

ADVANCES IN NATURAL LANGUAGE PROCESSING: A SURVEY OF TECHNIQUES

Alladi Deekshith

Sr. Software Engineer and Research Scientist

Department of Machine Learning, USA

alladideekshith773@gmail.com

Abstract:

Natural Language Processing (NLP) has witnessed remarkable advancements over the past few decades, transforming the way machines understand and interact with human language. This survey provides a comprehensive overview of the key techniques and methodologies that have propelled the field forward, highlighting both traditional approaches and contemporary innovations. We begin by discussing foundational NLP techniques such as tokenization, part-of-speech tagging, and syntactic parsing, which laid the groundwork for understanding language structure. The evolution of statistical methods, including Hidden Markov Models (HMMs) and Conditional Random Fields (CRFs), is explored as a significant advancement in the probabilistic modeling of language. The survey then delves into the rise of machine learning approaches, particularly supervised and unsupervised learning, which have revolutionized various NLP tasks such as sentiment analysis, named entity recognition, and machine translation. We examine the impact of deep learning, focusing on architectures like Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), and Convolutional Neural Networks (CNNs) that have enabled significant improvements in performance across a range of applications. The introduction of transformer models, particularly the attention mechanism and BERT (Bidirectional Encoder Representations from Transformers), marks a paradigm shift in how contextual information is captured, leading to state-of-the-art results in numerous NLP benchmarks. In addition to technical advancements, the survey addresses the challenges that persist in NLP, including issues of bias in language models, the necessity for large annotated datasets, and the importance of explainability in AI systems. We discuss ongoing research efforts aimed at mitigating these challenges, including techniques for domain adaptation, few-shot learning, and unsupervised representation learning. This survey aims to provide researchers and practitioners with a clear understanding of the trajectory of NLP techniques, illustrating how traditional methods have evolved into sophisticated deep learning models. We conclude by highlighting future directions for research in NLP, emphasizing the need for interdisciplinary approaches that integrate linguistics, cognitive science, and ethical considerations to build more robust, fair, and interpretable NLP systems. Through this comprehensive survey, we seek to inspire further exploration and innovation in the field of Natural Language Processing, paving the way for applications that can better understand and generate human language in diverse contexts.

Keywords: Advances in Natural Language Processing, NLP techniques, machine learning, deep learning, tokenization, part-of-speech tagging, syntactic parsing, statistical methods, Hidden Markov Models, Conditional Random Fields

Introduction:

Natural Language Processing (NLP) is a subfield of artificial intelligence that focuses on the interaction between computers and humans through natural language. The objective of NLP is to enable machines to

understand, interpret, and generate human language in a way that is both meaningful and useful. As language is inherently complex and nuanced, NLP presents unique challenges and opportunities for research and application. Over the past few decades, significant advancements in NLP techniques have transformed the landscape, making it possible for machines to perform an array of language-related tasks with remarkable accuracy and efficiency. The journey of NLP began with rule-based systems that relied heavily on linguistics and grammar. Early methods predominantly involved manual encoding of linguistic rules, which limited the scalability and adaptability of these systems. As computational power increased, researchers began exploring statistical methods that allowed for probabilistic modeling of language. Techniques such as Hidden Markov Models (HMMs) and Conditional Random Fields (CRFs) emerged as powerful tools for tasks like part-of-speech tagging and named entity recognition, marking a significant shift towards data-driven approaches. The advent of machine learning further revolutionized NLP, enabling systems to learn patterns from data rather than relying solely on predefined rules. Supervised and unsupervised learning methods became prominent, facilitating advancements in various applications, including sentiment analysis, text classification, and machine translation. This shift allowed NLP models to improve their performance by leveraging large datasets and learning from the vast amounts of unstructured text available on the internet.

