MILAN AROTE, 2025, 13:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

Natural Language Processing

Milan Arote
Tilak Maharashtra Vidyapeeth

Abstract- This paper provides a comprehensive review of recent advancements in Natural Language Processing (NLP), examining how deep learning architectures, particularly transformer-based models, have revolutionized the field. We analyze the current state of NLP technologies across various applications including machine translation, sentiment analysis, question answering, and text generation. Additionally, we discuss emerging challenges in the domain such as ethical considerations, multilingual capabilities, and computational efficiency. The paper concludes with a discussion of promising future research directions that may shape the next generation of NLP systems. Our analysis suggests that while significant progress has been made, considerable opportunities remain for improving context understanding, reducing hallucinations, and developing more resource-efficient models.

Moreover, Natural Language Processing (NLP), exploring its historical roots, fundamental techniques, diverse applications, and potential future directions. NLP, at the intersection of computer science, artificial intelligence, and linguistics, enables computers to understand, interpret, and generate human language. The paper delves into the evolution of NLP, starting from rule-based systems to the current era of machine learning and deep learning approaches. Key techniques such as text processing, syntactic and semantic analysis, and knowledge representation are discussed. Furthermore, the paper examines a wide array of NLP applications, including machine translation, sentiment analysis, chatbots, and information extraction. Finally, it explores emerging trends and future research directions in NLP, highlighting the potential for advancements in areas like contextual understanding, explainable AI, and multilingual processing.

Keywords- Natural Language Processing, Deep Learning, Transformer Models, Language Understanding, NLP Applications.

I. INTRODUCTION

Natural Language Processing (NLP) has undergone a remarkable transformation in recent years, evolving from rule-based systems to sophisticated deep learning models capable of understanding and generating human language with unprecedented accuracy. This evolution has been primarily driven by advances in neural network architectures, the availability of large-scale datasets, and increasing computational resources. The advent of transformer architectures in particular has resulted in state-of-the-art performance across almost all NLP tasks, leading to what many researchers consider a paradigm shift in the field.

1. Research Aims

This paper aims to: Provide a comprehensive overview of recent technological advancements in NLP Examine real-world applications and their societal impact Discuss ongoing challenges and limitations Explore promising future research directions.

3.Significance of the Study

The significance of NLP research extends beyond academic interest, as language technologies increasingly influence how we interact with machines, access information, and communicate globally. Understanding the current capabilities, limitations, and future trajectory of NLP is essential for researchers, practitioners, and policymakers alike.

II. LITERATURE REVIEW

1. Historical Development of NLP

The field of Natural Language Processing has evolved dramatically over several decades. Early approaches in the 1950s and 1960s were dominated by rule-based systems and symbolic methods, exemplified by Chomsky's work on formal grammars (Chomsky, 1957) and early machine translation efforts like the Georgetown-IBM experiment (Hutchins, 2004). These approaches relied heavily on hand-crafted rules and dictionaries, demonstrating initial promise but ultimately facing limitations in handling the complexity and ambiguity inherent in natural language.

The 1970s and 1980s saw a shift toward statistical methods. Works by Jelinek and Mercer (1980) at IBM introduced statistical approaches to speech recognition that would later influence text processing. This period also saw the development of probabilistic parsing techniques (Church, 1988), which began addressing some of the rigidity of purely rule-based approaches.

2. Statistical NLP and Early Machine Learning

The 1990s and early 2000s represented the golden era of statistical NLP. During this period, researchers developed statistical models that could learn patterns from data rather than relying solely on hand-crafted rules. Hidden Markov Models became widespread for sequential labeling tasks such as part-of-speech tagging (Brants, 2000), while probabilistic context-free grammars improved parsing accuracy (Collins, 1997).

Statistical machine translation emerged as a significant research area, with IBM's alignment models (Brown et al., 1993) and phrase-based systems (Koehn et al., 2003) setting new standards for translation quality. These approaches relied on parallel corpora and statistical inference to learn translation patterns automatically.

