Computer Modeling in Engineering & Sciences
Tech Science Press
Doi:10.32604/cmes.2025.0xxxxx

PAPER TYPE (ARTICLE/REVIEW/EDITORIAL, etc.)

LLMs for Low-Resource Languages: Pretraining, Adaptation

Firstname Lastname[1], Firstname Lastname[2] and Firstname Lastname[2,*]

[1]Affiliation 1 (Department/School/Faculty/Campus/…, University/…), City, Postcode, Country
[2]Affiliation 2, City, Postcode, Country
[*]Corresponding Author: Author's Name. Email: author@institute.xxx
Received: Day Month Year; Accepted: Day Month Year; Published: Day Month Year

ABSTRACT: This systematic review explores the recent progress (2020-2025) in the pretraining and adaptation of Large Language Models (LLMs) to Low-Resource Languages (LRLs). It combines major innovations on data-driven augmentation, parameter-efficient fine-tuning (e.g., LoRA, adapters, instruction tuning), and morphologically rich and underrepresented language script-sensitive tokenization. The results highlight the accumulating effectiveness of the culturally aware standards like IrokoBench and BLEnD, and show that the approaches to the lightweight adaptation eliminate many computational costs however maintain the language accuracy. Recent models like TACO and language-specific ones like UrduLLaMA and BengaliBERT can provide examples of how inclusive multilingual models can be scaled. Even though significant improvements have been made, endemic differences in quality of data, geographical coverage and reproducibility limit a fair deployment of LLM. The review focuses on the ethic in AI practice, the development of corpora through communities, and interdisciplinary research collaboration among computational linguist, social scientist, and digital humanists. The task of generating a diversified dataset, typology-conscious modeling strategies, and open-source multilingual benchmarks should be prioritised in future research as one of the possible solutions to the existing digital language gap worldwide.

KEYWORDS: Large Language Models (LLMs); Low-Resource Languages (LRLs); Parameter-Efficient Fine-Tuning (PEFT); Tokenization; Cross-Lingual Transfer; Multilingual Benchmarks

1 Introduction

The rise of Large Language Models (LLMs) in recent years has made them the core component of Natural Language Processing (NLP), with state-of-the-art performance on a wide variety of applications such as machine translation, question answering, and content summarization [1-3]. Such models as GPT-3, PaLM, and LLaMA-2 are endowed with terabytes of data that are sourced at internet scale, and this data is dominated by English and other well-resource languages [4]. These models show impressive results, however in the case of low-resource languages (LRLs) they do not work sufficiently and are characterized by a lack of large-scale digitized corpora, linguistic tools, and computational support [5].

Billions of people across the world speak low-resource languages and ensure that they are at the heart of cultural identity, education, and civic engagement [1]. However, the digital divide between the well- and badly-resourced languages is increasing to the extent that LLM development becomes more and more centralized in rich environments [6]. More recent efforts to ameliorate LRL representation have been the Glot500, BLOOM or SIB-200, with practical adaptation still under challenge. The body lacks a coherent appraisal of the procedures that are under development and testing to assist these languages [7]. The infrastructural asymmetries and data inequality have been conceptually traced continuously throughout the contexts of low resources, explaining the barriers to policy and capacity to implement inclusive development of LLM [8]. To democratise avenues of accessing the capabilities of the English dominant models, approaches that utilise the capabilities of linguistically diverse prompts have been suggested [9].

Moreover, this review article that attempts to address this gap by reviewing the existing (recent) literature on pretraining and adaptation of LLMs with low-resource languages in a systematic way [10]. It introduces a combination of approaches such as further pretraining, generate synthetic data, parameter efficient fine-tuning (PEFT), token writer, ground instruction tuning [11]. This is aimed at putting into relief their advantages, drawbacks as well as their applicability in various LRL instances. Additionally, recent critical reviews place methodological issues, as well as intrinsic limitations, in the context of current research of LLM, especially in terms of evaluation and reproducibility [12]. In such a way, this review added a value in the form of a comprehensive and up-to-date synthesis (2020-2025) of scalable and reproducible and typology-aware methods that are specifically customized to be used in low-resource language environments, through application to LLMs [13].

The main mobility of this review is to implement exponential data gains in data Strategies, Training Pipelines, and Parameter-Efficient Fine-Tuning (PEFT) through a unified Low-Resource Language LRL-based paradigm [14]. By focusing on the multilingual assessment and including the LRLs belonging to various linguistic groups Bantu, Dravidian, or Austroasiatic [15, 16], it stresses the typological inclusivity. Another crucial dimension is practical adaptation, which offers the guidelines toward adopting cost-efficient methods, such as Low-Rank Adaptation (LoRA), and instruction tuning, in the environment where finances or infrastructure are limited [17]. Moreover, the analysis includes an ethical and representational approach penetrating the concerns of cultural equity, representation, and bias of large language models (LLM) when applied to LRLs [18]. The interconnected, multilingual and multi-modular, and dialectal issues of failure to adapt the low-resource languages (LRLs) are requiring integrated approaches to the methodology in order to realize strong adaptation [19].

However, there are a number of issues that restrict the advancement in this area. Corpora imbalance is a serious problem with most large language models (LLMs) being trained on very skewed data towards English and other high resource languages [20]. Inefficiencies caused by tokenizers make the issue more critical especially the morphologically rich or non-Latin scripts, which are sometimes underrepresented [21]. The reproducibility and accessibility is also a problem because not all the LLMs and low-resource language (LRL) benchmarks are open-access and documented [22]. There are cultural misfit and ethical problems as well since LLMs often do not understand cultures delicately or create culturally insensitive products to LRLs [23]. Restricted computing capabilities are an added hurdle particularly to those researchers in areas where LRLs are widely used [24]. Finally, inadequate assessment systems among LRLs do not allow tracking progress consistently [25].

The main aim of this study is to review and in categorise the recent advancements (2020-2025) in the field of LLM pretraining and adaptation to LRLs in particular [26]. One of them is to determine the effectiveness of data-driven methods, including synthetic data generation, manually curated multilingual data pipelines and transfer learning with typologically similar but high-resource languages [27]. Due to the comparison of these approaches to the different linguistic families, the review emphasizes their advantages, drawbacks, and scalability.

One more fundamental goal is to assess the parameter-efficient fine-tuning architectures such as Low-Rank Adaptation (LoRA), adapter layers, and instruction tuning, to adapt LLMs to LRLs [28]. The approaches are especially useful in the case of computationally limited settings, and the article has reviewed the tendency of the methods to lower training expenses without worsening or improving the performance of multilingual models [29]. The research also deals with the issues of tokenizer and subword modeling concerning low-resource and morphologically rich languages [30]. Present tokenizers, which frequently are trained using Latin-script or English based corpora, but not account for LRLs. The approaches surveyed in this review include script sensitivity functions tokenizers, adaptive subword segmentation algorithms that attempt to represent a wider variety of orthographic systems [31]. Lastly, it suggests the possible avenues that researchers and developers working in low-budget environments can engage in or use LLMs in their languages [32].

The intended audience of this review is a wide-ranging group including NLP researchers active in the field of the multilingual and low-resource language systems, developers interested in the approaches to deploying the LLMs to under-representative languages, and the group of policymakers and financiers who want to focus on the future of the equitable and inclusive development of AI [2]. It is also relevant to linguists and digital humanists who deal with language preservation and the documentation of the endangered languages [33]. Additionally, the review is applicable to Global South developers who need low-compute solutions to implement LLM products in their native ecosystem [34].

2 Methodology

The methodology of this review was examined and synthesised in a systematic manner, the recent developments about the adaptation of Large Language Models (LLMs) to low-resource languages (LRLs) [35]. The field of the review is deliberately narrowed to the academic literature of 2020-2025, so that the analysis could be rather recent and topical in the needed manner [36]. This temporal restriction enables the review to pay special attention to recently advanced strategies and methodologies that appeared in the setting of the dynamic innovation of LLM research [37, 38]. Integrated LLM playgrounds that facilitate rapid prototyping and controlled fine-tuning have been proposed as pragmatic environments for multilingual experimentation [39].

The main aim is to provide important research questions that are fundamental in having some understanding of the state of LLM adaptation in LRLs [40, 10, 11]. Such questions comprise: What pretraining or adaptation techniques are currently employed for LRLs? How effective are these methods across diverse linguistic typologies, scripts, and corpus sizes? And what practical and ethical challenges arise when implementing these strategies in real-world LRL contexts? [17]. The literature review is designed to answer these questions and consists of a combination of both bibliographic databases (including Google Scholar, ACL Anthology and arXiv repositories) as well as some more recent benchmark documentation (Flores, MasakhaNER, and Taxi1500) [41]. These resources are strong bases to identify and discuss researches making contributions to the field.

Moreover, analyses of multi-epoch and two-stage multilingual training strategies indicate notable gains in low-resource settings when training schedules are carefully controlled. Analyses of multi-epoch and two-stage multilingual training strategies indicate notable gains in low-resource settings when training schedules are carefully controlled [42]. Self-supervised prompting has been shown to accelerate cross-lingual transfer without large labeled datasets [43, 44]. Code-switching curriculum learning frameworks show promise in sequencing multilingual exposure for improved transfer [45].

Figure 1 illustrates the distribution of data sources utilized by the surveyed studies, highlighting the relative contributions of various data types to the final corpus. Synthetic data proves to be the most common, followed by parallel corpora, which suggests the increased use of artificially data generating datasets to reduce the issue of data scarcity in low-resource languages (LRLs) [27]. Monolingual data and knowledge-graph-based data also represent quite important shares, and data gathered by the means of web crawling and other tools occupy relatively insignificant positions in the collection [46].

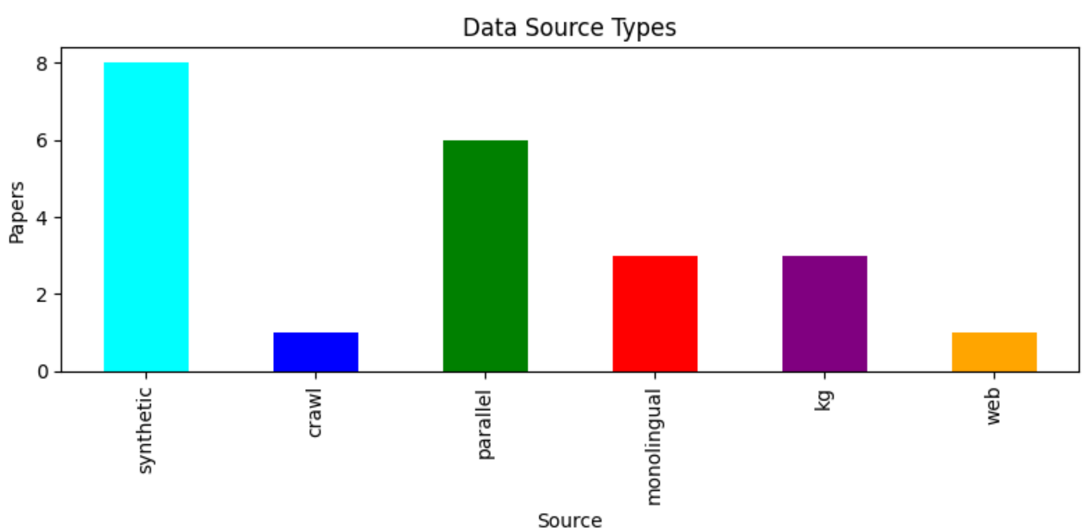

Figure 1: Types of Data Sources Used by Surveyed Studies

The entire review was given importance to reproducibility and accessibility and greater ranks were assigned to the studies that were freely published and well documented [23]. The discussed focus aims to meet a crucial gap in the field, where the presence of many popular large-language models (LLMs) in addition to LRL benchmarks can only be accessed via in-person communication, impeding the development and activity [24]. In addition to that, the review makes use of culturally sensitive assessment structures, so that the analysis of the deployment of LLM can be sensitive to contextual cultural and linguistic peculiarities in different cultural and linguistic settings [33]. Within the scope of identifying, adopting, and developing such structures, the review provides a methodologically sound and holistic study on the adaptation of the LLM to LRLs [47, 26].

2.1 Inclusion and Exclusion Criteria

To ensure the quality, relevance, and focus of the studies included in this review, a set of well-defined inclusion and exclusion criteria was applied during the selection process. The inclusion criteria were designed to capture studies that contribute meaningfully to the understanding and advancement of Large Language Models (LLMs) for low-resource languages (LRLs).

Inclusion Criteria

- Only studies published or preprinted between 2020 and 2025 were included to reflect the most recent and relevant advancements.
- Articles must be available through peer-reviewed venues (e.g., ACL, EMNLP) or reputable preprint platforms such as arXiv to ensure academic rigor and credibility.
- Studies must directly focus on the training, adaptation, or evaluation of LLMs specifically for low-resource languages—addressing key challenges like data scarcity, linguistic diversity, and compute constraints.
- Selected works must present either a quantitative or qualitative empirical evaluation of methods, offering measurable insights or novel strategies applicable to LRLs.

Exclusion Criteria

- Studies centered exclusively on high-resource languages (e.g., English, Mandarin) were excluded due to lack of relevance to LRL-specific challenges.
- Papers that mention LRLs however do not propose concrete methods or experiments were excluded for not offering actionable or replicable findings.
- Non-English articles were generally excluded unless a reliable translation or English-language summary was available, to maintain consistency and accessibility for a broad scholarly audience.

These inclusion and exclusion criteria eliminate several possible approaches of the review to make up a thorough yet focused study of the emerging methodology and strategies to develop LLMs in low-resource language environments. This level of systematization does not only increase the reliability of the results but also gives a clear guideline to the further research and the actual implementation of the findings in this field [10].

2.2 Data Collection and Review Process

The research was conducted sequentially throughout the process of data collection and review to incorporate high quality but relevant literature which is able to address the purposes of the study. This initial search strategy was to query major academic search engines, such as Google Scholar and ACL Anthology and arXiv using special terms i.e. low-resource language modeling, multilingual LLMs, continued pretraining, LoRA adaptation and synthetic corpora to LRLs. The identified keywords formed an attempt at including as many studies of the various avenues of Large Language Models (LLMs) in low-resource language (LRL) as possible. The search retrieved 812 articles in the first search and indicated how much interest and research was conducted regarding the field [7].

