



## Advances and Challenges in Developing Large Language Models for Low-Resource Languages

**Srreyansh Sethi\****American High School, California, USA***\*Corresponding Author:** Srreyansh Sethi, American High School, Fremont, California, USA.**Received:** April 15, 2024**Published:** April 30, 2024© All rights are reserved by **Srreyansh Sethi**.**Abstract**

The development and deployment of large language models (LLMs) have demonstrated significant success in numerous high-resource languages, transforming aspects of communication, business, and technology. However, the application of these advanced AI systems in low-resource languages (LRLs) presents a distinct set of challenges, notably due to the scarcity of data, economic constraints, and the complexity of linguistic diversity. This paper reviews recent advancements in the adaptation of LLMs for LRLs, highlighting the technological innovations and methodological approaches that aim to mitigate these challenges. We discuss the introduction of novel training techniques such as cross-lingual transfer learning, resource augmentation methods, and unsupervised learning strategies that enhance the performance and applicability of LLMs in LRL contexts. Key challenges are analyzed, including data scarcity, linguistic diversity, and the economic implications of deploying LLMs in LRL settings. Case studies are presented to demonstrate the practical implications and successes of these approaches, providing insights into their effectiveness and the ongoing challenges. This review underscores the importance of continuous innovation and the need for collaborative efforts to ensure that the benefits of AI and LLM technologies are accessible across all linguistic landscapes, thus promoting global digital inclusivity. Through a comprehensive analysis of current strategies and future directions, this paper aims to contribute to the growing field of computational linguistics and the development of equitable AI technologies.

**Keywords:** Language Models; Low-Resource Languages; AI; Computational Linguistics; Transfer Learning**Abbreviations**

LLMs: Large Language Models; LRLs: Low-Resource Languages; AI: Artificial Intelligence; NLP: Natural Language Processing; RLHF: Reinforcement Learning with Human Feedback; MMLMs: Massively Multilingual Language Models; ICL: In-Context Learning; X-ICL: Cross-Lingual In-Context Learning; NER: Named Entity Recognition; PEFT: Parameter-Efficient Fine Tuning; BLEU: Bilingual Evaluation Understudy

**Introduction**

The rapid advancement of artificial intelligence (AI) in recent years has been marked by the significant development and deployment of large language models (LLMs) like GPT-4. These models have increasingly found their way into various user-facing applications, reshaping interactions and accessibility of digital information. Ensuring the safety and efficacy of these models in diverse linguistic environments is crucial, particularly given the unique challenges presented by low-resource languages (LRLs) [1].

The integration of advanced AI systems across multiple languages underscores a critical area of concern: the significant disparities in technology access and effectiveness across different linguistic contexts. Low-resource languages, often characterized

by limited data availability and less commercial interest, face a particularly acute challenge. This disparity not only impacts the performance of AI systems but also affects the inclusivity and fairness of technological advancements.

Recent research highlights these challenges. For instance, studies have shown how unsafe inputs can effectively bypass conventional safety measures when translated into LRLs using tools like Google Translate, exposing significant gaps in the security framework of current AI systems [1]. Additionally, discussions on the economic disparities that arise in the deployment of LLMs for LRLs, where excessive token generation leads to higher operational costs, put these languages at a further disadvantage [2].

Addressing these issues, several innovative approaches have been proposed. For example, integrating visual contexts has been suggested to enhance language learning and processing in LRLs such as Urdu and Swahili, demonstrating how visually-derived supervision can compensate for textual data limitations [3]. Furthermore, exploring the adaptability of LLMs trained primarily in high-resource languages like English to LRLs such as Swedish underscores the potential of transfer learning [4].

This paper aims to delve deeper into these advancements, challenges, and the broader implications of deploying LLMs in low-resource settings. We explore how the field is evolving to make AI technologies accessible and effective across all languages, highlighting the need for innovative solutions that ensure equitable access to the benefits of technology. As we navigate through these discussions, the importance of developing robust, inclusive, and versatile AI applications becomes ever more apparent, setting the stage for a more equitable technological future.

