. (2022). NLP: State of the art, trends & challenges. Multimedia Tools Appl., 82. https://doi.org/10.1007/s11042-022-13428-4

[2] B, Priya & J.M, Nandhini & Thangavel, Gnanasekaran. (2021). “An Analysis of the Applications of Natural Language Processing in Various Sectors”. 10.3233/APC210109.

[3] Zaib, M., Sheng, Q. Z., & Zhang, W. E. (2021). Survey of pre-trained language models for conversational AI. arXiv. https://arxiv.org/abs/2104.10810

[4] Topal, M. O., Bas, A., & Van Heerden, I. (2021). Exploring Transformers in Natural Language Generation: GPT, BERT, and XLNet. ArXiv. https://arxiv.org/abs/2102.08036

[5] Alawida, M., Mejri, S., Mehmood, A., Chikhaoui, B., & Abiodun, O. I. (2023). Study of ChatGPT: Advances, limits & ethics in NLP and cybersecurity. Information, 14, 462. https://doi.org/10.3390/info14080462

[6] Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. "Natural Language Processing (Almost) from Scratch". J. Mach. Learn. Res. 12, null (2/1/2011), 2493–2537.

[7] Awad, Mariette & Khanna, Rahul. (2015). "Hidden Markov Model". 10.1007/978-1-4302-5990-9_5.

[8] Yoon Kim. 2014. "Convolutional Neural Networks for Sentence Classification". In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1746–1751, Doha, Qatar. Association for Computational Linguistics.

[9] Shen, G. Satta, and A. K. Joshi. Guided learning for bidirectional sequence classification. In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL), 2007.

[10] Shen and A. Sarkar. Voting between multiple data representations for text chunking. Advances in Artificial Intelligence, pages 389–400, 2005.

[11] K. Ando and T. Zhang. A framework for learning predictive structures from multiple tasks and unlabeled data. JMLR, 6:1817–1953, 11 2005.

[12] Hao, Tianyong & Huang, Zhengxing & Liang, Likeng & Weng, Heng & Tang, Buzhou. (2020). "Health Natural Language Processing: Methodology Development and Applications" (Preprint). 10.2196/preprints.23898.

[13] Albert Gatt and Emiel Krahmer. 2018. "Survey of the state of the art in natural language generation: core tasks, applications and evaluation". J. Artif. Int. Res. 61,1 (January 2018), 65–170.

[14] Bahdanau, K. Cho, and Y. Bengio. 2015. Neural Machine Translation by Jointly Learning to Align and Translate. In International Conference on Learning Representations (ICLR).

[15] Freitag, M., & Roy, S. (2018). Unsupervised natural language generation with denoising autoencoders. In EMNLP 2018, 3922–3929.

[16] Priya, B., Nandhini, J. M., & Gnanasekaran, T. (2021). NLP applications across sectors. Adv. Parallel Comput., 38, 598–602. https://doi.org/10.3233/APC210109

[17] Tan, Kian & Lee, Chin-Poo & Anbananthen, Kalaiarasi & Lim, Kian. (2022). RoBERTa-LSTM: A Hybrid Model for Sentiment Analysis with Transformers and Recurrent Neural Network. IEEE Access.

[18] Maqsood, A. B., Maag, A., Seher, I., & Sayfullah, M. (2021). Customer data extraction techniques based on natural language processing for e-commerce business analytics. *CITISIA 2021 - IEEE Conference on Innovative Technologies in Intelligent System and Industrial Application, Proceedings*. https://doi.org/10.1109/CITISIA53721.2021.9719914

[19] Qiming, X., Zheng, F., Chenwei, G., Xubo, W., Haopeng, Z., Zhi, Y., Zichao, L., & Changsong, W. (2024). Applications of Explainable AI in Natural Language Processing. Global Academic Frontiers, 2(3), 51–64. https://doi.org/10.5281/zenodo.12684705

