International Journal of Innovation Scientific Research and Review
Vol. 07, Issue, 04, pp.8145-8154, April 2025
Available online at http://www.journalijisr.com
SJIF Impact Factor 6.599

Research Article

ISSN: 2582-6131

DEEP LEARNING IN NLP: HOW NEURAL NETWORKS ARE CHANGING LANGUAGE MODELS

[1,]* Ayad Zedo Ismaeel and [2]Ibrahim Mahmood Ibrahim

[1]Akre University for Applied Sciences, Technical College of Informatics, Department of Information Technology, Duhok, Iraq.
[2]Computer Networks and Information Security.

Received 13th February 2025; Accepted 14th March 2025; Published online 20th April 2025

ABSTRACT

This article analyzes the progression of natural language processing (NLP) via deep learning, emphasizing the transformative impact of neural network topologies on language model creation. It examines the transition from conventional rule-based and statistical methodologies to data-driven deep learning techniques adept in identifying intricate language patterns. The document presents a thematic summary of principal applications in natural language processing, encompassing text classification, machine translation, sentiment analysis, and syntactic parsing, while examining the influence of diverse deep learning methodologies, including recurrent neural networks, convolutional architectures, and transformer-based models. This study assesses methodology, datasets, pre-processing approaches, and assessment criteria from a wide range of literature, highlighting both the strengths and persistent shortcomings of existing models. Special emphasis is placed on difficulties in semantic comprehension, adaptation to low-resource environments, and computing requirements. The study continues by delineating prospective avenues for future research, encompassing model efficiency, flexibility, and enhanced language and cognitive integration within neural architectures.

Keywords: Deep Learning, Natural Language Processing (NLP),Neural Networks, Transformer Models, Language Modeling, Machine Translation, Low-Resource Languages.

INTRODUCTION

The study begins with a history of NLP from statistical approaches to deep learning models. It shows how RNNs and transformers (e.g., BERT, GPT) improve NLP tasks like part-of-speech tagging and named entity recognition. The part emphasizes quantitative evaluation and shows real-world applications like sentiment analysis and machine translation, setting the stage for the paper's extensive discussion.[1] The article introduces AI's expanding importance in NLP, focusing on sentiment analysis and text summarization. RNNs and Transformers, annotated datasets for model training, and transfer learning to increase task-specific performance are stressed. Accuracy, ROC curve, and F1 score are also covered. The introduction establishes the research's goals, methods, and importance.[2] The paper's introduction discusses how deep learning has transformed Natural Language Processing (NLP), helping machines understand and synthesize human language. The applications include sentiment analysis, machine translation, text summarization, question answering, and speech recognition. Performance-impacting models include Transformers, BERT, GPT, and attention-based architectures. Some languages have limited data, huge models require computer power, and ethical issues like prejudice are also discussed in the beginning. It prepares for a deeper look at NLP advances and future directions.[3] provides a complete review of recent NLP language modeling approaches. The work seeks to connect ancient and sophisticated approaches like n-gram and hidden Markov models to modern systems like BERT, GPT, LLAMA, and Bard. Agglutinative languages like Arabic and Turkish provide modeling issues, and the research compares transformer-based and conventional methods. It also provides implementation tips for NLTK, TensorFlow, PyTorch, and Gensim. The introduction prepares for a detailed treatment of language modeling theory and practice.[4] NLP and hybrid and ensemble models are becoming more important in language comprehension tasks including sentiment analysis, translation, and text categorization. It uses mixed models like RNNs and BERT to boost performance and handle computing needs, over fitting, and interpretability. The study reviews these techniques' structures, metrics, and applications. The preface promotes the paper as a resource for NLP deep learning.[5] analyzes how deep learning transforms NLP. It covers multi-layer neural networks, TensorFlow, PyTorch, and NLP applications including word embedding, language modeling, and text synthesis. The introduction emphasises model performance optimisation and real-world applications including machine translation, conversation systems, and speech recognition. It provides a good basis for studying deep learning and NLP optimization.[6] The introduction emphasizes the rising relevance of Natural Language Processing (NLP) in artificial intelligence, which aims to help robots comprehend and synthesize human language. Application examples include machine translation, sentiment analysis, and speech recognition. Deep Learning (DL) has improved text categorization, sequence tagging, and text creation in NLP. The function of word and phrase embeddings in capturing semantic meaning is examined. Different categorization methods and DL's influence on text production are also covered in the beginning, setting the groundwork for contemporary NLP approaches.[7] The study emphasizes NLP's importance in artificial intelligence and deep learning's transformation of language challenges. It highlights how deep learning has transformed named entity recognition, parsing, and speech recognition. Deep learning lets models capture complex semantic and contextual correlations from vast datasets, resulting in impressive sentiment analysis, machine translation, and conversation system outcomes. It emphasizes deep learning's importance to NLP research and applications.[8]. The paper's introduction discusses NLP's transition from statistical approaches to deep learning models. It highlights how RNNs, BERT, and GPT improve NLP tasks like part-of-speech tagging and named entity identification. Quantitative evaluation and real-world applications like sentiment analysis and machine translation are also covered. The introduction prepares for a

***Corresponding Author: Ayad Zedo Ismaeel,**
1Akre University for Applied Sciences, Technical College of Informatics, Department of Information Technology, Duhok, Iraq.