In order to simplify the study selection process, the study selection was done according to PRISMA to be transparent and reproducible. Figure 2 below demonstrates that identification of relevant studies into the programming started by exhaustive searches on databases and grey literature. The initial search identified the total number of 812 records. The review finally comprised 140 studies after the process of removing duplicates and irrelevant studies. The PRISMA flowchart gives a comprehensive report of such selection process.

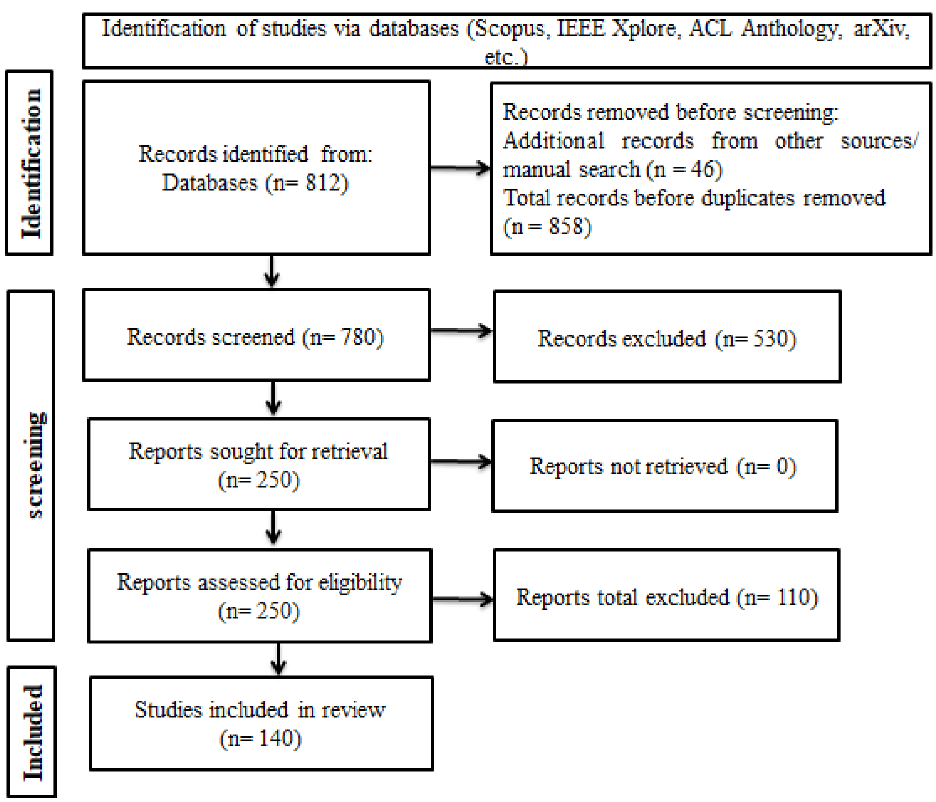

Figure 2: PRISMA Flowchart

This general data was logically categorised into the most significant thematic groups, with the introduction of a set of benchmarks to assess the performance of LLDMs, the development of fine-tuning with minimal parameters (PEFT) methods, including Low-Rank Adaptation (LoRA), and adaptation to different cultures [11]. The PRISMA flowchart offers a comprehensive description of the careful selection process, so that the review would report the latest and most significant developments in the field of the LLM adaptation into LRLs [7]. By utilising this deep data, the review integrates scalable, reustatable and typology sensitive methodologies that are specifically designed to fit low-resource language context and to undergo significant scalability gaps concerning data sparsity, linguistic heterogeneity, and computational inefficiency [31]. Such a method would both increase the accuracy of the results as well as it will create a solid platform to inform future studies on the subject and practical applications on the same in the future [23].

2.3 Literature Synthesis Framework

In order to cope with the difficulties and opportunities provided by applying large-scale models (LLM) to the low-resource languages (LRLs), the review takes a systematic approach that generalises recent developments, classifies methods, and assesses them in terms of their effectiveness. The methodology is organized and centered around four critical dimensions that being data strategies, model adaptation strategies, linguistic dimensions, and ethical dimensions, and thus covered in addressing the special needs of LRLs.

The basis of the performance of LLM was quality and variety of training data, and data scarcity was a major problem of LRLs. In order to reduce these shortcomings, this review examined a number of data-supported methods. Methods like synthetic data generation based on bilingual words lexicon [48] and using high resource languages in greater transfer [27] are discussed in terms of scalability and applicability. Additionally, the possibility of unified frameworks such as UnifiedCrawl [15] and Latxa [49] was examined in the light of aggregating and processing multilingual data in an efficient manner. The cross-lingual transfer through typologically related languages as well was reconsidered with Zhang et al. and Feng et al. studies pointing to the possibility of using related languages to enhance the work of LRL [5, 50]. The training of pre-trained LLMs into LRLs requires new methodologies that strike a balance between the cost and increase of performance [51].

Moreover, the reviewed articles explored parameter-efficient fine-tuning (PEFT) methods in the form of Low-Rank Adaptation (LoRA) [28], adapter layers [22], and instruction tuning [52], and each of them targets to minimize the cost of training models and not necessarily affect or diminish the ability of the models. Such strategies as continuous pre-training [37] and domain-adaptive learning [53, 54] were evaluated based on their efficiency in developing the ability of the LLM to improve when applied to LRLs. Besides, the studies by Toukmaji and Flanigan indicate that customized prompts have the power to induce in-context learning in LRLs without heavily relying on fine-tuning, which can be used as a low-cost solution in the context with resource limitations [34].

To achieve successful adaptation, it was important to speak about linguistic peculiarities of LRLs. Migration issues linked to scripts of morphological richness or non-Latin are addressed with adaptive tokenisation methods [30, 31], which include both script-sensitive tokenisers and dynamic sub-words segmentation algorithms. A study by Remy et al. highlights the essence of harmonization among vocabularies among languages to promote an excellent level of cross-linguistic understanding [55]. Additionally, these methods as VerbCraft [33] cost underline the need of models to support the complex grammar make-up of LRLs, thus guaranteeing linguistic inclusiveness and higher capacity. The cultural equity and representation in the development of LLM is one of the areas of the corner of this review. The debiasing and detoxification of multilingual LLMs [56] are mentioned and the need to create fair and inclusive language models emphasized. Research on the aspects of developing LLM to be culturally sensitive and prevent culturally insensitive results, such as the one conducted by Sundar et al. seeks to explain how a culturally aware application may be achieved [18]. Highlighted as a necessary element of a fair approach to the development of AI, especially in the context of low-resource settings, are open benchmarks and well-documented methods [22, 57].

In order to have a rigorous analytical framework, this review takes a holistic evaluation paradigm. As benchmarks, the performance of the LLM in various language groups (including Bantu, Dravidian, and Austroasiatic languages) is evaluated using BasqBBQ [58] and LEIA [21]. The review gives higher priorities to those methods that can be applied and reproducible to maintain their relevance in a resource-constrained setting [24]. The review can ensure high levels of typological inclusiveness and practicality by focusing on languages with a typological diversity to ensure that its results can be applied to a wide range of LRLs [15]. With elements of quantitative, qualitative and thematic analyses, this framework allows to shed light on a complete picture of the state of adaptation of Large Language Model (LLM) to low-resource languages (LRLs).

Quantitative Aggregation of Performance Metrics

Quantitative measures were aggregated where possible in numerical measures like BLEU scores, perplexity, and F1 scores to allow comparisons between studies directly. The metrics are common in the evaluation of language models and they give us a consistent measuring unit to measure up the performance of models using various datasets and languages. As an example, BLEU was especially good at evaluating machine translation tasks, whereas perplexity provided answers about fluency and coherence of the resulting text [59]. Human translation feedback loops applied in interactive systems such as ParroT improve alignment and translation quality [60]. The combination of these measurements allowed us to conclude which approaches exhibit the best results on a regular basis in situations involving LRL [1]. Figure 3 shows the most-used evaluation metrics across the surveyed literature.

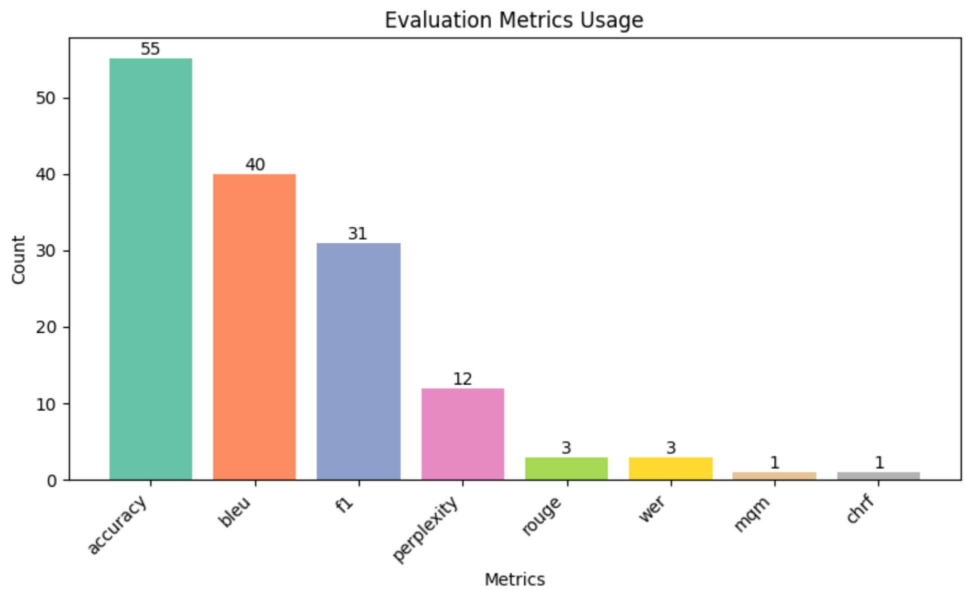

Figure 3: Most-used evaluation metrics across the surveyed literature.

Qualitative Assessment of Ethical Risks and Deployment Considerations

In addition technical performance, the question of applying LLMs in low-resource language environments raised much interest, ethical, and pragmatic review issues as well. Among the issues that were covered in the course of this qualitative assessment were the issues of cultural fairness, representation, and prejudice, when modeling was used to present adaptations to languages that have smaller digital presence and unusual sociolinguistics features [28, 36]. Other studies with regards to the importance of culturally conscious training information and refinement activities received special prominence whereas this factor is crucial in terms of the adoption of LLMs in an equitable and inclusive way [61]. The matters related to the constraints of the infrastructure, reproducibility, and access were also addressed, among them being how such matters can be mitigated in a practical context [14, 22].

By systematically reviewing these dimensions, this study aims to provide a comprehensive and up-to-date synthesis of scalable, reproducible, and typology-aware methods tailored for LRL environments.

3 Findings

The results of the systematic review were based on the review of 140 works devoted to applying Large Language Models (LLMs) to low-resource languages (LRLs). The flowchart presented in figure 2 shows the high level of rigor used in identifying, screening, and selection of these studies as is reflected by PRISMA flowchart. The results were categorised into some of the major themes: the creation of benchmarks to analyze the performance of LLMs, implementation of parameter-efficient fine-tuning (PEFT) techniques such as LoRA, as well as the relevance of culturally minded adaptation strategies. Altogether, these themes emphasize the developments and obstacles in the advancement of the LLM properties of underrepresented languages.

3.1 Key Themes in the Findings

Development of Benchmarks for LRLs

A substantial part of the research on LLMs discussed attempts to explore new standards of assessing the effectiveness of low-resource language tasks. Adelani et al. can be taken as an example: they proposed using a culturally based assessment system, IrokoBench, designed specifically to challenge African languages [1]. In an analogous operation, Myung et al. created BLEnD, a multicultural benchmark, whose goal was to evaluate the success of the LLM in different linguistic and cultural settings [7]. These guidelines offer standardized metrics on the adjustment of LRL as well as an understanding of how to improve it. Additionally, benchmarks like BasqBBQ [58] and LEIA [21] provide critical insights into social biases and cross-lingual knowledge transfer in LRLs. Figure 4 illustrates the timeline of major dataset/benchmark releases relevant to LRL-LLM research (2020–2025).

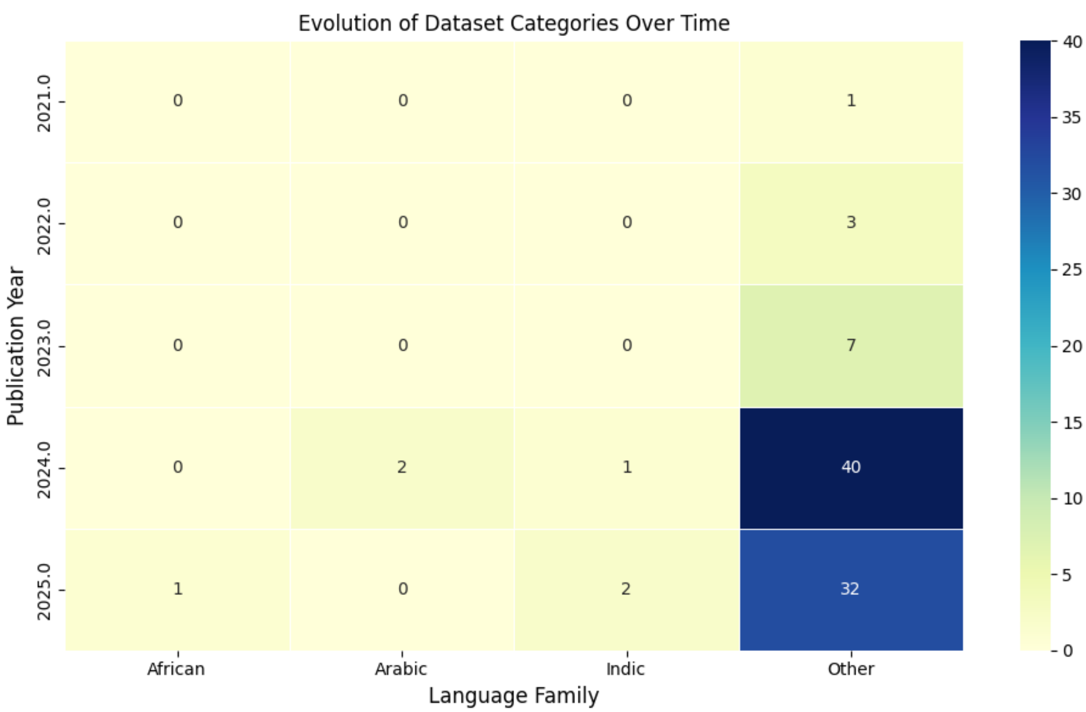

Figure 4: Timeline of major dataset / benchmark releases relevant to LRL-LLM research (2020–2025)

Application of Parameter-Efficient Fine-Tuning (PEFT) Methods

The other significant theme is the PEFT methods to fine-tune LLM on LRLs. Efficiency of the LoRA (Low-Rank Adaptation) was shown in the studies like Le et al. that proved helpful in enhancement of LLM classification in clinical NLP tasks with few data [10]. Khade et al. also discussed a possibility of using LoRA of multilingual South Asian LRLs, which is appropriate because of the low-compute adaption possibility [11]. These approaches have become efficient ways in fine-tuning LLMs under limited resources. Additionally, Khoboko et al. demonstrated the optimization of translation for low-resource languages through efficient fine-tuning with custom prompt engineering in large language models [28]. These approaches have become efficient ways in fine-tuning LLMs under limited resources. Figure 5 shows the frequency of adaptation strategies reported across the surveyed studies. Similarly, reported performance gains vs. compute cost (proxy) for PEFT and full fine-tuning across studies showed in Figure 6.

