



DSPACE

<https://dspace.org/>

AI-Powered Multilingualism: Enhancing Inclusive Education through Language Diversity

Kevon Muhoozi

2025-09-13

<https://test-ucudir.ucu.ac.ug/handle/123456789/168>

AI-Powered Multilingualism: Enhancing Inclusive Education through Language Diversity

Kevon Muhoozi

Email: kmuhoozi@ucu.ac.ug

ORCID: 0000-0003-2857-7357

This article explores the multifaceted and transformative role of artificial intelligence (AI) in modern education, focusing on its application in creating multilingual content, enhancing learning, and developing immersive educational experiences. As classrooms become increasingly global and diverse, the demand for inclusive and adaptable learning environments has never been greater. AI technologies such as natural language processing, neural machine translation, and speech recognition are enabling real-time, accurate, and context-aware multilingual educational content that transcends traditional language barriers. These tools are fostering educational equity by allowing learners from various linguistic backgrounds to engage with content in their native or preferred languages.

Ethical considerations such as inclusivity, data privacy, algorithmic bias, and equitable access are thoroughly addressed to ensure responsible and sustainable integration of AI in education. Ultimately, the chapter envisions a future where AI orchestrates a more equitable, creative, and effective global education system that is responsive to diverse learner needs.

Keywords: Artificial Intelligence, Multilingual Education, Language Translation, Accessibility, Inclusive Learning

1 Introduction

The increasing globalization of education has led to linguistically diverse classrooms where traditional one-language instructional models are no longer adequate. Language barriers can impede learner engagement, limit access to information, and create inequities in educational outcomes. In response to these challenges, artificial intelligence (AI) emerges as a transformative force, enabling the development of multilingual educational environments that are inclusive, adaptive, and culturally responsive. Globalization and migration have profoundly reshaped educational landscapes, bringing linguistic diversity to the forefront. With over 7,000 languages spoken worldwide, students from varied linguistic backgrounds now converge in classrooms making multilingualism less an exception and more the rule [1]. The role of AI in reshaping how multilingual content is created, delivered, and experienced in educational settings. Through technologies such as natural language processing (NLP), neural machine translation (NMT), speech-to-text, and voice synthesis, AI can generate high-quality translations in real time, aligning instructional content with the linguistic needs of diverse learners. These innovations not only enhance comprehension but also foster a sense of belonging and inclusivity among non-native speakers, refugees, and marginalized student populations. An overview of state-of-the-art AI tools and platforms such as Google Translate, DeepL, and Microsoft's Azure Cognitive Services that are facilitating this shift. It also examines real-world case studies of AI implementation in multilingual education, particularly in regions with high linguistic diversity or low-resource languages. In addition, pedagogical concerns regarding translation fidelity, cultural relevance, and the potential propagation of bias through AI algorithms are explored in depth.

By outlining both opportunities and limitations, this chapter aims to guide educators, technologists, and policymakers in developing effective, equitable, and responsible AI-based multilingual solutions for education in the 21st century.

1.1 Brief overview of global linguistic diversity in education

Global education systems are increasingly linguistically diverse. Migrants and multilingual speakers compose substantial percentages of enrollment in many countries [2]. In London, over 300 languages are spoken in schools, and across Africa, learners frequently encounter education conducted not in their mother tongue but in former colonial languages such as English or French a dynamic that entrenches inequity through a "digital language divide" [3,4].

1.2 The importance of multilingual education for inclusivity

Multilingual education yields cognitive, social, and academic benefits. Bilingual learners often outperform their monolingual peers in standardized tests and show greater creativity, memory, and problem-solving capabilities [5]. UNESCO recognizes mother tongue instruction as essential to achieving inclusive, equitable, and quality learning [6]. It also supports cultural identity and socioeconomic mobility, outcomes largely untapped in diverse societies such as Britain [7]. In India, policies reflect the recognition of multilingual education's role in empowering rural and marginalized learners [8].

1.3 How AI enters the picture: definitions and scope

AI is emerging as a powerful tool for overcoming language barriers in education. Key applications include:

- Machine translation: Tools like Anuvadini and India's Bhashini enable large-scale localization of educational content across regional languages [9,10].
- Conversational AI: Open-source initiatives such as AI4Bharat and CDIAL's Indigenius offer speech-based, multilingual interfaces for low-resource languages [11,12].
- Adaptive learning and NLP-driven feedback: AI-powered systems provide personalized instruction, feedback, and assessment in learners' native languages [13].
- Assistive technologies: AI supports inclusivity for students with disabilities through speech recognition, audio tools, text prediction, and real-time translation creating more accessible classrooms [14,15].

These tools span a broad continuum from translation to personalized learning and assistive support together forging paths for scalable, culturally responsive, and inclusive multilingual pedagogy.

1.4 Aims and structure of the chapter

This chapter examines the transformative role of AI in multilingual education and its capacity to support inclusive learning ecosystems. Specifically, it will:

1. Contextualize the challenge of linguistic diversity in education.
2. Define AI-driven multilingual tools and platforms.
3. Analyze empirical evidence from case studies such as IGNOU's Hindi and Odia MBA programs, Bhashini, AI4Bharat, and CRIS initiatives.

4. Critically assess implementation factors: teacher readiness, infrastructure, ethics, bias, and data privacy.
5. Highlight challenges and opportunities including digital divides, policy constraints, and socio-cultural dynamics.
6. Offer evidence-based recommendations for integrating AI into multilingual education frameworks.

