Computer Modeling in Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.075507

REVIEW

Advancing Large Language Models for Low-Resource Languages: A Systematic Review of Pretraining, Adaptation, and Ethical Challenges

Ismail Hossain[1], Mridul Banik[2], Fahmid Al Farid[3,4], Jia Uddin[5,*] and Hezerul bin Abdul Karim[4,*]

[1]Department of Computer Science, George Mason University, Fairfax, VA, USA
[2]Department of Computer Science, Colorado State University, Fort Collins, CO, USA
[3]Faculty of Computer Science and Informatics, Berlin School of Business and Innovation, Berlin, Germany
[4]Centre for Image and Vision Computing (CIVC), COE for Artificial Intelligence, Faculty of Artificial Intelligence and Engineering (FAIE), Multimedia University, Cyberjaya 63100, Selangor, Malaysia
[5]Artificial Intelligence and Big Data Department, Endicott College, Woosong University, Daejeon, Republic of Korea

*Corresponding Authors: Jia Uddin. Email: jia.uddin@wsu.ac.kr; Hezerul bin Abdul Karim. Email: hezerul.abdul@mmu.edu.my

Received: 03 November 2025; Accepted: 02 April 2026; Published: 27 July 2026

ABSTRACT: In recent years, the rapid advancement of Large Language Models (LLMs) has significantly transformed natural language processing (NLP), enabling impressive performance across a wide range of tasks. However, these developments have largely benefited high-resource languages, leaving many low-resource and underrepresented languages at risk of further digital marginalization. Addressing this imbalance is crucial to building more inclusive and culturally sustainable AI systems, which is motivating growing research interest in adapting LLMs for linguistically diverse and resource-scarce communities. This systematic review examines recent progress (2020–2025) in the pretraining and adaptation of LLMs for Low-Resource Languages (LRLs). Analysed 812 records obtained in the large databases and using PRISMA criteria, 140 core studies were identified. The innovations in data augmentation and parameter-efficient fine-tuning approaches can be outlined in this selection process. It combines major innovations on data-driven augmentation, parameter-efficient fine-tuning and morphologically rich and underrepresented language script-sensitive tokenization. The results highlight the growing effectiveness of culturally aware standards such as IrokoBench and BLEnD and show that approaches to lightweight adaptation eliminate high computational costs while maintaining language accuracy. The review focuses on the ethics in AI practice, the development of corpora through communities, and interdisciplinary research collaboration among computational linguists, social scientists, and digital humanists. The task of generating a diversified dataset, typology-conscious modelling strategies, and open-source multilingual benchmarks should be prioritized in future research as one possible solution to the existing digital language gap worldwide.

KEYWORDS: LLMs; low-resource languages (LRLs); parameter-efficient fine-tuning (PEFT); tokenization; cross-lingual transfer; multilingual benchmarks

1 Introduction

The rise of LLMs in recent years has made them the core component of NLP, with state-of-the-art performance on a wide variety of applications such as machine translation, question answering, and content summarization [1–3]. Recent studies have extended cross-lingual transfer techniques to code-focused LLMs, demonstrating that knowledge learned from high-resource programming languages can be effectively transferred to low-resource ones, despite structural and syntactic differences [4]. Models such as

GPT-3, PaLM, and LLaMA-2 are endowed with terabytes of data sourced at internet scale, which is dominated by English and other well-resourced languages [1]. These models show impressive results; however, in low-resource languages (LRLs), they do not perform well and are characterised by a lack of large-scale digitised corpora, linguistic tools, and computational support [5].

A large proportion of the world's population speaks low-resource languages, ensuring that they are at the heart of cultural identity, education, and civic engagement [5]. 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 improve LRL representation have included the Glot500, BLOOM, and 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 across contexts of limited resources, explaining the barriers to implementing inclusive development of LLMs [6]. To democratise access to the capabilities of English-dominant models, approaches that leverage the capabilities of linguistically diverse prompts have been proposed [8]. For instance, Adelani et al. introduced IrokoBench, a culturally grounded benchmark designed to evaluate LLM performance specifically on African languages, providing standardized evaluation criteria for underrepresented linguistic communities [9]. Similarly, Alhanai et al. proposed culturally adjusted fine-tuning methods and benchmarking protocols aimed at enhancing LLM capabilities for low-resource African languages [10]. In parallel, Cassano et al. demonstrated that cross-lingual transfer techniques can be extended to code-focused LLMs, showing that knowledge acquired from high-resource programming languages can improve performance on low-resource ones [4].

Moreover, this review article 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 [1]. It introduces a combination of approaches, including further pretraining, generating synthetic data, parameter-efficient fine-tuning (PEFT), tokenizer adaptation, and instruction tuning [3]. This aims to highlight their advantages, drawbacks, and applicability in various LRL instances. Additionally, recent critical reviews highlight methodological issues and intrinsic limitations in current research on LLMs, particularly in terms of evaluation and reproducibility [2]. In this way, this review adds value as a comprehensive and up-to-date synthesis (2020–2025) of scalable, reproducible, and typology-aware methods specifically tailored for use in low-resource language environments, through their application to LLMs [2].

The main focus of this review is to achieve exponential gains in data efficiency across Data Strategies, Training Pipelines, and PEFT through a unified LRL-based paradigm [11]. By focusing on multilingual assessment and including LRLs from various linguistic groups, such as Bantu, Dravidian, and Austroasiatic [9,12], it emphasises typological inclusivity. Another crucial dimension is practical adaptation, which provides guidelines for adopting cost-efficient methods, such as Low-Rank Adaptation (LoRA) and instruction tuning, in environments with limited finances or infrastructure [13]. Moreover, the analysis incorporates an ethical and representational approach that addresses concerns related to cultural equity, representation, and bias in LLMs when applied to LRLs [14]. The interconnected, multilingual, modular, and dialectal issues of failure to adapt LRLs require integrated methodologies to achieve strong adaptation [11]. However, several issues restrict advancement in this area. Corpus imbalance is a significant issue, as most LLMs are trained on highly skewed datasets that predominantly focus on English and other high-resource languages [6]. Inefficiencies caused by tokenisers make the issue more critical, especially for the morphologically rich or non-Latin scripts, which are sometimes underrepresented [15]. Reproducibility and accessibility are also problematic because not all LLMs and LRL benchmarks are open-access and documented [2]. There are also cultural misfits and ethical problems, as LLMs often fail to understand cultures delicately or create culturally insensitive products for LRLs [14]. Restricted computing capabilities pose an additional hurdle, particularly

for researchers in areas where LRLs are widely used [16]. Finally, inadequate assessment systems among LRLs do not allow for consistent tracking of progress [17].

Another fundamental goal is to assess parameter-efficient fine-tuning architectures, such as LoRA, adapter layers, and instruction tuning, to adapt LLMs to LRLs [18]. The approaches are especially useful in computationally limited settings, and the article reviews the tendency of the methods to reduce training expenses without compromising or improving the performance of multilingual models [19]. The research also addresses issues of tokenisation and subword modelling in low-resource, morphologically rich languages [15]. Present tokenisers are frequently trained using Latin-script or English-based corpora and do not adequately account for LRLs. The approaches surveyed in this review include script-sensitive functions, tokenisers, and adaptive subword segmentation algorithms that aim to represent a wider range of orthographic systems [15]. Lastly, it suggests possible avenues that researchers and developers working in low-budget environments can explore, utilising LLMs in their native languages [20].

The intended audience of this review is a wide-ranging group, including NLP researchers active in the field of multilingual and low-resource language systems, developers interested in the approaches to deploying the LLMs to underrepresented languages, and the group of policymakers and financiers who want to focus on the future of the equitable and inclusive development of AI [10]. It is also relevant to linguists and digital humanists who work on language preservation and the documentation of endangered languages [21]. Additionally, the review is applicable to Global South developers and educators navigating the adoption of AI technologies in resource-constrained institutional contexts [22].

Although there have been methodological improvements, machine translation for low-resource languages remains limited by data scarcity [23]. The use of generative language models risks deepening linguistic inequality; thus, there is a need to create systematic measures specific to languages with limited resources [23]. The given review will help the development of low-resource languages by drawing on a thorough investigation of pre-training practices and adaptation strategies [6]. The research uncovers the gaps characterised by ethics, pretraining, and adaptation, particularly the societal and participatory dimensions of low-resource language development [21]. Recent survey evidence further confirms that low-resource machine translation remains constrained by data scarcity, evaluation inconsistency, and domain mismatch, even in the presence of multilingual and LLMs [24,25]. LLMs, hence, enable intelligent question answering, personalization, and adaptive learning, including language-driven interaction in immersive virtual reality environments, where linguistic adaptability becomes essential for inclusive user experiences. The studies summarise the key issues in big language model pretraining, adaptation, and ethical concerns that drive research in low-resource machine translation [1,19,26]. It highlights the underrepresentation of LRLs, gaps in current pretraining, constraints on adoption strategies, and the resulting neglect.

LLMs have also been explored in immersive virtual reality environments, where their integration supports language-driven interaction and adaptive learning experiences [27].

The objectives of the study are as follows:

- It aims to find existing pre-training and adaptation system practices on low-resource languages.
- To assess their effectiveness in a variety of linguistic typologies, orthographic systems, and corpus sizes.
- To study the practical and ethical issues that lie in the application of low-resource language technologies in practice.

2 Methodology

The methodology of this review was systematically examined and synthesised, focusing on recent developments in adapting LLMs to LRLs [1]. The field of review is deliberately narrowed to the academic

literature of 2020–2025, allowing for a more recent and topical analysis in the required manner [26]. This temporal restriction enables the review to focus on recently advanced strategies and methodologies that have emerged in the context of dynamic innovation in LLM research [1,28,29]. Integrated LLM playgrounds that facilitate rapid prototyping and controlled fine-tuning have been proposed as pragmatic environments for multilingual experimentation [28].

The primary objective is to identify key research questions that will deepen understanding of LLM adaptation in LRLs [6]. Such questions comprise: What pre-training 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? [30].

The screening of the articles was performed independently by two authors, based on the defined inclusion and exclusion criteria. Titles and abstracts were sifted through in the first stage to remove apparent irrelevant research. The appropriate articles were identified, and in the next phase, the complete texts of the potentially eligible articles were accessed and reviewed. Disagreements among reviewers were resolved through discussion and consensus, ensuring uniformity and transparency in the study selection process. The general screening and selection process followed PRISMA guidelines, and the total number of included studies was recorded in a flow diagram. The PRISMA checklists are available in the supplementary file.

The data were extracted as follows: first author and year of publication; target language(s) and linguistic family; geographical region; model architecture; type of adaptation or pretraining strategy; data sources; evaluation benchmarks; performance metrics; computational requirements; and the availability of code or datasets. The discrepancies found during data extraction were eliminated by jointly verifying the articles. Specific attention was given to indicators of methodological clarity and reproducibility, such as open-source availability and experimental transparency.

The primary outcome measures: (1) How effectively can LLM adaptation strategies be applied to low-resource languages, and how well do they perform, assessed by task-specific metrics such as BLEU, F1-score, and perplexity? (2) The computational efficiency of adaptation methods, specifically parameter-efficient fine-tuning strategies applicable to LRLs.

The secondary outcomes included: (1) The strength of adaptation to a wide range of linguistic typologies and scripts; (2) The effects of tokenization and data-augmentation on morphologically rich or non-Latin-script languages; (3) The reproducibility and accessibility of the concept of developing the methods, and (4) The ethical, cultural, and representational concerns related to the implementation of LLMs in low-resource language conditions.

Fig. 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 is the most common, followed by parallel corpora, suggesting increased use of artificially generated data to address data scarcity in LRLs [12]. Monolingual data and knowledge-graph-based data also represent significant shares, while data gathered through web crawling and other tools occupy relatively insignificant positions in the collection [3].

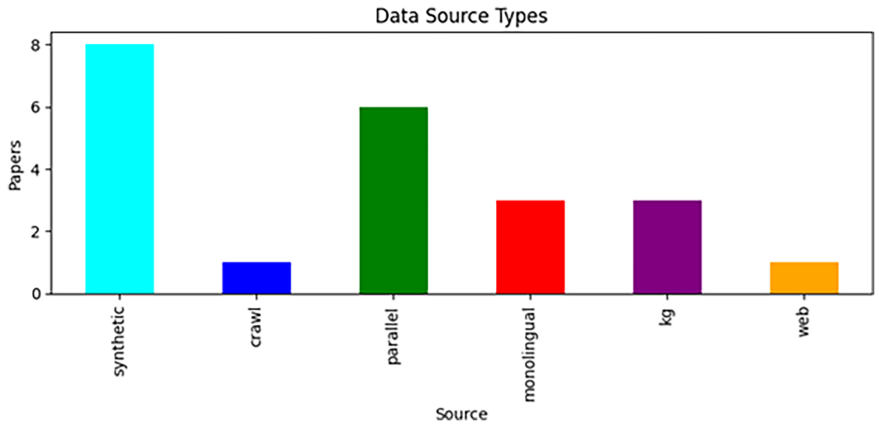

Figure 1: Types of data sources used by surveyed studies.

The entire review placed importance on reproducibility and accessibility, assigning higher ranks to studies that were freely published and well-documented [3]. The discussed focus aims to address a crucial gap in the field: the availability of many popular LLMs and LRL benchmarks is limited to in-person access, thereby impeding development and activity [2]. In addition, the review employs culturally sensitive assessment structures, enabling a nuanced analysis of LLMs deployment that is sensitive to the contextual cultural and linguistic peculiarities in diverse settings [31]. Within the scope of identifying, adopting, and developing such structures, the review presents a methodologically sound, holistic study of the adaptation of LLMs to LRLs [1,12]. Fig. 2 shows the step-wise overall selection process for maintaining reproducibility and transparency.

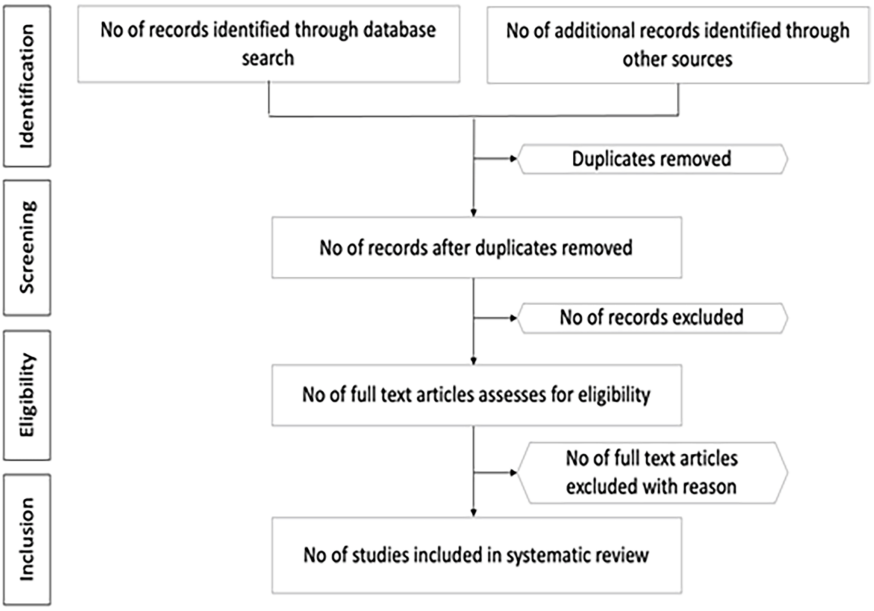

Figure 2: Flowchart of the article selection process.