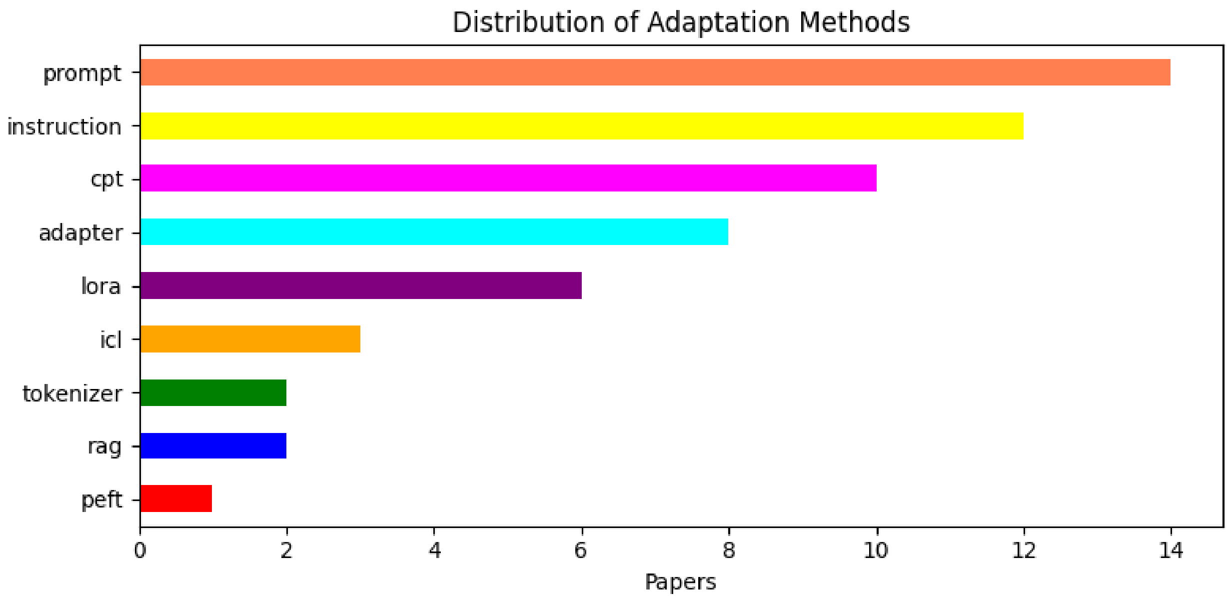

Figure 5: Frequency of adaptation strategies reported across the surveyed studies.

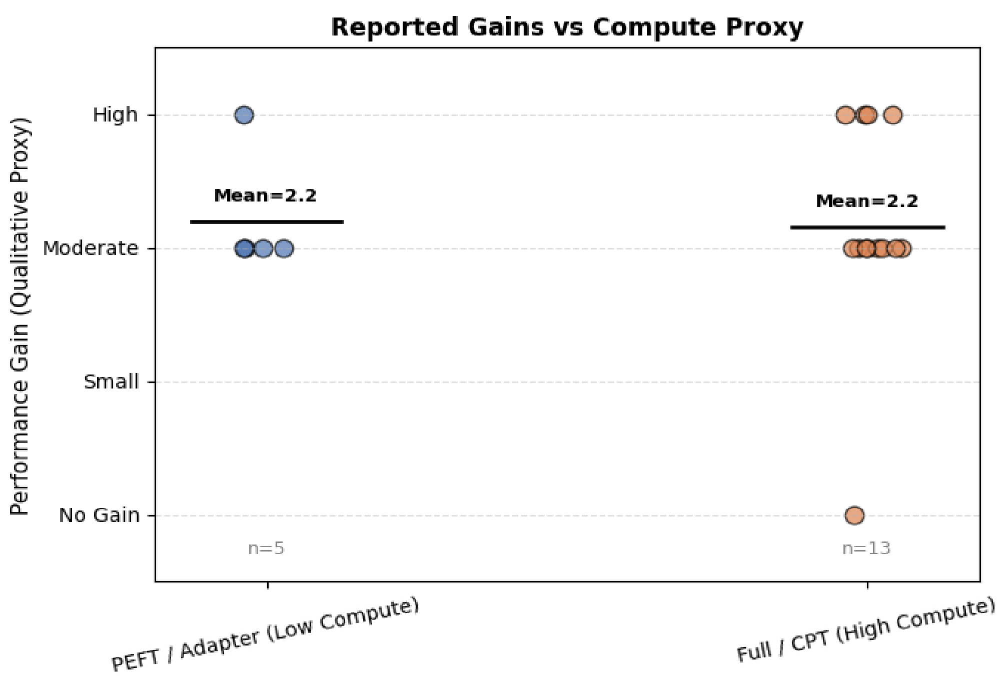

Figure 6. Performance gains vs. compute cost (proxy) for PEFT and full fine-tuning across studies

Additionally, Figure 7 displays the number of papers focused on specific tasks within the reviewed literature on LLM adaptation for low-resource languages. Translation is the most frequently addressed task (18 papers), followed by Named Entity Recognition (NER, 15 papers) and Part-of-Speech (POS) tagging (11 papers). Other tasks, such as speech, summarization, sentiment analysis, safety, dialogue, and classification, are less frequently studied. As an example, recent papers like Mao and Yu and Zhang et al. emphasize improvements in the translation of low-resource languages for better results, which are achieved with new methods, i.e. contrastive alignment instructions, books of code-augmented grammar, and so on [20, 62] . On the same note, such studies of named-entity recognition and part-of-speech tagging as considered by Subedi et al. and Shibu et al. were topical with respect to the fake news detection and linguistic pattern recognition applications [63, 64].

Additionally, Benkirane et al. and Shen et al. identified the existence of sentimental analysis and safety issues in their studies [65, 66]. These papers scrutinize cross-linguistic sentiment analysis and multilingual safety problems, individually. Purwarianti et al. investigate the concept of dialogue systems and summarization, with NusaDialogue framework solving the problem of dealing with languages that are underrepresented [61]. However, the given findings showed the range of different kinds of work that was done in the field of research on low resources languages, at the same time pointing at the areas that still need to be researched.

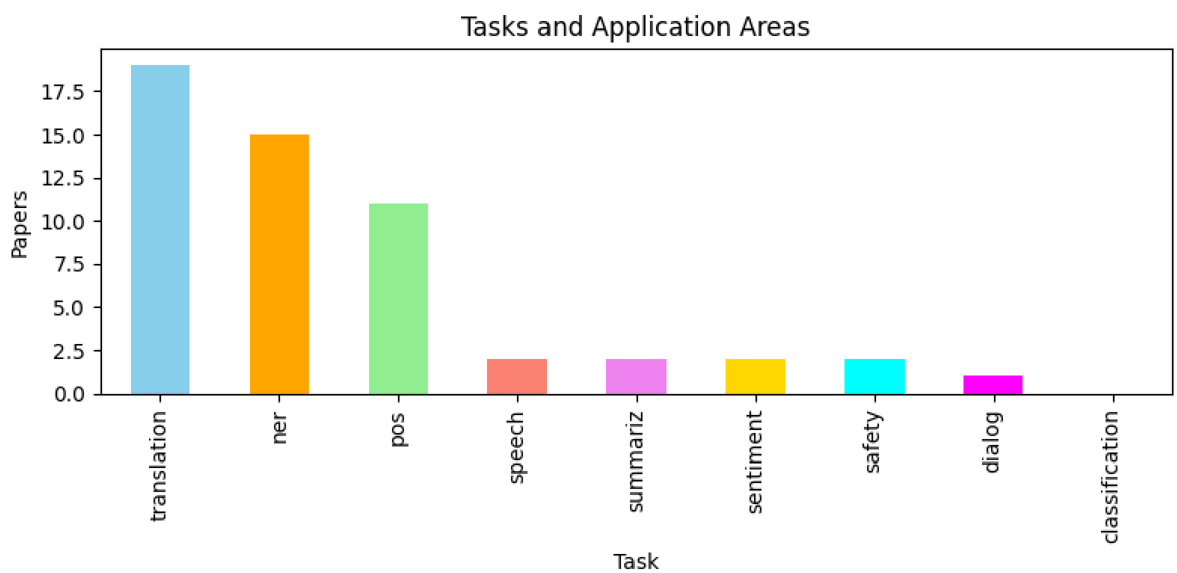

Figure 7: Tasks and Application Areas across surveyed studies

Cultural Sensitivity in Adaptation Strategies

Some studies highlighted the urgency of culturally-sensitive adaptation measures to solve such problems as prejudice and mismatch in LLM output. Specifically, the case studies in Moroccan Arabic dialectal adaptation that are proposed by Mekki et al., Atlas-Chat, and the culturally adapted fine-tuning approaches where the data trained on African LRLs were offered

[67, 68]. Such attempts indicate the necessity of language-sensitive context-sensitive and linguistically broad LLMs.

One of the most striking is the TACO framework suggested by Upadhayay and Behzadan that can achieve the relevant cross-lingual reasoning in LRLs by means of the translation-aided chain-of-thought procedures [69]. Additionally, Kazi et al. presented UrduLLaMA, the first spell-correcting morphologically rich language foundation model pipeline [70], and such fine-tuned solutions could be applied to combinations of domains too small to interest the model designers [71-73].

Questions of ethics have not been excluded in the recent studies either. The harm of bias and representational damage in the process of adapting LLMs to LRLs development, insisting on the change of responsible AI [74]. Besides, Furthermore, Shen et al. introduce methodologies for detecting safety challenges in multilingual contexts, emphasizing the need for equitable access to LLMs [66]. Zhong et al. introduce a socio-technical access model that can support access to LLMs to conduct humanities research in the Global South [75]. Figure 8 illustrates the proportion of studies within our review that explicitly address ethical considerations such as bias, toxicity, safety, or hallucination. A small but significant minority (5.2%) of the reviewed papers focus on these critical issues, highlighting an emerging, though still underdeveloped, area of research in LLM adaptation for low-resource languages.

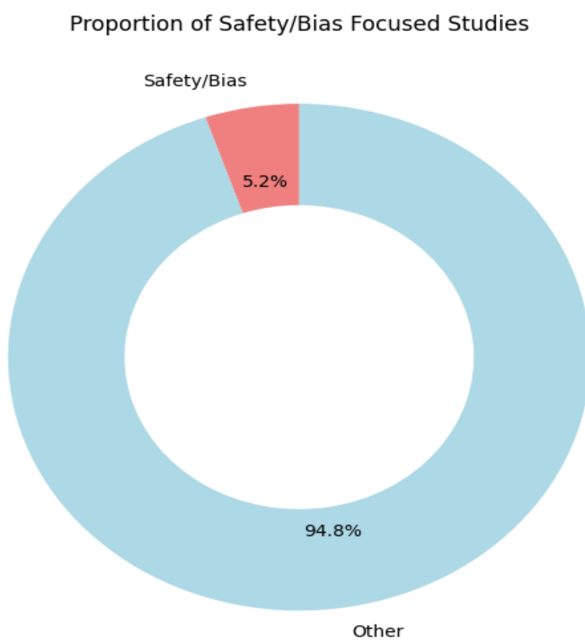

Figure 8: Proportion of safety/bias focused studies

Theoretical Contributions to LLM Adaptation for LRLs

The theoretical Contribution is an outline of the theoretical progress in the Large Language Models (LLMs) adaptation to the Low-Resource Languages (LRLs). It shows an extensive picture of the advances, techniques, and models brought forth by the new research, as well as the advancement and difficulties in this highly developing segment. Table 1 is a particular 20 study and gives us information in detail concerning the content of where the study is conducted and published and, the theoretical contribution that the study makes in the area of LLMs in relation to LRLs.

Table 1: Theoretical Contribution Summary for 20 studies

No.	Year	Study Country	Publication Country	Author(s) & Year	Theoretical Contribution
1	2024	Nigeria, Africa (Multi-country)	USA (arXiv)	Adelani et al. [1]	Proposes a new benchmark (IrokoBench) for evaluating LLMs on African languages; introduces culturally grounded evaluation framework.
2	2024	Latvia	USA (ACL)	Darģis et al. [4]	Provides evidence of open-source LLM performance in morphologically complex LRLs via real-world exams.
3	2025	Canada	USA (arXiv)	Le et al. [10]	Shows how LoRA improves LLM classification in clinical NLP with limited data—contribution to PEFT strategies.
4	2024	China	Netherlands (Elsevier)	Hu et al. [13]	Develops “Language Fusion via Adapters” theory for LRL speech recognition integration.
5	2024	Multinational	USA (arXiv)	Myung et al. [7]	Introduces BLEnD as a multicultural benchmark, contributing a new evaluation paradigm for LLMs across cultures.
6	2024	India	USA (arXiv)	Verma et al. [76]	Develops MILU benchmark; expands understanding of multitask learning for Indic LRLs.
7	2024	Morocco	USA (arXiv)	Shang et al. [14]	Presents Atlas-Chat as a case of dialectal adaptation; proposes culturally sensitive LLM adaptation strategies.
8	2024	Multinational (Africa)	USA (arXiv)	Alhanai et al. [2]	Suggests a culturally adjusted fine-tuning method and benchmarking protocol for African LRLs.
9	2022	Bangladesh	USA (arXiv)	Bhattacharjee et al. [77]	Introduces BanglaBERT; theoretical base for language-specific transformer adaptation.

