 ISSN: 2349-6002

Artificial Intelligence and Multilingualism in NEP-2020: A Comparative Linguistic Analysis of Indo-Aryan, Dravidian, and Global Languages

Smruti Vadher[1], Sejad Chokiya[2]
[1,2]*Postgraduate student in the Department of English at Maharaja Krishnakumarsinhji Bhavnagar University (MKBU)*
doi.org/10.64643/IJIRTV12I8-191210-459

***Abstract*—The National Education Policy (NEP-2020) places unprecedented emphasis on multilingualism, linguistic diversity, and the integration of modern technologies in higher education. In this context, Artificial Intelligence (AI) emerges as a transformative force in comparative linguistics, translation studies, and digital language learning. This study examines how AI-driven tools such as machine translation systems, natural language processing (NLP) models, and corpus-based linguistic analysis can support the comparative study of Indo-Aryan, Dravidian, and selected global languages. The paper explores structural, semantic, and phonological differences between these language families and evaluates the performance of AI tools in managing linguistic complexity, especially in low-resource Indian languages like Gujarati, Kannada, Tamil, and Odia. Using a mixed-method approach, the research integrates theoretical linguistic analysis with practical evaluation, including AI translation testing, corpus comparisons, and perception surveys of students and educators regarding AI-assisted multilingual learning. Findings reveal that while advanced AI models show significant accuracy with global languages, challenges persist in handling morphological richness, syntactic variation, and cultural nuance in Indian languages. The study highlights opportunities for AI-enhanced multilingual pedagogy aligned with NEP-2020, while emphasizing ethical, cultural, and pedagogical considerations. Ultimately, the research argues that AI, when thoughtfully integrated, can strengthen India's multilingual vision by enhancing translation quality, language accessibility, and cross-cultural communication.**

***Index Terms*—NEP-2020; Artificial Intelligence; Multilingualism; Comparative Linguistics; Indian Languages**

I. INTRODUCTION

Language is not merely a medium through which knowledge is transmitted; it is a living repository of culture, identity, memory, and social experience. In a country like India, where linguistic diversity is an everyday reality, education is inseparable from multilingual contexts. Each language carries its own worldview, shaping how learners think, interpret, and engage with knowledge. Recognizing this deep connection between language and learning, the National Education Policy (NEP) 2020 places strong emphasis on multilingualism, mother-tongue-based education, and respect for linguistic diversity in higher education. The policy envisions an inclusive educational framework where Indian languages coexist with global languages, ensuring that learning remains culturally rooted while being globally relevant. Within this vision, technology is positioned as a supportive force, one that enhances access and equity without undermining human values or cultural depth.

At the same time, the rapid development of Artificial Intelligence (AI) has begun to reshape the field of language studies. AI-driven tools such as machine translation systems, natural language processing models, and corpus-based analytical methods have created new possibilities for studying, comparing, and teaching languages. These technologies enable large-scale analysis of linguistic patterns and offer practical support in multilingual classrooms. However, their effectiveness is not uniform across languages. While global languages like English benefit from vast digital

resources and standardized usage, many Indian languages continue to face challenges related to grammatical complexity, rich morphology, cultural nuance, and limited digital representation. This uneven technological engagement raises important questions about linguistic equity and educational justice in the digital age. The present study addresses these concerns by examining how AI engages with linguistic diversity through a comparative analysis of Indo-Aryan, Dravidian, and global languages, combining theoretical perspectives with practical evaluation. By adopting a human-centred and ethically informed approach, the paper seeks to explore how AI can responsibly support the multilingual vision of NEP 2020 while preserving the cultural, linguistic, and moral dimensions of language learning.

II. RESEARCH QUESTIONS

This study asks how the National Education Policy (NEP) 2020 envisions multilingualism and the integration of technology in higher education, and how Artificial Intelligence based language tools function within this framework. It seeks to examine how effectively AI handles linguistic features such as grammar, meaning, and cultural context in Indian languages like Gujarati and Hindi in comparison with a global language such as English. The study also explores what corpus analysis and AI translation tests reveal about linguistic complexity across these languages, and how students and teachers perceive the role of AI in multilingual learning. Finally, it asks what pedagogical opportunities and ethical challenges arise from the use of AI in language education under NEP 2020.