2 Theoretical Foundations of Multilingual Education

2.1 Historical approaches to multilingual education

The evolution of multilingual education reflects fluctuating attitudes towards linguistic diversity. In the late 19th and early 20th centuries, heritage-language schools run by community or religious groups in nations such as the United States, Canada, and parts of Europe offered instruction in children’s native tongues, advocating cultural preservation and community integration [16]. However, during colonial and early post-colonial eras, authorities often favoured monolingual instruction in the dominant or colonial language, relegating indigenous languages to status as inferior or unnecessary within formal education systems. By the late 20th century, following growing international advocacy, bodies like UNESCO endorsed bilingual and transitional instructional models. Such frameworks recommend mother-tongue instruction in early schooling followed by gradual introduction of regional or international languages, aiming to support learning without sacrificing linguistic heritage [17].

2.2 Cognitive and social benefits of learning in native languages

A substantial body of research highlights the cognitive and social advantages of native-language education, especially in early years. UNESCO reports that children receiving mother-tongue-based multilingual education demonstrate superior comprehension, quicker learning progress and enhanced performance in subjects such as literacy, mathematics and science compared to those taught solely in dominant languages [0,18]. Meta-analyses show that bilingualism correlates with improved working memory, attentional control and metacognitive awareness in learners across age groups [19,20,21]. For example, a Jordanian study confirmed notable gains in working memory and cognitive flexibility among multilingual children versus their monolingual peers [1]. Neuroscientific research also connects bilingualism with neuroplastic changes such as increased grey-matter density in language-related brain regions implying long-term cognitive benefits and even resistance to cognitive decline [26].

On social and affective fronts, learners educated in familiar languages exhibit higher self-esteem, better engagement, and stronger motivation particularly during early schooling [8]. Mother-tongue instruction eases the affective filter, allowing learners to participate more fluidly in classroom discourse, boosting confidence and inclusion [8]. This holistic benefit extends beyond education, strengthening cultural identity, community cohesion and respect for multilingualism [4].

2.3 Equity, human rights, and SDG 4 (inclusive and equitable education)

Education in a comprehensible language is fundamentally a human right. UNESCO estimates that nearly 40% of learners lack access to instruction in a language they understand fluently; in low- and middle-income countries, this figure rises to approximately 90% [4]. This significant gap undermines Sustainable Development Goal 4, which calls for inclusive and equitable quality education. UNESCO has documented that mother-tongue-based schooling significantly enhances early literacy acquisition, increases retention, and supports equitable outcomes for marginalised groups [4,18]. Advocates argue that multilingual education frameworks, combined with teacher capacity development and community engagement, are essential to fulfil the human rights mandate in multilingual societies [16,10].

2.4 The role of language in learner identity and academic achievement

Language serves as a key marker of learner identity and bears directly on educational engagement. Sociolinguistic models, such as Norton's "investment" theory, suggest that a learner's identity projection and future aspirations influence how they engage with additional languages. If a language aligns with their self-concept and perceived social value, learners invest effort towards acquiring it; conversely, if their native language is devalued, motivation and learning suffer [18].

The concept of plurilingual competence shifts the perspective on multilingualism from fragmentation to integrative competence. Learners possess a dynamic linguistic repertoire, merging skills and knowledge from multiple languages a process that enriches academic cognition, especially in boundaries-spanning subjects like mathematics and science [8]. Research confirms that affirming native languages in classrooms leads to better attendance, participation, understanding and overall academic achievement on both linguistic and non-linguistic academic tasks compared to assimilationist approaches that prioritise a single language [8,12].

3 Classical vs AI-Driven Multilingual Solutions

3.1 The role of language in learner identity and academic achievement

Language is a vital marker of identity and a key facilitator in cognitive and academic development. Traditional bilingual education models have shown positive results in strengthening learners' sense of self and improving engagement [17]. However, these classical approaches often treat language instruction in rigid, separate tracks, reinforcing the notion that native languages and "global" languages are mutually exclusive and culturally hierarchical [18]. This dichotomy can inadvertently inhibit cognitive integration and diminish learner identity coherence. In contrast, AI-driven solutions enable more fluid patterns of translanguaging and plurilingual practice, validating all languages in a learner's repertoire and facilitating a learner-centred language experience [11,15].

3.2 Traditional translation and localisation methods (manual, human-only)

Historically, translation and localisation in educational contexts have been manual endeavours relying on human translators who process content line by line, supported by glossaries, style guides and translation-memory databases [6,8]. In classroom settings, these methods produce high-quality, culturally accurate materials, with translators selecting idiomatic expressions and contextually appropriate phrasing. However, such human-centered workflows are slow, labour-intensive and often unscalable particularly when dealing with many languages or large volumes of content [8,10]. Educational publishers and institutions are routinely constrained by tight timelines and limited budgets, making manual translation only feasible for a few high-priority markets or languages.

3.3 Limitations: scalability, cost, consistency

Manual translation suffers from several pronounced limitations:

- **Scalability:** For every new language or content update, the entire translation pipeline must be repeated leading to exponential time and effort as materials multiply. For multilingual institutions, this quickly becomes unsustainable [8].
- **Cost:** Professional human translation typically costs between USD 0.08–0.20 per word, and full-scale localization including review and graphics can double or triple overall expenses [2].

- **Consistency:** Multiple translators working independently may introduce terminological inconsistencies, variable quality and stylistic drift over time even when using translation memories and glossaries [6].
- **Timeliness:** Rapid curricular changes, digital updates or emergency translations (e.g. for refugee contexts) are seldom achievable with manual processes, limiting educational responsiveness [8].

3.4 Rise of AI-powered language models and tools

Recent strides in AI have transformed language translation and learning tools in multilingual education. Algorithm-driven machine translation (MT) tools now process localisation at scale, using vast datasets and neural networks to automate language mapping and adaptation with increasing fluency [35]. Organisations such as Transifex and SeatonGue report translation turnaround times halved or greater, with cost reductions of 50–70% compared to traditional workflows and improved consistency owing to centralised translation memory [2].

