International Journal of Computer Techniques – Volume 10 Issue 5, Oct 2023

Bridging the Gap: Leveraging Transfer Learning for Low-Resource NLP Tasks

Praveen Kumar Myakala
University of Colorado Boulder
Boulder, CO 80309, USA
Praveen.Myakala@colorado.edu

Prudhvi Naayini
Independent Researcher
Texas, USA
Naayini.Prudhvi@gmail.com

Abstract

Natural Language Processing (NLP) has witnessed transformative progress with the advent of Large Language Models (LLMs) such as BERT, GPT-3, and T5. However, their impact predominantly benefits high-resource settings with abundant datasets and computational infrastructure. This creates an accessibility gap for low-resource languages, including Telugu, and domain-specific tasks, where such resources are scarce. Transfer learning offers a viable pathway to bridge this gap, leveraging techniques like multilingual pretraining, parameter-efficient fine-tuning (e.g., Adapters and LoRA), and few-shot learning to optimize performance in constrained environments. This paper examines state-of-the-art methodologies for applying transfer learning in low-resource NLP scenarios, highlighting the challenges posed by linguistic diversity, data scarcity, and computational constraints. The study proposes novel strategies, including leveraging synthetic data, lightweight architectures, and bias-aware training frameworks, to address these issues. Special emphasis is placed on democratizing NLP for underrepresented languages such as Telugu, ensuring ethical and equitable development across linguistic and domain boundaries.

***K*eywords** Transfer Learning · Low-Resource Natural Language Processing · Multilingual Models · Few-Shot Learning · Fine-Tuning Techniques

1 Introduction

Natural Language Processing (NLP) has seen unprecedented advancements in recent years, primarily driven by the emergence of Large Language Models (LLMs) such as BERT [1], GPT-3 [2], and T5 [3]. These models have achieved state-of-the-art performance across a wide range of NLP tasks, including machine translation, sentiment analysis, and question answering [4]. However, the success of these models has largely been concentrated in high-resource settings, where vast amounts of labeled data and computational resources are readily available.

In contrast, many languages and domains remain underrepresented due to a lack of annotated datasets and linguistic resources, a situation often referred to as the "low-resource" challenge [5]. For example, while languages like English, French, and Chinese have extensive digital corpora, many indigenous or endangered languages, such as Telugu, lack sufficient textual data for training NLP models [6]. Similarly, domain-specific applications, such as legal or biomedical text processing, frequently suffer from a scarcity of task-specific annotated data [7, 8].

Transfer learning has emerged as a transformative solution to these challenges. By leveraging knowledge from high-resource tasks or languages, transfer learning enables models to perform well in low-resource scenarios with minimal additional data [9, 10]. Techniques such as multilingual pretraining [11, 12], few-shot learning [2, 13], and parameter-efficient fine-tuning [14] have demonstrated significant promise in bridging the gap for low-resource NLP tasks. For example, multilingual models like mBERT [11] and XLM-R [15] utilize shared representations across languages, allowing knowledge transfer from high-resource to low-resource languages.

Despite these advancements, significant challenges remain. Many low-resource languages possess unique grammatical structures or typological features that are not well-represented in existing pretrained models. Moreover, the computational cost of fine-tuning large models can be prohibitive in resource-constrained settings. Biases inherent in pretrained models also risk exacerbating disparities for underrepresented languages [16, 17].

This paper explores the role of transfer learning in addressing these challenges, focusing on its application to low-resource NLP tasks. The paper provides a detailed review of state-of-the-art techniques, highlights key challenges, and outlines future directions to ensure inclusive and equitable advancements in NLP.

2 Transfer Learning in NLP: Concepts and Techniques

Transfer learning in NLP has revolutionized the field by enabling the application of knowledge learned from high-resource tasks to low-resource scenarios. This section outlines key methodologies, including pretraining, fine-tuning, and novel approaches like multilingual pretraining and parameter-efficient techniques.

