

Lost in Translation, Lost in Nation: AI-Induced Linguistic Homogenization as a Threat to Rastravadah

Dr T. Jude Livingston*  

Abstract

In Indian philosophy, Rastravadah is not a single, unified identity but a collection of many identities. This idea depends on the coexistence of India's many languages, dialects, and ways of speaking. This paper examines a subtle threat to that foundation: how AI can make language more uniform. Large language models are mostly trained on widely used languages and tend to prefer the most standard forms. As a result, they often smooth out regional expressions, dialects, and cultural details in translation, writing, and daily communication. More people are using these tools to write or translate in Tamil, Malayalam, Hindi, and other Indian languages. Over time, their language starts to sound more average and less unique. This slow change weakens the special qualities that have shaped India's diverse national identity.

This article argues that this threat to Rastravadah comes from the tools we use every day, not from official policies. First, it explains how Indian nationhood is built on linguistic diversity. Next, it examines how AI can make language more uniform, focusing on training data and patterns in AI output. Finally, it discusses what this means for teaching and language policy, and suggests solutions like developing regional language models and teaching critical AI skills in language classes. The paper brings together ideas from philosophy, nationhood, and AI to discuss India's linguistic and cultural diversity.

Keywords: Artificial intelligence; linguistic homogenization; Rastravadah; Indian nationhood; linguistic diversity; large language models; dialect preservation; critical AI literacy.

Submitted: 16.07.2026

Revised: 18.08.2026

Accepted: 23.08.2026

Published 02.10.2026

*Dr T. Jude Livingston, Assistant Professor, Department of English, Nazareth Margoschis College at Pillaiyanmanai, 628617, Affiliated to Manonmaniam Sundaranar University, Tirunelveli- 627012, Tamil Nadu, India.

©2026 Dr T. Jude Livingston. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction, provided the original author and source are credited.

Introduction:

In Indian philosophy, a nation has never been imagined as speaking with just one voice. This idea has shaped debates about India's language policies since independence. Rabindranath Tagore argues that nationalism should not erase the diversity of the people it brings together, advocating instead for a form of social unity that allows different communities to retain their distinctiveness (115). Similarly, Jawaharlal Nehru viewed India as a composite culture in which diverse traditions interacted and contributed to a "variegated and yet essentially unified" cultural heritage (228).

Religion, since the beginning, has acquired an indomitable space in the Indian subcontinent, as it served as a psychological anchor to the tormented Indian mind, which was amidst poverty and a struggle for freedom. A greater number of educated Indians have adopted spirituality as a way of life, and this paradigm shift in human consciousness is shedding light on the way society is evolving. There is a growing interest in changing lifestyle patterns and the use of spiritual tools as a corrective measure for self-evolution. (Sharma and Mehrotra 225)

Rastravadah, the idea of nationhood, has been described, in these discussions, as a chorus, not a solo. This chorus exists because of its many languages, not in spite of them. Tamil, Malayalam, Bengali, Bhojpuri, Hindi, and many other languages and dialects do more than exist side by side in India; they make up the nation itself. Each language has its own connection to land, memory, and community, and no single national language could replace them without losing something. "There is a widespread assumption that it is language proficiency levels that mediate social inclusion and that linguistic assimilation is the high road to social inclusion" (Piller and Takahashi 371).

Objectives and Aim:

This article explores what happens to that chorus of voices when machines start to speak for people. Artificial intelligence, AI, especially large language models used for reading, writing, and translation, does not simply pass language through unchanged. These models are trained on data that is not evenly spread across all languages and styles, and they reflect those imbalances in their output. Millions of Indians now use these tools to write letters, translate documents, or finish sentences. These tools do more than help people communicate; they also change how people use language. They push idioms, grammar, and style toward whatever is most common in their training data.

This article calls this process 'AI-induced linguistic homogenization' and argues that, if ignored, it quietly but seriously threatens the foundation of Rastravadah as a celebration of diversity. The argument proceeds in four movements. The first section outlines the philosophical understanding of Indian nationhood as constituted through linguistic plurality. It draws on Tagore's critique of nationalism, Nehru's reflections on India's composite culture, and later theorisations of the nation as a

plural and narrated construct. The second section analyses the technical and empirical basis of AI-powered linguistic homogenization.