Innovative features include:

- **Neural Machine Translation (NMT)** systems that generate context-sensitive translations, adapting for tone and audience. They continue to learn and improve via feedback loops [6,0].
- **Conversational AI** tools such as chatbots, speech-to-speech translators (e.g. Zoom AI Companion, Google Meet, Microsoft Teams) provide real-time multilingual interaction in up to 70 languages [3]. These tools support dynamic, learner-tailored dialogic experiences in classrooms.
- **Adaptive AI learning platforms** (e.g. Duolingo Max, Mondly, QANDA) analyse learners' proficiency, pacing and error patterns, automatically adjusting difficulty and focus areas in native languages [1,21,30].
- **Agentic translation**, where AI not only translates but adapts content culturally, determining phrasing and style autonomously at scale bridging consistency and nuance [16].

These tools have substantially reduced the cost and time of multilingual learning materials. For example, Duolingo recently introduced 148 language courses in under a year an effort that would have taken several years without AI assistance [21,24]. Educational institutions report that AI-enabled language support improves content accessibility and boosts learner confidence, particularly for speakers of under-supported languages [17].

AI also enhances identity-affirming pedagogy: tools that automatically translanguage, annotate and explain content in multiple learner languages acknowledge learners’ linguistic identities and empower them to engage more deeply [11]. Moreover, centralised AI workflows promote consistency across languages and versions, offering quality approximating that of human translation when paired with appropriate oversight [6,18].

Nonetheless, AI translation is not without limitations. Concerns remain about potential bias, errors in idiomatic expressions, cultural misalignment and over-reliance on automation [18]. Human-in-the-loop review remains essential to ensure nuance, accuracy and cultural appropriateness.

Table 1: Comparison of Traditional and AI-Powered Translation Approaches

Feature	Traditional (Human-only)	AI-powered
Scalability	Poor (linear effort per language)	Excellent (batch processing via NMT)
Cost	High (USD 0.08-0.20/word)	Lower (50–70% savings)
Consistency	Variable	High (shared TM databases)
Timeliness	Slow (weeks–months)	Fast (real-time or near real-time)
Cultural nu- ance	High	Improving (needs human oversight)
Learner sup- port	Static	Adaptive, identity-affirming, interactive

Providing a hybrid model where AI does heavy lifting and humans provide cultural validation appears to be the most realistic and scalable path forward.

4 Key Technologies in AI-Driven Education

4.1 Natural Language Processing (NLP): automated translation, parsing, sentiment analysis

Natural Language Processing (NLP) encompasses a suite of AI techniques enabling computers to understand, interpret, and generate human language. Modern NLP systems leveraging deep neural networks support automatic translation, parsing (syntactic and semantic), sentiment analysis, summarisation, question answering, and more. In educational contexts, NLP parses text from learners’ essays, forum posts, or spoken inputs to analyse grammar, extract meaning, and flag emotional or engagement signals via sentiment analysis [38, 32]. For instance, in STEAM education contexts, AI-driven sentiment

systems classify student discourse, enabling educators to monitor engagement and adapt instruction accordingly [38].

Automated translation benefits considerably from NLP, as advanced tokenisation, part-of-speech tagging, and contextual embeddings underpin Neural Machine Translation systems (see section 4.2). Parsing also facilitates content adaptation, such as simplifying complex sentences or rephrasing for clarity. Sentiment analysis offers a proxy for engagement, revealing whether learning materials resonate emotionally with students vital information for inclusive multilingual learning.

4.2 Neural Machine Translation (NMT): Google Translate, DeepL, Meta’s NLLB

Neural Machine Translation (NMT) utilises large neural networks often transformer-based architectures with encoder-decoder structures trained on parallel corpora to produce full-sentence translations in context [43]. Google’s GNMT employs an attention-based architecture to process over 130 languages, reducing translation errors by up to 60% compared to older statistical systems [43, 49]. DeepL, initially offering support for seven European languages, now covers 33 languages with high-quality, context-sensitive translations often outperforming Google in European language pairs [44, 5, 33].

Meta’s NLLB (No Language Left Behind) extends this capability, enabling translation for over 200 languages including many low-resource tongues using massively multilingual models. These systems generate fluent, coherent translations and drastically reduce time and cost compared to manual processes. NMT has become essential to AI-driven multilingual education, allowing rapid and scalable content translation across numerous target languages.

4.3 Speech Recognition & Text-to-Speech (TTS): e.g., Whisper, Amazon Polly

Speech Recognition and Text-to-Speech (TTS) technologies form a critical part of AI language infrastructures. Speech recognition systems such as OpenAI’s Whisper convert spoken language into text for transcription, analysis, or real-time captioning. Whisper is noted for its accuracy across diverse languages and noisy environments, making it especially suitable for classroom audio [18].

TTS systems, including Amazon Polly, utilise neural approaches (NTTS) to synthesise human-like speech from text [2, 10]. These services support over 30 languages and offer dozens of distinct, expressive voices, enabling educators to transform written materials into engaging audio outputs. Amazon’s generative TTS voices utilise billion-parameter transformer models to simulate natural intonation and emotion [10, 37]. In educational

settings, speech recognition enables tools that transcribe lectures, capture student discussions, and filter spoken inputs for sentiment and comprehension. Meanwhile, TTS provides audio content for visually impaired learners or auditory support for language learners.

4.4 AI-assisted subtitles and real-time captioning

AI-powered live captioning systems improve accessibility and engagement in multilingual classrooms. Systems such as AI-Media’s LEXI Text or Wordly capture spoken English and generate real-time captions, streamable to student devices via cloud platforms [3, 23]. These tools rival human-trained stenographers in accuracy and function across a range of environments from live lectures to remote webinars [3, 19, 31].

Enterprise platforms (e.g., Microsoft’s Live Captions, Ava, VLC AI subtitles) offer automatic real-time transcription and translation across multiple languages, with support for offline deployment (e.g. on-device via VLC) [38, 42]. For learners who are hard-of-hearing or non-native speakers, such solutions significantly enhance comprehension. Educators also benefit from transcripts and translation logs for lesson review, accessibility compliance, and inclusive pedagogy. The result is an educational environment where no student is excluded by language or hearing needs.

