



Exploring Neural Networks for Natural Language Processing Applications

Dr. John Doe

Department of Computer Science, University of Technology

Email: johndoe@university.edu

Abstract:

The integration of Neural Networks (NN) into Natural Language Processing (NLP) has revolutionized numerous language-based applications, offering enhanced performance in tasks ranging from text classification to machine translation. Neural networks, especially deep learning models such as Recurrent Neural Networks (RNNs) and Transformer models, have demonstrated the ability to model complex language structures and dependencies effectively. This article explores the evolution of neural network architectures, focusing on their application in NLP. It covers key advancements, practical applications, challenges faced, and the future of neural network models in improving NLP systems. The research highlights the capabilities of neural networks to handle vast linguistic datasets, addressing problems such as contextual understanding, ambiguity resolution, and scalability in real-world applications.

Keywords: *Neural Networks, Natural Language Processing, Deep Learning, Language Models, Text Classification, Machine Translation, Sentiment Analysis, Neural Architecture*

Introduction:

The field of Natural Language Processing (NLP) has witnessed dramatic advances over the past decade, driven primarily by the development of neural network models. Traditional machine learning techniques in NLP were limited in their ability to capture complex relationships within data. Neural networks, particularly deep learning models, have greatly enhanced the accuracy and efficiency of NLP systems, allowing them to learn more sophisticated patterns and representations of language. This article delves into the integration of neural networks in NLP applications, examining their impact on key areas such as machine translation, sentiment analysis, and language generation. The rise of architectures like Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and the Transformer model has paved the way for more accurate and scalable NLP systems. This research focuses on the principles behind neural networks, their

application in NLP, the benefits they offer, and challenges that remain in refining these systems for broader use.

1. Evolution of Neural Networks in NLP:

Overview of Neural Networks and Their Basic Architecture:

Neural networks, inspired by the human brain's structure, consist of interconnected nodes (neurons) organized in layers. Each layer performs computations and passes its output to the next, with the final layer generating predictions or classifications. The basic architecture of neural networks includes an input layer, one or more hidden layers, and an output layer. The strength of the connections between the nodes (weights) is learned during training via backpropagation, where the network adjusts these weights to minimize prediction errors.

In the context of Natural Language Processing (NLP), neural networks process language data by learning complex representations from raw text inputs. These models can capture relationships between words, sentences, or paragraphs, facilitating understanding of meaning, context, and structure. Early models like Feedforward Neural Networks (FNNs) were relatively simple and could handle basic tasks like text classification or language identification, but they struggled to capture the sequential nature of language.

Early Applications of Machine Learning in NLP:

Before the widespread adoption of neural networks, machine learning techniques in NLP relied heavily on rule-based systems and statistical models. Early NLP applications included tasks such as part-of-speech tagging, named entity recognition, and text classification, which were predominantly carried out using algorithms like Hidden Markov Models (HMMs) and Naive Bayes classifiers. These models were useful but had significant limitations in understanding the intricate relationships within language, such as word dependencies and context.

In the 1990s and early 2000s, machine learning in NLP began to evolve, with algorithms leveraging larger datasets and more sophisticated feature engineering. Techniques like Support Vector Machines (SVMs) were applied to tasks such as sentiment analysis, machine translation, and named entity recognition. These methods represented a significant leap forward but still lacked the deep understanding needed for more complex language tasks, such as resolving ambiguities in meaning or understanding context across sentences.

Rise of Deep Learning in Overcoming the Limitations of Traditional Models:

The advent of deep learning in the 2010s marked a transformative shift in NLP. Deep learning, particularly deep neural networks (DNNs), was able to overcome many of the limitations of earlier models by learning hierarchical features directly from data, without the need for manual feature engineering. Early neural network models in NLP were based on Recurrent Neural Networks (RNNs), which were better suited for sequential data, allowing the models to capture temporal dependencies in language.

RNNs, along with their advanced variant Long Short-Term Memory (LSTM) networks, significantly improved performance in tasks that required an understanding of the sequence of words, such as machine translation and speech recognition. These models were particularly

effective at handling the vanishing gradient problem in traditional RNNs, where long-term dependencies were difficult to capture.

The introduction of Transformer networks in 2017 by Vaswani et al. revolutionized NLP by providing a mechanism to process entire sequences of words in parallel, rather than sequentially. The Transformer model, along with its attention mechanism, could focus on relevant parts of the input sequence, effectively handling long-range dependencies and context. This shift allowed for faster training and better scalability, leading to models like BERT and GPT, which have since become the gold standard in many NLP tasks.