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 make meaningful contributions to the understanding and advancement of LLMs for 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 LRLs, addressing key challenges such as data scarcity, linguistic diversity, and computational 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 centred exclusively on high-resource languages (e.g., English, Mandarin) were excluded because they were not relevant to LRL-specific challenges.
- Papers that mention LRLs, however, do not propose concrete methods or experiments and 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 to the review, allowing for a thorough yet focused study of the emerging methodology and strategies for developing LLMs in low-resource language environments. This level of systematization not only increases the reliability of the results but also gives a clear guideline to further research and the actual implementation of the findings in this field [32].

2.2 Data Collection and Review Process

The research was conducted sequentially throughout the data collection and review process to incorporate high-quality, relevant literature that addresses the study's purposes. This initial search strategy involved querying major academic search engines, such as Google Scholar, ACL Anthology, and arXiv, using specialised terms, including LRL modelling, multilingual LLMs, continued pretraining, LoRA adaptation, and synthetic corpora for LRLs. The identified keywords aimed to encompass as many studies as possible on the various avenues of LLMs in LRLs. The first search retrieved 812 articles and indicated the level of interest and research in the field [32].

To simplify the study selection process, it was conducted according to the PRISMA guidelines to ensure transparency and reproducibility. The PRISMA flowchart (Fig. 3) gives a comprehensive report of such a selection process. This figure illustrates the identification of relevant studies in the programming area, which was conducted through exhaustive searches of databases and grey literature. The initial search identified 812 records. The review finally comprised 140 studies after duplicates and irrelevant studies were removed. The final database search was conducted on September 2025. The full search strategies used for each database are provided in Appendix A.

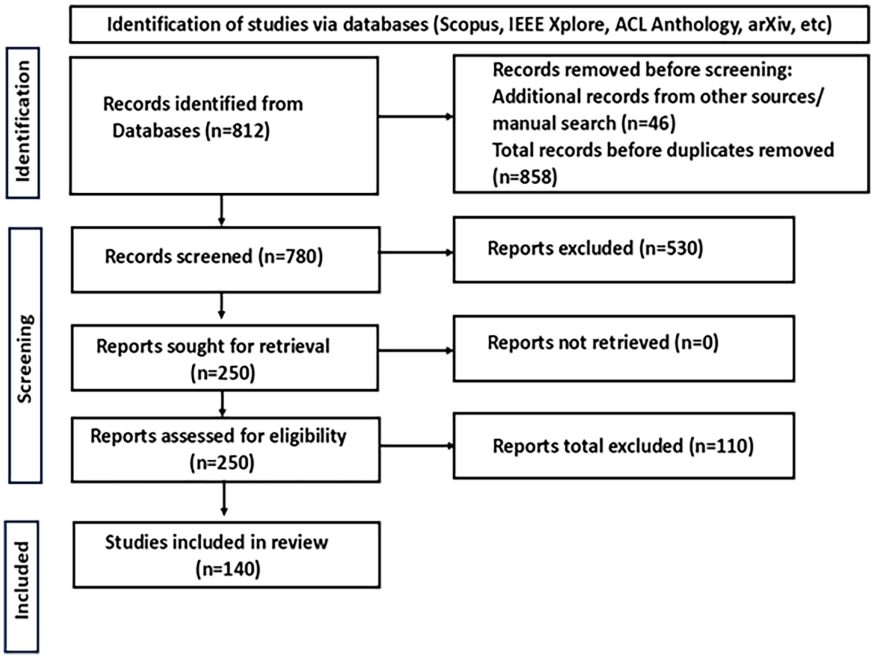

Figure 3: PRISMA flowchart.

This general data was logically categorised into the most significant thematic groups, along with the introduction of benchmarks to assess LLM performance. The development of fine-tuning methods with minimal parameters (PEFT) included LoRA and adaptation to different cultures [13]. The PRISMA flowchart provides a comprehensive description of the careful selection process, ensuring that the review reports the latest and most significant developments in LLM adaptation for LRLs [32]. By utilising this deep data, the review integrates scalable, reusable, and typology-sensitive methodologies specifically designed for low-resource language contexts and to overcome significant scalability gaps arising from data sparsity, linguistic heterogeneity, and computational inefficiency [6]. Such a method would both improve the accuracy of the results and provide a solid platform for informing future studies on the subject and for practical applications in the same area [3].

2.3 Literature Synthesis Framework

Studies, here, were grouped by adaptation strategy, linguistic family, and task. Narrative synthesis was used due to heterogeneity, with qualitative exploration across typology and tasks. Results were presented using tables and figures. No numerical data transformation was required. No formal sensitivity analysis was conducted.

The challenges of adapting LLMs to LRLs concern the need to go beyond descriptive summarization; the review evaluates empirical evidence on this study through a critical synthesis on four dimensions that are interrelated, such as data strategies, adaptation methods, linguistic factors, and ethical considerations. Although the survey reveals significant improvements through synthetic data generation and cross-lingual transfer [12,33], some have warned that excessive reliance on synthetic corpora can lead to semantic drift and task-specific deterioration, especially in sentiment analysis and named entity recognition (NER) systems [34,35]. This variation provides evidence that data augmentation is highly context-dependent rather than universally useful.

One area of research with a serious gap is the lack of comparative appraisals of data pipes across multiple language families. Models like UnifiedCrawl, Gemma2 and Latxa have proven effective for aggregating multilingual information [36–38], but there is limited empirical evidence of their generalizability to ultra-low-resource or non-standardised-script environments. The methodological approaches of many works focus on short-term performance improvement without longitudinal validation or a comprehensive analysis of errors, undermining assertions of robustness [2].

Similarly, results on PEFT methods are inconsistent. LoRA and adapter-based systems are also consistent in their ability to reduce computational cost [39,40], but various studies report reduced effectiveness in morphologically complex or typologically distant languages where full or hybrid fine-tuning is more effective than PEFT [37,41]. This points out a methodological weakness; most PEFT tests are based on small sets of benchmarks and are not cross-task consistent, which limits extraneous validity.

Research dedicated to the linguistic adaptation highlights the issue of tokenization as a bottleneck. Even though script-sensitive and morphology-aware tokenizers improve the quality of representation [15,20], the validity of evidence is not consistent because of the inconsistent evaluation indicators and the small size of the datasets used [23]. Studies on ethical and cultural adaptation, hence, depict disproportionate rigor in methodology, although frameworks for debiasing methods and culturally based standards are suggested [42,43]; their implementation and validation plans are not always rigorous.

To conclude, the general level of evidence in the literature is uneven. Many of these studies lack standardized benchmarks, repeatable pipelines, or transparent reporting [1,2], hindering the reproducibility of research results in LRL-oriented LLM adaptation.

This review defines four key dimensions to guide the development of large language models in low-resource language domains. Data strategies highlight the need to construct, expand, learn, transfer linguistic knowledge, and annotate a corpus by communities to address data scarcity. The techniques of adaptation focus on parameter-efficient fine-tuning, cross-linguistic transfer, and hybrid architectures to reduce the computational costs. The language aspect takes into account the morphological complexity, dialectal and cultural semantics. Ethical considerations include bias reduction, data privacy, cultural alignment, and fairness, which help ensure the deployment of large language models is responsible, inclusive, and sustainable for underrepresented languages.

2.3.1 Quantitative Aggregation of Performance Metrics

Quantitative metrics were aggregated across studies, particularly for underrepresented and extremely low-resource languages [44]. The combination of these measurements enabled us to determine which approaches consistently yield the best results in LRL situations [2]. Fig. 4 illustrates the most commonly used evaluation metrics across the surveyed literature.

2.3.2 Qualitative Assessment of Ethical Risks and Deployment Considerations

In addition to technical performance, the question of applying LLMs in LRL environments raised many interest, ethical, and pragmatic review issues. Among the issues covered in this qualitative assessment were cultural fairness, representation, and prejudice, particularly when modelling was used to present adaptations to languages with a smaller digital presence and unusual sociolinguistic features [14,31]. Other studies have highlighted the importance of culturally conscious training information and refinement activities, which are crucial for the equitable and inclusive adoption of LLMs [17]. The matters related to infrastructure constraints, reproducibility, and access were also addressed, including how to mitigate these issues in a practical context [2,6]. 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.

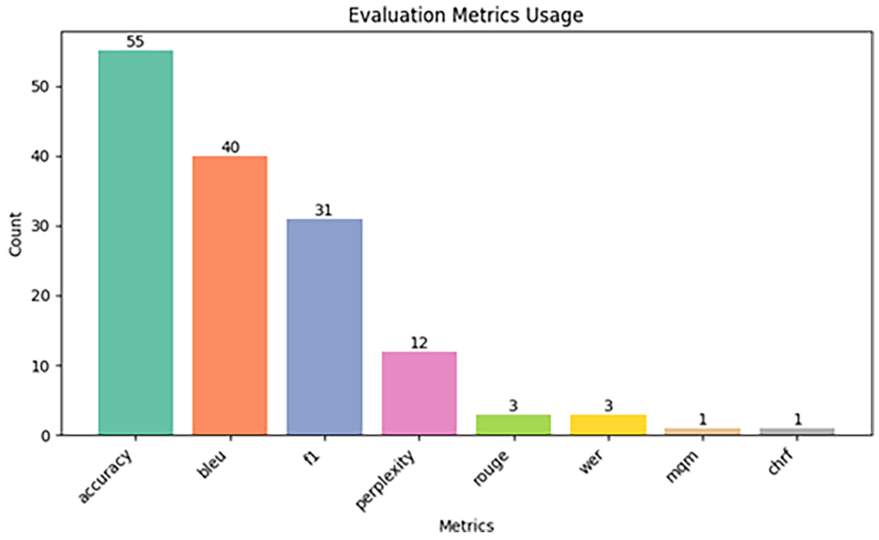

Figure 4: Most-used evaluation metrics across the surveyed literature.

2.4 Data Governance Frameworks

Community-curated corpora, consent-conscious data collection, and open-access licensing are among the data governance practices identified in studies. Participatory annotation and local stewardship are frameworks increasingly used to address extractive data practices; however, governance is uneven, with little to no record of provenance, ownership, and long-term maintenance plans, particularly for Indigenous and minority-language data.

2.5 Ethical–Technical Integration

Risk of bias was assessed qualitatively by examining openness of the dataset, transparency of the evaluation, and reproducibility indicators. Technical design decisions are closely linked to ethical risks. Biased or synthetic corpora used during fine-tuning are often known to increase representational errors in LRL settings, which have negative implications for sentiment, safety, and cultural accuracy. Experimental results show that parameter-efficient tuning transmits pretraining biases when not balanced by culturally grounded data, fairness-based objectives, or local assessment standards.

2.6 Limitations and Real-World Challenges

The major constraints are evidently benchmark fragmentation, inadequate evaluation metrics, and limited reproducibility due to closed models or datasets. Among the ethical risks, one may note misrepresentation of culture, amplification of bias, and asymmetry in access to computational resources. Practical implementation is also constrained by infrastructure limitations, a shortage of annotators, and mismatches between the model's results and local social-linguistic standards.

3 Findings

The results of the systematic review were based on a review of 140 works that applied LLMs to LRLs. The PRISMA flowchart presented in Fig. 3 illustrates the high level of rigor used in identifying, screening, and selecting these studies. The results were categorized into some of the major themes: the creation of benchmarks to analyze the performance of LLMs, implementation of PEFT techniques such as LoRA, and 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. The results of this systematic review also mostly focus on multilingual pre-training and transfer learning, usually at the expense of language-specific modelling. The paradigms of deep learning are established, and low-resource datasets remain rare and unevenly distributed. The technical problems are still in the form of intermittent performance, evaluation protocols, and dialectal dispersion. Moral issues, especially representational bias and cultural misalignment, are not adequately addressed. As a result, the literature highlights a pressing need for adaptive frameworks, standardized benchmarks, and methodologies that are explicitly low-resource, language-specific, and ethics-based.

3.1 Key Themes in the Findings

3.1.1 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 [9]. In a similar study, Myung et al. developed BLEnD, a multicultural benchmark for evaluating LLMs' performance across linguistic and cultural settings [45]. These guidelines offer standardized metrics on the adjustment of LRL as well as an understanding of how to improve it. Additionally, benchmarks like BasqBBQ [17] and LEIA [46] provide critical insights into social biases and cross-lingual knowledge transfer, an asymmetric phenomenon in LRLs [47]. Similarly, the CUTE dataset offers a multilingual resource specifically designed to enhance cross-lingual knowledge transfer in low-resource languages [48]. Fig. 5 illustrates the timeline of major dataset/benchmark releases relevant to LRL-LLM research (2020–2025).

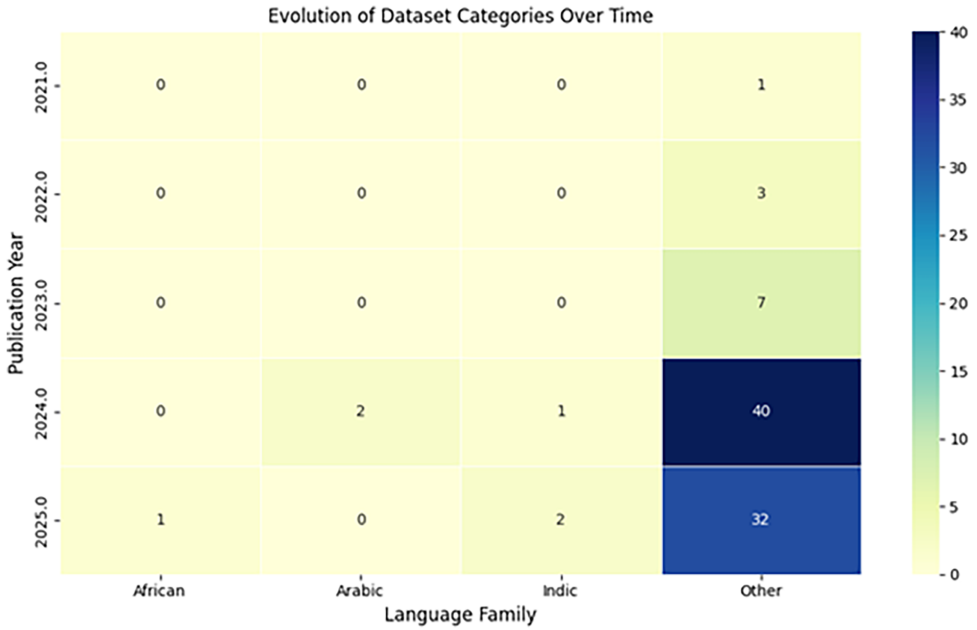

Figure 5: Timeline of major dataset/benchmark releases relevant to LRL-LLM research (2020–2025).

