Pak. j. life soc. Sci. (2025), 23(2): 1264-1275 E-ISSN: 2221-7630;P-ISSN: 1727-4915

Pakistan Journal of Life and Social Sciences
www.pjlss.edu.pk

https://doi.org/10.57239/PJLSS-2025-23.2.00254

RESEARCH ARTICLE

Language Policy in the Age of Artificial Intelligence: Implications for Urdu and Punjabi Language in Pakistan

Dr. Fayyaz Hussain[1], Dr. Samia Komal[2], Dr. Muhammad Irfan ul Haq[3], Dr. Naseem Ur Rehman[4], Nazia Anwar[5], Dr. Tariq Mahmood Hashmi[6*], Muhammad Asim Khan[7]

[1]Assistant Professor, Department of Punjabi, Government College University Faisalabad, Pakistan
[2]Assistant Professor, Department of Urdu, Emerson University Multan, Pakistan
[3]Assistant Professor, Institute of Punjabi and Cultural Studies, University of the Punjab, Lahore, Pakistan
[4]Assistant Professor, Department of Persian, University of the Punjab, Lahore, Pakistan
[5]Ph.D Scholar, Institute of Punjabi & Cultural Studies, University of the Punjab, Lahore, Pakistan
[6]Associate Professor, Department of Urdu, Government College University Faisalabad, Pakistan
[7]Visiting Lecturer, Department of Applied Linguistics, Government College University Faisalabad, Pakistan

ARTICLE INFO

Received: Oct 21, 2025

Accepted: Dec 25, 2025

Keywords

Language Policy
Artificial Intelligence
Critical Discourse Analysis
Urdu
Punjabi
Linguistic Marginalization
Digital Inequality

***Corresponding Author:**

drtariqhashmi@gcuf.edu.pk

ABSTRACT

This study has discussed the language policy in the era of artificial intelligence and how it would impact Urdu and Punjabi in Pakistan. The aim of the research was to learn how the artificial intelligence is influencing the use of languages, their visibility, and power and how current ideologies in language policy are duplicated by AI-based systems. It was a qualitative research design, which was informed by a comprehensive theoretical framework based on a combination of Language Policy and Planning, Critical Discourse Analysis (CDA), and perceptions of AI as a non-state language policy actor. The sources of data were language policy documents (1947-2025), AI-mediated language contents (translation tools, speech recognition systems, and digital learning platforms), and scholarly and policy-oriented academic writings about AI and linguistic marginalization. The thematic analysis of the data was performed within a CDA paradigm based on the patterns of representation, inclusion, exclusion, and the ideological framing of Urdu and Punjabi. The results indicate that past language dominance in favor of Urdu and against Punjabi is present in the digital and AI realm. Although Urdu has acquired more or less visibility and is functionally represented in AI applications, Punjabi is still a relatively underrepresented language, which is explained by policy silence, institutional support, and a lack of linguistic resources. In the research, it was also discovered that AI is a novel language policy mechanism that imposes ideological decisions into algorithms and infrastructures of the digital environment. The article makes a contribution to the field of language policy and AI studies by showing that the issue of linguistic inequality has also moved to the algorithmic-based systems in Pakistan, and that the solution to this problem lies in the involvement of inclusive language planning to guarantee linguistic justice in AI-based contexts.

1. INTRODUCTION

In Pakistan, language policy has always been interconnected with the problem of power, identity and nation-building. Since gaining independence in 1947, the state has been encouraging the use of Urdu as a national language in the country in order to foster unity in a multi-linguistic society. Yet, some long-term inequalities have also occurred because of this policy option, particularly regional languages, including Punjabi, which is the language of a majority of the population but is not represented in official spheres of life, such as education, administration, and a press (Khan et al., 2020). Consequently, language policy in Pakistan has established some degree of domination whereby Urdu is the symbol of power in the country, with Punjabi relegated.

This marginalization in the past is not a mere coincidence as evidenced by earlier research that claims it is ideologically motivated. The research is based on Critical Discourse Analysis (CDA), which

shows that the policy documents, political speeches, and educational policies are always focused on Urdu as a national identity symbol, and Punjabi is either excluded or represented as informal and cultural (Hussain et al., 2024; Hashmi et al., 2025). These forms of discourse influence the attitude of the people and limit the functional aspect of the Punjabi though it is socially dominating in daily communication.

Over the last several years, the intensive development of an artificial intelligence (AI) has created a new layer to the language policy. Artificial intelligence-based technologies like machine translation, speech recognition, chatbot, and online learning platforms are taking on more and more socialization, education, and government roles. The technologies are dependent on a lot of linguistic data, institutional support and investment in research. Consequently, they will favor those languages that already have political and symbolic influence (Cristaldi, 2025). The creation also creates significant questions regarding whether AI has the potential to disrupt the status quo of language inequalities or the fact that it creates them in the digital realm.

The studies conducted on AI and language technology in Pakistan indicate dissimilar results in the case of Urdu and Punjabi. The research on Urdu speech recognition and translation suggests that it is slowly improving, despite the fact that Urdu is a low-resource language on the international scale (Sharif et al., 2024; Safder et al., 2024). Punjabi on the contrary is direly underrepresented in AI. Even though AI has a great potential in Punjabi language revitalization, the absence of annotated datasets, dialect variation, script difference, and poor state backing bar its participation in the AI systems (Butt et al., 2025). These results indicate that not only technical issues can be used to explain digital exclusion, but several more important factors are policy silence and institutional inattention.

