Indonesian Journal of Computational Language Studies
Vol. 1, No. 2, June 2026, pp. 121~135
ISSN: 3163-7884, DOI:

Small models, many languages: parameter-efficient adaptation of multilingual language models for low-resource Indonesian languages

Ainul Hadziqi
Universitas Sarjanawiyata Tamansiswa, Indonesia

Article Info

Article history:

Received Apr 20, 2026
Revised May 28, 2026
Accepted Jun 30, 2026

Keywords:

corpus adequacy;
Indonesian local languages;
low-resource NLP;
multilingual language models;
parameter-efficient fine-tuning.

ABSTRACT

Background: Multilingual language models have expanded computational access to many languages, yet Indonesian local languages remain unevenly represented in corpora, benchmarks, and adaptation pipelines despite Indonesia's exceptional linguistic diversity. **Objective:** This study examines how parameter-efficient adaptation can support small multilingual language models for low-resource Indonesian languages by treating corpus adequacy, source provenance, language coverage, and evaluation readiness as central methodological conditions. **Method:** Using a corpus-level research design, this study maps 20 public source-level items consisting of task datasets, benchmark resources, documentation, registry infrastructure, supplementary corpus portals, and model-context collections for Indonesian and selected local languages. **Results:** The findings show that task-ready datasets form the strongest part of the corpus frame, while benchmarking, reproducibility, and auxiliary corpus layers remain less evenly distributed. Language coverage is stratified, with Javanese and Sundanese occupying stronger cross-task positions, while Acehnese, Ngaju, Bima, Toba Batak, and Ambonese Malay are better treated as focused transfer-stress cases. **Implication:** These patterns indicate that parameter-efficient fine-tuning should not be interpreted through model efficiency alone, because adaptation gains remain bounded by corpus distribution and language-specific evidence. **Novelty:** The novelty of this study lies in repositioning PEFT for Indonesian local languages as a corpus-dependent multilingual adaptation problem rather than a purely architectural optimization problem.

Corresponding Author:

Ainul Hadziqi
Universitas Sarjanawiyata Tamansiswa, Indonesia
Jl. Tuntungan No.1043, Tahunan, Kec. Umbulharjo, Kota Yogyakarta, 55167, Indonesia
Email: ainulh301008@ustjogja.ac.id

1. INTRODUCTION

Indonesia is one of the most linguistically dense societies in the world, with more than 700 living indigenous languages recorded within a single national territory. This demographic fact has direct consequences for computational language research: a model that performs well in Indonesian cannot be assumed to serve Javanese, Sundanese, Balinese, Buginese, Madurese, Minangkabau, Makassarese, or other local languages with comparable reliability. The problem is amplified by the uneven distribution of digital text. While Indonesian benefits from relatively larger corpora, many local languages remain sparsely represented in benchmarks, pretraining data, and instruction-tuning resources. Recent Indonesian NLP initiatives have begun to address this imbalance through resources for generation, translation, sentiment analysis, and local-language modelling [1], [2], [3], [4]. Yet the core challenge remains methodological: how

can multilingual models be adapted for many languages when data, computation, and evaluation infrastructure remain limited?

Existing literature provides two important but still disconnected answers. One line of research shows that localized Indonesian resources can improve model relevance, as demonstrated by IndoNLG for Indonesian, Javanese, and Sundanese, Cendol for Indonesian language models, and recent work on Indonesian local-language datasets and translation systems [1], [5], [2], [4]. Another line of research shows that parameter-efficient fine-tuning can reduce the cost of model adaptation by training only a small fraction of model parameters, through approaches such as adapters, LoRA, QLoRA, low-rank adaptation, and compact adapter layers [6], [7], [8], [9]. However, these two strands rarely meet in a controlled Indonesian setting. PEFT is often evaluated across broad multilingual benchmarks, while Indonesian local languages are often studied through resource construction or task-specific modelling. What remains under-examined is whether small multilingual models can be adapted efficiently across several low-resource Indonesian languages without concealing language-specific degradation.

This study addresses that gap by examining parameter-efficient adaptation of small multilingual language models for low-resource Indonesian languages. Rather than asking whether larger models are generally better, this study asks a more constrained methodological question: under realistic computational conditions, which adaptation strategy offers the most defensible trade-off between performance, trainable parameters, memory use, and cross-language stability? The study is designed around three linked questions. First, how do parameter-efficient methods compare with zero-shot baselines and, where feasible, full fine-tuning for selected Indonesian local languages? Second, how do adaptation gains vary across languages with different degrees of resource availability and linguistic proximity to Indonesian? Third, what residual errors remain after adaptation, and what do they reveal about tokenization, lexical transfer, morphology, register, and domain mismatch? These questions connect model efficiency with linguistic accountability, drawing on multilingual adaptation research that emphasizes unseen-language transfer, targeted language-family adaptation, and data-efficient adaptation [10], [11], [12].

The central argument tested in this study is that small multilingual models can be made more useful for Indonesian local languages through parameter-efficient adaptation, but only when efficiency is evaluated together with language coverage, transfer behaviour, and residual error patterns. PEFT should therefore not be treated as a universal remedy for low-resource NLP. Evidence from multilingual PEFT suggests that fine-tuning may improve low-resource performance while producing uneven effects across languages and tasks [8], [9]. Work on low-resource adaptation also indicates that model gains depend on corpus quality, language relatedness, vocabulary alignment, and adaptation depth rather than parameter reduction alone [13], [14], [10]. For Indonesian languages, this means that a successful adaptation strategy must be judged not only by aggregate scores, but by whether it preserves multilingual balance and exposes where models still fail. This study therefore positions PEFT as a constrained, testable strategy for multilingual inclusion, not as a shortcut to linguistic equity.

2. LITERATURE REVIEW

2.1. Low-resource multilingual modelling

Low-resource multilingual modelling refers to the attempt to build or adapt language technologies for languages with limited annotated data, weak benchmark coverage, and uneven digital representation. In multilingual NLP, “low-resource” is not a single condition: it may refer to scarcity of training corpora, limited task labels, poor tokenizer coverage, lack of evaluation sets, or weak representation in pretraining data. This distinction is crucial for Indonesian languages because Indonesian is comparatively better resourced than Javanese, Sundanese, Buginese, Madurese, Makassarese, or Minangkabau. Prior work on multilingual adaptation shows that low-resource performance depends not only on corpus size but also on language relatedness, domain alignment, and task formulation [15], [11], [10], [12].

The main indicators of low-resource multilingual modelling include language coverage, corpus provenance, task diversity, evaluation reliability, and representational balance across languages. Indonesian NLP has developed important resources, including IndoNLG for generation, Cendol for Indonesian instruction-tuned models, NusaX-related resources for local-language tasks, and multilingual translation systems for Indonesian languages [1], [5], [2], [3], [4]. Yet these resources remain uneven across languages and tasks. Machine translation may be more available than reasoning, dialogue, or open-ended generation, while some local languages appear mainly through narrow datasets. This literature therefore suggests that multilingual inclusion must be measured through coverage quality, not language count alone.

