AI-DRIVEN LANGUAGE REVITALIZATION:
AI-POWERED LANGUAGE LEARNING AS A CATALYST FOR LANGUAGE PRESERVATION

Mrityunjay Chauhan
22BAI71406
Dept. Of Engineering
Chandigarh University
Mohali, Punjab, India
mrityunjaychauhan0102@gmail.com
ORCID- 0009-0009-7760-4412

Atharva Durge
22BAI71399
Dept. Of Engineering
Chandigarh University
Mohali, Punjab, India
atharvadurge762@gmail.com
ORCID- 0009-0009-7760-4412

Shreya
22BAI7099
Dept. Of Engineering
Chandigarh University
Mohali, Punjab, India
shreyagupta895@gmail.com
ORCID- 0009-0004-6861-4612

Vishal
22BAI71398
Dept. Of Engineering
Chandigarh University
Mohali, Punjab, India line 5: email
iamvishalnishal@gmail.com
ORCID- 0009-0000-7905-0243

Abstract—**According to UNESCO, there are nearly 7,000 languages spoken worldwide [1], of which around 3,000 languages are in danger of disappearing before the end of the century [2]. With roughly 230 languages having already become extinct between the years 1950-2010 [3], collectively this represents a significant loss of linguistic and cultural diversity. This position paper aims to explore the potential of AI-based language learning approaches that promote early exposure and appreciation of languages to ultimately contribute to the preservation of endangered languages, thereby addressing the urgent need to protect linguistic and cultural diversity.**

Keywords— *AI-Artificial Intelligence, Deep Learning, Neural Networks, Natural Language Processing, Machine Learning, Language Preservation, Transfer Learning, Convolutional Neural Networks, Endangered languages.*

I. INTRODUCTION

In a world full of linguistic multiplicity, various languages fluctuate on the edge of extinction, endangered by the unstoppable force of globalization and the dominance of the various languages. We are pointing out this global competition, our model, titled "AI-Driven Language Revitalization" goals to leverage the changing power of digital technology and artificial intelligence to prevent and boost these at-risk languages. This introduction sheds light on the objectives, various applications and innovative characteristics encapsulated within our project's range.

Here our goal is to achieve multiple objectives associated with preserving, revitalising and commemorating endangered languages. We aim to make AI-driven solutions that preserve the language by utilising advanced natural language processing techniques and machine learning. With the help of these, we try to create digital repositories and interactive platforms which provide guardians of linguistic heritage and protect at-risk languages from the edge of forgetfulness.

In addition to this, our project's goal is to foster linguistic inclusivity and cultural differences by empowering access to educational stuff and language learning sources. By nurturing the capability of digital technology, we aim to foster a global ecosystem where people from different backgrounds communicate and engage with each other in meaningful language acquisition, linguistic exchange and cultural envelopment. Through this project, we focus on strengthening the communities to revitalise, reclaim and commemorate their linguistic heritage.

This research paper aims to review the role of AI in self-driving vehicles and the advancements made in this area. We will examine the different types of AI technologies used in self-driving cars and their effectiveness. We will also explore the challenges and limitations of the current AI-based self-driving systems and possible solutions for these challenges.

The goal of AI, or artificial intelligence, is to build smart computers that can mimic human behaviour and decision-making. AI is a subfield of computer science. It comprises the development of algorithms and models that enable computers to analyse data, draw conclusions, and do tasks that were previously only performed by intelligent humans, such as speech and picture recognition, natural language processing, and gaming. Local or weak AI and global or strong AI are the two main subcategories of artificial intelligence systems. Narrow AI is designed to carry out specific tasks or address specific problems, such as making product recommendations based on consumer behaviour or recognising financial fraud. Contrarily, general AI seeks to build machines that are capable of any intellectual activity that a person is capable of, including reasoning, problem-solving, and decision-making. Numerous industries, including healthcare, banking, education, transportation, and entertainment, can benefit from the use of AI technologies. Virtual assistants, self-driving cars, speech and image recognition systems, and recommendation engines are some of the most popular AI uses. Artificial intelligence (AI) has the potential to transform many industries and improve our

daily lives, but it also presents significant ethical and societal challenges, such as the effect on human autonomy and decision-making, privacy and security concerns, bias, and job displacement. As a result, it's critical to ensure that ethical principles and consideration of potential societal consequences are used to guide AI development and implementation.