The academic writing also points to the fact that AI cannot be considered an objective instrument. Rather, AI systems incorporate social and ideological presumptions by means of data selection, model design and prioritization of language. Research claims that AI is becoming a new, non-state language politics actor, which has an influence on which languages become visible and legitimate in the digital realm (Butt et al., 2025). This is equivalent to extending historical language hierarchies into algorithmic systems in the Pakistani context, which is strengthening the control of Urdu and marginalizes Punjabi.

Although the literature on language policy and AI is increasing, very few studies have analyzed the interaction of the two in Pakistan especially in CDA perspective. The majority of the current research is divided into either the adherence to the classic language policy or the technical facets of AI, but the connection of both to the wider power relations is not made. This study fills this gap by exploring the language policy during the era of artificial intelligence and its consequences to the Urdu and Punjabi. The research will examine language policy texts, AI-mediated language messages, and scholarly texts to demonstrate the way AI replicates or transforms linguistic inequalities and the way digital technologies are turning into the center of language management in Pakistan.

1.2 Problem Statement

Although Pakistan is a multilingual country, this policy has always favored Urdu over the regional languages since independence, which include Punjabi, resulting in linguistic inequality in the country. As artificial intelligence grows quickly in learning, communication, and government, these disparities are now being encompassed much more within domains of classic policy than in digital and AI-driven systems. The current literature indicates that although Urdu is slowly being represented in AI programs, Punjabi is still under the radar because of the lack of institutional support, poor policy intervention, and linguistic resources. Majority of research is on technical issues of AI or previous language policy, whereas there is little critical research on how language policy ideologies are replicated by the use of AI technologies. Subsequently, the role of AI as a new non-state language policy actor that affects the linguistic access, visibility and power has been under-researched in the Pakistani context. This study fills this gap by offering a critical analysis of the topic of language policy in the era of artificial intelligence and the way it applies to Urdu and Punjabi in Pakistan.

1.3 Significance of the Study

The present research is important as it expands the existence of the language policy to other state documents by revealing that artificial intelligence is now one of the forces controlling the linguistic hierarchies in Pakistan. Through the critical analysis of the language policy-AI-power intersection, the study brings to the fore the reproduction of domination of Urdu and the further marginalization of Punjabis by digital technologies. The results add to the language policy and Critical Discourse Analysis research as they present AI as a type of non-state language control mechanism. In practice, the research suggests findings to policymakers, educators, and those involved in developing AI about the necessity of the inclusive and ethically conscious language planning in the digital realms. It also provides a basis of the future investigation on linguistic justice and regional language integration in AI-mediated education, communication, and governance in Pakistan.

1.4 Research Questions

1. How do existing language policies in Pakistan shape the representation and use of Urdu and Punjabi in artificial intelligence–based language technologies?
2. In what ways do AI-driven systems reproduce or reinforce linguistic dominance of Urdu and marginalization of Punjabi in the digital space?
3. How is artificial intelligence emerging as a new language policy actor influencing linguistic equality and access for Urdu and Punjabi speakers in Pakistan?

2.Theoretical Framework

The theoretical basis of this study is the combination of Language Policy and Planning (LPP), Critical Discourse Analysis (CDA), and theoretical approaches to language and artificial intelligence. This framework can be used to understand the nature of interaction between language, power, and technology, and to understand the way Urdu and Punjabi are positioned differently in Pakistan in the era of AI.

2.1 Language Policy and Planning (LPP)

This study is largely based on the study of Language Policy and Planning since it gives an understanding of how languages are handled, encouraged or even ignored using state policies and institutional practices. The Pakistani language policy has a long history of focusing on Urdu as the national language, whereas regional languages such as Punjabi were not included in official arenas of life, including education and governance (Khan et al., 2020). Studies indicate that these policies are partisan rather than neutral as they are influenced by their political interest, historical backgrounds and political ideologies. The research on Punjabi marginalization claims that the language policies in Pakistan create effective hierarchies with Urdu, associated with nationalism and morality, the English, associated with power and mobility, and the Punjabi, associated with informal and rural life (Khan et al., 2025; Hashmi et al., 2025). LPP is used in this study to comprehend how these hierarchies are currently finding their way to digital and AI-driven spaces.

2.2 Critical Discourse Analysis (CDA)

The basic tool of analysis in this study is a critical Discourse Analysis. The socio-cultural aspect that is of interest to CDA is the way language is applied to create and sustain power, dominance, and inequality in our society. It presupposes that policy texts, media language, and technological discourses are ideologically constructed and are in the interest of a particular side. Past CDA-based research in Pakistan indicates that official documents and political discourse continue to revolve around the Urdu language and suppresses the Punjabi by using silence, exclusion, and limited representation (Hussain et al., 2024; Arshad, n.d.). On the same note, Hashmi et al. (2025) elaborate that policy and educational language usage supports social classes and cultural superiority. The *study* uses CDA to understand how analogous discursive strategies can be used in the AI-related discourse where some languages become visible and acceptable, and others stay unnoticed.