full assessment of deep learning's transformational NLP function.[9] describes how transformer architecture transforms NLP. Attention processes capture complicated linguistic patterns and transformers provide better contextual comprehension. The role of fine-tuning and transfer learning, adaptable pre-trained models, and multilingual task applications are highlighted. The introduction also discusses future research and issues, laying the groundwork for the in-depth study.[10] The introduction emphasizes deep learning's dramatic impact on NLP. Transformers, BERT, and GPT have improved sentiment analysis, machine translation, and speech recognition. It also covers data shortages, computing needs, and ethics. The introduction sets the stage for the paper's discussion of deep learning's influence and NLP's future.[11] describes how transformer-based deep learning architectures improved NLP. It emphasizes BERT and GPT as transformational tools and NLP's progress and model efficiency. The report filters relevant studies using inclusion and exclusion criteria in a systematic review. Comparing transformer applications, models, and datasets is another goal. Finally, the introduction acknowledges a significant NLP difficulty, laying the way for broader discussions throughout the article.[12] Transformers changed how machines perceive and produce words in natural language processing (NLP), as shown in the introduction. Their success relies on the self-attention mechanism, which captures deep contextual linkages in text. A multi-head attention and feed-forward layer design has improved model performance. Transformer models like BERT and GPT-3 set new NLP benchmarks, demonstrating their adaptability and power. The introduction prepares for a detailed look of how transformers have changed NLP.[13] stresses NLP's progress and ongoing issues. NLP is crucial to computer understanding of human language, because rule-based and statistical techniques have limits. The introduction credits deep learning, particularly RNNs, LSTMs, and transformers, for improving NLP performance. It also emphasizes model optimization to overcome over fitting and under fitting. The study aims to help academics by discussing current issues and viable solutions to improve NLP technology..[14] Since the mid-2000s, deep learning has transformed voice recognition and autonomous driving, thanks to academics like Geoffrey Hinton and Yoshua Bengio. It discusses vanishing gradients and how unsupervised pretraining solved deep network training problems. New technologies like Theano and TensorFlow have boosted progress. Deep learning in NLP is shifting toward sophisticated models like recursive and attention-based networks, according to the article. Key models and approaches are presented while analyzing how NLP needs have impacted current neural networks. The survey will cover key ideas, word embeddings, and NLP challenges.[15] The introduction describes Natural Language Processing (NLP) as a dynamic AI area that helps machines interpret and synthesize human language. It focuses NLP applications including machine translation, sentiment analysis, speech recognition, and question answering. Deep learning improves NLP tasks including text representation, categorization, sequence labeling, and creation. The relevance of embeddings, categorization kinds, language modeling, and style transmission is highlighted. The opening portrays deep learning as transforming current NLP.[16] emphasizes the necessity for efficient news categorization in an information-overloaded age. CNNs for local patterns, RNNs (including LSTMs) for sequential data, and Transformers (like BERT) for deep contextual semantics are introduced to demonstrate deep learning's transformational significance in text categorization. The report compares various approaches and discusses their merits, weaknesses, and news classification research directions.[17] Transformer-based language models like BERT and GPT2 have spurred AI and NLP advances, as noted in the introduction. These models have raised performance expectations, but the article criticizes a dearth of accessible, in-depth explanations in contemporary literature. A unified mathematical framework and visual aids and real-world examples are proposed to provide clear, thorough insights into neural language models. The study also underlines transformers' use in computer vision and time series analysis outside NLP.[18] focuses on the Transformer architecture and AI's rising significance on healthcare. Transformers, originally designed for NLP, are now used in clinical NLP, medical imaging, EHR analysis, and biosignal processing. The study discusses surgical guiding, adverse event prediction, diagnostic assistance, and medication synthesis. It also covers computational needs, interpretability, fairness, and ethics, preparing for a holistic healthcare Transformers analysis.[19] shows how deep learning has improved language understanding and generation in NLP. It discusses feature extraction using multi-layer neural networks with TensorFlow and PyTorch. Essential NLP tasks including word embedding, language modeling, and text creation are covered, along with text categorization and machine translation. The introduction emphasizes the relevance of optimization approaches for model performance, setting the way for a detailed discussion of strategies and applications.[20]

BACKGRAOUND THEORY

1. Foundations of Natural Language Processing and Deep Learning

Based on Hirschberg & Manning (2015)[21]

Natural Language Processing (NLP) has transitioned from rule-based systems to data-driven statistical models, ultimately progressing into the deep learning era. Hirschberg and Manning (2015) assert that this transition was propelled by enhanced computing capabilities, extensive annotated datasets, and progress in machine learning. Deep learning developed layered representations capable of learning hierarchical language structures autonomously, eliminating the need for manual feature engineering. Natural Language Processing (NLP) currently supports several applications such as speech recognition, sentiment analysis, and machine translation, driven by models like Recurrent Neural Networks (RNNs) and transformers that analyze sequential and contextual data.

2. Transfer Learning in Language Models

Based on Durrani et al. (2021)[22]

Transfer learning has emerged as a fundamental component of contemporary NLP, enabling pre-trained language models such as BERT, RoBERTa, and XLNet to generalize across many downstream tasks. Durrani et al. (2021) examine the effects of fine-tuning on the distribution of linguistic information among neural network layers. Their findings indicate that BERT preserves syntactic and semantic information in deeper levels, whereas models such as RoBERTa and XLNet transfer this knowledge to lower layers following fine-tuning. This redistribution impacts the model's interpretability and performance, as diagnostic classifiers indicate substantial structural alterations during adaptation.

3. Domain Adaptation in Neural Language Modeling

Based on Morioka et al. (2018)[23]

Domain adaptation tackles the issue of implementing neural language models in particular low-resource settings, such as multiparty dialogues. Morioka et al. (2018) introduced a dual-representation RNNLM architecture that distinguishes between domain-shared and domain-specific embeddings. This method enhances performance by preserving generic language characteristics while also capturing distinctive domain attributes. Experimental findings utilizing lecture and conversation corpora exhibit superior confusion and word mistake rates relative to conventional LSTM-based models.

4. Improving Language Models via Vocabulary Optimization

Based on Jascob (n.d.)[24]

Jascob presents a technique to improve language modeling by constructing "smarter" vocabularies using part-of-speech tagging and named entity recognition. This method categorizes words (e.g., time, place) rather than depending on the most common tokens, therefore diminishing ambiguity and enhancing generalization. The altered vocabulary structure results in reduced confusion and enhanced performance in neural models, particularly for infrequent or complex word expressions.

5. Adversarial Training for Robust Language Modeling

Based on Wang, Gong & Liu (2019)[25]

Wang et al. (2019) advocate for adversarial training as a regularization technique to mitigate over fitting in neural language models. Injecting adversarial noise into output embeddings during training enables the model to acquire more diversified and resilient representations. This method improves performance in tasks like as language modeling and machine translation, decreasing test perplexity and enhancing BLEU scores in benchmark datasets. The method's efficacy arises from its capacity to enhance variety without elevating computing complexity.

6. Neurosymbolic AI: Bridging Learning and Reasoning

Based on Kosaraju (2022)[26]

Kosaraju (2022) examines the constraints of contemporary deep learning models in addressing logical reasoning and abstraction. Neurosymbolic AI arises as a solution through the integration of deep neural networks with symbolic reasoning systems. This hybrid approach enhances performance on tasks necessitating intricate inference and contextual comprehension, including commonsense reasoning and low-resource inference. It connects statistical learning with cognitive-like logical thinking in natural language processing.

7. Critical Perspectives on Large Language Models

Based on Veres (2022)[27]

Veres (2022) challenges the assertion that large language models (LLMs) serve as comprehensive representations of human language. He contends that these models, developed using extensive corpora, are fundamentally "corpus models" rather than authentic representations of language skill. The author emphasizes the symbolic character of human language and asserts that generative grammar continues to provide crucial insights into syntax and semantics. Although LLMs excel at tasks like code production and translation, their performance does not reflect true comprehension of linguistic principles or cognitive thinking.

8. A Survey on Large Language Models for NLP

Based on Qin et al. (2024)[21]

Qin et al. (2024) present a comprehensive taxonomy of large language models (LLMs) in natural language processing (NLP), categorizing them into two paradigms: parameter-frozen (e.g., zero-shot and few-shot learning) and parameter-tuned (e.g., complete fine-tuning and parameter-efficient tuning). The survey examines the application of models such as GPT, BERT, and LLaMA in tasks like sentiment analysis, machine translation, and mathematical reasoning. The authors examine obstacles such as interpretability, ethical issues, and computing expenses, proposing avenues for future study in optimization methodologies and task-specific modifications.