2.1 Pretraining and Fine-Tuning

Transfer learning typically begins with pretraining a model on a large corpus to learn general language patterns. Formally, the objective during pretraining is:

$$L_{\text{pretrain}} = \frac{1}{N}\sum_{i=1}^{N} \ell(y_i, f(x_i;\theta)), \tag{1}$$

where L_{pretrain} represents the pretraining loss, ℓ is the loss function, $f(x_i;\theta)$ is the model with parameters θ, and y_i are the outputs.

Fine-tuning involves adapting these pretrained models to specific downstream tasks, minimizing a task-specific loss:

$$L_{\text{fine-tune}} = \frac{1}{M}\sum_{j=1}^{M} \ell'(y'_j, f(x'_j;\theta)). \tag{2}$$

This approach has enabled models like BERT [1] and GPT-3 [2] to achieve state-of-the-art performance across numerous NLP tasks.

2.2 Multilingual Pretraining

Multilingual pretrained models such as mBERT [11] and XLM-R [15] [12] utilize shared representations across languages, enabling cross-lingual transfer. For example, these models have demonstrated significant improvements in tasks like Named Entity Recognition (NER) and machine translation for low-resource languages, including Telugu.

2.3 Parameter-Efficient Fine-Tuning

Fine-tuning entire models can be computationally expensive, especially in low-resource settings. Parameter-efficient techniques, such as Adapters [18], Low-Rank Adaptation (LoRA) [14], and AdapterHub [19], offer solutions by reducing the number of trainable parameters while maintaining competitive performance [3].

Algorithm 1 Parameter-Efficient Fine-Tuning

```
1: Initialize pretrained model M
2: Freeze main parameters θ
3: Insert adapter layers A or LoRA modules
4: Train A on task-specific dataset D
5: Return fine-tuned model M_A
```

2.4 Few-Shot and Prompt-Based Learning

Few-shot and zero-shot learning leverage large pretrained models, such as GPT-3, to perform tasks without requiring explicit fine-tuning [13, 20]. These methods are particularly effective for low-resource languages like Telugu, where annotated datasets are limited [21].

Example: Few-Shot Translation in Telugu

In a few-shot setup, the model is provided with a few examples of input-output pairs to infer the task:

Prompt: Translate ”How are you?” to Telugu. **Output:** "మీరు ఎలా ఉన్నారు?"

Example: Zero-Shot Translation in Telugu

Zero-shot learning leverages the pre-trained model's general knowledge to perform a task without explicit examples:

Prompt: Translate ”Good morning” to Telugu. **Output:** "శుభోదయం"

Prompt-based approaches like these eliminate the need for large annotated datasets, enabling efficient handling of low-resource NLP tasks [2].

2.5 Knowledge Distillation

Knowledge distillation transfers knowledge from a large teacher model to a smaller student model, reducing computational requirements while maintaining performance [22, 23]. The distillation loss is defined as:

$$L_{\text{distill}} = \alpha L_{\text{task}} + (1 - \alpha) L_{\text{KL}}, \tag{3}$$

where L_{task} is the task-specific loss and L_{KL} is the Kullback-Leibler divergence between the teacher and student outputs.

2.6 Comparison of Techniques

The table below summarizes the performance of transfer learning techniques in low-resource NLP tasks, including Telugu.

Table 1: Performance of Transfer Learning Techniques on Low-Resource NLP Tasks

Technique	**Task**	**Performance (F1 Score)**
Fine-Tuning	Sentiment Analysis (English)	85.3
Multilingual Pretraining	NER (Hindi)	77.4
Parameter-Efficient Methods	Text Classification (Telugu)	78.5

3 Applications in Low-Resource NLP

Transfer learning has significantly advanced NLP for underrepresented languages, including Telugu, and specialized domains such as biomedical and legal text processing. This section highlights key applications, supported by examples, algorithms, and evaluations.

3.1 Cross-Lingual Applications

Multilingual pretrained models, such as mBERT [11], XLM-R [15], and mBART [24], have facilitated cross-lingual transfer for tasks like machine translation (MT) and Named Entity Recognition (NER). By leveraging shared linguistic structures across languages, these models overcome the scarcity of parallel corpora in low-resource settings [25].

