

Commentary

Foundation Models for Low-Resource Language Education (Vision Paper)

Zhaojun Ding¹, Zhengliang Liu¹, Hanqi Jiang¹, Yizhu Gao¹, Xiaoming Zhai¹, Tianming Liu¹, Ninghao Liu¹

1. University of Georgia, United States

Recent studies show that large language models (LLMs) are powerful tools for working with natural language, bringing advances in many areas of computational linguistics. However, these models face challenges when applied to low-resource languages due to limited training data and difficulty in understanding cultural nuances. Research is now focusing on multilingual models to improve LLM performance for these languages. Education in these languages also struggles with a lack of resources and qualified teachers, particularly in underdeveloped regions. Here, LLMs can be transformative, supporting innovative methods like community-driven learning and digital platforms. This paper discusses how LLMs could enhance education for low-resource languages, emphasizing practical applications and benefits.

1. Introduction

Recent research in large language models (LLMs) has demonstrated their profound capabilities in processing and generating natural language, leading to advancements across various domains of computational linguistics^{[1][2][3][4][5][6]}. Despite these advancements, applying LLMs in the study of low-resource languages presents significant challenges. These models often struggle with the limited availability of training data, which can lead to poorer performance and less effective language understanding and generation. Additionally, the complexities of accurately capturing the nuances and cultural contexts of less-documented languages further complicate their use in this area. To tackle the problem, recently multilingual models have been studied, extending language models to low-resource languages^{[7][8]}.

Education in the domain of low-resource languages often face significant challenges due to limited access to teaching materials, qualified instructors, and formal language programs^[9]. Typically, these languages are spoken in regions where educational infrastructure is less developed, making traditional classroom-based learning difficult to implement. As a result, generative AI tools such as LLMs have the potential to revolutionize educational initiatives^{[10][11][12][13]} such as community-led classes, oral tradition, and providing digital tools and automated platforms^{[13][14][15][16][17][18][19][20]}. In this vision paper, we discuss various aspects in low-resource language education that could be enhanced by the recent foundation models.

2. Foundation Model Development

This section explores the development of multilingual foundation models designed to extend the capabilities of NLP systems to low-resource languages. Both large language models and vision-language models (VLMs) are discussed in the context of pre-training, fine-tuning, and in-context learning. These approaches aim to overcome the data scarcity challenges inherent in low-resource languages, while also supporting educational and practical applications.

2.1. FM Pre-training

Pre-training forms the foundational stage for both LLMs and VLMs, where models learn to extract general linguistic or multimodal patterns from large-scale datasets. LLM pre-training typically employs loss functions such as masked language modeling (e.g., BERT^[21]) or autoregressive objectives (e.g., the GPT series^{[22][23][24]} and the Llama family of models^{[3][25][26]}). These techniques allow LLMs to capture complex syntactic, semantic, and contextual relationships in text.

LLM pre-training commonly^{[18][20][17]} employs objectives such as:

- **Masked Language Modeling (MLM):** This method masks random tokens in the input sequence and trains the model to predict these masked tokens during pre-training^[21]. BERT and many of its variants use MLM as their primary pre-training objective, enabling bidirectional context understanding.
- **Autoregressive Modeling:** This approach trains models to predict the next token in a sequence by learning patterns from vast amounts of text data^[24]. The model conditions each new token on all

previously generated tokens, enabling it to maintain coherence and capture long-range dependencies while generating text.

- **Denoising AutoEncoder (DAE):** This technique corrupts the input data with various noise functions (masking, deletion, permutation) and trains the model to reconstruct the original sequence^[27]. BART and T5 employ DAE-based objectives to learn robust text representations and generation capabilities.
- **Replaced Token Detection (RTD):** This task trains a discriminator to distinguish between original and synthetically replaced tokens in the input sequence^[28]. ELECTRA demonstrated that RTD enables more compute-efficient pre-training compared to MLM, particularly for smaller models.
- **Next Sentence Prediction (NSP) and Sentence Order Prediction (SOP):** NSP trains models to determine if two sentences appear consecutively in the original text^[21], while SOP focuses on predicting the correct ordering of consecutive sentences^[29]. ALBERT showed that SOP could be potentially more effective than NSP for learning inter-sentence coherence.

However, the representation of low-resource languages in existing datasets remains limited, often leading to suboptimal performance in these languages^[30].

For VLMs, pre-training involves aligning image and text embeddings within a unified semantic space. Techniques include the use of CLIP's contrastive learning framework^[31], where paired image-text data are used to maximize alignment, and latent diffusion models like Stable Diffusion^[32], which focus on text-to-image tasks. Models such as LLaVA^[33] and Kosmos-2^[34] enhance cross-modal understanding, enabling tasks such as visual question answering and captioning. However, these models predominantly rely on high-resource language datasets, leaving low-resource languages underserved.

Targeted pre-training strategies can significantly benefit low-resource languages. Continuous pre-training on multilingual corpora or domain-specific datasets enhances model adaptability^{[35][36]}. For VLMs, incorporating culturally relevant visual and textual data ensures better alignment for underrepresented languages. In educational settings, such pretrained models can generate learning materials, including multimodal content, that cater to diverse linguistic contexts, ultimately improving accessibility and inclusion.