III. RESEARCH OBJECTIVES

The primary objective of this study is to examine how the National Education Policy (NEP) 2020 conceptualizes multilingualism and defines the role of technology in higher education. It also seeks to analyse the effectiveness of Artificial Intelligence–based language tools when applied to Indian and global languages. Another important objective is to explore the linguistic challenges faced by AI systems in processing Indian languages, particularly in terms of grammar, morphology, meaning, and cultural context. In addition, the study aims to understand students' and teachers' perceptions of AI-assisted multilingual learning in higher education. Finally, the research seeks to identify the pedagogical opportunities as well as the ethical concerns associated with the use of AI in language education under the framework of NEP 2020.

IV. HYPOTHESES

The study proceeds with the hypothesis that Artificial Intelligence based language tools perform more effectively in global languages such as English than in Indian languages like Gujarati and Hindi due to differences in data availability, grammatical complexity, and cultural embeddedness. It is further hypothesized that while AI can support multilingual learning and translation under the framework of NEP 2020, it struggles to adequately capture linguistic nuance, morphology, and context in Indian languages. The study also assumes that students and teachers perceive AI as a useful supplementary tool rather than a replacement for human-guided language learning, and that responsible, ethically guided integration of AI can strengthen, but not replace, the multilingual and human-centred vision of NEP 2020.

V. MULTILINGUALISM AS AN EDUCATIONAL FOUNDATION IN NEP 2020

The National Education Policy (NEP) 2020 represents a decisive shift in Indian educational philosophy by positioning multilingualism as a foundational principle rather than a peripheral concern. Drawing on long-standing linguistic and cognitive research, the policy recognizes that learning through multiple languages enhances cognitive flexibility, deepens conceptual understanding, and strengthens cultural awareness. NEP 2020 views language not merely as an instrumental skill but as a medium through which knowledge systems, cultural traditions, and social identities are sustained and transmitted. By encouraging the use of Indian and regional languages alongside global languages, the policy seeks to democratize access to education and ensure that learners remain rooted in their linguistic and cultural contexts while participating in a global knowledge economy.

VI. TECHNOLOGY AS A SUPPORTIVE FORCE FOR LINGUISTIC INCLUSION

In parallel with its multilingual vision, NEP 2020 places strong emphasis on the integration of digital tools and technology-based learning in higher education. Influenced by national and international policy frameworks on education and technology, the policy presents technology as an enabling force rather than a substitute for human pedagogy. Digital platforms, language technologies, and translation tools are envisaged as mechanisms that can bridge linguistic divides, expand access to educational resources, and support learning across regions and languages. In a linguistically diverse country like India, such technological interventions are seen as essential for promoting linguistic equity and inclusivity. By combining multilingualism with thoughtful and ethically guided technological integration, NEP 2020 aims to create an education system that is flexible, inclusive, culturally sensitive, and responsive to the evolving demands of the future (UNESCO). This policy perspective reinforces the argument of the present study that technology, including AI, should function as an enabling support for multilingual education rather than as a replacement for human pedagogy. In the Indian context, where linguistic diversity is closely tied to cultural identity, such an approach ensures that technological innovation does not override linguistic equity or cultural sensitivity.

VII. ARTIFICIAL INTELLIGENCE AS A TRANSFORMATIVE TOOL IN LANGUAGE STUDIES

In recent years, Artificial Intelligence has emerged as a transformative force in the study of language, reshaping how languages are analyzed, compared, and taught in higher education. AI-based technologies such as machine translation systems, Natural Language Processing (NLP) models, and corpus-driven analytical tools have expanded the scope of linguistic research by enabling large-scale and systematic engagement with language data. Translation platforms like Google Translate, ChatGPT, and IndicTrans2 illustrate how AI can facilitate movement across languages (AI4Bharat). While such tools demonstrate the expanding technical capacity of AI in multilingual contexts, their effectiveness varies significantly across languages. The present study builds on this observation by showing that although AI performs efficiently for global languages, its performance declines when handling structurally complex and culturally embedded Indian languages. While NLP models such as BERT and MuRIL provide new ways of examining linguistic structures, semantic relations, and syntactic patterns across diverse language systems (Devlin).

