CulturalEcho: Multimodal Deep Learning Approaches to Language Preservation, Documentation, and Community-Led Revitalization

Eeman Majumder
Lead AIML Engineer
eeman.majumder@gmail.com

Abstract—The global linguistic landscape faces an unprecedented crisis, with numerous languages teetering on the brink of extinction. This alarming trend signifies not merely the loss of vocabulary and grammar, but a profound erosion of cultural heritage, historical knowledge, and unique worldviews. The rapid disappearance of these languages severs communities from their ancestral roots and diminishes the rich tapestry of human communication. Studies indicate that the digital realm, while offering opportunities, also presents a significant challenge, with a vast majority of languages struggling to transition into the digital age, potentially leading to a substantial decline. In response to this urgent situation, artificial intelligence (AI), particularly through the advancements in multimodal deep learning, offers a promising avenue for language preservation, documentation, and community-led revitalization. Multimodal deep learning, a sophisticated branch of AI, excels at processing and integrating information from various data types, such as text, audio, images, and video. This comprehensive approach is particularly relevant to language, which is inherently multimodal, encompassing not only spoken and written words but also visual cues, contextual settings, and a wealth of cultural information. By simultaneously analyzing these diverse streams of information, AI models can achieve a more profound understanding of language in its full richness. The application of these techniques aims to address critical challenges in language preservation, including the scarcity of resources, the limited availability of data for many endangered languages, and the intricate nature of oral traditions that often lack written documentation.

Index Terms—Language preservation, multimodal deep learning, endangered languages, cultural heritage, artificial intelligence

I. INTRODUCTION

Multimodal deep learning, a sophisticated branch of AI, excels at processing and integrating information from various data types, such as text, audio, images, and video [5]. This comprehensive approach is particularly relevant to language, which is inherently multimodal, encompassing not only spoken and written words but also visual cues, contextual settings, and a wealth of cultural information [6]. By simultaneously analyzing these diverse streams of information, AI models can achieve a more profound understanding of language in its full richness. The application of these techniques aims to address critical challenges in language preservation, including the scarcity of resources, the limited availability of data for many endangered languages, and the intricate nature of oral traditions that often lack written documentation [3].

The global linguistic landscape faces an unprecedented crisis, with numerous languages teetering on the brink of extinction [1]. This alarming trend signifies not merely the loss of vocabulary and grammar, but a profound erosion of cultural heritage, historical knowledge, and unique worldviews. The rapid disappearance of these languages severs communities from their ancestral roots and diminishes the rich tapestry of human communication. Studies indicate that the digital realm, while offering opportunities, also presents a significant challenge, with a vast majority of languages struggling to transition into the digital age, potentially leading to a substantial decline [1].

This paper explores how multimodal deep learning approaches can contribute to language preservation, documentation, and community-led revitalization efforts. We examine the potential of these technologies to address the challenges faced by endangered languages and discuss ethical considerations, case studies, and future directions in this emerging field.

II. DECODING THE SPOKEN WORD AND VISUAL WORLD: MULTIMODAL DEEP LEARNING FOR LANGUAGE DOCUMENTATION

Multimodal deep learning techniques present a powerful framework for a more comprehensive and nuanced approach to language documentation by integrating various data streams [7]. The ability to process and analyze audio recordings allows for the preservation of the spoken word, capturing not only the vocabulary but also the unique phonetic characteristics and prosody of endangered languages [4]. AI-powered transcription tools can automate the conversion of oral histories, traditional narratives, and everyday conversations into written text, significantly reducing the workload on linguists and community members involved in documentation efforts [4].

Beyond the auditory dimension, visual information plays a crucial role in communication, particularly in the context of endangered languages where non-verbal cues and cultural artifacts carry significant meaning [7]. Video recordings of speakers provide valuable insights into body language, facial expressions, and the visual context of language use. For signed languages, the visual modality is paramount, and multimodal deep learning can learn features from video data, potentially alongside audio or textual descriptions, to facilitate their documentation and understanding [9]. Furthermore, visual

documentation of cultural artifacts, rituals, and daily life can provide crucial contextual information that enriches the understanding of language within its cultural sphere [10].