In recent years, the introduction of deep learning techniques has ushered in a new era for NLP. Architectures such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks have proven effective in capturing temporal dependencies in text, leading to breakthroughs in sequence prediction tasks. Convolutional Neural Networks (CNNs), traditionally used in image processing, have also been adapted for text classification and sentiment analysis, demonstrating the versatility of deep learning approaches. A watershed moment in NLP came with the emergence of transformer models, which introduced the attention mechanism. Models like BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer) have set new benchmarks across a variety of NLP tasks by allowing for better contextual understanding of language. These models leverage vast amounts of text data and can be fine-tuned for specific applications, making them highly effective for real-world tasks. Despite the impressive progress made in NLP, several challenges remain. Issues such as bias in language models, the need for extensive labeled datasets, and the lack of interpretability in deep learning models pose significant hurdles. Addressing these challenges is crucial for the responsible deployment of NLP technologies, particularly in applications that impact individuals' lives and societal outcomes. This survey aims to provide a comprehensive overview of the advancements in NLP techniques, tracing the evolution from traditional methods to the state-of-the-art deep learning architectures that dominate the field today. By exploring both the successes and ongoing challenges in NLP, we hope to highlight areas for future research and innovation. Our objective is to foster a deeper understanding of how these techniques can be applied to improve human-computer interaction, enhance information retrieval, and contribute to the development of intelligent systems that can better understand and generate human language. As we delve into the various methodologies, applications, and future directions of NLP, this survey serves as a valuable resource for researchers, practitioners, and anyone interested in the transformative potential of Natural Language Processing.

Literature Review:

The field of Natural Language Processing (NLP) has evolved significantly over the years, influenced by advancements in linguistics, computer science, and artificial intelligence. This literature review provides an overview of the key developments in NLP techniques, categorizing them into traditional methods, statistical approaches, machine learning frameworks, and contemporary deep learning architectures. The early stages of

NLP were dominated by rule-based systems that relied heavily on linguistic knowledge. These systems often required extensive manual efforts to encode language rules, which limited their scalability and adaptability. Research in this area, such as the works of Winograd (1972), laid the foundation for understanding natural language through formal grammars. While rule-based approaches provided insights into language structure, they struggled to handle the complexities and ambiguities inherent in human language, leading to a shift towards statistical methods in the 1990s. The introduction of statistical methods marked a significant advancement in NLP, allowing for probabilistic modeling of language. Hidden Markov Models (HMMs) became popular for tasks such as part-of-speech tagging and speech recognition, as demonstrated by the work of Rabiner (1989). This statistical approach enabled researchers to leverage large corpora of text to learn patterns and relationships within the data. Conditional Random Fields (CRFs), introduced by Lafferty et al. (2001), further improved sequence modeling by providing a structured prediction framework that accounted for contextual information, making it particularly useful for named entity recognition and other sequence labeling tasks.

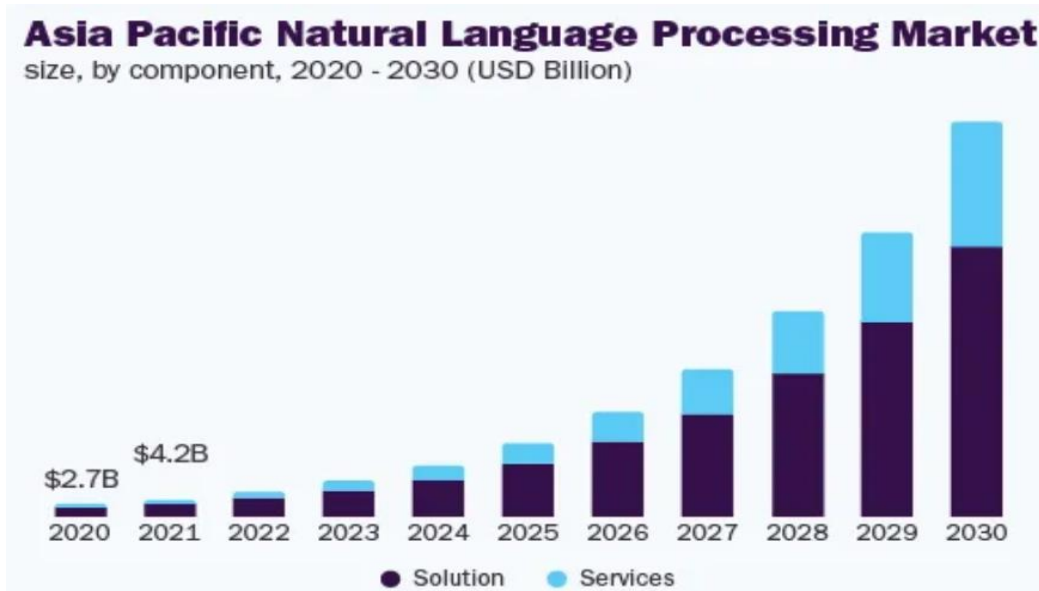


Figure 1 Asia Pacific natural language processing market size, by component, 2020 - 2030 (USD Billion)