The goal of the "deep learning" branch of artificial intelligence (AI) is to construct neural networks with multiple layers that are capable of challenging tasks like speech and picture recognition, natural language processing, and decision-making. A process known as training allows deep learning models to learn from huge amounts of data. During training, the neural network modifies its weights and biases to enhance its performance on a particular task. Feeding the neural network with labelled data during training involves providing input data (such as images, audio, and text) and matching output labels. (e.g., image categories, speech transcripts, text sentiment). This information is used by the network to discover patterns and features pertinent to the current job. The primary benefit of deep learning is the ability to autonomously identify patterns and features in unprocessed data without the need for explicit feature engineering. This enables models to be more precise and adaptable and to extend to new and unexplored data. Just a few well-known Deep Learning designs include Generative Adversarial Networks (GANs) for picture and text generation, Recurrent Neural Networks (RNNs) for natural language processing and sequence modelling, and Convolutional Neural Networks (CNNs) for image and video processing. Image captioning is the task of generating natural language descriptions for images. To achieve this, researchers have used deep learning techniques such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs). One of the most significant works in this area is "Understanding Deep Learning Techniques for Image Captioning" by Zhang et al. [3]. This paper presents a comprehensive survey of the different deep learning techniques used for image captioning, including attention-based models and reinforcement learning. Deep learning is utilised in several applications, including computer vision, speech recognition, natural language processing, recommendation systems, and gaming. Multi-modal deep learning for natural language processing was explored in a survey by Xu et al. [8]. The authors discussed the integration of multiple modalities, such as text, images, and audio, to improve the performance of NLP tasks. They also highlighted the potential applications of multi-modal deep learning in SDVs, such as incorporating. Deep Learning models demand a lot of data and processing power, which can be a problem in some fields. Deep Learning models can also be complicated and challenging to read, which presents significant moral and societal issues including bias and transparency.

Neural networks are a type of computing system that are inspired by the design and function of the human brain. They are an essential component in the branch of artificial intelligence known as deep learning. (AI). Layers of interconnected neurons, or nodes, make up neural networks, which process and send information. Every neuron takes input signals from other neurons, processes those signals using a mathematical function, and then generates an output signal that is sent to further neurons in the layer below. Weights on the synapses between neurons control how strongly a signal is sent. These weights are modified during the training process to optimize the neural network's performance on a particular task. A neural network's architecture is determined by the task that it is intended to carry out. For instance, a Convolutional Neural Network (CNN) for image classification uses convolutions and pooling operations in subsequent layers to extract features from the image. The input layer gets the raw pixel values for the image. A sequence of words is input into the input layer of a recurrent neural network (RNN) used for natural language processing while the following layers use recurrent connections to record the temporal associations between words. Numerous fields, including computer vision, speech recognition, natural language processing, and gaming, have effectively used neural networks. Additionally, they have proven successful in resolving complicated issues that are challenging to represent with conventional rule-based systems. Neural networks do, however, have significant drawbacks. For training, they need a lot of labelled data, and training can be computationally expensive. Furthermore, Neural Networks can be challenging to understand and comprehend, which raises questions regarding accountability and transparency.