Simultaneously, early machine learning techniques began making inroads into NLP. Maximum entropy models (Berger et al., 1996) and support vector machines (Joachims, 1998) demonstrated success across various classification tasks, including text categorization and named entity recognition.

3. The Neural Revolution in NLP

The paradigm shift toward neural methods began in earnest with the introduction of distributed word representations or "word embeddings." Pioneering work by Bengio et al. (2003) on neural language models laid the groundwork, but it was the introduction of Word2Vec by Mikolov et al. (2013) and GloVe by Pennington et al. (2014) that democratized the use of dense vector representations for words, capturing semantic relationships in a continuous vector space.Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks (Hochreiter & Schmidhuber, 1997), revolutionized sequence modeling in NLP. These architectures demonstrated remarkable capabilities in machine translation (Sutskever et al., 2014), leading to the development of sequence-to-sequence models. Bahdanau et al. (2014) further enhanced these models by introducing attention mechanisms, allowing the model to focus on different parts of the input sequence when generating outputs.The attention mechanism would later become the foundation for the transformer architecture introduced by Vaswani et al. (2017), which dispensed with recurrence entirely in favor of self-attention. This innovation enabled more efficient training and better modeling of long-range dependencies in text, catalyzing dramatic improvements across NLP tasks.

4, Pre-trained Language Models and Transfer Learning

A paradigm shift occurred with the introduction of pre-trained language models that leverage transfer learning. Howard and Ruder (2018) demonstrated with ULMFiT that language models pre-trained on large corpora could be fine-tuned for specific tasks with dramatic improvements over training from scratch.

This approach was further developed with models like ELMo by Peters et al. (2018), which provided contextualized word representations, capturing how a word's meaning might change depending on its context. The introduction of BERT by Devlin et al. (2019) marked another milestone, using masked language modeling to create bidirectional representations that achieved state-of-the-art results across numerous benchmarks.

GPT (Generative Pre-trained Transformer) models by Radford et al. (2018, 2019) and Brown et al. (2020) demonstrated the power of autoregressive language modeling at scale, with GPT-3 showing remarkable few-shot learning capabilities. T5 by Raffel et al. (2020) unified diverse NLP tasks into a text-to-text format, providing a consistent approach to transfer learning across task types.

5. Application-Specific Advances

Research in specific NLP applications has seen remarkable progress. In machine translation, the work of Wu et al. (2016) on Google's Neural Machine Translation system demonstrated that neural approaches could outperform traditional statistical methods. Johnson et al. (2017) extended this to multilingual translation within a single model.

For sentiment analysis, Liu (2012) provided a comprehensive survey of techniques, while works like Tang et al. (2015) demonstrated how deep learning could capture aspect-based sentiment. Question answering systems have evolved from early information retrieval approaches to sophisticated reading comprehension models, with SQuAD (Rajpurkar et al., 2016) becoming an influential benchmark dataset.

Text summarization has progressed from extractive methods (Nenkova & McKeown, 2011) to abstractive approaches using sequence-to-sequence models (Rush et al., 2015) and later transformer-based architectures (Lewis et al., 2020), enabling the generation of fluent summaries that may contain novel phrasing not found in the source document.

6. Ethical and Responsible NLP

Recent literature has increasingly focused on ethical considerations in NLP. Bolukbasi et al. (2016) demonstrated gender bias in word embeddings, while Caliskan et al. (2017) showed how language models could perpetuate human biases. These findings sparked research into bias detection and mitigation strategies (Gonen & Goldberg, 2019).Bender and Friedman (2018) proposed data statements for NLP, encouraging transparency about dataset composition and annotation. The environmental impact of large language models was highlighted by Strubell et al. (2019), prompting research into more efficient training and inference methods.Works by Bender et al. (2021) on the risks of large language models (the "stochastic parrots" paper) and discussions of algorithmic fairness by Blodgett et al. (2020) have further emphasized the importance of developing NLP technologies responsibly, with awareness of their societal implications.