2.3 Artificial Intelligence and Digital Language Inequality

This study considers artificial intelligence as a new and non-state language policy agent. The studies of AI and language indicate that the development of AI systems completely depends on the availability of data, funding, and institutional support, all of which prefer languages that are already dominant (Cristaldi, 2025). Due to that, AI tends to recreate the linguistic disparities that already exist rather than minimizing them. In the study, Butt et al. (2025) prove that despite the potential of

AI to help revitalize the Punjabi language, the absence of datasets, dialect diversity, script variation, and poor support of policies reduce its usefulness. Technical literature on Urdu and Punjabi also includes the fact that the two languages are also experiencing difficulty as low-resource languages, particularly with speech recognition and machine translating systems (Sharif et al., 2024; Srivastava et al., 2024). This model enables the research to connect technological constraints to more ideological and policy concerns.

This theoretical framework gives the solid foundation regarding the analysis of the language policy in the context of state documents as well as conceptualizing AI as the potent power that influences the language future of Pakistan.

LITERATURE REVIEW

The issue of language policy in Pakistan has been closely associated with issues of power, identity, and nation-building. The studies indicate that the Pakistani language planning has always been biased towards the promotion of the Urdu language as the national identity, with regional languages, such as Punjabi, being pushed out of such formal spheres as education, administration, and media (Khan et al., 2020). Historical studies point out that language policies were influenced more by political and ideological interests rather than by the realities of languages, which produced inequality in the long-term in a multilingual society.

A number of Critical Discourse Analysis (CDA) studies have revealed how texts associated with policies and political rhetoric form unequal language hierarchies. According to Hussain et al. (2024), Urdu is constantly represented as a moral and national language, whereas Punjabi is commonly associated with informal, rural, and cultural territories. The same results are also demonstrated by Hashmi et al. (2025), who state that selective inclusion and erasure are deployed in policy documents and political speeches in an attempt to preserve dominance of Urdu and English and depriving Punjabi of institutional assistance. Such practices of discourse naturalize the inequality of language and form the view of people on the value of language.

Socio-political and cultural viewpoints have also been looked upon in terms of the marginalization of Punjabi. According to Arshad, in the history, language policies in the post-independent period did not take into account Punjabi, even though it was the largest spoken language in Pakistan. The analysis based on the CDA of documents and interviews with officials shows that Punjabi speakers are also characterized by a poor educational background and a poor representation in the media, and it influences their linguistic activities. According to Khan et al. (2025), education policies like National Education Policy, provincial curriculum frameworks and others do not consider Punjabi, which further confirm its low position and endanger its further sustainability.

As artificial intelligence emerges, the language policy discussion is now going online. Butt et al. (2025) emphasize that AI technologies can contribute to the Punjab language revival, but this chance is not yet well-utilized because of the absence of annotated data, the presence of multiple dialects, and variations in scripts, as well as insufficiently developed assistance on the state level. In their research, they find that AI systems tend to replicate the current neglect of politics since Punjabi is underrepresented and underinvested in AI studies and uses. This is enough to imply that AI is not a neutral entity but it works within the pre-existing language ideologies.

The investigation of the AI and language technologies also indicates the structural issues concerning the low-resource languages of Urdu and Punjabi. Sharif et al. (2024) demonstrate that the Urdu speech recognition systems have low datasets and technological gaps, in contrast to the global languages. On the same note, Srivastava et al. (2024) justify that Punjabi machine translation is troubled by the morphological complexity and the absence of parallel corpora. Such technical constraints have a direct impact on the inclusion or exclusion of languages into the AI systems.

With a more general digital approach, Cristaldi (2025) claims that AI development language supports a few dominant languages worldwide, which further contributes to the digital divide between strong and marginalized languages. This point can apply to the Pakistani case where Urdu has become more digitally salient and Punjabi is digitally invisible. Another study by Hussain et al. (2023) also demonstrates that institutional discourse excludes Punjabi, yet digital platforms such as social media have turned into a domain of opposition, where speakers establish their demands in a fight and gain rights to language.

The available literature allows outlining that the language policy in Pakistan has long been discriminatory of Punjabi, and that the inequalities are currently being transferred to the AI-based digital realm. This is because whereas a number of studies concentrate on classic policy and discourses there are questions of AI and language technology in isolation, but there is a paucity of combined studies that conjoin language policy, AI, and power relations to both Urdu and Punjabi. This study fills this gap by discussing language policy in the era of artificial intelligence and how it impacts Urdu and Punjabi in Pakistan with special attention to how digital tools reproduce or replicate existing linguistic orders.

3.RESEARCH METHODOLOGY

The research design adopted in this study is a qualitative study to investigate the language policy in the era of artificial intelligence and the role of the same in the case of Urdu and Punjabi in Pakistan. A qualitative study design is appropriate since the research is centered on the topic of language, power, ideology, and representation that is best examined with the help of textual and discourse-based analysis, as opposed to quantification. A Critical Discourse Analysis (CDA) is used as a guiding approach to the study. The application of CDA to comprehend the construction, reinforcement, or challenge of linguistic hierarchies between Urdu and Punjabi involves the application of language policies and AI-related discourses. This method is suitable since prior research on the language policy and linguistic marginalization in Pakistan has effectively employed CDA to reveal the concealed ideology and power politics in the policy texts and online discourse (Khan et al., 2020; Hussain et al., 2024; Hashmi et al., 2025).

3.1 Data Sources

There are three primary sources of data used to gather the data of this study. The positioning of Urdu and Punjabi in the policies is made by analyzing official language policy texts, education policy documents, and other governmental materials. These documents assist in discovering continuities and shifts in language ideology in the multilingual setting in Pakistan. The language content of AI generated or AI assisted in Urdu and Punjabi is sampled. It contains the output of the language technologies like translation systems, speech recognizers and online learning system, which are already covered in the literature. It is concerned with the visibility, precision, and usability of the two languages in the field of AI. Contextual data is provided by selected scholarly materials on AI, language technology and linguistic marginalization because it is required to comprehend how AI is being positioned as a linguistic and policy actor, particularly with regard to low-resource languages such as Punjabi and Urdu (Butt et al., 2025; Sharif et al., 2024).