How computers and human language interact is a topic of interest in the "natural language processing" (NLP) branch of computer science and artificial intelligence (AI). The main goal of NLP is to enable computers to understand, interpret, and produce human language in meaningful and advantageous ways. NLP entails the creation of algorithms, models, and methods that let computers interpret the process as spoken language. This includes undertakings like Text classification: identifying a text's topic or category Analysis of a text's sentiment: and determining if it is favourable, negative, or neutral. Identification and classification of identified entities, such as individuals, locations, and organizations, in a text. Text that has been translated automatically from one language into another. Answering questions automatically based on text is known as question answering. Speech recognition is the process of turning spoken words into text. Text generation is the process of creating new text that is comparable to the current text in both style and content. Numerous applications, including chatbots, virtual assistants, recommendation engines, and automated content analysis, employ NLP extensively. NLP enables computers to comprehend and interpret human language more efficiently, creating new potential for organizations, governments, and people to use technology in more intuitive and natural ways. In natural language processing, researchers have also used deep learning techniques to achieve state-of-the-art performance in text classification. One of the most significant works in this area is "Universal Language Model Fine-tuning for Text Classification" by Howard and Ruder [4]. This paper introduces a novel technique for fine-tuning pre-trained language models on specific classification tasks, achieving state-of-the-art performance on several benchmark datasets.

Another important work in natural language processing is "Deep Contextualized Word Representations" by Peters et al. [5]. This paper presents a method for learning context-dependent word embeddings, which has achieved state-of-the-art performance on several natural language understanding tasks.

In the field of artificial intelligence (AI), machine learning is concerned with developing statistical models and algorithms that enable computer systems to improve on their own over time. In conventional programming, a programmer creates code that tells a computer how to approach a specific problem. Machine learning systems, in contrast, learn from data without explicit programming. As a result, the system learns to identify patterns in data rather than being explicitly coded, and it becomes better over time as it is exposed to more data. Reinforcement learning, unsupervised learning, and supervised learning are the three primary categories of machine learning algorithms. Supervised learning is a method for building models from labelled data when the desired output is already known. The model can predict outcomes for freshly arrived data by finding trends in the tagged data. Unsupervised learning is the process of creating a model using unlabelled data when the desired output is unknown. The model picks up on the structure and patterns in the data. Reward-based decision-making is taught to a model through reinforcement learning. When a game is won or a goal is accomplished, for example, the model learns to operate in a way that maximizes the reward signal. Image identification, audio recognition, natural language processing, recommendation systems, fraud detection, and autonomous vehicles are only a handful of the numerous applications for machine learning.

The term "data analytics" refers to the procedure of looking over and analysing huge data sets to glean valuable insights and make judgments. To find patterns, trends, and relationships in the data, several statistical and computational techniques must be used. Data analytics aims to offer businesses and organizations use information that can aid in making wise decisions and enhancing performance. A few of the steps that make up data analytics include data collection, data cleaning, data transformation, data modelling, and data visualisation. These procedures are essential for ensuring the data's correctness and dependability and for making it simpler to evaluate. As more and more businesses amass massive volumes of data, data analytics has grown in significance in today's corporate environment. Businesses can learn more about consumer behaviour, market trends, and other significant elements that can guide their decision-making by evaluating this data. Data analytics is also utilized in the medical, financial, and scientific industries to find patterns and trends that might help with research and decision-making.

II. Literature Review

A survey of deep learning in natural language processing by Wang et al. [6] explored the state-of-the-art deep learning techniques and their applications in various NLP tasks, including machine translation, sentiment analysis, and speech recognition. The authors also highlighted the challenges and opportunities for future research in the field.

Preserving endangered native languages is crucial to maintaining cultural heritage, and it is especially important to introduce young children to their native language in a way that goes beyond basic communication and emphasizes language appreciation. For example, Artificial Intelligence (AI) has the potential to be utilized for the creation of bilingual storybooks that can be written in two languages, including the native language and a second language, such as English. In addition, generative AI like DALL-E or Midjourney can be used to develop illustrations that can highlight culturally unique expressions and aspects of the native language. Through an analysis of the story's text, AI can create vivid visual representations of the characters and settings that are inclusive of clothing, food, and decorations specific to the culture. By creating interactive images, the AI-generated illustrations can evoke inquisitiveness in children, prompting them to ask questions about the new elements they are seeing. Furthermore, the AI has the capacity to explain each aspect of the image while simultaneously weaving it into the storyline. By exposing children to the poetry, stories, fables, and other artistic expressions of their native language, they can develop a deeper understanding and appreciation for the phonology, grammar, and script of the language.