7. Current Challenges and Research Gaps

Despite significant advances, several challenges persist in NLP research. Contextual understanding remains limited, with models struggling to maintain coherence over long texts (Beltagy et al., 2020). Common sense reasoning and world knowledge integration continue to be active areas of research (Bosselut et al., 2019).

Multilingual NLP faces significant resource disparities, with low-resource languages receiving comparatively little attention (Joshi et al., 2020). Cross-cultural nuances and linguistic diversity present ongoing challenges for creating truly global NLP systems.

The computational requirements of state-of-the-art models raise concerns about accessibility and environmental impact (Patterson et al., 2021), while the tendency of large language models to generate plausible but factually incorrect information ("hallucinations") remains a significant limitation for critical applications (Maynez et al., 2020).

Evaluation methodologies have also been critiqued, with concerns that current benchmarks may not adequately measure genuine language understanding (Schlangen, 2021). The gap between impressive performance on standard benchmarks and real-world robustness continues to be a central challenge in the field.

III Technological Advancements in NLP

1. Evolution of NLP Models

Natural language processing has evolved through several distinct paradigms:

Rule-Based Systems: Early NLP systems relied heavily on handcrafted rules and lexicons. While effective for highly specific tasks, these approaches lacked flexibility and required enormous human effort to maintain.

- **StatisticalMethods**: The introductionof statistical techniques, particularly probabilistic methods like Hidden Markov Models and Conditional Random Fields, enabled more robust language processing.
- **Word Embeddings**: The development of word embedding techniques (Word2Vec, GloVe) represented a significant step forward by capturing semantic relationships between words in vector space.
- **Recurrent Neural Networks**: LSTMs and GRUs improved sequence modeling capabilities, allowing for better handling of context in language.
- **Attention Mechanisms**: The introduction of attention allowed models to focus on relevant parts of input sequences, dramatically improving performance on various tasks
- **Transformer Architecture**: The transformer architecture, introduced in "Attention is All You Need" (Vaswani et al., 2017), revolutionized NLP by enabling highly parallelizable training and more effective modeling of long-range dependencies.

Large Language Models: The scaling of transformer models to billions of parameters (GPT, T5, BERT, and their successors) has led to remarkable improvements across virtually all NLP benchmarks.

- **Pre-trained Language Models**

Pre-trained language models have become the foundation of modern NLP systems. These models are trained on vast corpora of text using self-supervised objectives can be fine-tuned for specific downstream tasks with relatively small amounts of labeled data.

- **Key developments include:**

BERT (Bidirectional Encoder Representations from Transformers): Introduced bidirectional context understanding through masked language modeling.

- **GPT (Generative Pre-trained Transformer):** Demonstrated impressive text generation capabilities using autoregressive language modeling.
- **T5 (Text-to-Text Transfer Transformer)**: Reformulated all NLP tasks into a unified text-to-text format.
- **BART and other encoder-decoder models**: Combined bidirectional encoding with autoregressive decoding for tasks like summarization and translation.

Multilingual models: Models like mBERT, XLM-R, and BLOOM advanced cross-lingual transfer learning.

- **Multimodal models:** CLIP, DALL-E, and similar systems integrated language understanding with visual processing.

3.Few-Shot and Zero-Shot Learning

A particularly significant development has been the emergence of few-shot and zero-shot capabilities in large language models:

- **Few-shot learning:** The ability to perform tasks with only a handful of examples.
- **Zero-shot learning**: The ability to perform tasks without any task-specific examples, using only natural language instructions.

These capabilities emerged primarily from scale and have transformed how NLP
systems are deployed, reducing the need for extensive task-specific datasets and enabling rapid adaptation to new use cases.