2.2. FM Fine-tuning

Fine-tuning adapts pre-trained models to specific tasks or domains by training them on task-specific data and optimizing loss functions to refine their performance. This process allows foundation models to address particular requirements, such as low-resource language processing or multimodal applications.

Fine-tuning typically uses a supervised learning approach, where the model's predictions are compared against labeled data. The most commonly used loss function is cross-entropy^[37], which measures the difference between predicted probabilities and the true labels. By minimizing cross-entropy loss, the model learns to generate accurate outputs for the target task while mitigating overfitting through techniques like dropout, weight decay, or early stopping. Optimization algorithms, such as AdamW^[38], are often employed to fine-tune model parameters efficiently.

Instruction Fine-tuning for LLMs and VLMs

Instruction fine-tuning adapts pre-trained models to align with specific instructions and tasks, enhancing their generalization to unseen scenarios. For LLMs, this involves training on datasets containing task-specific instructions paired with responses. This approach, as seen in models like InstructGPT^[2] and FLAN-T5^[39], enables effective zero-shot and few-shot task generalization. Such fine-tuning often leverages datasets encompassing diverse tasks, such as answering questions, performing translations, and summarizing text. For low-resource languages, instruction-tuning can be pivotal. By training on carefully curated datasets of low-resource languages (e.g., Quechua or Cherokee), LLMs can address linguistic gaps and support education through task-based language exercises tailored to underrepresented communities.

In the context of VLMs, instruction fine-tuning expands the capability of models to follow multimodal instructions. Recent developments like LLaVA^[33] and Flamingo^[40] have demonstrated how visual instruction-tuning aligns language and vision encoders with task-specific goals, such as image captioning and visual question answering. For instance, by fine-tuning on datasets with paired image-text instructions, these models can describe images in underrepresented languages or perform tasks such as cultural-specific object identification. This is particularly impactful for education, where VLMs can enable immersive learning experiences, providing visual aids and engaging storytelling in local languages.

Applications in Education

Fine-tuned models open up significant possibilities for personalized and localized educational tools. Instruction-tuned LLMs can assist students in learning low-resource languages by generating culturally specific exercises, correcting grammar, or providing contextually appropriate examples. Similarly, fine-tuned VLMs can enable immersive learning experiences by creating visual aids, interactive storytelling, or image-based quizzes tailored to the learner's linguistic and cultural background. These adaptations bridge the gap between global AI technologies and local educational needs, fostering inclusivity and accessibility in resource-constrained environments.

2.2.1. Preference Alignment with RLHF

Preference alignment refines foundation models by aligning their outputs with human preferences, enhancing their utility for specific applications. Two prominent approaches for preference alignment are Proximal Policy Optimization (PPO)^[41] and Direct Preference Optimization (DPO)^[42], each with unique strengths.

Proximal Policy Optimization (PPO)

PPO is a reinforcement learning method central to Reinforcement Learning from Human Feedback (RLHF)^{[43][2]}. This process involves training a reward model on human-labeled rankings of model outputs, followed by fine-tuning the foundation model to maximize these rewards. PPO ensures stable updates to the model while aligning it with desired stylistic and behavioral traits, such as cultural sensitivity or conversational norms.

For low-resource languages, PPO can adapt models to handle specific cultural contexts or linguistic subtleties, making it suitable for applications like culturally aware chatbots or educational AI companions. However, the approach requires substantial human-labeled data for training the reward model, which may pose challenges in low-resource settings.

Direct Preference Optimization (DPO)

DPO provides a more streamlined alternative by bypassing the need for a reward model and reinforcement learning loop. Instead, it directly fine-tunes the model on preference-labeled data using a contrastive loss to maximize the likelihood of preferred outputs over less-preferred ones. This simplicity reduces computational overhead and eliminates the instability risks inherent in PPO.

A significant advantage of DPO is its ability to reduce labeling costs. Unlike PPO, which requires ranked outputs for reward model training, DPO leverages binary or limited comparative preference data, minimizing the burden of annotation. This efficiency makes it particularly valuable for low-resource languages, where obtaining extensive labeled datasets is often infeasible. For instance, DPO can be used to train models that generate culturally sensitive text or conversational agents with minimal labeled examples.

Applications in Low-Resource Language Education:

Both PPO and DPO can enhance educational tools for low-resource languages, but DPO's efficiency and lower data requirements make it especially impactful. By reducing the need for extensive labeling, DPO enables the rapid adaptation of AI systems to underserved languages and cultural contexts. In education, it supports the development of AI companions that provide personalized feedback, culturally sensitive guidance, or tailored lesson plans, fostering inclusivity and accessibility in resource-constrained environments.

In low-resource languages, RLHF can adjust models to reflect cultural sensitivity, conversational norms, and localized communication styles. Applications include AI learning companions tailored to specific linguistic contexts, culturally aware educational chatbots, and training systems for language-specific customer service.

2.3. FM In-context Learning

In-context learning (ICL)^[44] allows pre-trained models to generalize to new tasks without additional parameter updates by leveraging few-shot examples provided as input prompts. This capability is particularly valuable for low-resource languages, where annotated data is scarce. Properly designed prompts enable LLMs to recognize patterns and generate relevant outputs, reducing the reliance on large labeled datasets.