Corpus-based analysis further strengthens linguistic inquiry by allowing researchers to work with extensive collections of texts and observe patterns of usage, variation, and linguistic change over time. AI systems depend heavily on the availability of large and balanced datasets, which directly influences their performance across languages. This limitation becomes especially visible in low-resource Indian languages, where insufficient digital data restricts accurate modelling and translation (Haddow et al.). The findings of the present study support this scholarly observation, as AI translation tests and survey responses consistently revealed meaning loss and reduced cultural sensitivity in Indian language outputs. This suggests that AI's limitations are not merely technical but are deeply connected to linguistic richness and cultural specificity. While global languages benefit from extensive corpora and standardized usage, many Indian languages continue to remain underrepresented in AI training datasets, limiting the accuracy and depth of AI-based analysis. More importantly, the capacity of AI to process linguistic data does not automatically result in cultural or contextual understanding. Languages are embedded within historical experience, social practice, and cultural meaning dimensions that cannot be fully captured through computational patterns alone. As a result, meaningful interpretation of language, particularly in multilingual and culturally diverse contexts, continues to depend on human judgment and critical awareness. This necessary balance between technological capability and human insight lies at the core of contemporary language research and is central to the ethical and responsible integration of AI in multilingual education.

VIII. FOCUS ON INDO-ARYAN, DRAVIDIAN, AND GLOBAL LANGUAGES

Indian Language Families and Linguistic Diversity

This study adopts a comparative linguistic perspective that moves beyond the analysis of a single language or a narrowly defined linguistic context. By bringing Indian and global languages into dialogue, the paper seeks to understand how different language systems interact with AI-based tools in multilingual educational settings. Central to this comparison are the two major Indian language families Indo-Aryan and Dravidian which together represent the core of India's linguistic diversity. Indo-Aryan languages such as Hindi, Gujarati, Marathi, and Bengali constitute a significant part of the Indian linguistic landscape and are historically rooted in Sanskritic traditions. These languages are marked by inflectional morphology, flexible word order, and rich semantic expression. (Cardona)

Dravidian languages, including Tamil, Kannada, Malayalam, and Telugu, represent a distinct linguistic tradition with their own structural and cultural characteristics. These languages are typically agglutinative in nature and exhibit complex syntactic organization and strong continuity with regional cultures and literary traditions. Examining these two language families together allows the study to highlight the structural richness and cultural depth that characterize Indian multilingualism (Steever).

IX. GLOBAL LANGUAGES AND DIGITAL REPRESENTATION

Alongside Indian languages, the study includes global languages particularly English to establish a broader comparative framework. Global languages have benefited from long processes of standardization and extensive digital documentation, which make them more readily accessible to AI-based language technologies. As a result, AI tools tend to perform more efficiently when processing these languages. (Comrie)

In contrast, many Indian languages, despite having large speaker populations and rich grammatical systems, remain underrepresented in digital corpora and AI training datasets. By placing Indo-Aryan, Dravidian, and global languages side by side, the study examines how AI tools respond to languages shaped by different historical trajectories, grammatical structures, and levels of digital availability. This comparative approach not only reveals shared linguistic patterns but also exposes significant disparities in technological engagement, thereby enhancing the analytical depth and originality of the research.

X. AI AND LINGUISTIC DIFFERENCES ACROSS LANGUAGES

Structural and Cultural Variation in Languages

Languages differ profoundly in their internal structures and expressive possibilities. Variations in syntax, phonology, morphology, and semantics shape how each language organizes thought and conveys meaning. These linguistic characteristics are closely tied to cultural practices and historical contexts, making each language a unique system of expression. When such diversity is introduced into AI-based language tools, these differences become particularly visible. Artificial Intelligence systems are primarily trained on statistical patterns derived from large digital datasets, and therefore their performance often reflects the availability and uniformity of such data rather than the full complexity of language.