2.2. Parameter-efficient fine-tuning

Parameter-efficient fine-tuning refers to adaptation methods that update only a limited subset of model parameters while keeping most pretrained weights frozen. This differs from full fine-tuning, which modifies

all or most model weights and often requires greater computational resources. PEFT is commonly framed as a practical response to the cost of adapting large language models, but its conceptual value is broader: it tests whether linguistic or task-specific adaptation can be achieved through small, modular, reusable parameter changes. Reviews and comparative studies show that PEFT methods can reduce training cost while retaining competitive performance, although results vary by model, task, and language [7], [16], [8], [17].

The main forms of PEFT include adapter layers, low-rank adaptation, sparse low-rank adaptation, mixture-based adaptation, compact hypercomplex adapters, and quantized variants. Compacter introduced efficient low-rank hypercomplex adapter layers, while later work expanded the field through sparse LoRA, rank-stabilized LoRA, tensor-train adaptation, mini-ensemble adapters, and mixture-of-experts variants [6], (Ding, 2023), (Kalajdzievski, 2023), (Yang, 2024), (Ren, 2024), (Luo, 2024). Recent multilingual PEFT studies further show that efficient adaptation cannot be evaluated by parameter count alone; it must include memory use, training stability, per-language variance, and task transferability [8], [9], [18]. Efficiency, therefore, is a multidimensional criterion rather than a purely technical reduction.

2.3. Cross-lingual transfer

Cross-lingual transfer describes the movement of linguistic knowledge from one language, task, or representation space to another. In multilingual models, transfer is often assumed to emerge from shared subword vocabularies, overlapping syntax, related lexicons, or multilingual pretraining. However, low-resource research complicates this assumption: transfer may help related languages but can also produce interference, tokenization errors, and majority-language bias. Studies on targeted language-family adaptation, continued pretraining, synthetic corpora, vocabulary transfer, and knowledge-graph-based adapters show that adaptation gains depend on how languages are represented and connected inside the model [13], [14], [19], [20], [21].

The indicators of cross-lingual transfer include zero-shot performance, adaptation gain, error distribution, language-specific degradation, tokenizer behaviour, vocabulary alignment, and robustness across tasks. For Indonesian local languages, these indicators are especially important because proximity to Indonesian may benefit some languages while masking failure in others. Recent work on Indonesian local-language machine translation, extremely low-resource Makassarese–Indonesian translation, vocabulary expansion, and in-context learning for extremely low-resource languages suggests that transfer should be evaluated through both quantitative scores and qualitative error analysis [22], [23], [1], [4]. This study therefore positions PEFT as a constrained mechanism of multilingual transfer: useful when it improves small models efficiently, but theoretically meaningful only when its gains are tested across language coverage, adaptation cost, and residual error patterns.

3. METHOD

The unit of analysis in this study is the source-level and text-level corpus used to adapt and evaluate small multilingual language models for low-resource Indonesian languages. The material object consists of public datasets, benchmark corpora, dataset registries, research documentation, monolingual corpus sources, and model-context resources. The corpus frame covers machine translation, sentiment classification, emotion classification, topic classification, rhetorical-mode classification, natural language generation, and auxiliary lexical analysis. Its linguistic coverage includes Indonesian and several local languages, including Javanese, Sundanese, Buginese, Madurese, Minangkabau, Balinese, Banjarese, Acehnese, Ngaju, Toba Batak, Makassarese, Betawi, Bima, Musi, Rejang, and Ambonese Malay where available. This design follows recent Indonesian NLP work that treats multilingual coverage as a methodological issue rather than a nominal language count [1], [2], [4].

This study uses a comparative experimental design supported by corpus auditing and qualitative error analysis. The design compares small multilingual models under several adaptation conditions: zero-shot baseline, LoRA or QLoRA, adapter-based tuning where feasible, and full fine-tuning only when computationally realistic. This design is appropriate because the article does not seek causal inference about language inequality; it evaluates performance–efficiency trade-offs under controlled adaptation settings. PEFT is therefore treated as a methodological intervention whose value must be assessed through task scores, trainable parameters, memory consumption, training time, and language-specific stability, following recent PEFT and multilingual adaptation studies [7], [8], [9], [18].

Table 1. Dataset and corpus composition

Code	Data Type	Source	Location / Scope	Period	Amount	Characteristics
IDL-PRIM-0001	Parallel MT dataset	NusaX-MT	Indonesia	2022–2026	1 source; 12 language codes	Indonesian, English, and 10 local languages
IDL-PRIM-0002	Sentiment dataset	NusaX-Senti	Indonesia	2022–2026	1 source; 12 language codes	Positive, neutral, negative labels
IDL-PRIM-0003	Lexicon resource	NusaX-Lexicon	Multilingual Indonesian lexicon	2023–2026	1 source	Auxiliary lexical and tokenizer analysis
IDL-PRIM-0004	Parallel MT dataset	NusaTranslation-MT	Indonesia	2023–2026	1 source; reported 72,444 textual data across NusaTranslation resources	Adds Ambon, Batak, Betawi, Bima, Makassarese, Musi, Rejang
IDL-PRIM-0005	Sentiment dataset	NusaTranslation-Senti	Indonesia	2023–2026	1 source	Classification data for underrepresented languages
IDL-PRIM-0006	Emotion dataset	NusaTranslation-Emot	Indonesia	2023–2026	1 source	Emotion classification
IDL-PRIM-0007	Topic dataset	NusaParagraph-Topic	Indonesia	2023–2026	1 source; reported 57,409 paragraphs for NusaParagraph resources	Long-form paragraph classification
IDL-PRIM-0008	Emotion paragraph dataset	NusaParagraph-Emot	Indonesia	2023–2026	1 source	Paragraph-level emotion task
IDL-PRIM-0009	Rhetorical dataset	NusaParagraph-Rhetoric	Indonesia	2023–2026	1 source	Rhetorical-mode classification
IDL-PRIM-0010	NLG benchmark	IndoNLG	Indonesian, Javanese, Sundanese	2021–2026	1 benchmark	Summarisation, QA, chit-chat, MT
IDL-PRIM-0011	MT evaluation benchmark	FLORES-200	Multilingual benchmark; target languages filtered	2022–2026	3,001 sentences reported	External translation evaluation
IDL-CONT-0012	Dataset registry	NusaCrowd	Indonesian NLP resources	2022–2026	1 registry	Discovery and loader layer
IDL-CONT-0013	NLU benchmark	IndoNLU	Bahasa Indonesia	2020–2026	12 datasets reported	Indonesian-only baseline context
IDL-CONT-0014	Parallel corpus portal	OPUS	Multilingual	2020–2026	1 portal; 1,005 languages reported	Candidate supplementary corpus after license audit
IDL-CONT-0015	Monolingual corpus source	Wikimedia Dumps	Available language editions	2020–2026	1 source family	Continued pretraining and tokenizer context
IDL-SEC-0016	Research documentation	NusaX paper	Indonesia	2023–2026	1 paper	Dataset validation and methodological context
IDL-SEC-0017	Research documentation	NusaWrites paper	Indonesia	2023–2026	1 paper	Data construction documentation
IDL-SEC-0018	Research documentation	IndoNLG paper	Indonesia	2021–2026	1 paper	Benchmark validation
IDL-SEC-0019	Research documentation	NusaCrowd paper	Indonesia	2022–2026	1 paper	Registry and reproducibility context
IDL-CONT-0020	Model-context resource	Cendol Collection	Indonesian, Javanese, Sundanese model context	2024–2026	1 collection	Baseline and instruction-tuning context