For VLMs, ICL extends to multimodal prompts that combine images with textual instructions. This facilitates zero-shot or few-shot learning for tasks like image captioning or visual question answering in low-resource languages. Models like Kosmos-2^[34] and DeepSeek-VL^[45] demonstrate strong zero-shot capabilities, making them suitable for applications where task-specific data are unavailable.

The intuitive and flexible nature of ICL is a significant advantage in educational contexts. LLMs can be prompted to create exercises, translate content, or summarize lessons in low-resource languages,

while VLMs can answer visual questions or provide contextual understanding of imagery. This adaptability supports educators and learners in resource-constrained environments, democratizing access to AI-driven educational tools.

3. Language Education Modules

With foundational models for low-resource languages, we can generate comprehensive lessons and learning materials spanning various aspects^[30], including vocabulary and pronunciation, grammar, interactive exercises, cultural integration, and video generation. This section will discuss methodologies for leveraging foundational models for various language education modules, demonstrating how foundation model agents can effectively support education in low-resource languages.

3.1. Vocabulary and Pronunciation

To build vocabulary resources, the foundation model agents can harness their ability to understand and generate contextualized text in low-resource language^[46]. More specifically, these model agents can produce:

- **Word Lists:** Starting with foundational word sets, models can extrapolate to generate synonyms, antonyms, and contextually related terms. For instance, a low-resource language with a small dictionary can be expanded by inferring semantically linked words using multilingual embeddings in large language models.
- **Contextual Examples:** Learners often benefit from seeing new vocabulary used in meaningful contexts. Foundation models can generate diverse, practical examples by embedding target words within sentences, dialogues, or short narratives. These examples provide learners with a deeper understanding of how words function in real-life scenarios^[47].
- **Thematic Categorization:** Vocabulary lists can be organized into thematic areas like family, nature, or technology. Foundation models trained on multilingual data can identify domain-specific terms and present them in learner-friendly ways, such as flashcards or interactive quizzes.

Pronunciation mastery is often a hurdle in language learning, especially for learners of low-resource languages with limited audio examples. Foundation model agents equipped for speech and multimodal processing provide innovative solutions^{[48][49]}:

- **Phoneme-to-Audio Generation:** Text-to-speech (TTS) systems based on foundation models can generate audio samples for specific words or phrases. Foundation models or fine-tuned systems can provide native-like pronunciation examples, even when trained on small low-resource language datasets.
- **Phonetic Guides and Feedback:** By integrating speech recognition, models can evaluate learners' pronunciation. For instance, a learner speaking a word can receive feedback on accuracy, intonation, and stress patterns. This fosters self-paced learning without needing a native speaker's constant availability.
- **Regional Variations:** For low-resource languages with significant dialectal differences, foundation models can be fine-tuned to generate region-specific pronunciations, offering learners exposure to the diversity within the language. This is particularly valuable for languages where phonetic, tonal, or lexical differences between regions are substantial.

The TF-IDF (Term Frequency-Inverse Document Frequency) ^[50] method can enhance vocabulary and pronunciation learning by identifying words that are both frequent and contextually important. For vocabulary, TF-IDF can prioritize words that are common in a learner's target domain (e.g., travel or business), helping learners focus on terms with high relevance. In pronunciation practice, TF-IDF can highlight less common but important words or those with regional variations, guiding targeted practice. This method ensures that learners engage with vocabulary that is both practical and specific to their needs, optimizing their learning experience.

Foundation model agents can also be improved through tokenization, a critical component for processing and understanding language. Tokenization is the process of breaking down text into smaller units, or tokens, which are processed by language models. This is essential for understanding how different words and phrases are constructed in low-resource languages, especially when dealing with agglutinative or morphologically complex languages. By tokenizing text properly, foundation models can capture linguistic nuances more accurately and generate vocabulary lists that reflect the language's true structure, rather than overgeneralizing based on sparse data.

3.2. Grammar

Grammar is fundamental to mastering any language, yet low-resource languages often lack comprehensive instructional materials. The foundation model agents can automate the creation of grammar explanations, exercises, contextualized examples, etc., all tailored to the specific needs of

learners. By harnessing the power of these foundation model agents, we can provide learners of low-resource languages with personalized, accessible, and engaging grammar content.

- **Grammar Rules Extraction and Explanation:** Foundation models can analyze multilingual datasets to extract and explain grammatical patterns for low-resource languages. By identifying syntactic structures such as word order, tense, and subject-verb agreement, foundation models can generate grammar guides tailored to the unique rules of a given low-resource language. These models can produce detailed explanations of fundamental grammatical concepts, from simple noun-adjective agreements to more complex structures, offering step-by-step guidance for learners.
- **Cross-Lingual Grammar Transfer:** Foundation models can also facilitate cross-lingual grammar transfer by comparing the grammatical structures of related languages. For example, a learner who speaks English may benefit from grammar lessons comparing English and a low-resource language, focusing on shared and divergent grammatical features. This comparison allows learners to leverage their existing linguistic knowledge, accelerating their learning process.