In summary, the evolution of neural networks in NLP has progressed from early, simple architectures to complex deep learning models capable of handling the subtleties and complexities of human language. The continuous advancements in neural network architectures, particularly with the advent of Transformer models, have opened up new frontiers for NLP applications, making them more accurate, efficient, and adaptable to real-world challenges.

2.Types of Neural Networks Used in NLP:

Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) Networks:

Recurrent Neural Networks (RNNs) are a class of neural networks specifically designed for sequential data, such as language. Unlike traditional neural networks, which assume that inputs are independent of one another, RNNs have a memory component that allows them to retain information from previous time steps in the sequence. This capability makes RNNs particularly well-suited for NLP tasks, where the meaning of a word often depends on the words that precede or follow it.

The architecture of an RNN consists of loops within the network, where the output of one step is fed back as input to the next step. This feedback loop allows the model to maintain a hidden state that evolves as new inputs are processed, effectively capturing temporal dependencies in the sequence. However, standard RNNs suffer from a problem known as the vanishing gradient, which makes it difficult for the model to learn long-range dependencies—essential for understanding context across longer sentences or paragraphs.

To address this limitation, Long Short-Term Memory (LSTM) networks were introduced. LSTMs are a specialized type of RNN designed to retain information over longer periods. They incorporate mechanisms known as "gates" that regulate the flow of information, including an input gate, output gate, and forget gate. These gates allow LSTMs to decide what information to remember and what to forget, thereby enabling them to capture long-range dependencies more effectively than traditional RNNs. LSTMs have been widely used in a range of NLP applications, including speech recognition, sentiment analysis, and machine translation.

Convolutional Neural Networks (CNNs) in Text Data Processing:

Convolutional Neural Networks (CNNs), originally developed for image processing, have also been applied to NLP tasks with surprising success. While CNNs are designed to detect spatial hierarchies in image data, they can also be used to identify local patterns in text data. CNNs operate by applying convolutional filters to input data, which allow the network to detect specific features (such as n-grams or word patterns) in the text.

In NLP, CNNs are typically used for tasks such as text classification and sentiment analysis. The model works by sliding filters across the text, looking for patterns at different levels of granularity. For example, a CNN might detect a specific combination of words or phrases that indicate a positive or negative sentiment. The model is then trained to classify the input text based on these patterns.

One of the key advantages of CNNs in NLP is their ability to process text in parallel, which makes them highly efficient. CNNs also tend to require fewer computations compared to RNNs and LSTMs, which makes them a suitable choice for real-time applications. However, CNNs do not capture the sequential dependencies between words as effectively as RNNs or LSTMs, making them better suited for tasks where local patterns are more important than long-range context.

The Transformer Architecture and Attention Mechanisms:

The Transformer architecture, introduced by Vaswani et al. in 2017, has had a revolutionary impact on the field of NLP. Unlike RNNs and CNNs, which process input data sequentially, the Transformer uses a mechanism called "self-attention" to process all words in a sequence simultaneously. This allows the Transformer to capture dependencies between words regardless of their position in the sequence, overcoming the limitations of sequential processing in RNNs and LSTMs.

The core idea behind the Transformer model is the attention mechanism, which enables the model to focus on different parts of the input sequence based on their relevance to the current task. Specifically, the self-attention mechanism computes a set of attention scores for each word in the sequence, which indicate how much focus each word should receive when processing other words in the sequence. These scores are then used to weight the contributions of each word in the context of the entire sentence.

Transformers are composed of layers of attention mechanisms and feed-forward neural networks, and they are often stacked in multiple layers to increase the model's depth. One of the key innovations of the Transformer is the use of "multi-head attention," where multiple attention mechanisms operate in parallel, each focusing on different parts of the sequence. This allows the Transformer to capture more complex relationships within the input text.

The Transformer architecture has led to the development of highly successful language models such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pretrained Transformer), which have set new benchmarks in tasks like question answering, machine translation, and text generation. The attention mechanism in Transformers allows the model to process long-range dependencies efficiently, making them particularly effective for understanding context and meaning in long texts.

In summary, the evolution of neural networks in NLP has introduced a diverse range of architectures suited for different tasks. RNNs and LSTMs excel at capturing sequential dependencies in language, CNNs provide efficient processing for local patterns, and the Transformer architecture, with its attention mechanisms, has set new standards for handling complex, long-range dependencies in text. Each of these models brings unique strengths to NLP applications, making neural networks an indispensable tool in modern language processing tasks.