3.1.2 Application of PEFT Methods

The other significant theme is the PEFT methods for fine-tuning LLMs on LRLs. The efficiency of LoRA has been demonstrated in studies such as those by Le et al., which have shown that it can enhance LLM classification in clinical NLP tasks with limited data [39]. Khade et al. also discussed the

possibility of using LoRA for multilingual South Asian LRLs, which is appropriate given LoRA's low-compute adaptation [13]. These approaches have become efficient ways to fine-tune LLMs with limited resources. Additionally, Guo et al. demonstrated the optimization of translation for low-resource languages through efficient fine-tuning with custom prompt engineering in LLMs [49]. Similarly, Jiao et al. proposed ParroT, a framework that leverages human translation and feedback to tune LLMs for translation during interactive chat, demonstrating improved translation quality in resource-constrained settings [50]. Fig. 6 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 are shown in Fig. 7.

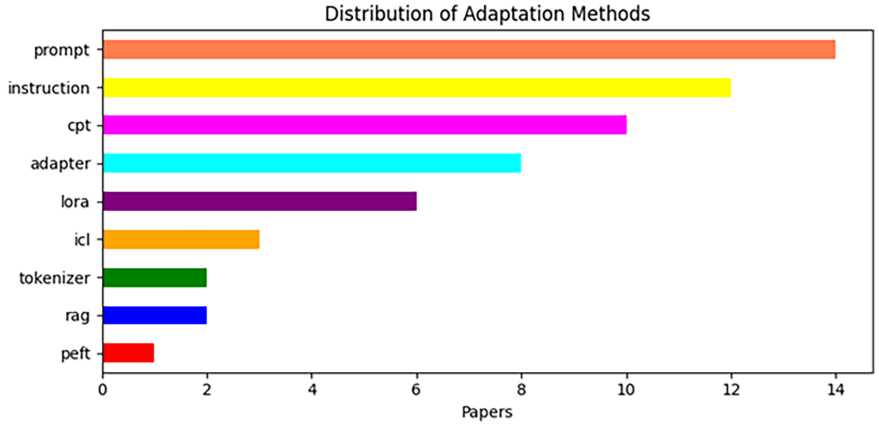

Figure 6: Frequency of adaptation strategies reported across the surveyed studies.

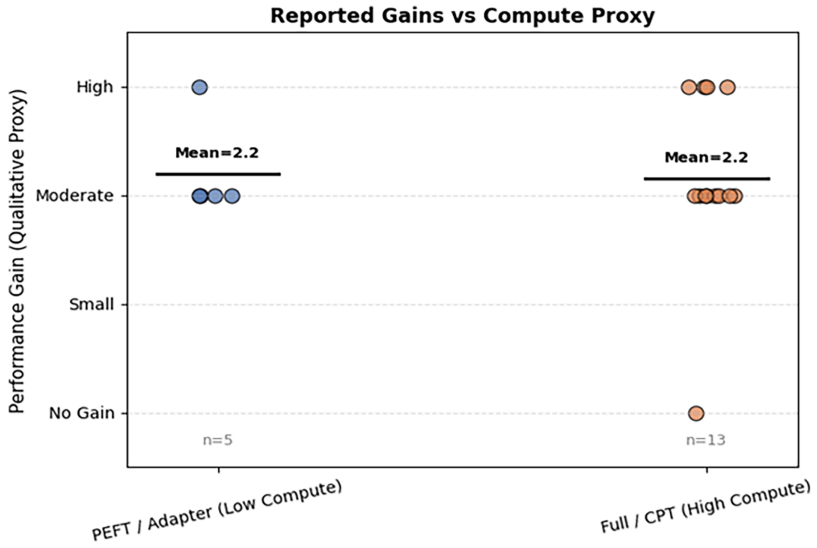

Figure 7: Performance gains vs. compute cost (proxy) for PEFT and full fine-tuning across studies.

Additionally, Fig. 8 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, summarisation, sentiment analysis, safety, dialogue, and classification, and fake news detection 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 [51,52]. On the same note, Huang et al. and Shibu et al. explored the application of LLMs to fake news detection in low-resource settings, demonstrating how language modelling techniques can support content verification and linguistic pattern recognition [53,54].

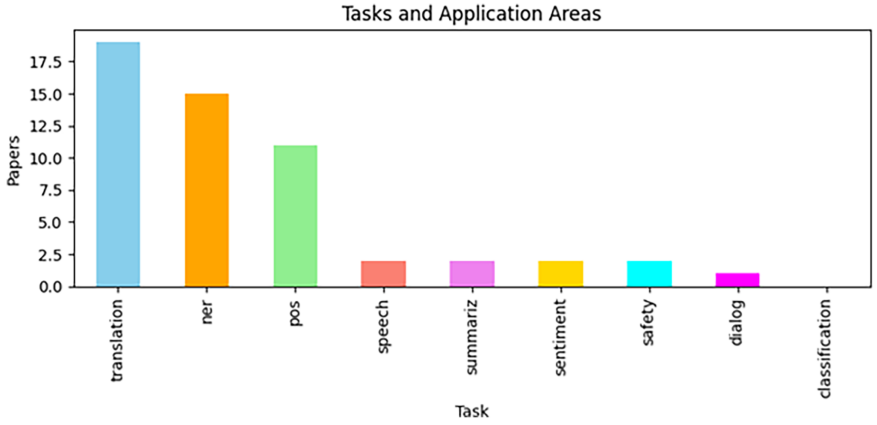

Figure 8: Tasks and application areas across surveyed studies.

Additionally, Song et al., Shen et al., and Mekki et al. identified sentiment analysis and safety issues in their studies [55–57]. Mekki et al. studied the use of NileChat to develop linguistically diverse and culturally aware LLMs for local communities [57], whereas Purwarianti et al. investigated dialogue systems and summarisation, with the NusaDialogue framework addressing the challenge of dealing with underrepresented languages [44,58]. However, the findings revealed a range of work conducted in the field of research on low-resource languages, while also highlighting areas that still require further research.

3.1.3 Cultural Sensitivity in Adaptation Strategies

Some studies have highlighted the urgency of culturally sensitive adaptation measures to address problems such as prejudice and mismatches in LLM output. Specifically, Shang et al. introduced Atlas-Chat for Moroccan Arabic dialectal adaptation [59], while Mekki et al. proposed NileChat to develop linguistically diverse and culturally aware LLMs for local communities [57]. These culturally adapted fine-tuning approaches utilised data from African LRLs. For instance, Adelani et al. compared LLM prompting with cross-lingual transfer performance on Indigenous and LR Brazilian languages, aiming to underscore the need for language- and context-sensitive, linguistically broad LLMs [60]. One of the most notable is the TACO framework proposed by Upadhayay and Behzadan, which enables relevant cross-lingual reasoning in LRLs through translation-aided chain-of-thought procedures [61]. Additionally, Fiaz et al. presented UrduLLaMA, the first spell-correcting morphologically rich language foundation model pipeline [62]. Another empirical study also indicates that fine-tuned language models retain the capacity for continual learning, allowing incremental adaptation to new tasks or languages when updates are carefully constrained [63,64]. Such fine-tuned solutions could also be applied to combinations of domains too small to interest the model designers [14,65,66].

Ethical questions have also not been excluded in recent studies. Concerns about bias and representational harm in the adaptation of LLMs to LRLs have underscored the need for responsible AI practices [67]. Furthermore, Shen et al. introduce methodologies for detecting safety challenges in multilingual contexts,

emphasizing the need for equitable access to LLMs [56]. Separately, Upadhayay and Behzadan examine how new language learning can expose vulnerabilities in LLM safety mechanisms, underscoring the risks of multilingual deployment [68]. Zhong et al. and De Klerk and McLean discuss the opportunities and challenges of deploying LLMs in humanities research for low-resource languages [67], while De Klerk and McLean examine the adoption of AI technologies in Global South higher education contexts [22]. Fig. 9 illustrates the proportion of studies in our review that explicitly address ethical considerations, including bias, toxicity, safety, and hallucinations. A small but significant minority (5.2%) of the reviewed papers focus on these critical issues, highlighting an emerging yet still underdeveloped area of research in LLM adaptation for low-resource languages.

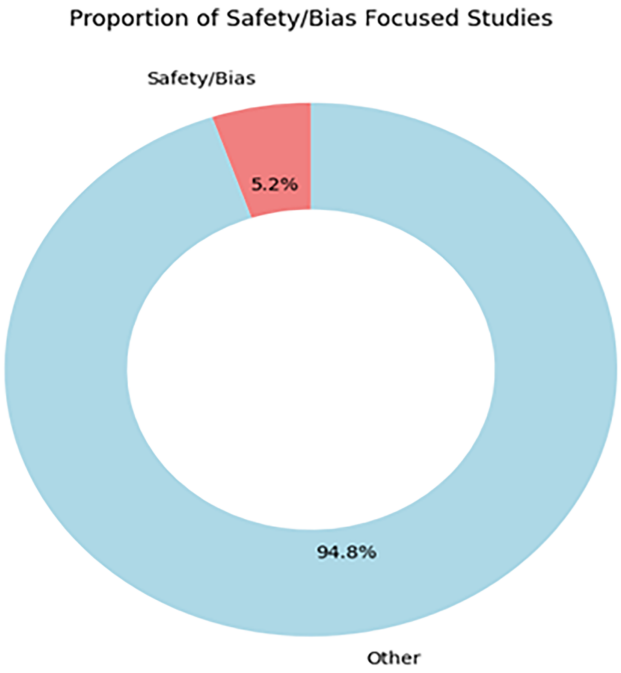

Figure 9: Proportion of safety/bias-focused studies.

3.1.4 Theoretical Contributions to LLM Adaptation for LRLs

The theoretical contribution outlines the theoretical progress in adapting LLMs to LRLs. It provides an extensive overview of the advances, techniques, and models emerging from the new research, along with the challenges and opportunities in this rapidly evolving field. Table 1 presents a detailed overview of 20 studies, providing information on their content, location, and publication, as well as the theoretical contributions they make to 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 (NAACL 2025)	Adelani et al. [9]	Proposes a new benchmark (IrokoBench) for evaluating LLMs on African languages; introduces culturally grounded evaluation framework.
2	2024	Latvia	USA (ACL)	Darģis et al. [69]	Provides evidence of open-source LLM performance in morphologically complex LRLs via real-world exams.
3	2025	Canada	USA (arXiv)	Le et al. [39]	Shows how LoRA improves LLM classification in clinical NLP with limited data—contribution to PEFT strategies.
4	2024	China	Netherlands (Elsevier)	Hu et al. [70]	Develops "Language Fusion via Adapters" theory for LRL speech recognition integration.
5	2024	Multinational	Canada (NeurIPS 2024)	Myung et al. [45]	Introduces BLEnD as a multicultural benchmark, contributing a new evaluation paradigm for LLMs across cultures.
6	2024	India	USA (arXiv)	Verma et al. [71]	Develops MILU benchmark; expands understanding of multitask learning for Indic LRLs.
7	2024	Morocco	USA (arXiv)	Shang et al. [59]	Presents Atlas-Chat as a case of dialectal adaptation; proposes culturally sensitive LLM adaptation strategies.
8	2024	Multinational (Africa)	USA (arXiv)	Alhanai et al. [10]	Suggests a culturally adjusted fine-tuning method and benchmarking protocol for African LRLs.

(Continued)

Table 1 (continued)

No.	Year	Study Country	Publication Country	Author(s) & Year	Theoretical Contribution
9	2022	Bangladesh	USA (Findings NAACL 2022)	Bhattacharjee et al. [72]	Introduces BanglaBERT; theoretical base for language-specific transformer adaptation.
10	2023	Multinational	Canada (ACL 2023)	ImaniGooghar et al. [7]	Proposes Glot500 multilingual corpus; advances scalable pretraining theory across 500 languages.
11	2024	Global	USA (arXiv)	Lin et al. [73]	Develops Mala-500 and explores massively multilingual adaptation; contributes to scaling LLM adaptation theory.
12	2023	Global	USA (arXiv)	Upadhayay and Behzadan [61]	Proposes TACO: Chain-of-thought + translation framework for cross-lingual reasoning.
13	2025	Pakistan	USA (arXiv)	Fiaz et al. [62]	Proposes UrduLLaMA—fine-tuned foundation model pipeline for a morphologically rich LRL.
14	2024	India	USA (ACL Anthology)	Khade et al. [13]	Studies LoRA PEFT for multilingual South Asian LRLs; highlights low-compute adaptation theory.
15	2023	Global	India	Researcher and Kalluri [30]	Proposes a conceptual framework for LLM adaptation ethics in LRLs.
16	2025	Kazakhstan	Switzerland (MDPI)	Kadyrbek et al. [74]	Suggests small-scale LLMs with Direct Preference Optimization for resource-constrained LRLs.
17	2024	Pakistan	USA (CHiPSAL 2025)	Tahir et al. [75]	Benchmarks LLMs in Urdu NLP; expands PEFT + multilingual tuning theoretical base.
18	2021	Mongolia	Germany (SpringerOpen)	Byambadorj et al. [76]	Applies cross-lingual transfer theory to low-resource TTS systems.

(Continued)

Table 1 (continued)

No.	Year	Study Country	Publication Country	Author(s) & Year	Theoretical Contribution
19	2025	Kenya	USA (arXiv)	Ngugi [77]	Introduces targeted lexical injection into early layers of LLMs to improve alignment in LRLs.
20	2024	Global South	USA (arXiv)	Zhong et al. [67]	Analyzes socio-technical challenges of using LLMs in humanities research for low-resource languages.

The geographical distribution of theoretical contributions to LLM adaptation for LRLs reveals significant insights into the research landscape. Additionally, Fig. 10 shows the number of papers from different countries or regions in the reviewed literature on LLM adaptation for LRLs. China leads with 23 contributions, followed by the USA (11), India (9), Japan, the 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. Fig. 11 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 global interest in addressing the challenges of LLM adaptation for LRLs, particularly in culturally and linguistically diverse regions.

