Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 11, pp 1-10, November-2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.311001
Available Online at https://www.jaiet.com

Localised Data, Unlocalised Justice: Data Localisation, AI and Tribal Communities in India

S.Sathia

Research Scholar, School of Law, Pondicherry University, Pondicherry, India

Abstract: India has made data localisation a central plank of its digital sovereignty agenda, and the state is now feeding locally held data into artificial intelligence (AI) systems for welfare delivery, land administration, policing and court management. Debates on localisation have focused on national security, law enforcement access and economic growth. Debates on tribal data have focused on consent and Gram Sabha authority in Scheduled Areas. This article brings the two together and asks a question neither literature has asked: when data about Scheduled Tribes is kept inside India and processed by state AI, who actually gains control over it? Using doctrinal analysis of the Digital Personal Data Protection Act, 2023, the DPDP Rules, 2025 and the Supreme Court's draft Regulations for Use of Artificial Intelligence in Courts, 2026, read with documented cases from Jharkhand, Telangana, Madhya Pradesh and Gujarat, the article identifies a *localisation paradox*: localisation shields tribal data from foreign extraction but concentrates it in state hands without giving communities any corresponding voice. It further shows that the remedy most often proposed, Gram Sabha consent, reaches only tribal people living in notified Scheduled Areas and leaves out a large share of India's Scheduled Tribes. The article proposes a two-track model of community data governance, territorial and non-territorial, and suggests concrete safeguards for welfare and judicial AI.

Keywords: Data localisation; Artificial intelligence; Scheduled Tribes; Indigenous data sovereignty; DPDP Act 2023; Judicial AI; Gram Sabha.

I. Introduction

In September 2017, eleven yearold Santoshi Kumari died in Simdega district of Jharkhand. It is said that her family's ration card had been cancelled months earlier because it was not linked to Aadhaar (Saha, 2018). According to the Right to Food Campaign, the family had not eaten for six days before her death and had complained to local officials about the card several times without any help (Thomson Reuters Foundation, 2017). State officials maintained that she died of malaria, not starvation, and an inquiry was ordered, yet the State's welfare minister himself conceded that no ration card should be cancelled merely because it lacked an Aadhaar link (Thomson Reuters Foundation, 2017). However, her death was not caused by artificial intelligence. It was caused by a database, a linking rule and an administrative system that treated a missing data point as proof that a family did not exist. That distinction matters less than it seems. The databases that excluded families like hers are the same databases on which India is now building algorithmic and AI-enabled systems of governance

Over the same decade, India has made data localisation a pillar of its digital policy. The Justice B. N. Srikrishna Committee recommended that a copy of personal data be stored in India, largely so that law enforcement could reach it without depending on foreign providers (Committee of Experts, 2018). The Reserve Bank of India required payment data to be stored locally (Reserve Bank of India, 2018). The Digital Personal Data Protection Act, 2023 (DPDP Act) and the DPDP Rules notified on 13 November 2025 now frame how personal data is handled, while the IndiaAI Mission pools government datasets for AI development. Localisation is usually defended as protecting Indian citizens from foreign surveillance and foreign corporations.

This article asks what that protection means for one group of citizens in particular: India's Scheduled Tribes. Tribal communities are not only individual data principals. They hold collective rights over land, forests and customary governance under the Fifth and Sixth Schedules of the Constitution, the Panchayats (Extension to the Scheduled Areas) Act, 1996 (PESA) and the Forest Rights Act, 2006 (FRA).

1.1 What existing writing has done, and what it has not

A small but growing body of work already connects AI and tribal communities in India. A 2026 article in *AI & Society* proposes a human-centric AI framework that treats tribal communities as co-decision-makers rather than end users

© 2026 JAIET All Rights Reserved CC BY 4.0 License

Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 11, pp 1-10, November-2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.311001
Available Online at https://www.jaiet.com

("Designing human-centric AI," 2026). Existing studies argues that digital infrastructures weaken Gram Sabha autonomy and calls for digital free, prior informed consent for minorities and scheduled groups ("From territorial to digital control," 2025). A recent commentary on forest governance makes a similar case for expanding Gram Sabha powers under PESA to cover AI and surveillance technologies (Raj, 2026). Dutta and Mazumdar (2025) document how Indian civil society activists contest a state-centred idea of data sovereignty.