## Theoretical background

### Overview of large language models

Large Language Models (LLMs) serve as cornerstones in the advancement of artificial intelligence, driving numerous breakthroughs across various domains. These models, which leverage extensive datasets and advanced training methodologies, have demonstrated impressive capabilities in a range of applications. However, the effectiveness of LLMs in low-resource languages has not kept pace with their achievements in more commonly used languages [5].

Alam., *et al.* critically examine the challenges faced by LLMs when applied to medium-to-low-resource languages. They point out that despite the utilization of advanced training techniques such as Reinforcement Learning with Human Feedback (RLHF) and instruction tuning, the performance of LLMs is often constrained by the lack of sufficient and representative data in these languages [5]. This limitation not only affects the accuracy but also the applicability of LLMs in diverse linguistic contexts.

Addressing the challenge of generalization to underrepresented languages, Cahyawijaya., *et al.* have developed InstructAlign, a novel approach that aligns the learned behaviors from high-resource languages with those of new, low-resource counterparts. This method ensures that the models retain their multitasking capabilities while expanding their linguistic reach, thus enhancing their usability across a broader spectrum of languages [6].

Expanding on the theme of linguistic inclusivity, Sitaram., *et al.* discuss the evolution of Massively Multilingual Language Models (MMLMs). These models are specifically designed to support a wide array of languages, thereby significantly enhancing linguistic accessibility and improving model performance across a global range of linguistic varieties. This development represents a substantial step forward in making AI technologies more universally applicable and effective [7].

Despite these advancements, Shen., *et al.* highlight a critical issue concerning the inconsistency in LLMs' safety across different languages. They note a significant discrepancy in the safety of responses when interacting with low- versus high-resource languages. This observation underscores the urgent need for im-

proving the quality and quantity of pre-training data across diverse languages to ensure that safety and reliability are maintained uniformly across all linguistic interactions [8].

These insights collectively underscore the need for continued innovation in the development of LLMs that are not only powerful and versatile but also equitable in their linguistic capabilities. As LLMs continue to evolve, it is imperative that they do so in a manner that bridges the gap between high-resource and low-resource languages, thus democratizing the benefits of AI technology worldwide.

### Defenition and scope of low-resource Languages

The scope of low-resource languages (LRLs) within the framework of Large Language Models (LLMs) is multifaceted, involving more than just linguistic diversity. These languages present unique challenges that encompass data availability, model training, and equitable technology deployment. Alam., *et al.* draw attention to how the distribution of language data in models like Bloom is often skewed, negatively affecting their performance on a variety of language processing tasks in low-resource contexts. This imbalance demonstrates the critical need for a broader distribution of data that better represents linguistic diversity globally [5].

Building on this understanding, Mao and Yu explore the limitations inherent in traditional machine translation instruction methods, such as MTInstruct, which often fail to provide adequate cross-lingual signals for LRLs. Their research suggests a pivot towards innovative approaches like AlignInstruct, a discriminator-based method designed to enforce cross-lingual alignment, thereby enhancing the model's ability to handle translations involving low-resource languages more effectively [9].

Further emphasizing the importance of inclusivity in AI, Sitar- am., *et al.* provide a comprehensive tutorial on the imperative of extending LLMs to encapsulate LRLs, addressing both technological and linguistic barriers. Their work underscores the significance of linguistic diversity in fostering fairness and inclusivity in AI applications, highlighting the ethical implications of AI technologies [7].

However, as Shen., *et al.* point out, LLMs trained specifically on LRLs tend to be more prone to generating unsafe or irrelevant responses compared to those trained on high-resource languages. This discrepancy underscores the complex challenge of achieving cross-lingual safety and reliability in AI, stressing the need for advanced training strategies that mitigate these risks [8].