LITERATURE REVIEW

1. Fathi and Shoja (2018)

Fathi and Shoja (2018) presented an extensive chapter on the function of deep neural networks (DNNs) in natural language processing (NLP). They delineated the historical evolution from rule-based systems to machine learning, culminating in deep learning methodologies. The authors highlighted that deep learning enables representation learning, allowing models to autonomously acquire abstract features from raw data, hence obviating the necessity for human feature engineering. The chapter thoroughly examines word vector representation, feed forward and convolutional neural networks, regularization methods, and their applications in language modeling. This study functions as a fundamental introduction to the ways in which deep learning revolutionizes crucial NLP tasks via hierarchical representations and data-driven learning methodologies.

2. Yang (n.d.)

Yang (n.d.) examined the diverse applications of deep learning in natural language processing, concentrating on optimization methodologies. The article delineated fundamental deep learning ideas, including neural network architectures and prominent frameworks such as TensorFlow and PyTorch. Critical NLP tasks—such as word embedding, text classification, and speech recognition—were thoroughly examined, alongside its execution utilizing sophisticated models like RNNs, LSTMs, GRUs, and Transformers. Additionally, Yang explored optimization methodologies like model compression, self-supervised learning, federated learning, and transfer learning. The work provides a comprehensive overview connecting model structures to their practical applications, highlighting performance optimization and flexibility.

3. Shi (2017)

Shi (2017) performed a comprehensive assessment of neural network language models (NNLMs), outlining several topologies and improvement techniques. The study evaluated the performance and application of feed forward, recurrent, and bidirectional RNNs across several NLP workloads. Enhancements such as significance sampling, lexical categories, and caching techniques were subjected to experimental evaluation. A critical analysis addressing the theoretical constraints of neural network language models regarding architecture and knowledge representation, positing that these models embody probabilistic approximations rather than authentic semantic or grammatical knowledge. The work is significant for its rigorous examination of the inherent limitations of NNLMs and its empirical validation of optimization methodologies.

4. Wilcox et al. (2019)

Wilcox et al. (2019) examined the capacity of neural language models, namely LSTMs, to describe hierarchical syntactic patterns. The study assessed models' capacity to handle long-distance dependencies using experimental setups featuring center-embedding and syntactic islands. The authors discovered that whereas LSTMs might mitigate and occasionally reinstate grammatical expectations, they frequently did not completely restore the original syntactic condition. This work is important for examining the cognitive validity of deep learning models in language challenges, investigating whether neural models truly acquire hierarchical grammar or depend on superficial pattern recognition.

5. Torfi et al. (2021)

Torfi et al. (2021) conducted a comprehensive assessment of developments in NLP propelled by deep learning. The article classified NLP tasks, such as part-of-speech tagging, named entity identification, and semantic parsing, and elucidated how deep neural networks improved these domains. It encompassed both supervised and unsupervised learning paradigms, emphasizing the significance of structures like RNNs, CNNs, and transformers. The review situated NLP within the wider AI framework and emphasized the significance of representation learning. This study serves as a comprehensive assessment that consolidates significant advancements and obstacles in the amalgamation of deep learning with language processing.

6. Michel (2021)

Michel (2021) tackled a significant real-world issue in natural language processing: distributional shift. The dissertation delineated several forms of data distribution changes (e.g., domain, label, or input distribution shifts) and assessed their effects on deep learning models. Michel presented techniques utilizing distributionally robust optimization (DRO) to improve the robustness of NLP systems. A significant contribution encompasses benchmarks and adaptive methodologies for sustaining performance when models face novel or fluctuating data. This discovery substantially enhances the resilience and generalizability of neural language models, particularly in dynamic contexts.

7. Min et al. (2021)

Min et al. (2021) examined current advancements in natural language processing (NLP) driven by huge pre-trained language models (PLMs), such as BERT, GPT, and T5. They classified PLM use into three paradigms: pre-training followed by fine-tuning, prompt-based learning, and text creation. The poll further addressed PLM-driven data production for training enhancement and model refinement. The authors analyzed advantages and disadvantages, encompassing computational expenses and ethical issues. This research is crucial for comprehending the current paradigm change in NLP and the emergence of PLMs as universal frameworks for diverse language problems.

8. Sordoni et al. (2023)

Sordoni et al. (2023) presented the notion of Deep Language Networks (DLNs), which aggregate numerous large language models (LLMs) and consider prompts as adjustable parameters. They employed variational inference to optimize prompts for multi-layer structures, using the intermediate outputs as latent variables. Their research indicated that DLNs can attain performance comparable to GPT-4, even when constructed from smaller models. This research is groundbreaking in conceptualizing LLMs as composable, learnable layers, hence facilitating the development of more modular and efficient NLP pipelines.

9. Gangar et al. (2023)

Gangar et al. (2023) created a neural machine translation (NMT) system employing the Transformer architecture for translating Hindi to English. To tackle the constraints posed by Hindi as a low-resource language, they utilized techniques such as back-translation and byte pair encoding (BPE) to improve data augmentation and vocabulary development. The authors attained a BLEU score of 24.53 by testing 10 setups, establishing a benchmark on the IIT Bombay corpus. This study demonstrates how deep learning may be customized for low-resource languages by strategic preprocessing and architectural decisions.

10. Saleh and Paquelet (2024)

Saleh and Paquelet (2024) provided a comprehensive mathematical tutorial on neural language models, emphasizing feed forward neural networks (FFNNs), recurrent neural networks (RNNs), and transformers. They provide a cohesive framework to comprehend model topologies, training objectives, and parameter configurations. The study analyzed BERT and GPT-2 models in detail, incorporating parameter formulae and original code implementations. The study also examined applications beyond natural language processing, including computer vision and time series analysis. This lesson is distinguished by its clarity and rigor, rendering it an exceptional instructional resource for comprehending the internal mechanics of contemporary language models.

11. Xu (2023)[28]

Xu (2023) highlights the revolutionary influence of deep learning in natural language processing (NLP), especially in improving computing efficiency and semantic comprehension of language-related tasks. The research underscores the enhancement of text categorization, named entity identification, and speech recognition through the utilization of feed forward neural networks that employ multi-layer representations. Xu elucidates the significance of unsupervised pre-training in feature extraction and classification tasks, facilitating enhanced generalization from extensive unstructured datasets. The study asserts that deep learning is a crucial technique for enhancing the practical application of NLP across many domains by augmenting the depth and contextual understanding of models.