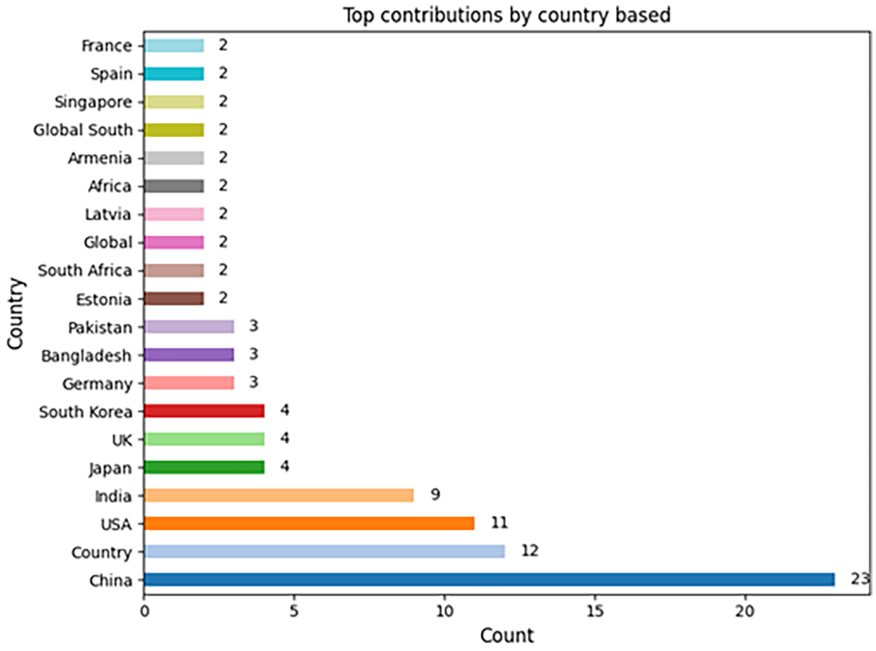

Figure 10: Top contributions by country.

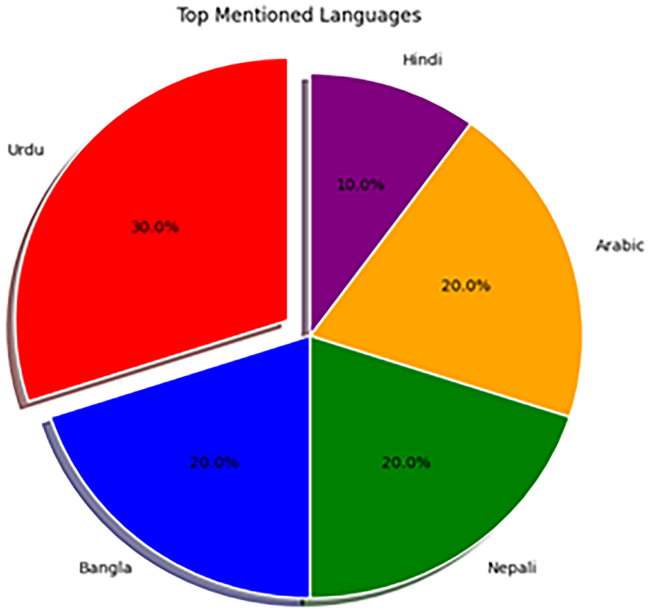

Figure 11: Top mentioned languages.

Moreover, the analysis of publication trends reveals significant growth in research on LLM adaptation for LRLs over the past five years. As illustrated in Fig. 12, the number of publications increased steadily from 2020 to 2025, reaching a peak of several papers in 2024. However, there was a slight decline in 2025. This trend highlights the growing importance of LRL-specific research and indicates potential shifts in scholarly focus.

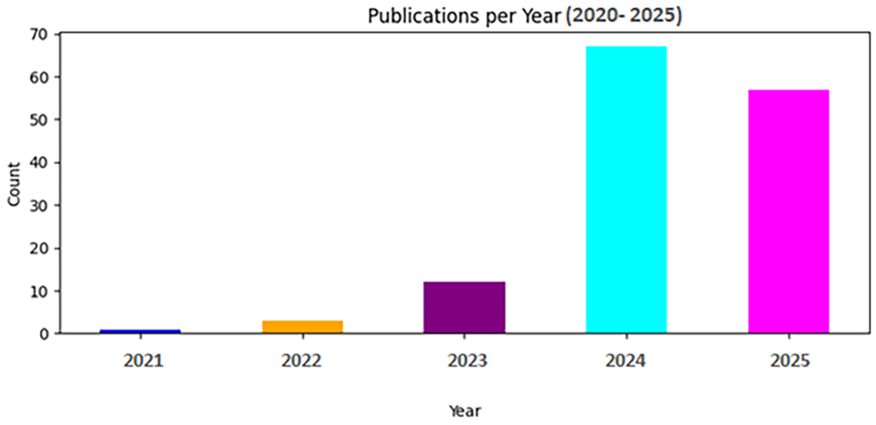

Figure 12: Annual number of publications on LLMs for low-resource languages (2020–2025).

This section provides an overview of the adaptation techniques available to tune LLMs on LRLs in various regions. Table 2 summarizes 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, it ranks first with the highest adaptation count: LoRA used three times, Adapter layers once, Tokenizer Tuning once, and Instruction Tuning twice, totalling seven. Africa is next with five adaptations, but fewer cases are observed in Europe, North America, and the Global/Mixed category, due to differences in attention to the topic and resource allocation across these regions [9,13].

Table 2: Adaptation methods by region based on the 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

Fig. 13 presents the same information in a visual format, focusing on regional differences in adaptation technique use. The visual representation enables us to identify patterns, such as the popularity of LoRA as a preferred approach across regions, the underutilization of the Adapter layer, and the lack of Tokenizer Tuning. Collectively, these data serve as a reminder of the diversity across regions in how they address LRL challenges and the necessity of adjusting adaptation measures to the linguistic and computational requirements of the given district. This discussion can be considered informative regarding activities aimed at improving LLM performance in the use of underrepresented languages worldwide. Fig. 14 compares two metrics for four major low-resource languages: Breadth (number of papers) and Depth (average text length in the "Items/Scale" field).

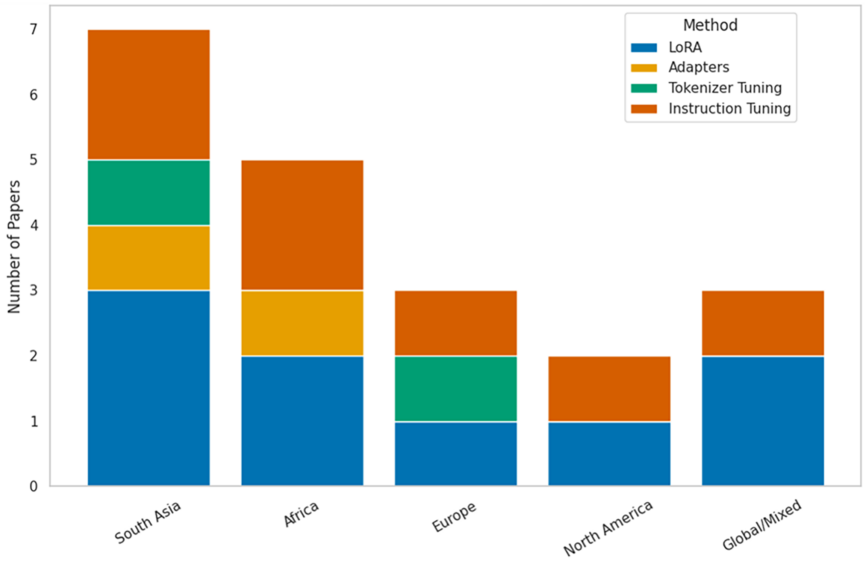

Figure 13: Adaptation techniques by region.

Fig. 14 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 may be less comprehensive than those for Hindi. The variety of methods established in the current review highlights the challenge of adapting LLMs to LRLs. Despite recent advancements, such as Mala-500 [73] and Language Fusion via Adapters [70], problems persist in scaling multilingual adaptation and ensuring ethical fairness. In particular, Bhattacharjee et al. emphasise the necessity of transformer adaptation to a global language [72], whereas Byambadorj et al. apply the cross-lingual transfer theory to text-to-speech systems with limited resources [76]. Fig. 15 shows the modalities studied in the literature (text, speech, multimodal, dialogue).

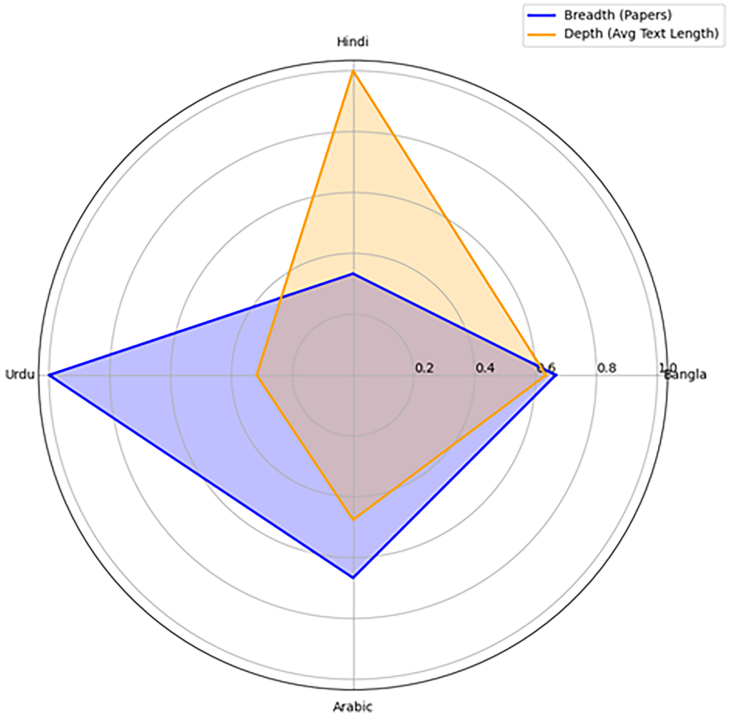

Figure 14: Breadth (study count) vs. depth (avg. dataset size) across language families.

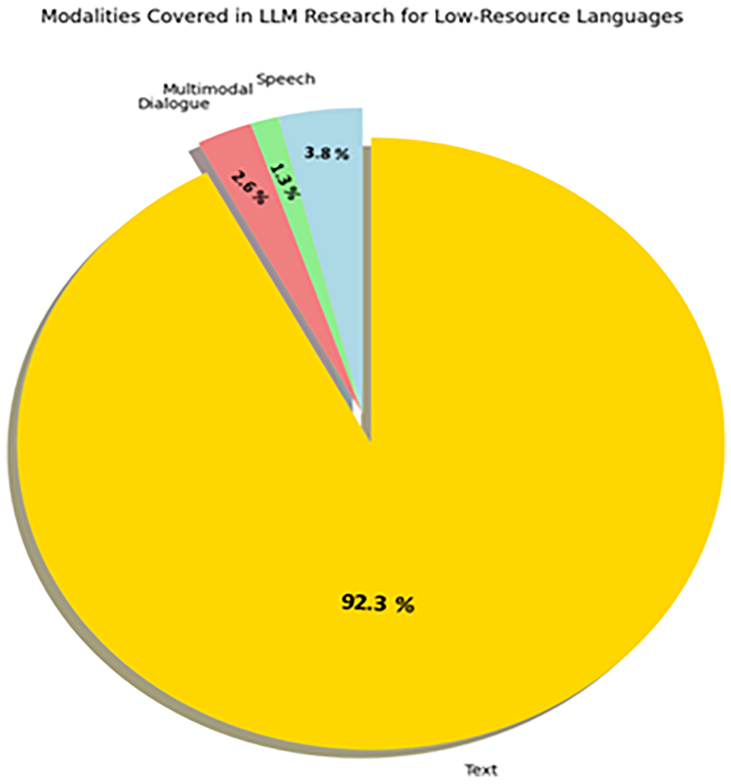

Figure 15: Modalities studied in the literature (text, speech, multimodal, dialogue).

There is also a regional imbalance in research focus. Africa [10], South Asian [13], and Pakistani [75] studies emphasise the importance of contextual solutions, whereas global projects, such as Glot500, focus on bridging the gap between communities with different languages in this area [7]. This review highlights the theoretical and practical developments in adapting LLMs for LRL applications, focusing on the value of benchmarks, efficient or cost-effective adaptation, and ethical concerns. Important advances include the creation of culturally based frameworks [9], creative fine-tuning strategies [19], and cross-lingual reference pattern transfer systems [77,78]. Despite these, there are still pending issues, such as scalability, resource, and ethical risks, that should be addressed in future research. By filling these gaps, the researchers will be able to provide even greater accessibility and influence of LLMs in linguistically underrepresented contexts.

4 Discussion

4.1 Data-Centric Adaptation and Pre-Training Strategies

Cross-cultural evaluations further demonstrate that LLM outputs may encode culturally implicit assumptions, leading to divergent interpretations and risks when models are deployed across sociolinguistically diverse communities [31]. Recent studies have led to a unified view that data quality and linguistic coverage are key factors in the performance of large language models in low-resource settings. LRL adaptation needed domain-specific, curated data pipelines that rely on in-house and synthetic data (e.g., small datasets curated in-house). In contrast to high-resource models, which rely on large and widely heterogeneous internet-scale corpora [1,79]. Continued pretraining [79,80] and new-language training [81] have been shown to be effective for maintaining linguistic coherence and improving transfer learning in 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. Toraman (2024) and Wang et al. (2025) demonstrate that adapting open-source LLMs such as LLaMA2 to specific target languages improves performance through continued pretraining and language-specific tuning, while highlighting practical challenges in language transfer [82,83]. Nahin et al. (2025) introduce TituLLMs, a family of Bangla LLMs with extensive benchmarking [84], and Wang et al. (2024) analyze how cross-lingual alignment naturally emerges during training, offering insights into the mechanisms that enable multilingual generalization [85]. Another example of an engineering approach to teaching LLMs new languages is Sambalingo, which targets language learning by adding specific alignment and a set of successive data [81].

Compensating for corpus scarcity in LRLs increasingly relies on synthetic data generation and parameter-efficient language adaptation techniques. Approaches such as language-adaptive pretraining and efficient transfer of pretrained models enable extension to new languages with limited resources [80,86]. In addition, synthetic dataset construction has been used to support reasoning and linguistic generalization in low-resource settings, although performance may vary depending on task design and data quality [33,87]. Retrieval-augmented prompting has proven particularly effective for low-resource machine translation by grounding generation in retrieved bilingual exemplars, significantly improving translation quality in extremely low-resource scenarios [88]. Complementary strategies such as soft prompt tuning further enhance retrieval-augmented frameworks by improving dense retrieval effectiveness within LLM-based systems [89]. Empirical evaluations in cross-lingual sentiment analysis and very low-resource translation tasks highlight the importance of careful adaptation strategies and benchmarking to ensure robust multilingual performance [90–92].