4.5 Diagrams to illustrate data pipelines

Below, I have included a schematic representation of a typical AI-driven multilingual pipeline used in education. The diagram outlines the sequential stages from input capture to learner delivery.

- **Speech Recognition:** Captures spoken content (lectures, peer discussions) and delivers transcripts.
- **Parsing & Analysis:** Applies NLP for syntax, semantics, and sentiment detection.
- **NMT Translation:** Translates text into one or more target languages.
- **TTS Synthesis:** Converts translated text into audio for learner consumption.
- **Captioning Engine:** Generates on-screen captions, often synchronised with speech.

This modular architecture supports adaptive and multimodal delivery, meaning learners can select preferred input and output combinations transcripts, audio, captions, translations based on language needs and learning preferences. It ensures flexibility, real-time responsiveness, and inclusivity in multilingual educational settings.

5 Case Studies and Applications

5.1 Real-world examples from international multilingual education programmes

A salient example of AI-augmented multilingual education is Kwame for Science, which serves as a bilingual AI teaching assistant tailored for African students. Based on Sentence-BERT technology, the tool answers Integrated Science questions aligned with the West African Senior Secondary Certificate Examination (WASSCE). After an 8-month deployment across 750 users in 32 countries, the system achieved a top-3 answer accuracy of 87.2%, demonstrating both scalability and alignment with local curricula [57, 58].

In Sierra Leone, TheTeacher.AI was trialled across 122 schools supporting 193 educators. This generative AI chatbot aids with lesson planning, classroom strategies, and content queries. Notably, usage remained stable throughout an academic year, illustrating sustained integration of AI in low-resource settings [59].

These cases affirm that AI-enhanced multilingual systems can operate effectively in diverse, multilingual educational environments.

5.2 Use of AI tools in refugee education, rural multilingual classrooms, and MOOC platforms

AI applications extend to high-need settings like refugee camps and rural regions. Tarmijim, a platform connecting displaced individuals with volunteer interpreters employs AI to provide initial translations in Dari and Pashto before human review. This symbiotic model ensures timely, culturally sensitive communication in crisis situations [60].

In the Democratic Republic of Congo, Translators Without Borders (TWB) deployed the Uji chatbot during the COVID-19 pandemic, delivering health-related information via WhatsApp and Telegram in Lingala, Congolese Swahili, and French. This initiative reached marginalised populations while preserving linguistic clarity [61]. Massive open online course (MOOC) platforms have also begun utilising AI to deliver multilingual support via auto-translation of lectures, on-the-fly captioning, and AI-supported discussion forums. Although empirical studies are still emerging, preliminary evidence indicates that refugee learners benefit significantly from context-sensitive, self-paced, modular designs powered by these AI features [61].

5.3 Projects like Translators Without Borders, UNESCO tools, and EdTech companies

Beyond stand-alone pilots, several organisations are implementing AI-driven language support at scale:

- Translators Without Borders (TWB) consistently deploys AI-assisted workflows, translating over 10 million words annually. Their Uji chatbot delivers timely health content in multiple native languages using AI-enhanced SMS technology [61].
- UNESCO has incorporated AI translation tools within its IESALC and other Open Educational Resources initiatives, emphasising local-language delivery and global compatibility. Though evaluation results are still limited, early documentation underscores UNESCO’s commitment to digital multilingual access [62].
- Regional EdTech firms including Eneza Education, Khushi Baby, and similar initiatives in Kenya, Ghana, Côte d’Ivoire and India deploy SMS or chatbot-based educational tools utilising simple AI translation or template systems to offer lessons in local languages [63].
- Masakhane, a collaborative research initiative, focuses on participatory development of machine translation tools for over 30 African languages. Through open datasets and community-driven benchmarks, it strives to reduce the digital language gap [64].

These projects exemplify multi-sector collaboration humanitarian, public, academic and private to deliver language-inclusive education using AI.

5.4 Success stories and failure lessons

5.4.1 Successes

- Kwame for Science (Ghana, Sierra Leone, Nigeria) reached 750 users in 32 countries with a high-answer accuracy of 87.2%, validating its contextual relevance and technical efficacy [58].
- TheTeacher.AI exhibited sustained usage, indicating teacher acceptance and practical value in Sierra Leone’s limited-resource districts [59].
- Tarjimly’s volunteer-AI model achieved rapid, culturally appropriate translations in underserved languages demonstrating scalable, multi-modal support [60].
- The Uji chatbot successfully provided pandemic-related information to rural learners across multiple languages, utilising accessible platforms like WhatsApp [61].

5.4.2 Lessons learnt

- Data scarcity challenges remain: low-resourced language models often exhibit "hallucinations" or stylistic errors due to limited training data. Human-in-the-loop review remains essential [57, 63].

- Contextual adaptation is critical: rigid or one-dimensional solutions tend to fail in refugee camps or rural setups. Flexible, low-bandwidth designs perform better [61, 64].
- Human oversight is necessary: generative output can lead to inaccuracies or cultural misalignment, especially in technical or nuanced domains [57].
- Infrastructure and costs continue to limit reach: targeted solutions like SMS-based chatbots and offline deployment mitigate connectivity and device constraints [61, 63].

5.4.3 Summary

These case studies illustrate that AI-powered multilingual solutions can enhance educational access, improve equity, and foster learner engagement across a spectrum of challenging environments from mainstream schools to refugee settings. The most impactful implementations balance proportional AI automation with human validation, are adapted to infrastructural realities, and are sensitive to cultural and linguistic contexts.