Contrasting with conventional wisdom, Holmström., *et al.* challenge the notion that monolingual models inherently outperform multilingual ones in applications involving LRLs. Their research presents compelling evidence that multilingual models can effectively extend their capabilities to languages like Swedish without

the need for language-specific pre-training. This finding advocates for a more nuanced understanding of how LLMs can be optimized across different linguistic landscapes, suggesting that multilingual approaches may offer a more resource-efficient pathway to inclusivity in language technology [10].

Together, these studies highlight the complex landscape of low-resource languages in the realm of LLMs, illustrating the various dimensions—from technical to ethical—that must be navigated to harness the full potential of AI in linguistically diverse settings.

## Key challenges

### Data scarcity

Data scarcity in low-resource languages significantly hampers the efficacy of Large Language Models (LLMs) in recognizing and responding to harmful inputs. A concerning scenario where GPT-4's safety filters can be bypassed with a 79% success rate when inputs are translated into low-resource languages, a stark contrast to less than 1% in English, illustrates a critical shortfall in the robustness of current LLM safety mechanisms across different linguistic contexts [1].

To tackle the acute problem of data scarcity, Kholodna., *et al.* propose an innovative active learning framework that leverages LLM annotations to selectively curate informative samples. This approach not only addresses the paucity of data but also enhances the training efficiency of models, reducing the amount of data required to achieve effective learning outcomes [4].

Further exploring solutions to overcome data limitations, Patwa., *et al.* introduce a method to generate synthetic data to augment the scant real data available for low-resource languages. This strategy enriches the training dataset, significantly improving the model's learning efficiency and overall effectiveness in language processing tasks [4].

Highlighting a similar issue in the realm of programming, Casano., *et al.* discuss the disproportionate distribution of training data available for popular languages like Python compared to low-resource programming languages such as OCaml and Racket. This disparity not only underscores the challenges in data availability but also impacts the development of technology in less common linguistic frameworks [11].

### Linguistic diversity

Alam., *et al.* contribute to this discussion by evaluating the performance of ChatGPT across various tasks and languages, demonstrating a noticeable performance degradation in low-resource languages. Their findings emphasize the substantial impact of inadequate datasets on LLM performance outside of English-speaking settings, reinforcing the necessity for comprehensive linguistic data [5].

Further integrating diversity, Muraoka., *et al.* explore the potential of cross-lingual Vokenization, a technique that integrates visual data during the training of language models. This innovative method not only aligns visual representations with textual tokens but also facilitates a more nuanced and effective transfer of knowledge between high-resource and low-resource languages, potentially revolutionizing how LLMs handle diverse linguistic inputs [12].

### Technological limitations

A significant technological limitation is the economic burden imposed on users of low-resource languages (LRLs) due to the tokenization process in commercial LLMs. Nag., *et al.* point out that LRLs generate more tokens than high-resource languages when processed by LLMs, thereby incurring greater costs. This economic disparity poses a barrier to the equitable access to technology, emphasizing the need for more efficient processing methods for LRLs [2].

Furthermore, addressing the technological limitations of LLMs in low-resource language contexts, Nguyen., *et al.* highlight the challenges posed by sparse data and limited linguistic diversity in training corpora. They propose an unsupervised prompting method that leverages the strong generative capabilities of LLMs trained on high-resource languages to bootstrap performance in low-resource contexts. This method showcases that even without direct supervision, strategic prompting techniques can coax LLMs to achieve competitive performance, demonstrating a viable path to circumventing data scarcity [3].

These insights into the technological hurdles emphasize the complexity of deploying LLMs in diverse linguistic landscapes and underscore the importance of innovative solutions to overcome these barriers, thereby ensuring the inclusivity and effectiveness of language technologies globally.

### Recent advancements

Recent advancements in the field of Large Language Models (LLMs) have dramatically expanded their utility, particularly in the context of low-resource languages. These innovations reflect a targeted effort to bridge the gap between high-resource and low-resource linguistic applications, facilitating broader accessibility and functional parity across languages.