The Department of School Education & Literacy has initiated Bhasha Sangam - A Celebration of Linguistic Diversity which marks the appreciation of the unique symphony of languages of our country. To celebrate the unique characteristic of our country, Bhasha Sangam provides an opportunity for schools and educational institutions (BIETs, DIETs, CTEs/IASEs, SCERTs, SIEs, School Boards, Directorates of School Education, etc.) to provide multilingual exposure to students in Indian Languages. The objective is to familiarize every child with simple dialogues in all the 22 languages under Schedule VIII of the Constitution of India, taking up one language on each working day, to enhance linguistic tolerance and promote national integration.[2]

III. Applications

The implementation of our project broadens across a variety of domains surrounding cultural preservation, language revival, and technological advancement. We seek to promote interactive language learning experiences customized to the desired needs and tastes of different learners through the enhancement of AI-powered platforms and various web-based tools. It also visualizes the inclusivity of AI-driven translation and documentation services, enabling never-ending communication and spreading knowledge across linguistic challenges. Additionally, this project delves into fostering the power of digital media and content formation to increase the voices of linguistic Neighbourhoods and increase their cultural histories. we aim to spark a rebirth of cultural representation and linguistic innovation by taking initiatives of multimedia production, and online radio stations. Furthermore, our project is made to measure the incorporation of AI-driven solutions in various educational blueprints and government policies to preserve and promote endangered languages.

Utilizing AI-generated content that includes heritage and culturally specific elements can help promote a sense of pride and identity in their cultural heritage, which is particularly important for children who may be growing up in a multicultural environment. This can help foster a sense of connection and belonging, which can have positive implications for their overall well-being. It could facilitate connections between children from the same cultural background, promoting socialization and the formation of friendships based on shared cultural experiences. This could help foster a sense of community and belonging, which can be particularly important for children who may be living in areas where they are in the minority

IV. Machine Translation and Revitalization

Machine Translation, an integral facet of Artificial Intelligence (AI), serves as a potent ally in the preservation of endangered languages. Its ability to bridge linguistic barriers by translating texts from endangered languages into more widely spoken ones facilitates the dissemination of knowledge and cultural insights across diverse communities. By harnessing AI-driven translation tools, endangered languages gain visibility and accessibility, ensuring their perpetuation in an increasingly interconnected world. AI-powered translation tools enable the translation of texts, documents, and literature from endangered languages into major languages, thereby preserving invaluable knowledge, stories, and cultural heritage embedded within these languages. Moreover, Machine Translation plays a pivotal role in language revitalization efforts. Interactive applications, language learning platforms, and AI-driven Chabot facilitate language learning and engagement with endangered languages. These tools create immersive experiences for learners, allowing them to actively engage with the language through interactive exercises, conversations, and cultural activities. The integration of Machine Translation in language revitalization programs amplifies access to educational resources in endangered languages. By translating educational materials, including textbooks, websites, and multimedia content, AI-driven translation tools ensure that learners have access to comprehensive educational resources in their native or endangered languages. Communities speaking endangered languages can utilize these translation tools to create digital content, translate literature, and communicate their cultural narratives to a wider audience. However, challenges persist in Machine Translation, including linguistic nuances, idiomatic expressions, and contextual accuracy. Endangered languages often possess intricate grammatical structures and cultural nuances that require meticulous translation, posing challenges for AI algorithms. Addressing these challenges remains crucial for ensuring accurate and culturally sensitive translations. Machine Translation, driven by AI, plays a pivotal role in both preserving endangered languages and revitalizing their usage. Its ability to facilitate translation, enhance access to educational resources, foster community engagement, and empower language communities underscores its significance in the broader effort to safeguard linguistic diversity.

