Journal of Advanced Research in Signal Processing and Applications
Volume 6, Print Issue 1 - 2024, Pg. No. 16-22
Peer Reviewed Journal

Review Article

Next Word Prediction Using ML and DA

Jahanvi Arora

Assistant Professor in CSE/IT.

I N F O

E-mail Id:
jahanvi@pcte.edu.in
Orcid Id:
https://orcid.org/0009-0002-9927-2390
How to cite this article:
Arora J. Next Word Prediction Using ML and DA. *J Adv Res Sig Proc Appl* 2024; 6(1): 16-22.

Date of Submission: 2024-03-03
Date of Acceptance: 2024-04-07

A B S T R A C T

Next-word prediction is a natural language processing task that involves predicting the most likely next word given a sequence of words. It is a fundamental problem in many applications, such as speech recognition, machine translation, and text generation. The main research problem for building a next-word prediction system is to develop an accurate and efficient model that can predict the next word in a given sequence of words. The model needs to be able to handle the complexity of natural language, including the variability in word order, syntax, and semantics. It should also be able to learn from a large corpus of text data and generalize well to new data. There are several challenges associated with building a next-word prediction system, including dealing with out-of-vocabulary words, handling long-term dependencies, and avoiding overfitting. These challenges require careful design choices, such as the choice of neural network architecture, training data, and hyperparameters. Overall, the research problem for building a next-word prediction system is to develop a model that can accurately predict the next word in a sequence of words, while also being robust to the complexities and variability of natural language. This is a challenging problem that requires expertise in natural language processing, machine learning, and deep learning, and has many potential applications in fields such as artificial intelligence, speech recognition, and text generation.

Keywords: Next-Word, Speech Recognition, Machine Translation, Artificial Intelligence

Introduction

The innate human ability to foresee upcoming words in conversation serves as a foundational element in natural language processing, molding our interactions with technology and facilitating smooth communication. Anticipatory language models, particularly those specialized in predicting subsequent words, have become central across diverse applications, ranging from predictive text on smartphones to intelligent virtual assistants adept at comprehending and responding to spoken commands.

Throughout the years, these anticipatory language models have undergone significant evolution, progressing from initial rule-based and statistical approaches to the prevalent dominance of sophisticated deep learning architectures. Foreseeing the next word involves not just linguistic prowess but also the dynamic interplay of computational algorithms and neural networks, striving to capture the nuances of language.

This comprehensive review embarks on an exploration of the historical landscape of anticipatory language models, mapping their development and discerning the pivotal role they play in enhancing human-computer interaction. As we delve into the intricacies of employed methodologies,

our aim is to unravel the strengths and limitations of each paradigm, encompassing classical statistical models to the advanced techniques of deep learning. The objective is to provide readers with a nuanced understanding of the evolution of these models, the challenges they face, and the promising avenues for future exploration.

The significance of next-word prediction goes beyond mere convenience in autocomplete suggestions; it mirrors the core of how we communicate and comprehend language. By deciphering the mechanisms that underlie anticipatory language models, we gain insights not only into the advancements made in artificial intelligence but also into the essence of linguistic cognition. This review stands as a comprehensive resource for researchers, practitioners, and enthusiasts eager to navigate the intricate domain of next-word prediction and its transformative impact on natural language processing.

Evolution of Approaches to Predicting Subsequent Words

The early stages of developing subsequent word prediction techniques were dominated by statistical language models, with a notable focus on employing n-gram models and Hidden Markov Models (HMMs). This extensive section meticulously scrutinizes the merits and drawbacks inherent in these traditional methodologies, going beyond evaluating their effectiveness in simpler contexts to investigating the challenges posed by intricate semantic dependencies.

Statistical language models, particularly exemplified by n-gram models, emerged as foundational frameworks in the initial landscape of subsequent word prediction. These models operate based on the probability distribution of word sequences, estimating the likelihood of the next word by considering the preceding n-1 words. The straightforwardness of this approach facilitated widespread adoption, providing computationally efficient solutions for predicting subsequent words. However, the limitation of n-gram models lies in their incapacity to capture extensive dependencies and the nuanced semantic relationships characteristic of natural language's richness.