The integration of contextual information further enhances the depth of language documentation [7]. Metadata, such as speaker demographics, recording location, and cultural setting, provides valuable context for interpreting linguistic data. By leveraging AI to analyze these contextual cues alongside audio and visual information, a richer and more complete understanding of language use can be achieved [7]. This holistic approach moves beyond simply transcribing words to capturing the intricate interplay between language, culture, and environment.

Several architectures and models are being developed to harness the power of multimodal deep learning for language analysis [9]. For instance, the Audio-Video Language Network (AVLnet) demonstrates the potential of self-supervised learning to create shared audio-visual representations directly from raw video inputs, eliminating the need for manual text annotations [11]. Deep neural networks have also shown success in learning features across multiple modalities, such as audio and video, and in creating shared representations that can bridge the gap between different forms of communication [9]. More advanced models like AVicuna integrate audio and visual information with large language models, ensuring temporal alignment and enabling a more comprehensive understanding of multimodal content [12].

TABLE I
COMPARISON OF TRADITIONAL VS. MULTIMODAL DEEP LEARNING APPROACHES FOR LANGUAGE DOCUMENTATION

Feature	Traditional Approaches	Multimodal Deep Learning
Data Types Processed	Primarily Text and Audio	Audio, Visual, Contextual
Ability to Capture Nuances	Limited	Higher potential due to integrated data
Data Requirements	Can vary	Can leverage complementary information
Complexity of Implementation	Lower for basic tasks	Higher
Community Engagement Potential	Can be high (e.g., oral history projects)	Offers new avenues through interactive tools

III. BRIDGING THE DATA DIVIDE: TRANSFER LEARNING FOR LOW-RESOURCE LANGUAGE PRESERVATION

A significant impediment to applying advanced natural language processing (NLP) techniques to endangered languages is the scarcity of available data [13]. Transfer learning has emerged as a crucial strategy to overcome this challenge, allowing knowledge gained from training models on well-resourced languages to be applied to low-resource scenarios [4]. This approach is particularly vital for endangered languages, which often lack the extensive datasets required for training complex AI models from scratch.

Various transfer learning methodologies are relevant to NLP for low-resource languages. Fine-tuning pre-trained multilingual models involves taking models that have been trained on a vast amount of text in numerous languages and adapting them to a specific endangered language using a smaller, targeted dataset [13]. Models like multilingual BERT (mBERT) and XLM-R, which are trained on a wide array of languages, provide a strong foundation for cross-lingual knowledge transfer

[13]. These models can be further refined on the limited data available for an endangered language to improve performance on specific tasks.

Zero-shot and few-shot learning are powerful transfer learning techniques that enable models to perform tasks in a new language with minimal or even no labeled data [15]. By leveraging the knowledge acquired from training on other languages, these methods allow AI to understand and process endangered languages with remarkable efficiency, even when training data is extremely scarce. Cross-lingual embeddings play a key role in these techniques by creating shared semantic spaces where words and concepts from different languages are mapped together based on their meaning [15]. This shared representation facilitates the transfer of knowledge across linguistic boundaries.

Several instances demonstrate the application of transfer learning in the context of endangered languages. For example, machine learning models, likely utilizing transfer learning, have been employed to transcribe audio recordings of Maori oral traditions, contributing significantly to the preservation of this language [4]. Similarly, AI's role in transcribing and translating oral histories and stories from speakers of endangered languages often relies on the principles of transfer learning to achieve meaningful results despite limited language-specific data [8].

TABLE II
APPLICATIONS OF TRANSFER LEARNING FOR ENDANGERED LANGUAGES

Language	Technique Used	Application	Source
Maori	Transfer Learning (Implied)	Transcribing oral traditions	[4]
Various	Transfer Learning (Implied)	Translating oral histories and stories	[8]
Many	Fine-tuning mBERT/XLM-R	Various NLP tasks	[13]
Various	Zero-shot/Few-shot Learning	Translation, Language Understanding	[15]

IV. NURTURING LANGUAGE REVIVAL: AI-POWERED GENERATION OF CULTURALLY SENSITIVE TEACHING MATERIALS

AI holds considerable potential in assisting with the generation of language teaching materials for endangered languages, offering scalable and customizable solutions [3]. AI can create a wide range of resources, including reading passages tailored to different proficiency levels, interactive exercises, and engaging quizzes, all adapted to the specific needs of individual learners [4]. Furthermore, AI-powered language learning platforms, such as chatbots and virtual tutors, can simulate real-life conversations, providing immersive and interactive environments for learners to practice speaking and understanding the target language [3].