V. Methodology

The research aims to address the challenge of language revitalization by developing a machine-learning model for English-to-Marwari translation. Our objective is to develop a model that can translate English phrases into Marwari smoothly and correctly, promoting communication and protecting the Marwari language.

A small dataset comprising English-Marwari translation pairs was created. This dataset was crucial for training and evaluating the translation model. The dataset included common phrases and sentences in English along with their corresponding translations in Marwari. Care was taken to ensure diversity in the dataset, covering various linguistic aspects and scenarios likely to be encountered in practical usage.

The data gathering and preparation method for the English to Marwari translation model included many important processes designed to ensure the dataset's quality and usefulness for training and assessment. Our major objective was to collect a comprehensive set of English-Marwari translation pairs that could address a wide variety of linguistic settings and circumstances found in ordinary communication. To do this, we gathered information from a variety of credible sources, including existing bilingual corpora, real conversations, and contributions from native Marwari speakers and language specialists. Once collected, the data was meticulously curated to ensure relevance, accuracy, and variety. To ensure the inclusion of culturally relevant information and idiomatic expressions unique to the Marwari language and culture, typical English phrases and sentences were paired with their equivalent translations in Marwari. To ensure that translations were accurate and to identify any discrepancies or mistakes, human reviewers who were fluent in both Marwari and English were employed. Furthermore, particular focus was placed on resolving linguistic difficulties that arise while translating from English to Marwari, including variations in sentence construction, word order, grammatical conventions, and colloquial idioms. To guarantee that the translated materials accurately conveyed the message, techniques including contextual translation and cultural subtleties were used. Ultimately, preparatory techniques like text normalisation and tokenization were used to prepare the carefully chosen dataset into an appropriate input format. The resulting dataset, characterized by its quality, diversity, and suitability, served as the foundation for training the translation model, enabling it to produce accurate and culturally appropriate translations from English to Marwari.

The process of selecting the best base model for fine-tuning the English to Marwari translation project required careful evaluation of several aspects. We aimed to identify a pre-trained model that would work well for the translation problem and be consistent with the transfer learning methodology. We selected the pre-trained Helsinki-NLP/opus-mt-en-mr model after weighing our possibilities. its proven effectiveness in low-resource language translation tasks and its compatibility with the MarianMTModel architecture, which aligns with our project requirements.

The Helsinki-NLP model is a good option for fine-tuning because of its promising performance in accurately collecting common phrases and linguistic subtleties in Marwari. The architecture and parameters of the model are useful for adapting to our translation requirement, offering us a strong basis for more improvement through fine-tuning. We chose the Helsinki-NLP/opus-mt-en-mr model as our foundation model to take advantage of its advantages and capabilities to create a strong English-to-Marwari translation model that can translate words accurately and fluently while maintaining the linguistic and cultural subtleties of the Marwari language.

test set	BLEU	chr-F
Tatoeba.en.mr	22.0	0.397

Table-1 Benchmarks for Helsinki-NLP/opus-mt-en-mr

First, the pre-trained Helsinki-NLP/opus-mt-en-mr model is loaded using the MarianMTModel and Marian Tokenizer classes from the transformer's library. After that, the dataset containing English sentences paired with their corresponding Marwari translations is prepared for fine-tuning. This dataset provides the base for training the model to improve its translation capabilities specifically for Marwari. Every pair of sentences is tokenized using the tokenizer, converting the text into numerical tokens that the model can comprehend. The data is then formatted into input-output pairs suitable for training. The fine-tuning process involves iterative training of the model using the prepared dataset. The model's parameters are adjusted during training to minimize translation errors, with optimization algorithms like AdamW utilized to update the parameters and reduce translation loss. This process occurs over multiple epochs, with each epoch consisting of iterations through the entire dataset. Evaluation and validation are performed periodically to monitor the model's performance and prevent overfitting, ensuring that the model generalizes well to unseen data and produces accurate translations beyond the training set. Upon completion of the fine-tuning process, the fine-tuned model's parameters are saved for future use. This specialized model is capable of accurately translating English text into Marwari, catering specifically to the requirements of the language revitalization research project.