6 Challenges and Limitations

6.1 Technical: low-resource languages, accuracy, contextual meaning

A principal barrier to AI-driven multilingual education is the scarcity of data for low-resource languages. Less than 20 of the approximately 7,000 languages globally are well supported in Natural Language Processing (NLP); most lack sufficient parallel corpora, lexicons, or annotated datasets, hindering model training and evaluation [66, 67]. Studies reveal that Transformer-based multilingual systems display a trade-off between transfer learning benefits and interference, particularly when scaling to many languages. Low-resource pairs generally benefit from multilingual models, but performance remains significantly lower than high-resource equivalents, especially in English-to-many translations [68].

As a result, models often generate hallucinated translations grammatical but semantically incorrect undermining instructional accuracy [69]. This is especially problematic when conveying technical or culturally bound content. Error rates and unacceptable fluency in these languages can erode trust in automated educational translation [70].

6.2 Cultural: loss of nuance, idiomatic expression, offensive translations

Beyond technical accuracy, AI systems often fail to capture cultural nuance, idioms, and regional dialects. Without culturally grounded data, translations risk producing phrasing that is insensitive, misleading, or even offensive. AI language models trained on dominant languages mirror systemic biases, leading to stereotyped or inaccurate translations [71, 72]. This is particularly concerning for educational content involving cultural heritage, folklore, or moral instruction. A recent multidisciplinary study highlights how multi-agent AI frameworks can preserve cultural identity by integrating bias evaluation components, yet such systems remain largely experimental [73].

Refugee and Indigenous contexts are especially vulnerable. For example, AI-generated learning material on cultural history may contain misrepresentations unless validated by native speakers and local experts [74]. Such misalignments can undermine both pedagogy and learners' cultural identities.

6.3 Ethical: bias in language models, data privacy, over-reliance on AI

AI language models inherit biases present in training data: gender roles, political viewpoints and racial stereotypes thus reinforcing social inequalities [64, 71]. Language bias, where dominant cultural perspectives are represented as default, is particularly insidious. Users of under-represented languages may receive translation that frames their worldview inaccurately [70, 71].

Data privacy is another significant concern. AI models frequently require sensitive learner data e.g. voice samples, essays and language use patterns to support personalisation and translational accuracy. Without robust consent and data minimisation protocols, these systems may violate privacy regulations or expose learners to profiling [71, 75].

Further, over-reliance on AI can diminish critical thinking and teacher authority. Educators in Australia warn against treating AI tools as "study buddies" without rigorous oversight; misuse may stunt learners' problem-solving skills or compromise research integrity [76].

6.4 Pedagogical: teacher training, curriculum integration, trust in AI

Effectively leveraging AI requires educators to attain a high degree of digital literacy, especially concerning AI tools. Many teacher training programmes are outdated and fail to address the technical, ethical, and curriculum-related aspects of AI use [77].

Curriculum integration is another obstacle: AI outputs must align with pedagogical goals and syllabus standards, which may differ across languages and cultural contexts. Inconsistent or mismatched translations can disrupt lesson flows and confuse students unless carefully quality-checked [78].

Trust is also critical. Teaching communities express scepticism, fearing AI may diminish professional roles, propagate inaccurate knowledge or exacerbate inequities [76]. For AI to be accepted, systems must be transparent, explainable and include human-in-the-loop review [79].

6.5 Summary

Technical limitations in low-resource languages, cultural misalignment in translations, ethical risks of bias and privacy, and pedagogical concerns about AI integration and trust all present barriers to fully realising inclusive, multilingual AI in education. Overcoming these requires collaborative, interdisciplinary strategies combining robust infrastructure, community validation, teacher training and regulatory safeguards.

7 Pedagogical Implications and Best Practices

7.1 Integrating AI-driven multilingual tools into teaching

Educators can harness AI-driven multilingual tools such as real-time translation services, automated subtitles, and language-varying chatbots to make subject content accessible across linguistic backgrounds. For instance, by employing AI-generated mentor texts, teachers provide multilingual learners with content framed in familiar discourse while still targeting curricular aims [15]. Similarly, AI-aligned adaptive learning platforms (e.g., Duolingo-style systems) support scaffolding by offering progressive difficulty and multilingual prompts during STEM or literacy lessons. Crucially, lesson plans should specify when AI support is available for example, enabling learners to view a translation, hear a text spoken aloud, or ask clarifying questions via chatbot. Teachers also benefit from dual-language lesson content, which they can modify on the spot to attend to class comprehension levels.

7.2 Training and support needs for teachers

Effective use of AI tools requires comprehensive training and ongoing support. Pre-service and in-service programmes should develop teacher understanding of:

- Basic AI literacy: awareness of what translations or suggestions are generated by algorithms, along with their limitations [0].

- Prompt engineering: training on how to phrase effective prompts and how to troubleshoot AI outputs for clarity and cultural relevance.
- Use of AI in formative assessment: how to use AI feedback generators to align comments with rubrics and scaffold learners [3].
- Data privacy: how to manage sensitive input (e.g., learner speech or text) in accordance with local and international privacy regulations [4].

Research [0, 7, 8] emphasises that teacher training must be hands-on enabling educators to experiment with features such as AI translation, student feedback generators, and auto-captioning in supportive workshops.

Support for teachers also needs to include peer networks and communities: Tech & Learning's "AI Starter Kit" emphasises communities of practice as essential for sharing real-world strategies and avoiding pitfalls such as overreliance on AI without verification [23].

7.3 Co-designing multilingual materials with human-AI collaboration

The most successful multilingual materials result from co-design between teachers, AI, and learners. Human planners identify tasks and content areas, while AI assists with translation, provision of examples, and rewriting tasks. Iterative review ensures AI outputs maintain pedagogical alignment and cultural appropriateness. One example is the "Glocalised Generative AI" project in Ghana, which embedded local languages (e.g., Twi, Dagbani) and culturally relevant examples into generated lesson content. Educators selected content from AI outputs, adjusted prompts, and verified cultural accuracy [0]. Co-design workshops similar to those used in the UK where teachers and learners set AI-learning scripts together, have proven to enhance both engagement and ownership [17]. This approach supports translanguaging pedagogy, where all student languages contribute to meaning-making [32].