### Techniques for resource augmentation

InstructAlign emerged as a pivotal development in language model training techniques, crafted to integrate the nuances of low-resource languages without compromising proficiency in well-resourced environments. This approach not only deepens the model's linguistic comprehension but also preserves its existing performance benchmarks, as demonstrated by Cahyawijaya., *et al.* [6]. Such cross lingual instruction tuning proves its worth by offering a seamless blend of multilingual adaptability and robust training fidelity.

Following this, the adaptMLLM technology, introduced by Lankford., *et al.* [13], marks a significant stride in customizing multilingual language models for low-resource languages. This framework has been instrumental in elevating machine translation outputs, significantly enhancing BLEU scores for languages like Irish and Marathi, and illustrating the profound impact of precise model tuning in a multilingual context.

Robinson., *et al.* expand on this discourse through a comprehensive analysis across 204 languages, highlighting a performance discrepancy where traditional machine translation systems excel over LLMs in high-resource scenarios but falter in low-resource settings [14]. This discrepancy underscores the crucial need for advancements that specifically address linguistic diversity.

Further exploring the realm of in-context learning, Cahyawijaya., *et al.* discover that while conventional label alignment methods falter, an innovative query alignment strategy can remarkably enhance LLM applications in linguistically underrepresented contexts, thus broadening the scope of LLM usability [15].

Nag., *et al.* address the operational efficiency of LLMs, focusing on reducing token generation in low-resource languages—a factor that often escalates operational costs. Their exploration into pre-processing techniques like code-mixing and transliteration underscores a commitment to cost-effective LLM deployment without sacrificing performance [2].

Shen., *et al.* critically evaluate the effectiveness of instruction tuning methods like RLHF and SFT, revealing their limited efficacy in low-resource contexts and suggesting an inherent pre-training shortfall in current LLM frameworks [8].

Lastly, Holmström., *et al.* and Patwa., *et al.* highlight how synthetic data and Parameter-Efficient Fine Tuning (PEFT) can revolutionize LLM training for low-resource languages. These methods not only mitigate the reliance on extensive annotated datasets but also demonstrate a scalable strategy for enhancing LLM applicability in diverse linguistic settings [4,10].

All the materials and methods that are used to complete the study should be mentioned.

### Transfer learning approaches

Transfer learning has emerged as a pivotal strategy to combat data scarcity in low-resource languages. However, as highlighted by Yong., *et al.*, simplistic application of translation tools can inadvertently compromise safety mechanisms designed for high-resource languages, indicating the need for more sophisticated transfer methodologies [1].

Cahyawijaya., *et al.* delve into cross-lingual in-context learning (X-ICL), presenting novel alignment techniques that substantially

amplify the efficacy of transfer learning, thereby enriching the LLMs' operational scope in multilingual contexts [15].

The integration of LLMs within an active learning framework for data annotation, as discussed by Kholodna., *et al.*, epitomizes a forward-thinking approach in transfer learning, effectively applying learned knowledge to facilitate NER tasks in low-resource languages [4].

### Unsupervised and sem-supervised learning methods

As the dependency on supervised training methodologies wanes, the potential of unsupervised and semi-supervised approaches comes to the forefront, particularly in settings where labeled data is a rarity. Alam., *et al.* explore how these methods can bridge significant gaps, enhancing the generalization capabilities of LLMs across varied linguistic landscapes [5].

The adaptMLLM framework and the pioneering work by Zhao., *et al.* in federated learning exemplify innovative approaches that not only respect user privacy but also boost performance across languages through federated prompt tuning. This strategy highlights a significant reduction in training parameters while maintaining accuracy, showcasing a viable path forward for multilingual [16].

These advancements collectively underscore a transformative phase in LLM development, characterized by increased inclusivity, enhanced efficiency, and broader linguistic applicability, setting a promising trajectory for future innovations in the field.