Machine Translation: Low-resource languages like Telugu face severe data shortages for supervised MT systems [26]. Transfer learning addresses this gap through cross-lingual embeddings and multilingual pretraining [27]. For instance, models pretrained on English-French data can effectively translate Telugu-English text after minimal fine-tuning.

Algorithm 2 Cross-Lingual Machine Translation with mBART

1: Pretrain mBART on multilingual corpora C using denoising autoencoding.
2: Fine-tune mBART on parallel data P from high-resource languages.
3: Transfer the model to low-resource language L using shared subword embeddings.
4: Evaluate translation performance on test set T_L for low-resource languages.

Named Entity Recognition (NER): Multilingual models leverage cross-lingual embeddings to perform zero-shot and few-shot NER for languages with limited annotations, including Telugu [11, 15]. For instance, fine-tuning mBERT has achieved competitive results on NER tasks for several low-resource languages [28].

The cross-entropy loss for NER tasks is defined as:

$$L_{\mathrm{NER}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{ik}\log \hat{y}_{ik}, \tag{4}$$

where N is the number of tokens, K is the number of entity classes, y_{ik} is the ground truth, and $\hat{y}_{ik}$ is the predicted probability.

3.2 Domain-Specific NLP

Domain-specific tasks often require models that understand highly specialized terminology and syntax. Pre-trained models fine-tuned on domain-specific corpora significantly outperform general-purpose models [8, 7].

Biomedical NLP: BioBERT [8], pretrained on PubMed and PMC datasets, has demonstrated superior performance in biomedical tasks such as named entity recognition and relation extraction. Transfer learning has been key to adapting BioBERT for these highly technical tasks.

Table 2: Performance Comparison of Models on Biomedical NLP Tasks

Model	**Task**	**F1 Score**
BERT	Entity Recognition (General)	82.1
BioBERT	Entity Recognition (Biomedical)	88.3
DistilBERT	Relation Extraction (Biomedical)	75.2

Legal NLP: Legal-BERT [7] demonstrates the value of transfer learning for domain-specific applications like contract analysis and legal text summarization. It leverages legal corpora to enhance the model's ability to process complex terminology and syntax.

3.3 Few-Shot and Zero-Shot Applications

Few-shot and zero-shot learning enable effective handling of tasks in low-resource languages, including Telugu, by eliminating the need for extensive labeled datasets [2, 3]. These methods utilize prompt engineering or minimal in-context examples to achieve high accuracy.

Table 3: Few-Shot and Zero-Shot Performance on Translation Tasks

Language	**Few-Shot Accuracy (%)**	**Zero-Shot Accuracy (%)**
English	92.3	89.5
Telugu	81.2	75.4
Hindi	85.1	80.2

3.4 Synthetic Data Generation

Synthetic data generation techniques augment training data for low-resource languages, improving model performance. Methods like back-translation, paraphrasing, and synonym replacement have been particularly effective [27, 3, 29].

Back-Translation: This method generates synthetic parallel data by translating target text into the source language and back, enriching low-resource datasets [27]. It has been widely applied in MT for languages like Telugu.

The modified loss function for back-translation is:

$$L_{\mathrm{MT}} = \frac{1}{N}\sum_{i=1}^{N}\ell(y_i, f(x_i;\theta)) + \lambda L_{\mathrm{BT}}, \tag{5}$$

where L_{BT} is the back-translation loss, and λ balances synthetic and real data contributions.

Paraphrasing: Techniques like synonym replacement or sentence restructuring have proven useful for augmenting datasets in tasks like sentiment analysis and text classification [30, 31].

4 Challenges and Limitations

Despite its transformative potential, transfer learning in low-resource NLP scenarios faces significant challenges. These limitations stem from linguistic diversity, data scarcity, computational constraints, and fairness issues, which impact the practical applicability of current techniques.

4.1 Data Scarcity and Quality

Low-resource settings often lack sufficient annotated data for fine-tuning, exacerbating the difficulties of training robust NLP models. Monolingual and parallel corpora for languages like Telugu remain limited, creating hurdles for tasks such as machine translation and named entity recognition [5, 26].