12. Oralbekova et al. (2023)[29]

Oralbekova et al. (2023) present an extensive review on the progression of language models, encompassing traditional statistical models such as n-grams and Hidden Markov Models, as well as sophisticated neural architectures including BERT, GPT, and LLAMA. The authors analyze the structural design and optimization of these models, while tackling issues associated with agglutinative languages like Arabic and Turkish. They provide a comparison examination of prominent NLP libraries, such as TensorFlow, PyTorch, and Gensim, for usability and efficacy. This work functions as an essential reference for the contemporary landscape of language modeling tools and methodologies, particularly in multilingual and cross-lingual situations.

13. Zhang and Li (2024)[30]

Zhang and Li (2024) critically examine the intrinsic limits of contemporary brain language models in encapsulating profound semantic meaning. Although models like BERT and GPT-2 excel in several superficial NLP tasks, the authors contend that they do not comprehend propositional logic and non-linguistic communication purpose. They highlight the distinction between language structure and semantics, encouraging researchers to create hybrid systems that amalgamate logical thinking with external information repositories. This paper substantially advances the discussion on natural language understanding by identifying deficiencies in existing models' comprehension and suggesting pathways for theoretical and architectural improvements.

14. Weng et al. (n.d.)[31]

Weng et al. (n.d.) investigate the amalgamation of Deep Convolutional Neural Networks (DCNNs), generative adversarial networks (GANs), and machine learning methods to augment natural language processing (NLP) skills. The research illustrates enhanced precision in essential tasks like segmentation, text classification, and part-of-speech tagging. The authors demonstrate how GANs may improve text production and translation, while soft computing increases adaptation in unpredictable language contexts. The extensive architecture demonstrated illustrates how the integration of several AI methodologies may surmount conventional NLP constraints, yielding more resilient and effective models.

15. Goldstein et al. (2022)[32]

Goldstein et al. (2022) examine the relationship between the hierarchical architecture of deep language models and the temporal processing of language in the human brain. Utilizing ECoG recordings and contextual embeddings from GPT2-XL, they demonstrate a robust correlation between the model's layer-wise activations and neural responses in higher-order language regions. This neuro computational study demonstrates how deep language models reflect human understanding by progressively gathering contextual knowledge. The scientists find unique brain pathways triggered by uncertain words, indicating that although DLMs simulate human thinking, they still lack specific cognitive aspects.

16. Brownlee (2017)[33]

Brownlee (2017) offers a comprehensive approach for constructing deep learning models for natural language processing with Python. The book addresses fundamental ideas like word embeddings, recurrent neural networks, convolutional neural networks, and sequence modeling. The author guides readers through sentiment analysis, text categorization, and the development of neural language models using libraries such as Keras, employing structured examples. Brownlee underscores the "drop-in" capability of deep models to supplant conventional approaches and demonstrates how feature learning and end-to-end frameworks facilitate scalable NLP systems. This book serves as a significant resource for practitioners pursuing practical experience in NLP model creation.

17. Goyal et al. (2018)[34]

Goyal, Pandey, and Jain (2018) provide an extensive overview of constructing NLP systems with deep learning in Python. The book integrates theoretical principles with practical applications, including essential elements such RNNs, LSTMs, encoder-decoder architectures, and chatbot systems. The authors instruct readers on sentiment classification and neural language modeling, employing TensorFlow and Keras for model development. Their work is especially beneficial for novices seeking to comprehend the mathematical foundations and practical application of deep NLP models.

18. Yih, He, and Gao (2015)

Yih, He, and Gao (2015) provide a tutorial on deep learning and continuous representations for natural language processing, including both fundamental and advanced neural architectures. The seminar explores semantic embeddings, recurrent and convolutional models, as well as applications such as semantic parsing, information retrieval, and question answering. The authors examine structured models such as DSSM and highlight the need of end-to-end optimization. This lesson is distinguished by its equitable emphasis on theoretical foundations and practical implementation, rendering it indispensable for comprehending how neural embeddings and deep networks transform semantic comprehension in natural language processing.

19. He, Gao, and Deng (2014)[35]

Gao and Deng (2014) delineate the influence of deep learning on natural language processing and associated domains via a systematic tutorial. They introduce architectures like CNNs, RNNs, LSTMs, and deep structured semantic models, highlighting their efficacy in voice recognition, slot filling, and semantic search. A significant addition is their examination of semantic embeddings and sub-word modeling, which enhance contextual comprehension and linguistic generalization. The authors present a comprehensive review of deep architectures designed for NLP, emphasizing the interdisciplinary integration of language comprehension and machine learning.

20. Arkhangelskaya and Nikolenko (2023)[36]

Arkhangelskaya and Nikolenko (2023) present a comprehensive overview of deep learning in natural language processing, detailing the progression of neural architectures from initial feed forward and recurrent networks to sophisticated attention-based and memory-augmented models. The paper examines fundamental tasks like syntactic parsing, sentiment analysis, and machine translation, while presenting contemporary systems designed for natural language processing. The authors underscore the importance of unsupervised learning, word embeddings, and character-level models in attaining generalization. This study is especially significant for delineating the conceptual and technological evolution of NLP systems utilizing deep learning.

Deep Learning in NLP: How Neural Networks Are Changing Language Models

Deep learning has rapidly changed natural language processing (NLP) by bringing new ways to interpret, generate, and analyze human language. From early feed forward and recurrent models to modern transformer-based systems like BERT and GPT, neural network architectures have transformed language models in academic and industrial applications. These models outperform rule-based and statistical methods and provide scalable, data-driven solutions to complicated language challenges. A comparison table of current literature summarises major contributions to critically analyze how various brain architectures have affected modern NLP. Each entry describes the NLP job, dataset, preprocessing methodologies, assessment metrics, and approach strengths and weaknesses. This structured comparison illuminates deep learning's development, efficacy, and limitations in language modeling across settings.

Table1: "Natural Language Processing Deep Learning Comparison: Tasks, Methods, and Evaluation"