Despite recent advances, data imbalance across languages remains substantial. Efficient adaptation strategies for LRLs increasingly rely on structured corpus selection and graph-based propagation methods, which enable small multilingual models to achieve competitive performance without large-scale retraining [93]. Similarly, benchmarking efforts for underrepresented languages, such as Latvian, reveal that many regional languages remain insufficiently covered in current large language models, highlighting persistent evaluation and coverage gaps in multilingual NLP research [94]. Overall, the literature indicates that structured cross-lingual transfer and targeted corpus adaptation are emerging as practical strategies for extending multilingual coverage in resource-scarce settings. Table 3 summarizes the key data-centric pretraining and adaptation strategies identified across the reviewed studies.

Table 3: Summary of data-centric pretraining and adaptation strategies across reviewed studies (2020–2025).

Approach/Strategy	Representative Studies	Data Type & Source	Key Contributions	Observed Limitations
Continued Pretraining on multilingual corpora	Csaki et al. [81], Csaki et al. [80], Gurgurov et al. [93]	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)	Ghazaryan et al. [33], Subedi et al. [12]	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. [83]; Cahyawijaya et al. [95]	HRL → LRL transfer using multilingual encoders	Boosts typological generalization	Transfer bias from dominant languages
Community-curated LRL corpora	Adelani et al. [60], Skadiņa et al. [94]	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 major advances of PEFT is its ability to adapt large language models to low-resource settings under limited computational constraints. Methods such as LoRA, adapters, prefix tuning, and instruction tuning enable task- or language-specific specialization without updating the full parameter set, forming the basis of parameter-efficient adaptation strategies for multilingual models [70,96]. Recent work demonstrates that adapter-based and merging strategies can effectively facilitate cross-lingual transfer while maintaining computational efficiency [96]. In addition, seed-free synthetic instruction-tuning frameworks further reduce annotation costs in low-resource languages, as shown in Thai, enabling scalable adaptation pipelines [97]. Empirical studies of LoRA-based approaches report substantial reductions in GPU memory usage while achieving near-baseline performance in multilingual tasks [39,98].

However, comparative evidence suggests that the efficiency of PEFT methodologies depends heavily on language characteristics and the availability of training data. Cross-lingual adaptation strategies combining prompting, continued pretraining, and adapter-based interventions show varying effectiveness across languages with different typological properties and corpus sizes [99,100]. Language adapters in particular provide modular intervention points that enable controlled cross-lingual transfer and parameter reuse without full model retraining [100,101]. Empirical case studies further demonstrate that lightweight adaptation schemes can effectively extend large models to new languages. For example, LLaMA adaptation to Persian shows that structured fine-tuning and selective parameter updates enable competitive performance in previously unsupported languages [102]. Similarly, Turkish-focused adaptation efforts illustrate practical strategies for low-resource customization and benchmarking of generative models [82,103], and instruction-finetuned multilingual models such as Aya have demonstrated strong cross-lingual generalization across over

100 languages, including many low-resource ones [104]. Studies on morphologically rich and typologically diverse languages, including Ukrainian and Basque, indicate that structured adaptation and task-specific fine-tuning can stabilize performance in complex linguistic settings [37,41]. In contrast, instruction tuning and prompt-based transfer have been shown to support strong zero-shot generalization in unseen LRLs, particularly in cross-lingual NLI and transfer benchmarks [18,105,106]. Despite these advances, several surveys and benchmarking studies highlight the lack of consistent evaluation frameworks and benchmarks for low-resource multilingual NLP, which complicates systematic comparison of adaptation strategies, including PEFT approaches [1,26].

Kuulmets et al. (2024) demonstrate that cross-lingual knowledge transfer techniques can effectively teach LLaMA new low-resource languages by leveraging representations learned from high-resource languages, reducing the need for large-scale monolingual data [107]. Tao et al. (2024) further show that model-merging strategies can improve performance in low-resource languages by combining complementary capabilities of multiple models, offering an efficient alternative to full retraining or extensive fine-tuning [108]. These approaches reflect a broader trend toward lightweight adaptation strategies that reuse existing multilingual representations rather than training new models from scratch. Additional work explores techniques such as cross-lingual transfer mechanisms and prompting-based adaptation to extend multilingual capabilities in low-resource settings [109–113]. Table 4 presents a comparative overview of the performance trade-offs across these PEFT techniques.

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. [114], Miyano and Arase [19]	˜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	Gurgurov et al. [40], Schlenker et al. [115]	3%–5% additional params	Stable across morphologically rich languages	Modular and reusable	Larger memory footprint than LoRA
Prefix/prompt tuning	Singh et al. [99], Toukmaji and Flanigan [116]	0.1%–1%	Strong zero-shot transfer	Minimal training time	Sensitive to initialization
Instruction tuning	Kuulmets et al. [107]; Zhao et al. [96]	Variable	Highest generalization across unseen LRLs	Task flexibility; cultural adaptability	Requires curated instruction data
Model merging/active forgetting	Aggarwal et al. [111], Tao et al. [108]	—	Preserves multilingual alignment	Prevents overfitting	Implementation complexity

4.3 Tokenization, Script Diversity, and Typological Representation

Morphology-aware modeling approaches and language-specific adaptation strategies have been investigated to capture linguistic structure more effectively in low-resource settings, particularly for languages with complex inflectional or agglutinative systems [20,117]. These efforts highlight the importance of developing language-aware preprocessing and representation techniques to ensure that multilingual models can generalize across diverse linguistic environments.

Recent evaluation studies indicate that in-context learning can enable large language models to acquire capabilities in highly under-resourced languages while requiring minimal parameter updates [118–120]. Pivot-based feature conversion approaches also offer promising zero-resource translation strategies by projecting representations through an intermediate high-resource language, thereby facilitating cross-lingual transfer [121]. In addition, alignment techniques have been shown to enhance few-shot in-context learning, improving cross-lingual generalization in multilingual LLMs [122]. Comparative studies further explore combinations of prompting, translation, and instruction tuning strategies to improve domain-specific adaptation in low-resource languages [116].

Other research investigates embedding-level alignment strategies, including adapter-based interventions and cross-lingual instruction tuning, to improve semantic consistency across languages with different scripts and typological properties [123,124]. However, studies also emphasize that preprocessing and representation choices remain critical in multilingual modeling, as segmentation and representation errors at the subword level can propagate during downstream fine-tuning and adaptation [103]. Table 5 summarizes representative tokenization and representation strategies discussed in the literature, highlighting the importance of morphology-aware modeling for languages with diverse scripts and complex morphological structures.

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

Tokenization Framework	Core Studies	Supported Scripts	Linguistic Benefit	Drawbacks
BPE (Baseline)	Dargis et al. [69]	Latin, Cyrillic	Widely adopted, efficient	Breaks agglutinative morphemes
SentencePiece/ UnigramLM	Remy et al. [15]	Multiscript	Balanced vocabulary control	Weak for tonal systems
Byte-level encoding	Lin et al. [73]	Any UTF-8	Script-agnostic; reversible	Longer sequences; compute cost
Script-sensitive tokenizers	Rust et al. [117]	Bantu, Dravidian, Indic	Maintains morphological integrity	Requires linguistic annotation
Transliteration-based models	Zhuang et al. [125], Hong et al. [110]	Non-Latin → Latin	Enables cross-script sharing	Possible phoneme loss

4.4 Evaluation, Reproducibility, and Benchmark Disparities

Evaluation and reproducibility remain major challenges in research on LLMs for low-resource languages. Many studies rely on task-specific metrics such as BLEU, F1, perplexity, or accuracy, which makes cross-task and cross-language comparisons difficult and limits systematic assessment of multilingual model performance [1,2]. These inconsistencies are particularly evident in datasets for African and other underrepresented languages, where annotation quality and corpus granularity vary significantly across benchmarks [9,44]. Recent work has sought to improve multilingual evaluation by developing dedicated benchmark frameworks. Evaluation suites such as BUFFET and MEXA provide structured approaches for assessing cross-lingual transfer and alignment across diverse languages, while frameworks such as LEIA support cross-lingual knowledge transfer through structured data augmentation [46,126,127]. In addition,

reference-less evaluation methods have been proposed for extremely low-resource machine translation scenarios where gold-standard translations are unavailable, although such approaches also reveal limitations in LLM performance under severe data scarcity [128,129]. Further research explores methods for rapidly extending LLM capabilities to unseen languages. Techniques such as prompt-based or on-the-fly language learning demonstrate that models can bootstrap functional linguistic knowledge with minimal additional training data [130]. However, studies on language-specific LLM development emphasize that effective deployment in underrepresented languages still requires careful dataset design, evaluation protocols, and infrastructure support [131,132]. Investigations into extremely low-resource language adaptation also show that multilingual models can generalize to new linguistic environments, but their performance remains strongly dependent on the availability of even limited domain-specific data [133]. Overall, these findings highlight the ongoing need for standardized multilingual benchmarks, consistent evaluation protocols, and reproducible experimental pipelines to enable reliable comparison of adaptation strategies for low-resource languages. Table 6 provides a comparative summary of evaluation metrics and benchmark availability across the surveyed LRL studies.

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	Lin et al. [73], Nag et al. [79]	BLEU, ChrF	200 languages/1500 pairs	Limited low-resource test sets
MasakhaNER/ BLEnD	Myung et al. [45], Adelani et al. [60]	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 [17], Purwarianti et al. [44]	Bias, Fairness Scores	Basque, Indonesian	Limited benchmarking tools
Cross-lingual Embedding Analysis	Wen-Yi and Mimno [134]	Token Embedding Similarity	Multilingual Typologies	Absent standardized pipeline

4.5 Cultural Fairness, Accessibility, and Ethical Adaptation

Beyond technical optimization, it is increasingly recognized that cultural awareness and ethical reflexivity must be integrated into the adaptation of LLMs for LRLs. Researchers have highlighted that English-dominant training corpora risk reinforcing linguistic hegemony and representational bias, particularly in socio-culturally sensitive contexts [10,22]. To address these concerns, ethnographically informed frameworks emphasize community-centered data governance, participatory evaluation processes, and locally grounded dataset development [21,22]. Several recent initiatives demonstrate the value of collaborative and culturally grounded model development. Projects such as NileChat and ELEVATE-ID illustrate how partnerships with local communities can support the creation of linguistically appropriate and culturally meaningful AI applications [57,135–137]. Similarly, multilingual multi-channel modeling approaches have shown improvements in tasks such as cross-lingual hate-speech detection, highlighting the potential societal benefits of culturally aware LLM development [137].

However, research also shows that ethical alignment does not transfer uniformly across languages. Studies on multilingual value alignment indicate that moral and cultural concepts vary across linguistic contexts, emphasizing the importance of locale-specific evaluation of model behavior [14,138]. To mitigate risks such as toxicity, bias, and harmful outputs, inference-time optimization techniques for detoxification and debiasing have been proposed as practical complements to traditional fine-tuning approaches, especially in low-resource or sensitive linguistic settings where retraining large models may be impractical [42]. In addition, prompt-based translation strategies demonstrate that carefully designed prompts can produce competitive performance in low-resource translation scenarios without extensive parameter updates [139].

Cultural adaptation is also closely connected to infrastructural accessibility. Empirical studies across African, South Asian, and Southeast Asian contexts suggest that parameter-efficient techniques, such as LoRA-based adaptation combined with instruction tuning, can reduce computational barriers, enabling institutions with limited resources to deploy multilingual models more effectively [13,19,62]. Nevertheless, significant ethical risks remain. Concerns about bias, misinformation propagation, and hallucination persist in multilingual settings, motivating the development of dedicated frameworks for hallucination detection and responsible deployment of LLMs in underrepresented languages [14,56,57,60,140]. Overall, current research highlights a growing emphasis on fairness auditing, cultural representation, and participatory data governance in multilingual AI development. As summarized in Table 7, ethical adaptation frameworks increasingly complement technical strategies, underscoring the need for LLM systems that are not only computationally efficient but also socially responsible and culturally grounded.

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	Adelani et al. [9], Shang et al. [59]	Cultural representation & bias audit	Community-sourced evaluation	Improves inclusivity in African contexts
NileChat/ ELEVATE-ID	Mekki et al. [57], Gusmita et al. [136]	Participatory data governance	Local co-creation of datasets	Enhances authenticity & trust
BLEnD Fairness Protocol	Myung et al. [45]	Bias detection & fairness metrics	Embedded during fine-tuning	Promotes balanced representation
Low-Compute Democratic Adaptation	Mahfuz et al. [131], Subedi et al. [12]	Accessibility & inclusion	LoRA + instruction tuning in low-resource setups	Democratizes LLM deployment

4.6 Findings and Theoretical Implications Synthesis

In this review, the overall certainty of evidence across the synthesized studies was assessed qualitatively due to the heterogeneity of study designs. Integrating findings across the examined domains reveals three major patterns. First, recent advances in cross-lingual transfer and parameter-efficient adaptation provide an important foundation for scalable multilingual applications, including automated assessment systems. Studies on semantic-aware transfer and graph-based prompting demonstrate that multilingual and low-resource language models can achieve strong cross-lingual generalization with reduced reliance

on language-specific representations, supporting more equitable performance across diverse linguistic settings [141,142]. At the same time, negative findings suggest that language-specific neurons do not consistently enhance cross-lingual transfer, reinforcing the importance of shared semantic representations for multilingual scalability [143]. Probabilistic and graph-driven prompting methods further extend these capabilities to unsupervised low-resource translation scenarios, highlighting the feasibility of multilingual generalization under minimal supervision. Building on these modeling advances, recent work in educational measurement demonstrates that automated scoring systems can leverage multilingual robustness to provide scalable, quality-controlled support for constructed-response assessment in large-scale international evaluations, reducing dependence on extensive human rater recruitment across language versions [144–146].

Second, equitable adaptation has increasingly been operationalized through parameter-efficient fine-tuning (PEFT), which enables scalable model customization without the high computational costs associated with full model retraining [70,98].

Third, the field is gradually adopting a more decolonial and community-centered perspective, in which linguistic justice and digital inclusion are considered integral components of technical design and deployment [10,22].

Taken together, these findings confirm that (RQ1) pre-training and adaptation for low-resource languages are most effective when multilingual alignment is combined with careful corpus curation; (RQ2) PEFT methods balance efficiency and performance under limited computational resources; and (RQ3) equitable LLM adaptation requires culturally sensitive and reproducible evaluation frameworks. Collectively, these insights outline a multi-level roadmap for advancing typology-aware, ethically grounded, and resource-efficient LLM adaptation strategies in the coming decade (Table 8).

Table 8: Comparative charts summarizing findings across studies.