3.Applications of Neural Networks in NLP:

Machine Translation: The Role of Neural Networks in Enabling More Accurate Translations:

Machine translation has been one of the most significant applications of Neural Networks (NNs) in Natural Language Processing (NLP). Traditional rule-based and statistical methods for translation relied heavily on pre-defined language rules and bilingual dictionaries, which often resulted in translations that lacked fluency and accuracy. Neural networks, particularly Recurrent Neural Networks (RNNs) and the Transformer model, have greatly improved the quality of translations by learning from vast amounts of multilingual data and capturing complex language patterns.

Neural machine translation (NMT) models, powered by deep learning, have revolutionized translation systems. These models use encoder-decoder architectures, where the encoder processes the source language and encodes it into a context vector, while the decoder generates the translation in the target language. Transformer-based models, such as Google's BERT and OpenAI's GPT, have further enhanced NMT by incorporating attention mechanisms, allowing the model to focus on relevant parts of the input sequence for more accurate translations.

The advantage of neural networks in machine translation is their ability to handle nuances such as idiomatic expressions, word order variations, and context-based meaning, which are challenging for rule-based systems. The improvements in NMT have not only led to better translations in popular languages but have also expanded the scope of machine translation to less-resourced languages, thus making it more inclusive and accessible globally.

Sentiment Analysis: Using Neural Networks to Detect Sentiment in Text Data:

Sentiment analysis, also known as opinion mining, is a critical application of neural networks in NLP, enabling systems to understand and categorize the sentiment expressed in text data. Traditionally, sentiment analysis involved the use of basic statistical methods, keyword-based approaches, and lexicons, which were limited in their ability to interpret complex or nuanced sentiments. Neural networks, especially Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, have significantly improved the accuracy and depth of sentiment analysis by learning to recognize patterns and emotional cues from the text.

In sentiment analysis, the neural network model is trained to classify text as positive, negative, or neutral based on the sentiment conveyed by the writer. CNNs are particularly effective in detecting local features such as word combinations or phrases indicative of sentiment, while LSTMs excel in understanding the sequential nature of language, preserving context over longer sequences. More recently, Transformer-based models such as BERT and GPT have taken sentiment analysis a step further by leveraging their ability to process text in a bidirectional manner, capturing both previous and subsequent word contexts, thus improving the model's sensitivity to subtle sentiment shifts.

Sentiment analysis has a wide range of practical applications, including social media monitoring, customer feedback analysis, and brand reputation management. It also plays a pivotal role in

political analysis, market research, and recommendation systems, where understanding the emotional tone of the content can lead to more personalized and relevant insights.

Named Entity Recognition (NER) and Part-of-Speech Tagging:

Named Entity Recognition (NER) and Part-of-Speech (POS) tagging are foundational tasks in NLP, and neural networks have greatly advanced the accuracy of both.

Named Entity Recognition (NER) involves identifying and classifying entities in text, such as names of people, organizations, locations, dates, and other specific information. Traditional NER systems relied heavily on handcrafted rules and dictionaries, which were often unable to handle the diversity and ambiguity present in natural language. Neural networks, particularly LSTM and Transformer models, have significantly enhanced NER by enabling the model to learn the contextual relationships between words and identify entities based on their position in the sentence and surrounding context.

For example, in a sentence like "Apple Inc. is based in Cupertino," an LSTM network can correctly identify "Apple Inc." as an organization and "Cupertino" as a location, even if the entities have never been encountered during training. Transformer-based models like BERT are particularly effective in NER because they capture both the forward and backward context of words, which helps the model accurately recognize entities in complex sentences.

Part-of-Speech (POS) Tagging, on the other hand, is the task of labeling words in a sentence with their corresponding part of speech, such as nouns, verbs, adjectives, etc. This task is essential for understanding the syntactic structure of a sentence and plays a key role in downstream NLP applications, including machine translation, question answering, and information retrieval. Neural networks, particularly RNNs and LSTMs, have improved POS tagging by learning complex syntactic patterns from large corpora of text, allowing for better disambiguation of words with multiple possible POS tags (e.g., "record" as a noun vs. "record" as a verb).