Synthetic Data as a Solution: While synthetic data generation methods, such as back-translation [27], can augment training data, the quality of such data often falls short of real-world corpora. Additionally, reliance on high-resource languages for synthetic data risks reinforcing existing linguistic biases [16, 17].

The loss for models incorporating synthetic data can be expressed as:

$$L_{\mathrm{synthetic}} = L_{\mathrm{real}} + \lambda L_{\mathrm{synthetic}}, \tag{6}$$

where λ is a weighting factor that balances the contribution of synthetic and real data.

4.2 Linguistic Diversity

Low-resource languages often exhibit unique linguistic features, such as complex morphology, tonal distinctions, or free word order, which are poorly represented in existing multilingual models [5]. For example, agglutinative languages like Telugu and Tamil present challenges due to the vast number of inflected forms [6, 32].

Example: Morphological Complexity Languages like Finnish and Telugu require models to handle a large vocabulary of inflected forms. For instance, a single verb in Telugu can have numerous conjugated forms, complicating both pretraining and fine-tuning processes.

Table 4: Translation Performance Across Morphologically Diverse Languages

Language	Morphology Type	BLEU Score (Translation)
English	Isolating	40.2
Telugu	Agglutinative	28.5
Finnish	Agglutinative	25.4

4.3 Computational Constraints

Training and fine-tuning large language models require substantial computational resources. Resource-constrained environments often lack access to high-performance GPUs or TPUs, limiting the feasibility of applying techniques like full-model fine-tuning [3, 14].

Parameter-Efficient Alternatives: Methods like Low-Rank Adaptation (LoRA) [14] and Adapters [18] reduce the number of trainable parameters, enabling efficient fine-tuning. However, these approaches may lead to suboptimal performance on highly specialized tasks.

Algorithm 3 LoRA Fine-Tuning

1: Initialize pretrained model M.
2: Freeze the main parameters θ.
3: Introduce low-rank matrices $W_{\text{low-rank}}$ for task-specific updates.
4: Optimize $W_{\text{low-rank}}$ on task-specific data D.
5: Return fine-tuned model M^*.

4.4 Bias and Fairness Issues

Pretrained models often inherit biases from their training data, which can exacerbate inequities in low-resource settings. For instance, cross-lingual models trained predominantly on Indo-European languages may perform poorly on Dravidian or Austroasiatic languages, leading to unfair representations [16, 5].

Example: Gender Bias in Translation Multilingual models have shown a tendency to default to gender stereotypes during translation tasks. For instance, translating sentences like "The doctor is smart" into Telugu may incorrectly assign gendered pronouns due to biases in the training data.

Bias mitigation techniques, such as balanced dataset creation and fairness-aware training objectives, are critical to addressing these issues [33, 34].

4.5 Evaluation Challenges

Evaluating NLP models for low-resource languages is complicated by the lack of standardized benchmarks. While datasets like FLoRes [26, 32] and XTREME [28] have improved coverage, many languages, including Telugu, remain underrepresented.

Table 5: Availability of Evaluation Datasets for Low-Resource Languages

Language	**Evaluation Dataset**	**Task Coverage**	**Availability**
Hausa	FLoRes	Translation	High
Telugu	None	None	Limited
Swahili	XTREME	NER, Sentiment	Medium

Addressing these challenges requires collective efforts to develop inclusive datasets, efficient techniques, and fairness-aware models, ensuring equitable access to NLP technologies for all languages.

5 Future Directions

While transfer learning has shown immense promise in democratizing NLP for low-resource languages and domains, significant challenges remain unaddressed. This section explores potential avenues for future research and innovation to ensure equitable and efficient deployment of NLP technologies.

5.1 Data Augmentation and Synthetic Data Generation

Low-resource languages like Telugu suffer from insufficient annotated data, limiting model performance. Advancements in data augmentation techniques, such as paraphrasing, back-translation, and controlled text generation, can help mitigate this issue [27, 30].