Dimension	Representative Approaches	Synthesized Findings Across Studies	Key Strengths	Key Limitations
Data strategies	Synthetic data generation, cross-lingual transfer, community-curated corpora	Synthetic augmentation and typology-aligned transfer consistently improve LRL performance, particularly in translation and NER, while community-curated datasets enhance cultural validity	Scalable data expansion; improved linguistic coverage	Noise amplification; uneven representation of minority languages
Adaptation methods	LoRA, adapters, instruction tuning, continued pretraining	Parameter-efficient fine-tuning achieves 95%–98% of full fine-tuning performance with significantly reduced computational cost across diverse LRLs	Compute-efficient; suitable for low-resource settings	Performance varies by language typology and task
Tokenization & scripts	Script-sensitive tokenizers, byte-level encoding, transliteration-based models	Morphology-aware and script-sensitive tokenization improves semantic integrity in non-Latin and agglutinative languages, enhancing downstream task performance	Better representation of morphologically rich languages	Increased preprocessing and vocabulary design complexity

(Continued)

Table 8 (continued)

Dimension	Representative Approaches	Synthesized Findings Across Studies	Key Strengths	Key Limitations
Evaluation & benchmarks	FLORES-200, MasakhaNER, BLEnD, IrokoBench	Culturally grounded benchmarks reveal performance gaps and biases overlooked by generic metrics, especially in African and Indic languages	Fairer and context-aware evaluation	Fragmentation limits cross-study comparability
Ethical considerations	Bias audits, participatory data governance, fairness-aware fine-tuning	Fine-tuning on biased or synthetic corpora propagates cultural and social distortions unless mitigated by localized data and evaluation frameworks	Improved cultural fairness and trust	Ethical practices remain inconsistently operationalized

5 Conclusion

The present systematic review represents an integrated overview of the current developments in the adaptation of LLMs to LRLs that are dedicated to data-centric pretraining, parameter-efficient fine-tuning, tokenization, and ethical issues. Results from 2020–2025 indicate that data augmentation, multilingual transfer, and typology-trained pretraining are highly beneficial for covering language, whereas parameter-efficient techniques, including LoRA, adapters, and instruction tuning, are especially effective at producing 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 participatory data governance, are essential to establishing inclusivity and linguistic justice in AI development. All these findings together substantiate the claim that the equitable implementation of LLMs entails a dual commitment to both technical scalability and sociocultural contextualization.

To build high-quality, LRL datasets, it is possible to use community-curated corpora, participatory annotation, cross-linguistic transfer to typologically related languages, and limited synthetic data built on authentic linguistic material.

Recent works, such as MedUniC and G2D, reduce bias by aligning vision-language representations across languages, addressing modality imbalance and data imbalance, and pretraining multilingual large language models more fairly, which are biased toward medicine and radiography.

PEFT techniques like LoRA tend to work well; however, morphologically complex or typologically distant languages, which may include low-resource languages, often require extensive or hybrid fine-tuning to learn language subtleties that low-rank adaptation poorly captures.

To guarantee the reproducibility of benchmarks, open-access datasets, annotation guidelines, transparent documentation, and culture-based evaluation structures are required, even for non-public or inconsistently annotated low-resource language collections.

Future studies should pursue hybrid adaptation pipelines that combine synthetic corpus generation, cross-lingual continual learning, and parameter-efficient fine-tuning across a wider range of typologically diverse LRLs. An increased focus on multilingual benchmarks, based on open-source data and curated by the local community, will encourage transparency and inclusiveness. Furthermore, within the scope of

interdisciplinary cooperation, computational linguists, ethicists, and local stakeholders should collaborate to ensure cultural fairness and the development of sustainable AI ecosystems. Possibly the most crucial step towards narrowing the global gap in LLM development and ensuring truly equitable outcomes will be advancing reproducible evaluation schemes and democratizing access to multilingual resources.

Acknowledgement: The authors would like to thank Multimedia University, Cyberjaya, Selangor, Malaysia, for supporting this research.

Funding Statement: This research was funded by Multimedia University, Cyberjaya, Selangor, Malaysia (Grant Number: PostDoc MMUI/240029).

Author Contributions: Conceptualization, Ismail Hossain and Mridul Banik; formal analysis/investigation, Ismail Hossain, Mridul Banik and Fahmid Al Farid; writing—original draft preparation, Ismail Hossain, Mridul Banik, Fahmid Al Farid and Jia Uddin; writing—review and editing, Jia Uddin and Hezerul bin Abdul Karim; supervision, Jia Uddin and Hezerul bin Abdul Karim; funding acquisition, Hezerul bin Abdul Karim. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Not applicable.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Supplementary Materials: The supplementary material is available online at https://www.techscience.com/doi/10.32604/cmes.2026.075507/s1. The PRISMA checklists are available in the supplementary file.

Appendix A

Example of: Search Strategy (Google Scholar)

("large language model" OR "large language models" OR LLM OR GPT OR "foundation model")

AND ("low-resource language" OR "low resource language" OR "under-resourced language" OR "minority language" OR "indigenous language")

AND (pretraining OR "continued pretraining" OR adaptation OR fine-tuning OR "parameter-efficient fine-tuning" OR PEFT OR LoRA OR tokenization OR "cross-lingual transfer")

Filters applied:

- Publication years: 2020–2025
- Language: English
- Document types: journal articles, conference papers, and reputable preprints

Example of: Search Strategy (ACL Anthology)

("large language models" OR LLM OR GPT OR LLaMA) AND ("low-resource languages" OR "under-resourced languages") AND (fine-tuning OR adaptation OR pretraining OR LoRA OR PEFT OR "instruction tuning" OR tokenization)

Filters applied:

- Years: 2020–2025
- Peer-reviewed conference and workshop papers

Example of: Search Strategy (arXiv)

("large language models" OR LLM) AND ("low-resource languages" OR "under-resourced languages") AND (pretraining OR adaptation OR "parameter-efficient fine-tuning" OR LoRA OR "instruction tuning")

Filters applied:

- Categories: cs.CL, cs.AI
- Years: 2020–2025

References

1. 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.
2. Asgari R, Moradi M, Yan K, Colwell D, Samwald M. A critical review of methods and challenges in large language models. Comput Mater Contin. 2025;82(2):1681–98. doi:10.32604/cmc.2025.061263.
3. Gain B, Bandyopadhyay D, Ekbal A. Bridging the linguistic divide: a survey on leveraging LLMs for machine translation. arXiv:2504.01919. 2025. doi:10.48550/arXiv.2504.01919.
4. Cassano F, Gouwar J, Lucchetti F, Schlesinger C, Freeman A, Anderson CJ, et al. Knowledge transfer from high-resource to low-resource programming languages for code LLMs. Proc ACM Program Lang. 2024;8(OOPSLA2):677–708. doi:10.1145/3689735.
5. Joshi P, Santy S, Budhiraja A, Bali K, Choudhury M. The state and fate of linguistic diversity and inclusion in the NLP world. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics; 2020 Jul 5–10; Online. p. 6282–93. doi:10.18653/v1/2020.acl-main.560.
6. Pava J, Meinhardt C, Uz Zaman H, Friedman T, Truong S, Zhang D, et al. Mind the (Language) Gap: mapping the challenges of LLM development in low-resource language contexts; 2025 [cited 2026 Jan 1]. Available from: https://hai.stanford.edu/policy/mind-the-language-gap-mapping-the-challenges-of-llm-development-in-low-resource-language-contexts.
7. ImaniGooghari A, Lin P, Kargaran AH, Severini S, Jalili Sabet M, Kassner N, et al. Glot500: scaling multilingual corpora and language models to 500 languages. arXiv:2305.12182. 2023.
8. Nguyen XP, Aljunied M, Joty S, Bing L. Democratizing LLMs for low-resource languages by leveraging their English dominant abilities with linguistically-diverse prompts. In: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics; 2024 Aug 11–15; Bangkok, Thailand. Stroudsburg, PA, USA: ACL; 2024. doi:10.18653/v1/2024.acl-long.192.
9. 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; 2025 Apr 29–May 4; Albuquerque, New Mexico. Stroudsburg, PA, USA: ACL; 2025. p. 2732–57. doi:10.18653/v1/2025.naacl-long.139.
10. Alhanai T, Kasumovic A, Ghassemi MM, Zitzelberger A, Lundin JM, Chabot-Couture G. Bridging the gap: enhancing LLM performance for low-resource African languages with new benchmarks, fine-tuning, and cultural adjustments. Proc AAAI Conf Artif Intell. 2025;39(27):27802–12. doi:10.1609/aaai.v39i27.34996.
11. Alam F, Chowdhury SA, Boughorbel S, Hasanain M. LLMs for low resource languages in multilingual, multimodal and dialectal settings. In: Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: Tutorial Abstracts; 2024 Mar 21; St. Julian's, Malta. Stroudsburg, PA, USA: ACL; 2024. p. 27–33. doi:10.18653/v1/2024.eacl-tutorials.5.
12. Subedi B, Regmi S, Bal B, Acharya P. Exploring the potential of LLMs for low-resource languages: a study on named-entity recognition (NER) and part-of-speech (POS) tagging for Nepali language. In: Proceedings of the Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024); 2024 May 20–25; Torino, Italia. p. 6974–9.

13. 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 2025); 2025 Jan 19; Abu Dhabi, UAE. p. 217–22.
14. Aksoy M. Whose morality do they speak? Unraveling cultural bias in multilingual language models. Nat Lang Process J. 2025;12(44):100172. doi:10.1016/j.nlp.2025.100172.
15. Remy F, Delobelle P, Avetisyan H, Khabibullina A, de Lhoneux M, Demeester T. Trans-tokenization and cross-lingual vocabulary transfers: language adaptation of LLMs for low-resource NLP. arXiv:2408.04303. 2024.
16. Corral A, Antero IS, Saralegi X. Pipeline analysis for developing instruct LLMs in low-resource languages: a case study on Basque. In: Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies; 2025 Apr 29–May 4; Albuquerque, New Mexico. Stroudsburg, PA, USA: ACL; 2025. p. 12636–55. doi:10.18653/v1/2025.naacl-long.629.
17. Saralegi X, Zulaika M. BasqBBQ: a QA benchmark for assessing social biases in LLMs for Basque, a low-resource language. In: Proceedings of the 31st International Conference on Computational Linguistics; 2025 Jan 19–24; Dhabi, UAE. p. 4753–67.
18. Chirkova N, Nikoulina V. Zero-shot cross-lingual transfer in instruction tuning of large language models. In: Proceedings of the 17th International Natural Language Generation Conference; 2024 Sep 23–27; Tokyo, Japan. Stroudsburg, PA, USA: ACL; 2024. p. 695–708. doi:10.18653/v1/2024.inlg-main.53.
19. Miyano R, Arase Y. Adaptive LoRA merge with parameter pruning for low-resource generation. In: Proceedings of the Findings of the Association for Computational Linguistics: ACL 2025; 2025 Jul 27–Aug 1; Vienna, Austria. Stroudsburg, PA, USA: ACL; 2025. p. 19353–66. doi:10.18653/v1/2025.findings-acl.990.
20. Avetisyan H, Broneske D. VerbCraft: morphologically-aware Armenian text generation using LLMs in low-resource settings. In: Proceedings of the 3rd Workshop on Resources and Representations for Under-Resourced Languages and Domains (RESOURCEFUL-2025); 2025 Mar 2; Tallinn, Estonia. p. 111–9.
21. Nekoto W, Marivate V, Matsila T, Fasubaa T, Fagbohungbe T, Akinola SO, et al. Participatory research for low-resourced machine translation: a case study in African languages. In: Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2020; 2020 Nov 16–20; Online. Stroudsburg, PA, USA: ACL; 2020. p. 2144–60. doi:10.18653/v1/2020.findings-emnlp.195.
22. De Klerk M, McLean N. Defining a technocritical approach to AI adoption in the Global South: perspectives from higher education. Eduvos Res. 2024 [cited 2026 Jan 1]. Available from: https://www.eduvos.com/researchpaper.pdf.
23. McGiff J, Nikolov N. Overcoming data scarcity in generative language modelling for low-resource languages: a systematic review. arXiv:2505.04531. 2025.
24. Peng C, Ma Z. A review of machine translation techniques for low-resource languages. J Lit Art Stud. 2025;15(9):725–31. doi:10.18653/v1/2025.loresmt-1.12.
25. Tafa TO, Hashim SZM, Othman MS, Alhussian H, Nasser M, Abdulkadir SJ, et al. Machine translation performance for low-resource languages: a systematic literature review. IEEE Access. 2025;13(4):72486–505. doi:10.1109/access.2025.3562918.
26. Liu S, Best M. A survey of NLP progress in Sino-Tibetan 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 May 3–10; Albuquerque, New Mexico. Stroudsburg, PA, USA: ACL; 2025. p. 7804–25. doi:10.18653/v1/2025.naacl-long.396.
27. Izquierdo Domenech J, Linares Pellicer J, Ferri Molla I. Virtual reality and language models, a new frontier in learning. Int J Interact Multimed Artif Intell. 2024;8(5):46–54. doi:10.9781/ijimai.2024.02.007.
28. 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.
29. 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; 2025 Jan 20; Abu Dhabi, UAE. p. 50–7.