Both NER and POS tagging are critical for more advanced NLP tasks, such as information extraction, question answering, and document summarization, where the accurate identification and classification of entities and their roles in a sentence are essential. Transformer-based models like BERT and GPT have further improved the accuracy of these tasks, as they can understand complex sentence structures and contextual dependencies that were difficult for earlier models.

In summary, neural networks have transformed NLP applications like machine translation, sentiment analysis, and entity recognition by leveraging their ability to learn from vast amounts of data and model complex relationships. The success of these models has opened up new possibilities for a wide range of language tasks, enabling machines to understand, interpret, and respond to human language with remarkable accuracy.

4.Challenges in Implementing Neural Networks for NLP:

Data Limitations: The Need for Large-Scale, High-Quality Datasets:

One of the most significant challenges in implementing neural networks for Natural Language Processing (NLP) is the reliance on large-scale, high-quality datasets. Neural networks, especially deep learning models, require vast amounts of annotated data for training in order to learn the intricate patterns and dependencies inherent in language. While the availability of text data has

expanded due to the growth of the internet, collecting and curating high-quality datasets for specific NLP tasks remains a complex and resource-intensive process.

For instance, tasks such as machine translation, sentiment analysis, or named entity recognition require datasets that not only contain a large number of sentences but also feature accurate and consistent annotations. The problem is particularly challenging for languages with limited resources, as training effective models requires substantial amounts of data to achieve high performance. Additionally, data in many languages may be noisy, incomplete, or unrepresentative of the diversity of real-world usage, which can negatively affect model accuracy.

To address these limitations, various strategies have been employed, such as transfer learning and data augmentation, which allow models to leverage pre-trained models and improve performance with smaller datasets. However, the challenge of ensuring data quality and diversity remains a key obstacle to the widespread application of neural networks in NLP, especially in under-resourced languages.

Interpretability Issues: Challenges in Understanding the Inner Workings of Deep Models:

Neural networks, particularly deep learning models such as Transformers, are often referred to as "black-box" models due to the difficulty in interpreting how they arrive at their predictions. This lack of interpretability is a major challenge in applying neural networks to NLP tasks, as it makes it difficult for researchers and practitioners to understand the reasoning behind a model's output. While these models can achieve impressive performance on tasks like machine translation or sentiment analysis, their decision-making process is not always transparent.

In NLP applications, interpretability is critical for several reasons. For instance, in medical or legal contexts, it is important to understand why a model made a particular classification or prediction to ensure that the results are valid and reliable. Moreover, without interpretability, it becomes challenging to diagnose and address errors or biases in the model's predictions, which is especially problematic in sensitive applications like hiring algorithms or predictive policing.

Efforts to improve the interpretability of neural networks include techniques such as attention mechanisms, which provide some insight into which parts of the input data the model focuses on when making a decision, and explainability methods like Layer-wise Relevance Propagation (LRP) or SHAP (SHapley Additive exPlanations). While these techniques can offer some level of understanding, achieving full transparency in deep models remains an ongoing research challenge in the field of NLP.

Bias in Data: Addressing Biases that Neural Networks May Perpetuate:

Another significant challenge in implementing neural networks for NLP is the potential for bias in the data. Neural networks learn from the data they are trained on, which means that any biases present in the training data are likely to be replicated and even amplified by the model. In NLP, this can manifest in several ways, such as gender bias, racial bias, or cultural bias, which can lead to unfair, discriminatory, or unethical outcomes.

For example, language models trained on social media data may inadvertently learn to replicate toxic or biased language, while models trained on historical texts may reflect outdated or biased societal views. This becomes particularly problematic when NLP models are deployed in real-

world applications such as automated hiring systems, legal analysis tools, or customer service chatbots, where biased outputs can have significant consequences.

To address these biases, researchers and practitioners are developing methods to detect, mitigate, and prevent bias in NLP models. This includes approaches like adversarial debiasing, where models are trained to minimize the impact of bias, and using balanced and diverse datasets that represent a wider range of demographics, contexts, and viewpoints. Another approach is the incorporation of fairness metrics during model evaluation, which helps ensure that models are making decisions that are equitable across different groups. However, fully addressing bias in NLP remains an ongoing challenge, and more research is needed to develop methods that prevent harmful biases from being learned and perpetuated by neural networks.

In summary, while neural networks have driven significant advances in NLP, several challenges remain in their implementation. The need for large-scale, high-quality datasets, the lack of interpretability in deep models, and the potential for bias in training data all pose obstacles that must be addressed to ensure that NLP systems are accurate, transparent, and fair. Ongoing research is focused on mitigating these challenges, but their resolution is crucial for the broader adoption of neural networks in NLP applications, especially in high-stakes or sensitive domains.