The training and testing phases play pivotal roles in assessing and optimizing the model's performance. The training phase involves iteratively exposing the model to the prepared dataset consisting of English sentences paired with their corresponding Marwari translations. During training, the model learns to adjust its parameters to minimize translation errors and accurately map English input sequences to Marwari output sequences. This process occurs over multiple epochs, with each epoch representing a complete pass through the training dataset. Within each epoch, the dataset is typically divided into batches, allowing the model to process multiple sentence pairs simultaneously. For each batch, the model computes translation predictions for the English input sentences and compares them to the ground truth Marwari translations using a predefined loss function. Optimization algorithms, such as AdamW, are then employed to update the model's parameters based on the computed loss, gradually improving the model's translation capabilities. After completing the training phase, the model undergoes testing to evaluate its performance on unseen data. This testing phase is essential for assessing the model's generalization ability and ensuring that it can accurately translate English sentences into Marwari beyond the training dataset. Typically, a separate dataset, known as the test set, is used for evaluation. The test set contains English sentences paired with their corresponding Marwari translations, but these examples were not seen by the model during training. During testing, the fine-tuned model generates Marwari translations for the English sentences in the test set. The quality of these translations is then assessed using various evaluation metrics, such as BLEU score, which measures the similarity between the model-generated translations and the reference translations in the test set.
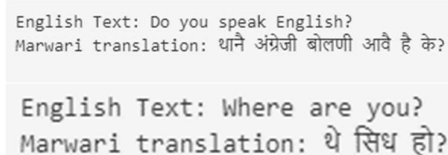
In the model evaluation and testing phase, the performance of the fine-tuned model was assessed using appropriate evaluation metrics. Metrics such as the loss function value (Transformers models all have a default task-relevant loss function), translation accuracy, and fluency were utilized to gauge the effectiveness of the model in generating accurate and linguistically sound Marwari translations. The model was subjected to testing with unseen English input sentences to generate corresponding Marwari translations. These generated translations were manually reviewed and compared against human-generated translations to assess correctness and linguistic quality accurately. This manual review process helped validate the model's performance and identify any areas for improvement. Upon satisfactory performance evaluation, the fine-tuned model was prepared for deployment. This involved saving both the model weights and tokenizer configurations using the save pretrained method, ensuring that the trained model could be easily accessed and utilized for future tasks.

The deployed model held significant potential for integration into language revitalization initiatives, where it could facilitate English-to-Marwari translation tasks. By making the model accessible to users, such as Marwari speakers and language enthusiasts, it could contribute to the preservation and promotion of the Marwari language in digital communication and educational settings.

RESULTS

Our findings reveal that the model, trained on a dataset consisting of 75 translations, achieved an accuracy of 22%. This performance metric underscores the model's effectiveness in translating English sentences to Marwari, albeit with certain limitations due to the relatively small size of the training dataset. Despite these constraints, the results demonstrate the potential of transfer learning in addressing language translation tasks, particularly in low-resource language settings. Further research endeavours are warranted to explore strategies for augmenting the training data and refining the model architecture to improve translation accuracy and robustness.

English Text: Do you speak English?
Marwari translation: थानै अंग्रेजी बोलणी आवै है के?

English Text: Where are you?
Marwari translation: थे सिध हो?

