Frontiers in Social Sciences Research Volume 2, Issue 2, February 2026
ISSN: 3108-0324
DOI: https://doi.org/10.24113/em0kjk03

AI, Algorithms, and Linguistic Bias: How Technology Shapes Language Use in the Digital Age

Ravi Pawar

Research Scholar

Institute for Excellence in Higher Education (IEHE)

Bhopal, Madhya Pradesh, India

ravipawar9825@gmail.com

Abstract

This paper examines how artificial intelligence (AI), machine learning algorithms, and automated digital systems shape linguistic practices, reinforce or challenge linguistic hierarchies, and influence communication in contemporary society. As digital platforms increasingly mediate human interaction, algorithms determine what content becomes visible, which linguistic varieties are privileged, and how users adapt their language to gain visibility and engagement. The study explores algorithmic bias in search engines, social media feeds, voice assistants, and automated moderation systems, highlighting how these technologies reproduce existing social inequalities related to class, caste, gender, and ethnicity. Drawing on sociolinguistic theories of language ideology, linguistic capital, and digital discourse, the paper argues that AI-driven communication environments are not neutral but deeply ideological. They shape linguistic norms, influence identity performance, and regulate public discourse. The findings underscore the need for critical sociolinguistic engagement with AI systems to ensure equitable, inclusive, and culturally sensitive digital communication.

Received 25 January, Accepted 19 February 2026

Keywords: Artificial Intelligence; Algorithmic Bias; Sociolinguistics; Digital Communication; Linguistic Inequality; Language Ideology; Machine Learning; Moderation Systems; Voice Assistants.

Introduction

Artificial intelligence and algorithmic systems have become central to digital communication. From search engines and social media feeds to voice assistants and automated moderation tools, AI shapes how people interact, what they see, and how they express themselves. Sociolinguistics, traditionally concerned with human-to-human communication, must now grapple with machine-mediated communication, where algorithms influence linguistic norms and social hierarchies.

This paper explores how AI systems shape language use and linguistic inequality. It examines:

- How algorithms privilege certain linguistic varieties
- How automated moderation misinterprets non-standard or marginalized speech
- How voice assistants reinforce dominant accents
- How predictive text and autocorrect influence linguistic creativity
- How AI-driven platforms shape identity and discourse

The study argues that AI systems reflect the biases of their creators and training data, reproducing linguistic hierarchies and shaping digital communication in subtle but powerful ways.

AI, Language, and Power: A Sociolinguistic Perspective

Sociolinguistics recognizes that language is tied to power, identity, and social structure. AI systems, trained on large datasets, inherit the linguistic ideologies embedded in those datasets.

1. Language Ideology in AI Systems

AI systems often reflect dominant linguistic ideologies:

- Standard English is treated as "correct."
- Non-standard varieties are flagged as errors.
- Regional accents are misrecognized.
- Vernacular expressions are misclassified as inappropriate.

These patterns mirror offline linguistic hierarchies.

2. Linguistic Capital and Digital Prestige

Pierre Bourdieu's concept of linguistic capital becomes crucial in digital spaces. Users with proficiency in standardized English often gain:

- Higher visibility
- More credibility
- Better algorithmic ranking

AI systems thus amplify existing inequalities.

Algorithmic Bias in Social Media Platforms

Social media algorithms determine what content users see. These algorithms are not neutral; they privilege certain linguistic patterns.

1. English Dominance in Algorithmic Ranking

Platforms often prioritize English-language content because:

- It has higher global engagement
- It aligns with training data
- It is easier for AI to process

This marginalizes regional languages and hybrid varieties like Hinglish.

2. Hashtag Visibility and Linguistic Inequality

Hashtags in English trend more easily than those in vernacular languages. For example:

- #Love wins over #Pyar
- #Motivation trends more than #Prerna

This shapes how users choose language for visibility.

3. Algorithmic Suppression of Activist Language

AI moderation systems sometimes suppress activist or political speech, especially when expressed in non-standard varieties. For example:

- Dalit activism posts may be flagged due to assertive vernacular expressions.
- Feminist posts using reclaimed terms may be misclassified as abusive.

This demonstrates how algorithms regulate discourse.

Automated Moderation and Linguistic Misinterpretation

Automated moderation systems struggle with linguistic nuance, especially in multilingual contexts.

1. Misclassification of Vernacular Speech

AI may misinterpret:

- Sarcasm
- Irony
- Code-mixing
- Regional idioms

For example, a humorous Hinglish insult among friends may be flagged as harassment.

2. Cultural Context and Semantic Ambiguity

AI lacks cultural context. Words that are harmless in one culture may be flagged in another.