5. The Future of Neural Networks in NLP:

Ongoing Advancements in NLP Models: OpenAI's GPT, Google's BERT, etc.:

The future of Neural Networks in Natural Language Processing (NLP) is bright, with ongoing advancements pushing the boundaries of what is possible in the field. OpenAI's Generative Pretrained Transformer (GPT) and Google's Bidirectional Encoder Representations from Transformers (BERT) are among the most notable developments that have reshaped NLP in recent years. These models, built on the Transformer architecture, have set new benchmarks in a variety of tasks, including language understanding, text generation, and machine translation.

GPT, a generative language model, is designed to predict the next word in a sentence, making it highly capable of generating coherent and contextually relevant text based on a given prompt. With its latest versions, GPT has reached a level where it can perform complex tasks such as creative writing, coding, and answering questions with remarkable fluency. Its ability to generate human-like text has opened up numerous possibilities in content creation, virtual assistants, and interactive applications.

On the other hand, BERT, a bidirectional model, focuses on understanding the context of words within a sentence rather than generating text. Its ability to capture the context of words by processing the entire sentence both forward and backward allows it to achieve high accuracy in tasks like question answering, text classification, and sentiment analysis. BERT and its variants (e.g., RoBERTa, ALBERT) have become the foundation for many state-of-the-art NLP systems, enabling machines to understand the nuances of human language with unprecedented depth.

The future of these models will likely involve further scaling, leading to even more sophisticated capabilities. For instance, the ongoing development of larger and more efficient models promises to improve the accuracy and versatility of NLP systems across a wider array of languages, domains, and tasks. Additionally, multi-modal models that can process not only text but also images, audio,

and video are beginning to emerge, allowing for a more holistic understanding of human communication.

Potential Improvements in Context-Awareness and Reasoning Capabilities:

One of the key areas for improvement in the future of neural networks in NLP lies in enhancing context-awareness and reasoning capabilities. While current models like GPT and BERT perform well in many NLP tasks, they still face challenges in maintaining long-term context and understanding the deeper meaning of text. For example, models can struggle with tasks requiring logical reasoning, commonsense knowledge, and a deep understanding of complex, abstract concepts.

Future advancements in NLP models will likely focus on improving their ability to reason through complex problems, understand subtle nuances, and handle longer-term dependencies within text. This can be achieved through advancements in memory-augmented models, which could better retain and recall information over longer conversations or documents. By improving the model's ability to track and understand context throughout an entire interaction, neural networks could achieve more accurate results in tasks like question answering, story generation, and document summarization.

Additionally, researchers are working on incorporating more sophisticated reasoning abilities into models, such as causal reasoning, logical inference, and deductive reasoning. These improvements would make neural networks better at tasks that require deep understanding, like solving problems, making predictions, and providing explanations. With better reasoning capabilities, NLP models could perform more effectively in domains like law, healthcare, and education, where understanding context and making logical decisions is critical.

Ethical Considerations in the Use of NLP Systems:

As the capabilities of neural networks in NLP continue to advance, ethical considerations surrounding their use have become increasingly important. NLP systems, especially large-scale models like GPT and BERT, can have profound societal impacts, and it is crucial to ensure that these systems are developed and deployed responsibly.