10	2023	Multinational	USA (arXiv)	Imani et al. [78]	Proposes Glot500 multilingual corpus; advances scalable pretraining theory across 500 languages.
11	2024	Global	USA (arXiv)	Lin et al. [79]	Develops Mala-500 and explores massively multilingual adaptation; contributes to scaling LLM adaptation theory.
12	2023	Global	USA (arXiv)	Upadhayay and Behzadan [69]	Proposes TACO: Chain-of-thought + translation framework for cross-lingual reasoning.
13	2025	Pakistan	USA (arXiv)	Fiaz et al. [80]	Proposes UrduLLaMA—fine-tuned foundation model pipeline for a morphologically rich LRL.
14	2024	India	USA (ACL Anthology)	Khade et al. [11]	Studies LoRA PEFT for multilingual South Asian LRLs; highlights low-compute adaptation theory.
15	2023	Global	India	Kalluri [36]	Proposes a conceptual framework for LLM adaptation ethics in LRLs.
16	2025	Kazakhstan	Switzerland (MDPI)	Kadyrbek et al. [81]	Suggests small-scale LLMs with Direct Preference Optimization for resource-constrained LRLs.
17	2024	Pakistan	USA (arXiv)	Tahir et al. [82]	Benchmarks LLMs in Urdu NLP; expands PEFT + multilingual tuning theoretical base.
18	2025	Kenya	USA (arXiv)	Ngugi [83]	Introduces targeted lexical injection into early layers of LLMs to improve alignment in LRLs.
19	2021	Mongolia	Germany (SpringerOpen)	Byambadorj et al. [84]	Applies cross-lingual transfer theory to low-resource TTS systems.
20	2024	Global South	USA (arXiv)	Zhong et al. [75]	Theorizes challenges of LLMs in humanities; proposes socio-technical access model for LRLs.

The geographical distribution of theoretical contributions to LLM adaptation for low-resource languages (LRLs) reveals significant insights into the research landscape. Additionally, Figure 9 displays the number of papers contributed by different countries or regions in the reviewed literature on LLM adaptation for low-resource languages. China leads with 23 contributions, followed by the USA (11), India (9), Japan, UK, and South Korea (each with 4). Other countries, including Germany, Bangladesh, and Pakistan, also show significant contributions.

Furthermore, in South Asia, the focus is on adapting LLMs for languages like Urdu, Hindi, Bangla, and Nepali, which are among the most frequently mentioned in the literature. Figure 10 illustrates the proportion of papers focusing on specific low-resource languages. Urdu is the most frequently mentioned language (30%), followed by Arabic, Bangla, and Nepali (each 20%), and Hindi (10%). This reflects the strong regional emphasis on South Asian languages in LLM adaptation research. However, this distribution also underscores the global interest in addressing the challenges of LLM adaptation for LRLs, with particular emphasis on culturally and linguistically diverse regions [85].

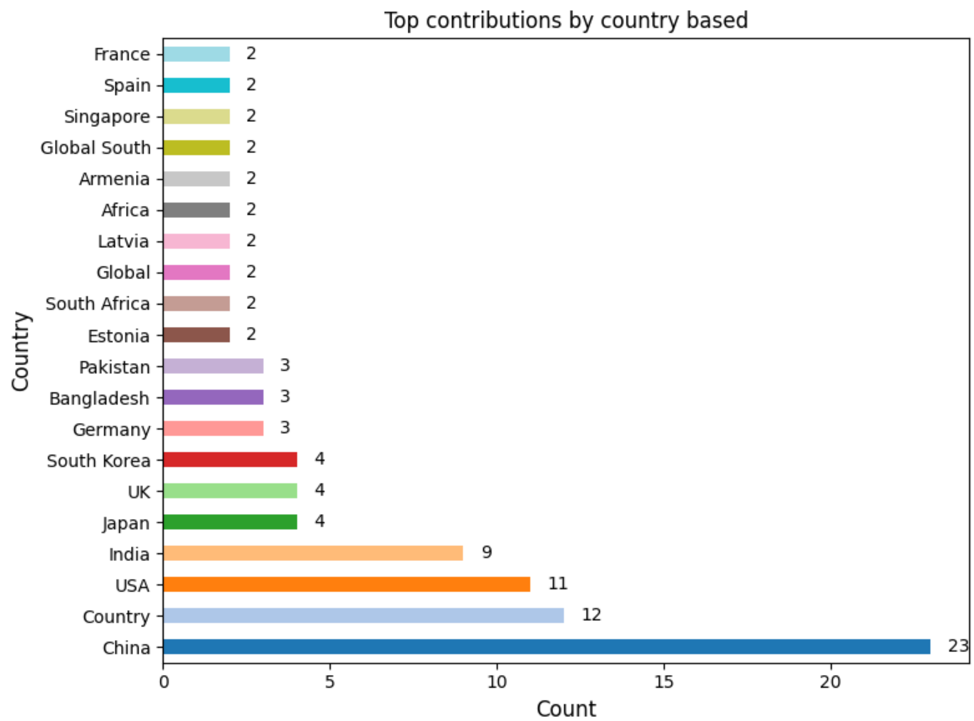

Figure 9: Top Contributions by Country

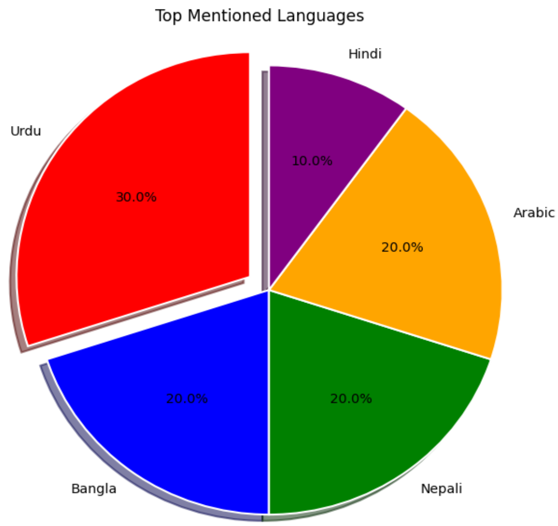

Figure 10. Top Mentioned Languages

Moreover, the analysis of publication trends reveals significant growth in research on LLM adaptation for low-resource languages (LRLs) over the past five years. As illustrated in Figure 11, the number of publications increased steadily from 2020 to 2025, peaking at several papers in 2024. However, there was a slight decline in 2025. This trend underscores the increasing importance of LRL-specific research while also suggesting potential fluctuations in scholarly focus.

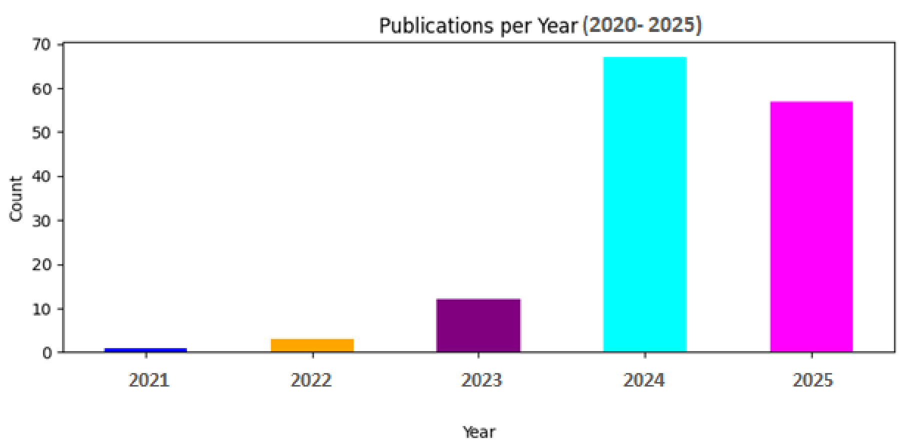

Figure 11: Annual number of publications on LLMs for low-resource languages (2020–2025)

The overview of the adaptation techniques available to tune Large Language Models (LLMs) on low-resource languages (LRLs) in various regions. Table 2 summarises the allocations of four primary adaptation techniques, LoRA, Adapter layers, Tokenizer Tuning and Instruction Tuning, across five areas, namely South Asia, Africa, Europe, North America, and Global/Mixed. In, South Asia ranks first with maximum adaptation count, including the usage of LoRA three times, Adapter layers once, Tokenizer

Tuning once, and Instruction Tuning two times, amounting to a sum of seven. Africa is next with five adaptations, but fewer cases are observed in Europe, North America, and Global/Mixed due to differences in the levels of attention to the topic and allocation of resources in these territories [1, 11].

Table 2: Adaptation Methods by Region based on top 20 studies

Region	LoRA	Adapter	Tokenizer Tuning	Instruction Tuning	Total
South Asia	3	1	1	2	7
Africa	2	1	0	2	5
Europe	1	0	1	1	3
North America	1	0	0	1	2
Global/Mixed	2	0	0	1	3

To supplement the Table 2, Figure 12 provides the same information in a visual manner with the focus on regional differences in the usage of adaptation techniques. The visual representation allows to spot the patterns, e.g., the popularity of LoRA as one of the preferred approaches in various regions, the lack of the use of the Adapter layer and Tokenizer Tuning there. Collectively, the table and the figure serve as a reminder of the diversity in the regions concerning the methods in dealing with LRL challenges and the necessity to adjust the measures of adaptation with the linguistic and computational requirements of the given district [86]. This discussion can be considered as informative regarding the activities to improve LLM performance in the use of underrepresented languages all over the world [7, 10]. Figure 13 compares two metrics for four major low-resource languages: Breadth (number of papers) and Depth (average text length in the "Items / Scale" field).

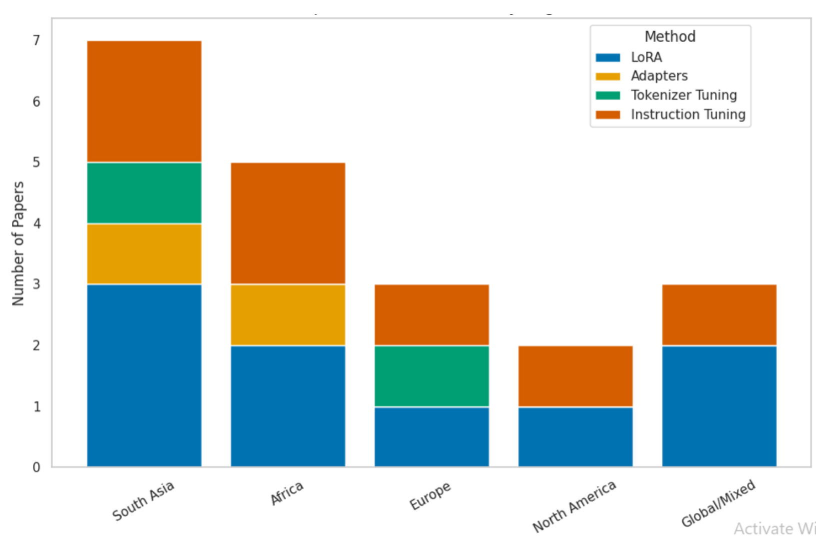

Figure 12: Adaptation Techniques by Region

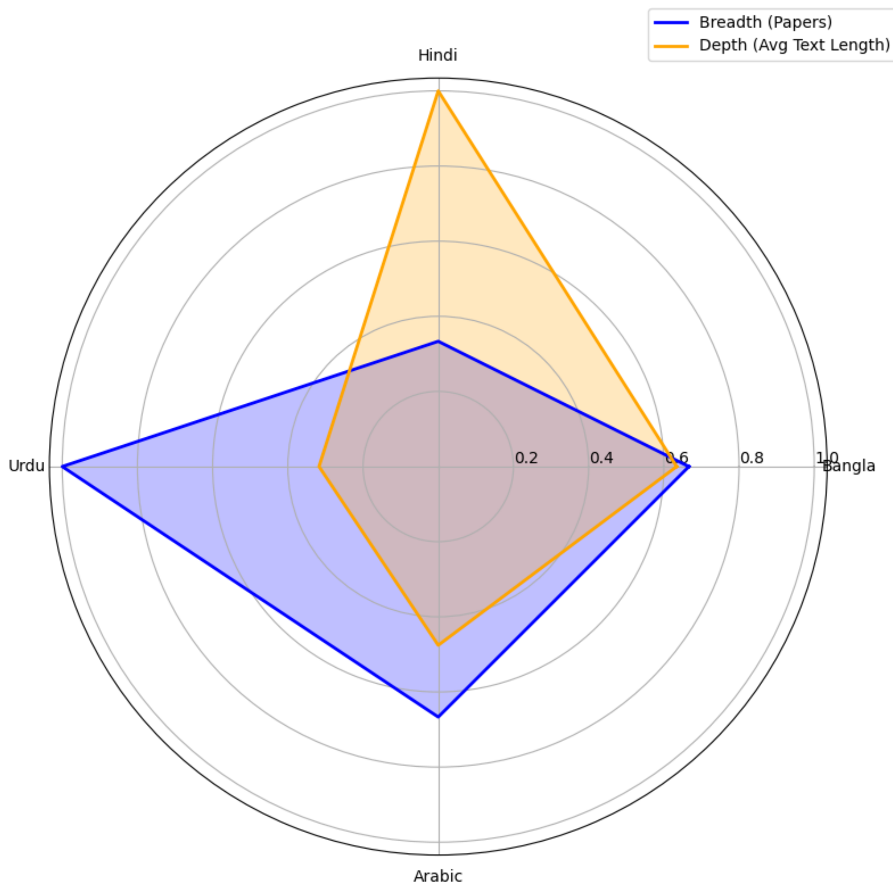

Figure 13: Breadth (study count) vs. depth (avg. dataset size) across language families

Figure 13 reveals that while Urdu is the most widely studied language (highest breadth), Hindi is associated with the most detailed or complex resources (highest depth). This highlights a potential gap: languages like Urdu may be researched more often, but the resources used might be less comprehensive compared to those for Hindi.