This article is organised into four main parts. First, it explains how Indian nationhood is built on linguistic diversity, drawing on ideas from Tagore, Nehru, and later thinkers. Second, it examines how AI-driven language technology leads to more uniform language, focusing on how this happens through everyday tools rather than official policies. Third, it discusses how losing linguistic uniqueness also weakens the philosophical idea of a diverse nation. Fourth, it looks at what can be done in teaching and policy to push back against this trend. This study ends by suggesting that, to protect the nation's philosophy, we may first need to protect the unique ways people speak, against the Western model of the nation-state, in which a single language and culture are imposed as the price of political unity.

Tagore argued that India's power rested in its capacity to hold multiple cultural and linguistic streams within a single civilizational frame (115). This is not multiculturalism as an afterthought or a concession. It is, in Tagore's framing, constitutive of what makes India. Jawaharlal Nehru extended this thought in his reflections on India's civilizational continuity. He described the country as a palimpsest of successive cultural layers, each inscribed over the other without fully erasing what came before (Nehru 228). On this view, the country's unity was less a matter of sameness than of accommodation. Later scholarship on Indian nationhood has reinforced this reading. Sunil Khilnani's account of the idea of India describes it as a continually negotiated and unfinished project. The Indian nation, in this reading, is less a fixed essence. Instead, it is an ongoing argument among its constituent plurality, conducted in many languages at once (Khilnani 3).

This philosophical position echoes Benedict Anderson's theorisation of the nation as an imagined community. Anderson's account is not India-specific, but it is instructive here. A nation, on this view, is constructed and sustained through print language and shared narrative forms (Anderson 4). A nation is held together via the stories its members can tell about themselves, in a shared, though not necessarily singular, print vernacular. A nation is imagined partly through language. The health of that imagining then depends on the health and variety of the languages available to it.

It is necessary to highlight some features of Indian philosophy and its relationship to spirituality. India has a rich philosophical heritage over several thousand years. These ancient philosophical traditions included many schools of thought and a substantial body of intellectual argumentation about reality, the human person, and their interrelatedness. (Inbadas 320)

Homi Bhabha's account of the nation goes further still. He describes it as a site of narrative ambivalence, always being written and rewritten by its margins as much as its centre. A nation's coherence, on this view, is not threatened by linguistic multiplicity. Multiplicity is instead the very medium through which the nation is continually renarrated (Bhabha 4).

Bringing these ideas together, this article defines Rastravadah as Indian nationhood built on many languages and cultures, not just one dominating the rest. It is a constantly changing mix in which each language and culture is unique and important. Each one adds its own way of telling the nation's story. With this understanding, this article examines how AI language technology might threaten this diversity.

Analysis and Discussion:

The Mechanics of AI-Induced Linguistic Homogenization

Large language models are trained on massive quantities of text scraped from the internet, digitised books, and other available corpora. This process is not linguistically neutral. Scholarship on natural language processing has repeatedly shown that training data represents the world's languages in wildly disproportionate quantities. A small number of languages, chiefly English and a handful of other global languages, account for the overwhelming majority of digitised text. The remainder of the world's several thousand languages are represented only sparsely, if at all (Joshi et al. 6282–93). This includes most Indian languages beyond Hindi and a few others. Even within the better-resourced Indian languages, the digitised text that does exist skews heavily toward standardised registers. These are often urban and print-mediated: news writing, government documents, and formal literature. They are not the spoken idiom, regional dialect, or colloquial register most people use day to day.

The consequence is a systematic bias built into the very statistical foundations of these models. When a model is asked to translate, complete, or correct text within Tamil, it favours one outcome. It is more likely to produce the standardised, textbook register that dominates its training data. It is less likely to produce the regional idiom of Tirunelveli or the coastal Thoothukudi belt. Su Lin Blodgett and her collaborators make a related argument. What gets labelled a language model's fluency frequently reflects whose language was digitised, and whose was not (Blodgett et al. 5454–76). The model's output, on this view, quietly encodes a hierarchy among dialects and registers, even since it seems to offer neutral linguistic assistance. Emily Bender and Batya Friedman have similarly advocated for greater transparency about the provenance of training data. Without such transparency, homogenising bias can pass unnoticed as a simple technical limitation (Bender and Friedman 587–604).