30. Researcher I, Kalluri K. Adapting LLMs for low resource languages-techniques and ethical considerations. Interantional J Sci Res Eng Manag. 2024;8(12):1–6. doi:10.55041/isjem00140.
31. Shan X, Xu Y, Wang Y, Lin YS, Bao Y. Cross-cultural implications of Large language models: an extended comparative analysis. In: HCI International 2024-Late breaking papers. Cham, Switzerland: Springer; 2025. p. 106–18. doi:10.1007/978-3-031-76806-4_8.
32. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA, 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev. 2021;10(1):89. doi:10.1186/s13643-021-01626-4.
33. Ghazaryan G, Arakelyan E, Minervini P, Augenstein I. SynDARin: synthesising datasets for automated reasoning in low-resource languages. In: Proceedings of the 31st International Conference on Computational Linguistics; 2025 Jan 19–24; Abu Dhabi, UAE. p. 6459–66.
34. Li Z, Zhu H, Lu Z, Yin M. Synthetic data generation with large language models for text classification: potential and limitations. In: Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing; 2023 Dec 6–10; Singapore. Stroudsburg, PA, USA: ACL; 2023. p. 10443–61. doi:10.18653/v1/2023.emnlp-main.647.
35. Bhadauria D, Sierra Munera A, Krestel R. The effects of data quality on named entity recognition. In: Proceedings of the Ninth Workshop on Noisy and User-generated Text (W-NUT 2024); 2024 Nov 16; Miami, FL, USA. p. 79–88.
36. Tessema B, Kedia A, Chung T. UnifiedCrawl: aggregated Common Crawl for affordable adaptation of LLMs on low-resource languages. arXiv:2411.14343. 2024. doi:10.48550/arXiv.2411.14343.
37. Etxaniz J, Sainz O, Miguel 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 Aug 11–16; Bangkok, Thailand. Stroudsburg, PA, USA: ACL; 2024. p. 14952–72. doi:10.18653/v1/2024.acl-long.799.
38. Šliogeris V, Daniušis P, Nakvosas A. Full-parameter continual pretraining of Gemma2: insights into fluency and domain knowledge. arXiv:2505.05946. 2025.
39. Le T, Nguyen T, Nguyen Ha V, Chatzinotas S, Jouvet P, Noumeir R. The impact of LoRA adapters for LLMs on clinical NLP classification under data limitations. arXiv:2407.19299v2. 2025.
40. 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 Aug 15; Bangkok, Thailand. Stroudsburg, PA, USA: ACL; 2024. p. 63–74. doi:10.18653/v1/2024.kallm-1.7.
41. Radchenko V, Drushchak N. Improving named entity recognition for low-resource languages using large language models: a Ukrainian case study. In: Proceedings of the Fourth Ukrainian Natural Language Processing Workshop (UNLP 2025); 2025 Jul 31–Aug 1. Online. Stroudsburg, PA, USA: ACL; 2025. p. 27–35. doi:10.18653/v1/2025.unlp-1.3.
42. Yang Z, Yi X, Li P, Liu Y, Xie X. Unified detoxifying and debiasing in language generation via inference-time adaptive optimization. arXiv:2210.04492. 2022.
43. Neplenbroek V, Bisazza A, Fernández R. Cross-lingual transfer of debiasing and detoxification in multilingual LLMs: an extensive investigation. In: Proceedings of the Findings of the Association for Computational Linguistics: ACL 2025; 2025 Jul 27–Aug 1; Vienna, Austria. Stroudsburg, PA, USA: ACL; 2025. p. 2805–30. doi:10.18653/v1/2025.findings-acl.145.
44. 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. In: Proceedings of the 2nd Workshop in South East Asian Language Processing; 2025 Jan 19; Abu Dhabi, UAE. p. 82–100.
45. Myung J, Lee N, Zhou Y, Jin J, Putri R, Antypas D, et al. BLEnD: a benchmark for LLMs on everyday knowledge in diverse cultures and languages. arXiv:2406.09948. 2024.
46. Yamada I, Ri R. LEIA: facilitating cross-lingual knowledge transfer in language models with entity-based data augmentation. In: Proceedings of the Findings of the Association for Computational Linguistics ACL 2024; 2024 Aug 11–16; Bangkok, Thailand. Stroudsburg, PA, USA: ACL; 2024. p. 7029–39. doi:10.18653/v1/2024.findings-acl.419.

47. Zhang C, Liao Z, Feng Y. Cross-lingual transfer of cultural knowledge: an asymmetric phenomenon. In: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics; 2025 Jul 27–Aug 1; Vienna, Austria. Stroudsburg, PA, USA: ACL; 2025. p. 147–57. doi:10.18653/v1/2025.acl-short.13.
48. Zhuang W, Sun Y. CUTE: a multilingual dataset for enhancing cross-lingual knowledge transfer in low-resource languages. In: Proceedings of the 31st International Conference on Computational Linguistics (COLING 2025); 2025 Jan 19–24. Abu Dhabi, UAE; 2025. p. 10037–46.
49. Guo P, Ren Y, Hu Y, Li Y, Zhang J, Zhang X, et al. Teaching LLMs to translate on low-resource languages with textbook prompting. In: Proceedings of the Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024); 2024 May 20–25; Torino, Italy. p. 15685–97.
50. Jiao W, Huang JT, Wang W, He Z, Liang T, Wang X, et al. ParroT: translating during chat using large language models tuned with human translation and feedback. In: Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2023; 2023 Dec 6–10; Singapore. Stroudsburg, PA, USA: ACL; 2023. p. 15009–20. doi:10.18653/v1/2023.findings-emnlp.1001.
51. Mao Z, Yu Y. Tuning LLMs with contrastive alignment instructions for machine translation in unseen, low-resource languages. In: Proceedings of the Seventh Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2024); 2024 Aug 16; Bangkok, Thailand. Stroudsburg, PA, USA: ACL; 2024. p. 1–25. doi:10.18653/v1/2024.loresmt-1.1.
52. 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. In: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics; 2025 Jul 27–Aug 1; Vienna, Austria. Stroudsburg, PA, USA: ACL; 2025. p. 3977–97. doi:10.18653/v1/2025.acl-long.202.
53. Huang D, Ng D, Wang Z, Pen H, Cambria E. Evaluating the impact of LLM-manipulated content on fake news detection. In: Trends and applications in knowledge discovery and data mining. Singapore: Springer; 2025. p. 375–86. doi:10.1007/978-981-96-8197-6_28.
54. Shibu H, Datta S, Miah M, Sami N, Chowdhury M, Islam M, et al. From scarcity to capability: empowering fake news detection in low-resource languages with LLMs. In: Proceedings of the First Workshop on Natural Language Processing for Indo-Aryan and Dravidian Languages; 2025 Jan 20. Abu Dhabi, UAE; 2025. p. 100–7.
55. Song J, Huang Y, Zhou Z, Ma L. Multilingual blending: large language model safety alignment evaluation with language mixture. In: Proceedings of the Findings of the Association for Computational Linguistics: NAACL 2025; 2025 Apr 29–May 4; Albuquerque, NM, USA. Stroudsburg, PA, USA: ACL; 2025. p. 3433–49. doi:10.18653/v1/2025.findings-naacl.191.
56. 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. In: Proceedings of the Findings of the Association for Computational Linguistics ACL 2024; 2024 Aug 11–16; Bangkok, Thailand. Stroudsburg, PA, USA: ACL; 2024. p. 2668–80. doi:10.18653/v1/2024.findings-acl.156.
57. Mekki A, Atou H, Nacar O, Shehata S, Abdul-Mageed M. NileChat: towards linguistically diverse and culturally aware LLMs for local communities. arXiv:2505.18383. 2025.
58. Park G, Hwang S, Lee H. Low-resource cross-lingual summarization through few-shot learning with large language models. In: Proceedings of the Seventh Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2024); 2024 Aug 16; Bangkok, Thailand. Stroudsburg, PA, USA: ACL; 2024. p. 57–63. doi:10.18653/v1/2024.loresmt-1.6.
59. Shang G, Abdine H, Khoubrane Y, Mohamed A, Abbahaddou Y, Ennadir S, et al. Atlas-Chat: adapting LLMs for low-resource Moroccan Arabic dialect. arXiv:2409.17912. 2024.
60. Adelani DI, Doğruöz AS, Coneglian A, Ojha AK. Comparing LLM prompting with Cross-lingual transfer performance on indigenous and low-resource Brazilian languages. In: Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024); 2024 Jun 20; Mexico City, Mexico. Stroudsburg, PA, USA: ACL; 2024. p. 34–41. doi:10.18653/v1/2024.americasnlp-1.5.
61. Upadhayay B, Behzadan V. Taco: enhancing cross-lingual transfer for low-resource languages in LLMs through translation-assisted chain-of-thought processes. arXiv:2311.10797. 2023.

62. Fiaz L, Tahir M, Shams S, Hussain S. UrduLLaMA 1.0: dataset curation, preprocessing, and evaluation in low-resource settings. arXiv:2502.16961. 2025.
63. Scialom T, Chakrabarty T, Muresan S. Fine-tuned language models are continual learners. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing; 2022 Dec 7–11; Abu Dhabi, UAE. Stroudsburg, PA, USA: ACL; 2022. p. 6288–301. doi:10.18653/v1/2022.emnlp-main.410.
64. 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. 2026;58(5):1–42. doi:10.1145/3735633.
65. Liu D, Niehues J. Middle-layer representation alignment for cross-lingual transfer in fine-tuned LLMs. In: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics; 2025 Jul 27–Aug 1; Vienna, Austria. Stroudsburg, PA, USA: ACL; 2025. p. 15979–96. doi:10.18653/v1/2025.acl-long.778.
66. Yoo H, Park C, Yun S, Oh A, Lee H. Code-switching curriculum learning for multilingual transfer in LLMs. In: Proceedings of the Findings of the Association for Computational Linguistics: ACL 2025; 2025 Jul 27–Aug 1; Vienna, Austria. Stroudsburg, PA, USA: ACL; 2025. p. 7816–36. doi:10.18653/v1/2025.findings-acl.407.
67. Zhong T, Yang Z, Liu Z, Zhang R, Liu Y, Sun H, et al. Opportunities and challenges of LLMs for low-resource languages in humanities research. arXiv:2412.04497. 2024.
68. Upadhayay B, Behzadan V. Tongue-tied: breaking LLMs safety through new language learning. In: Proceedings of the 7th Workshop on Computational Approaches to Linguistic Code-Switching; 2025 May 3; Albuquerque, NM, USA. Stroudsburg, PA, USA: ACL; 2025. p. 32–47. doi:10.18653/v1/2025.calcs-1.5.
69. Darģis R, Bārzdiņš G, Skadiņa I, Saulite 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 Nov 15; Miami, FL, USA. Stroudsburg, PA, USA: ACL; 2024. p. 289–93. doi:10.18653/v1/2024.nlp4dh-1.28.
70. 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.
71. Verma S, Rahman Khan M, Kumar V, Murthy R, Sen J. MILU: a multi-task Indic language understanding benchmark. arXiv:2411.02538. 2024.
72. Bhattacharjee A, Hasan T, Ahmad W, Samin K, Islam M, Iqbal A, et al. BanglaBERT: language model pretraining and benchmarks for low-resource language understanding evaluation in Bangla. arXiv:2101.00204. 2022.
73. Lin P, Ji S, Tiedemann J, Martins A, Schütze H. Mala-500: massive language adaptation of large language models. arXiv:2401.13303. 2024.
74. 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.
75. Tahir M, 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 2025); 2025 May 3–4; Albuquerque, NM, USA. p. 17–34.
76. Byambadorj Z, Nishimura R, Ayush A, Ohta K, Kitaoka N. Text-to-speech system for low-resource language using cross-lingual transfer learning and data augmentation. EURASIP J Audio Speech Music Process. 2021;2021(1):42. doi:10.1186/s13636-021-00225-4.
77. Ngugi S. Targeted lexical injection: unlocking latent cross-lingual alignment in lugha-llama via early-layer LoRA fine-tuning. arXiv:2506.15415. 2025.
78. Zhang Y, Chen N. PPT: a minor language news recommendation model via cross-lingual preference pattern transfer. In: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics; 2025 Jul 27–Aug 1; Vienna, Austria. Stroudsburg, PA, USA: ACL; 2025. p. 10728–45. doi:10.18653/v1/2025.acl-long.526.
79. 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 Apr 29–May 4; Albuquerque, NM, USA. Stroudsburg, PA, USA: ACL; 2025. p. 304–17. doi:10.18653/v1/2025.naacl-industry.25.

80. Csaki Z, Pawakapan P, Thakker U, Xu Q. Efficiently adapting pretrained language models to new languages. arXiv:2311.05741. 2023.
81. Csaki Z, Li B, Li JL, 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 Nov 16; Miami, FL, USA. Stroudsburg, PA, USA: ACL; 2024. p. 1–21. doi:10.18653/v1/2024.mrl-1.1.
82. 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); 2024 Nov 16; Miami, FL, USA. Stroudsburg, PA, USA: ACL; 2024. p. 30–44. doi:10.18653/v1/2024.mrl-1.3.
83. 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; 2025 Jan 19–24; Abu Dhabi, UAE. p. 7195–208.
84. Nahin SK, Nandi RN, Sarker S, Muhtaseem QS, Kowsher M, Shill AC, et al. TituLLMs: a family of bangla LLMs with comprehensive benchmarking. In: Proceedings of the Findings of the Association for Computational Linguistics: ACL 2025; 2025 Jul 27–Aug 1; Vienna, Austria. Stroudsburg, PA, USA: ACL; 2025. p. 24922–40. doi:10.18653/v1/2025.findings-acl.1279.
85. Wang H, Minervini P, Ponti E. Probing the emergence of cross-lingual alignment during LLM training. In: Proceedings of the Findings of the Association for Computational Linguistics ACL 2024; 2024 Aug 11–16; Bangkok, Thailand. Stroudsburg, PA, USA: ACL; 2024. p. 12159–73. doi:10.18653/v1/2024.findings-acl.724.
86. Iyer V, Malik B, Zhu W, Stepachev P, Chen P, Haddow B, et al. Exploring very low-resource translation with LLMs: the University of Edinburgh's submission to AmericasNLP 2024 translation task. In: Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024); 2024 Jun 17; Mexico City, Mexico. p. 209–20.
87. Zhu H, Liang Y, Xu W, Xu H. Evaluating LLMs for in-context learning of linguistic patterns in unseen low-resource languages. In: Proceedings of the 1st Workshop on Language Models for Low-Resource Languages; 2025 Jan 19–20; Abu Dhabi, UAE. p. 414–26.
88. Merx R, Mahmudi A, Langford K, de Araujo L, Vylomova E. Low-resource machine translation through retrieval-augmented LLM prompting: a study on the Mambai language. In: Proceedings of the 2nd Workshop on Resources and Technologies for Indigenous, Endangered and Lesser-resourced Languages in Eurasia (EURALI) @ LREC-COLING 2024; 2024 May 25; Torino, Italy. p. 1–11.
89. Peng Z, Wu X, Wang Q, Fang Y. Soft prompt tuning for augmenting dense retrieval with large language models. Knowl Based Syst. 2025;309(4):112758. doi:10.1016/j.knosys.2024.112758.
90. 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. In: Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media Analysis; 2024 Aug 15–16; Bangkok, Thailand. p. 141–52.
91. Zhang W, Deng Y, Liu B, Pan S, Bing L. Sentiment analysis in the era of large language models: a reality check. In: Proceedings of the Findings of the Association for Computational Linguistics: NAACL 2024; 2024 Jun 16–21; Mexico City, Mexico. Stroudsburg, PA, USA: ACL; 2024. p. 3881–906. doi:10.18653/v1/2024.findings-naacl.246.
92. Lamin NZ, Aziz AA. Cross-lingual sentiment analysis in low-resource languages: a recent review on tasks, methods and challenges. Int J Adv Comput Sci Appl. 2025;16(11). doi:10.14569/ijacsa.2025.0161144.
93. 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 Jul 28–30; Vienna, Austria. Stroudsburg, PA, USA: ACL; 2025. p. 355–95. doi:10.18653/v1/2025.acl-srw.24.
94. Skadiņa I, Bakanovs B, Darģis R. First steps in benchmarking Latvian in large language models. In: Proceedings of the 3rd Workshop on Resources and Representations for Under-Resourced Languages and Domains (RESOURCEFUL-2025); 2025 Jan 19; Abu Dhabi, UAE. p. 86–95.