IV. APPLICATIONSOF NLP

1. **Machine Translation**

Neural machine translation (NMT) has largely replaced statistical approaches, with transformer-based models achieving near-human performance on many language pairs. Recent advances include:

•Document-level translation that maintains context across sentences
•Multilingual models capable of translating between hundreds of languages
•Unsupervised translation methods that require minimal parallel data
•Real-time speech-to-speech translation systems
Despite these advances, challenges remain in handling low-resource languages, domain-specific terminology, and culturally-specific expressions.

Information Extraction and Retrieval

NLP has transformed how information is extracted and retrieved from large text collections:
•**Named Entity Recognition (NER):** Identifying and classifying entities such as people, organizations, and locations in text.
•**Relation Extraction**: Determining relationships between entities.
•**Event Extraction:** Identifying events and their participants from text.
•**Semantic Search:** Retrieving information based on meaning rather than keyword matching.
.**Retrieval-Augmented Generation (RAG):** Combining information retrieval with text generation for factual accuracy.
These technologies power applications ranging from academic research tools to business intelligence systems.
3 .Conversational AI and Question Answering
The development of sophisticated dialogue systems and question-answering technologies has been driven by advances in:
•**Open-domain question answering:** Answering questions without domain restrictions.
•**Reading comprehension:** Extracting answers from specific passages.
•**Dialogue state tracking**: Maintaining context through multi-turn conversations.
•**Response generation**: Producing coherent, contextually appropriate responses.
These capabilities have enabled more natural human-computer interaction through virtual assistants, customer service bots, and educational tools.

4. Text Summarization and Generation

Automatic text summarization has progressed significantly:
•**Extractive summarization**: Selecting important sentences from source documents.
•**Abstractive summarization**: Generating novel text that captures key information.
•**Query-focused summarization:** Creating summaries tailored to specific information needs.
•**Multi-document summarization**: Synthesizing information across multiple sources.

Text generation capabilities have expanded to include:
•Creative writing (stories, poetry, scripts)
•Report generation from structured data
•Product descriptions
•Email drafting
•Code generation

5. Sentiment Analysis and Opinion Mining

Sentiment analysis has evolved from simple polarity detection to nuanced understanding of emotions, opinions, and attitudes:
•**Aspect-based sentiment analysis:** Identifying sentiments toward specific aspects of products or services.
•**Emotion detection:** Recognizing complex emotional states beyond positive/negative sentiment.
•**Stance detection:** Determining positions on contentious topics.

•**Opinion summarization**: Aggregating opinions across multiple sources.

These technologies enable businesses to monitor brand perception, analyze customer feedback, and track public opinion.

V .CHALLENGES AND LIMITATIONS

1. Ethical Considerations

The widespread deployment of NLP systems raises important ethical concerns:
•**Bias and fairness:** Language models often reflect and amplify societal biases present in training data.
•**Privacy concerns**: Models may memorize sensitive information from training data.
•**Misuse potential**: Advanced text generation capabilities can be used for misinformation, spam, or impersonation.
•**Transparency and explainability**: Many models function as "black boxes" with limited interpretability.
.**Environmental impact:** Training large language models requires significant computational resources and energy.

2. Technical Challenges

Despite impressive advances, several technical challenges persist:
•**Long-context understanding**: Most models struggle with very long documents or conversations.
•**Common sense reasoning:** Models often lack fundamental world knowledge that humans take for granted.
•**Causal reasoning**: Understanding cause-effect relationships remains difficult.
•**Computational efficiency**: State-of-the-art models are increasingly resource-intensive.
•**Hallucinations:** Models frequently generate plausible but factually incorrect information.
•**Evaluation metrics**: Current metrics often fail to capture nuanced aspects of language understanding and generation.

3. Multilingual and Cross-cultural Challenges

The diversity of human languages presents unique challenges:
•**Language resource disparity:** Low-resource languages lack sufficient data for robust model training.
•**Cross-lingual transfer**: Effectively transferring knowledge between languages remains difficult.
•**Cultural context**: Models often miss culturally-specific meanings and references.
•**Linguistic diversity**: Languages vary dramatically in structure, making unified approaches challenging.