3.3. Interactive Exercises

The foundation model agents could offer various types of interactive exercises that engage learners in practicing the language in a controlled environment, providing immediate feedback and adjustments based on their responses^[51]. The flexibility of these exercises ensures that learners can practice different aspects of language acquisition in a way that suits their individual needs. Below are several specific use cases that illustrate how foundation model agents can generate interactive exercises for low-resource language learning.

- **Matching Exercises for Vocabulary Learning:** Learners are presented with a set of vocabulary words and their corresponding images or definitions. The learner must match each word with its correct meaning or image. The foundation models adapt the difficulty based on the learner's progress, offering more challenging words or a wider variety of images as they advance.
- **Dialogue Simulation Exercises for Pronunciation Learning:** The foundation models can simulate real-life conversations, where learners interact with an AI-based character in a scenario (e.g., ordering food at a restaurant). The learner's speech input is evaluated in real-time, and the system provides feedback on pronunciation, fluency, and appropriateness of responses. This type of exercise helps learners practice speaking in a practical, contextual setting.

- **Error Correction Exercises for Grammar Learning:** Learners are presented with sentences that contain common grammatical errors (such as subject-verb agreement or tense inconsistencies), and they are tasked with identifying and correcting them. Immediate feedback is provided, explaining why a particular correction is necessary.

3.4. *Culture Integration*

Integrating cultural education into language learning is essential, particularly for low-resource languages. Language and culture are deeply intertwined, and foundation model agents—spanning text, voice, and vision—can generate materials that highlight how language operates within its cultural context, including folklore, history, art, and traditions^[52].

Foundation model agents can produce stories, songs, or dialogues that reflect the cultural heritage of the language’s speakers, exposing learners to folklore and idiomatic expressions. These materials not only teach language but also reveal cultural values and worldviews. For instance, learners can engage with traditional narratives that embody cultural meanings.

Moreover, foundation model agents can generate historical content that provides context on the language’s evolution and its speakers’ experiences. They can translate or summarize historical texts, highlight significant events or figures, and showcase scientific or artistic contributions, helping learners understand the language’s cultural and historical roots.

Visual content created by these foundation model agents can complement text and voice lessons, featuring landmarks, art, and indigenous knowledge. Interactive maps, art galleries, and multimedia on local practices, like traditional medicine or sustainable agriculture, allow learners to connect the language to the community’s way of life.

By integrating culture into low-resource language learning, foundation model agents offer a comprehensive approach that connects language contents with cultural identity, enriching the learners’ understanding and experience.

3.5. *Video Generation*

In the context of low-resource languages, one of the most significant barriers to education is the lack of access to qualified educators and learning buddies. Foundation model agents offer solutions to this problem, particularly through the generation of video content featuring virtual teachers and learning

buddies^[53]. These resources provide immersive, interactive learning experiences that cater to the needs of learners in underserved linguistic communities.

Virtual Teachers for Immersive Learning

Virtual teachers are AI agents that deliver lessons tailored to a student's level and progress. Using a combination of text-based foundation models, voice synthesis, and computer vision technologies, these virtual teachers can simulate human-like interactions. For instance, voice models can generate speech in the target language while video models display facial expressions and gestures that align with cultural context, helping students not only learn the language but also understand its social and cultural nuances.

These AI-driven teachers enable scalable and personalized education in areas where human teachers are scarce. By adjusting lesson complexity based on the learner's proficiency, virtual teachers can introduce new vocabulary and improve pronunciation over time. Additionally, they can be multilingual, supporting low-resource language learners from diverse language backgrounds.

The integration of non-verbal cues like body language and facial expressions in video lessons further enhances communication, helping students associate spoken words with actions. This immersive approach aids both comprehension and retention, making virtual teachers a powerful tool for language acquisition in under-resourced areas.

Learning Buddies for Enhanced Engagement

In addition to virtual teachers, learning buddies—AI agents designed to interact and engage with students—play a crucial role in fostering motivation and providing ongoing support. Learning buddies can offer encouragement, answer questions, and guide students through practice exercises. These AI agents can hold conversations, giving learners opportunities to practice speaking and listening in dynamic, real-life contexts^[54].

Learning buddies also provide personalized feedback, offering corrections when necessary and reinforcing new concepts. This adaptive approach builds student confidence and accelerates language mastery. Furthermore, learning buddies can be customized to reflect cultural nuances, making lessons not only linguistically accurate but also culturally relevant, thereby deepening students' connection to the language.

4. Adaptive Learning in Language Models and Education

Adaptive learning in low-resource language education should primarily focus on creating personalized learning experiences based on individual student knowledge levels and progress. A foundation model-based system could analyze each learner's current understanding and mastery of language concepts to recommend customized learning paths, content, and objectives. This personalization would ensure that students receive appropriately challenging material and targeted practice opportunities that align with their learning goals and current capabilities.

Such a system could continuously evaluate student performance to adjust these recommendations dynamically. When students would demonstrate mastery of certain concepts, the system might advance them to more challenging material. Conversely, if students were to struggle with particular aspects, the system could provide additional practice and support in those areas. This dynamic adjustment would help maintain an optimal learning pace for each student while ensuring thorough understanding of fundamental concepts.