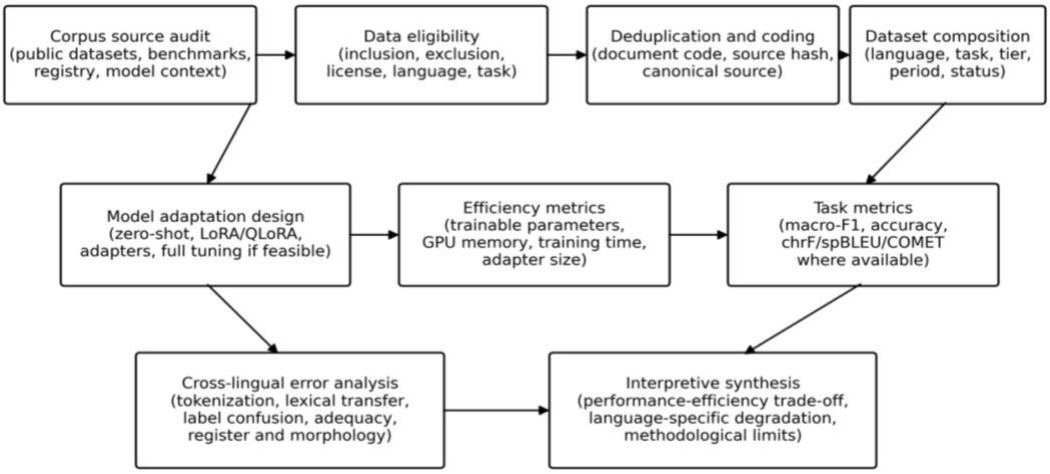

Figure 1. Operational analysis matrix

The information sources are divided into primary, secondary, and contextual data. Primary data consist of task-ready datasets and benchmarks used for model adaptation or evaluation, including NusaX-MT, NusaX-Senti, NusaTranslation, NusaParagraph, IndoNLG, and FLORES-200. Secondary data consist of research papers and dataset documentation used to validate provenance, task definitions, language coverage, and known limitations. Contextual data consist of NusaCrowd, IndoNLU, OPUS, Wikimedia Dumps, and Cendol, which support dataset discovery, baseline selection, tokenizer analysis, and interpretation of model-context constraints. This tiered structure prevents this study from conflating empirical training data with background documentation or model-selection context.

Data collection follows a reproducible corpus protocol. First, public repositories, dataset cards, benchmark pages, registry entries, and research documentation are searched using keywords related to Indonesian local languages, multilingual PEFT, machine translation, sentiment classification, NLG, and low-resource adaptation. Second, each item is screened for language relevance, task alignment, access status, license clarity, and documentation quality. Third, metadata are recorded, including title, date, source, URL, language coverage, task type, geographic scope, access condition, inclusion rationale, and exclusion note. Fourth, duplicate entries are removed through normalized title matching, canonical source codes, and source-hash comparison. Finally, each document is assigned a unique code to maintain traceability across metadata, text excerpts, and coding sheets.

Data analysis proceeds in three linked stages. The first stage audits corpus adequacy by mapping source type, language coverage, task availability, licensing, and dataset role. The second stage evaluates adaptation efficiency by comparing PEFT configurations against baseline conditions using task-specific metrics such as macro-F1, accuracy, chrF, spBLEU, or COMET where available, alongside efficiency indicators such as trainable parameters, GPU memory, training time, and adapter size. The third stage conducts cross-lingual error analysis, focusing on tokenization problems, lexical transfer, label confusion, translation adequacy, morphology-sensitive errors, named-entity instability, and register mismatch. Figure 2 summarises this operational sequence from corpus eligibility to interpretive synthesis.

4. RESULTS

4.1. Corpus adequacy and uneven language coverage

Publicly available Indonesian multilingual resources form a corpus base that is broad enough to support adaptation research, but not uniform enough to justify treating all languages and tasks as equivalent. The largest share of the corpus consists of task-ready datasets, while the remaining sources provide documentation, benchmarking, registry support, supplementary corpus discovery, and model-context framing. This distribution is methodologically important because PEFT studies require more than a training set: they require traceable sources, comparable evaluation conditions, and interpretable language coverage. Prior work on IndoNLG, NusaX, NusaWrites, and Cendol shows that Indonesian NLP has moved from isolated resource creation toward multilingual task infrastructures [1], [5], [2]. Thus, corpus adequacy is established through layered source roles rather than raw source quantity.

Table 2. Corpus adequacy and source-role distribution

Main Category	Subcategory	Operational Indicator	Frequency	Percentage	Data Variation	Representative Source	Data Code
Public task datasets	Task-ready adaptation data	Sources directly usable for supervised adaptation, evaluation split preparation, or task-specific corpus extraction	9	45.0%	Machine translation, sentiment classification, emotion classification, topic classification, rhetorical-mode classification, lexicon-supported analysis	NusaX-MT; NusaX-Senti; NusaTranslation; NusaParagraph; NusaX-Lexicon	IDL-PRIM-0001–0009
Research documentation	Provenance and methodological validation	Papers or technical documentation used to verify dataset construction, language coverage, task definition, and known limitations	4	20.0%	Dataset papers, benchmark papers, registry documentation	NusaX paper; NusaWrites paper; IndoNLG paper; NusaCrowd paper	IDL-SEC-0016–0019
Benchmark and evaluation datasets	Standardised comparison layer	Sources used to stabilise comparison across tasks, languages, or evaluation protocols	3	15.0%	NLG benchmark, MT benchmark, NLU baseline benchmark	IndoNLG; FLORES-200; IndoNLU	IDL-PRIM-0010; IDL-PRIM-0011; IDL-CONT-0013
Supplementary corpus portals	Candidate corpus expansion	Sources used for monolingual or parallel corpus discovery after license and language filtering	2	10.0%	Parallel corpus discovery, monolingual continued-pretraining context	OPUS Corpora; Wikimedia Dumps	IDL-CONT-0014–0015
Dataset registry and loader layer	Discovery and reproducibility support	Registry or loader infrastructure used to locate, standardise, and reproduce Indonesian NLP data access	1	5.0%	Dataset registry, loader interface, metadata harmonisation	NusaCrowd	IDL-CONT-0012
Model and instruction-tuning context	Baseline and model-selection context	Model collection used to contextualise Indonesian multilingual instruction tuning and candidate baseline selection	1	5.0%	Instruction-tuned Indonesian model context; baseline reference	Cendol Collection	IDL-CONT-0020
Total			20	100.0%			