Simultaneously, Hidden Markov Models (HMMs) gained traction, especially in the realm of speech recognition, as a probabilistic framework modeling transitions between hidden states representing linguistic phenomena. Despite their success in certain applications, HMMs grappled with the intricate nature of language, often falling short in capturing context beyond immediate transitions. The rigid assumption of independence between observed and hidden states presented challenges in encapsulating the dynamic and contextual nature of language, particularly in the face of complex sentence structures.

A thorough examination of the strengths of these classical approaches reveals their effectiveness in managing straightforward contexts. In scenarios where linguistic patterns follow predictable structures and dependencies are short-range, n-gram models and HMMs demonstrated commendable accuracy. Their simplicity and computational efficiency positioned them as preferred choices in the early stages of language modeling endeavors, contributing valuable insights into subsequent word prediction.

However, as the demand for more nuanced language understanding grew, the limitations of these classical methodologies became increasingly evident. Their inability to discern intricate semantic dependencies and their struggle to adapt to the dynamic nature of language hindered their effectiveness in more complex linguistic contexts. Long-range dependencies, prevalent in situations requiring a nuanced understanding of language, posed significant challenges, prompting the exploration of more advanced methodologies.

The challenges associated with intricate semantic dependencies proved particularly challenging. Classical approaches often faltered in scenarios where words derived meaning from their context, demanding a comprehension that extended beyond the immediate surrounding words. The limitations of n-gram models and HMMs became apparent when tasked with predicting words in sentences with layered meanings, metaphors, or idiomatic expressions. This shortfall in their ability to decipher contextual nuances underscored the imperative need for a paradigm shift in subsequent word prediction models.

While the initial endeavors in subsequent word prediction were characterized by the prevalence of statistical language models like n-grams and HMMs, their effectiveness in managing complex linguistic contexts was limited. The merits of these classical approaches lay in their simplicity and efficiency, particularly suited for uncomplicated scenarios. Yet, their constraints, particularly in capturing long-range dependencies and intricate semantic relationships, spurred the transition towards more sophisticated and adaptive methodologies, as we shall explore in subsequent sections.

Advancements in Machine Learning Techniques for Anticipatory Language Models

The evolution of anticipatory language models took a significant leap forward with the integration of machine learning techniques. This transformative shift harnessed the capabilities of algorithms such as decision trees, random forests, and support vector machines, ushering in a new era of innovation. In this extensive section, we delve into a comprehensive exploration of how these methodologies were applied to predict the subsequent word based on contextual features. The primary objective is to scrutinize the intricacies of how these techniques aim to overcome

the inherent limitations present in traditional approaches.

Machine learning, serving as a pivotal catalyst in shaping anticipatory language models, introduced a diverse set of algorithms that departed from the deterministic nature of rule-based and statistical methods. Decision trees, characterized by their hierarchical decision nodes, emerged as crucial tools in capturing intricate decision-making processes. Constructed based on contextual features, these trees facilitated a nuanced understanding of the relationships between words and their subsequent counterparts. The introduction of random forests, an ensemble of decision trees, further elevated predictive accuracy by amalgamating insights from multiple trees.

Support vector machines (SVMs), another cornerstone of machine learning, played an invaluable role in the anticipatory language models landscape. SVMs excel in delineating decision boundaries in high-dimensional spaces, showcasing proficiency in discerning intricate patterns within linguistic contexts. The application of SVMs in predicting the next word involved identifying optimal hyperplanes, and maximizing the margin between different word categories. This unique feature allowed SVMs to navigate the complexities of language with heightened precision.

This section thoroughly explores the utilization of these machine learning methodologies in predicting subsequent words. Decision trees, with their ability to form branching structures based on contextual cues, offered a detailed approach to understanding the dependencies between words. The hierarchical nature of decision trees allowed for the exploration of various linguistic nuances, contributing to more accurate predictions.

The introduction of random forests marked a paradigm shift, introducing an ensemble approach that mitigated the risk of overfitting inherent in individual decision trees. By amalgamating predictions from multiple trees, random forests presented a more robust framework for anticipatory language models. The collaborative nature of random forests allowed for a comprehensive exploration of the contextual landscape, capturing a broader spectrum of linguistic intricacies.