The variety of methods established in the current review highlights the challenge of the adaptation of LLMs to LRLs. Despite the latest advancement, like Mala-500 [79], Language Fusion via Adapters [13], there exist still problems with making multilingual adaptation scalable and maintain ethical fairness. In particular, Bhattacharjee et al. emphasise the necessity of transformer adaptation to a global language [77], whereas Chen et al. and Byambadorj et al. implement the concept of cross-lingual transfer theory on text-to-speech systems with low resources [84, 87] . Figure 14 shows the modalities studied in the literature (text, speech, multimodal, dialogue).

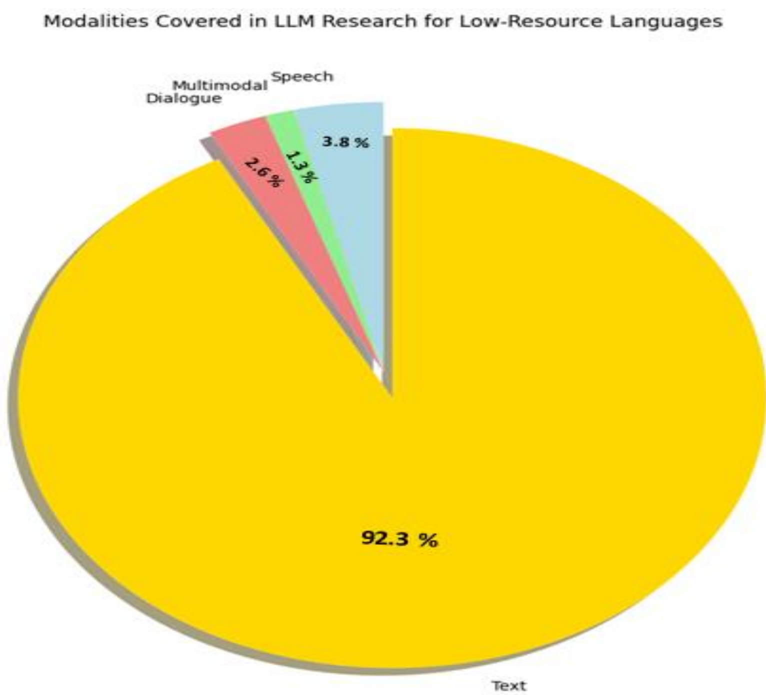

Figure 14: Modalities studied in the literature (text, speech, multimodal, dialogue)

There also exists regional imbalance in areas of research focus. Africa [2], South Asian [11] and Pakistani [80] studies underline the importance of contextual solutions, whereas the global projects such as Glot500 in this area [79] focus on bridging the gap between communities with different languages.

This review highlights theoretical and practical developments that surrounding the adaptation of LLMs to suit the application of LRLs lie in the value of benchmarks, efficient or cost-effective adaptation, and ethical concerns. Important advances entail the creation of culturally-based frameworks [1], creative fine-tuning strategies [7], as well as cross-lingual inference systems [88, 89]. In spite of them, there are still pending issues such as scalability, resource, and ethical risks that should be addressed in research in the future [36, 75]. With the filling of these gaps, the researchers will be capable of providing an even greater accessibility and influence of LLMs in linguistically underrepresented contexts.

4 Discussion

4.1 Data-Centric Adaptation and Pre-training Strategies

Recent studies lead to a unified point of view that data quality and the linguistic coverage were the beginning factors of the performance of a large-language-model in a low-resource condition. Low-resource language adaptation needed domain-specific and curated data pipelines that rely on in-house and synthetic data (e.g., small datasets that have been curated). In contrast to high-resource models, which rely on large and widely heterogeneous internet-scale corpora [1, 10, 40]. Continued pretraining [41, 47] and domain-adaptive data mixing [90, 91] had shown to be effective when it comes to maintaining linguistic coherence and improving the transfer learning of similar high-resource languages.

Syntactic generalization was also facilitated by typologically aligned multilingual pretraining (aligning Hindi and Urdu, or Swahili and Bantu variables), and catastrophic forgetting was also reduced further comparative analyses of multilingual and language-specific corpora are necessary [17, 92-94]. An example of an engineering direction of teaching LLM in new languages is Sambalingo, which targets teaching new languages by adding specific alignment and a set of successive data [95]. Already existing works on efficient adaptation techniques outline modular plans to adapt pre-trained models to new languages that have not been observed in comparable compute models on limited compute budgets [96].

Compensating corpus scarcity by means of synthetic corpus generators, with the assistance of the back-translation, paraphrasing, and lexicon driven data generation, constitutes a key compensatory aspect [48, 97]. However, synthetic augmentation alleviates the problem of data sparsity, but often causes distributional artifacts, which can affect downstream tasks, e.g. named-entity recognition or sentiment analysis [63; 98, 99]. Retrieval-augmented prompting improves translation performance by grounding generated outputs in retrieved bilingual exemplars [100, 101]. It has been shown that a combination of synthetic and monolingual authentic data should be used instead of the application of either type alone and provide better results [102, 103]. In addition, graph-based techniques of corpus enrichment also improve over random sampling since they yield better propagation of semantic consistency across multilingual embeddings [22, 104].

Despite these developments, the level of data imbalance is high. Seed-free synthetic data pipelines for instruction tuning demonstrated in Thai offer scalable augmentation without heavy annotation costs [105]. According to a number of studies, even such a large multilingual corpus as Glot500, SIB-200, or MasakhaNER is disproportionately represented by major African and Indo-European languages, and many regional languages are underrepresented [68, 106]. A summary of the most evident data strategies found in the reviewed literature, Table 3 indicates that cross-lingual knowledge transfer and synthetic corpus generation are increasingly used as the most generalizable methods of increasing the coverage of low-resource languages.

Table 3: Summary of Data-Centric Pretraining and Adaptation Strategies Across Reviewed Studies (2020–2025)

Approach / Strategy	Representative Studies	Data Type & Source	Key Contributions / Outcomes	Observed Limitations
Continued Pretraining on Multilingual Corpora	Adelani et al. [1], Li et al. [90], Gurgurov et al. [41]	Web-scale parallel and typologically aligned corpora	Improved cross-lingual transfer, reduced catastrophic forgetting	High compute demand; limited minority coverage
Synthetic Data Generation (Back-Translation, Paraphrasing)	Yong et al. [48], Zhu et al. [98], Subedi et al. [63]	Machine-generated sentences and translated pairs	Mitigates data sparsity; enhances NER and QA performance	Risk of noise and semantic drift

Cross-Lingual Knowledge Transfer	Wang et al. [92]; Cahyawijaya et al. [17]	HRL → LRL transfer using multilingual encoders	Boosts typological generalization	Transfer bias from dominant languages
Graph-Based Corpus Expansion	Lavrinovics et al. [104], Gurgurov et al. [22]	Semantic graph-linked documents	Maintains coherence across scripts	Complex preprocessing pipeline
Community-Curated LRL Corpora	Adelani et al. [68], Skadiņa et al. [106]	Locally sourced parallel texts	Enhances representational equity	Small-scale; uneven domain coverage

4.2 Parameter-Efficient Fine-Tuning (PEFT) and Adaptation Techniques

One of the biggest advances to Parameter-efficient fine-tuning (PEFT) is an attempt to generalize huge models to diminutive resources through limited computational limits. Low-rank adaptation (LoRA), adapters, prefix tuning, and instruction tuning represent some of the methods that allow the practical use of specialization and do not require retraining the entire parameter set [13, 107, 108]. Such LoRA-based improvements, including those, have shown significant performance improvement in multilingual tasks with significantly lower GPU usage, by a factor of up to 80% with almost baseline performance [35, 28].

However, comparative data show that the efficiency of any PEFT methodology relies on the language typology and the size of the corpus to a great extent. Language adapters provide modular intervention points that permit fine-grained cross-lingual control and efficient reuse [109, 110]. Additionally, Lightweight adaptation schemes are shown to be effective in typologically divergent languages and empirical studies illustrating LLaMA adaptation to Persian support the used method accordingly [111]. Morphologically complex languages, such as Basque, Tamil, and Turkish, are researched and demonstrate that stable propagation of gradient is reached by adapter layers in contrast to prefix tuning because of morphological modularity [49, 112]. Turkish-focused adaptation work exemplifies practical strategies for benchmarking and low-resource model customization [113]. On the other hand, instruction tuning provides the best zero-shot generalization in low resource languages that are not seen [34, 114, 115]. Even in these advantages, most authors highlight that there are no standardized measures to determine PEFT efficiency when operating in low-resource languages [38, 68].

Other innovations, such as model merging [116] and the active forgetting mechanisms [117] where cross-lingual parameters selectively are kept in the active removing stage of fine-tuning to avoid over impression, are not excluded. Graph-based adaptation [41] and cross-lingual adapter fusion [57] are the further extensions of these paradigms, which form a multilingual match based on intermediate representations. Empirical results also show that LoRA still can be applied to single-language adaptation, and higher in multi-language scenarios implanting adaptors and instruction tuning outperform [118, 107].

Table 4: Performance Trade-Offs Across Parameter-Efficient Fine-Tuning (PEFT) Techniques

Technique	Key References	Parameter Efficiency	Performance Trend	Strengths	Limitations
LoRA (Low-Rank Adaptation)	Hu et al. [13], Miyano and Arase [35]	~0.5–2% additional params	95–98% of full fine-tuning accuracy	Minimal compute cost; fast deployment	May underperform on highly inflected LRLs
Adapters / Adapter Fusion	Etxaniz et al. [49], Schlenker et al. [57]	3–5% additional params	Stable across morphologically rich languages	Modular and reusable	Larger memory footprint than LoRA
Prefix / Prompt Tuning	Singh et al. [108], Toukmaji and Flanigan, [34]	0.1–1%	Strong zero-shot transfer	Minimal training time	Sensitive to initialization
Instruction Tuning	Kuulmets et al. [115]; Zhao et al. [107]	Variable	Highest generalization across unseen LRLs	Task flexibility; cultural adaptability	Requires curated instruction data
Model Merging / Active Forgetting	Aggarwal et al. [117], Tao et al. [116]	—	Preserves multilingual alignment	Prevents overfitting	Implementation complexity

4.3 Tokenization, Script Diversity, and Typological Representation

A major portion of the research made in the area of adapting large-language models to actually sensible languages is dedicated to tokenization, which is still a bottleneck in working with morphologically or non-Latin written languages. The vast majority of tokenizers (based on BPE or SentencePiece architectures) break words in aggluistic and tonally speaking languages, which destroys the semantic integrity [4, 55]. Recent papers suggest script-sensitive tokenizers and adaptive sub-word vocabularies that are willing to maintain linguistic integrity in the Bantu, Dravidian, and Turkic families [21, 119]. Encoders that are based on transliteration at the byte-level and reversibility [30, 120] are more effective in the determinism of typological subtleties than tokenizers.

The current evaluation shows that in-context learning may be used to provide teaching in the languages that are highly under-resourced and require the little parameter fine-tuning [121]. Moreover, pivot-based feature conversion procedures offer quite promising zero-resource translation routes and convert representations by converting them into an intermediate rich-resource language [122]. Few-shot in-context learning can be enhanced through alignment, and it leads to better cross-lingual generalization of multilingual large language models [123]. Bringing together code-switching curriculum with embedding-transfer strategy also alleviates domain-specific adaptation of multilingual large language model [31].

Comparative research also proves that the embedding realignment of multilinguals i.e. through language adapters or through the span of transliterators improves semantic overlap of distant scripts [25, 50]. However, there is also an opinion that tokenization has to be co-trained with pretraining goals, since mistakes on a subword scale may only be magnified during fine-tuning [56, 108]. Table 5 provides a summary of tokenization designs that have been put forward in significant literature, which points to the relevance of morphology-aware and phoneme-receptive modeling to script-diverse languages.

Table 5: Comparative analysis of tokenization and subword modeling strategies for morphologically rich languages.

Tokenization Framework	Core Studies	Supported Scripts	Linguistic Benefit	Drawbacks
BPE (Baseline)	Darģis et al. [4]	Latin, Cyrillic	Widely adopted, efficient	Breaks agglutinative morphemes
SentencePiece / UnigramLM	Remy et al. [55], Feng et al. [50]	Multiscript	Balanced vocabulary control	Weak for tonal systems
Byte-Level Encoding	Lyu et al. [120]	Any UTF-8	Script-agnostic; reversible	Longer sequences; compute cost
Script-Sensitive Tokenizers	Yamada & Ri [21], Zhuang et al. [30]	Bantu, Dravidian, Indic	Maintains morphological integrity	Requires linguistic annotation
Transliteration-Based Models	Wu et al. [25], Neplenbroek et al. [56]	Non-Latin → Latin	Enables cross-script sharing	Possible phoneme loss

4.4 Evaluation, Reproducibility, and Benchmark Disparities

Another issue that has remained constant in the history of low-resource language-oriented large-language-model development is the lack of standardized evaluation and reproducibility procedures. Most of the works use not cross-task-com-comparable metrics such as BLEU, F1, perplexity, or accuracy [124, 125]. The discrepancies in the benchmark are specific in the case of African and indigenous languages, where the quality of the annotation differs significantly as well as the granularity of a corpus [68, 58]. Included in an encouraging direction towards multilingual reproducibility is the introduction of evaluation suites, a subset of which includes BUFFET [126], MEXA [127], and LEIA [21], which, however, are less comprehensive than is desired. Methods that enable on-the-fly instruction of unseen languages demonstrate rapid bootstrapping abilities for LLMs in low-resource settings [128].

Research highlights that cross-lingual alignment because XSim or CrossEval scores have a better comparability than measures based on monolingual metrics of tasks especially in low-resource languages that are typologically in proximity to each other [27, 129]. The Irish-based adaptation effort underlines viability of LLMs in ultra-scarce data environments [130]. Extremely low-resource Finno-Ugric adaptations validate generalization limits [131]. However, the unavailability of open repositories and missing documentation still hamper reproducibility [1, 7]. Others suggest culturally-based testbeds like BasqBBQ and NusaDialogue as unbiased and fair evaluators [58, 61].

Table 6: Comparative summary of evaluation metrics and benchmark availability for LRL studies.