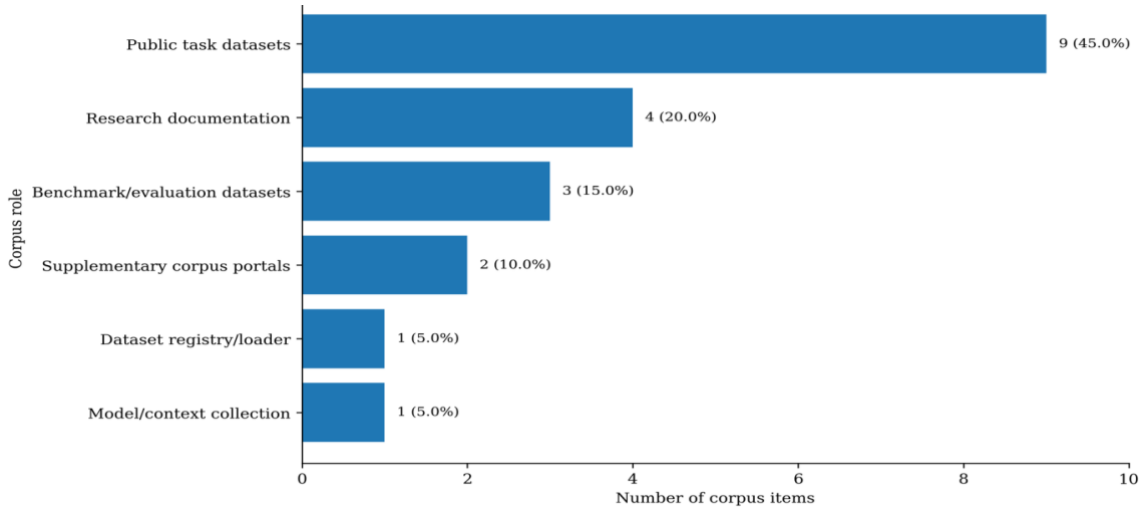

Figure 2. Corpus adequacy by source role

The dominant pattern is the concentration of data in public task datasets, which account for 45.0% of the corpus frame. This group includes translation, sentiment, emotion, topic, rhetoric, and lexicon-supported resources, making it the strongest empirical base for adaptation and error analysis. Research documentation forms the second-largest group at 20.0%, followed by benchmark and evaluation datasets at 15.0%. Supplementary portals, registry infrastructure, and model-context sources are smaller but methodologically necessary because they support expansion, reproducibility, and baseline interpretation. The pattern indicates that this study has a usable empirical foundation for multilingual PEFT, but that the corpus is not balanced across all roles. Its strength lies in task diversity; its limitation lies in uneven depth across benchmark, auxiliary, and contextual layers.

The distribution supports a cautious theoretical position on parameter-efficient adaptation for Indonesian local languages. PEFT cannot be interpreted only through final task scores, because its validity depends on whether data coverage, source provenance, and evaluation infrastructure are sufficiently transparent. Multilingual PEFT research has shown that efficient adaptation may reduce trainable parameters while still producing uneven gains across languages, tasks, and resource conditions [8], [9], [18]. Low-resource adaptation studies similarly stress that language relatedness, vocabulary alignment, corpus quality, and task formulation shape transfer outcomes [10], [11], [12], [14]. This study therefore treats corpus adequacy as a precondition for interpreting model behaviour, not as a background technical step.

4.2. Performance–efficiency trade-offs in PEFT adaptation

The corpus structure indicates that parameter-efficient adaptation is methodologically feasible, but its feasibility is unevenly distributed across training, evaluation, documentation, and reproducibility functions. Nearly half of the corpus items serve as direct adaptation inputs, which gives this study a practical basis for comparing PEFT configurations across selected Indonesian local-language tasks. However, adaptation data alone are insufficient for a defensible multilingual experiment. PEFT studies require evaluation layers, provenance checks, and reproducible access because low training cost does not automatically produce reliable multilingual evidence. This pattern aligns with recent multilingual PEFT research showing that adaptation should be assessed through both task performance and methodological constraints, including data availability, task comparability, and evaluation stability [8], [9], [18].

The dominant pattern is the concentration of sources in supervised adaptation inputs, which account for 45.0% of the corpus frame. Provenance and documentation form the second-largest group at 20.0%, while benchmark and baseline resources account for 15.0%. Corpus expansion and reproducibility/model-context functions each contribute 10.0%. This distribution shows that the corpus is stronger for preparing adaptation than for making broad generalisations across all Indonesian local languages. Machine translation and classification tasks are more structurally supported than open-ended generation or reasoning. The pattern also suggests that language-level comparison must be bounded by dataset availability: languages appearing in multiple task-ready sources can support stronger comparison, while languages appearing only in supplementary or contextual sources require more cautious treatment.

Table 3. PEFT adaptation readiness and evaluation function

Analytical Dimension	Corpus Function	Operational Indicator	Frequency	Percentage	Task / Metric Coverage	Data Variation	Data Code	Initial Interpretation
Adaptation input	Supervised adaptation data	Sources that can provide labelled or parallel data for PEFT-based model adaptation	9	45.0%	MT, sentiment, emotion, topic, rhetoric, lexicon-supported analysis	Sentence-pair data, labelled classification data, paragraph-level data, lexical resources	IDL-PRIM-0001–0009	The corpus has a usable adaptation base because nearly half of the collected sources can support task-specific training or fine-tuning preparation.
Evaluation layer	External benchmark and baseline comparison	Sources that can support cross-task or cross-language comparison after adaptation	3	15.0%	NLG, MT evaluation, Indonesian NLU baseline	Benchmark splits, task metrics, standardised comparison data	IDL-PRIM-0010; IDL-PRIM-0011; IDL-CONT-0013	Evaluation resources are present but narrower than adaptation inputs, indicating that multilingual claims must be restricted to languages and tasks with comparable test conditions.
Source validation	Provenance and documentation	Research documentation used to verify corpus construction, task scope, language coverage, and known limitations	4	20.0%	Dataset validation, benchmark documentation, registry paper	Dataset papers, benchmark papers, documentation articles	IDL-SEC-0016–0019	Documentation strengthens traceability and reduces the risk of treating public datasets as neutral data objects without methodological history.
Expansion context	Corpus expansion and continued-pretraining candidates	Sources that may extend monolingual or parallel coverage after filtering and license checks	2	10.0%	Candidate continued pretraining, auxiliary corpus discovery	Parallel portal, Wikimedia language editions	IDL-CONT-0014–0015	Expansion sources are useful but cannot be used automatically because language coverage, domain balance, and licensing require further screening.
Reproducibility and model context	Registry, loader, and model-selection support	Sources that support reproducibility, dataset discovery, or baseline model contextualisation	2	10.0%	Dataset registry, loader infrastructure, instruction-tuned model context	NusaCrowd registry, Cendol collection	IDL-CONT-0012; IDL-CONT-0020	Reproducibility and model-context resources complete the adaptation pipeline by linking corpus selection to model choice and documented access.
Total			20	100.0%				