However, the generation of language teaching materials for endangered languages necessitates a deep consideration of cultural nuances and context [3]. Language is inextricably linked to culture, and effective teaching materials must go beyond mere linguistic instruction to incorporate the rich cultural heritage embedded within the language [3]. Without careful attention, AI-generated content risks reflecting biases present in its training data or failing to capture the subtle yet

crucial cultural undertones that give language its true meaning [18].

Several methods and frameworks can be employed to ensure cultural sensitivity in AI-generated content. Incorporating culturally specific datasets during the training of AI models is paramount [19]. This involves utilizing diverse datasets that represent the specific language, its dialects, and the cultural context in which it is spoken. Engaging community members and language experts throughout the development and validation process is equally crucial [18]. Their expertise can guide the AI in understanding and representing cultural nuances accurately and can help to identify and rectify any potential biases or misrepresentations. Moreover, developing AI models that can understand and adapt to different cultural contexts is essential for creating materials that resonate authentically with the target audience [20]. This includes the ability to handle idiomatic expressions, cultural references, and varying norms of formality and politeness.

V. Navigating the Ethical Landscape: Responsible AI Deployment in Linguistic Communities

The deployment of AI technologies within vulnerable linguistic communities necessitates careful consideration of ethical implications and potential risks [21]. AI models trained on biased data can inadvertently perpetuate cultural stereotypes or fail to perform equitably for certain dialects or accents [21]. Concerns regarding the ownership and control of linguistic data, as well as the potential for misrepresentation or cultural appropriation, must also be addressed [23].

Establishing robust ethical frameworks is crucial to guide the responsible development and deployment of AI in this sensitive context [24]. These frameworks should prioritize key principles such as community consent and ownership of data, ensuring that linguistic communities have control over how their language is used and represented [23]. Transparency and accountability are also essential, requiring that AI algorithms and decision-making processes are understandable and that mechanisms are in place to address errors or biases [22]. Furthermore, fairness and inclusivity must be central to the development process, ensuring that AI models benefit all members of the community and do not exacerbate existing inequalities [22]. Respect for cultural values and norms is paramount, demanding that AI systems are designed to be culturally sensitive and avoid perpetuating harmful stereotypes or misrepresentations [22].

TABLE III
ETHICAL PRINCIPLES, RISKS, AND MITIGATION STRATEGIES

Ethical Principle	Potential Risk	Mitigation Strategy
Community Consent	Data misuse, lack of control over language representation	Informed consent protocols, community involvement in data governance
Transparency	Opaque algorithms, lack of understanding of AI decisions	Explainable AI techniques, clear documentation of model development
Fairness and Inclusivity	Bias perpetuation, unequal performance across dialects	Diverse datasets, bias detection and correction methods
Cultural Respect	Cultural misrepresentation, perpetuation of stereotypes	Culturally specific datasets, community involvement in model validation

VI. Finding Linguistic Kinships: Identifying Semantic Similarities Across Languages with AI

A key research area involves the development of computational models and techniques to identify semantic similarities between endangered languages and well-documented ones [4]. This endeavor can unlock valuable linguistic insights and facilitate the development of cross-lingual resources.

One approach involves the use of cross-lingual word embeddings, which aim to map words from different languages into a shared semantic space based on their meaning [17]. Techniques like machine translation can also be leveraged to compare the semantic content of languages by translating between them and analyzing the relationships between the translated texts [26]. Ontology-based approaches offer another avenue, utilizing structured knowledge representations to identify conceptual overlaps and relationships between languages [27].

Identifying semantic similarities holds numerous potential benefits. It can facilitate language learning by highlighting cognates and semantically related words between a learner's native language and the endangered language. It can also aid in linguistic research, providing valuable data for historical linguistics and comparative studies. Furthermore, it can contribute to the development of cross-lingual resources such as dictionaries and translation tools, even for languages with limited parallel data [28].

VII. Empowering Voices: Designing Community-Centered AI Systems for Language Revitalization

Effective language preservation and revitalization efforts require AI systems that are designed with the needs and expertise of the linguistic communities at their core [1]. This community-centered approach ensures that AI tools are culturally relevant, ethically sound, and truly empower local speakers in their efforts to safeguard their linguistic heritage.