3.2 Sampling Technique

The research employs purposive selection because the sources of data are chosen according to the criteria of relevance to the language policy, AI, and Urdu and Punjabi relations. Any texts and materials not directly dealing with language planning, digital technologies, and linguistic representation in Pakistan will not be included.

3.3 Data Analysis Procedure

The thematic analysis is conducted to analyze the data in a framework of CDA. The steps of the process include the following:

- Careful reading of policy texts, and texts with AI in order to determine recurring words, phrases, and patterns.
- Determination of the main themes that include language dominance, marginalization, digital inclusion, technological bias, and linguistic visibility.
- Comparison of discursive practices such as inclusion, exclusion, framing and normalization applied to Urdu and Punjabi.
- Explanation of results as far as power relations between national and regional language are concerned.

This approach will enable the study to connect the existing discourse of language policy to the new language activities that are driven by AI.

3.4 Ethical Considerations

All the data employed in this study are publicly available documents and texts. There are no direct participants who are human beings. Hence, no consent risks or confidentiality risks are involved. The academic integrity is maintained by ensuring that the content is cited and representation of original sources is done faithfully.

3.5 Scope and Limitations

The research is confined to the Urdu and Punjabi in the Pakistani setting and is subjective to qualitative understanding and not technical analysis of AI systems. Although the results offer profound understanding of language policy and the discourse of AI, these results are not meant to be used to generalize about all regional languages or worldwide AI systems. Overall, the proposed methodology can offer a systematic and critical analysis of the intersection of language policy and artificial intelligence and the way this intersection will impact the future of Urdu and Punjabi within the realm of digital and politics in Pakistan.

4. Data Analysis

This part gives the explanation of the analysis of the data gathered to discuss the topic of language policy in the era of artificial intelligence and its effects on Urdu and Punjabi in Pakistan. Data analysis is directed by a Critical Discourse Analysis (CDA) approach, which will enable the investigation of how language, power, and ideology work in both traditional texts of policies and new digital and AI-mediated ones. This section aims to establish the trends of representation, inclusion, and exclusion of the Urdu and Punjabi, and demonstrate the way in which the language ideologies of the past are propagated with the help of artificial intelligence.

The analysis will be arranged into three primary sources of information. As the first step, the language policy documents (19472025) are analyzed to follow the state and institutional level positioning of Urdu and Punjabi. Second, a language content mediated by AI is studied based on the results of translation systems, speech recognition systems, and online learning platforms to evaluate the language visibility, accuracy, and functional scope in AI settings. Third, the fieldwork of academic and policy-driven discourse is examined to know how scholars and policymakers represent AI as a linguistic and policy-making agent, especially when it comes to low-resource language. A combination of these sources offers a complete picture of the intersection of lingual politics and artificial intelligence to create linguistic stratification of modern-day Pakistan.

Table 1. Language policy documents (1947–2025)

Policy Period	Key Policy Documents	Position of Urdu	Position of Punjabi	Dominant Language Ideology	Discursive Pattern (CDA)
1947–1956	Early state declarations, Constitutions (1947–1956)	Declared as symbol of national unity and identity	Completely absent	Nation-building through one language	Erasure of regional languages
1957–1971	Education policies, political speeches	Strengthened as sole national language	Ignored despite majority speakers	Linguistic nationalism	Centralization and exclusion
1972–1988	1973 Constitution, education reforms	Reaffirmed as national language	Mentioned only as regional language	Unity over diversity	Symbolic recognition without power
1989–2008	National education policies	Medium of instruction emphasized	Excluded from formal education	Modernization through Urdu and English	Marginalization through silence
2009–2017	National Education Policy (2009), Punjab Curriculum Policy (2015)	Institutional dominance in schooling	Excluded from curriculum	Utility and prestige ideology	Policy omission and devaluation
2018–2025	Digital education initiatives, technology-related policy texts	Preferred in digital governance	No clear digital policy support	Technological progress with dominant languages	Digital invisibility of Punjabi

A comparative study of 1947 to 2025 language policy documents reveals that there was a steady trend of giving greater privileges to Urdu at the expense of being progressive and systematic in diminishing Punjabi. Since the very beginning of policy documents, the Urdu is oriented as the national symbol, and Punjabi is neglected or diminished to local or cultural identities. The continuity

of the language ideology is high as this trend has not been affected by the various political eras. The policy documents also resort to discursive acts of silence and exclusion in an attempt to marginalize Punjabi repeatedly. In spite of the recognition of regional languages, they do not receive institutional positions in education, administration and digital governance. The education policies also make Urdu superior particularly through the provision of it as the main medium of instruction whereas Punjabi, the most spoken language, has not been included in the curricula. This disparity is carried into the digital and technological sphere in the near future (20182025). As Urdu gains presence in the digital governance and artificial intelligence-related projects, Punjabi is still not present in the policy debate regarding technology and artificial intelligence. It means that AI and digitalization are strengthening the existing language hierarchies, instead of questioning them. In general, it can be concluded that the language policy in Pakistan is characterized by ideological persistence, and Urdu still has a symbolic and practical authority, and Punjabi is still a peripheral language. This lends credence to the belief that even in the world of artificial intelligence, language policy still recreates linguistic inequality rather than advanced multilingual inclusion.