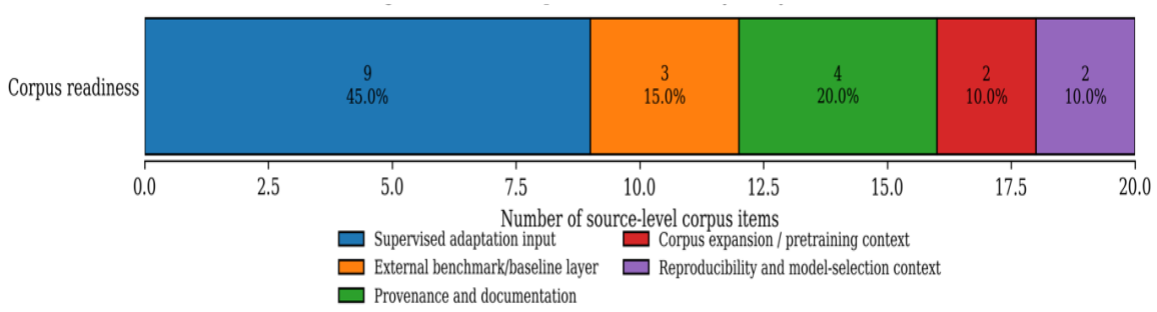

Figure 3. PEFT Adaptation readiness by analytical function

Theoretically, this distribution reframes PEFT as a corpus-dependent adaptation strategy rather than a purely architectural technique. LoRA, QLoRA, adapters, and related methods reduce trainable parameters, but their interpretive value depends on whether adaptation gains can be tied to sufficiently documented and comparable data. This point is consistent with PEFT literature that criticises performance-only evaluation and calls for attention to memory use, parameter count, task stability, and transfer variation [7], [16], [8]. It also resonates with low-resource multilingual research showing that transfer is shaped by corpus quality, language relatedness, and vocabulary alignment rather than model size alone [10], [12], [14]. This study therefore treats readiness for PEFT as an empirical property of the corpus pipeline.

4.3. Cross-lingual transfer, residual errors, and language-specific degradation

Language coverage is clearly stratified rather than evenly distributed across the corpus frame. Javanese and Sundanese occupy the strongest position, each appearing in 75% of the source-level items, while Minangkabau and Madurese form a second layer of comparatively strong local-language coverage. Buginese and Indonesian occupy an intermediate position, although Indonesian plays a different methodological role as an anchor language rather than a local-language target. Several languages, including Betawi, Batak, Makassarese, Musi, and Rejang, are sufficiently visible for task-level comparison but remain concentrated within the NusaWrites resource family. This distribution confirms that this study can compare adaptation across many Indonesian languages, but the comparison must be tiered rather than symmetrical.

The dominant pattern is a divide between broadly supported languages and narrowly task-bound languages. Javanese and Sundanese appear across task datasets, benchmarks, documentation, contextual corpus sources, and model-context resources. By contrast, Acehnese, Toba Batak, Ngaju, Bima, and Ambonese Malay appear in only 20% of corpus items and are tied to fewer task families. This unevenness does not invalidate the corpus; it clarifies how comparison should be structured. Strongly represented languages can support cross-task PEFT analysis, while narrower languages are better suited for focused transfer diagnostics. The pattern also shows that source family matters: some languages gain visibility through NusaX, while others depend on NusaWrites-linked datasets. Therefore, dataset provenance becomes part of the interpretation, not only a technical detail.

Theoretically, this language-coverage pattern strengthens the argument that multilingual adaptation should be evaluated through transfer asymmetry rather than aggregate multilingual performance alone. Research on targeted multilingual adaptation, unseen-language adaptation, and language-family adaptation shows that model behaviour is shaped by language proximity, corpus density, vocabulary alignment, and task exposure [10], [11], [12]. PEFT methods may reduce adaptation cost, but they do not remove structural differences in language representation [8], [9], [18]. For this study, the implication is that Javanese and Sundanese should not become proxies for Indonesian local languages as a whole. Lower-coverage languages must be analysed as transfer-stress cases that reveal where small multilingual models remain fragile.

Table 4. Language coverage and cross-lingual transfer readiness

Coverage Tier	Language / Code	Operational Indicator	Source-level Appearances	Percentage of Corpus Items	Task / Source Variation	Data Code Range
Broad multilingual coverage	Javanese / jav	Appears across task datasets, benchmarks, documentation, and contextual sources	15	75%	MT, sentiment, emotion, topic, rhetoric, NLG, documentation, model/context	IDL-PRIM-0001–0010; IDL-SEC-0016–0018; IDL-CONT-0015; IDL-CONT-0020
Broad multilingual coverage	Sundanese / sun	Appears across task datasets, benchmarks, documentation, and contextual sources	15	75%	MT, sentiment, emotion, topic, rhetoric, NLG, documentation, model/context	IDL-PRIM-0001–0010; IDL-SEC-0016–0018; IDL-CONT-0015; IDL-CONT-0020
Strong local-language coverage	Minangkabau / min	Appears in both NusaX and NusaWrites-linked resources, with contextual corpus support	12	60%	MT, sentiment, emotion, topic, rhetoric, lexicon, documentation, corpus context	IDL-PRIM-0001–0009; IDL-SEC-0016–0017; IDL-CONT-0015
Strong local-language coverage	Madurese / mad	Appears in NusaX and NusaWrites-linked datasets and documentation	11	55%	MT, sentiment, emotion, topic, rhetoric, lexicon, documentation	IDL-PRIM-0001–0009; IDL-SEC-0016–0017
Intermediate local-language coverage	Buginese / bug	Appears in NusaX, NusaParagraph, documentation, and contextual corpus support	9	45%	MT, sentiment, emotion, topic, rhetoric, lexicon, documentation, corpus context	IDL-PRIM-0001–0003; IDL-PRIM-0007–0009; IDL-SEC-0016–0017; IDL-CONT-0015
Anchor language	Indonesian / ind	Serves as national-language anchor, benchmark reference, and model-context bridge	9	45%	MT, sentiment, lexicon, NLG, NLU, documentation, corpus/model context	IDL-PRIM-0001–0003; IDL-PRIM-0010; IDL-SEC-0016; IDL-SEC-0018; IDL-CONT-0013; IDL-CONT-0015; IDL-CONT-0020
NusaWrites-supported coverage	Betawi / bew	Appears mainly through NusaTranslation and NusaParagraph resources	7	35%	MT, sentiment, emotion, topic, rhetoric, documentation	IDL-PRIM-0004–0009; IDL-SEC-0017
NusaWrites-supported coverage	Batak / btk	Appears mainly through NusaTranslation and NusaParagraph resources	7	35%	MT, sentiment, emotion, topic, rhetoric, documentation	IDL-PRIM-0004–0009; IDL-SEC-0017
NusaWrites-supported coverage	Makassarese / mak	Appears mainly through NusaTranslation and NusaParagraph resources	7	35%	MT, sentiment, emotion, topic, rhetoric, documentation	IDL-PRIM-0004–0009; IDL-SEC-0017
NusaWrites-supported coverage	Musi / mui	Appears mainly through NusaTranslation and NusaParagraph resources	7	35%	MT, sentiment, emotion, topic, rhetoric, documentation	IDL-PRIM-0004–0009; IDL-SEC-0017