Ref	NLP Task / Application	Dataset	Preprocessing	Metrics Used	Strengths	Limitations
[37]	Language modeling, analogy tasks, categorization, and word embeddings	Gigawords, Common Crawl, Wikipedia2014, and synthetic corpora for examples.	One-hot encoding, co-occurrence matrices, SVD, skip-gram, window-based context, stop word elimination	Precision, cosine similarity, correlation with human assessments (intrinsic/extrinsic evaluation)	Describes word representation evolution from symbolic to vector-based models (word2vec, GloVe), illustrates analogy extraction, and examines optimization and training specifics.	Word ambiguity (polysemy), difficulties capturing subtleties in discrete representations, corpus quality and hyper parameters affect performance.
[38]	Speech recognition, conversation systems, summarization, machine translation, sentiment analysis, and text categorization	Wikipedia, news databases, and back-translated information (from a variety of sources, not all of which are listed)	Stopword elimination, synonym substitution, back translation, dynamic context embedding (BERT),	"Accuracy, performance benchmarks (e.g., on classification, translation tasks)" box	Covers several NLP problems using DL, different architectures (RNN, LSTM, GRU, Transformer), and optimization tactics including federated learning and model compression.	Some datasets or experiments not described; real-world deployment difficulties (e.g., resource restrictions, privacy) addressed but not resolved.
[39]	Language modeling, FNNLM, RNNLM, LSTM-RNNLM comparison, importance sampling, word classes, caching, BiRNN enhancements	Amazon Book and Electronic Reviews, Brown Corpus	Caching, back propagation, reversed sequences, field-specific training, word class clustering	Perplexity (PPL)	Analyzes NNLM designs, optimization methods, and model architecture and knowledge representation restrictions.	Lack of language comprehension, dynamic learning, and domain generalization; costly training and inflexible architecture
[40]	Central-embedding, filler–gap relationships, syntactic islands	Wikipedia, One Billion Word Benchmark, British National Corpus	Hand-crafted syntactic stimuli, psycholinguistic testing (surprisal analysis), embedding depth control, gap/filler combinations	Using linear mixed-effects models, surprising (inverse log-probability) statistics	Shows that LSTM and Transformer LMs may approach stack-like syntactic hierarchical behavior; LSTMs generalize more consistently under controlled experiments.	Expectation recovery imprecise; Transformer underperforms LSTM in some syntactic circumstances despite bigger capacity; hierarchical generalizations susceptible to lexical and structural changes.
[41]	Machine translation, sentiment analysis, parsing, SRL, QA, NER, coreference resolution, event extraction, text categorization	AG News, IMDB, SST, CoNLL-2003, SQuAD, WSJ-PTB, CNN/DM, WMT 2014, etc.	Character-level encoding, word embeddings, attention, encoder-decoder, contextualized embeddings, reinforcement learning	F1-score, accuracy, BLEU, ROUGE, METEOR, perplexity	Comprehensive study of tasks and models (CNN, RNN, LSTM, BERT, Transformers, GANs, RL); benchmarks and state-of-the-art.	Training issues (exposure bias, huge action space), labeled data dependence, long-range dependencies, generalization
[42]	Robust machine translation, adversarial evaluation, continuous learning, domain adaptation	Reddit, TED Talks, KFTT, JESC, MultiNLI, Amazon, toxicity detection datasets, MTNT, WMT 2015/2017	Contradictory input generation, BPE tokenization, language model filtering, multilingual embedding alignment, trajectory regularization	CVaR, divergence metrics (KL, Wasserstein), classification accuracy, ongoing learning score, BLEU, perplexity, proxy A-distance	Introduces MTNT dataset for noisy MT, tests adversarial resilience, and suggests parametric DRO and trajectory regularization to reduce catastrophic forgetting.	Some approaches require synthetic benchmarks; adversarial assessment is challenging to generalize; DRO uses many resources; adaptation is still subject to domain-specific quirks
[43]	Classification, labeling, structure prediction, QA, NER, emotion, IE, summary	Wikipedia, Books Corpus, Open WebText, Common Crawl, and domain-specific corpora for tuning	Pretraining (masked LM, autoregressive LM, seq2seq), fine-tuning, prompt-based learning, text creation, data augmentation, template-based tuning, and adapter layers	Accuracy, F1 score, BLEU, ROUGE, perplexity, and task-specific metrics	A comprehensive summary of PLM paradigms: Pre-train/fine-tune, prompt learning, text production; demonstrates effective fine-tuning, adaption approaches, and template design methodologies.	PLMs are resource-intensive; their efficacy depends on timely quality and template design; domain shift and scalability restrictions continue in low-resource circumstances.

[44]	Few-shot reasoning, categorization, linguistic logic, timely engineering optimization, modular NLP	BigBench-Hard, BBH, Mpqa, Trec, Subject, Disaster, Airline, Date, Navigation, and Logic.7	Prompt initialization, backward templates, latent variable modeling, chain-of-thought (CoT), variational inference, and template scoring	Accuracy (averaged), confidence intervals, ELBO for variational bounds	Introduces Deep Language Networks (DLN-1, DLN-2) with prompt-based modular LLM layers; achieves state-of-the-art performance via stacked prompting; integrates instruction tweaking, CoT, and in-context learning	Requires LLM API calls for both layers, complicated scoring and sampling, substantial variation in outcomes; optimization may stagnate or overfit on little datasets; Practical deployment may be resource demanding.
[45]	Language modelling, embeddings, transformer internals, CV, time series applications	NLP datasets, CV datasets (e.g., ViT), and time series benchmarks are mentioned.	Cross-entropy training, embedding tying, context windowing, positional encoding, transformer and RNN architecture description, one-hot embedding mapping	Cross-entropy loss, parameter counts, qualitative comparisons, embedding projection graphs	Unifies FFNN, RNN, and Transformer LMs under a similar mathematical framework; thorough anatomy with diagrams and from-scratch GPT2 code; applies NLP ideas to CV and TS.	Focused on architecture-level understanding; minimal empirical benchmarking; does not investigate training modifications or task-specific applications thoroughly.
[46]	POS tagging in Hindi and Nepali utilizing transfer and multitask learning	Hindi: LDC-IL corpus; Nepali: online Nepali POS corpus manually annotated.	FastText character-level embeddings, cross-lingual mapping, CNN-BLSTM-CRF architecture, gender and plurality tags, multitask setup	Accuracy, accuracy, recall, F1-score (POS tagging performance)	Cross-lingual embeddings (Hindi-Nepali) increase POS tagging; contrasts monolingual, mapped, and joint embeddings; gives deep architectural insights (CNN-BLSTM-CRF).	Multitask learning hurts Hindi POS performance; combined Hindi-Nepali training doesn't improve outcomes; data scarcity restricts low-resource languages.
[47]	Machine translation, named entity recognition, semantic analysis, conversation systems, text categorization	No dataset names; vague discussion of large-scale text data.	FNNs, word embeddings, Seq2Seq for MT, feature mapping, syntactic/semantic analysis, gradient descent optimization	Classification accuracy, semantic labeling quality, qualitative analysis	Clear explanation of NLP tasks (POS, MT, NER, text classification), use of standard neural models like FFNN and RNN, NLP pipeline, and model evolution.	Without actual evidence or benchmarks; Conceptual/theoretical; lacks comparison to Transformers.
[48]	Document classification, embedding analysis, multilingual NLP, sentiment analysis, language modeling	Different corpora: BookCorpus, Wikipedia, Reddit (WebText), Twitter (MARBERT), Turkish, Arabic, Uzbek, Korean.	Pretraining (MLM, NSP, permutation LM), embedding (Word2Vec, GloVe, FastText), model-specific tokenization (WordPiece, BPE), character n-grams, subword approaches, language-specific	Accuracy, F1-score, model parameter comparison, sentiment, NER, classification benchmarks	Comprehensive comparison of traditional (n-gram, HMM) and current (BERT, GPT, XLNet, FastText) models, focusing on agglutinative languages and multilingual tools; Word2Vec, CBOW, GloVe, Transformer-based PLM	High computational cost of models like BERT; contextual limits of classical models; resource needs for multilingual adaptation; challenges for low-resource and morphologically rich languages
[49]	LM semantic comprehension, formal language processing, logic recognition, and compositionality evaluation	Synthetic corpora with semantic transparency and logic-based limitations; no standard datasets.	To train symbols, use logical operators (¬, ∧, ∨), propositional logic corpora, transformer architecture modifications (e.g., UHAT, AHAT), hard/soft attention modeling, and SFST	Logical inference, AC0 class identification, compositional generalization, qualitative analysis model correctness	Comprehensive overview of LMs' semantics, logic, compositionality, and attention processes; compares Transformer versions and presents new benchmarks (e.g., UHAT versus AHAT).	Attention processes and positional encoding restrict transformers' logical symbol differentiation, generalization in complicated structures, semantic depth, and formal language recognition.
[50]	Text classification, sentiment analysis, NER, machine translation, text production, word segmentation, POS tagging	Movie review datasets, Wikipedia (chosen subsets), synthetic segmentation, classification, translation corpora	DCNN integration with CRF, GAN, soft computing, statistical models (HMM, CRF), back propagation, TF-IDF, semantic similarity score, NER entity labeling and assessment (macro/micro F1).	ROUGE, BLEU, segmentation precision, processing time (seconds), accuracy, recall	Improves standard NLP modules using DCNN, GAN, and ML; shows empirical advances in segmentation (+10% accuracy), recall (+4%), classification, translation, and efficiency. scalable, modular architecture	Complex computational complexity; GAN optimization and domain-specific tweaking; limited pretrained LM and large-scale benchmark exploration
[51]	Language neural encoding, GPT2-XL layers and speech comprehension brain activity	Custom spoken story ("Monkey in the Middle", NPR 2017) + 48-layer GPT2-XL contextual embeddings	ECG recordings, PCA embedding reduction, 10-fold CV, prediction success broken down by top-1 vs. top-5 accuracy.	Predicted and actual neuronal activity correlation, P-values, FDR correction, Pearson/Spearman correlation, bootstrapping	Displays a temporal mapping between layer-wise DLM processing (GPT2-XL) and cortical dynamics in IFG, aSTG, and TP; greatest alignment in high-order language regions; New encoding benchmarks	Results primarily confined to correct predictions; early sensory areas (mSTG) have poor mapping; GPT2-XL architecture differs from brain serial processing; no direct behavioral consequences indicated.