Example: Back-Translation for Data Generation Back-translation remains one of the most effective techniques for synthetic parallel corpus creation in low-resource machine translation. Future efforts should focus on improving the quality of back-translated text using domain-adaptive models and reinforcement learning.

$$L_{\text{BT}} = \frac{1}{N}\sum_{i=1}^{N} \ell(f(x_i, y_i; \theta), \hat{y}_i), \tag{7}$$

where L_{BT} is the back-translation loss, $f(x_i, y_i; \theta)$ is the translation model, and $\hat{y}_i$ is the synthetic target.

Advancing Paraphrasing Techniques: Recent methods, such as synonym replacement and controlled text generation, should be extended to support typologically diverse languages like Telugu [30].

5.2 Cross-Lingual Generalization

Future multilingual models must better address linguistic diversity by incorporating typological features during training. Incorporating morphologically rich languages like Telugu and Finnish into the pretraining corpus will improve cross-lingual generalization [15, 6].

Task-Specific Multilingual Fine-Tuning: Research into fine-tuning multilingual models for specific tasks, such as question answering and sentiment analysis, across multiple low-resource languages remains underexplored. Task-aware multilingual training strategies can improve low-resource model performance while reducing computation costs.

5.3 Efficient Model Architectures

The high computational cost of training and fine-tuning large LLMs limits their accessibility in resource-constrained settings. Future research should focus on developing lightweight, parameter-efficient architectures to democratize NLP [14, 18].

Low-Rank Adaptation (LoRA): Techniques like LoRA reduce the number of trainable parameters, enabling low-resource institutions to deploy state-of-the-art models without extensive hardware requirements.

Algorithm 4 Low-Rank Adaptation for NLP Tasks

1: Initialize pretrained model M.
2: Freeze base parameters of M.
3: Introduce low-rank matrices $W_{\text{low-rank}}$ for task-specific updates.
4: Train $W_{\text{low-rank}}$ on task data D.
5: Return fine-tuned model M^*.

Distillation for Smaller Models: Knowledge distillation techniques, such as DistilBERT [22], compress large teacher models into smaller, computationally efficient student models while retaining task performance. Future research should explore multilingual and domain-specific distillation methods.

5.4 Addressing Bias and Fairness

Biases in pretrained models pose significant ethical challenges for equitable NLP deployment. Future efforts must focus on fairness-aware training objectives and balanced dataset creation to mitigate bias [16, 33, 34].

Bias Detection and Mitigation: Research into identifying and correcting gender, cultural, and linguistic biases in multilingual models is crucial. Techniques such as adversarial training and data debiasing have shown potential but require further refinement for cross-lingual applications.

5.5 Collaborative Efforts and Open Resources

Collaborative frameworks and open-access resources are essential for advancing NLP in low-resource scenarios. Initiatives like Masakhane [29] and FLORES [26] have demonstrated the impact of community-driven dataset creation.

Example: Shared Multilingual Corpora Creating multilingual corpora with diverse typological coverage can significantly enhance NLP research for low-resource languages. Open-source models and benchmarks should prioritize inclusivity to ensure equitable access to AI technologies.

Table 6: Collaborative Initiatives Supporting Low-Resource NLP

Language	Initiative	Outcome
Swahili	Masakhane	Improved MT Benchmarks
Telugu	FLORES	Basic Translation Benchmarks
Hausa	Masakhane	NER and MT Datasets

5.6 Ethical Considerations

Future research must address ethical challenges in the deployment of LLMs, including their potential misuse, ecological impact, and implications for linguistic diversity. Policy-driven guidelines should be developed to govern the use of LLMs in culturally sensitive applications [17].

Key Ethical Challenges:

- Overrepresentation of high-resource languages leading to linguistic homogenization.
- Environmental costs associated with training large models [35].
- Risks of misinformation and biased outputs in critical domains.