### Case studies

Successful implementations of Large Language Models (LLMs) in low-resource settings provide crucial insights into the effectiveness of recent advancements and innovative methods. These case studies highlight the transformative impact of LLMs across various linguistic landscapes, showcasing both the challenges and opportunities in enhancing language technologies for global inclusivity.

A notable implementation involves the use of the InstructAlign method by Cahyawijaya., *et al.*, which demonstrated significant performance improvements on Indonesian local languages. By enhancing F1 scores by 5-10% while maintaining capabilities in high-resource languages, this method exemplifies the practical effectiveness of continual crosslingual instruction in real-world applications [6].

Lankford., *et al.* provide another compelling example with the adaptMLLM framework, which fine-tunes Multilingual Language Models (MLLMs) for better translation between English and low-resource languages such as Irish and Marathi. Their efforts resulted in BLEU score increases of up to 40.5 points, illustrating the potential of targeted fine-tuning in multilingual contexts [13].

Robinson, *et al.* utilized the FLORES-200 benchmark to expose stark performance disparities between high-resource and low-resource languages when translated using LLMs like ChatGPT. Their empirical evidence helps both researchers and end-users understand the practical utility of LLMs across diverse linguistic landscapes [14].

Kholodna, *et al.* showcased how an active learning loop incorporating LLM annotations can be applied to Named Entity Recognition (NER) tasks in African languages. This approach significantly reduced costs and achieved near-state-of-the-art performance with far fewer data requirements [4].

Zhao, *et al.* presented a case study on Federated Prompt Tuning, demonstrating its effectiveness in handling multilingual datasets under strict privacy constraints. This method effectively addresses issues related to non-IID data distributions and shows substantial improvements in NLP tasks for low-resource languages [17].

Finally, Cassano, *et al.* provided a case study where they fine-tuned Code LLMs for languages like Julia, Lua, OCaml, R, and Racket using the MultiPL-T approach. Their models achieved state-of-the-art performance on benchmarks designed to test code generation capabilities in these languages, further highlighting the broad applicability of LLMs in specialized domains [11].

These case studies underscore the practical implications of recent advancements in LLM technology, illustrating how strategic innovations can dramatically enhance the accessibility and efficacy of language models in low-resource settings.

## Discussion

The evolving landscape of Large Language Models (LLMs) in handling low-resource languages (LRLs) illuminates a compelling narrative of technological progress coupled with ongoing challenges. This discussion synthesizes insights from recent studies, highlighting the path forward and pinpointing the areas that require urgent attention to ensure the equitable development of AI technologies across diverse linguistic landscapes.

Yong, *et al.* underscore the necessity of redefining AI safety protocols to encompass a multilingual perspective, signaling a shift away from English-centric safety measures. This highlights a broader risk where linguistic inequalities could compromise the safety of all LLM users, advocating for a holistic approach to red-teaming that includes multiple languages, particularly those considered low-resource [1].

Echoing a similar concern for inclusivity, Alam, *et al.* call for enhanced evaluation and benchmarking systems that accurately reflect the capabilities of LLMs in diverse linguistic settings. Without such advancements, the full potential of LLMs remains untapped,

underscoring the need for more refined assessment tools that cater to the subtleties of LRLs [5].

The contributions of Cahyawijaya, *et al.* provide a roadmap for future LLM enhancements. By focusing on crosslingual alignment and the integration of in-context learning (ICL) methods, they pave the way for more robust frameworks that support continuous learning and accommodate extensive language diversity [6,15]. These developments suggest that future LLM designs could greatly benefit from embedding mechanisms that facilitate both the scalability and effectiveness of multilingual AI applications.

Nag, *et al.* highlight an economic perspective, addressing the cost barriers associated with using LLMs in multilingual contexts. Their findings urge a reconsideration of billing and utilization strategies for LLMs to make AI technologies more accessible, especially in low-resource settings, promoting equitable access across all language groups [2].