[20] Zhang, T., & Liao, S. (2024). Translation system for deaf-mute individuals using NLP. In ICAICE 2024 (pp. 658–661). https://doi.org/10.1109/ICAICE63571.2024.10864057

[21] Zheng, H., Xu, K., Zhou, H., Wang, Y., & Su, G. (2024). Medication Recommendation System Based on Natural Language Processing for Patient Emotion Analysis. *Academic Journal of Science and Technology*, *10*(1), 62-68. https://doi.org/10.54097/v160aa61

[22] Oyewole, A. T., Adeoye, O. B., Addy, W. A., Okoye, C. C., Ofodile, O. C., Ugochukwu, C. E., et al. (2024). Automating financial reporting with NLP: A review and case analysis. WJARR, 21(3), 575–589. https://doi.org/10.30574/wjarr.2024.21.3.0688

[23] Boulieris, P., Pavlopoulos, J., Xenos, A., & Vassalos, V. (2024). Fraud detection with natural language processing. *Machine Learning*, *113*(8), 5087–5108. https://doi.org/10.1007/s10994-023-06354-5

[24] Zhong, H., Xiao, C., Tu, C., Zhang, T., Liu, Z., & Sun, M. (2020). *How Does NLP Benefit Legal System: A Summary of Legal Artificial Intelligence*. http://arxiv.org/abs/2004.12158

[25] Yaghoobian, H., Arabnia, H. R., & Rasheed, K. (2021). *Sarcasm Detection: A Comparative Study*. http://arxiv.org/abs/2107.02276

[26] Treviso, M., Lee, J.-U., Ji, T., van Aken, B., Cao, Q., Ciosici, M. R., et al. (2022). Efficient methods for NLP: A survey. arXiv. http://arxiv.org/abs/2209.00099

[27] Hovy, D., & Prabhumoye, S. (2021). Five sources of bias in natural language processing. *Language and Linguistics Compass*, *15*(8). https://doi.org/10.1111/lnc3.12432

[28] Khattak, W. A. (n.d.). *Ethical Considerations and Challenges in the Deployment of Natural Language Processing Systems in Healthcare*. https://www.researchgate.net/publication/372406349

[29] Charpentier, L. G. G., & Samuel, D. (2024). *GPT or BERT: why not both?* http://arxiv.org/abs/2410.24159

[30] Qin, L., Chen, Q., Feng, X., Wu, Y., Zhang, Y., Li, Y., Li, M., Che, W., & Yu, P. S. (2024). *Large Language Models Meet NLP: A Survey*. http://arxiv.org/abs/2405.12819

[31] Santy, S., Rani, A., & Choudhury, M. (2021). *Use of Formal Ethical Reviews in NLP Literature: Historical Trends and Current Practices*. http://arxiv.org/abs/2106.01105

[32] Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. *ArXiv*. https://arxiv.org/abs/1909.11942

[33] Wang, W., Bi, B., Yan, M., Wu, C., Bao, Z., Xia, J., Peng, L., & Si, L. (2019). StructBERT: Incorporating Language Structures into Pre-training for Deep Language Understanding. *ArXiv*. https://arxiv.org/abs/1908.04577

[34] Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *ArXiv*. https://arxiv.org/abs/1810.04805

[35] Dai, Z., Yang, Z., Yang, Y., Carbonell, J., Le, Q. V., & Salakhutdinov, R. (2019). Transformer-XL: Attentive Language Models Beyond a Fixed-Length Context. *ArXiv*. https://arxiv.org/abs/1901.02860

[36] Pookduang, P., Klangbunrueang, R., Chansanam, W., & Lunrasri, T. (2025). Advancing sentiment analysis: Evaluating RoBERTa against traditional and deep learning models. ETASR, 15(1), 20167–20174. https://doi.org/10.48084/etasr.9703

[37] Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., & Le, Q. V. (2019). XLNet: Generalized Autoregressive Pretraining for Language Understanding. *ArXiv*. https://arxiv.org/abs/1906.08237