Table 2. AI-Mediated language content (Urdu vs. Punjabi)

AI Domain	Type of AI Tool	Representation of Urdu	Representation of Punjabi	Key Issues Identified	CDA Interpretation
Machine Translation	Urdu–English / English–Urdu tools	Widely available and functional	Very limited or inconsistent	Punjabi lacks parallel corpora; low accuracy	Linguistic privilege of Urdu
Machine Translation	English–Punjabi tools	Indirect support via Urdu	Poor quality and unstable outputs	Dialect variation; script issues	Structural marginalization
Speech Recognition	ASR systems	Partial but improving accuracy	Mostly absent or unreliable	Lack of annotated speech data	Digital exclusion
Educational AI Platforms	AI-based learning apps	Used as instructional language	Rarely supported	Policy-driven language preference	Institutional neglect
AI Chatbots / Assistants	Text-based interaction	Recognized and processed	Often unsupported or ignored	Low training data for Punjabi	Algorithmic invisibility
NLP Resources	Datasets and corpora	Moderate availability	Severe shortage	No state-backed dataset creation	Policy failure in AI planning

The comparison of the AI-mediated language content indicates the undeniable disproportion of the Urdu and Punjabi in the field of artificial intelligence. Urdu has a greater presence in AI methods, namely translation software, speech recognition, and e-learning. Even though Urdu is considered a low resource language in the world even now, it is getting relatively higher institutional and technological support as compared to Punjabi. Punjabi is also not much visible in AI systems. Support of Punjabis is poor, patchy and imprecise where it is present. The lack of annotated datasets, variation of dialect, and variability of scripts, and the unavailability of support at a policy level are the key causes. These are technical issues that are not neutral, but they are also indicative of long-term political and institutional disregard of Punjabi. Critically speaking, AI systems are replicating the current ideologies of language. Urdu is regarded as an official and educational language and can be used in governance and technology whereas Punjabi is excluded, except in the informal and cultural context. This is similar to the previous language policies in which Punjabi was not included in the formal field.

The results are also indicative of AI being a new language policy actor. Although there are no particular policy statements, AI tools choose what languages are usable, visible, and valuable. Here, the Punjabi speakers are digitally marginalized, and the Urdu speakers will have access to the AI-based communication and learning. Altogether, this data demonstrates that AI technologies are spreading the traditional language disparities into the digital realm. In the absence of an inclusive language planning and specific AI policy, the Punjabi will be left on the sidelines of AIs in future.

Table 3. Academic and Policy-Oriented discourse on AI and language

Source	Focus of the Study	Framing of AI	Position of Urdu	Position of Punjabi	Key Discursive Theme

Butt et al. (2025)	AI and Punjabi language revitalization	AI as tool with policy dependency	Indirectly prioritized	Target of revitalization but neglected	Policy-driven inequality
Sharif et al. (2024)	Urdu automatic speech recognition	AI as technical system	Low-resource but advancing	Not addressed	Technological imbalance
Cristaldi (2025)	Global language gap in AI	AI as gatekeeper of access	Minor global presence	Digitally invisible	Digital linguistic injustice
Srivastava et al. (2024)	English–Punjabi translation	AI as technical challenge	Reference language	Complex and under-resourced	Structural exclusion
Khan et al. (2025)	Language marginalization in Pakistan	AI as extension of policy ideology	Symbolic power	Marginalized	Ideological continuity
Hashmi et al. (2025)	Language ideology and power	AI implied in policy discourse	Dominant	Excluded	Discursive silence

The analysis of academic and policy-oriented language reveals that the concept of artificial intelligence is becoming more and more represented as an effective linguistic and policy agent, and not a technological instrument. The academic sources always emphasize that the development of AI is conditional upon the state patronage, institutional priorities, and the presence of linguistic data, which puts already prevailing languages in advantage. The research dedicated to Punjabi reveals that AI has a promising future in terms of language revitalization, yet the opportunity is not being exploited because of the poor policy intervention, absence of datasets, and political oversight (Butt et al., 2025). The Punjabi language is also frequently presented as one that requires support more than a language to be involved in AI systems. This framing places the Punjabi in the dependency and vulnerability part of the AI ecosystem.

Urdu on the other hand has been positioned as a low resource but an advancing language. The studies of speech recognition and translation point out the progressive technological advances in Urdu AI, although the issues persist (Sharif et al., 2024). This demonstrates that Urdu enjoys an advantage of comparatively greater research focus and institutional concern than that of Punjabi. Considering the critical discourse, these studies indicate continuation between the traditional language policy and the discourse of AI. Historically privileged languages are still given the priority to study in AI research, and minor languages are underrepresented. AI thus brings hierarchies within language into the new digital space. Comprehensively, the results indicate that an academic and policy-driven discourse discusses AI as a non-state controller of language use, which forms access, visibility and legitimacy. In the absence of an inclusive language planning, AI will probably increase the linguistic inequality of low-resource languages such as Punjabi, as well as solidify the position of dominance of Urdu over others.