[52]	Sentiment analysis, POS tagging, language modeling, info retrieval	English, Chinese, Twitter, Turkish, and multilingual IR data	Feed forward (POS/parse trees); CNN with k-max pooling (sentiment); RNN/LSTM sequence modeling and POS tagging; multilingual training	Up to 90% parsing, 87% sentiment, 60–80% tagging, 150% quicker parsing	Investigates NLP model evolution from FFNN to CNN/LSTM, demonstrating task fit (e.g., CNN for sentiment, RNN for sequences) and multilingual applicability.	Lacks comparison with newer PLMs (BERT, GPT); accuracy in some tasks (IR) lower than SOTA; largely covers classical models.
[53]	QA, computer vision, text classification, emotion analysis, language modeling, machine translation, speech recognition, text summarizing	German-English translation, undefined NLP and voice datasets (literature-based).	Literature analysis; architectural breakdown (RNN, LSTM, GRU versus Transformer, BERT); self-attention vs. repetition; contextual situations	There are no empirical measures, however text-based benchmarks like processing speed and translation accuracy are explored.	Compares RNNs and Transformers side-by-side; links model capabilities to application settings (time-series versus parallelism); describes evolution, use cases, and future prospects.	No new empirical data; uses current research; no task-specific experimental validation; broad NLP overview, not task-specific
[54]	POS, NER, emotion, machine translation, content generation	MT: standard corpora; NER; generic NLP benchmark tasks; sentiment: BERT, LSTM datasets (unspecified).	CNN for local features, RNN/LSTM for sequences, Transformer/BERT for bidirectional context, CNN+RNN hybrid networks, pretraining, fine-tuning, pruning, quantization cell	AUC (model-by-task comparison table), Accuracy, F1 Score, Precision, Recall	Evolution from statistical to neural to transformer-based models; quant analysis of metrics and efficiency; hybrid models; real-world case studies (sentiment, MT, generation); table performance	No new experiments; datasets are cited; high review emphasis with low implementation detail; literature-synthesised performance
[55]	Low-resource language neural machine translation (NMT) (Kurdish Sorani)	KurdNet (WordNet), Tanzil (Quran), TED Talks, and Auta (different genres); a total of 199,375 Kurdish-English phrase pairings that are aligned	Tokenization, subword alignment, dictionary-based bilingual pairing, and combined datasets; GPU-trained over 100 epochs	Training time (GPU vs. CPU), BLEU (0.45), loss (5.109), and qualitative translation comparison (ergative structure vs. Google Translate)	First Kurdish NMT system based on transformers; excellent BLEU results on combined corpus; demonstrates resilience when faced with complicated sentence structures (e.g., ergative); open-source implementation	Short training data for good generalizability; CPU training inefficient; low BLEU on particular corpora; translation mistakes in idiomatic and lengthy phrases

This review employs a comparison table designed to systematically arrange and assess the contributions of selected research publications in deep learning for natural language processing (NLP). Each row of the table represents a unique scientific work, while each column delineates a certain component essential for thorough examination. The initial column delineates the NLP task or application examined in the study. This encompasses several language-related tasks, including sentiment analysis, machine translation, question answering, part-of-speech tagging, and named entity identification, facilitating task-based comparisons among models. The second column examines the dataset(s) utilized in the research, specifying whether commonly used benchmark datasets, domain-specific corpora, or low-resource language datasets were employed. This elucidates the study's scope, generalizability, and possible limits. The third column delineates the preprocessing procedures employed, which are essential for converting raw text into forms appropriate for model training. These approaches may encompass tokenization, sub word segmentation, stop word elimination, embedding creation, or architecture-specific methods like as fine-tuning or transfer learning. The fourth column delineates the assessment measures employed to gauge model performance. Depending on the job, these metrics may include accuracy, F1-score, BLEU, ROUGE, METEOR, perplexity, or surprisal. These metrics assess the efficacy and efficiency of the models in tackling the specified NLP challenges. The last two columns delineate the strengths and limitations noted in each research. The strengths encapsulate the fundamental contributions or advances, like architectural originality, enhanced accuracy, resilience to noisy data, or efficacy in low-resource environments. Conversely, the constraints recognize the obstacles or deficiencies that persist, including factors such as elevated computing datasets. Collectively, these categories offer a comprehensive and nuanced perspective of the contemporary research environment in deep learning-driven NLP.