Beyond individual learning paths, adaptive learning would also encompass the technical evolution of language models themselves. These models should be able to update and improve their knowledge base through mechanisms like continuous pre-training and fine-tuning, particularly important for low-resource languages where traditional training data might be limited. This dual nature - personalized education and model evolution - could create an effective framework for low-resource language education that responds to both individual learner needs and evolving linguistic landscapes.

4.1. Content Generation and Adaptation

Model Evolution

Traditional adaptive learning systems face limitations with pre-authored content and fixed question banks^[55]. In contrast, foundation models for low-resource languages should be able to continuously evolve their knowledge base through interaction with new linguistic data. This evolution would allow dynamic content generation that could reflect current language usage patterns and cultural contexts.

Educational Content

The adaptive capability should extend to the generation of varied, contextually appropriate learning materials that could be adjusted based on learner needs. This would overcome the traditional

limitation of manual content updates and restricted practice variety^[56]. The system could generate exercises, examples, and explanations in real-time that would maintain cultural relevance and support progressive language acquisition.

Individualized Content

A foundation model-based system could construct customized learning paths by analyzing each student's progress milestones and learning objectives. When a student's goal would be professional communication, the content sequence might emphasize relevant vocabulary and formal language patterns. For basic conversational goals, the path could prioritize common phrases and daily expressions. Such goal-oriented content sequencing would help ensure that learning materials align with each student's specific language objectives.

4.2. Feedback Mechanisms

Interactive Guidance

Unlike traditional systems constrained to predefined responses^[57], foundation models should be able to provide nuanced, contextual feedback through natural language understanding. This capability would lead to more sophisticated error analysis and targeted remediation strategies for language learners.

Cultural Sensitivity

The feedback system should be able to incorporate cultural context awareness. This feature would ensure that corrections and suggestions align with both linguistic accuracy and cultural appropriateness. This approach would address the limitation of traditional systems that cannot recognize or respond to culturally influenced language usage patterns.

Personalized Feedback

The feedback mechanism should be able to adapt to each student's position on their learning path and current learning objectives. For students working toward reading proficiency, feedback could emphasize character recognition and comprehension strategies. For those focused on speaking skills, the system might provide more pronunciation guidance. This alignment between feedback and

learning goals would help students progress more effectively toward their target language proficiency levels.

4.3. Student Modeling

Comprehensive Assessment

Traditional modeling approaches often struggle with complex learning trajectories^{[58][59]}. Foundation models should be able to track multiple dimensions of language acquisition at once, from vocabulary and grammar to pronunciation and cultural understanding. This comprehensive tracking would create more detailed learner profiles.

Adaptive Pathways

These systems should be able to analyze patterns in learner interactions and performance to adjust learning paths for individual needs. This approach would address the traditional limitation of insufficient consideration of affective states^[60] and motivation in learning. The result would be a more personalized and engaging educational experience.

4.4. Integration Considerations

Technical Implementation

When integrating foundation models for low-resource language education, scalability could potentially be a primary concern. The system architecture must support efficient model deployment while maintaining quick response times for interactive learning sessions. Additionally, quality control mechanisms must be implemented to ensure the accuracy and appropriateness of generated content, particularly crucial for low-resource languages where training data may be limited.

Educational Workflow

The integration of foundation models must complement existing language teaching methodologies. This includes developing interfaces that allow educators to monitor and guide the learning process, ensuring that automated adaptations align with pedagogical goals. Assessment mechanisms need to balance automated evaluation with human oversight, particularly for nuanced aspects of language acquisition like cultural competency^[61].

Resource Management

For low-resource language contexts, considerations must be made for environments with limited technological infrastructure. This includes developing offline functionality and optimizing model performance for resource-constrained settings. The system should be designed to gracefully handle intermittent connectivity while maintaining learning continuity.

Quality Assurance

Maintaining educational effectiveness requires robust quality assurance processes. This includes automated testing of generated content, collection of user feedback, and regular assessment of learning outcomes. For low-resource languages, additional verification may be needed to ensure cultural and linguistic accuracy, potentially involving community experts in the review process.

Community Engagement

Successful integration requires active participation from the language community. This includes establishing channels for native speakers and cultural experts to provide feedback and validation of educational content. Such engagement helps ensure that the adaptive learning system remains aligned with community needs and cultural values while supporting language preservation efforts.

5. Conclusion and Future Challenges

Incorporating large language models into low-resource language education poses significant educational challenges and opportunities. One major issue is the development of teaching methodologies to effectively utilize LLM capabilities, such as generating language exercises or providing feedback. Educators must develop pedagogical strategies that leverage LLM strengths while addressing their limitations, such as occasional inaccuracies or lack of context-specific nuances. Additionally, there is the challenge of integrating these technologies into existing educational frameworks in a way that complements traditional teaching methods rather than replacing them. This integration also requires substantial training for educators, who must become adept at using LLMs and understanding their potential biases and shortcomings. Furthermore, evaluating the effectiveness of LLMs in improving language competencies in diverse educational settings such as adaptive learning^[62] remains a critical challenge, necessitating ongoing research and adaptation to ensure that these tools are making a positive impact on learning outcomes.