Languages with relatively stable sentence structures, standardized grammatical usage, and extensive digital representation tend to be processed more efficiently by AI systems. In contrast, languages characterized by flexible syntax, rich inflectional or agglutinative morphology, and culturally embedded meanings present greater challenges (Yule). This theoretical insight is particularly relevant to the present study, as it explains why AI systems struggle with Indian languages that exhibit flexible syntax and rich morphology. The analysis confirms that linguistic structure plays a crucial role in determining AI performance, reinforcing the need for human-guided interpretation in multilingual education. These structural and cultural variations reveal the limits of AI's ability to treat all languages equally.

AI Performance across Indian and Global Languages

This study examines how AI responds to linguistic diversity by comparing its performance across selected Indian and global languages. Global languages such as English are generally handled more effectively by AI tools due to the availability of large,

well-curated datasets and predictable grammatical patterns shaped by long histories of standardization. As a result, AI-based translation and analysis tend to be more accurate and fluent in such languages.

In contrast, Indian languages such as Gujarati and Tamil often expose limitations in AI performance, particularly in areas involving complex word formation, contextual interpretation, and culturally specific expressions. These challenges suggest that linguistic difficulty for AI is not merely a technical limitation but is deeply connected to the nature of language itself. By analyzing these disparities, the paper moves beyond a purely technological evaluation of AI and foregrounds the linguistic and cultural factors that influence how languages are represented, processed, and sometimes misinterpreted in AI-driven systems.

XI. METHODOLOGY

Theoretical Framework

The present study adopts a mixed-method research approach in order to examine the role of Artificial Intelligence in multilingual education within the framework of the National Education Policy (NEP) 2020. The theoretical component of the research is grounded in comparative linguistics and policy analysis. It draws upon established linguistic scholarship on Indo-Aryan and Dravidian languages to examine differences in syntax, morphology, semantics, and phonology, while also engaging with global languages such as English for comparative reference. Alongside linguistic theory, the study undertakes a critical reading of NEP 2020 to understand its emphasis on multilingualism, linguistic equity, and the integration of digital technologies in higher education. This theoretical foundation provides the conceptual lens through which AI's role in language learning and translation is evaluated.

Data Collection Methods

The practical component of the study consists of three interconnected methods: AI translation testing, corpus comparison, and a perception-based survey. These methods were selected to examine both the technical performance of AI tools and the human responses to their use in multilingual educational contexts, thereby aligning empirical data with the research objectives and hypotheses of the study. AI translation tests were conducted using commonly available tools such as Google Translate, ChatGPT, and IndicTrans2 to examine how effectively they translate between English, Gujarati, and Hindi. Particular attention was given to grammatical accuracy, preservation of meaning, idiomatic expressions, and cultural nuance. In addition, brief corpus-based observations were used to compare structural patterns across languages, highlighting differences in word formation, sentence structure, and lexical variation.

To complement these analytical methods, a structured survey was conducted to capture human perspectives on AI-assisted multilingual learning. The survey included 40 multiple-choice questions and was administered to 41 respondents, comprising students, teachers, and research scholars from higher education institutions. The questionnaire was designed to assess language background, awareness of NEP 2020, frequency and purpose of AI tool usage, perceived accuracy of AI translations in English, Hindi, and Gujarati, cultural sensitivity of AI outputs, and ethical concerns related to AI in education. The structural design of the survey ensured clarity and consistency, enabling respondents from diverse linguistic backgrounds to participate meaningfully.

The survey results, presented through graphical representations, indicate strong awareness of NEP 2020, widespread use of AI tools for academic purposes, higher confidence in AI performance for English, and noticeable concern regarding meaning loss and cultural accuracy in Gujarati translations. These findings directly support the study's objectives by demonstrating both the uneven performance of AI across languages and the perception that AI should function as a supportive rather than substitutive educational tool. The complete survey instrument and response charts are included as Appendix A for reference. By integrating AI-based testing with perception-driven data, the mixed-method approach strengthens the reliability, analytical depth, and relevance of the study, in alignment with the multilingual and human-centered vision of NEP 2020.