Recommendations

The use of deep learning into natural language processing has produced significant outcomes; nonetheless, numerous key avenues might improve future research and practical applications. Initially, it is essential to emphasize model efficiency and scalability, particularly in low-resource settings where access to advanced technology and annotated data is constrained. Creating lightweight, energy-efficient models or employing methods such as model pruning, quantization, and knowledge distillation might enhance the accessibility of state-of-the-art NLP models.

Secondly, semantic comprehension and logical reasoning continue to be significant constraints in contemporary brain networks. Researchers want to investigate hybrid models that integrate neural networks with symbolic reasoning or neurosymbolic frameworks to bridge these gaps and advance toward enhanced language understanding.

Third, it is advised that further research emphasizes resilience and flexibility to domain transitions, hostile inputs, and real-world unpredictability. Techniques like domain adaptation, ongoing learning, and adversarial training can facilitate effective generalization of models outside their training settings.

Fourth, focus should be placed on multilingual and cross-lingual skills, particularly in supporting underrepresented and morphologically complex languages. This encompasses the development of enhanced tokenization algorithms, the utilization of cross-lingual embeddings, and the creation of inclusive datasets. The ethical ramifications and interpretability of huge language models require further scrutiny. Research should focus on creating transparent models that reduce bias, safeguard user privacy, and offer reasons for predictions, especially in critical applications.

Ultimately, evaluation approaches have to be diverse. Although accuracy and BLEU scores are valuable, it is essential to include additional metrics such as robustness, fairness, and human alignment to provide a more comprehensive assessment of NLP performance.

By implementing these ideas, the NLP community may enhance language models to be more potent, accountable, generalizable, and aligned with human communication requirements.

CONCLUSION

The use of deep learning into natural language processing has resulted in significant progress in computer comprehension and generation of human language. The comparative study of recent research indicates that neural architectures, especially transformer-based models, have established new standards in several NLP tasks. These models have exhibited outstanding performance in domains such as machine translation, sentiment analysis, and question answering, mostly owing to their capacity to grasp contextual linkages and their scalability with extensive datasets. Nonetheless, the examination also uncovers enduring difficulties. Constraints in extrapolating to low-resource languages, addressing logical reasoning, and maintaining computing efficiency persist as significant issues. Moreover, several cutting-edge models continue to face challenges with interpretability, ethical alignment, and domain applicability. As the discipline advances, focus must transition to creating models that are not only precise but also efficient, comprehensible, and resilient across many language situations. In conclusion, although deep learning has significantly transformed NLP, ongoing research is essential to rectify its existing deficiencies. Future trajectories indicate a shift towards more modular, scalable, and human-aligned language models that transcend superficial comprehension to achieve profound semantic and cognitive processing.

REFERENCES

[1] Y. Chai, L. Jin, S. Feng, and Z. Xin, “Evolution and advancements in deep learning models for natural language processing,” Applied and Computational Engineering, vol. 77, pp. 144–149, 2024.

[2] K. Pritam, “Advanced NLP Techniques for Sentiment Analysis and Text Summarization Using RNNs and Transformers,” Int. J. Res. Appl. Sci. Eng. Technol, vol. 12, pp. 1485–1494, 2024.

[3] A. Pandey, V. Ramesh, P. Mittal, M. Elangovan, and G. Deepa, “Exploring advancements in deep learning for natural language processing tasks,” The Scientific Temper, vol. 14, no. 04, pp. 1316–1323, 2023.

[4] D. Oralbekova, O. Mamyrbayev, M. Othman, D. Kassymova, and K. Mukhsina, “Contemporary approaches in evolving language models,” Applied Sciences, vol. 13, no. 23, p. 12901, 2023.

[5] J. Jia, W. Liang, and Y. Liang, “A review of hybrid and ensemble in deep learning for natural language processing,” arXiv preprint arXiv:2312.05589, 2023.

[6] Z. Yang, “Deep Learning Applications in Natural Language Processing and Optimization Strategies,” Journal of Modern Education and Culture, vol. 1, no. 2, 2024.

[7] A. Pandey, V. Ramesh, P. Mittal, M. Elangovan, and G. Deepa, “Exploring advancements in deep learning for natural language processing tasks,” The Scientific Temper, vol. 14, no. 04, pp. 1316–1323, 2023.

[8] Z. Xu, “Research on deep learning in natural language processing,” Advances in Computer and Communication, vol. 4, no. 3, 2023.

[9] Y. Chai, L. Jin, S. Feng, and Z. Xin, “Evolution and advancements in deep learning models for natural language processing,” Applied and Computational Engineering, vol. 77, pp. 144–149, 2024.

[10] F. Yvon, “Transformers in natural language processing,” in ECCAI Advanced Course on Artificial Intelligence, Springer, 2021, pp. 81–105.

[11] A. Pandey, V. Ramesh, P. Mittal, M. Elangovan, and G. Deepa, “Exploring advancements in deep learning for natural language processing tasks,” The Scientific Temper, vol. 14, no. 04, pp. 1316–1323, 2023.

[12] G. Tucudean, M. Bucos, B. Dragulescu, and C. D. Caleanu, “Natural language processing with transformers: a review,” PeerJComput Sci, vol. 10, p. e2222, 2024.

[13] L. Tunstall, L. Von Werra, and T. Wolf, Natural language processing with transformers. “ O’Reilly Media, Inc.,” 2022.

[14] J. Yao and B. Yuan, “Optimization Strategies for Deep Learning Models in Natural Language Processing,” Journal of Theory and Practice of Engineering Science, vol. 4, no. 05, pp. 80–87, 2024.

[15] E. O. Arkhangelskaya and S. I. Nikolenko, “Deep Learning for Natural Language Processing: A Survey,” Journal of Mathematical Sciences (United States), vol. 273, no. 4, pp. 533–582, Jul. 2023, doi: 10.1007/s10958-023-06519-6.

[16] P. Goyal, S. Pandey, and K. Jain, “Deep learning for natural language processing,” New York: Apress, 2018.

[17] G. Geddam, G. Dharmaraju, G. P. Kumar, M. B. Ketha, and A. Lakshmanarao, “Exploring Deep Learning Approaches for News Classification with CNNs, RNNs and Transformers,” in 2024 First International Conference on Innovations in Communications, Electrical and Computer Engineering (ICICEC), IEEE, 2024, pp. 1–5.

[18] M. Saleh and S. Paquelet, “Anatomy of neural language models,” arXiv preprint arXiv:2401.03797, 2024.

[19] S. Nerella et al., “Transformers and large language models in healthcare: A review,” ArtifIntell Med, p. 102900, 2024.

[20] Z. Yang, “Deep Learning Applications in Natural Language Processing and Optimization Strategies,” Journal of Modern Education and Culture, vol. 1, no. 2, 2024.