References

1. [△]Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, Aleman FL, Almeida D, Altenschmidt J, Altman S, Anadkat S, et al. (2023). "Gpt-4 technical report". arXiv preprint arXiv:2303.08774. Available from: <http://arxiv.org/abs/2303.08774>.
2. ^{a, b, c}Ouyang L, Wu J, Jiang X, Almeida D, Wainwright C, Mishkin P, Zhang C, Agarwal S, Slama K, Ray A, et al. (2022). "Training language models to follow instructions with human feedback". *Advances in neural information processing systems*. 35: 27730–27744.
3. ^{a, b}Touvron H, Lavril T, Izacard G, Martinet X, Lachaux MA, Lacroix T, Rozière B, Goyal N, Hambro E, Azhar F, et al. (2023). "Llama: Open and efficient foundation language models". arXiv preprint arXiv:2302.13971. Available from: <https://arxiv.org/abs/2302.13971>.
4. [△]Gemma Team, Mesnard T, Hardin C, Dadashi R, Bhupatiraju S, Pathak S, Sifre L, Rivière M, Kale MS, Lowe J, et al. (2024). "Gemma: Open models based on gemini research and technology". arXiv preprint arXiv:2403.08295. Available from: <https://arxiv.org/abs/2403.08295>.
5. [△]Jiang AQ, Sablayrolles A, Mensch A, Bamford C, Chaplot DS, de las Casas D, Bressand F, Lengyel G, Lample G, Saulnier L, et al. 2023. "Mistral 7b". arXiv preprint arXiv:2310.06825. Available from: <https://arxiv.org/abs/2310.06825>.
6. [△]Bai J, Bai S, Chu Y, Cui Z, Dang K, Deng X, Fan Y, Ge W, Han Y, Huang F, et al. (2023). "Qwen technical report". arXiv preprint arXiv:2309.16609. Available from: <https://arxiv.org/abs/2309.16609>.
7. [△]Chang TA, Arnett C, Tu Z, Bergen BK (2023). "When is multilinguality a curse? language modeling for 250 high-and low-resource languages". arXiv preprint arXiv:2311.09205. Available from: <https://arxiv.org/abs/2311.09205>.
8. [△]Qin L, Chen Q, Zhou Y, Chen Z, Li Y, Liao L, Li M, Che W, Yu PS (2024). "Multilingual large language model: A survey of resources, taxonomy and frontiers". arXiv preprint arXiv:2404.04925. Available from: <https://arxiv.org/abs/2404.04925>.
9. [△]Kohnke L, Moorhouse BL, Zou D (2023). "Chatgpt for language teaching and learning". *Relc Journal*. 54 (2): 537–550.
10. [△]Lee GG, Shi L, Latif E, Gao Y, Bewersdorff A, Nyaaba M, Guo S, Wu Z, Liu Z, Wang H, et al. 2023. "Multimodality of AI for education: Towards artificial general intelligence." arXiv preprint arXiv:2312.06037. Available from: <https://arxiv.org/abs/2312.06037>.

11. ^aTian J, Hou J, Wu Z, Shu P, Liu Z, Xiang Y, Gu B, Filla N, Li Y, Liu N, et al. (2024). "Assessing large language models in mechanical engineering education: A study on mechanics-focused conceptual understanding". arXiv preprint arXiv:2401.12983. Available from: <https://arxiv.org/abs/2401.12983>.
12. ^aPeng S, Zhao H, Jiang H, Li Y, Xu S, Pan Y, Wu Z, Liu Z, Lu G, Guan L, et al. (2024). "LLMs for coding and robotics education". arXiv preprint arXiv:2402.06116.
13. ^{a, b}Latif E, Zhou Y, Guo S, Gao Y, Shi L, Nayaaba M, Lee G, Zhang L, Bewersdorff A, Fang L, et al. (2024). "A systematic assessment of OpenAI O1-preview for higher order thinking in education". arXiv preprint arXiv:2410.21287. Available from: <https://arxiv.org/abs/2410.21287>.
14. ^aJiang H, Pan Y, Chen J, Liu Z, Zhou Y, Shu P, Li Y, Zhao H, Mihm S, Howe LC, et al. Oraclesage: Towards unified visual-linguistic understanding of oracle bone scripts through cross-modal knowledge fusion. a rXiv preprint arXiv:2411.17837. 2024.
15. ^aPeng S, Chen J, Liu Z, Wang H, Wu Z, Zhong T, Li Y, Zhao H, Jiang H, Pan Y, et al. Transcending language boundaries: Harnessing llms for low-resource language translation. arXiv preprint arXiv:2411.11295. [arXiv:2411.11295](https://arxiv.org/abs/2411.11295).
16. ^aZhong T, Liu Z, Pan Y, Zhang Y, Zhou Y, Liang S, Wu Z, Lyu Y, Shu P, Yu X, et al. (2024). "Evaluation of openai o1: Opportunities and challenges of agi". arXiv preprint arXiv:2409.18486. Available from: <https://arxiv.org/abs/2409.18486>.
17. ^{a, b}Liu Z, He X, Liu L, Liu T, Zhai X (2023). "Context matters: A strategy to pre-train language model for science education." In: International Conference on Artificial Intelligence in Education. Springer. p. 666–674.
18. ^{a, b}Liu Y, He H, Han T, Zhang X, Liu M, Tian J, Zhang Y, Wang J, Gao X, Zhong T, et al. (2024). "Understanding LLMs: A comprehensive overview from training to inference". arXiv preprint arXiv:2401.02038. Available from: <https://arxiv.org/abs/2401.02038>.
19. ^aLiu Z, Li Y, Cao Q, Chen J, Yang T, Wu Z, Hale J, Gibbs J, Rasheed K, Liu N, et al. (2023). "Transformation vs tradition: Artificial general intelligence (agi) for arts and humanities". arXiv preprint arXiv:2310.19626. Available from: <https://arxiv.org/abs/2310.19626>.
20. ^{a, b}Ce Zhou, Li Q, Li C, Yu J, Liu Y, Wang G, Zhang K, Ji C, Yan Q, He L, et al. (2024). "A comprehensive survey on pretrained foundation models: A history from BERT to ChatGPT". International Journal of Machine Learning and Cybernetics. pages 1–65.
21. ^{a, b, c}Devlin J (2018). "Bert: Pre-training of deep bidirectional transformers for language understanding". arXiv preprint arXiv:1810.04805.