5. DISCUSSION

As evident in the analysis, the hierarchies that were instituted in the language policy of Pakistan since 1947 are reflected in the era of artificial intelligence. Language policy documents which are regularly in place put Urdu as the language of national unity, administration and education and Punjabi as absent or functionally represented. The ideological basis of this historical favoritism of Urdu directly predetermines the treatment of languages in digital and AI space. The results show that AI systems are not able to act without this policy history. They pass on and replicate the same language ideologies instead. The Urdu language is more visible and usable in AI systems like translation systems, learning platforms and digital governance whereas Punjab is not. This persistence proves that AI does not disrupt language inequality but, instead, it is projecting state-induced language hierarchies into the digital realm.

The major result of this research is that artificial intelligence is viewed as a new language policy actor despite the lack of legal or policy guidelines. AI technologies are also effective in determining the usable languages, learnable languages and legitimate languages in the digital spaces, through their decisions on data availability, system design, and language support. The language content analysis in terms of AI mediators demonstrates that Urdu is partially supported in various AI domains, but Punjabi is underdeveloped or not supported at all. This trend is not associated with institutional apathy but administrative blindness. Artificial intelligence is based on the use of data, financial support, and research priorities all of which are influenced by political and institutional decisions. In this respect, language policy results are imposed by AI without any legislation. This observation

undermines the existing concept of language policy as a process that is controlled by a state. The study has shown that algorithm systems are currently playing the policy roles and they influence language access and updated participation in daily life.

The results indicate that there is systematic structural marginalization of Punjabi in AI settings. The Punjabi language has a low presence in the machine translation system, is mostly not supported by speech recognition applications, and is not supported by AI-based learning platforms. Such constraints are usually elaborated in the form of technical issues like dialect difference, script difference, and unavailability of corpora. Nonetheless, the critical examination reveals that such technical explanations cover a further ideological and political concern. The absence of Punjabi data sets cannot be attributed to chance but is symptomatic of historical exclusion of Punjabi in the education, administration and funding of research. This places Punjabi in the AI age with a poor institutional base, and hence digitally vulnerable. This proves that digital marginalization is a continuation of socio-political marginalization and not a technological failure.

These inequalities are further enhanced by academic and policy-oriented discourse. The Urdu language is presented in terms of low resources that are growing and enriching with the help of AI studies. Punjabi, however, is presented as a language under crisis and reliant on resuscitation, and is limited by structural factors. This framing variation is a major one. It makes Urdu worthy of investment and innovation, and Punjabi made an issue that should be dealt with instead of a language that should be empowered. This kind of discourse influences the research agenda, funding priorities, and policy concerns, which directly impact the way in which the languages are integrated in AI systems. Symbolic power is reproduced in this framing, though, in a critical viewpoint. Punjabi is made to appear linguistically intricate and institutionally feeble whereas Urdu is made to appear manageable and nationally significant.

Despite the potential of digital platforms to problematize the traditional language hierarchies, the results indicate that AI-related digital spaces are reproducing the linguistic inequality instead of challenging it at the moment. Though there are some grassroots digital activism, it does not have a large impact on the development of AI or policy direction. The absence of institutional support makes the role of Punjabi in AI weak and informal. This is one of the key contradictions of technology supposed to be neutral and inclusive is in fact escalating exclusion. The prospect of AI as a democratizing agent is still pending to come to fruition in the case of the marginalized languages.

The cumulative results indicate that the language policy in Pakistan has taken a different turn as linguistic power is exercised by the technological systems. In case of the existing trends, Punjabi can become even more marginalized, not only in schools and media but also in digital governance, education, and even the communication mediated by AI. It states that the linguistic justice of the AI era should be based on explicit and inclusive language planning where the regional languages are considered both in policy and in the technology design. Otherwise, AI remains a silent and potent tool of linguistic exclusion in the future.

On the whole, the results of the present research indicate that the language policy ideology remains a commanding factor in determining the language performance within the artificial intelligence systems. The notion of AI does not act independent of historical and institutional language policy; in fact, it is a de facto language policy agent where the existing forms of power relations are incorporated into digital infrastructures. It is analyzed that Punjabi is still structurally and digitally marginalized because of long-term policy inattention and failure to be supported by institutions, whereas Urdu has been preserved in the form of both symbolic and functional supremacy in the educational, governmental, and AI-mediated communication. These trends allow showing that the development of technology does not necessarily lead to linguistic equality since AI systems replicate the existing hierarchies in more traditional areas of policy. Through this transformation, the study provides an important contribution to language policy research and AI studies by revealing that the equality of linguistic representatives in Pakistan ceased to be combated in the classroom and in the constitution and shifted to algorithms and data sets.

The results of the present research are consistent with and far-reaching the previous studies on Urdu, Punjabi, and artificial intelligence. The previous research is primarily dedicated to the technical performance, system design, and linguistic issues, whereas the current research predicts the language policy, ideology, and power relations as the background of the AI performance.

Previous studies mostly consider AI as a technical tool. As a case study, Kumar and Bansal (2017) examine machine translators on Punjabi and Urdu and explain that the absence of a parallel corpus, script diversity, and data paucity is a significant difficulty. Likewise, Singh et al. (2016) demonstrate that Urdu to Punjabi translation can be highly accurate with the help of statistical models, which are carefully trained. These studies indicate that AI has the potential to assist both languages in case there is an adequate amount of the linguistic support. However, the present research concludes that technical capability is not the only indicator that can be used to conclude the inclusion of language. Whereas previous research deals with how translation can be enhanced, this study has clarified why Punjabi is not represented in the mainstream AI systems though technically feasible. The results indicate that in the context of AI language outcomes weak policy support, institutional neglect, and historical marginalization have a decisive role. Accordingly, the present study repositions AI as a non-state language policy actor instead of a neutral instrument.