Nguyen, *et al.* and Zhao, *et al.* introduce innovative methods to bypass traditional barriers in language model training. Nguyen, *et al.*'s use of unsupervised prompting to enhance LLM capabilities without relying on extensive labeled datasets exemplifies how AI can transcend linguistic barriers more efficiently [3]. Concurrently, Zhao, *et al.*'s exploration of federated learning frameworks showcases how linguistic and physical barriers can be overcome, promoting linguistic diversity and social equality through more inclusive language technologies [17].

Shen, *et al.* and Holmström, *et al.* both address the practical aspects of deploying LLMs across different languages. Shen's emphasis on the necessity for inclusive datasets that enhance safety across all languages calls for a systematic overhaul of how data are curated and utilized in AI training [8]. Conversely, Holmström's findings suggest that extensive pre-training on specific low-resource languages may not always be essential, advocating for more adaptable and resource-efficient training approaches [10].

Lastly, Patwa, *et al.* demonstrate how synthetic data and fine-tuning can be strategically used to tackle the challenges in NLP for low-resource settings. This approach not only conserves resources but also sets a precedent for how other domains within NLP could leverage similar methodologies to enhance AI performance in linguistically diverse environments [4].

In summary, while significant strides have been made in enhancing the functionality of LLMs in low-resource languages, persistent challenges such as economic barriers, the need for inclusive training practices, and the development of robust multilingual evaluation frameworks remain. Addressing these issues will be crucial for the advancement of equitable and effective AI technologies worldwide. As this field continues to evolve, the integration



of innovative learning approaches and the refinement of existing methodologies will be pivotal in shaping the future of multilingual AI applications.

## Conclusion

This review systematically explores significant strides and persistent challenges in developing and deploying large language models (LLMs) for low-resource languages (LRLs). Advances in this field are enhancing linguistic accessibility and pushing the boundaries of what AI can achieve across diverse linguistic landscapes. Discussions center around innovative methodologies such as cross-lingual transfer learning, resource augmentation, and unsupervised learning strategies, showcasing their potential to mitigate data scarcity and enhance the functional capabilities of LLMs in LRLs.

Key advancements include the development of techniques that enable LLMs to learn effectively from limited data, overcoming major barriers in technology deployment in low-resource settings. Innovations like the InstructAlign and adaptMLLM frameworks have shown promising results in bridging the gap between high-resource and low-resource languages, offering improved translation and interaction capabilities [6,13]. Moreover, the integration of unsupervised and semi-supervised learning methods signifies a shift toward more adaptive, efficient, and cost-effective AI applications that can serve a global audience [3,17].

The discussed case studies emphasize the practical implications of these advancements, where real-world applications have seen measurable improvements in performance and accessibility. These success stories validate theoretical approaches and highlight the transformative impact of LLMs in enhancing digital inclusivity.

However, challenges remain, particularly in ensuring the safety, reliability, and economic feasibility of LLM deployments across various languages. The economic implications of tokenization, the need for robust multilingual safety protocols, and the uneven distribution of technological benefits underscore the necessity for continued innovation and concerted efforts toward equitable AI development [2,8].

Future research should focus on refining these technologies and methodologies to address the nuanced demands of LRLs. This includes improving data usage efficiency, enhancing model adaptability without extensive retraining, and developing more sophisticated evaluation metrics that reflect the diverse capabilities of LLMs across different linguistic contexts [5].

In conclusion, while significant progress has been made, the journey toward fully realizing the potential of LLMs in low-resource languages continues. Through persistent effort, collaborative research, and innovative thinking, the benefits of AI can be truly democratized, ensuring that no language community is left

behind in the digital era. The implications of this work are profound, suggesting a future where language technology not only transcends linguistic barriers but also fosters global connectivity and understanding.