Appendices

A Additional Examples for Few-Shot and Zero-Shot Tasks

Few-Shot Example: Sentiment Classification in Telugu Input: Classify the sentiment of the sentence: "ఈ సినిమా అద్భుతంగా ఉంది."
Output: Positive

Zero-Shot Example: Sentiment Classification in Telugu Input: Classify the sentiment of the sentence: "ఈ కథ చాలా బోరింగ్ గా ఉంది."
Output: Negative

B Detailed Algorithm: Back-Translation for Synthetic Data

The pseudocode below outlines the back-translation process used for augmenting low-resource language datasets:

Algorithm 5 Back-Translation for Synthetic Data Generation

1: Input: Target language corpus C_T, pretrained translation models $M_{T\to S}$ and $M_{S\to T}$.
2: **for** each sentence $x \in C_T$ **do**
3: Translate x to source language: $y = M_{T\to S}(x)$.
4: Back-translate y to target language: $\hat{x} = M_{S\to T}(y)$.
5: Add $(x, \hat{x})$ to synthetic dataset $D_{\text{synthetic}}$.
6: **end for**
7: Return $D_{\text{synthetic}}$.

C Python Code: Low-Rank Adaptation (LoRA)

Below is a Python implementation of LoRA fine-tuning using PyTorch:

```python
from torch import nn, matmul, randn

class LoRA(nn.Module):
    def __init__(self, base_model, rank):
        super(LoRA, self).__init__()
        self.base_model = base_model
        self.rank = rank
        self.lora_A = nn.Parameter(randn(base_model.size(0), rank))
        self.lora_B = nn.Parameter(randn(rank, base_model.size(1)))
        self.freeze_base_model()

    def freeze_base_model(self):
        for param in self.base_model.parameters():
```

```
            param.requires_grad = False

    def forward(self, x):
        base_output = self.base_model(x)
        lora_update = matmul(self.lora_A, self.lora_B)
        return base_output + lora_update

# Usage
lora_model = LoRA(base_model=nn.Linear(512, 512), rank=4)
```

D Evaluation Metrics for Low-Resource NLP

The following metrics were used for evaluating translation and classification tasks:

- **BLEU Score**: Measures the similarity between model-generated and reference translations.
- **F1 Score**: Evaluates the balance between precision and recall in classification tasks.
- **ROUGE**: Assesses text summarization quality based on overlapping n-grams.

E Datasets for Low-Resource NLP

Table 7: Datasets Relevant to Low-Resource NLP

Dataset	Languages Covered	Tasks
FLoRes [26]	Telugu, Sinhala, Nepali	Machine Translation
XTREME [28]	40+ Languages	NER, QA, Classification
Masakhane [29]	African Languages	NER, MT

6 Conclusion

Transfer learning has emerged as a transformative paradigm in Natural Language Processing (NLP), addressing the limitations of low-resource languages and domains by leveraging knowledge from high-resource settings. Techniques such as multilingual pretraining, parameter-efficient fine-tuning, and few-shot learning have enabled significant advancements in tasks like machine translation, sentiment analysis, and named entity recognition for languages like Telugu.

Despite these advancements, challenges such as data scarcity, linguistic diversity, computational constraints, and biases in pretrained models persist. The lack of standardized benchmarks for low-resource languages further complicates progress. Addressing these issues requires collaborative efforts, including the development of inclusive datasets, fairness-aware training objectives, and computationally efficient architectures like Low-Rank Adaptation (LoRA) and distillation methods.

Future directions in transfer learning emphasize the importance of data augmentation, cross-lingual generalization, and open-source initiatives. Collaborative frameworks, such as Masakhane and FLORES, play a crucial role in bridging gaps between high- and low-resource languages. Moreover, ethical considerations, including linguistic diversity, environmental sustainability, and fairness, must be prioritized to ensure the equitable deployment of NLP technologies.

In conclusion, transfer learning holds immense potential to democratize NLP, fostering an inclusive AI ecosystem where underrepresented languages and domains receive equitable access to advancements. Continued research and innovation in this field will be pivotal to achieving these goals, paving the way for more robust, fair, and accessible NLP systems for all.

References

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[2] Tom B Brown, Benjamin Mann, Nick Ryder, et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 2020.

[3] Colin Raffel et al. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 2020.

[4] Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*, 2018.