VI. FUTURE RESEARCH DIRECTIONS

1. More Efficient Models

Research is increasingly focused on developing more efficient NLP technologies:
•**Model distillation**: Creating smaller models that retain the capabilities of larger ones.
•**Parameter-efficient fine-tuning**: Methods like adapters and prompt tuning that modify only a small subset of parameters.
•**Sparse activation:** Models that activate only relevant parts for specific inputs.
•**Quantization**: Reducing the precision of model weights without significant performance loss.
•**Neural architecture search:** Automatically discovering efficient architectures.

2. Multimodal Integration

The integration of language with other modalities represents a promising direction:
•**Vision-language models**: Combining text and image understanding.
•**Audio-language processing**: Integrating speech and language processing.
•**Embodied AI**: Connecting language understanding with physical interaction.
These multimodal approaches may better capture the grounded nature of human language understanding.

3. Knowledge Integration
Enhancing models with structured knowledge and reasoning capabilities:
•**Knowledge graph integration**: Explicitly connecting language models with structured knowledge sources.
•**Retrieval-augmented methods**: Dynamically accessing external knowledge during inference.
•**Neuro-symbolic approache**s: Combining neural networks with symbolic reasoning.
•**Tool use**: Enabling models to interact with external tools and APIs.

4. Trustworthy AI
Developing more trustworthy NLP systems through:
•**Bias detection and mitigation**: Techniques to identify and reduce harmful biases.
•**Explainable NLP:** Methods for interpreting model decisions.
•**Robust evaluation**: More comprehensive evaluation frameworks that account for reliability, fairness, and safety.
•**Red teaming and adversarial testing**: Systematic approaches to identifying potential failure modes.
Chapter Seven: Conclusion

VII. CONCLUSION

Natural Language Processing has experienced unprecedented progress in recent years, primarily driven by advances in deep learning architectures and the scaling of pre-trained language models. These technological developments have enabled applications that are transforming how humans interact with computers and access information.
Despite these advances, significant challenges remain in developing NLP systems that truly understand language with the depth and nuance that humans do. Future research directions focusing on efficiency, multimodal integration, knowledge grounding, and trustworthiness promise to address current limitations and expand the capabilities of NLP technologies.
As the field continues to evolve, interdisciplinary collaboration between computer scientists, linguists, cognitive scientists, and domain experts will be essential for creating NLP systems that are not only powerful but also responsible, fair, and beneficial to society.

REFERENCES

1. .Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (pp. 5998-6008).
2. .Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of NAACL-HLT (pp. 4171-4186).
3. .Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877-1901.
4. .Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research, 21(140), 1-67.
5. .Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.
6. .Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610-623).
7. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9459-9474.
8. .Radfrd, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021). Learning transferable visual models from