Support vector machines brought a different dimension to the anticipatory language modeling landscape. Leveraging the power of high-dimensional spaces, SVMs excelled in discerning subtle relationships between words. The optimization of hyperplanes in SVMs facilitated a precise demarcation of word categories, allowing for the effective prediction of subsequent words based on their contextual surroundings.

The application of machine learning methodologies in predicting subsequent words addressed several limitations inherent in traditional approaches. Unlike rule-based or statistical models, machine learning algorithms inherently adapt to the intricacies of language patterns. The ability to discern complex relationships, adapt to varying contexts, and generalize from training data contributed to a more dynamic and accurate anticipatory language modeling paradigm.

Despite the notable advancements facilitated by machine learning approaches, challenges persisted. The interpretability of decision trees, the potential for overfitting in random forests, and the parameter sensitivity of SVMs were aspects that demanded ongoing consideration. As the field progressed, researchers and practitioners continued to refine these methodologies, seeking a delicate balance between complexity and interpretability to optimize anticipatory language models for diverse linguistic scenarios.

The integration of machine learning approaches marked a transformative phase in anticipatory language models. Decision trees, random forests, and support vector machines provided a sophisticated toolkit to navigate the complexities of language, offering nuanced insights into contextual dependencies and contributing to more accurate predictions of subsequent words. The journey into the realm of machine learning paved the way for further innovations, setting the stage for the exploration of deep learning methodologies, as we will uncover in subsequent sections.

Revamping Anticipatory Language Models through Deep Learning Innovations

The landscape of anticipatory language models is undergoing a profound transformation, driven by the recent surge in deep learning. In this extensive section, we conduct a comprehensive analysis of key deep learning architectures, specifically exploring recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and transformers. We delve into their remarkable capabilities to capture contextual information and dependencies across extended sequences, with a dedicated focus on the influential impact of pre-trained language models such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer) in achieving cutting-edge performance.

Recurrent Neural Networks (RNNs)

The introduction of recurrent neural networks marked a pivotal moment in the evolution of anticipatory language models. Departing from the conventional treatment of each input independently, RNNs introduced the concept of sequential information processing. This architectural shift empowered RNNs to comprehend dependencies across sequences, making them particularly suitable for

tasks involving sequential data, such as language modeling.

RNNs showcase dynamic capabilities in maintaining an internal memory of past inputs, allowing them to consider context beyond the immediate preceding words. However, a persistent challenge in the form of vanishing gradients limited their effectiveness in capturing long-range dependencies. The vanishing gradient problem emerges as training signals diminish exponentially during backpropagation, hampering the model's ability to learn from distant past inputs.

Long Short-Term Memory Networks (LSTMs)

To address the vanishing gradient problem, long short-term memory networks (LSTMs) emerged as a sophisticated evolution of RNNs. LSTMs introduced a memory cell that selectively stores or discards information over long sequences, facilitating the retention of relevant contextual information. This architectural enhancement significantly bolstered the model's ability to capture dependencies across extended sequences, establishing LSTMs as a cornerstone in anticipatory language models.

The intricate gating mechanisms within LSTMs, encompassing input, forget, and output gates, enable the controlled flow of information within the network. This not only mitigates the vanishing gradient problem but also allows LSTMs to discern and remember crucial contextual cues over extended sequences. Consequently, LSTMs played a pivotal role in achieving more accurate predictions, especially in scenarios where long-term dependencies are integral.

Transformers

The advent of transformers marked a paradigm shift in deep learning, particularly within the domain of natural language processing. Unlike sequential models such as RNNs and LSTMs, transformers introduced a self-attention mechanism that enables the model to weigh the importance of different words within a sequence simultaneously. This parallelization of attention across the entire sequence empowers transformers to capture global dependencies more efficiently than their sequential counterparts.

The transformer architecture, popularized by models like BERT and GPT, comprises encoder and decoder layers. The encoder's self-attention mechanism excels in capturing contextual information, while the decoder facilitates the generation of subsequent words in a sequence. Originally designed for machine translation, this architecture has demonstrated remarkable efficacy in various language-related tasks, including next word prediction.