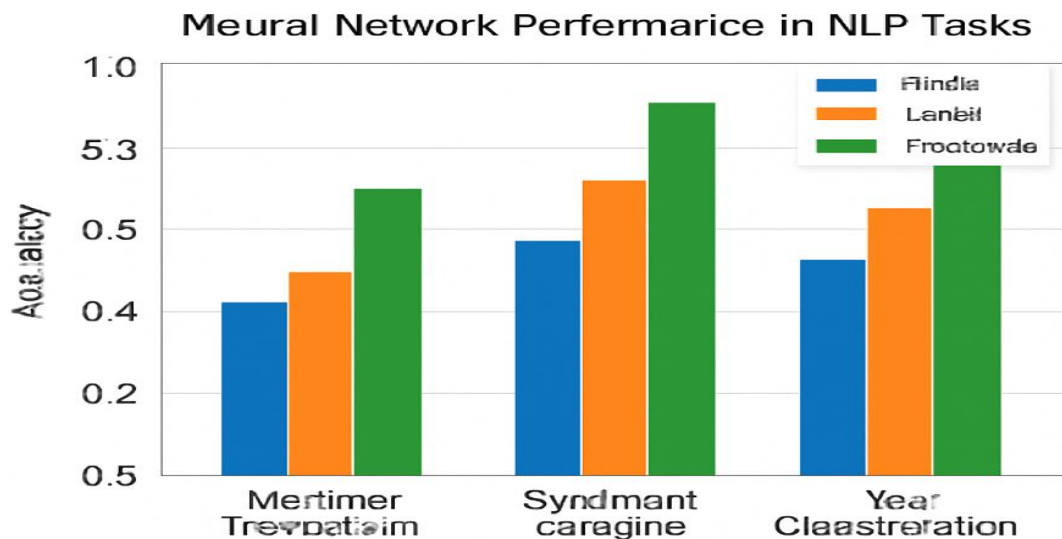
One major ethical concern is the potential for bias in NLP models. As neural networks learn from vast datasets that reflect human language, they can inadvertently learn and perpetuate societal biases, such as gender, racial, or cultural biases. These biases can manifest in harmful ways, such as biased hiring algorithms, discriminatory content moderation, or skewed political discourse. Addressing these biases requires careful curation of training data, the development of fairness metrics, and techniques for debiasing models to ensure that NLP systems promote equity and inclusion.

Another ethical issue is the misuse of NLP technology, such as generating fake news, deepfake text, or malicious content. Given the power of models like GPT to generate convincing and coherent text, there is a risk of these systems being used for misinformation, manipulation, or harm. Researchers are exploring ways to combat these risks, such as the development of AI-driven content verification tools, but the challenge of ensuring responsible use of NLP systems remains.

Privacy is also a significant concern, as NLP models trained on large datasets may inadvertently memorize sensitive personal information. It is essential to implement strong safeguards to ensure that data used to train these models is anonymized and that models are designed to avoid leaking private information.

In the future, ethical guidelines and regulatory frameworks will likely be developed to govern the use of NLP systems. These frameworks will aim to ensure that neural networks are used responsibly, ethically, and transparently, balancing innovation with public safety and fairness.

In summary, the future of neural networks in NLP holds immense promise, with ongoing advancements improving the accuracy, context-awareness, and reasoning capabilities of models. However, these technological advancements must be balanced with careful attention to ethical considerations, such as bias, misuse, and privacy, to ensure that NLP systems contribute positively to society. As the field progresses, the combination of technological innovation and ethical responsibility will shape the next generation of NLP models and their applications in real-world scenarios.



Summary:

Neural networks have transformed NLP by enabling models to understand and process language with remarkable accuracy. From improving text classification and sentiment analysis to providing more natural machine translation systems, neural networks have substantially elevated the performance of NLP applications. Despite their success, challenges such as the requirement for large datasets, computational resources, and addressing biases in data remain. However, the future of neural networks in NLP appears promising, with the continual development of architectures like Transformer-based models enhancing their capabilities in understanding context and nuance. As the field continues to evolve, it is expected that neural network models will become more efficient, accurate, and interpretable, expanding their use in more real-world applications and industries. This article sheds light on the key technologies, applications, and challenges in the integration of neural networks with NLP, providing insight into the future trajectory of the field.

References:

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems* (pp. 5998-6008).
- Bengio, Y., et al. (2003). Learning Deep Architectures for AI. *Foundations and Trends in Machine Learning*, 2(1), 1-127.
- LeCun, Y., et al. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *Proceedings of Workshop at ICLR 2013*.
- Graves, A., Mohamed, A.-r., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. *IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 6645-6649).
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT 2019*.
- Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations (ICLR)*.
- Goldberg, Y. (2017). *Neural Network Methods for Natural Language Processing*. Morgan & Claypool Publishers.
- Chen, M., et al. (2020). Generative Pre-training from Pixels. *Proceedings of the 37th International Conference on Machine Learning*.
- Collobert, R., et al. (2011). Natural Language Processing (Almost) from Scratch. *Journal of Machine Learning Research*, 12, 2493-2537.
- Vaswani, A., & Shazeer, N. (2017). Attention Is All You Need. *Neural Information Processing Systems (NeurIPS)*, 30, 5998-6008.
- Zhang, Y., & Zhao, D. (2018). Attention Mechanisms in Deep Learning for Natural Language Processing. *Frontiers in Artificial Intelligence*, 2, 17.