Fig1. Resulting translation outputs

This project underscores the transformative potential of transfer learning in natural language processing, particularly in the context of language revitalization. By employing transfer learning techniques, we have leveraged the knowledge encoded in pre-trained language models and adapted it to the specific task of English to Marwari translation. Transfer learning enabled us to overcome the challenge of limited data availability for Marwari by fine-tuning a pre-trained translation model on a small dataset of English-Marwari sentence pairs. This approach not only expedited the training process but also allowed the model to capture the linguistic patterns and nuances of the Marwari language. Moreover, the use of transfer learning has broader implications for language preservation and promotion efforts. By reducing the dependency on large, annotated datasets and lowering the barriers to entry for language technology development, transfer learning democratizes access to

advanced AI capabilities and empowers language communities to reclaim and revitalize their linguistic heritage.

CHALLENGES AND FUTURE PROSPECTS

While the integration of Artificial Intelligence (AI) presents promising solutions for preserving endangered languages, it is not without its challenges. One significant obstacle lies in the scarcity of resources and linguistic data available for training AI models. Endangered languages often lack extensive digital resources or substantial corpora necessary to train AI algorithms effectively. This scarcity impedes the development of robust AI tools tailored to these languages, limiting their accuracy and effectiveness. These languages exhibit diverse dialects, variations, and nuances within their structures, presenting hurdles for AI systems accustomed to more standardized linguistic patterns. As a result, developing AI tools that accurately capture these variations remains a considerable challenge in preserving linguistic diversity. Ethical considerations also loom large in AI-based language preservation initiatives. Questions regarding intellectual property rights, cultural representation, and ownership of linguistic data demand careful consideration. It is essential to navigate these ethical concerns sensitively, ensuring that AI technologies do not inadvertently exploit or misrepresent endangered language communities and their heritage. Despite these challenges, the future prospects for AI in preserving endangered languages are promising. Technological advancements continue to refine AI algorithms, offering prospects for more accurate and culturally sensitive language preservation tools. Developments in Natural Language Processing (NLP) and machine learning techniques hold the potential to address linguistic variations and intricacies within endangered languages. Furthermore, collaborative efforts between linguists, technology experts, and indigenous communities are pivotal in overcoming challenges and shaping the future of AI-powered language preservation. Community engagement is critical for gathering linguistic data, refining AI models, and ensuring that preservation efforts align with the cultural values and needs of the language speakers. Looking ahead, the future of AI-driven language preservation holds potential breakthroughs in addressing linguistic documentation and revitalization challenges. Enhanced AI models that account for dialectal variations, sophisticated language-learning platforms, and more context-aware translation systems are on the horizon. These advancements promise more accurate transcriptions, culturally sensitive translations, and immersive language learning experiences. The collaborative synergy between AI technology and human expertise offers a promising trajectory for the preservation of endangered languages. AI serves as a powerful tool that, when ethically applied and integrated with community engagement, can play a pivotal role in not only preserving but also revitalizing endangered languages for generations to come. This comprehensive exploration touches upon the challenges faced in using AI to preserve endangered languages while highlighting the potential future advancements and the crucial role of collaborative efforts in shaping the preservation landscape.