Participatory design is a fundamental aspect of this approach, involving community members in every stage of AI system development, from identifying needs and priorities to testing and deploying the final tools [31]. This collaborative process ensures that the AI solutions are tailored to the specific context and goals of the community. Respect for local knowledge and practices is also crucial, integrating traditional linguistic knowledge and cultural practices into the AI models [30]. This acknowledges the invaluable expertise held within the community and ensures that the AI systems are grounded in authentic linguistic and cultural understanding.

The primary focus of community-centered AI should be on addressing the specific benefits and needs identified by the community itself [3]. This ensures that the AI tools are practical and contribute meaningfully to the community's language preservation and revitalization goals. Furthermore, designing systems for sustainability and local ownership is essential for long-term success [34]. This involves creating tools that the community can maintain and manage independently, fostering a sense of ownership and responsibility. The overall shift in

this paradigm is from viewing AI as a replacement for human expertise to recognizing its role in augmenting and supporting community-led efforts, where the unique knowledge and cultural understanding of human speakers remain central [4].

VIII. Stories of Success: Case Studies in Community-Led AI Language Preservation

Several inspiring case studies demonstrate the successful application of AI in language preservation through community empowerment and collaboration with language experts [1]. The Living Dictionaries project, for instance, exemplifies a community-driven initiative where individuals collaboratively document their endangered languages through an online platform [1]. Google Arts & Culture's Woolaroo app provides an open-source tool designed to engage younger generations with Indigenous languages through interactive exploration [1].

Google's partnership with the Centre of Excellence for the Dynamics of Language (CoEDL) showcases a collaboration between a technology company and academic researchers to develop automatic speech recognition systems for languages with very few speakers, directly supporting documentation efforts [3]. The City of Boston's use of generative AI to provide real-time translation services for their 311 system in fourteen languages demonstrates AI's potential to enhance community engagement and inclusivity [3].

The development of the CHOCO Choctaw language corpus and the Masheli bilingual chatbot, in collaboration with the Choctaw Nation, illustrates a community-driven approach to creating resources for language learning and conversational practice [35]. Initiatives like Indigenous in AI and Indigi-Genius' Lakota AI Code Camp actively involve Indigenous engineers and youth in developing AI tools for language preservation and cultural documentation [36]. The success of these projects highlights the importance of community involvement, cultural sensitivity, and the effective integration of AI tools to address specific needs in documentation, language learning, and community engagement [1].

IX. Tackling the Data Deficit: Challenges and Opportunities in Limited Dataset Scenarios

Working with endangered languages often presents the significant challenge of limited and specialized datasets [3]. The lack of extensive written or spoken corpora can hinder the training of robust AI models, potentially leading to inaccuracies and an inability to fully capture the linguistic nuances of these languages [3]. This data scarcity impacts model performance, generalization capabilities, and the ability to accurately represent the complexities of language [3].

Despite these challenges, several opportunities and strategies exist for overcoming the data deficit. Data augmentation techniques, such as back translation and the generation of synthetic data, can help to expand the available training data [16]. Leveraging multimodal data, by combining audio, visual, and contextual information, can provide richer insights even when data is limited in each individual modality [9]. Targeted data collection efforts, in close collaboration with linguistic communities, can prioritize the gathering of high-quality data relevant to specific tasks [1]. Furthermore, the application of transfer learning and few-shot learning techniques, as discussed earlier, allows models to learn effectively from limited data by leveraging knowledge from other languages. The emergence of models specifically designed for low-resource scenarios, such as small language models (SLMs), offers a promising avenue for creating cost-effective and efficient tools for indigenous communities facing data limitations [32].

X. Conclusion: A Multimodal Future for Language Preservation

The findings of this analysis underscore the transformative potential of multimodal deep learning in addressing the urgent crisis of endangered language loss. By integrating audio, visual, and contextual data, AI offers a pathway to achieve a more comprehensive and nuanced understanding of language [7]. The critical role of transfer learning in overcoming the inherent data scarcity associated with endangered languages cannot be overstated, enabling the application of advanced NLP techniques even with limited resources [4]. However, the successful deployment of these technologies hinges on a community-centered and ethically grounded approach, ensuring that the needs and values of linguistic communities are prioritized at every stage [1].