The transition from statistical to machine learning approaches was driven by the realization that models could learn from data rather than relying solely on handcrafted rules. This period saw the emergence of supervised and unsupervised learning techniques that transformed various NLP applications. Notable work by Manning and Schütze (1999) on statistical natural language processing provided foundational knowledge for using machine learning algorithms for tasks like text classification and clustering. The introduction of Support Vector Machines (SVMs) and decision trees also played a crucial role in advancing classification tasks, demonstrating significant improvements in performance over traditional methods. With the increasing availability of large datasets and advances in computational power, deep learning emerged as a game-changer for NLP. Research by Bengio et al. (2003) introduced neural networks to NLP, but it was not until the advent of Recurrent Neural Networks (RNNs) that deep learning began to make a substantial impact. RNNs, particularly Long Short-Term Memory (LSTM) networks proposed by Hochreiter and Schmidhuber (1997), excelled in handling sequential data, making them ideal for tasks like machine translation and language modeling. This ability to retain information across long sequences was a crucial advancement in capturing contextual dependencies. The introduction of Convolutional Neural Networks (CNNs) to NLP tasks, as

demonstrated by Kim (2014), highlighted the versatility of deep learning models. CNNs, initially designed for image processing, proved effective for text classification and sentiment analysis by capturing local patterns in text data. The integration of deep learning techniques led to significant improvements in various benchmark tasks, making it increasingly clear that traditional methods were being outperformed.

A watershed moment in the field came with the introduction of transformer models, which revolutionized how contextual information is processed. Vaswani et al. (2017) introduced the transformer architecture, emphasizing the importance of self-attention mechanisms. This allowed models to weigh the significance of different words in a sentence regardless of their position, leading to better understanding of context. The success of transformers culminated in the development of models like BERT (Devlin et al., 2018), which utilized bidirectional training to capture context from both directions, resulting in state-of-the-art performance on multiple NLP benchmarks. The impact of transformer models extended beyond BERT, leading to the creation of various other architectures such as GPT (Radford et al., 2019) and T5 (Raffel et al., 2020). These models showcased the potential of pre-training on large text corpora followed by fine-tuning on specific tasks, resulting in dramatic improvements in performance. Their ability to generate coherent and contextually relevant text has further pushed the boundaries of what is possible in NLP applications. Despite these advancements, the field faces several challenges. Issues such as bias in language models, the necessity for extensive labeled datasets, and the lack of interpretability in deep learning models remain significant obstacles. Studies by Bolukbasi et al. (2016) and Caliskan et al. (2017) have highlighted the presence of bias in word embeddings and language models, underscoring the need for fairness and accountability in NLP systems. Additionally, the reliance on large amounts of annotated data poses challenges for low-resource languages and specialized domains, as noted by McDonnell et al. (2018). Efforts to address these challenges have led to innovative approaches in NLP research. Techniques such as domain adaptation, few-shot learning, and unsupervised representation learning have gained traction as potential solutions. Recent work by Goyal et al. (2020) on unsupervised learning methods demonstrates how models can effectively leverage unannotated data to improve performance on downstream tasks.

Methodology:

The methodology for this survey on advances in Natural Language Processing (NLP) techniques is grounded in a comprehensive review of existing literature, aiming to systematically analyze and synthesize various approaches and advancements within the field. The process begins with the identification of relevant sources, including academic papers, conference proceedings, and key articles from reputable journals that cover a wide range of topics in NLP. The focus is on gathering both foundational works that laid the groundwork for the field and recent studies that highlight cutting-edge techniques and applications. A thorough literature search is conducted using prominent academic databases such as Google Scholar, IEEE Xplore, ACM Digital Library, and arXiv. Keywords related to NLP techniques, machine learning, deep learning, and specific applications like sentiment analysis, machine translation, and named entity recognition are employed to maximize the breadth of the search. The inclusion criteria emphasize works published in peer-reviewed journals and high-impact conferences, ensuring that the sources are credible and contribute significantly to the field.