Benchmark / Dataset	Key Sources	Metrics Used	Coverage & Languages	Reproducibility Issues
FLORES-200 / Taxi1500	Adelani et al. [1], Nag et al. [47]	BLEU, ChrF	200 languages / 1500 pairs	Limited low-resource test sets
MasakhaNER / BLEnD	Myung et al. [7], Adelani et al. [68]	F1, Accuracy	17 African languages	Annotation inconsistency
LEIA / BUFFET / MEXA	Asai et al. [126], Kargaran et al. [127]	Factuality, Faithfulness	50+ multilingual models	Partial openness
BasqBBQ / NusaDialogue	Saralegi and Zulaika [58], Purwarianti et al. [61]	Bias, Fairness Scores	Basque, Indonesian	Limited benchmarking tools
CrossEval / XSim	Lee et al. [27], Wen-Yi and Mimno, [129]	Alignment Score	Multilingual Typologies	Absent standardized pipeline

4.5 Cultural Fairness, Accessibility, and Ethical Adaptation

Along with technical optimization, it is increasingly being understood that culture and ethical reflexivity has to be interwoven into large-language-model adaptation to low-resource languages. Many researches warn that English-based corpora pose the risk of continuing linguistic hegemony and representational bias especially in sociocultural sensitive contexts [74, 2]. The ethnographically informed frameworks [36, 83] promote the data governance based on community, and the evaluation of participatory frameworks. As an example, the local cooperation, as in NileChat [67] and ELEVATE-ID [132], can result in the creation of linguistically and culturally believable artificial-intelligence applications.

Another area that overlaps with cultural adaptation is infrastructural accessibility. Recent empirical results (African, South Asian, Southeast Asian) have described that low-compute LoRA-based adaptation becomes democratic using instruction tuning to expose low-funded institutions to model deployment [63, 68, 133]. Additionally, investigations of value alignment show that moral and cultural dimensions do not transfer uniformly across languages, underscoring the need for locale-aware evaluation [134]. Moral issues of prejudice, misinformation spreading, and hallucination [104, 66] are still acute, and frameworks of multilingual hallucination-detection have been demanded [65]. Multilingual multi-channel models have demonstrated improved hate-speech detection across languages, illustrating important social benefits [135-138]. Table 7 provides a cross-comparison of the models of ethical adaption and the rising emphasis of both fairness audits and representational balance in low-resource-language projects today.

Table 7: Overview of ethical and fairness-focused adaptation frameworks across multilingual LLMs

Framework / Initiative	Main References	Ethical Dimension	Implementation Mechanism	Outcome / Relevance
IrokoBench / Atlas-Chat	Kalluri [36], Zhong et al. [75]	Cultural representation & bias audit	Community-sourced evaluation	Improves inclusivity in African contexts
NileChat / ELEVATE-ID	Mekki et al. [67], Gusmita et al. [132]	Participatory data governance	Local co-creation of datasets	Enhances authenticity & trust
AI Ethics Guidelines	Ngugi [83]	Global policy framework	Standardized ethical compliance	Encourages equitable AI adoption
BLEnD Fairness Protocol	Adelani et al. [68]	Bias detection & fairness metrics	Embedded during fine-tuning	Promotes balanced representation
Low-Compute Democratic Adaptation	Mahfuz et al. [133], Subedi et al. [63]	Accessibility & inclusion	LoRA + instruction tuning in low-resource setups	Democratizes LLM deployment

4.6 Findings and Theoretical Implications Synthesis

Integrating the findings in all the domains considered, there are three major patterns. To begin with, the data-driven augmentation is here to stay: systematic empirical experiments always indicate that synthetic and cross-lingual pipelines synthesize outperform the downstream task performance in cases where typologically aligned corpora are used [139, 140]. Second, equitable adaptation has now become popularized as parameter-efficient fine-tuning because it can be deployed at scalable cost without being prohibitively expensive [13, 35]. Third, the discipline is becoming decolonial and community-focused, where linguistic justice and digital inclusion are termed as a part of technical design [2, 74].

In addition, these findings confirm that (RQ1) pretraining and adaptation for LRLs are most successful when combining multilingual alignment with corpus curation; (RQ2) PEFT methods balance efficiency and performance under limited compute environments; and (RQ3) equitable LLM adaptation necessitates culturally sensitive and reproducible evaluation frameworks. Together, these insights establish a multi-level roadmap for advancing typology-aware, ethically grounded, and resource-efficient LLM adaptation strategies in the coming decade.

5 Conclusion

The present systematic review represents an integrated overview of the current developments in the adaptation of Large Language Models (LLM) to Low-Resource Languages (LRLs) that are dedicated to data-centric pretraining, parameter-efficient fine-tuning, tokenization and ethical issues. Results of augmentation in 2020-2025 indicate data augmentation, multilingual transfer, and typology trained pretraining are greatly beneficial in covering language, whereas parameter-efficient techniques include LoRA, adapters, and instruction tuning are especially able to produce significant performance improvements with limited computational budgets. The current drawbacks of tokenization, differences in script models, and inappropriate evaluation metrics emphasize caution on having a standardized multilingual benchmark and pipeline research probability. Moreover, the frameworks of ethical and cultural fairness supported by the participatory data governance are essential to the establishment of inclusiveness and linguistic justice in AI-development. All these findings together substantiate the fact that equitable implementation of LLM implies that there should be dual commitment to technical scalability and sociocultural contextualization. Future studies need to pursue hybrid pipelines of adaptation which combine synthetic corpus generation, cross-lingual continual learning and parameter-efficient fine-tuning to the wider scope of typologically different LRLs. Increased focus on multilingual benchmarks based on open-source and addressed by the local community in the curation of its diverse data will encourage transparency and inclusiveness. Besides, within the scope of interdisciplinary cooperation, computational linguists, ethicists, and local stakeholders should collaborate to guarantee cultural fairness and sustainable AI ecosystems. Possibly the most crucial step towards expanding the global large language model development gap and ensuring really equitable outcomes in large language model development will be the advancement of reproducible evaluation schemes and democratized access to multilingual resources.

References

1. Adelani DI, Ojo J, Azime IA, Zhuang JY, Alabi JO, He X, et al. IrokoBench: A new benchmark for African languages in the age of large language models. In: Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies. 2025a;1:2732–57. doi:10.18653/v1/2025.naacl-long.139
2. Alhanai T, Kasumovic A, Ghassemi M, Zitzelberger A, Lundin J, Chabot Couture G. Bridging the gap: Enhancing LLM performance for low-resource African languages with new benchmarks, fine tuning, and cultural adjustments. arXiv preprint. 2024;2412.12417. doi:10.48550/arXiv.2412.12417
3. Cassano F, Gouwar J, Lucchetti F, Schlesinger C, Freeman A, Anderson CJ, Guha A. Knowledge transfer from high-resource to low-resource programming languages for code LLMs. Proc ACM Program Lang. 2024;8(OOPSLA2):677–708. Available from: https://doi.org/10.1145/3689735
4. Darģis R, Bārzdiņš G, Skadiņa I, Grūzītis N, Saulīte B. Evaluating open-source LLMs in low-resource languages: Insights from Latvian high school exams. In: Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities. 2024;289–93. doi:10.18653/v1/2024.nlp4dh-1.28
5. Zhang C, Lin J, Liu X, Zhang Z, Feng Y. Read it in two steps: Translating extremely low-resource languages with code-augmented grammar books. Proc 63rd Annual Meeting of the Association for Computational Linguistics. 2025a;1:3977–97.

6. Park G, Hwang S, Lee H. Low-resource cross-lingual summarization through few-shot learning with large language models. Proc 7th Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2024);57–63. https://doi.org/10.18653/v1/2024.loresmt-1.6
7. Myung J, Lee N, Zhou Y, Jin J, Putri RA, Antypas D, et al. BLEnD: A benchmark for LLMs on everyday knowledge in diverse cultures and languages. arXiv preprint. 2024;2406.09948. doi:10.48550/arXiv.2406.09948
8. Pava J, Meinhardt C, Uz Zaman HB, Friedman T, Truong ST, Zhang D, et al. Mind the (language) gap: Mapping the challenges of LLM development in low-resource language contexts. 2025. Available from: https://hai.stanford.edu/policy/mind-the-language-gap-mapping-the-challenges-of-llm-development-in-low-resource-language-contexts
9. Nguyen XP, Aljunied SM, Joty S, Bing L. Democratizing LLMs for low-resource languages by leveraging their English-dominant abilities with linguistically-diverse prompts. Proc Annual Meeting of the Association for Computational Linguistics. 2023. Available from: https://doi.org/10.48550/arXiv.2306.11372
10. Le TD, Nguyen TT, Nguyen Ha V, Chatzinotas S, Jouvet P, Noumeir R. The impact of LoRA adapters for LLMs on clinical NLP classification under data limitations. arXiv preprint. 2025;2407.19299v2. doi:10.48550/arXiv.2407.19299
11. Khade O, Jagdale S, Phaltankar A, Takalikar G, Joshi R. Challenges in adapting multilingual LLMs to low-resource languages using LoRA PEFT tuning. In: Proceedings of the First Workshop on Challenges in Processing South Asian Languages (CHiPSAL). 2024;217–22. Available from: https://aclanthology.org/2025.chipsal-1.22
12. Moradi M, Yan K, Colwell D, Samwald M, Asgari R. A critical review of methods and challenges in large language models. Comput Mater Continua 2025;82(2). doi:10.32604/cmc.2025.061263
13. Hu Q, Zhang Y, Zhang X, Han Z, Liang X. Language fusion via adapters for low-resource speech recognition. Speech Commun. 2024;158:103037. doi:10.1016/j.specom.2024.103037
14. Shang G, Abdine H, Khoubrane Y, Mohamed A, Abbahaddou Y, Ennadir S, et al. Atlas-Chat: Adapting large language models for low-resource Moroccan Arabic dialect. arXiv preprint. 2024;2409.17912. doi:10.48550/arXiv.2409.17912
15. Tessema BM, Kedia A, University TC, Researcher I. UnifiedCrawl: Aggregated Common Crawl for affordable adaptation of LLMs on low-resource languages. arXiv 2024. Available from: https://arxiv.org/abs/2411.14343
16. Shu P, Chen J, Liu Z, Wang H, Wu Z, Zhong T, Li Y, Zhao H, Jiang H, Pan Y, Zhou Y, Owl C, Zhai X, Liu N, Saunt C, Liu T. Transcending language boundaries: Harnessing LLMs for low-resource language translation. arXiv 2024. Available from: https://arxiv.org/abs/2411.11295
17. Cahyawijaya S, Lovenia H, Fung P. LLMs are few-shot in-context low-resource language learners. In: Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2024;1:405–33. doi:10.18653/v1/2024.naacl-long.24
18. Sundar A, Williamson S, Metcalf K, Theobald BJ, Seto S, Fedzechkina M. Steering into new embedding spaces: analyzing cross-lingual alignment induced by model interventions in multilingual language models. Proc 63rd Annu Meet Assoc Comput Linguist (ACL 2025). 2025;1:2375–400. doi:10.18653/v1/2025.acl-long.118.

19. Alam F, Chowdhury SA, Boughorbel S, Hasanain M. LLMs for low-resource languages in multilingual, multimodal and dialectal settings. Proc 18th Conf European Chapter of the Association for Computational Linguistics: Tutorial Abstracts;27–33. 2024. Available from: https://doi.org/10.18653/v1/2024.eacl-tutorials.5
20. Mao Z, Yu Y. Tuning LLMs with contrastive alignment instructions for machine translation in unseen, low-resource languages. Proc 7th Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2024);1–25. https://doi.org/10.18653/v1/2024.loresmt-1.1
21. Yamada I, Ri R. LEIA: Facilitating cross-lingual knowledge transfer in language models with entity-based data augmentation. Findings Assoc Comput Linguist: ACL 2024;7029–7039. doi:10.18653/v1/2024.findings-acl.419
22. Gurgurov D, Hartmann M, Ostermann S. Adapting multilingual LLMs to low-resource languages with knowledge graphs via adapters. In: Proceedings of the 1st Workshop on Knowledge Graphs and Large Language Models (KaLLM 2024). 2024;63–74. doi:10.18653/v1/2024.kallm-1.7
23. Gain B, Bandyopadhyay D, Ekbal A. Bridging the linguistic divide: A survey on leveraging large language models for machine translation. arXiv [Preprint]. 2025. Available from: https://arxiv.org/abs/2504.01919
24. Corral A, Sarasua I, Saralegi X. Pipeline analysis for developing instruct LLMs in low-resource languages: A case study on Basque. Proc 2025 Conf North American Chapter of the Association for Computational Linguistics: Human Language Technologies;1:12636–55. https://doi.org/10.18653/v1/2025.naacl-long.629
25. Wu L, Wei HR, Yang B, Lu W. From English to second-language mastery: Enhancing LLMs with cross-lingual continued instruction tuning. Proc ACL 2025;1:23006–23023.
26. Pan S, Tian Z, Ding L, Huang Z, Wen Z, Li D. POMP: Probability-driven meta-graph prompter for LLMs in low-resource unsupervised neural machine translation. Proc 62nd Annual Meeting of the Association for Computational Linguistics 2024;1:9976–92. https://doi.org/10.18653/v1/2024.acl-long.537
27. Lee S, Hong S, Moon H, Lim H. Semantic aware linear transfer by recycling pre-trained language models for cross-lingual transfer. Findings ACL 2025;16180–16193. doi:10.18653/v1/2025.findings-acl.832
28. Khoboko PW, Marivate V, Sefara J. Optimizing translation for low-resource languages: Efficient fine-tuning with custom prompt engineering in large language models. Mach Learn Appl 2025;20:100649. doi:10.1016/j.mlwa.2025.100649
29. Chirkova N, Nikoulina V. Zero-shot cross-lingual transfer in instruction tuning of large language models. Proc 17th Int Nat Lang Gener Conf (INLG 2024). 2024;695–708. doi:10.18653/v1/2024.inlg-main.53.
30. Zhuang W, Sun Y, Zhao X. Enhancing cross-lingual transfer through reversible transliteration: a Huffman-based approach for low-resource languages. Proc 63rd Annu Meet Assoc Comput Linguist (ACL 2025). 2025;1:16299–313. doi:10.18653/v1/2025.acl-long.795.
31. Hong S, Lee S, Moon H, Lim HS. MIGRATE: cross-lingual adaptation of domain-specific LLMs through code-switching and embedding transfer. Proc 31st Int Conf Comput Linguist (COLING 2025). 2025 Jan;9184–93.
32. Adeyemi M, Oladipo A, Pradeep R, Lin J. Zero-shot cross-lingual reranking with large language models for low-resource languages. Proc 62nd Annual Meeting of the Association for Computational Linguistics 2023;2:650–6. doi:10.18653/v1/2024.acl-short.59
33. Avetisyan H, Broneske D. VerbCraft: Morphologically-aware Armenian text generation using LLMs in low-resource settings. Proc 3rd Workshop on Resources and Representations for Under-Resourced Languages

and Domains (RESOURCEFUL-2025);111–9. Available from: https://aclanthology.org/2025.resourceful-1.25.pdf