XII. KEY FINDINGS

The findings of the study reveal a clear difference in the performance of Artificial Intelligence tools when applied to global languages as compared to Indian languages. AI-based language systems demonstrate a

higher level of accuracy, fluency, and consistency when working with global languages such as English and other widely used international languages. This effectiveness can largely be attributed to the availability of extensive digital datasets, long-standing standardization, and sustained technological development associated with these languages. As a result, AI tools are better equipped to process grammatical structures, lexical patterns, and semantic relationships in global languages.

In contrast, the study finds that AI systems face significant challenges when applied to Indian languages such as Gujarati, Tamil, and Odia. These languages are relatively low-resource in the digital domain, meaning that large and balanced datasets required for effective AI training are limited. Moreover, Indian languages often exhibit complex grammatical structures, rich morphological systems with multiple word forms, and flexible syntactic patterns. Cultural meanings, idiomatic expressions, and context-specific usage further complicate AI processing, as such elements cannot be fully captured through statistical patterns alone. These limitations are reflected in AI translation tests, corpus observations, and survey responses, all of which indicate frequent meaning loss, grammatical inconsistencies, and reduced cultural sensitivity in AI outputs for Indian languages. Collectively, these findings highlight that AI's struggles with Indian languages are not merely technical shortcomings but are deeply connected to linguistic richness and cultural depth.

XIII. PEDAGOGICAL OPPORTUNITIES

The findings of this study suggest that Artificial Intelligence holds significant pedagogical potential for strengthening multilingual education in higher education, particularly within the framework of the National Education Policy (NEP) 2020. When used thoughtfully, AI-based tools can support multilingual classrooms by facilitating communication across languages and enabling students to engage with learning materials in languages they are more comfortable with. This is especially valuable in linguistically diverse classrooms, where students come from different language backgrounds and may experience barriers to learning through a single dominant language.

AI-assisted translation tools can play an important role in bridging linguistic gaps by helping students access academic content across languages. Such tools can assist learners in understanding complex concepts, translating reference materials, and engaging with texts written in global languages like English. In this way, AI contributes to making learning more accessible, particularly for students who are more proficient in Indian or regional languages. This aligns closely with NEP 2020's emphasis on inclusion, equity, and mother-tongue-based learning (UNESCO). Moreover, AI has the potential to connect Indian languages with global knowledge systems. By enabling translation and cross-linguistic exchange, AI tools can help Indian language users participate more actively in global academic and intellectual discourse. At the same time, AI can support the preservation and continued use of Indian languages in higher education by integrating them into digital and technological platforms. However, the study also emphasizes that these pedagogical benefits can be realized only when AI is used as a supportive tool under human guidance. Teachers remain central to interpreting meaning, ensuring cultural sensitivity, and maintaining ethical standards. Thus, AI's pedagogical value lies not in replacing traditional education but in complementing it, in harmony with the inclusive and human-centered vision of NEP 2020.

XIV. CONCLUSION

The present study has examined the relationship between multilingualism, Artificial Intelligence, and higher education within the framework of the National Education Policy (NEP) 2020. By combining theoretical insights from comparative linguistics with practical evaluation through AI translation tests, corpus observations, and a perception-based survey, the research highlights both the possibilities and limitations of AI in multilingual education. The findings demonstrate that while AI performs effectively in global languages due to extensive digital resources and standardization, it continues to face significant challenges with Indian languages because of their grammatical complexity, rich morphology, cultural embeddedness, and limited digital representation.

At the same time, the study argues that Artificial Intelligence can play a constructive and meaningful

role in advancing multilingualism in India when its use is guided by the inclusive and human-centred principles outlined in NEP 2020. AI-based language tools have the potential to support translation, widen access to educational resources, and enable interaction across linguistic and cultural boundaries, thereby helping Indian languages connect more effectively with global knowledge systems. However, the value of AI in language education lies in its supportive rather than autonomous function. Language is deeply embedded in cultural experience, historical memory, and social context dimensions that extend beyond computational patterns. While AI can assist in processing and analysing language data, it cannot replace the interpretive judgment, ethical awareness, and cultural sensitivity that human educators bring to the learning process. The study therefore concludes that the successful realization of NEP 2020's multilingual vision depends on a balanced integration of AI and human insight, where technology enhances rather than diminishes the linguistic and cultural richness of Indian education.