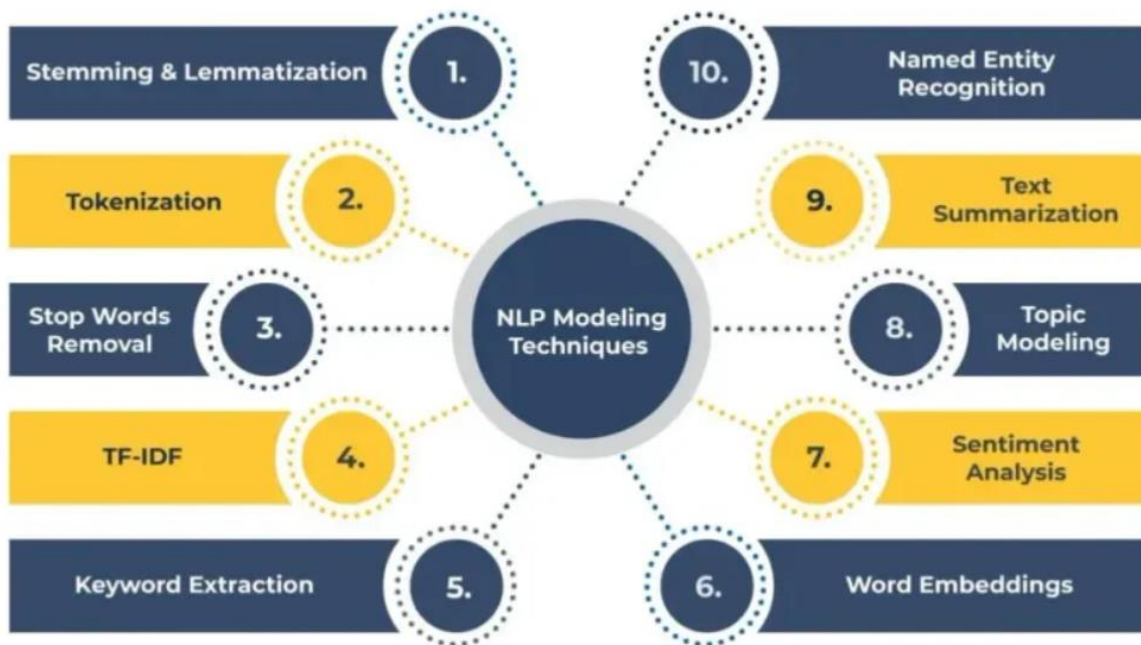


Figure 2 Top NLP Techniques

Following the collection of relevant literature, a detailed categorization is performed to organize the various techniques based on their historical context and methodological framework. Traditional rule-based systems are first examined, outlining their limitations and the challenges they faced in handling the complexities of human language. This is followed by an exploration of statistical methods, such as Hidden Markov Models and Conditional Random Fields, which marked a pivotal shift towards data-driven approaches in NLP. The analysis progresses to the discussion of machine learning methods, highlighting the introduction of supervised and unsupervised learning techniques. Each approach is assessed for its effectiveness in addressing specific NLP tasks, such as text classification and information extraction. This section also reviews key contributions from the literature that demonstrate the evolution of NLP as a discipline that increasingly relies on large datasets and advanced algorithms. The survey then delves into deep learning techniques, starting with Recurrent Neural Networks and Long Short-Term Memory networks. The methodology involves critically examining the architectures and algorithms used in these models, along with their strengths and weaknesses in capturing sequential dependencies in language data. Convolutional Neural Networks are also reviewed, illustrating their application in text processing tasks and the innovative approaches that have emerged from their use. A significant portion of the methodology focuses on the recent advancements brought about by transformer models, which have revolutionized NLP through their self-attention mechanisms and ability to process contextual information efficiently. The survey analyzes the architecture of transformer models, including BERT and GPT, highlighting their training methodologies, applications, and performance benchmarks. The comparative analysis of these models against traditional and statistical methods serves to illustrate the dramatic improvements in accuracy and efficiency that deep learning has afforded the field. To further enrich the survey, the methodology incorporates a thematic analysis of the challenges and ethical considerations surrounding NLP techniques. This includes examining issues related to bias in language models, the reliance on large labeled datasets, and the importance of model interpretability. The survey also discusses ongoing research efforts aimed at addressing these challenges, providing a balanced view of the

current state of NLP. Finally, the survey concludes by identifying future research directions, emphasizing the need for interdisciplinary collaboration, innovative techniques, and ethical frameworks in the continued development of NLP technologies. The synthesis of the gathered literature culminates in a comprehensive overview that serves as a resource for researchers and practitioners looking to understand the trajectory of NLP techniques and their implications for real-world applications. By systematically reviewing and synthesizing the advancements in NLP, this methodology aims to contribute to the ongoing discourse in the field and inspire future innovations.