This bias operates cumulatively rather than through any single event. This bias builds up over time, not through one big event. For example, a student in Thoothukudi might use a translation tool to write an email, and the tool changes their Tamil into a more standard form than they would use naturally. A writer might accept an AI suggestion that sounds more like the model's average Tamil than their own local voice. A language learner might mostly use AI-based exercises and start to think the model's version of the language is the only correct one. They miss out on the richness of dialects that a local teacher could provide. No single action erases a

dialect, but when this happens millions of times over many years, language use slowly shifts toward whatever the training data favours. Indian languages and English. It occurs within a single language as well. Regional and colloquial forms of Tamil, Malayalam, or Hindi are quietly displaced. Their more standardised, textbook counterparts take over whenever an AI tool is the mediating hand. The mechanism, in other words, does not require the disappearance of a language. It requires only the narrowing of the range of ways in which that language may be spoken, written, and imagined. This narrowing can occur even as the language's raw usage statistics appear healthy.

This trend also shows up in voice technology. More people in rural and semi-urban Tamil Nadu use speech recognition and voice assistants to access digital services. These systems are usually trained on urban, standard Tamil. So, someone speaking the coastal Thoothukudi dialect or Tirunelveli Tamil is more likely to be misunderstood. To be recognised, they often change how they speak to match the system's preferred style. Over time, this adjustment can lead to real changes in how communities talk, especially among younger people who use these systems more often. The dialect isn't banned, but it slowly fades as people use the technology.

From Language Loss to Nation Loss: Theorising the Threat

As Anderson describes, when language becomes more uniform, it limits the different ways people can imagine and talk about the nation (4). Bhabha says the nation is always being retold from its edges (4). If language is made more uniform, those unique edges lose their voice. This means fewer viewpoints from which to tell the nation's story, whether those are regional dialects, local phrases, or community-specific ways of speaking. The nation is still imagined, but in fewer and more similar voices. This trend also runs counter to what Nehru and Khilnani saw as key to Indian nationhood. In the past, India added new languages and cultures on top of old ones without erasing them. But AI-driven uniformity does the opposite; it blends these layers into one average style. Over time, the rich mix of languages becomes just one standard version.

Another important notion is that this threat comes from the tools we use every day, not from official policies. In the past, governments sometimes made language uniform through rules, but now it happens quietly through technology. No one tells Tamil speakers to use a standard version, but translation apps, messaging apps, and AI writing tools all push people toward it by default. Because this change is technical, not political, it doesn't spark the same resistance as past language policies. This hidden nature makes the threat to Rastravadah even more serious. It looks like a simple convenience, not a threat.

This does not mean that AI language technology is used to push nationalism or uniformity on purpose. The real issue is that intent doesn't matter here. Technology can weaken the commitment to diversity just by being easy to use and widely adopted. No one has to plan for this to happen. That's what makes this an infrastructural threat, not an ideological one. It also means that solutions need to be

practical, not just about making arguments.

A simple example can make this idea clearer. Imagine a college student in Thoothukudi writing a formal letter in Tamil using an AI writing tool. The tool is mostly trained on standard, printed Tamil, so it changes the student's local words and grammar to more neutral ones. The letter sounds fluent, but it no longer feels personal. If this happens again and again during the student's education, they may end up more comfortable writing in the AI's version of Tamil than in the way their family speaks. Now, think about millions of students across India using similar tools. What seems like a small convenience in one case becomes a big issue for the whole nation. This is how a hidden, everyday process can have a huge impact.

Pedagogical and Policy Implications

If the threat to Rastravadah comes from everyday tools and systems, then the solutions must also focus on technology and teaching, not just on philosophical debate. Based on the earlier analysis, three main actions are suggested.

One necessary intervention is investment in regional-language model development. Such development must take dialectal and colloquial variation seriously as data, not simply as noise to be cleaned away before training. One important step is to invest in developing language models for regional languages. This means treating dialects and everyday speech as valuable data, not just as mistakes to remove. It requires making an effort to digitise spoken language, regional literature, and community-specific phrases. These should be included in the data used to train language models for Indian languages. The goal is for these models to show the full variety of each language, not just the standard written form. Personal linguistic voice is often pushed into a more standardised or foreign register. They need the confidence and the vocabulary to resist that smoothing where it matters. This is particularly relevant to departments of English, such as the one convening this seminar. Such departments are well placed to model a pedagogy that uses AI tools without being quietly altered by them.