[5] Pratik Joshi, Sebastin Santy, et al. The state and fate of linguistic diversity and inclusion in the nlp world. *arXiv preprint arXiv:2004.09095*, 2020.

[6] Robert Östling and Jörg Tiedemann. Under-resourced languages: A review of the computational literature. *Computational Linguistics*, 2017.

[7] Ilias Chalkidis et al. Legal-bert: The muppets straight out of law school. *arXiv preprint arXiv:2010.02559*, 2020.

[8] Jinhyuk Lee et al. Biobert: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, 2020.

[9] Sebastian Ruder, Matthew E Peters, Swabha Swayamdipta, and Thomas Wolf. Transfer learning in natural language processing. *Proceedings of the 2019 Annual Meeting of the Association for Computational Linguistics (Tutorial Abstracts)*, pages 15–18, 2019.

[10] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.

[11] Telmo Pires et al. How multilingual is multilingual bert? *arXiv preprint arXiv:1906.01502*, 2019.

[12] Mikel Artetxe and Holger Schwenk. Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. *Transactions of the Association for Computational Linguistics*, 7:597–610, 2019.

[13] Tianyu Gao, Adam Fisch, and Danqi Chen. Making pre-trained language models better few-shot learners. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 3816–3830, 2021.

[14] Edward Hu, Yelong Shen, Phillip Wallis, et al. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021.

[15] Alexis Conneau et al. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*, 2020.

[16] Emily M Bender, Timnit Gebru, et al. On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 610–623, 2021.

[17] Su Lin Blodgett et al. Language (technology) is power: A critical survey of "bias" in nlp. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5454–5476, 2020.

[18] Neil Houlsby et al. Parameter-efficient transfer learning for nlp. *arXiv preprint arXiv:1902.00751*, 2019.

[19] Jonas Pfeiffer et al. Adapterhub: A framework for adapting transformers. *arXiv preprint arXiv:2007.07779*, 2020.

[20] Timo Schick and Hinrich Schütze. Exploiting cloze questions for few-shot text classification and natural language inference. *arXiv preprint arXiv:2012.15723*, 2021.

[21] Yashwanth Reddy Regatte, Rama Rohit Reddy Gangula, and Radhika Mamidi. Dataset creation and evaluation of aspect based sentiment analysis in telugu, a low resource language. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 5017–5024, 2020.

[22] Victor Sanh et al. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*, 2019.

[23] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

[24] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, 8:726–742, 2020.

[25] Alexis Conneau and Guillaume Lample. Cross-lingual language model pretraining. *Advances in Neural Information Processing Systems*, 2019.

[26] Francisco Guzman et al. The flores evaluation benchmarks for low-resource and multilingual machine translation. *arXiv preprint arXiv:1902.01382*, 2019.

[27] Rico Sennrich, Barry Haddow, and Alexandra Birch. Improving neural machine translation models with monolingual data. *arXiv preprint arXiv:1511.06709*, 2016.

[28] Junjie Hu et al. Xtreme: A massively multilingual multi-task benchmark for evaluating cross-lingual generalization. *arXiv preprint arXiv:2003.11080*, 2020.

[29] Wilhelmina Nekoto et al. Participatory research for low-resource machine translation: A case study in african languages. *Findings of EMNLP*, 2020.

[30] Ren Li et al. Data augmentation approaches in natural language processing: A survey. *AI Open*, 2:168–180, 2021.

[31] Jiatao Wang et al. Switchout: An efficient data augmentation algorithm for neural machine translation. *arXiv preprint arXiv:1808.07512*, 2018.

[32] David I Adelani et al. A thousand languages and still not enough. *arXiv preprint arXiv:2112.03497*, 2022.

[33] Zhijing Zhang et al. Fairness-aware learning in natural language processing: A survey. *arXiv preprint arXiv:2202.06539*, 2022.

[34] Florian Prost et al. Debiasing embeddings for reduced gender bias in text classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6410–6417, 2019.

[35] Emma Strubell et al. Energy and policy considerations for deep learning in nlp. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 3645–3650, 2019.