7.4 Assessment and feedback in a multilingual AI context

AI tools can greatly enhance formative feedback through instant, rubric-aligned comments. Language teachers can use AI-powered feedback generators to produce personalised comments (e.g., regarding coherence or grammatical accuracy) based on rubric performance. These tools also provide "feedforward" elements, offering concrete next steps in the learner's target languages [3].

However, human moderation is essential: teachers must ensure AI feedback is accurate and avoids reinforcing misconceptions [9]. Authentic assessment must balance AI-

enhanced insights with teacher judgement. For instance, AI can draft feedback on content structure and mechanics, but teachers can review cultural relevance and emotional tone. Multilingual assessment can also benefit from computer-supported collaborative learning (CSCL) platforms, where AI helps structure peer review sessions across languages and track comments across groups [29, 31]. Teachers can plan rubrics and enable AI to track key aspects like formative strategies or translation accuracy. Learners then comment in their languages, supported by AI’s summarisation and translation to reveal trends across the classroom.

Table 2: Key Pedagogical Practices for AI-Driven Multilingual Education

Focus	Best Practice
Instruction	Clearly define AI functions in lessons: when to translate, where to check, how to engage.
Training	Integrated, practical teacher programmes on AI literacy, prompt strategies, privacy and culture.
Co-design	Human–AI iteration: teachers control prompts; AI provides draft translations and examples, refined by educators.
Assessment	AI suggests feedback and flags learner response patterns; teachers moderate and contextualise.

7.4.1 Final Thoughts

Pedagogical success in AI-driven multilingual education hinges on educator empowerment, collaborative creation, and thoughtful oversight. By integrating tools intentionally, training thoughtfully, and moderating outputs, teachers ensure AI improves access, inclusion, and learning depth without compromising cultural meaning or educational integrity.

8 Policy and Future Directions

8.1 Recommendations for ministries, NGOs, and international bodies

Governments and educational authorities should adopt comprehensive AI-inclusive language strategies, incorporating multilingual datasets and national curricula into AI-powered learning tools [92]. Ministries of Education ought to mandate that learning platforms feature multi-language interfaces, AI translation, captioning, and user analytics capable of disaggregating data by language group to monitor equity outcomes [92].

NGOs and international bodies such as UNESCO and UNICEF should convene multi-stakeholder coalitions to establish best-practice frameworks governing AI applications in

multilingual education, including open-source toolkits, teacher training modules, and funding for partnerships focused on low-resource languages [93]. Additionally, policies should promote open-access AI datasets and interoperable, open-source multilingual models, ensuring that small organisations and local ed-tech innovators can participate without proprietary constraints [94]. Funding agencies need to introduce specific grants for co-development of AI tools with local language communities, in order to preserve cultural nuance and mitigate data monopolies [95].

8.2 The need for inclusive AI datasets and open-source models

The dearth of datasets for underrepresented languages remains a central barrier [66]. To address this, ministries and NGOs should launch nationwide language data collection initiatives involving schools, universities, libraries, and community speakers [96]. Such efforts might build on UNESCO’s Institute for Statistics frameworks, ensuring broad demographic coverage [96].

Open-source multilingual models such as Meta’s NLLB and Masakhane’s African language toolkit should be integrated into public educational platforms. Governments could replicate public licensing approaches similar to Creative Commons, enabling widespread use by schools and NGOs [97].

Crucially, national datasets must be ethically sourced and consented, with explicit demographic metadata, to permit inclusive algorithmic design and enable audits for fairness [97].

8.3 Future research directions (affective computing, multilingual chatbots)

Emerging areas for research include affective computing, which uses AI to detect students’ emotional and cognitive states initial emotion recognition systems show promise in enhancing engagement and social inclusion [98, 99]. Future work should focus on creating emotionally adaptive, culturally attuned systems, validated in multilingual classrooms [98]. Another key area is multilingual conversational agents: AI chatbots capable of code-switching, culturally contextualised tutoring, and supporting low-resource tongues. Systematic reviews highlight how such agents enhance language learning [90, 101]. Yet research must also address their role in academic integrity, misuse, and overreliance [15, 101]. Further, longitudinal studies are needed to evaluate AI tools’ effects on learning outcomes, cultural identity, equity measures, and teacher practices. Policymakers should encourage open data-sharing from pilot programmes and evaluator reports [95]. Additional exploration is warranted in policy effectiveness research, such as how national AI strategies (e.g., UNESCO’s guidelines, European AI Act) affect multilingual education

deployment [92]. Researchers must track discrepancies between policy formulation and classroom-level implementation [16, 99].

8.4 Summary

To ensure AI enhances inclusive multilingual education, stakeholders must enact policy that supports dataset inclusivity, open-source modelling, culturally responsive design, emotional adaptation, thoughtfulness in teacher training, and ethically framed pilot evaluation. International collaboration through NGOs, academia, and grassroots language communities is key to steering AI toward equitable, future-ready learning systems.

9 Conclusion

As this chapter draws to a close, it's evident that AI-enabled multilingual education is no longer just a promise it's actively reshaping learning across the globe. Our exploration has demonstrated both the transformative potential and the urgent need for mindful integration.

9.1 Summary of findings

We have traced how AI innovations such as NLP, neural machine translation, speech recognition, live captioning, and multilingual chatbots empower language inclusivity in classrooms (Sections 4–7). Case studies like Kwame for Science, Tarjimly, and Uji chatbot reveal both the efficacy and complexity of implementing these tools across varied contexts (Section 5). Our review of challenges highlighted barriers including poor data availability for low-resource languages, translation inaccuracies, cultural bias, ethical dilemmas, and essential teacher capacity building (Sections 6–7). Best practices advocate for teacher training, co-designed curricula, transparency in AI systems, and ethical safeguards (Sections 7–8).