3. Impact on Marginalized Communities

Marginalized groups often use linguistic styles that deviate from standardized norms. Automated moderation disproportionately targets:

- AAVE (African American Vernacular English)

- Dalit vernacular expressions
- LGBTQ+ slang
- Regional dialects

This results in digital silencing.

Voice Assistants and Accent Bias

Voice assistants such as Siri, Alexa, and Google Assistant rely on speech recognition models trained on dominant accents.

1. Recognition Bias

These systems often struggle with:

- Indian English accents
- Regional Indian accents
- Non-native English speakers
- Code-mixed speech

Users must modify their speech to be understood, reinforcing linguistic inequality.

2. Accent Prestige and Digital Identity

Users may adopt "neutral" or "global" accents when interacting with AI, reflecting internalized linguistic hierarchies.

3. Exclusion of Vernacular Languages

Many Indian languages are poorly supported by voice assistants, limiting accessibility for non-English speakers.

Predictive Text, Autocorrect, and Linguistic Creativity

Predictive text systems influence how people write.

1. Standardization Pressure

Autocorrect often "corrects" non-standard or hybrid forms:

- Hinglish words

- Regional spellings
- Slang
- Creative spellings

This pressures users to conform to standardized English.

2. Influence on Youth Language

Predictive text shapes youth communication by:

- Suggesting certain phrases
- Limiting linguistic experimentation
- Reinforcing dominant linguistic norms

3. Loss of Vernacular Expression

Autocorrect may erase culturally specific expressions, reducing linguistic diversity.

Search Engines and Linguistic Visibility

Search engines determine which linguistic varieties appear in top results.

1. English-Centric Search Results

Search engines prioritize English content, even for queries in regional languages.

2. SEO and Linguistic Capital

Search Engine Optimization (SEO) favors:

- Standard English
- Short, clear sentences
- Keywords aligned with global trends

This shapes how content creators write.

3. Marginalization of Regional Knowledge

Local knowledge in regional languages may be buried under English-dominant results.

AI in Education and Linguistic Inequality

AI-driven educational tools influence language learning.

1. English as the Default

Most AI learning platforms prioritize English, reinforcing its prestige.

2. Bias in Assessment Tools

Automated essay scoring systems may penalize:

- Code-mixing
- Non-standard grammar
- Regional idioms

3. Impact on Students

Students may internalize linguistic insecurity, believing their natural speech is "incorrect."

AI, Identity, and Digital Self-Presentation

AI systems shape how users present themselves online.

1. Algorithmic Persona Construction

Users adapt their language to:

- Gain visibility
- Avoid moderation
- Fit platform norms

2. Linguistic Self-Censorship

Fear of moderation leads to:

- Avoidance of vernacular expressions
- Avoidance of political language
- Sanitized communication

3. Reinforcement of Dominant Identities

AI systems privilege dominant linguistic identities, marginalizing others.

Ethical Concerns and Sociolinguistic Implications

AI-driven linguistic environments raise ethical questions.

1. Whose Language Counts?

AI systems often reflect:

- Western linguistic norms
- Urban middle-class speech
- Standardized English

This marginalizes diverse linguistic communities.

2. Transparency and Accountability

AI companies rarely disclose:

- Training data sources
- Linguistic biases
- Moderation criteria

3. Need for Inclusive AI Design

Sociolinguists must collaborate with technologists to ensure:

- Multilingual training data
- Culturally sensitive moderation
- Accent-inclusive speech recognition

Conclusion

AI and algorithmic systems profoundly shape linguistic practices in the digital age. They influence what languages are visible, how users express themselves, and which linguistic varieties gain prestige. While AI offers opportunities for linguistic innovation and accessibility, it also reinforces existing inequalities related to class, caste, gender, and ethnicity. Sociolinguistic engagement with AI is essential to ensure that digital communication remains inclusive, diverse, and equitable. As AI continues to evolve, understanding its linguistic implications becomes crucial for shaping a fair and culturally sensitive digital future.

Conflict of Interest: The corresponding author, on behalf of second author, confirms that there are no conflicts of interest to disclose.

Works Cited

Bender, Emily M., and Batya Friedman. “Data Statements for Natural Language Processing.” Transactions of the ACL, vol. 6, 2018, pp. 587–604.

Bourdieu, Pierre. Language and Symbolic Power. Harvard University Press, 1991.

Crawford, Kate. Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press, 2021.

Eubanks, Virginia. Automating Inequality. St. Martin’s Press, 2018.

Floridi, Luciano. The Ethics of Artificial Intelligence. Oxford University Press, 2019.

Noble, Safiya Umoja. Algorithms of Oppression. NYU Press, 2018.

Tufekci, Zeynep. Twitter and Tear Gas. Yale University Press, 2017.

van Dijck, José. The Culture of Connectivity. Oxford University Press, 2013.