## Bibliography

1. Yong Zheng-Xin., *et al.* "Low-Resource Languages Jailbreak GPT-4". ArXiv abs/2310.02446 (2023): n.
2. Nag Arijit., *et al.* "Cost-Performance Optimization for Processing Low-Resource Language Tasks Using Commercial LLMs". ArXiv abs/2403.05434 (2024): n.
3. Nguyen Xuan-Phi., *et al.* "Democratizing LLMs for Low-Resource Languages by Leveraging their English Dominant Abilities with Linguistically-Diverse Prompts". ArXiv abs/2306.11372 (2023): n.
4. Patwa Parth., *et al.* "Enhancing Low-Resource LLMs Classification with PEFT and Synthetic Data". (2024).
5. Alam Firoj., *et al.* "LLMs for Low Resource Languages in Multilingual, Multimodal and Dialectal Settings". *\_Proceedings of the 18<sup>th</sup> Conference of the European Chapter of the Association for Computational Linguistics: Tutorial Abstracts\_*, edited by Mohsen Mesgar and Sharid Loáiciga, Association for Computational Linguistics, (2024): 27-33.
6. Cahyawijaya Samuel., *et al.* "InstructAlign: High-and-Low Resource Language Alignment via Continual Crosslingual Instruction Tuning". *Proceedings of the First Workshop in South East Asian Language Processing* (2023): n.
7. Sitaram Sunayana., *et al.* "Everything You Need to Know about Multilingual LLMs: Towards Fair, Performant and Reliable Models for Languages of the World". *\_Proceedings of the 61<sup>st</sup> Annual Meeting of the Association for Computational Linguistics (Volume 6: Tutorial Abstracts)*, edited by Yun-Nung (Vivian) Chen, Margot Margot, and Siva Reddy, Association for Computational Linguistics, (2023): 21-26.
8. Shen Lingfeng., *et al.* "The Language Barrier: Dissecting Safety Challenges of LLMs in Multilingual Contexts". ArXiv abs/2401.13136 (2024): n.
9. Mao Zhuoyuan and Yen Yu. "Tuning LLMs with Contrastive Alignment Instructions for Machine Translation in Unseen, Low-resource Languages". ArXiv abs/2401.05811 (2024): n.
10. Holmström Oskar Jenny Kunz and Marco Kuhlmann. "Bridging the Resource Gap: Exploring the Efficacy of English and Multilingual LLMs for Swedish". *\_Proceedings of the Second Workshop on Resources and Representations for Under-Resourced Languages and Domains (RESOURCEFUL-2023)\_*, edited by Nikolai Ilinykh., *et al.*, Association for Computational Linguistics, (2023): 92-110.

11. Cassano Federico., *et al.* "Knowledge Transfer from High-Resource to Low-Resource Programming Languages for Code LLMs". ArXiv abs/2308.09895 (2023): n.
12. Muraoka Masayasu., *et al.* "Cross-Lingual Transfer of Large Language Model by Visually-Derived Supervision Toward Low-Resource Languages". Proceedings of the 31st ACM International Conference on Multimedia, Association for Computing Machinery, (2023): 3637-3646.
13. Lankford S., *et al.* "adaptMLLM: Fine-Tuning Multilingual Language Models on Low-Resource Languages with Integrated LLM Playgrounds". *Information* 14 (2023): 638.
14. Robinson Nathaniel R., *et al.* "ChatGPT MT: Competitive for High- (but Not Low-) Resource Languages". ArXiv abs/2309.07423 (2023): n.
15. Cahyawijaya Samuel., *et al.* "LLMs Are Few-Shot In-Context Low-Resource Language Learners". ArXiv abs/2403.16512 (2024): n.
16. Zhao Wanru., *et al.* "Breaking Physical and Linguistic Borders: Multilingual Federated Prompt Tuning for Low-Resource Languages". International Workshop on Federated Learning in the Age of Foundation Models in Conjunction with NeurIPS 2023 (2023).
17. Andersland Michael. "Amharic LLaMA and LLaVA: Multimodal LLMs for Low Resource Languages". ArXiv abs/2403.06354 (2024): n.
18. Merx Raphael., *et al.* "Low-Resource Machine Translation through Retrieval-Augmented LLM Prompting: A Study on the Mambai Language". (2024).