NusaWrites-supported coverage	Rejang / rej	Appears mainly through NusaTranslation and NusaParagraph resources	7	35%	MT, sentiment, emotion, topic, rhetoric, documentation	IDL-PRIM-0004–0009; IDL-SEC-0017
NusaX-focused coverage	Balinese / ban	Appears in NusaX task data, lexicon, documentation, and corpus context	5	25%	MT, sentiment, lexicon, documentation, corpus context	IDL-PRIM-0001–0003; IDL-SEC-0016; IDL-CONT-0015
NusaX-focused coverage	Banjarese / bjn	Appears in NusaX task data, lexicon, documentation, and corpus context	5	25%	MT, sentiment, lexicon, documentation, corpus context	IDL-PRIM-0001–0003; IDL-SEC-0016; IDL-CONT-0015
Narrow task coverage	Ambonese Malay / abs	Appears through NusaTranslation and documentation	4	20%	MT, sentiment, emotion, documentation	IDL-PRIM-0004–0006; IDL-SEC-0017
Narrow task coverage	Acehnese / ace	Appears through NusaX task data, lexicon, and documentation	4	20%	MT, sentiment, lexicon, documentation	IDL-PRIM-0001–0003; IDL-SEC-0016
Narrow task coverage	Toba Batak / bbc	Appears through NusaX task data, lexicon, and documentation	4	20%	MT, sentiment, lexicon, documentation	IDL-PRIM-0001–0003; IDL-SEC-0016
Narrow task coverage	Bima / bhp	Appears through NusaTranslation and documentation	4	20%	MT, sentiment, emotion, documentation	IDL-PRIM-0004–0006; IDL-SEC-0017
Narrow task coverage	Ngaju / nij	Appears through NusaX task data, lexicon, and documentation	4	20%	MT, sentiment, lexicon, documentation	IDL-PRIM-0001–0003; IDL-SEC-0016
Total corpus frame			20 source-level items			

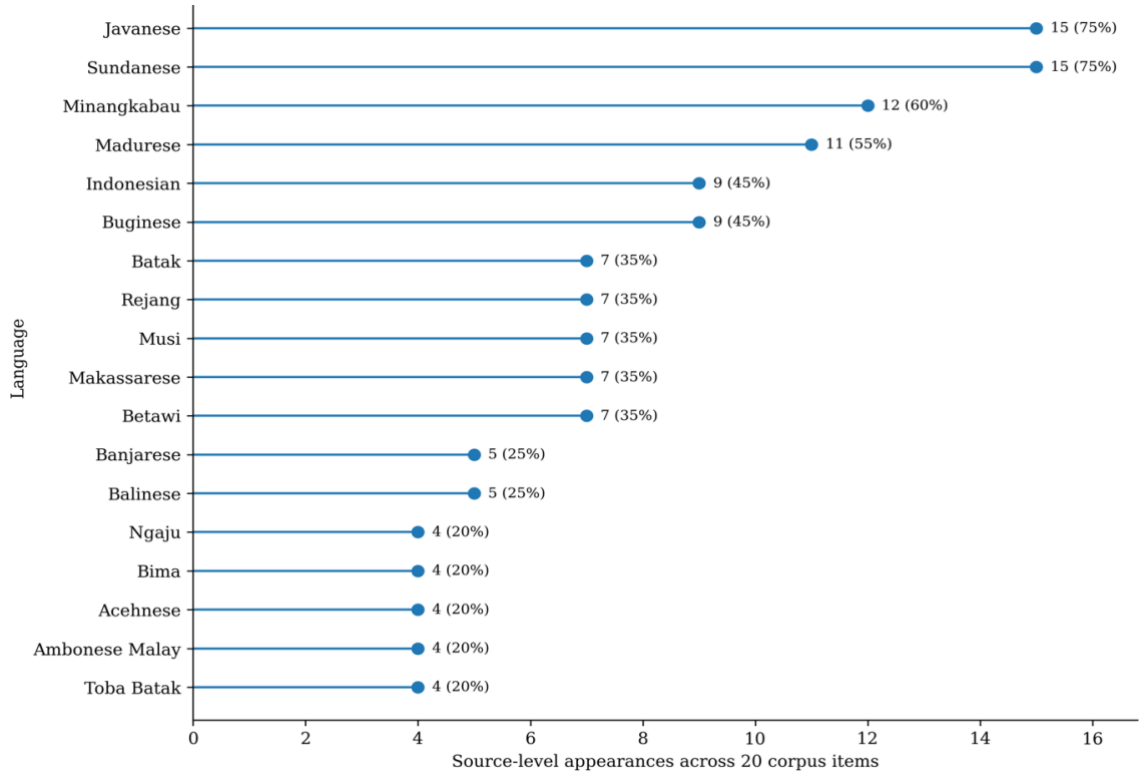

Figure 4. Language coverage across the corpus frame

5. DISCUSSION

The corpus-level evidence has a direct implication for how multilingual PEFT should be framed in Indonesian low-resource NLP. The main function of the assembled corpus is not merely to supply data for fine-tuning, but to define the evidential boundary within which claims about adaptation can be made. Because task-ready datasets account for the strongest part of the corpus frame, this study can reasonably investigate supervised adaptation for machine translation, sentiment classification, emotion classification, topic classification, and rhetorical-mode analysis. Its dysfunction lies in the uneven distribution of benchmarks, documentation, and auxiliary sources across languages. This matters because PEFT research often foregrounds algorithmic efficiency while treating data conditions as secondary. The present evidence suggests the opposite: adaptation efficiency is interpretable only when the corpus infrastructure is visible. This position supports recent work arguing that multilingual PEFT must be evaluated beyond aggregate task scores [8], [9], [18].