Future research should focus on developing more robust multimodal models specifically tailored for low-resource languages, refining techniques for ensuring cultural sensitivity and mitigating biases in AI-generated content, and fostering stronger, more collaborative partnerships between AI researchers, linguists, and the linguistic communities themselves [18]. By embracing a multimodal future for language preservation, we can harness the power of AI to safeguard the world's rich linguistic heritage, ensuring its survival and vitality for generations to come.

References

[1] "AI-Powered Preservation of Endangered Languages," https://www.historica.org/blog/ai-powered-preservation-of-endangered-languages, accessed on March 24, 2025.

[2] "A Sociocognitive Approach to Language Death, Identity Loss, and Preservation in the Age of Artificial Intelligence - Greenwich Academic Literature Archive (GALA)," https://gala.gre.ac.uk/id/eprint/46509/7/46509_ DAY %20Endangered languages _A sociocognitive approach_ to language death identity loss _and preservation.pdf, accessed on March 24, 2025.

[3] "Machine Learning for Endangered Language Preservation - TRENDS Research & Advisory," https://trendsresearch.org/insight/machine-learning-for-endangered-language-preservation/, accessed on March 24, 2025.

[4] "Generative AI and Large Language Models in Language Preservation: Opportunities and Challenges - arXiv," https://arxiv.org/html/2501.11496v1, accessed on March 24, 2025.

[5] "Multimodal Deep Learning — Papers With Code," https://paperswithcode.com/task/multimodal-deep-learning, accessed on March 24, 2025.

[6] "Multimodal Deep Learning: Definition, Examples, Applications - V7 Labs," https://www.v7labs.com/blog/multimodal-deep-learning-guide, accessed on March 24, 2025.

[7] "CM^2: Cross-Modal Contextual Modeling for Audio-Visual Speech Enhancement," https://openreview.net/forum?id=EO2hZTtK3M, accessed on March 24, 2025.

[8] "Reviving Indigenous Languages using Machine Learning - Insights2TechInfo," https://insights2techinfo.com/reviving-indigenous-languages-using-machine-learning/, accessed on March 24, 2025.

[9] "(PDF) Multimodal Deep Learning - ResearchGate," https://www.researchgate.net/publication/221345149_Multimodal_Deep_Learning, accessed on March 24, 2025.

[10] "Integrating Visual Context Into Language Models for Situated Social Conversation Starters," https://www.computer.org/csdl/journal/ta/5555/01/10599883/1YD8cCzVKj6, accessed on March 24, 2025.

[11] "AVLnet: Learning Audio-Visual Language Representations from Instructional Videos - ISCA Archive," https://www.isca-archive.org/interspeech_2021/rouditchenko21_interspeech.pdf, accessed on March 24, 2025.

[12] "AVicuna: Audio-Visual LLM with Interleaver and Context-Boundary Alignment for Temporal Referential Dialogue - arXiv," https://arxiv.org/html/2403.16276v1, accessed on March 24, 2025.

[13] "Natural language processing applications for low-resource languages - Cambridge University Press & Assessment," https://www.cambridge.org/core/journals/natural-language-processing/article/natural-language-processing-applications-for-lowresource-languages/7D3DA31DB6C01B13C6B1F698D4495951, accessed on March 24, 2025.

[14] "Transfer Learning in Natural Language Processing (NLP): A Game-Changer for AI Models — by Hassaan Idrees — Medium," https://medium.com/@hassaanidrees7/transfer-learning-in-natural-language-processing-nlp-a-game-changer-for-ai-models-b8739274bb02, accessed on March 24, 2025.

[15] "(PDF) AI-Augmented Dynamic Language Adaptation for Low-Resource Languages: A Transfer Learning Solution for Optimized NLP Performance - ResearchGate," https://www.researchgate.net/publication/384861943_AI-Augmented_Dynamic_Language_Adaptation_for_Low-Resource_Languages_A_Transfer_Learning_Solution_for_Optimized_NLP_Performance, accessed on March 24, 2025.

[16] "A Guide to Low-Resource Natural Language Processing," https://www.marstranslation.com/blog/a-guide-to-low-resource-natural-language-processing, accessed on March 24, 2025.

[17] "Evaluating Semantic Textual Relatedness in Low-resource Languages using Various Embedding - ACL Anthology," https://aclanthology.org/2024.semeval-1.65.pdf, accessed on March 24, 2025.

[18] "The Challenges Of Preserving Cultural Nuance In AI-Generated Content," https://internationalachieversgroup.com/localization-ai/the-challenges-of-preserving-cultural-nuance-in-ai-generated-content/, accessed on March 24, 2025.