Influence of Pre-trained Language Models

The impact of pre-trained language models like BERT and GPT cannot be overstated in recent anticipatory language model advancements. Trained on extensive and diverse textual data, these models learn rich contextual representations that encapsulate the intricacies of language. Their pre-trained nature allows them to transfer knowledge from pre-training tasks to downstream tasks, such as subsequent word prediction, with exceptional success.

BERT, adopting a bidirectional context understanding approach, captures information from both preceding and succeeding words in a sequence. This bidirectional contextualization significantly enhances its ability to accurately predict subsequent words. Similarly, GPT, rooted in a generative approach, leverages learned context to generate coherent and contextually appropriate sequences, including predicting subsequent words in a given context.

The fine-tuning of these pre-trained models on specific language tasks further refines their ability to predict subsequent words, achieving state-of-the-art performance in natural language understanding and generation. The knowledge distilled during pre-training empowers these models to capture intricate contextual dependencies, leading to more nuanced and accurate anticipatory language models.

The recent surge in deep learning has not only revolutionized but also redefined anticipatory language models. The transition from sequential architectures like RNNs to more advanced models like LSTMs and transformers has significantly enhanced the models' capacity to capture contextual information and dependencies across extended sequences. The pivotal role played by pre-trained language models, exemplified by BERT and GPT, underscores the impact of leveraging vast amounts of pre-existing linguistic knowledge. This amalgamation of advanced architectures and pre-trained models stands as a testament to the continuous evolution and refinement of anticipatory language models, propelling them towards unprecedented levels of performance and understanding.

Navigating Complexity: Overcoming Challenges in Anticipatory Language Models

Despite considerable advancements in anticipatory language models, a set of enduring challenges and limitations requires ongoing exploration and innovation. This section explores the complexities presented by data scarcity, multilingual adaptability, and real-time contextual integration, offering strategies to navigate and alleviate these hurdles.

Data Scarcity

A foundational challenge for anticipatory language models lies in the accessibility and diversity of training data. The efficacy of learning and prediction relies on comprehensive, varied, and representative datasets. However, obtaining such datasets, particularly in certain languages and domains,

proves challenging. This scarcity impedes the model's ability to generalize across diverse linguistic contexts, resulting in suboptimal performance.

To address data scarcity, researchers are exploring inventive techniques. Data augmentation, for instance, involves artificially expanding the training dataset by introducing variations in existing data, exposing the model to a wider array of linguistic patterns. Transfer learning leverages knowledge from pre-training on a large dataset in one domain to enhance performance in a different, potentially smaller, domain. Additionally, domain adaptation tailors the model to specific linguistic domains, mitigating the impact of data scarcity within those domains.

Multilingual Adaptability

Anticipatory language models encounter challenges in understanding and predicting words across multiple languages. While pre-trained models like BERT demonstrate proficiency in various languages, achieving robust performance across all languages remains a formidable task, especially for languages with limited digital representation and linguistic resources.

Addressing multilingual adaptability involves incorporating diverse language datasets during pre-training to ensure the model becomes adept at capturing language-specific nuances. Ongoing research also explores the development of language-agnostic models capable of seamlessly adapting to diverse linguistic structures, irrespective of the language involved. Enhancing the multilingual capabilities of anticipatory language models demands deliberate efforts to broaden language representation in training data and refine model architectures for greater language agnosticism.

Real-time Contextual Integration

A critical challenge faced by anticipatory language models is the seamless integration of real-time contextual information for accurate predictions. In dynamic conversational settings or rapidly evolving contexts, models must adapt swiftly to new information. Traditional models, particularly those with sequential architectures, may struggle to capture the nuances of evolving conversations, leading to delayed or inaccurate predictions.

To enhance real-time contextual integration, researchers are exploring architectures that dynamically update internal representations based on incoming contextual information. Mechanisms such as attention, akin to those in transformers, enable the model to dynamically focus on relevant parts of the input sequence, allowing anticipatory language models to react promptly to evolving contexts and provide more accurate, contextually relevant predictions.

While anticipatory language models have made significant progress, persistent challenges demand ongoing attention and innovative solutions. Creative approaches such as data augmentation and transfer learning tackle the issue of data scarcity. Multilingual adaptability requires intentional efforts to broaden language representation in training data and explore language-agnostic model architectures. Real-time contextual integration necessitates dynamic architectures capable of promptly adapting to evolving contexts. Confronting these challenges ensures the continuous refinement of anticipatory language models, enhancing their applicability across diverse linguistic scenarios and real-world conversational dynamics.