The underlying structure behind this pattern is a layered resource ecology rather than a single dataset problem. Indonesian NLP has developed through several partially connected infrastructures: task datasets, benchmark corpora, registry systems, dataset papers, and model-context collections. This explains why the corpus is strongest where public task datasets have been actively curated, but thinner where standardised evaluation or comparable documentation is needed. The relationship should not be read causally in a strict experimental sense; rather, source-role concentration is associated with what can and cannot be claimed from the adaptation pipeline. Work on IndoNLG, NusaX, NusaWrites, and Cendol shows similar asymmetry: these resources expand Indonesian multilingual NLP but do not remove differences in language coverage, task availability, or evaluation maturity [1], [5], [2], [4]. Thus, this study positions corpus adequacy as a methodological condition for PEFT interpretation.

The second implication concerns adaptation readiness. The results indicate that this study has a workable base for PEFT experiments, but not an equally strong base for every adaptation claim. The function of the corpus is strongest at the point of supervised adaptation: enough sources exist to prepare task-specific training or evaluation data. Its dysfunction appears when adaptation input is compared with benchmark comparability and reproducibility support. This matters because PEFT is frequently praised for reducing trainable parameters, memory burden, and storage cost, yet these efficiencies can be misleading when evaluation conditions are uneven. LoRA, QLoRA, adapters, and related methods may produce practical

gains, but their relevance to low-resource Indonesian languages depends on whether these gains can be compared under consistent corpus conditions. This finding aligns with Xu et al. (2023) [7], Pu et al. (2023) [16], and Aggarwal et al. (2024) [8], who caution against evaluating PEFT through performance alone.

The reason for this imbalance lies in the different temporalities of model engineering and dataset infrastructure. PEFT methods can be implemented rapidly once a model and dataset are available, but multilingual dataset ecosystems develop more slowly, through collection, documentation, licensing, annotation, benchmarking, and community validation. This structural gap helps explain why adaptation inputs may be more visible than robust evaluation layers. It also explains why resource-efficient methods do not automatically resolve low-resource conditions: they reduce the cost of parameter updates, not the cost of building comparable multilingual evidence. Studies on unseen-language adaptation and targeted multilingual adaptation reach a similar conclusion, showing that transfer depends on representation, vocabulary alignment, and task exposure, not only on the number of trainable parameters [10], [11], [12]. This study therefore treats PEFT readiness as an interaction between architecture, corpus quality, and evaluation design.

The third implication is theoretical: Indonesian local languages cannot be treated as a flat multilingual category. The language-coverage results show a clear stratification between broadly represented languages, moderately supported languages, and narrowly task-bound languages [24], [25]. The function of this stratification is analytically productive because it allows this study to compare different levels of transfer pressure. Javanese and Sundanese can support broader cross-task analysis, while languages such as Acehnese, Ngaju, Bima, Toba Batak, and Ambonese Malay are better positioned as focused stress cases for low-resource transfer. The dysfunction emerges when stronger languages are allowed to stand in for Indonesian local languages as a whole. This would reproduce a familiar problem in multilingual NLP: apparent inclusion at the language-list level while evaluation remains dominated by better-resourced languages. The corpus therefore supports a tiered interpretation of multilinguality rather than a celebratory count of covered languages.

This stratification is best explained by resource concentration, language visibility, and source-family dependence. Languages that appear across multiple source families have stronger evidential support because their performance can be interpreted across tasks and documentation layers. Languages concentrated in a single resource family may still be valuable, but their adaptation results must be read with greater caution. This pattern resonates with research on cross-lingual transfer showing that multilingual models often benefit languages with stronger representation, closer lexical or typological relations, and better vocabulary alignment [20], [21], [13], [14]. At the same time, it complicates optimistic claims in PEFT literature by showing that small adaptation modules do not neutralise unequal linguistic infrastructure. This study therefore contributes a restrained theoretical position: PEFT may make adaptation more feasible, but multilingual adequacy remains governed by corpus distribution, transfer asymmetry, and residual language-specific error.

6. CONCLUSION

This study shows that parameter-efficient adaptation for low-resource Indonesian languages should be understood first as a corpus-dependent methodological problem, not merely as a model-optimization problem. The most important lesson is that small multilingual models can only be meaningfully adapted when corpus adequacy, source provenance, language coverage, and evaluation readiness are made explicit. The strength of this study lies in shifting attention from aggregate multilingual performance to the conditions that make such performance interpretable. By combining corpus-role mapping, PEFT readiness analysis, and language-coverage stratification, this study updates the discussion on multilingual NLP by treating Indonesian local languages as unevenly resourced linguistic systems rather than as interchangeable low-resource labels.

This study is limited by its corpus-level design and does not yet report task-level model performance, memory use, training time, or residual error rates from completed PEFT experiments. Its conclusions therefore concern methodological readiness and evidential boundaries, not final claims about model superiority. Future research should implement controlled LoRA, QLoRA, adapter, and full fine-tuning comparisons across the selected languages and tasks, using consistent splits and transparent hyperparameters. Further work should also include qualitative error analysis, tokenizer diagnostics, and language-specific evaluation protocols, especially for languages with narrower coverage. Such research would clarify where small multilingual models genuinely support Indonesian linguistic diversity and where adaptation remains constrained by resource imbalance.

ACKNOWLEDGMENTS

The author thanks the reviewers and editorial team for their constructive feedback on this manuscript.

FUNDING INFORMATION

Author states no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

Ainul Hadziqi: conceptualization, methodology, formal analysis, writing – original draft, writing – review and editing.

CONFLICT OF INTEREST STATEMENT

Author states no conflict of interest.

AI USE DECLARATION

The author used artificial intelligence tools to assist with the generation of illustrative figures and with the formatting and layout of this manuscript. AI was not used to generate the substantive content, conceptualization, data, analysis, or conclusions of the research, all of which remain the sole responsibility of the author.

INFORMED CONSENT

Not applicable – this study did not involve direct human participants. Data consisted solely of publicly accessible corpora, benchmark resources, and documentation.

ETHICAL APPROVAL

This research did not involve human or animal subjects; it analyzed publicly accessible corpora, benchmark resources, and documentation, and therefore no institutional review board approval was required. Where research does involve human subjects, the author affirms that such research would be conducted in full compliance with all relevant national regulations and institutional policies, in accordance with the tenets of the Declaration of Helsinki, and would be approved by the author's institutional review board or equivalent committee.

DATA AVAILABILITY

The corpus of documents analyzed in this study are publicly accessible and available from the corresponding author upon reasonable request.

REFERENCES

[1] S. Cahyawijaya, G. I. Winata, B. Wilie, K. Vincentio, X. Li, A. Kuncoro, et al., "IndoNLG: Benchmark and resources for evaluating Indonesian natural language generation," *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 8875–8898*, 2021, doi: 10.18653/v1/2021.emnlp-main.699.