34. Toukmaji C, Flanigan J. Prompt, translate, fine-tune, re-initialize, or instruction-tune? Adapting LLMs for in-context learning in low-resource languages. arXiv. 2025. doi:10.48550/arXiv.2506.19187
35. Miyano R, Arase Y. Adaptive LoRA merge with parameter pruning for low-resource generation. Findings Assoc Comput Linguist: ACL 2025;19353–66. doi:10.18653/v1/2025.findings-acl.990
36. Kalluri K. Adapting LLMs for low-resource languages: Techniques and ethical considerations. Int Sci J Sci Res Eng Manag. 2023;8(12):1–6. doi:10.55041/ISJEM00140
37. Shi H, Xu Z, Wang H, Qin W, Wang W, Wang Y, et al. Continual learning of large language models: A comprehensive survey. ACM Comput Surv. 2025. doi:10.1145/3735633
38. Qin L, Chen Q, Zhou Y, Chen Z, Li Y, Liao L, et al. A survey of multilingual large language models. Patterns. 2025;6(1):101118. doi:10.1016/j.patter.2024.101118
39. Lankford S, Afli H, Way A. adaptMLLM: Fine-tuning multilingual language models on low-resource languages with integrated LLM playgrounds. Information. 2023;14(12):638. doi:10.3390/info14120638
40. Joshi R, Singla K, Kamath A, Kalani R, Paul R, Vaidya U, et al. Adapting multilingual LLMs to low-resource languages using continued pre-training and synthetic corpus: A case study for Hindi LLMs. In: Proceedings of the First Workshop on Natural Language Processing for Indo-Aryan and Dravidian Languages. 2024;50–7. Available from: https://aclanthology.org/2025.indonlp-1.6.pdf
41. Gurgurov D, Vykopal I, van Genabith J, Ostermann S. Small models, big impact: Efficient corpus and graph-based adaptation of small multilingual language models for low-resource languages. In: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics. 2025;4:355–95. doi:10.18653/v1/2025.acl-srw.24
42. Akimoto K, Oyamada M. Optimizing low-resource language model training: Comprehensive analysis of multi-epoch, multi-lingual, and two-stage approaches. arXiv 2024. Available from: https://doi.org/10.48550/arXiv.2410.12325
43. Rathore V, Deb A, Chandresh A, Singla P. SSP: Self-supervised prompting for cross-lingual transfer to low-resource languages using large language models. Findings Assoc Comput Linguist: EMNLP 2024;15081–102. https://doi.org/10.18653/v1/2024.findings-emnlp.886
44. Mondal SK, Sen S, Singhania A, Jyothi P. Language-specific neurons do not facilitate cross-lingual transfer. Proc 6th Workshop Insights Negative Results NLP. 2025;46–62.
45. Yoo H, Park C, Yun S, Oh A, Lee H. Code-switching curriculum learning for multilingual transfer in LLMs. Findings ACL 2025;7816–7836. doi:10.18653/v1/2025.findings-acl.407
46. Zhuang W, Sun Y. CUTE: A multilingual dataset for enhancing cross-lingual knowledge transfer in low-resource languages. Proc 31st Int Conf Comput Linguist (COLING 2025). 2025;10037–46.
47. Nag A, Chakrabarti S, Mukherjee A, Ganguly N. Efficient continual pre-training of LLMs for low-resource languages. In: Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies. 2025;3:304–17. doi:10.18653/v1/2025.naacl-industry.25

48. Yong ZX, Menghini C, Bach SH. LexC-Gen: Generating data for extremely low-resource languages with large language models and bilingual lexicons. Findings Assoc Comput Linguist: EMNLP 2024. 2024;13990–1009. doi:10.18653/v1/2024.findings-emnlp.818.
49. Etxaniz J, Sainz O, Perez N, Aldabe I, Rigau G, Agirre E, et al. Latxa: An open language model and evaluation suite for Basque. In: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics. 2024;1:14952–72. doi:10.18653/v1/2024.acl-long.799
50. Feng Z, Cao H, Xu W, Zhao T. Word-level cross-lingual structure in large language models. Proc 31st Int Conf Comput Linguist (COLING 2025). 2025;2026–37. Available from: https://aclanthology.org/2025.coling-main.138/
51. Šliogeris V, Daniušis P, Nakvosas A. Full-parameter continual pretraining of Gemma2: Insights into fluency and domain knowledge. arXiv [Preprint]. 2025. doi: 10.48550/arXiv.2505.05946
52. Cahyawijaya S, Lovenia H, Yu T, Chung W, Fung P. InstructAlign: High-and-low resource language alignment via continual cross-lingual instruction tuning. In: Proceedings of the First Workshop in South East Asian Language Processing. 2023;55–78. doi:10.18653/v1/2023.sealp-1.5
53. Duwal S, Prasai S, Manandhar S. Domain-adaptive continual learning for low-resource tasks: Evaluation on Nepali. Proc COLING Workshops. 2024. doi:10.48550/arXiv.2412.13860
54. Zhukova A, Matt CE, Gipp B. Efficient domain-adaptive continual pretraining for the process industry in the German language. In: *International Conference on Text, Speech, and Dialogue*. Cham: Springer Nature Switzerland; 2025; 77–89. doi: 10.1007/978-3-032-02551-7_8
55. Remy F, Delobelle P, Avetisyan H, Khabibullina AM, de Lhoneux M, Demeester T. Trans-tokenization and cross-lingual vocabulary transfers: Language adaptation of LLMs for low-resource NLP. arXiv 2024. doi:10.48550/arXiv.2408.04303
56. Neplenbroek V, Bisazza A, Fernández R. Cross-lingual transfer of debiasing and detoxification in multilingual LLMs: an extensive investigation. Findings Assoc Comput Linguist: ACL 2025. 2024;2805–30. doi:10.18653/v1/2025.findings-acl.145.
57. Schlenker J, Kunz J, Anikina T, Neumann G, Ostermann S. Only for the unseen languages, say the llamas: on the efficacy of language adapters for cross-lingual transfer in English-centric LLMs. Proc 63rd Annu Meet Assoc Comput Linguist (ACL 2025). 2025 Jul;4:849–71. doi:10.18653/v1/2025.acl-srw.62.
58. Saralegi X, Zulaika M. BasqBBQ: A QA benchmark for assessing social biases in LLMs for Basque, a low-resource language. Proc 31st Int Conf Computational Linguistics;4753–67. Available from: https://aclanthology.org/2025.coling-main.318.pdf
59. Guo P, Ren Y, Hu Y, Li Y, Zhang J, Zhang X, Huang HY. Teaching large language models to translate on low-resource languages with textbook prompting. *P*roc Joint Int Conf Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024);15685–97. Available from: https://aclanthology.org/2024.lrec-main.1362.pdf
60. Jiao W, Huang JT, Wang W, He Z, Liang T, Wang X, Tu Z. ParroT: Translating during chat using large language models tuned with human translation and feedback. Findings Assoc Comput Linguist: EMNLP 2023;15009–20. doi:10.18653/v1/2023.findings-emnlp.1001

61. Purwarianti A, Adhista D, Baptiso A, Mahfuzh M, Sabila Y, Adila A, et al. NusaDialogue: Dialogue summarization and generation for underrepresented and extremely low-resource languages. Proc 2nd Workshop in South East Asian Language Processing 2025;82–100.
62. Zhang C, Liao Z, Feng Y. Cross-lingual transfer of cultural knowledge: An asymmetric phenomenon. Proc 63rd Annual Meeting of the Association for Computational Linguistics. 2025b;2:147–57.
63. Subedi B, Regmi S, Bal BK, Acharya P. Exploring the potential of large language models (LLMs) for low-resource languages: A study on named-entity recognition (NER) and part-of-speech (POS) tagging for Nepali language. Proc Joint Int Conf Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024);6974–9. Available from: https://aclanthology.org/2024.lrec-main.611.pdf
64. Shibu HM, Datta S, Miah MS, Sami N, Chowdhury MS, Islam MS. From scarcity to capability: Empowering fake news detection in low-resource languages with LLMs. Proc 1st Workshop on Natural Language Processing for Indo-Aryan and Dravidian Languages 2025;100–7. Available from: https://aclanthology.org/2025.indonlp-1.12/
65. Benkirane K, Gongas L, Pelles S, Fuchs N, Darmon J, Stenetorp P, et al. Machine translation hallucination detection for low and high resource languages using large language models. Findings Assoc Comput Linguist: EMNLP 2024;9647–65. doi:10.18653/v1/2024.findings-emnlp.564
66. Shen L, Tan W, Chen S, Chen Y, Zhang J, Xu H, et al. The language barrier: Dissecting safety challenges of LLMs in multilingual contexts. Findings Assoc Comput Linguist: ACL 2024;2668–80. doi:10.18653/v1/2024.findings-acl.156
67. Mekki AE, Atou H, Nacar O, Shehata S, Abdul-Mageed M. NileChat: Towards linguistically diverse and culturally aware LLMs for local communities. arXiv 2025. doi:10.48550/arXiv.2505.18383
68. Adelani DI, Doğruöz AS, Coneglian A, Ojha AK. Comparing LLM prompting with cross-lingual transfer performance on Indigenous and low-resource Brazilian languages. Proc 4th Workshop Nat Lang Process Indigenous Lang Americas (AmericasNLP 2024). 2024b;34–41. doi:10.18653/v1/2024.americasnlp-1.5.
69. Upadhayay B, Behzadan V. Taco: Enhancing cross-lingual transfer for low-resource languages in LLMs through translation-assisted chain-of-thought processes. arXiv preprint. 2023;2311.10797. doi:10.48550/arXiv.2311.10797
70. Kazi S, Rahim M, Khoja SA. Crossing language boundaries: evaluation of large language models on Urdu–English question answering. Proc 1st Workshop Nat Lang Process Indo-Aryan Dravidian Langs. 2025 Jan;141–51.
71. Scialom T, Chakrabarty T, Muresan S. Fine-tuned language models are continual learners. Proc Empirical Methods in Natural Language Processing (EMNLP 2022);2528–15378. doi:10.18653/v1/2022.emnlp-main.410
72. Liu D, Niehues J. Middle-layer representation alignment for cross-lingual transfer in fine-tuned LLMs. Proc 63rd Annu Meet Assoc Comput Linguist (ACL 2025). 2025;1:15979–96. doi:10.18653/v1/2025.acl-long.778.
73. Xie J, Zhang Z, Zeng S, Hilliard J, An G, Tang X, et al. Leveraging large language models for infectious disease surveillance: Using a web service for monitoring COVID-19 patterns from self-reporting tweets. J Med Internet Res 2025;27(1):e63190. doi:10.2196/63190
74. Aksoy M. Whose morality do they speak? Unraveling cultural bias in multilingual language models. Nat Lang Process J 2025;100172. doi:10.1016/j.nlp.2025.100172

75. Zhong T, Yang Z, Liu Z, Zhang R, Liu Y, Sun H, et al. Opportunities and challenges of large language models for low-resource languages in humanities research. arXiv preprint. 2024;2412.04497. doi:10.48550/arXiv.2412.04497
76. Verma S, Rahman Khan MSU, Kumar V, Murthy R, Sen J. MILU: A multi-task Indic language understanding benchmark. arXiv preprint. 2024;2411.02538. doi:10.48550/arXiv.2411.02538
77. Bhattacharjee A, Hasan T, Ahmad WU, Samin K, Islam MS, Iqbal A, et al. BanglaBERT: Language model pretraining and benchmarks for low-resource language understanding evaluation in Bangla. arXiv preprint. 2022;2101.00204. doi:10.18653/v1/2022.findings-naacl.98
78. Imani A, Lin P, Kargaran AH, Severini S, Sabet MJ, Kassner N, et al. Glot500: Scaling multilingual corpora and language models to 500 languages. arXiv preprint. 2023;2305.12182. doi:10.48550/arXiv.2305.12182
79. Lin P, Ji S, Tiedemann J, Martins AF, Schütze H. Mala-500: Massive language adaptation of large language models. arXiv preprint. 2024;2401.13303. doi:10.48550/arXiv.2401.13303
80. Fiaz L, Tahir MH, Shams S, Hussain S. UrduLLaMA 1.0: Dataset curation, preprocessing, and evaluation in low-resource settings. arXiv preprint. 2025;2502.16961. doi:10.48550/arXiv.2502.16961
81. Kadyrbek N, Tuimebayev Z, Mansurova M, Viegas V. The development of small-scale language models for low-resource languages, with a focus on Kazakh and direct preference optimization. Big Data Cogn Comput. 2025;9(5):137. doi:10.3390/bdcc9050137
82. Tahir MH, Shams S, Fiaz L, Adeeba F, Hussain S. Benchmarking the performance of pre-trained LLMs across Urdu NLP tasks. In: Proceedings of the First Workshop on Challenges in Processing South Asian Languages (CHiPSAL). 2024;17–34. doi:10.48550/arXiv.2405.15453
83. Ngugi S. Targeted lexical injection: Unlocking latent cross-lingual alignment in Lugha-Llama via early-layer LoRA fine-tuning. arXiv preprint. 2025;2506.15415. doi:10.48550/arXiv.2506.15415
84. Byambadorj Z, Nishimura R, Ayush A, et al. Text-to-speech system for low-resource language using cross-lingual transfer learning and data augmentation. J Audio Speech Music Proc. 2021;42. doi:10.1186/s13636-021-00225-4
85. Jadhav S, Shanbhag A, Thakurdesai A, Sinare R, Joshi R. On limitations of LLM as annotator for low-resource languages. Proc 8th Int Conf Nat Lang Speech Process (ICNLSP 2025). 2024;277–82. Available from: https://aclanthology.org/2025.icnlsp-1.27.pdf
86. Kew T, Schottmann F, Sennrich R. Turning English-centric LLMs into polyglots: How much multilinguality is needed? Findings EMNLP 2024;13097–13124.
87. Chen L, Guan S, Huang X, Wang WJ, Xu C, Guan Z, et al. Cross-lingual multimodal sentiment analysis for low-resource languages via language family disentanglement and rethinking transfer. Findings Assoc Comput Linguist: ACL 2025. 2025 Jul;6513–22. doi:10.18653/v1/2025.findings-acl.338.
88. Zhang Y, Chen N. PPT: A minor language news recommendation model via cross-lingual preference pattern transfer. Proc 63rd Annual Meeting of the Association for Computational Linguistics. 2025;1:10728–45.
89. Upadhayay B, Behzadan V. Tongue-tied: Breaking LLMs safety through new language learning. Proc 7th Workshop on Computational Approaches to Linguistic Code-Switching 2025;32–47. doi:10.18653/v1/2025.calcs-1.5