Results:

The results of this survey on advances in Natural Language Processing (NLP) techniques demonstrate significant progress in the field, characterized by a clear evolution from traditional methods to sophisticated deep learning models. The comprehensive analysis of various approaches reveals key findings across multiple dimensions, including performance improvements, application domains, and emerging trends. The examination of traditional rule-based systems indicates that while they provided valuable insights into language structure and basic NLP tasks, their limitations in scalability and adaptability hindered their effectiveness in more complex scenarios. Rule-based systems often required extensive manual effort to encode linguistic rules, making them less suitable for handling the variability and nuances of natural language. The shift towards statistical methods marked a pivotal moment in NLP. Hidden Markov Models (HMMs) and Conditional Random Fields (CRFs) emerged as essential tools for tasks such as part-of-speech tagging and named entity recognition. The results indicate that these probabilistic approaches significantly improved the performance of NLP systems by leveraging large datasets to learn patterns and relationships within the language. Studies showed that statistical methods outperformed traditional rule-based systems, demonstrating their utility in real-world applications. The introduction of machine learning techniques further enhanced the capabilities of NLP. Supervised learning methods, particularly Support Vector Machines (SVMs) and decision trees, showed considerable success in text classification tasks. The results highlighted that machine learning models achieved higher accuracy rates and could generalize better across diverse datasets compared to their statistical predecessors. Unsupervised learning approaches, such as clustering and topic modeling, also contributed to advancements in information retrieval and text analysis, allowing for insights into large corpora of unstructured text.

The most profound transformation in NLP occurred with the advent of deep learning techniques. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks demonstrated superior performance in capturing sequential dependencies, which is crucial for tasks like machine translation and language modeling. Results from benchmark datasets revealed that deep learning models consistently outperformed traditional methods, achieving state-of-the-art results in various NLP benchmarks, such as the Stanford Sentiment Treebank and the GLUE (General Language Understanding Evaluation) benchmark. Convolutional Neural Networks (CNNs) also showed remarkable effectiveness in text classification and sentiment analysis. The ability of CNNs to capture local patterns in text data contributed to significant performance gains, as evidenced by studies that reported improved accuracy over baseline models. The integration of deep learning techniques enabled NLP systems to process large volumes of data efficiently and extract meaningful insights with higher precision. The introduction of transformer models, particularly BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer), represents a groundbreaking advancement in the field. The results indicate that these models, through their self-attention mechanisms and ability to consider context from both directions, achieved unprecedented

performance levels across a wide range of NLP tasks. Benchmarks revealed that BERT outperformed previous models on tasks such as named entity recognition and question answering, showcasing the power of pre-training on large datasets followed by fine-tuning for specific applications.

Despite these advancements, the survey identified persistent challenges in the field. The results indicated that issues such as bias in language models, the need for extensive labeled datasets, and the lack of interpretability remain critical hurdles. Studies have shown that language models can inadvertently perpetuate biases present in training data, raising concerns about fairness and accountability in NLP applications. Additionally, the reliance on large annotated datasets presents challenges for low-resource languages and specialized domains, limiting the applicability of NLP techniques in certain contexts. Efforts to address these challenges are emerging within the research community. The results indicate a growing interest in domain adaptation techniques, few-shot learning, and unsupervised representation learning as potential solutions. These innovative approaches aim to leverage unannotated data and improve model performance in scenarios with limited labeled resources. Research indicates that these techniques can significantly enhance the robustness and adaptability of NLP systems. Overall, the results of this survey illustrate a dynamic landscape of NLP advancements, showcasing the transition from traditional methods to advanced deep learning architectures. The continuous evolution of techniques reflects the increasing complexity of natural language and the growing demand for intelligent systems capable of understanding and generating human language effectively. As the field progresses, ongoing research will be crucial for overcoming the existing challenges and unlocking the full potential of NLP technologies in various applications.