22. [△]Radford A (2018). "Improving language understanding by generative pre-training."
23. [△]Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I, et al. (2019). "Language models are unsupervised multitask learners." *OpenAI blog*. 1 (8): 9.
24. ^{△, b}Brown TB (2020). "Language models are few-shot learners". *arXiv preprint arXiv:2005.14165*.
25. [△]Touvron H, Martin L, Stone K, Albert P, Almahairi A, Babaei Y, Bashlykov N, Batra S, Bhargava P, Bhosale S, et al. (2023). "Llama 2: Open foundation and fine-tuned chat models". *arXiv preprint arXiv:2307.09288*. Available from: <https://arxiv.org/abs/2307.09288>.
26. [△]Dubey A, Jauhri A, Pandey A, Kadian A, Al-Dahle A, Letman A, Mathur A, Schelten A, Yang A, Fan A, et al. (2024). "The llama 3 herd of models". *arXiv preprint arXiv:2407.21783*. Available from: <https://arxiv.org/abs/2407.21783>.
27. [△]Lewis M (2019). "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension." *arXiv preprint arXiv:1910.13461*. Available from: <https://arxiv.org/abs/1910.13461>.
28. [△]Clark K (2020). "Electra: Pre-training text encoders as discriminators rather than generators". *arXiv preprint arXiv:2003.10555*. Available from: <https://arxiv.org/abs/2003.10555>.
29. [△]Lan Z. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*. 2019.
30. ^{△, b}Magueresse A, Carles V, Heetderks E (2020). "Low-resource languages: A review of past work and future challenges". *arXiv preprint arXiv:2006.07264*.
31. [△]Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, et al. (2021). "Learning transferable visual models from natural language supervision." In *International Conference on Machine Learning*, pages 8748–8763. PMLR.
32. [△]Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B (2022). "High-resolution image synthesis with latent diffusion models." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pages 10684–10695.
33. ^{△, b}Liu H, Li C, Wu Q, Lee YJ (2024). "Visual instruction tuning". *Advances in Neural Information Processing Systems*. 36.
34. ^{△, b}Peng Z, Wang W, Dong L, Hao Y, Huang S, Ma S, Wei F (2023). "Kosmos-2: Grounding multimodal large language models to the world". *arXiv preprint arXiv:2306.14824*. Available from: <https://arxiv.org/abs/2306.14824>.

35. [△]Xie Y, Aggarwal K, Ahmad A (2023). "Efficient continual pre-training for building domain specific large language models". arXiv preprint arXiv:2311.08545. Available from: <https://arxiv.org/abs/2311.08545>.
36. [△]Liu Y, Han T, Ma S, Zhang J, Yang Y, Tian J, He H, Li A, He M, Liu Z, et al. (2023). "Summary of chatgpt-related research and perspective towards the future of large language models". *Meta-Radiology*. Page 100017.
37. [△]Li Z, Chen C, Xu T, Qin Z, Xiao J, Sun R, Luo ZQ (2024). "Entropic distribution matching in supervised fine-tuning of llms: Less overfitting and better diversity". arXiv preprint arXiv:2408.16673. Available from: <https://arxiv.org/abs/2408.16673>.
38. [△]Loshchilov I (2017). "Decoupled weight decay regularization". arXiv preprint arXiv:1711.05101. Available from: [arXiv:1711.05101](https://arxiv.org/abs/1711.05101).
39. [△]Chung HW, Hou L, Longpre S, Zoph B, Tay Y, Fedus W, Li Y, Wang X, Dehghani M, Brahma S, et al. (2024). "Scaling instruction-finetuned language models". *Journal of Machine Learning Research*. 25 (70): 1–53.
40. [△]Alayrac JB, Donahue J, Luc P, Miech A, Barr I, Hasson Y, Lenc K, Mensch A, Millican K, Reynolds M, et al. (2022). "Flamingo: a visual language model for few-shot learning". *Advances in Neural Information Processing Systems*. 35: 23716–23736.
41. [△]Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O (2017). "Proximal policy optimization algorithms". arXiv preprint arXiv:1707.06347. Available from: <https://arxiv.org/abs/1707.06347>.
42. [△]Rafailov R, Sharma A, Mitchell E, Manning CD, Ermon S, Finn C (2024). "Direct preference optimization: Your language model is secretly a reward model". *Advances in Neural Information Processing Systems*. 36.
43. [△]Kaufmann T, Weng P, Bengs V, Hüllermeier E (2023). "A survey of reinforcement learning from human feedback". arXiv preprint arXiv:2312.14925. Available from: <https://arxiv.org/abs/2312.14925>.
44. [△]Dong Q, Li L, Dai D, Zheng C, Ma J, Li R, Xia H, Xu J, Wu Z, Liu T, et al. (2022). "A survey on in-context learning". arXiv preprint arXiv:2301.00234. Available from: <https://arxiv.org/abs/2301.00234>.
45. [△]Lu H, Liu W, Zhang B, Wang B, Dong K, Liu B, Sun J, Ren T, Li Z, Yang H, et al. (2024). "Deepseek-vl: towards real-world vision-language understanding". arXiv preprint arXiv:2403.05525. Available from: <https://arxiv.org/abs/2403.05525>.
46. [△]Guo P, Ren Y, Hu Y, Li Y, Zhang J, Zhang X, Huang H (2024). "Teaching large language models to translate on low-resource languages with textbook prompting". *Proceedings of the 2024 Joint International*

Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). pages 15685–15697.

47. [△]Cahyawijaya S, Lovenia H, Fung P (2024). "LLMs are few-shot in-context low-resource language learners". *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*. pages 405–433.
48. [△]Li Y, Kim S, Wu Z, Jiang H, Pan Y, Jin P, Song S, Shi Y, Liu T, Li Q, et al. Echopulse: ECG controlled echocardiograms video generation. *arXiv preprint arXiv:2410.03143*. 2024.
49. [△]Lajszczak M, Cambara G, Li Y, Beyhan F, van Korlaar A, Yang F, Joly A, Martín-Cortinas Á, Abbas A, Michalski A, et al. Base tts: Lessons from building a billion-parameter text-to-speech model on 100k hours of data. *arXiv preprint arXiv:2402.08093*. 2024. Available from: <https://arxiv.org/abs/2402.08093>.
50. [△]Aizawa A (2003). "An information-theoretic perspective of tf-idf measures". *Information Processing & Management*. 39 (1): 45–65.
51. [△]Xiao C, Xu SX, Zhang K, Wang Y, Xia L (2023). "Evaluating reading comprehension exercises generated by llms: A showcase of chatgpt in education applications." In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*, pages 610–625.
52. [△]Cheng Li, Mengzhou Chen, Jindong Wang, Sunayana Sitaram, Xing Xie (2024). "Culturellm: Incorporating cultural differences into large language models". *arXiv preprint arXiv:2402.10946*. Available from: <https://arxiv.org/abs/2402.10946>.
53. [△]Xu T, Tong R, Liang J, Fan X, Li H, Wen Q (2024). "Foundation models for education: Promises and prospects". *arXiv preprint arXiv:2405.10959*. Available from: <https://arxiv.org/abs/2405.10959>.
54. [△]Nyaaba M, Shi L, Nabang M, Zhai X, Kyeremeh P, Ayoberd SA, Akanzire BN (2024). "Generative ai as a learning buddy and teaching assistant: Pre-service teachers' uses and attitudes". *arXiv preprint arXiv:2407.11983*. Available from: <https://arxiv.org/abs/2407.11983>.
55. [△]Larsson JA, White B (2014). *Learning analytics: From research to practice*, volume 13. Springer.
56. [△]Aleven V, Roll I, McLaren BM, Koedinger KR (2016). "Help helps, but only so much: Research on help seeking with intelligent tutoring systems". *International Journal of Artificial Intelligence in Education*. 26: 205–223.
57. [△]Ma W, Adesope OO, Nesbit JC, Liu Q (2014). "Intelligent tutoring systems and learning outcomes: A meta-analysis". *Journal of Educational Psychology*. 106 (4): 901.
58. [△]Corbett AT, Anderson JR (1994). "Knowledge tracing: Modeling the acquisition of procedural knowledge". *User modeling and user-adapted interaction*. 4: 253–278.

59. [^]Chi M, VanLehn K, Litman D, Jordan P (2011). "An evaluation of pedagogical tutorial tactics for a natural language tutoring system: A reinforcement learning approach". *International Journal of Artificial Intelligence in Education*. 21 (1-2): 83–113.
60. [^]Rachel Carlos Duque Reis, Seiji Isotani, Carla Lopes Rodriguez, Kamila Takayama Lyra, Patrícia Augustin Jaques, Ig Ibert Bittencourt (2018). "Affective states in computer-supported collaborative learning: Studying the past to drive the future." *Computers & Education*. 120: 29–50.
61. [^]Steele R (2000). "Language learning and intercultural competence". *Language policy and pedagogy*. 100: 193.
62. [^]Wen Q, Liang J, Sierra C, Luckin R, Tong R, Liu Z, Cui P, Tang J (2024). "AI for education (AI4EDU): advancing personalized education with LLM and adaptive learning." In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 6743–6744.

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.