CONCLUSION

The endeavour to preserve endangered languages stands as an indispensable task in safeguarding the rich tapestry of human cultural heritage. In this quest, the integration of Artificial Intelligence (AI) emerges as a beacon of hope, offering innovative solutions to counter the looming threat of linguistic extinction. Throughout this exploration, we have witnessed how AI-driven technologies revolutionize language preservation efforts, from documentation and transcription to translation and revitalization. AI plays a pivotal role in Language Documentation, offering advanced tools for recording, transcribing, and analysing oral languages. These technologies facilitate the meticulous archiving of endangered languages, capturing the essence of oral traditions, dialects, and linguistic nuances that shape cultural identities. Moreover, the collaborative nature of AI-powered tools encourages community involvement, empowering indigenous speakers to actively participate in preserving their linguistic heritage. The integration of Automatic Speech Recognition (ASR) technology amplifies language preservation endeavours by transcribing oral languages into written forms. Despite challenges posed by dialectal variations and tonal intricacies, ASR stands as a transformative tool immortalizing oral traditions, narratives, and linguistic nuances. By bridging the gap between oral and written forms, ASR facilitates the integration of endangered languages into educational materials, digital archives, and communication platforms. Machine Translation, another facet of AI, serves as a potent ally in preserving linguistic diversity. Its ability to translate texts from endangered languages into widely spoken ones ensures the perpetuation of invaluable knowledge and cultural insights. Moreover, Machine Translation aids in language revitalization efforts through interactive applications, language learning platforms, and educational resources, fostering engagement and active usage of endangered languages. However, these endeavours are not without challenges. Resource limitations, linguistic complexities, and ethical considerations pose significant hurdles in harnessing AI for language preservation. Scarcity of linguistic data, dialectal variations, and ethical dilemmas demand thoughtful navigation and collaborative approaches to ensure culturally sensitive and accurate preservation efforts. Yet, the future prospects for AI in language preservation are promising. Technological advancements promise more refined AI models capable of addressing linguistic variations and nuances within endangered languages. Collaborative efforts between linguists, technology experts, and indigenous communities offer a path forward, ensuring that AI-driven preservation aligns with cultural values and community needs. In conclusion, AI stands as a catalyst for change, offering unprecedented opportunities to protect and celebrate the linguistic diversity that defines humanity's cultural mosaic. The convergence of AI and language preservation heralds a new era where no language is left behind but instead thrives as a testament to the rich tapestry of human expression and heritage

REFERENCES

[1] Languages UNESCO WAL. (n.d.). En.wal.unesco.org. https://en.wal.unesco.org/languages

[2] *Bhasha Sangam - Celebrating the Linguistic Diversity of India| National Portal of India.* (n.d.). https://www.india.gov.in/spotlight/bhasha-sangam-celebrating-linguistic-diversity-india

[3] UNESCO Institute for lifelong Learning. (2022, February 21). A decade to prevent the disappearance of 3,000 languages. UNESCO.

https://www.iesalc.unesco.org/en/2022/02/21/a-decade-to-prevent-the-disappearance-of-3000-languages/

[4] Strochlic, N. (2018, April 16). Saving the World's Dying and Disappearing Languages. Culture; National Geographic.

https://www.nationalgeographic.com/culture/article/saving-dying-disappearing-languages-wikitongues-culture

[5] Y. Zhang, J. Wang, J. Zhao, Y. LeCun, and M. Paluri, "Understanding Deep Learning Techniques for Image Captioning," IEEE Transactions on Multimedia, 2021.

[6] Howard, J., & Ruder, S. (2018, January 1). *Universal Language Model Fine-tuning for Text Classification.* https://doi.org/10.18653/v1/p18-1031

[7] Peters, M., Neumann, M., Iyyer, M., & Zettlemoyer, L. (2018, February 14). *Deep contextualized word representations.* ResearchGate. https://www.researchgate.net/publication/323217640_Deep_contextualized_word_representations

[8] X. Wang, J. Zhang, Y. Jiang, Z. Dong, and Q. Liu, "A Survey of Deep Learning in Natural Language Processing," IEEE Transactions on Neural Networks and Learning Systems, 2021.

[9] W. Guo, J. Wang and S. Wang, "Deep Multimodal Representation Learning: A Survey," in IEEE Access, vol. 7, pp. 63373-63394, 2019, doi: 10.1109/ACCESS.2019.2916887.

[10] https://magazine.amstat.org/blog/2021/09/01/stats4good_ai/ 2.

[11] https://deepgram.com/learn/language-ai-cultural-preservation 3.

[12] https://mindmatters.ai/2020/09/surprising-ways-ai-can-help-recover-lost-languages/ 4.

[13] https://www.forbes.com/sites/cognitiveworld/2018/11/23/turning-to-ai-to-save-endangeredlanguages/?sh=6db9e6896f45 5.

[14] https://www.fairplanet.org/story/embracing-artificial-intelligence-to-preserve-dyinglanguages