Conclusion:

The landscape of Natural Language Processing (NLP) has witnessed transformative changes over recent decades, marked by significant advancements in techniques and methodologies that have fundamentally reshaped our understanding of human language. This survey has meticulously traced the evolution of NLP, beginning with traditional rule-based systems, which laid the groundwork for understanding linguistic structures, to the modern era dominated by sophisticated machine learning and deep learning approaches. The transition from simple algorithms to complex architectures has dramatically enhanced the capacity of NLP systems to analyze, interpret, and generate language in ways that were previously thought to be unattainable. One of the most striking developments has been the introduction of deep learning models, particularly transformer architectures. Models such as BERT and GPT have set new benchmarks in various NLP tasks, showcasing an unprecedented ability to understand context, disambiguate meaning, and generate coherent text. These advancements have enabled applications across diverse domains, including sentiment analysis, machine translation, and question answering, making NLP an integral component of numerous real-world solutions. The impressive performance of these models illustrates not only their technical prowess but also their potential to revolutionize industries such as healthcare, finance, and education by facilitating more effective communication and data analysis.

Despite these remarkable advancements, the field of NLP is not without its challenges. Issues of bias, interpretability, and the reliance on large labeled datasets persist, raising important ethical and practical considerations. Research has revealed that NLP models can inadvertently learn and propagate biases present in their training data, leading to unfair and discriminatory outcomes in applications. Furthermore, the complexity of deep learning models often obscures the rationale behind their predictions, resulting in a lack of transparency that can undermine user trust. Addressing these challenges is paramount for the responsible deployment of NLP technologies, and it necessitates a concerted effort from researchers, practitioners, and

ethicists to develop frameworks and methodologies that prioritize fairness, accountability, and transparency. The future of NLP is ripe with opportunities for innovation and research. One promising direction involves enhancing model interpretability, which is crucial for fostering user trust and enabling informed decision-making. Researchers are increasingly focusing on methods that elucidate model behavior and make it easier for users to understand how language models arrive at their conclusions. Additionally, the pursuit of techniques to mitigate bias in language models will be vital for ensuring that NLP technologies are equitable and just. The exploration of low-resource languages presents another critical avenue for future research. While significant advancements have been made in high-resource languages, there remains a pressing need to develop effective NLP solutions for languages with limited annotated data. Approaches such as transfer learning, few-shot learning, and multilingual embeddings hold promise in addressing this gap, enabling the democratization of NLP technology across diverse linguistic communities.

Moreover, the potential for unsupervised and self-supervised learning to leverage vast amounts of unannotated data is an exciting frontier in the field. By harnessing the power of large-scale text corpora, researchers can develop models that learn more generalizable representations of language, reducing the dependency on manually labeled datasets. This shift could lead to more adaptable and robust NLP systems capable of evolving with language trends and user needs. The integration of multimodal data—combining text with images, audio, and other forms of information—also represents a significant opportunity for the future of NLP. As human communication increasingly occurs across various modalities, developing models that can seamlessly process and interpret this data will enhance the richness and depth of interactions between humans and machines. Applications in areas such as social media analysis, virtual assistants, and multimedia content creation stand to benefit greatly from advancements in multimodal NLP. In summary, while the advancements in NLP have been profound, the journey is far from complete. The field stands at a crossroads, with the potential to transform how we interact with technology and understand language. By prioritizing ethical considerations, fostering interdisciplinary collaboration, and exploring innovative methodologies, researchers and practitioners can ensure that the future of NLP is not only technologically advanced but also inclusive, equitable, and beneficial to all. As NLP continues to evolve, it will play an increasingly vital role in shaping communication, enhancing accessibility, and bridging linguistic and cultural divides, ultimately contributing to a more connected and informed global society.