We also need to change how we judge success in language technology. Right now, fluency is measured by how closely an AI model matches a standard version of the language. "Artificial intelligence (AI) has gained tremendous attention recently as a potential tool to innovate teaching and learning in educational settings" (Ji et al. 48). However, the idea of applying AI in education is not entirely new. This means that making language more uniform is seen as a good thing. Instead, we should judge language models by how well they preserve and reflect dialect variety, not just how standard they are.

Teachers can also take simple steps in their own classrooms. For example, they can ask students to put a piece of their regional writing through an AI translation or paraphrasing tool, then look at what changed. Students can talk about these changes and, if needed, put back what was lost. This kind of exercise takes only class time and doesn't require new data or policies. It helps students become more aware of how AI affects language. Instead of just talking about the risks of AI,

students can see the effects for themselves. This makes the importance of Rastravadah more real than abstract discussions usually do.

None of these interventions requires rejecting AI language technology outright. This article's argument is not technophobic. It is, rather, an argument for treating language technology as a site of philosophical consequence. It deserves the same scrutiny that Indian philosophical and political thought has historically applied to more overt questions of national language policy.

Conclusion:

This article has argued that Rastravadah faces a subtle but real threat from AI-induced linguistic homogeneity. This paper has shown that Rastravadah faces a subtle but real risk from AI making language more uniform. In Indian philosophy, Rastravadah means a nation made up of many voices, not just one. The threat doesn't come from laws or force, but from the everyday use of translation tools, autocomplete, and AI writing assistants. These tools are trained on data that favours standard language over the wide variety of dialects and local speech that make up India's identity. Because this is a technical, not an ideological, threat, it is easy to miss and hard to regulate. Still, there are ways to respond: investing in language technology that includes dialects, teaching critical AI skills in language classes, and rethinking how we judge language quality in AI. To protect the nation as a group of many voices, we may first need to protect the unique ways each voice sounds.

Works Cited

- Anderson, Benedict. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Rev. ed., Verso, 2006.
- Bender, Emily M., and Batya Friedman. "Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science." *Transactions of the Association for Computational Linguistics*, vol. 6, 2018, pp. 587-604.
- Bhabha, Homi K. *The Location of Culture*. Routledge, 1994.
- Blodgett, Su Lin, et al. "Language (Technology) Is Power: A Critical Survey of 'Bias' in NLP." *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 5454-76.
- Inbadas, Hamilton. "Indian Philosophical Foundations of Spirituality at the End of Life." *Mortality*, vol. 23, no. 4, 2018, pp. 320-33. <https://doi.org/10.1080/13576275.2017.1351936>.
- Ji, Hyangeun, et al. "A Systematic Review of Conversational AI in Language Education: Focusing on the Collaboration with Human Teachers." *Journal of Research on Technology in Education*, vol. 55, no. 1, 2023, pp. 48-63. <https://doi.org/10.1080/15391523.2022.2142873>.

- Joshi, Pratik, et al. "The State and Fate of Language Variation and Inclusion in the NLP World." *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 6282–93.
- Khilnani, Sunil. *The Idea of India*. Penguin Books, 1997.
- Nehru, Jawaharlal. *The Discovery of India*. Oxford UP, 1946.
- Piller, Ingrid, and Kimie Takahashi. "Linguistic Diversity and Social Inclusion." *International Journal of Bilingual Education and Bilingualism*, vol. 14, no. 4, 2011, pp. 371–81. <https://doi.org/10.1080/13670050.2011.573062>.
- Sharma, Suhani, and Sakshi Mehrotra. "The Spiritual Revolution in India from the Ritual to the Spiritual Indian: Factors Shaping Spiritual Seeking in Modern Indians." *Journal of Spirituality in Mental Health*, vol. 20, no. 3, 2018, pp. 225–38. <https://doi.org/10.1080/19349637.2017.1411218>.
- Tagore, Rabindranath. *Nationalism*. Macmillan, 1917.