95. 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 Jun 16–21; Mexico City, Mexico. Stroudsburg, PA, USA: ACL; 2024. p. 405–33. doi:10.18653/v1/2024.naacl-long.24.
96. Zhao Y, Zhang W, Wang H, Kawaguchi K, Bing L. AdaMergeX: cross-lingual transfer with large language models via adaptive adapter merging. In: Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies; 2025 Apr 29–May 4; Albuquerque, NM, USA. Stroudsburg, PA, USA: ACL; 2025. p. 9785–800. doi:10.18653/v1/2025.naacl-long.493.
97. Pengpun P, Udomcharoenchaikit C, Buaphet W, Limkonchotiwat P. Seed-free synthetic data generation framework for instruction-tuning LLMs: a case study in Thai. In: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics; 2024 Aug 11–16; Bangkok, Thailand. p. 445–64.
98. Akimoto K, Oyamada M. Optimizing low-resource language model training: comprehensive analysis of multi-epoch, multi-lingual, and two-stage approaches. arXiv:2410.12325. 2024.
99. Singh V, Krishna A, Karthika N, Ramakrishnan G. A three-pronged approach to cross-lingual adaptation with multilingual LLMs. arXiv:2406.17377. 2024. doi:10.48550/arxiv.2406.17377.
100. Kunz J, Holmström O. The impact of language adapters in cross-lingual transfer for NLU. In: Proceedings of the 1st Workshop on Modular and Open Multilingual NLP (MOOMIN 2024); 2024 Mar 21; St Julian's, Malta. Stroudsburg, PA, USA: ACL; 2024. p. 24–43. doi:10.18653/v1/2024.moomin-1.4.
101. Pfeiffer J, Vuli I, Gurevych I, Ruder S. MAD-Xl an adapter-based framework for multi-task cross-lingual transfer. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP); 2020 Nov 16–20; Online. Stroudsburg, PA, USA: ACL; 2020. p. 7654–73. doi:10.18653/v1/2020.emnlp-main.617.
102. Sani S, Sadeghi P, Vu T, 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 Jan 19–24; Abu Dhabi, UAE. p. 8868–84.
103. Acikgoz E, Erdogan M, Yuret D. Bridging the Bosphorus: advancing Turkish LLMs through strategies for low-resource language adaptation and benchmarking. arXiv:2405.04685. 2024.
104. Üstün A, Aryabumi V, Yong Z, Ko WY, D'souza D, Onilude G, et al. Aya model: an instruction finetuned open-access multilingual language model. In: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics; 2024 Aug 11–16; Bangkok, Thailand. Stroudsburg, PA, USA: ACL; 2024. p. 15894–939. doi:10.18653/v1/2024.acl-long.845.
105. 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. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics; 2022 May 22–27; Dublin, Ireland. Stroudsburg, PA, USA: ACL; 2022. p. 6279–399. doi:10.18653/v1/2022.acl-long.435.
106. Adeyemi M, Oladipo A, Pradeep R, Lin J. Zero-shot cross-lingual reranking with large language models for low-resource languages. In: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics; 2024 Aug 11–16; Bangkok, Thailand. Stroudsburg, PA, USA: ACL; 2024. p. 650–6. doi:10.18653/v1/2024.acl-short.59.
107. Kuulmets HA, Purason T, Luhtaru A, Fishel M. Teaching llama a new language through cross-lingual knowledge transfer. In: Proceedings of the Findings of the Association for Computational Linguistics: NAACL 2024; 2024 Jun 16–21; Mexico City, Mexico. Stroudsburg, PA, USA: ACL; 2024. p. 3309–22. doi:10.18653/v1/2024.findings-naacl.210.
108. Tao M, Zhang C, Huang Q, Ma T, Huang S, Zhao D, et al. Unlocking the potential of model merging for low-resource languages. In: Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2024; 2024 Jun 16–21; Miami, FL, USA. Stroudsburg, PA, USA: ACL; 2024. p. 8705–20. doi:10.18653/v1/2024.findings-emnlp.508.
109. Shu P, Chen J, Liu Z, Wang H, Wu Z, Zhong T, et al. Transcending language boundaries: harnessing LLMs for low-resource language translation. arXiv:2411.11295. 2024. doi:10.48550/arXiv.2411.11295.

110. Hong S, Lee S, Moon H, Lim H. MIGRATE: cross-lingual adaptation of domain-specific LLMs through code-switching and embedding transfer. In: Proceedings of the 31st International Conference on Computational Linguistics (COLING 2025); 2025 Jan 19–24; Abu Dhabi, UAE. p. 9184–93.
111. Aggarwal D, Sathe A, Sitaram S. Exploring pretraining via active forgetting for improving cross-lingual transfer for decoder language models. arXiv:2410.16168. 2024.
112. Wang Z, Li J, Zhou H, Weng R, Wang J, Huang X, et al. Investigating and scaling up code-switching for multilingual language model pre-training. In: Proceedings of the Findings of the Association for Computational Linguistics: ACL 2025; 2025 Jul 27–Aug 1; Vienna, Austria. Stroudsburg, PA, USA: ACL; 2025. p. 11032–46. doi:10.18653/v1/2025.findings-acl.575.
113. Rathore VK, Deb A, Chandresh AK, Singla P, Mausam. SSP: self-supervised prompting for cross-lingual transfer to low-resource languages using large language models. In: Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2024; 2024 Nov 12–16; Miami, FL, USA. Stroudsburg, PA, USA: ACL; 2024. p. 15081–102. doi:10.18653/v1/2024.findings-emnlp.886.
114. Hu E, Shen Y, Wallis P, Allen-Zhu Z, Li Y, Wang S, et al. LoRA: low-rank adaptation of large language models. In: Proceedings of the International Conference on Learning Representations (ICLR 2022); 2022 Apr 25–29; Online.
115. 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. In: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics; 2025 Jul 28–30; Vienna, Austria. Stroudsburg, PA, USA: ACL; 2025. p. 849–71. doi:10.18653/v1/2025.acl-srw.62.
116. Toukmaji C, Flanigan J. Prompt, translate, fine-tune, re-initialize, or instruction-tune? Adapting LLMs for in-context learning in low-resource languages. arXiv:2506.19187. 2025.
117. Rust P, Pfeiffer J, Vulić I, Ruder S, Gurevych I. How good is your tokenizer? on the monolingual performance of multilingual language models. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing; 2021 Aug 1–6; Online. Stroudsburg, PA, USA: ACL; 2021. p. 3118–35. doi:10.18653/v1/2021.acl-long.243.
118. Lyu S, Deng Y, Liu G, Qi Z, Wang R. Transferable modeling strategies for low-resource LLM tasks: a prompt and alignment-based approach. In: Proceedings of the 2025 7th International Conference on Artificial Intelligence Technologies and Applications (ICAITA); 2025 Jun 27–29; Wenzhou, China. p. 225–9. doi:10.1109/icaita67588.2025.11137773.
119. Deng Y. Transfer methods for large language models in low-resource text generation tasks. J Comput Sci Softw Appl. 2024;4(6). doi:10.5281/zenodo.15392270.
120. Pei R, Liu Y, Lin P, Yvon F, Schuetze H. Understanding in-context machine translation for low-resource languages: a case study on Manchu. In: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics; 2025 Jul 27–Aug 1; Vienna, Austria. Stroudsburg, PA, USA: ACL; 2025. p. 8767–88. doi:10.18653/v1/2025.acl-long.429.
121. Li L, Hu W, Luo M. PNMT: zero-resource machine translation with pivot-based feature converter. Comput Mater Contin. 2025;84(3):5915–35. doi:10.32604/cmc.2025.064349.
122. Tanwar E, Dutta S, Borthakur M, Chakraborty T. Multilingual LLMs are better cross-lingual in-context learners with alignment. In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics; 2023 Jul 9–14; Toronto, ON, Canada. Stroudsburg, PA, USA: ACL; 2023. p. 6292–307. doi:10.18653/v1/2023.acl-long.346.
123. Wu L, Wei HR, Yang B, Lu W. From English to second language mastery: enhancing LLMs with cross-lingual continued instruction tuning. In: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics; 2025 Jul 27–Aug 1; Vienna, Austria. Stroudsburg, PA, USA: ACL; 2025. p. 23006–23. doi:10.18653/v1/2025.acl-long.1121.
124. 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. In: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics; 2025 Jul 27–Aug 1; Vienna, Austria. Stroudsburg, PA, USA: ACL; 2025. p. 2375–400. doi:10.18653/v1/2025.acl-long.118.

125. Zhuang W, Sun Y, Zhao X. Enhancing cross-lingual transfer through reversible transliteration: a Huffman-based approach for low-resource languages. In: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics; 2025 Jul 27–Aug 1; Vienna, Austria. Stroudsburg, PA, USA: ACL; 2025. p. 16299–313. doi:10.18653/v1/2025.acl-long.795.
126. Asai A, Kudugunta S, Yu X, Blevins T, Gonen H, Reid M, et al. BUFFET: benchmarking LLMs for few-shot cross-lingual transfer. In: Proceedings of NAACL-HLT 2024; 2024 Jun 16–21; Miami, FL, USA. p. 1771–800.
127. Kargaran AH, Modarressi A, Nikeghbal N, Diesner J, Yvon F, Schuetze H. MEXA: multilingual evaluation of English-centric LLMs via cross-lingual alignment. In: Proceedings of the Findings of the Association for Computational Linguistics: ACL 2025; 2025 Jul 27–Aug 1; Vienna, Austria. Stroudsburg, PA, USA: ACL; 2025. p. 27001–23. doi:10.18653/v1/2025.findings-acl.1385.
128. Sindhujan A, Kanojia D, Orasan C, Qian S. When LLMs struggle: reference-less translation evaluation for low-resource languages. In: Proceedings of the 1st Workshop on Language Models for Low-Resource Languages; 2025 Jan 19–20; Abu Dhabi, UAE. p. 437–59.
129. Court S, Elsner M. Shortcomings of LLMs for low-resource translation: retrieval and understanding are both the problem. In: Proceedings of the 9th Conference on Machine Translation; 2024 Nov 15–16; Miami, FL, USA. p. 1332–54.
130. Zhang C, Liu X, Lin J, Feng Y. Teaching LLMs an unseen language on the fly. arXiv:2402.19167. 2024.
131. Mahfuz T, Dey S, Naswan R, Adil H, Sayeed K, Shahgir H. Too late to train, too early to use? A study on necessity and viability of low-resource Bengali LLMs. In: Proceedings of the 31st International Conference on Computational Linguistics; 2025 Jan 19–24; Abu Dhabi, UAE. p. 1183–200.
132. Tejaswi A, Gupta N, Choi E. Exploring design choices for building language-specific LLMs. In: Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2024; 2024 Nov 12–16; Miami, FL, USA. Stroudsburg, PA, USA: ACL; 2024. p. 10485–500. doi:10.18653/v1/2024.findings-emnlp.614.
133. Tran K, 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 Aug 16; Bangkok, Thailand. p. 193–202.
134. Wen-Yi A, Mimno D. Hyperpolyglot LLMs: cross-lingual interpretability in token embeddings. In: Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing; 2023 Dec 6–10; Singapore. Stroudsburg, PA, USA: ACL; 2023. p. 1124–31. doi:10.18653/v1/2023.emnlp-main.71.
135. Purason T, Kuulmets HA, Fishel M. LLMs for extremely low-resource finno-Ugric languages. In: Proceedings of the Findings of the Association for Computational Linguistics: NAACL 2025; 2025 Apr 30–May 4; Albuquerque, NM, USA. Stroudsburg, PA, USA: ACL; 2025. p. 6677–97. doi:10.18653/v1/2025.findings-naacl.373.
136. Gusmita RH, Firmansyah AF, Zahera HM, Ngonga Ngomo AC. ELEVATE-ID: extending large language models for end-to-end entity linking evaluation in Indonesian. Data Knowl Eng. 2026;161(3):102504. doi:10.1016/j.datak.2025.102504.
137. Kia MA, Samiee D. From monolingual to multilingual: enhancing hate speech detection with multi-channel language models. Procedia Comput Sci. 2024;246:2704–13. doi:10.1016/j.procs.2024.09.401.
138. 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? In: Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2024; 2024 Nov 12–16; Miami, FL, USA. Stroudsburg, PA, USA: ACL; 2024. p. 1771–93. doi:10.18653/v1/2024.findings-emnlp.96.
139. Zhang B, Haddow B, Birch A. Prompting large language model for machine translation: a case study. In: Proceedings of the 40th International Conference on Machine Learning. Vol. 202 of Proceedings of Machine Learning Research; 2023 Jul 23–29; Honolulu, HI, USA. p. 41092–110.
140. Feng S, Shi W, Wang Y, Ding W, Ahia O, Li SS, et al. Teaching LLMs to abstain across languages via multilingual feedback. In: Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing; 2024 Nov 12–16; Miami, FL, USA. Stroudsburg, PA, USA: ACL; 2024. p. 4125–50. doi:10.18653/v1/2024.emnlp-main.239.

141. Lee S, Hong S, Moon H, Lim H. Semantic aware linear transfer by recycling pre-trained language models for cross-lingual transfer. In: Proceedings of the Findings of the Association for Computational Linguistics: ACL 2025; 2025 Jul 27–Aug 1; Vienna, Austria. Stroudsburg, PA, USA: ACL; 2025. p. 16180–93. doi:10.18653/v1/2025.findings-acl.832.
142. Pan S, Tian Z, Ding L, Zheng H, Huang Z, Wen Z, et al. POMP: probability-driven meta-graph prompter for LLMs in low-resource unsupervised neural machine translation. In: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics; 2024 Aug 11–16; Bangkok, Thailand. Stroudsburg, PA, USA: ACL; 2024. p. 9976–92. doi:10.18653/v1/2024.acl-long.537.
143. Mondal SK, Sen S, Singhania A, Jyothi P. Language-specific neurons do not facilitate cross-lingual transfer. In: Proceedings of the Sixth Workshop on Insights from Negative Results in NLP; 2025 May 3–4; Albuquerque, NM, USA. Stroudsburg, PA, USA: ACL; 2025. p. 46–62. doi:10.18653/v1/2025.insights-1.6.
144. 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(2):100375. doi:10.1016/j.caeai.2025.100375.
145. Jung JY, Tyack L, von Davier M. Combining machine translation and automated scoring in international large-scale assessments. Large Scale Assess Educ. 2024;12(1):10. doi:10.1186/s40536-024-00199-7.
146. Shin HJ, Yamamoto K, He Q, von Davier M. Automatic scoring of constructed responses in PISA using the machine-supported coding system. In: Innovative digital-based international large-scale assessments. Cham, Switzerland: Springer Nature; 2025. p. 299–320. doi:10.1007/978-3-031-90951-1_12.