natural language supervision. In International Conference on Machine Learning (pp. 8748-8763).
9. Wei, J., Bosma, M., Zhao, V. Y., Guu, K., Yu, A. W., Lester, B., ... & Le, Q. V. (2022). Finetuned language models are zero-shot learners. In International Conference on Learning Representations.
10. .Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., ... & Sifre, L. (2022). Training compute-optimal large language models. arXiv preprint arXiv:2203.15556.
11. .Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., ... & Wen, J. R. (2023). A survey of large language models. arXiv preprint arXiv:2303.18223.
12. .Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.
13. Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.
14. Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016). SQuAD: 100,000+ questions for machine comprehension of text. arXiv preprint arXiv:1606.05250.
15. Chomsky, N. (1957). Syntactic structures. The Hague: Mouton.
16. .Hutchins, J. (2004). The Georgetown-IBM experiment demonstrated in January 1954. In Conference of the Association for Machine Translation in the Americas (pp. 102-114).
17. .Jelinek, F., & Mercer, R. L. (1980). Interpolated estimation of Markov source parameters from sparse data. In Proceedings of the Workshop on Pattern Recognition in Practice.
18. .Church, K. W. (1988). A stochastic parts program and noun phrase parser for unrestricted text. In Second Conference on Applied Natural Language Processing (pp. 136-143).
19. .Brants, T. (2000). TnT: A statistical part-of-speech tagger. In Sixth Applied Natural Language Processing Conference (pp. 224-231).
20. Collins, M. (1997). Three generative, lexicalised models for statistical parsing. In 35th Annual Meeting of the Association for Computational Linguistics (pp. 16-23).
21. Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., & Mercer, R. L. (1993). The mathematics of statistical machine translation: Parameter estimation. Computational linguistics, 19(2), 263-311.
22. .Koehn, P., Och, F. J., & Marcu, D. (2003). Statistical phrase-based translation. In Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (pp. 127-133).
23. .Berger, A. L., Pietra, V. J. D., & Pietra, S. A. D. (1996). A maximum entropy approach to natural language processing. Computational linguistics, 22(1), 39-71.
24. .Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. In European conference on machine learning (pp. 137-142).
25. .Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. Journal of machine learning research, 3, 1137-1155.
26. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint ariv:1301.3781.
27. .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).
28. .Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.
29. Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. Advances in neural information processing systems, 27.
30. .Howard, J., & Ruder, S. (2018). Universal language model fine-tuning for text classification. arXiv preprint arXiv:1801.06146.

31. .Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. arXiv preprint arXiv:1802.05365.
32. Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.
33. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI blog, 1(8), 9.
34. .Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., ... & Dean, J. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv preprint arXiv:1609.08144.
35. Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., ... & Hughes, M. (2017). Google's multilingual neural machine translation system: Enabling zero-shot translation. Transactions of the Association for Computational Linguistics, 5, 339-351.
36. Tang, D., Qin, B., & Liu, T. (2015). Document modeling with gated recurrent neural network for sentiment classification. In Proceedings of the 2015 conference on empirical methods in natural language processing (pp. 1422-1432).
37. .Nenkova, A., & McKeown, K. (2011). Automatic summarization. Foundations and Trends in Information Retrieval, 5(2–3), 103-233.
38. .Rush, A. M., Chopra, S., & Weston, J. (2015). A neural attention model for abstractive sentence summarization. arXiv preprint arXiv:1509.00685.
39. .Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. Advances in neural information processing systems, 29.
40. Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. Science, 356(6334), 183-186.
41. 41.Gonen, H., & Goldberg, Y. (2019). Lipstick on a pig: Debiasing methods cover up systematic gender biases in word embeddings but do not remove them. arXiv preprint arXiv:1903.03862.
42. .Bender, E. M., & Friedman, B. (2018). Data statements for natural language processing: Toward mitigating system bias and enabling better science. Transactions of the Association for Computational Linguistics, 6, 587-604.
43. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. arXiv preprint arXiv:1906.02243.
44. Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (technology) is power: A critical survey of "bias" in NLP. arXiv preprint arXiv:2005.14050.
45. .Beltagy, I., Peters, M. E., & Cohan, A. (2020). Longformer: The long-document transformer. arXiv preprint arXiv:2004.05150.
46. Bosselut, A., Rashkin, H., Sap, M., Malaviya, C., Celikyilmaz, A., & Choi, Y. (2019). COMET: Commonsense transformers for automatic knowledge graph construction. arXiv preprint arXiv:1906.05317.
47. .Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. arXiv preprint arXiv:2004.09095.
48. .Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L. M., Rothchild, D., ... & Dean, J. (2021). Carbon emissions and large neural network training. arXiv preprint arXiv:2104.10350.
49. Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On faithfulness and factuality in abstractive summarization. arXiv preprint arXiv:2005.00661.
50. Schlangen, D. (2021). Targeting the benchmark: On methodology in current natural language processing research. arXiv preprint arXiv:2007.04792.