A number of earlier studies warrant the linguistic complexity as the key obstacle. Kumar and Bansal (2017) and Singh et al. (2016) point to the script variations (Shahmukhi and Gurmukhi), morphology, and alignment issues. Bhatti et al. (2025) go on to clarify that AI dialogue systems have issues with pragmatics, honorific and regional speech regulations in Pakistani languages. The given study does not negate such difficulties but claims that complexity in linguistics turns into a justification but not a cause of exclusion. The results reveal that the complexity of Punjabi is mentioned in scholarly discussions multiple times, whereas the lack of state financing, exclusion of the curriculum, and the lack of national planning of AI languages are not paid much attention. This study thus reverses the focus of the explanation in terms of language to structural and ideological marginalization.

Other previous studies define AI language development as functional or strategic. The article on multilingual translation systems by Srimal and Kumar emphasizes the national security issue and intelligence concerns where languages are treasured due to their surveillance capability and strategic benefit but not because of cultural or social inclusion. In this framing the languages are the instruments of power of the state. On the contrary, the present research is based on the language of justice approach. It concludes that AI systems favor languages which already possess symbolic and institutional force. This is the advantage that Urdu has, with Punjabi being digitally invisible. The study claims that this kind of prioritization creates inequality instead of intended inclusive national development.

According to the recent research of Ahmed et al. (2025) and Bhatti et al. (2025), AI has the potential of documenting the undocumented and under-resourced Pakistani languages. These research studies recommend culturally enhanced datasets and context-sensitive models as a way of enhancing language documentation and preservation. The present research concurs with this possibility but adds a serious limitation: documentation that is not integrated with a policy is still weak. The results indicate that the Punjabi-related AI projects are still one-off projects, instead of being included in a consistent national policy. This is in contrast to the Urdu oriented AI research which has a better institutional support. The study therefore claims that revitalization programs cannot be effective without digital language planning and policy recognition.

Earlier studies tend to refer to Urdu as well as Punjabi as low-resource languages (Kumar and Bansal, 2017; Singh et al., 2016). Nevertheless, the results of the present research indicate that there is a significant difference: Urdu is low-resource and valued though, Punjabi is low-resource and disregarded. This is a major distinction. Urdu is gaining more and more recognition in the speech recognition, translation and artificial intelligence dialogue systems, with Punjabi being a sideline. The study reveals that the status of resources is not sufficient to explain the inequality, ideological value and policy support is what makes the difference between the languages with low resources being attended to or not.

It is a research that unites the Critical Discourse Analysis (CDA) and artificial intelligence (AI) by in-texting AI as a technical system but a discursive and ideological framework that creates language hierarchies. In CDA terms, power is effected through texts, silences and representational patterns of policy and institutional discourse. These results indicate that comparable processes are now in play on the AI systems with data availability, model design, and language support becoming discursive decisions favoring particular languages and discriminating against others. There is scholarly and

policy-based research that demonstrates that Urdu receives legitimacy and use in AI settings whereas Punjabi is underrepresented because of the poor institutional support and lack of policy support (Butt et al., 2025; Hussain et al., 2024; Hashmi et al., 2025). In this regard, AI is a location where discourse is codified, and ideology is incorporated into algorithms. According to Cristaldi (2025), AI is a filter of linguistic access, and it determines the individuals who may have full access to digital life. Through the combination of CDA and AI analytical processes, this study shows how algorithmic systems replicate historic language ideologies, which carry the state-based language policy into the realm of digital governance. This summary enables the definition of AI as a non-state language policy mechanism in which power is manifested via technical infrastructures which seem to be neutral but which in fact, systematically support linguistic inequality.

5.1 Implications and Policy Recommendations

This research has significant theoretical, practical and policy implications on governance of the language in Pakistan in the era of artificial intelligence. The analysis reveals that linguistic inequality is no longer created exclusively by state policies and education systems but is becoming a part of AI-mediated technologies hence algorithms and data systems have become new sources of language power. This means that language policy has to extend beyond the traditional spheres to digital and artificial intelligence settings as the primary location of language planning. To resolve these issues, policymakers need to officially consider regional languages like Punjabi as a part of the national language and AI strategies, facilitate the creation of state-supported linguistic datasets, and focus on the low-resource languages when distributing resources to AI studies. There should also be reforms in the educational system to incorporate Punjabi in school curricula and AI-based learning systems to make its use in the fields of knowledge and technology a normal aspect. Furthermore, linguistically fair and ethically sensitive design methods which allow cultural relevance, sensitivity to dialect, and accessibility should be embraced by the AI developers. To make sure that digital innovation facilitates linguistic inclusion instead of repeating the patterns of historical inequalities, collaboration between policymakers, academics, AI researchers, educators, and language communities should take place.