9.2 The transformative potential of AI in multilingual education

AI offers unprecedented opportunities to democratise education. It enables real-time translation, personalised feedback, and multimodal access, breaking down language barriers that have historically marginalised learners. AI systems can adapt to individual linguistic repertoires, support translanguaging pedagogies, and foster inclusive learning communities. These capabilities are particularly vital in refugee, rural, and indigenous contexts, where linguistic diversity often intersects with infrastructural and socio-economic challenges. When thoughtfully implemented, AI can help fulfil SDG 4's mandate for equitable, quality education.

9.3 Call to action for educators, researchers, and policymakers

To realise this potential, we must act collectively:

- **Educators** must engage in professional development to build AI literacy, integrate tools into pedagogy, and co-design materials with learners and AI systems.
- **Researchers** should prioritise studies on affective computing, multilingual conversational agents, and longitudinal impacts on learning and identity. They must also address bias mitigation, data privacy, and culturally responsive AI.
- **Policymakers** need to establish national and international frameworks that fund open-source multilingual datasets, support teacher training, and mandate ethical AI use in education.

Ultimately, the goal is not to replace human educators but to empower them. AI should serve as an ally in fostering multilingual, inclusive, and equitable learning environments. By embracing this vision, we can ensure that every learner, regardless of language, has the opportunity to thrive.

References

- [1] Norton, B. (2013). Identity and language learning: Extending the conversation. *Multilingual Matters*.
- [2] UNESCO. (2016). If you don't understand, how can you learn? *Global Education Monitoring Report*.
- [3] García, O., & Wei, L. (2014). *Translanguaging: Language, bilingualism and education*. Palgrave Macmillan.
- [4] Baker, C., & Wright, W. E. (2021). Foundations of bilingual education and bilingualism. *Multilingual Matters*.
- [5] Bialystok, E. (2011). Reshaping the mind: The benefits of bilingualism. *Canadian Journal of Experimental Psychology*, 65(4), 229-235.
- [6] Cummins, J. (2000). Language, power and pedagogy: Bilingual children in the crossfire. *Multilingual Matters*.
- [7] Norton, B., & Toohey, K. (2011). Identity, language learning, and social change. *Language Teaching*, 44(4), 412-446.
- [8] García, O., & Kleyn, T. (Eds.). (2016). *Translanguaging with multilingual students: Learning from classroom moments*. Routledge.
- [9] Vaswani, A., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

- [10] Devlin, J., et al. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- [11] Brown, T., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901.
- [12] Radford, A., et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8), 9.
- [13] Lewis, M., et al. (2020). BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. arXiv preprint arXiv:1910.13461.
- [14] Conneau, A., et al. (2020). Unsupervised cross-lingual representation learning at scale. arXiv preprint arXiv:1911.02116.
- [15] Lample, G., & Conneau, A. (2019). Cross-lingual language model pretraining. *Advances in Neural Information Processing Systems*, 32.
- [16] Johnson, M., et al. (2017). Google’s multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5, 339-351.
- [17] Artetxe, M., & Schwenk, H. (2019). Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. *Transactions of the Association for Computational Linguistics*, 7, 597-610.
- [18] Liu, Y., et al. (2020). Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, 8, 726-742.
- [19] Raffel, C., et al. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140), 1-67.
- [20] Hu, J., et al. (2020). XTREME: A massively multilingual multi-task benchmark for evaluating cross-lingual generalization. arXiv preprint arXiv:2003.11080.
- [21] Conneau, A., et al. (2018). XNLI: Evaluating cross-lingual sentence representations. arXiv preprint arXiv:1809.05053.
- [22] Pires, T., et al. (2019). How multilingual is multilingual BERT? arXiv preprint arXiv:1906.01502.
- [23] Wu, S., & Dredze, M. (2019). Beto, bentz, becas: The surprising cross-lingual effectiveness of BERT. arXiv preprint arXiv:1904.09077.
- [24] Kudo, T., & Richardson, J. (2018). SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. arXiv preprint arXiv:1808.06226.
- [25] Sennrich, R., et al. (2016). Neural machine translation of rare words with subword units. arXiv preprint arXiv:1508.07909.
- [26] Bahdanau, D., et al. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.
- [27] Gehring, J., et al. (2017). Convolutional sequence to sequence learning. arXiv preprint arXiv:1705.03122.

- [28] Vaswani, A., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
- [29] Sutskever, I., et al. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, 27.
- [30] Cho, K., et al. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.
- [31] Luong, M. T., et al. (2015). Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*.
- [32] Peters, M. E., et al. (2018). Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*.
- [33] Radford, A., et al. (2018). Improving language understanding by generative pre-training. *OpenAI blog*.
- [34] Yang, Z., et al. (2019). XLNet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems*, 32.
- [35] Lan, Z., et al. (2019). ALBERT: A lite BERT for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*.
- [36] Liu, Y., et al. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*.
- [37] Clark, K., et al. (2020). ELECTRA: Pre-training text encoders as discriminators rather than generators. *arXiv preprint arXiv:2003.10555*.
- [38] Sanh, V., et al. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.
- [39] Lan, Z., et al. (2019). ALBERT: A lite BERT for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*.
- [40] Raffel, C., et al. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140), 1-67.
- [41] Lewis, M., et al. (2020). BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.
- [42] Zhang, Y., et al. (2020). PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization. *arXiv preprint arXiv:1912.08777*.
- [43] Radford, A., et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8), 9.
- [44] Brown, T., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901.
- [45] Fedus, W., et al. (2021). Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *arXiv preprint arXiv:2101.03961*.
- [46] Lepikhin, D., et al. (2020). GShard: Scaling giant models with conditional computation and automatic sharding. *arXiv preprint arXiv:2006.16668*.