ACKNOWLEDGEMENT

The authors acknowledge the use of Artificial Intelligence–based tools, including ChatGPT (OpenAI), Google Gemini, and AI-assisted translation tools such as Google Translate and IndicTrans, during the preparatory stages of this research. These tools were used solely for supportive purposes such as idea generation, preliminary drafting, language refinement, translation testing, and structural organization of the paper. All interpretations, analyses, arguments, and conclusions presented in this study are entirely the original work of the authors. The use of AI did not replace independent critical thinking, scholarly judgment, or analytical responsibility, which remained central throughout the research process.

REFERENCES

[1] AI4Bharat, et al. "IndicTrans2: Towards High-Quality and Accessible Machine Translation Models for All 22 Scheduled Indian Languages." arXiv (Cornell University), May 2023, doi:10.48550/arxiv.2305.16307.

[2] Alaparthi, Shivaji, and Manit Mishra. "Bidirectional Encoder Representations From Transformers (BERT): A Sentiment Analysis Odyssey." arXiv (Cornell University), July 2020, doi:10.48550/arxiv.2007.01127.

[3] Cardona, George, and Danesh Jain. The Indo-Aryan Languages. Routledge, 2007.

[4] Comrie, Bernard. The World's Major Languages. 2009.

[5] Deumert, Ana. "Multilingualism." Cambridge University Press eBooks, 2011, pp. 261–82, doi:10.1017/cbo9780511997068.021.

[6] Girish Nath Jha. 2010. The TDIL Program and the Indian Langauge Corpora Intitiative (ILCI). In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), Valletta, Malta. European Language Resources Association (ELRA).

[7] Haddow, Barry, et al. "Survey of Low-Resource Machine Translation." Computational Linguistics, vol. 48, no. 3, Jan. 2022, pp. 673–732, doi:10.1162/coli_a_00446.

[8] Home | Official Website of Linguistic Data Consortium for Indian Languages. www.ldcil.org.

[9] Khanuja, Simran, et al. "MuRIL: Multilingual Representations for Indian Languages." arXiv (Cornell University), Mar. 2021, doi:10.48550/arxiv.2103.10730.

[10] MacSwan, Jeff. "A Multilingual Perspective on Translanguaging." American Educational Research Journal, vol. 54, no. 1, Feb. 2017, pp. 167–201, doi:10.3102/0002831216683935.

[11] Nic, Laxminarayan Prajapati. About National Education Policy | Government of India, Ministry of Education. www.education.gov.in/en/nep/about-nep.

[12] Project: State of the Education Report India 2025: Mother Tongue Based and Multilingual ED Details | Core Data Portal. 30 Sept. 2025, core.unesco.org/en/project/201IND1001.

[13] Roy, Anna, et al. "Responsible AI for All: Adopting the Framework – a Use Case Approach on Facial Recognition Technology." Discussion Paper, Nov. 2022, www.niti.gov.in/sites/default/files/2022-11/Ai_for_All_2022_02112022_0.pdf?trk=public_post_comment-text.

[14] Shapiro, Michael C., et al. Language and Society in South Asia. Final Report. report, BR-110012HH, Institute of International Studies

(DHEW/OE), Washington, D.C., 1975, files.eric.ed.gov/fulltext/ED127806.pdf.
[15] Steever, Sanford B. The Dravidian Languages. Taylor and Francis, 1998.
[16] Yule, George. The Study of Language. Cambridge UP, 2020.

APPENDIX A: SURVEY INSTRUMENT AND DATA OVERVIEW

[1] The survey consisted of 40 structured multiple-choice questions administered to 41 respondents, including postgraduate students, teachers, and research scholars. The questionnaire was designed to examine awareness of NEP 2020, frequency of AI tool usage, perceived accuracy of AI translations in English, Hindi, and Gujarati, cultural sensitivity, and ethical concerns related to AI use in education.
[2] The data were analysed using descriptive methods to identify patterns and trends. The findings indicate higher confidence in AI tools for global languages such as English, while concerns were expressed regarding meaning loss and cultural accuracy in Indian languages. These responses support the research objectives and hypotheses of the study.