[21] J. Hirschberg and C. D. Manning, “Advances in natural language processing.” [Online]. Available: http://science.sciencemag.org/

[22] N. Durrani, H. Sajjad, and F. Dalvi, “How transfer learning impacts linguistic knowledge in deep NLP models?”

[23] ICASSP : 2018 IEEE International Conference on Acoustics, Speech and Signal Processing : proceedings : April 15-20, 2018, Calgary Telus Convention Center, Calgary, Alberta, Canada. Institute of Electrical and Electronics Engineers, 2018.

[24] B. Jascob, “Improving Language Model Performance with Smarter Vocabularies.”

[25] L. Deng and Y. Liu, “Deep Learning in Natural Language Processing.”

[26] D. Kosaraju, “Integrating Reasoning with Learning: Neurosymbolic AI’s Impact on NLP,” International Journal of Science and Healthcare Research, vol. 7, no. 4, pp. 360–365, Nov. 2024, doi: 10.52403/ijshr.20221051.

[27] C. Veres, “Large Language Models are Not Models of Natural Language: They are Corpus Models,” IEEE Access, vol. 10, pp. 61970–61979, 2022, doi: 10.1109/ACCESS.2022.3182505.

[28] Z. Xu, “Research on Deep Learning in Natural Language Processing,” Advances in Computer and Communication, vol. 4, no. 3, pp. 196–200, Jul. 2023, doi: 10.26855/acc.2023.06.018.

[29] D. Oralbekova, O. Mamyrbayev, M. Othman, D. Kassymova, and K. Mukhsina, “Contemporary Approaches in Evolving Language Models,” Dec. 01, 2023, Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/app132312901.

[30] X. Zhang and B. Li, “A Review of the Limitations of Language Models in NLP,” Applied and Computational Engineering, vol. 97, no. 1, pp. 102–107, Nov. 2024, doi: 10.54254/2755-2721/97/20241301.

[31] S. Wen-tau Yih, X. He, and J. Gao, “Deep Learning and Continuous Representations for Natural Language Processing Outline Part I. Background of neural network learning architectures Background: A review of deep learning theory and applications in relevant fields Advanced architectures for modeling language structure Common problems and concepts in language processing: Why deep learning is needed.”

[32] A. Goldstein et al., “The Temporal Structure of Language Processing in the Human Brain Corresponds to The Layered Hierarchy of Deep Language Models,” Oct. 2023, doi: 10.1101/2022.07.11.499562.

[33] J. Brownlee Disclaimer, “Deep Learning for Natural Language Processing Develop Deep Learning Models for Natural Language in Python Acknowledgements Copyright Deep Learning for Natural Language Processing →,” 2017.

[34] P. Goyal, S. Pandey, and K. Jain, Deep Learning for Natural Language Processing. Apress, 2018. doi: 10.1007/978-1-4842-3685-7.

[35] X. He, J. Gao, and L. Deng, “Deep Learning for Natural Language Processing and Related Applications,” 2014.

[36] E. O. Arkhangelskaya and S. I. Nikolenko, “Deep Learning for Natural Language Processing: A Survey,” Journal of Mathematical Sciences (United States), vol. 273, no. 4, pp. 533–582, Jul. 2023, doi: 10.1007/s10958-023-06519-6.

[37] E. Fathi and B. Maleki Shoja, “Deep Neural Networks for Natural Language Processing,” in Handbook of Statistics, vol. 38, Elsevier B.V., 2018, pp. 229–316. doi: 10.1016/bs.host.2018.07.006.

[38] Z. Yang, “Deep Learning Applications in Natural Language Processing and Optimization Strategies.”

[39] D. Shi, “A Study on Neural Network Language Modeling,” Aug. 2017, [Online]. Available: http://arxiv.org/abs/1708.07252

[40] E. Wilcox, R. Levy, and R. Futrell, “Hierarchical Representation in Neural Language Models: Suppression and Recovery of Expectations,” Jun. 2019, [Online]. Available: http://arxiv.org/abs/1906.04068

[41] A. Torfi, R. A. Shirvani, Y. Keneshloo, N. Tavaf, and E. A. Fox, “Natural Language Processing Advancements By Deep Learning: A Survey.”

[42] P. Michel, “Learning Neural Models for Natural Language Processing in the Face of Distributional Shift,” Sep. 2021, [Online]. Available: http://arxiv.org/abs/2109.01558

[43] B. Min et al., “Recent Advances in Natural Language Processing via Large Pre-Trained Language Models: A Survey,” Nov. 2021, [Online]. Available: http://arxiv.org/abs/2111.01243

[44] K. Gangar, H. Ruparel, and S. Lele, “Hindi to English: Transformer-Based Neural Machine Translation,” Sep. 2023, doi: 10.1007/978-981-33-4909-4_25.

[45] M. Saleh, “Anatomy of Neural Language Models.” [Online]. Available: https://github.com/b-com/AnatomyNLM.

[46] D. Yadav, “Exploring transfer learning for Deep NLP systems on rarely annotated languages.”

[47] Z. Xu, “Research on deep learning in natural language processing,” Advances in Computer and Communication, vol. 4, no. 3, 2023.

[48] D. Oralbekova, O. Mamyrbayev, M. Othman, D. Kassymova, and K. Mukhsina, “Contemporary Approaches in Evolving Language Models,” Dec. 01, 2023, Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/app132312901.

[49] X. Zhang and B. Li, “A Review of the Limitations of Language Models in NLP,” Applied and Computational Engineering, vol. 97, no. 1, pp. 102–107, Nov. 2024, doi: 10.54254/2755-2721/97/20241301.

[50] C. Weng et al., “A Thorough Investigation into the Application of Deep CNN for Enhancing Natural Language Processing Capabilities.”

[51] A. Goldstein et al., “The Temporal Structure of Language Processing in the Human Brain Corresponds to The Layered Hierarchy of Deep Language Models,” Oct. 2023, doi: 10.1101/2022.07.11.499562.

[52] Z. Zhao, “Research on the evolution of neural networks models in NLP,” Applied and Computational Engineering, vol. 4, no. 1, pp. 690–695, May 2023, doi: 10.54254/2755-2721/4/2023385.

[53] X. Li, “Comparative analysis and prospect of RNN and Transformer,” Applied and Computational Engineering, vol. 75, no. 1, pp. 178–184, Jul. 2024, doi: 10.54254/2755-2721/75/20240535.

[54] Y. Chai, L. Jin, S. Feng, and Z. Xin, “Evolution and advancements in deep learning models for Natural Language Processing,” Applied and Computational Engineering, vol. 77, no. 1, pp. 144–149, Jul. 2024, doi: 10.54254/2755-2721/77/20240674.

[55] S. Badawi, “Transformer-Based Neural Network Machine Translation Model for the Kurdish Sorani Dialect,” UHD Journal of Science and Technology, vol. 7, no. 1, pp. 15–21, Jan. 2023, doi: 10.21928/uhdjst.v7n1y2023.pp15-21.

****