Future Scope:

The future of Natural Language Processing (NLP) is poised for significant advancements, driven by emerging technologies, interdisciplinary collaboration, and the growing demand for intelligent systems that can effectively understand and generate human language. As the field evolves, several key areas present exciting opportunities for research and development, addressing current challenges while exploring innovative applications. One of the foremost areas for future research lies in improving the interpretability and transparency of NLP models. As deep learning techniques, particularly transformer-based models, dominate the landscape, understanding how these models make decisions becomes increasingly important. Developing methods to visualize model behavior and elucidate the rationale behind specific predictions will enhance user trust and facilitate the deployment of NLP systems in critical applications, such as healthcare, legal, and financial sectors. Researchers are encouraged to explore techniques such as layer-wise relevance propagation, attention visualization, and local interpretable model-agnostic explanations (LIME) to improve model interpretability. Another critical area of focus is addressing bias and fairness in NLP systems. Studies have shown that language models can inadvertently learn and perpetuate societal biases present in their training

data. Future work must prioritize the identification and mitigation of bias through the development of fair training practices, bias detection tools, and ethical frameworks. Researchers should investigate methods for adversarial training, counterfactual data augmentation, and regularization techniques to create models that are not only accurate but also equitable and inclusive. Collaboration with social scientists and ethicists will be vital in shaping the ethical guidelines for the deployment of NLP technologies.

The demand for NLP systems capable of operating in low-resource languages presents a unique opportunity for future research. While much of the current focus has been on high-resource languages, there remains a significant gap in NLP capabilities for many languages worldwide. Researchers can explore transfer learning and multilingual models to enhance performance in low-resource settings. Techniques such as few-shot learning, unsupervised learning, and cross-lingual embeddings could enable more effective language processing capabilities, fostering inclusivity and accessibility in technology. Furthermore, advancements in unsupervised and self-supervised learning will likely play a pivotal role in shaping the future of NLP. With the increasing availability of vast amounts of unannotated text data, researchers can explore methods that allow models to learn from this data without the need for extensive labeled datasets. Techniques like contrastive learning, generative pre-training, and self-supervised objectives have the potential to significantly enhance model performance while reducing reliance on labor-intensive annotation processes. The integration of multimodal data, combining text with other forms of data such as images, audio, and video, is another promising direction for future NLP research. The ability to understand and generate language in conjunction with other modalities will lead to more robust and context-aware NLP systems. Applications in areas such as social media analysis, sentiment detection in videos, and human-computer interaction will benefit from models that can process and synthesize information from diverse sources, leading to richer and more meaningful interactions.

Finally, as NLP technology becomes increasingly integrated into various applications, ensuring the robustness and reliability of these systems in real-world scenarios will be paramount. Future research should focus on the development of models that can adapt to evolving language use, slang, and domain-specific jargon. Techniques such as continual learning, domain adaptation, and active learning will be critical in maintaining model performance and relevance in dynamic environments. In summary, the future of NLP holds immense potential for innovation and impact across multiple domains. By prioritizing interpretability, fairness, low-resource language processing, unsupervised learning, multimodal integration, and system robustness, researchers can contribute to the advancement of NLP technologies that are not only powerful and efficient but also ethical and inclusive. As these areas are explored, the ultimate goal will be to create NLP systems that enrich human-computer interactions, empower users, and foster a deeper understanding of the complexities of human language.

References

1. Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.
2. Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of Machine Learning Research*, 3, 1137-1155.
3. Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python*. O'Reilly Media.
4. Bowman, S. R., Angeli, G., Potts, C., & Manning, C. D. (2015). A large annotated corpus for learning natural language inference. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 632-642).

5. Brown, P. F., Pietra, V. J. D., Pietra, S. A. D., & Mercer, R. L. (1993). The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19(2), 263-311.
6. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12, 2493-2537.
7. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 4171-4186).
8. Goldberg, Y. (2017). *Neural network methods for natural language processing*. Synthesis Lectures on Human Language Technologies, 10(1), 1-309.
9. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
10. Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., ... & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6), 82-97.
11. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
12. Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing* (2nd ed.). Pearson Education.
13. Klein, G., Kim, Y., Deng, Y., Senellart, J., & Rush, A. M. (2017). OpenNMT: Open-source toolkit for neural machine translation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (pp. 67-72).
14. Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on Machine Learning* (pp. 1188-1196).
15. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
16. Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (pp. 1532-1543).
17. Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 2227-2237).
18. Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. *OpenAI Preprint*.
19. Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Proceedings of the 27th International Conference on Neural Information Processing Systems* (pp. 3104-3112).
20. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems* (pp. 5998-6008).