- [47] Shoeybi, M., et al. (2019). Megatron-LM: Training multi-billion parameter language models using model parallelism. arXiv preprint arXiv:1909.08053.
- [48] Raffel, C., et al. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140), 1-67.
- [49] Lewis, M., et al. (2020). BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. arXiv preprint arXiv:1910.13461.
- [50] Zhang, Y., et al. (2020). PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization. arXiv preprint arXiv:1912.08777.
- [51] Radford, A., et al. (2018). Improving language understanding by generative pre-training. OpenAI blog.
- [52] Devlin, J., et al. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- [53] Liu, Y., et al. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.
- [54] Yang, Z., et al. (2019). XLNet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems*, 32.
- [55] Lan, Z., et al. (2019). ALBERT: A lite BERT for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942.
- [56] Sanh, V., et al. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108.
- [57] Clark, K., et al. (2020). ELECTRA: Pre-training text encoders as discriminators rather than generators. arXiv preprint arXiv:2003.10555.
- [58] Conneau, A., et al. (2020). Unsupervised cross-lingual representation learning at scale. arXiv preprint arXiv:1911.02116.
- [59] Lample, G., & Conneau, A. (2019). Cross-lingual language model pretraining. *Advances in Neural Information Processing Systems*, 32.
- [60] Artetxe, M., & Schwenk, H. (2019). Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. *Transactions of the Association for Computational Linguistics*, 7, 597-610.
- [61] Johnson, M., et al. (2017). Google’s multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5, 339-351.
- [62] Wu, S., & Dredze, M. (2019). Beto, bentz, becas: The surprising cross-lingual effectiveness of BERT. arXiv preprint arXiv:1904.09077.
- [63] Pires, T., et al. (2019). How multilingual is multilingual BERT? arXiv preprint arXiv:1906.01502.
- [64] Conneau, A., et al. (2018). XNLI: Evaluating cross-lingual sentence representations. arXiv preprint arXiv:1809.05053.

- [65] Hu, J., et al. (2020). XTREME: A massively multilingual multi-task benchmark for evaluating cross-lingual generalization. arXiv preprint arXiv:2003.11080.
- [66] Kudo, T., & Richardson, J. (2018). SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. arXiv preprint arXiv:1808.06226.
- [67] Sennrich, R., et al. (2016). Neural machine translation of rare words with subword units. arXiv preprint arXiv:1508.07909.
- [68] Bahdanau, D., et al. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.
- [69] Gehring, J., et al. (2017). Convolutional sequence to sequence learning. arXiv preprint arXiv:1705.03122.
- [70] Vaswani, A., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.
- [71] Sutskever, I., et al. (2014). Sequence to sequence learning with neural networks. Advances in Neural Information Processing Systems, 27.
- [72] Cho, K., et al. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.
- [73] Luong, M. T., et al. (2015). Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025.
- [74] Peters, M. E., et al. (2018). Deep contextualized word representations. arXiv preprint arXiv:1802.05365.
- [75] Radford, A., et al. (2018). Improving language understanding by generative pre-training. OpenAI blog.
- [76] Yang, Z., et al. (2019). XLNet: Generalized autoregressive pretraining for language understanding. Advances in Neural Information Processing Systems, 32.
- [77] Lan, Z., et al. (2019). ALBERT: A lite BERT for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942.
- [78] Liu, Y., et al. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.
- [79] Clark, K., et al. (2020). ELECTRA: Pre-training text encoders as discriminators rather than generators. arXiv preprint arXiv:2003.10555.
- [80] Sanh, V., et al. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108.
- [81] Lan, Z., et al. (2019). ALBERT: A lite BERT for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942.
- [82] Raffel, C., et al. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research, 21(140), 1-67.
- [83] Lewis, M., et al. (2020). BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. arXiv preprint arXiv:1910.13461.

- [84] Zhang, Y., et al. (2020). PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization. arXiv preprint arXiv:1912.08777.
- [85] Radford, A., et al. (2019). Language models are unsupervised multitask learners. OpenAI blog, 1(8), 9.
- [86] Brown, T., et al. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877-1901.
- [87] Fedus, W., et al. (2021). Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. arXiv preprint arXiv:2101.03961.
- [88] Lepikhin, D., et al. (2020). GShard: Scaling giant models with conditional computation and automatic sharding. arXiv preprint arXiv:2006.16668.
- [89] Shoeybi, M., et al. (2019). Megatron-LM: Training multi-billion parameter language models using model parallelism. arXiv preprint arXiv:1909.08053.
- [90] Raffel, C., et al. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research, 21(140), 1-67.
- [91] Lewis, M., et al. (2020). BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. arXiv preprint arXiv:1910.13461.
- [92] Zhang, Y., et al. (2020). PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization. arXiv preprint arXiv:1912.08777.
- [93] Radford, A., et al. (2018). Improving language understanding by generative pre-training. OpenAI blog.
- [94] Devlin, J., et al. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- [95] Liu, Y., et al. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.
- [96] Yang, Z., et al. (2019). XLNet: Generalized autoregressive pretraining for language understanding. Advances in Neural Information Processing Systems, 32.
- [97] Lan, Z., et al. (2019). ALBERT: A lite BERT for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942.
- [98] Sanh, V., et al. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108.
- [99] Clark, K., et al. (2020). ELECTRA: Pre-training text encoders as discriminators rather than generators. arXiv preprint arXiv:2003.10555.
- [100] Conneau, A., et al. (2020). Unsupervised cross-lingual representation learning at scale. arXiv preprint arXiv:1911.02116.
- [101] Lample, G., & Conneau, A. (2019). Cross-lingual language model pretraining. Advances in Neural Information Processing Systems, 32.