90. Li Z, Ji S, Luo H, Tiedemann J. Rethinking multilingual continual pretraining: Data mixing for adapting LLMs across languages and resources. arXiv [Preprint]. 2025a. doi:10.48550/arXiv.2504.04152
91. Toraman C. Adapting open-source generative large language models for low-resource languages: A case study for Turkish. In: Proceedings of the Fourth Workshop on Multilingual Representation Learning (MRL 2024). 2024a;30–44. doi:10.18653/v1/2024.mrl-1.3
92. Wang S, Xie Y, Ding B, Gao J, Zhang Y. Language adaptation of large language models: An empirical study on LLaMA2. In: Proceedings of the International Conference on Computational Linguistics. 2025a;7195–208. Available from: https://aclanthology.org/2025.coling-main.480/
93. Nahin SK, Nandi RN, Sarker S, Muhtaseem QS, Kowsher M, Shill AC, et al. TituLLMs: A family of Bangla LLMs with comprehensive benchmarking. In: Findings of the Association for Computational Linguistics: ACL 2025. 2025;24922–40. doi:10.18653/v1/2025.findings-acl.1279
94. Wang H, Minervini P, Ponti EM. Probing the emergence of cross-lingual alignment during LLM training. Findings of the Association for Computational Linguistics: ACL 2024;12159–73.
95. Csaki Z, Li B, Li J, Xu Q, Pawakapan P, Zhang L, et al. Sambalingo: Teaching large language models new languages. In: Proceedings of the Fourth Workshop on Multilingual Representation Learning (MRL 2024). 2024;1–21. doi:10.18653/v1/2024.mrl-1.1
96. Csaki Z, Pawakapan P, Thakker U, Xu Q. Efficiently adapting pretrained language models to new languages. arXiv preprint. 2023;2311.05741. doi:10.48550/arXiv.2311.05741
97. Ghazaryan G, Arakelyan E, Minervini P, Augenstein I. SynDARin: Synthesising datasets for automated reasoning in low-resource languages. *Proc 31st Int Conf Computational Linguistics*;6459–66.
98. Zhu H, Liang Y, Xu W, Xu H. Evaluating large language models for in-context learning of linguistic patterns in unseen low-resource languages. Proc 1st Workshop on Language Models for Low-Resource Languages 2025;414–426.
99. Zhu X, Gardiner S, Roldán T, Rossouw D. The model arena for cross-lingual sentiment analysis: A comparative study in the era of large language models. Proc 14th Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media Analysis 2024;141–52.
100. Merx R, Mahmudi A, Langford K, de Araujo LA, Vylomova E. Low-resource machine translation through retrieval-augmented LLM prompting: A study on the Mambai language. Proc 2nd Workshop on Resources and Technologies for Indigenous, Endangered and Lesser-resourced Languages in Eurasia (EURALI) @ LREC-COLING 2024;1–11. Available from: https://aclanthology.org/2024.eurali-1.1/
101. Peng Z, Wu X, Wang Q, Fang Y. Soft prompt tuning for augmenting dense retrieval with large language models. Knowl-Based Syst 2025;309:112758. doi:10.1016/j.knosys.2024.112758
102. Ruciński S. Efficient language adaptive pre-training: Extending state-of-the-art large language models for Polish. arXiv [Preprint]. 2024. Available from: https://doi.org/10.48550/arXiv.2402.09759
103. Iyer V, Malik B, Zhu W, Stepachev P, Chen P, Haddow B, Birch A. Exploring very low-resource translation with LLMs: The University of Edinburgh's submission to AmericasNLP 2024 translation task. Proc 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024);209–20. Available from: https://openreview.net/pdf?id=b4H7plMwVU

104. Lavrinovics E, Biswas R, Bjerva J, Hose K. Knowledge graphs, large language models, and hallucinations: An NLP perspective. J Web Semant 2025;85:100844. doi:10.1016/j.websem.2024.100844
105. Pengpun P, Udomcharoenchaikit C, Buaphet W, Limkonchotiwat P. Seed-free synthetic data generation framework for instruction-tuning LLMs: A case study in Thai. Proc 62nd Annual Meeting of the Association for Computational Linguistics;4:445–64. 2024. Available from: https://aclanthology.org/2024.acl-srw.50.pdf
106. Skadiņa I, Bakanovs B, Darģis R. First steps in benchmarking Latvian in large language models. Proc 3rd Workshop on Resources and Representations for Under-Resourced Languages and Domains (RESOURCEFUL-2025);86–95.
107. Zhao Y, Zhang W, Wang H, Kawaguchi K, Bing L. AdaMergeX: cross-lingual transfer with large language models via adaptive adapter merging. Proc 2025 Conf North Am Chapter Assoc Comput Linguist: Hum Lang Technol (NAACL-HLT 2025). 2025;1:9785–800. doi:10.18653/v1/2025.naacl-long.493.
108. Singh V, Krishna A, Karthika NJ, Ramakrishnan G. A three-pronged approach to cross-lingual adaptation with multilingual LLMs. arXiv [Preprint]. 2024. Available from: https://arxiv.org/abs/2406.17377
109. Kunz J, Holmström O. The impact of language adapters in cross-lingual transfer for NLU. Proc MOOMIN 2024;24–43.
110. Schmidt FD, Borchert P, Vulić I, Glavaš G. Self-distillation for model stacking unlocks cross-lingual NLU in 200+ languages. Findings Assoc Comput Linguist: EMNLP 2024;6724–43. doi:10.18653/v1/2024.findings-emnlp.394
111. Sani SM, Sadeghi P, Vu TT, Yaghoobzadeh Y, Haffari G. Extending LLMs to new languages: A case study of Llama and Persian adaptation. In: Proceedings of the 31st International Conference on Computational Linguistics. 2025;8868–84. Available from: https://aclanthology.org/2025.coling-main.594/
112. Toraman C. Llamaturk: Adapting open-source generative large language models for low-resource language. arXiv preprint. 2024b;2405.07745. doi:10.48550/arXiv.2405.07745
113. Acikgoz EC, Erdogan M, Yuret D. Bridging the Bosphorus: Advancing Turkish large language models through strategies for low-resource language adaptation and benchmarking. arXiv [Preprint]. 2024. doi:10.48550/arXiv.2405.04685
114. Ebrahimi A, Mager M, Oncevay A, Chaudhary V, Chiruzzo L, Fan A, et al. AmericasNLI: evaluating zero-shot natural language understanding of pretrained multilingual models in truly low-resource languages. Proc 60th Annu Meet Assoc Comput Linguist (ACL 2022). 2022;1:6279–99. doi:10.18653/v1/2022.acl-long.435.
115. Kuulmets HA, Purason T, Luhtaru A, Fishel M. Teaching Llama a new language through cross-lingual knowledge transfer. Findings NAACL 2024;3309–3322.
116. Tao M, Zhang C, Huang Q, Ma T, Huang S, Zhao D, Feng Y. Unlocking the potential of model merging for low-resource languages. Findings Assoc Comput Linguist: EMNLP 2024;8705–20. doi:10.18653/v1/2024.findings-emnlp.508
117. Aggarwal D, Sathe A, Sitaram S. Exploring pretraining via active forgetting for improving cross-lingual transfer for decoder language models. arXiv [Preprint]. 2024. doi:10.48550/arXiv.2410.16168

118. Wang Z, Li J, Zhou H, Weng R, Wang J, Huang X, Huang S. Investigating and scaling up code-switching for multilingual language model pre-training. Findings of the Association for Computational Linguistics: ACL 2025b;11032–46.
119. Feng S, Shi W, Wang Y, Ding W, Ahia O, Li SS, et al. Teaching LLMs to abstain across languages via multilingual feedback. Proc 2024 Conf Empir Methods Nat Lang Process (EMNLP 2024). 2024;4125–50. doi:10.18653/v1/2024.emnlp-main.239.
120. Lyu S, Deng Y, Liu G, Qi Z, Wang R. Transferable modeling strategies for low-resource LLM tasks: A prompt and alignment-based approach. Proc 7th Int Conf Artificial Intelligence Technologies and Applications (ICAITA 2025);225–9. IEEE; 2025. doi:10.1109/ICAITA67588.2025.11137773
121. Li Y, Zhao Z, Scarton C. It's all about in-context learning! Teaching extremely low-resource languages to LLMs. arXiv [Preprint]. 2025b. doi:10.48550/arXiv.2508.19089
122. Li L, Hu W, Luo M. PNMT: Zero-resource machine translation with pivot-based feature converter. *Computers, Materials & Continua*. 2025c;84(3). doi:10.32604/cmc.2025.064349
123. Tanwar E, Dutta S, Borthakur M, Chakraborty T. Multilingual LLMs are better cross-lingual in-context learners with alignment. Proc ACL 2023;1:6292–6307. doi:10.18653/v1/2023.acl-long.346
124. Sindhujan A, Kanojia D, Orasan C, Qian S. When LLMs struggle: Reference-less translation evaluation for low-resource languages. Proc 1st Workshop on Language Models for Low-Resource Languages;437–59. Available from: https://aclanthology.org/2025.loreslm-1.33.pdf
125. Court S, Elsner M. Shortcomings of LLMs for low-resource translation: Retrieval and understanding are both the problem. Proc 9th Conf on Machine Translation;1332–54. 2024. Available from: https://aclanthology.org/2024.wmt-1.125.pdf
126. Asai A, Kudugunta S, Yu XV, Blevins T, Gonen H, Reid M, et al. BUFFET: Benchmarking large language models for few-shot cross-lingual transfer. Proc NAACL-HLT 2024;1:1771–1800. doi:10.18653/v1/2024.naacl-long.100
127. Kargaran AH, Modarressi A, Nikeghbal N, Diesner J, Yvon F, Schütze H. MEXA: Multilingual evaluation of English-centric LLMs via cross-lingual alignment. Findings ACL 2025;27001–27023. doi:10.18653/v1/2025.findings-acl.1385
128. Zhang C, Liu X, Lin J, Feng Y. Teaching large language models an unseen language on the fly. arXiv [Preprint]. 2024. doi:10.48550/arXiv.2402.19167
129. Wen-Yi AW, Mimno D. Hyperpolyglot LLMs: cross-lingual interpretability in token embeddings. Proc 2023 Conf Empir Methods Nat Lang Process (EMNLP 2023). 2023;1124–31. doi:10.18653/v1/2023.emnlp-main.71.
130. Tran KT, O'Sullivan B, Nguyen H. Irish-based large language model with extreme low-resource settings in machine translation. In: Proceedings of the Seventh Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2024). 2024;193–202. Available from: https://aclanthology.org/2024.loresmt-1.20.pdf
131. Purason T, Kuulmets HA, Fishel M. LLMs for extremely low-resource Finno-Ugric languages. Findings Assoc Comput Linguist: NAACL 2025;6677–97. doi:10.18653/v1/2025.findings-naacl.373

132. Gusmita RH, Firmansyah AF, Zahera HM, Ngomo ACN. ELEVATE-ID: Extending large language models for end-to-end entity linking evaluation in Indonesian. Data Knowl Eng 2025;161:102504. doi:10.1016/j.datak.2025.102504

133. Mahfuz T, Dey SK, Naswan R, Adil H, Sayeed KS, Shahgir HS. Too late to train, too early to use? A study on necessity and viability of low-resource Bengali LLMs. Proc 31st Int Conf Computational Linguistics;1183–200.

134. Xu S, Dong W, Guo Z, Wu X, Xiong D. Exploring multilingual concepts of human values in large language models: Is value alignment consistent, transferable and controllable across languages? Findings EMNLP 2024;1771–1793.

135. Kia MA, Samiee D. From monolingual to multilingual: Enhancing hate speech detection with multi-channel language models. Procedia Comput Sci 2024;246:2704–2713. doi:10.1016/j.procs.2024.09.401

136. Tejaswi A, Gupta N, Choi E. Exploring design choices for building language-specific LLMs. In: Findings of the Association for Computational Linguistics: EMNLP 2024. 2024;10485–500. doi:10.18653/v1/2024.findings-emnlp.614

137. Zhang B, Haddow B, Birch A. Prompting large language model for machine translation: A case study. *Proc 40th International Conference on Machine Learning (ICML 2023)*. Proceedings of Machine Learning Research. 2023;202:41092–110. Available from: https://proceedings.mlr.press/v202/zhang23m.html

138. Pei R, Liu Y, Lin P, Yvon F, Schütze H. Understanding in-context machine translation for low-resource languages: A case study on Manchu. Proc 63rd Annual Meeting of the Association for Computational Linguistics. 2025;1:8767–88. Available from: https://doi.org/10.18653/v1/2025.acl-long.429

139. Sato M, Obara Y, Souma N, Kuramitsu K. CL-HumanEval: a benchmark for evaluating cross-lingual transfer through code generation. Proc 38th Pac Asia Conf Lang Inf Comput (PACLIC 2024). 2024 Dec;656–64.

140. Jung JY, Tyack L, von Davier M. Towards the implementation of automated scoring in international large-scale assessments: Scalability and quality control. Comput Educ Artif Intell 2025;8:100375. doi:10.1016/j.caeai.2025.100375