6. CONCLUSION

This study has explored how language policy works in the time of artificial intelligence and what it means to Urdu and Punjabi in Pakistan by combining the language policy analysis with the Critical Discourse Analysis (CDA) and works related to AI. The results indicate that the linguistic hierarchies developed by the state language policy since 1947 have not diluted with the technological advances, but on the contrary, have been replicated and reinforced by AI systems. Urdu still retains a symbolic and functional hegemony in policy texts, education and in digital platforms, and Punjabi is an outsider even though it is the predominant language in the nation. The review shows that AI is not an impartial technological instrument. Instead, it is considered a non-state language policy agent that influences the visibility, usability and legitimacy of language. Algorithms are embedded with ideological preferences on choices of datasets, language support and system design, and are often not subject to policy debate. This means that Punjabi has been treated as a language of digital invisibility in translation engines, speech recognition applications, and other AI-driven learning platforms but Urdu enjoys a comparatively more significant institutional and research presence. This article connects the CDA and AI to demonstrate that the circulation of power is no longer supported only by the policy texts but also by the code. Linguistic exclusion does not exist in classrooms, or even in constitutions or media speech any more, it is trapped in algorithms and digital infrastructures that are becoming paramount arbiters of daily interactions. The use of AI-based systems will only exacerbate linguistic inequality with no intention to change the situation, pushing the local languages, like Punjabi, further to the periphery. Finally, the study concludes that the future of linguistic equality in Pakistan lies in the fact that the language policy has shifted to the algorithmic realm. The continuation of AI without planning language inclusion will be strengthening past inequalities. Nevertheless, through active policy intervention, AI also can be used as an instrument of multilingual inclusion, cultural safeguarding, and linguistic justice, so that both Urdu and Punjabi have a role to play in the digital future of Pakistan.

7. REFERENCES

Ahmed, W., Bhatti, Z. I., & Shah, R. H. (2025). The role of AI in documenting pragmatic markers and semantic structures in undocumented Pakistani languages. *Journal of Arts and Linguistics Studies, 3*(2), 1891–1906.

Arshad, A. (n.d.). *Language, partition, and politics: A historical examination of Punjabi language policy in post-independence Punjab, Pakistan*.

Bhatti, Z. I., Muhammad, G., Aslam, S., & Salamat, A. (2025). Managing contextual shifts: Pragmatics and semantics in AI dialogue systems for Pakistani language documentation. *Kashf Journal of Multidisciplinary Research, 2*(02), 1–14.

Butt, M., Ashraf, M. J., Sanjarani, K. M., Khan, M. A., & Arshaad, T. (2024). The politics of language and the fall of Dhaka: A discourse historical analysis. *Migration Letters, 21*(S14), 1065–1084.

Butt, M., Rouf, A., Irfan ul Haq, M., Kalyan, K. S., Rasool, A. G., & Khan, M. A. (2025). Artificial intelligence and Punjabi language revitalization: Opportunities, challenges, and policy interventions. *Pakistan Journal of Life and Social Sciences, 23*(2), 967–984.

Cristaldi, M. P. E. (2025). *You shall speak my language: Mind the language gap in the era of AI*. Proceedings Book, 11.

Hashmi, T. M., Hyder, S., Gujjar, M., Khan, H., Gondal, M. I., & Khan, M. A. (2025). Language, ideology and policy: A CDA of power relations among Urdu, Punjabi, and English in Pakistan. *Pakistan Journal of Life & Social Sciences, 23*(2).

Hussain, A. P., Khan, M. A., & Zahoor, Z. (n.d.). *Cultural capital and linguistic marginalization: A critical discourse analysis of Punjabi and Saraiki languages in Pakistan*.

Hussain, F., Butt, M., Sarwar, T., & Khan, M. A. (2024). The evolution of language politics: A CDA study of Urdu and Punjabi in Pakistan. *History of Medicine, 10*(2), 956–980.

Khan, H., Butt, M., Ishaq, S., Wattoo, Z. H., Yaseen, R., & Khan, M. A. (2025). Marginalization of Punjabi in Pakistan: Exploring language. *Pakistan Journal of Life and Social Sciences, 23*(2), 70–86.

Khan, M. A., Mahjabeen, A., Hussain, F., & ShehrYar. (2020). Socio-political dimensions of language policy and planning: A case study of Pakistan's multilingual landscape (1947–2020). *PalArch's Journal of Archaeology of Egypt/Egyptology, 17*(12), 1906–1923.

Kumar, A., & Bansal, N. (2017, September). Machine translation survey for Punjabi and Urdu languages. In *2017 3rd International Conference on Advances in Computing, Communication & Automation (ICACCA) (Fall)* (pp. 1–11). IEEE.

Safder, I., Abu Bakar, M., Zaman, F., Waheed, H., Aljohani, N. R., Nawaz, R., & Hassan, S. U. (2024). Transforming language translation: A deep learning approach to Urdu–English translation. *Journal of Ambient Intelligence and Humanized Computing, 15*(10), 3651–3662.

Sharif, M., Abbas, Z., Yi, J., & Liu, C. (2024). From statistical methods to pre-trained models: A survey on automatic speech recognition for resource scarce Urdu language. *arXiv Preprint*, arXiv:2411.14493.

Singh, U., Goyal, V., & Lehal, G. S. (2016). Urdu to Punjabi machine translation: An incremental training approach. *International Journal of Advanced Computer Science and Applications, 7*(4).

Srimal, P. K., & Kumar, M. K. (n.d.). *Language translation model for four languages–Urdu, Pashto, Punjabi and Nepali*.

Srivastava, M., Malik, N., Gupta, A., Chhabra, M., Khepra, S., & Abidi, A. I. (2024). Translation English to Punjabi: A concise review of significant approaches. In *2024 International Conference on Computational Intelligence and Computing Applications (ICCICA)* (Vol. 1, pp. 483–488). IEEE.

Tripathi, A. K. (2025). Challenges of language survival in digital perspectives: Case and context of India. In *The Routledge handbook of endangered and minority languages* (pp. 171–187). Routledge.