[2] J. A. Lopo and R. Tanone, "Constructing and expanding low-resource and underrepresented parallel datasets for Indonesian local languages," *ArXiv, abs/2404.01009*, 2024, doi: 10.48550/arxiv.2404.01009.

[3] W. Wongso, A. Joyoadikusumo, B. S. Buana, and D. Suhartono, "Many-to-many multilingual translation model for languages of Indonesia," *IEEE Access*, vol. 11, pp. 91385–91397, 2023, doi: 10.1109/access.2023.3308818.

[4] W. Tan and K. Zhu, "NusaMT-7B: Machine translation for low-resource Indonesian languages with large language models," *ArXiv, abs/2410.07830*, 2024, doi: 10.48550/arxiv.2410.07830.

[5] S. Cahyawijaya, H. Lovenia, F. Koto, R. A. Putri, E. Dave, J. Lee, et al., "Cendol: Open instruction-tuned generative large language models for Indonesian languages," *ArXiv, abs/2404.06138, 14899–14914*, 2024, doi: 10.48550/arxiv.2404.06138.

[6] "Compacter: Efficient low-rank hypercomplex adapter layers," , pp. 1022–1035, 2021.

[7] L. Xu, H. Xie, S. Qin, X. Tao, and F. Wang, "Parameter-efficient fine-tuning methods for pretrained language models: A critical review and assessment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 48, pp. 6107–6126, 2023, doi: 10.1109/tpami.2026.3657354.

[8] D. Aggarwal, A. Sathe, I. Watts, and S. Sitaram, "MAPLE: Multilingual evaluation of parameter efficient finetuning of large language models," *ArXiv, abs/2401.07598*, 2024, doi: 10.48550/arxiv.2401.07598.

[9] T. Su, X. Peng, S. Thillainathan, D. Guzmán, S. Ranathunga, and E.-S. A. Lee, "Unlocking parameter-efficient fine-tuning for low-resource language translation," *ArXiv, abs/2404.04212, 4217–4225*, 2024, doi: 10.48550/arxiv.2404.04212.

[10] P.-H. Chen and Y.-N. Chen, "Efficient unseen language adaptation for multilingual pre-trained language models," *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, 18983–18994*, 2024, doi: 10.18653/v1/2024.emnlp-main.1057.

[11] Z. Csaki, P. Pawakapan, U. Thakker, and Q. Xu, "Efficiently adapting pretrained language models to new languages," *ArXiv, abs/2311.05741*, 2023, doi: 10.48550/arxiv.2311.05741.

[12] C. M. Downey, T. Blevins, D. Serai, D. Parikh, and S. Steinert-Threlkeld, "Targeted multilingual adaptation for low-resource language families," *ArXiv, abs/2405.12413, 15647–15663*, 2024, doi: 10.48550/arxiv.2405.12413.

[13] D. Gurgurov, M. Hartmann, and S. Ostermann, "Adapting multilingual LLMs to low-resource languages with knowledge graphs via adapters," *Proceedings of the Workshop on Knowledge Augmented Methods for Natural Language Processing, 1–7*, 2024, doi: 10.18653/v1/2024.kallm-1.7.
[14] D. Gurgurov, I. Vykopal, J. Van Genabith, and S. Ostermann, "Small models, big impact: Efficient corpus and graph-based adaptation of small multilingual language models for low-resource languages," *ArXiv, abs/2502.10140, 355–395*, 2025, doi: 10.48550/arxiv.2502.10140.
[15] J. O. Alabi, D. I. Adelani, M. Mosbach, and D. Klakow, "Adapting pre-trained language models to African languages via multilingual adaptive fine-tuning," *Proceedings of the 29th International Conference on Computational Linguistics, 4336–4349*, 2022.
[16] G. Pu, A. Jain, J. Yin, and R. Kaplan, "Empirical analysis of the strengths and weaknesses of PEFT techniques for LLMs," *ArXiv, abs/2304.14999*, 2023, doi: 10.48550/arxiv.2304.14999.
[17] N. J. Prottasha, U. R. Chowdhury, S. Mohanto, T. Nuzhat, A. A. Sami, M. S. Ali, et al., "PEFT A2Z: Parameter-efficient fine-tuning survey for large language and vision models," *ArXiv, abs/2504.14117*, 2025, doi: 10.48550/arxiv.2504.14117.
[18] O. Khade, S. Jagdale, A. Phaltankar, G. Takalikar, and R. Joshi, "Challenges in adapting multilingual LLMs to low-resource languages using LoRA PEFT tuning," *ArXiv, abs/2411.18571*, 2024, doi: 10.48550/arxiv.2411.18571.
[19] R. Joshi, K. Singla, A. Kamath, R. Kalani, R. Paul, U. Vaidya, et al., "Adapting multilingual LLMs to low-resource languages using continued pre-training and synthetic corpus," *ArXiv, abs/2410.14815, 50–57*, 2024, doi: 10.48550/arxiv.2410.14815.
[20] S. Purkayastha, S. Ruder, J. Pfeiffer, I. Gurevych, and I. Vulić, "Romanization-based large-scale adaptation of multilingual language models," *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 7996–8005*, 2023, doi: 10.48550/arxiv.2304.08865.
[21] F. Remy, P. Delobelle, H. Avetisyan, A. Khabibullina, M. De Lhoneux, and T. Demeester, "Trans-tokenization and cross-lingual vocabulary transfers: Language adaptation of LLMs for low-resource NLP," *ArXiv, abs/2408.04303*, 2024, doi: 10.48550/arxiv.2408.04303.
[22] D. Isnaeni, N. Afra, I. Budi, and M. T. Uliniansyah, "Fine-tuning pretrained language models for extremely low-resource machine translation: The case of Makassarese-Indonesian," *2025 International Conference on Computer, Control, Informatics and Its Applications (IC3INA), 208–213*, 2025, doi: 10.1109/ic3ina68387.2025.11325367.
[23] J. Sha, M. Zhu, C. Feng, and Y. Shang, "VEEF-Multi-LLM: Effective vocabulary expansion and parameter efficient finetuning towards multilingual large language models," *7963–7981*, 2025.
[24] H. Kang, J. Ni, and H. Yao, "Ever: Mitigating hallucination in large language models through real-time verification and rectification," *arXiv*, 2023, doi: 10.48550/arXiv.2311.09114.
[25] G. Kim, J. Park, J. Kim, S. Han, K. Son, and I. Jang, "SERC: LDPC-inspired semantic error correction for retrieval-augmented generation," , 2026.

BIOGRAPHY OF AUTHOR

Ainul Hadziqi is affiliated with Universitas Sarjanawiyata Tamansiswa, Indonesia. His academic studies are related to language, sociolinguistics, and related humanities fields. His scholarly interests include media discourse, sociolinguistics, inequality narratives, language ideology, and public communication. His research frequently investigates how social issues are framed in media language and how discourse reflects broader structures of power and marginalization. He can be contacted at email: ainulh301008@ustjogja.ac.id