In-depth exploration of Evaluation Metrics for Anticipatory Language Models

Effectively evaluating the performance of anticipatory language models requires a detailed examination of robust evaluation metrics. This section undertakes a thorough exploration of commonly used metrics such as perplexity, accuracy, and F1 score, elucidating their significance while acknowledging their limitations across diverse contexts.

Perplexity

Perplexity serves as a fundamental metric for assessing language models, quantifying how well a model predicts a given dataset. A lower perplexity indicates superior predictive performance and a betowaasp of underlying language patterns. However, the practical relevance of perplexity may be limited, as it may not directly correlate with user satisfaction or the real-world utility of the model. Evaluating perplexity requires a broader consideration of its alignment with user-centric criteria and the specific goals of the application.

Accuracy

Widely adopted, accuracy measures the proportion of correctly predicted words or tokens, providing a straightforward assessment of correctness. However, accuracy may lack nuance, especially in tasks with imbalanced classes or when certain errors carry more significant consequences. In the context of anticipatory language models, accuracy offers a basic understanding of correctness but may fall short in providing a comprehensive assessment of the model's proficiency, particularly in scenarios with intricate contextual dependencies.

F1 Score

The F1 score, a composite metric of precision and recall, aims to offer a balanced evaluation of a model's performance. Precision measures the accuracy of positive predictions, while recall assesses the model's ability to capture all relevant instances. The F1 score becomes particularly valuable when seeking a balance between precision and recall. However, in the realm of anticipatory language models, where the implications of false positives or false

negatives may vary, interpreting the F1 score requires a nuanced consideration of the specific task and its potential consequences.

BLEU Score

In addition to the aforementioned metrics, the Bilingual Evaluation Understudy (BLEU) score plays a crucial role in evaluating language generation tasks. It quantifies the similarity between generated and reference text, providing insights into the model's ability to produce linguistically coherent and contextually relevant output. However, BLEU has its limitations, especially when dealing with diverse and nuanced language use, as it may not fully capture the richness and creativity inherent in human-generated text.

While perplexity, accuracy, F1 score, and BLEU score represent pivotal evaluation metrics for anticipatory language models, a nuanced understanding of their relevance and limitations is essential. The choice of metrics should align with the specific objectives and challenges of the given task. Future advancements in evaluation methodologies hold promise in further refining our ability to comprehensively assess anticipatory language models. This includes considering both quantitative metrics and qualitative aspects such as linguistic quality and user satisfaction. A holistic evaluation approach will contribute to the continuous improvement and application of anticipatory language models in diverse real-world scenarios.

7. Future Trajectories in Anticipatory Language Models: A Glimpse into Upcoming Research Areas:

As we bring this paper to a close, it is crucial to turn our gaze toward the future and contemplate potential directions that may shape the research landscape of anticipatory language models. This section delves into emerging trends, placing particular emphasis on the integration of multimodal information and underlining the significance of ethical considerations.

Integration of Multimodal Information

The evolution of anticipatory language models is intrinsically linked to the seamless integration of multimodal information. Traditional models have predominantly operated within the confines of textual data, limiting their understanding of the broader context that includes visual and auditory cues. The future promises a paradigm shift with the integration of images, videos, and audio signals, enriching the contextual understanding of language models. This shift towards multimodal capabilities holds the potential to enable anticipatory language models to comprehend and predict user intent across diverse and nuanced scenarios.

Researchers are actively exploring novel architectures that can effectively process and synthesize information from multiple modalities. The convergence of linguistic and visual cues, for instance, can significantly enhance the model's predictive capabilities by considering both textual and visual contexts. Similarly, the integration of speech and language processing is a frontier that holds promise, allowing anticipatory models to decipher the nuances of spoken language. The exploration of these multimodal approaches signifies a transformative era where anticipatory language models transcend the limitations of textual understanding, evolving into more comprehensive and contextually aware systems.

Ethical Considerations in Language Modeling

As anticipatory language models become more sophisticated and ubiquitous, addressing ethical considerations emerges as a critical area for future research. There is a pressing need to develop ethical frameworks and guidelines that govern the deployment and usage of these models. Ethical considerations encompass a broad spectrum, including privacy concerns, biases in language generation, and the potential societal impact of these models.