[19] "How Generative AI Impacts Non-English Content Creation? - Identrics," https://identrics.ai/generative-ai-and-its-impact-on-non-english-content/, accessed on March 24, 2025.

[20] "Adapting Models to Handle Cultural Variations in Language and Context - Welocalize," https://www.welocalize.com/insights/adapting-models-to-handle-cultural-variations-in-language-and-context/, accessed on March 24, 2025.

[21] "The Impact of AI on Enhancing Students' Intercultural Communication Competence at the University Level: A Review Study - ERIC," https://files.eric.ed.gov/fulltext/EJ1457270.pdf, accessed on March 24, 2025.

[22] "Navigate a Maze of AI Ethical Issues in Language Processing - Waywithwords.net," https://waywithwords.net/resource/ai-ethical-issues-language-processing/, accessed on March 24, 2025.

[23] "Generative AI and Large Language Models in Language Preservation: Opportunities and Challenges - arXiv," https://arxiv.org/html/2501.11496, accessed on March 24, 2025.

[24] "AI Ethics Statement - SIL International," https://www.sil.org/ai-ethics-statement, accessed on March 24, 2025.

[25] "Localized Ethical Frameworks: Aligning AI with Regional Cultural and Societal Norms — AIGN," https://aign.global/ai-ethics-consulting/patrick-upmann/localized-ethical-frameworks-aligning-ai-with-regional-cultural-and-societal-norms/, accessed on March 24, 2025.

[26] "Can Lexical Semantics Improve Translation In Low-resource And Endangered Languages?," https://www.researchgate.net/publication/368642499_Can_Lexical_Semantics_Improve_Translation_In_Low-resource_And_Endangered_Languages, accessed on March 24, 2025.

[27] "Language engineering for the Semantic Web: a digital library for endangered languages. - Information Research," https://informationr.net/ir/9-3/paper176.html, accessed on March 24, 2025.

[28] "Neural Transfer Learning for Truly Low-Resource Natural Language ...," https://ethesis-old.helsinki.fi/repository/handle/123456789/47712, accessed on March 24, 2025.

[29] "The Potential of AI in Language Preservation and Revival - Teachflow," https://teachflow.ai/the-potential-of-ai-in-language-preservation-and-revival/, accessed on March 24, 2025.

[30] "(PDF) Multimodal Generative AI for African Language Preservation ...," https://www.researchgate.net/publication/387668586_Multimodal_Generative_AI_for_African_Language_Preservation_A_Framework_for_Language_Documentation_and_Revitalization, accessed on March 24, 2025.

[31] "AI for Community: Preserving Culture and Tradition - 1st Edition - Ira," https://www.routledge.com/AI-for-Community-Preserving-Culture-and-Tradition/Ardalan-Banifatemi-Gonzalez-Ingram-Moradinezhad-Williams/p/book/9781032846620, accessed on March 24, 2025.

[32] "Can small language models revitalize Indigenous languages? - Brookings Institution," https://www.brookings.edu/articles/can-small-language-models-revitalize-indigenous-languages/, accessed on March 24, 2025.

[33] "The Role of AI in Supporting Indigenous Languages - KMAN Publication Inc.," https://www.journals.kmanpub.com/index.php/aitechbesosci/article/download/2875/4148/14577, accessed on March 24, 2025.

[34] "National Consultation on Indigenous Languages: A Call for Preservation and Technology Integration — UNESCO," https://www.unesco.org/en/articles/national-consultation-indigenous-languages-call-preservation-and-technology-integration, accessed on March 24, 2025.

[35] "Using Artificial Intelligence to Preserve Indigenous Languages - USC Institute for Creative Technologies - University of Southern California," https://ict.usc.edu/news/essays/using-artificial-intelligence-to-preserve-indigenous-languages/, accessed on March 24, 2025.

[36] "How Indigenous Engineers Are Using AI to Preserve Their Culture - NBCU Academy," https://nbcuacademy.com/how-indigenous-engineers-are-using-ai-to-preserve-their-culture/, accessed on March 24, 2025.

[37] "Opportunities and Challenges of Large Language Models for Low-Resource Languages in Humanities Research - arXiv," https://arxiv.org/html/2412.04497v2, accessed on March 24, 2025.