Privacy concerns stem from the extensive personal data processed by language models. Future research should prioritize the development of robust privacy-preserving mechanisms, ensuring responsible and transparent handling of user data. Addressing biases in language generation is another crucial dimension. Anticipatory language models must be trained on diverse and representative datasets to mitigate biases and ensure fair and inclusive predictions.

Furthermore, researchers should delve into the societal impact of anticipatory language models. Understanding how these models influence communication dynamics, information dissemination, and user behavior is essential. This research could inform the development of models that contribute positively to societal well-being, fostering responsible and ethical use.

Envisioning the future trajectory of anticipatory language models entails navigating the integration of multimodal information and conscientiously addressing ethical dimensions. The research pathway in these directions not only promises to augment the technical prowess of language models but also ensures their conscientious and ethical application in real-world scenarios.

The journey towards more advanced anticipatory language models is intricately woven with our capacity to seamlessly incorporate diverse modalities of information and uphold ethical standards. By navigating these uncharted frontiers, the field of anticipatory language modeling is poised to make significant contributions to natural language understanding, human-computer interaction, and societal harmony. This forward-looking perspective serves as a comprehensive guide, steering researchers and practitioners towards a

future where anticipatory language models evolve into potent, responsible, and contextually savvy tools.

Concluding Perspectives

In conclusion, this comprehensive review provides an intricate grasp of anticipatory language models, navigating the diverse terrains of traditional, machine learning, and deep learning paradigms. By meticulously dissecting the evolutionary trajectory of these models and addressing contemporary challenges with precision, the primary objective of this paper is to serve as a catalyst, propelling further advancements in this pivotal realm of natural language processing.

Commencing with an exploration of traditional methodologies, including statistical language models and Hidden Markov Models, the review unraveled the inherent strengths and limitations shaping their foundational principles. The subsequent foray into machine learning techniques, incorporating decision trees, random forests, and support vector machines, illuminated a trajectory toward overcoming the constraints embedded in classical approaches.

The profound impact of deep learning models, specifically recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and transformers, was elucidated, underscoring their remarkable ability to capture contextual intricacies and dependencies across extended sequences. The transformative potential of pre-trained language models like BERT and GPT in achieving state-of-the-art performance highlighted the paradigm shift witnessed within the dynamic landscape of deep learning.

As the review expanded its horizons, contemplating the challenges and limitations confronted by anticipatory language models, it simultaneously set the stage for future exploration. The integration of multimodal information and the ethical considerations permeating the landscape emerged as prominent focal points for subsequent research. By weaving a narrative that traverses historical footprints, contemporary landscapes, and future horizons, this review seeks not only to document the journey of anticipatory language models but also to inspire and guide the trajectory of advancements in this burgeoning field.

Initiating a deep dive into the challenges faced by these models, the narrative propels the discourse beyond the current state, envisioning a future where anticipatory language models seamlessly integrate multimodal information. This transformative leap aims to enrich their contextual understanding, enabling more nuanced predictions in diverse scenarios. Additionally, the emphasis on ethical considerations underscores the evolving responsibility researchers bear in deploying these models. The spotlight on privacy, bias mitigation, and societal impact signifies a conscientious approach to technology development, emphasizing the need for responsible and ethically aligned anticipatory language models.

In essence, this review functions as more than a mere retrospective; it serves as a guidepost for the future. By encapsulating the evolution, current landscape, and potential trajectories, it lays the foundation for researchers, practitioners, and enthusiasts to chart new frontiers in anticipatory language models. The narrative crafted herein is not just a reflection but a call to action, beckoning the community to contribute to the continuous evolution of these models, ensuring they become not just technically sophisticated but ethically robust and contextually versatile tools in the vast landscape of natural language processing.

References

1. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems.
2. Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP).
3. Vaswani, A., et al. (2017). Attention is all you need. In Advances in neural information processing systems.
4. Devlin, J., et al. (2019). BERT: Pre-training of deep bi-directional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.
5. Brown, T. B., et al. (2020). Language models are few-shot learners. In Advances in neural information processing systems.