These contributions share two features. First, none of them treats data localisation itself as the object of analysis; localisation appears, if at all, as background. Second, the remedy they converge on, Gram Sabha consent, is territorial. It works only where PESA and the Fifth Schedule apply. This article departs from existing writing on both counts.

1.2 Questions, argument and method

The article addresses three questions. Does data localisation protect tribal communities, expose them, or both? Where is the harm already visible as the state moves from databases to AI? And what governance model would protect tribal people both inside and outside Scheduled Areas?

The argument has two parts. The first is the *localisation paradox*: keeping tribal data inside India removes one risk, foreign extraction, while concentrating control in state agencies that the DPDP Act largely exempts. The second is the *territorial gap*: Gram Sabha consent, though valuable, cannot by itself protect the many tribal people who live outside notified Scheduled Areas, including those who appear in police, prison and court records far from their villages.

The method is doctrinal and qualitative. The article reads the DPDP Act, the DPDP Rules, the Supreme Court's draft Regulations for Use of Artificial Intelligence in Courts, 2026 and relevant judgments alongside documented cases reported by courts, government sources, civil society audits and investigative journalism. It does not claim statistical proof that localisation causes harm. Its claim is narrower and, I think, more useful: that the present legal allocation of control over tribal data leaves communities without a say at exactly the moment AI is scaling.

Section 2 sets out India's localisation framework. Section 3 explains why tribal data raises distinct constitutional questions. Section 4 examines four sites where AI and tribal lives meet. Section 5 develops the localisation paradox, and Section 6 draws lessons from other jurisdictions. Section 7 proposes a two-track governance model, and Section 8 concludes.

II. Data Localisation in India: Purposes and Architecture

India does not have a single, general data localisation law. What it has is a layered set of policies, sectoral directions and statutory powers which, taken together, keep a large and growing share of Indians' data within national jurisdiction. Understanding why these measures were adopted is essential, because the stated purposes tell us whose interests localisation was designed to serve.

2.1 The stated rationales

The most detailed official justification remains the 2018 report of the Committee of Experts chaired by Justice B. N. Srikrishna. The Committee recommended that at least one serving copy of personal data be stored in India. Its principal reason was that Indian law enforcement agencies faced delays in obtaining data held abroad, since requests routed through mutual legal assistance treaties could take many months (Committee of Experts, 2018; Burman & Sharma, 2021). The report also mentioned protection from foreign surveillance and the growth of a domestic digital economy, including an artificial intelligence ecosystem.

Critics pointed out at the time that the Committee localised data for law enforcement but did not reform the procedures by which the state itself accesses that data. In other words, localisation was designed to make data more reachable for the Indian state, without a matching increase in safeguards for the people the data describes. That observation is the starting point of this article's argument.

Sectoral rules followed. In April 2018, the Reserve Bank of India directed payment system operators to store all payment data only in India (Reserve Bank of India, 2018). Government cloud procurement has long preferred infrastructure hosted within the country. More recently, sovereign AI has become an explicit policy objective, with public statements linking local storage to building models trained on Indian data.

2.2 The DPDP Act, 2023 and the DPDP Rules, 2025

The DPDP Act does not itself impose a general localisation mandate. Section 16 permits transfers of personal data abroad except to countries the Central Government notifies as restricted, and it preserves stricter sectoral rules. The Act therefore leaves localisation largely to sectoral regulators and to the Government's notification power.

Two other features of the Act matter more for tribal communities. First, the Act is built around the individual "data principal". Consent, notice, correction and erasure are rights of

© 2026 JAIET All Rights Reserved CC BY 4.0 License

Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 11, pp 1-10, November-2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.311001
Available Online at https://www.jaiet.com

individuals. There is no concept of a community as a rights-holder over data that describes it collectively. Second, Section 17 allows the Central Government to exempt notified state instrumentalities from most of the Act's obligations in the interests of sovereignty, security, public order and related grounds. Such bodies may then process personal data without the consent, purpose limitation and storage limitation duties that bind private entities. Section 7 additionally permits the State to process personal data without fresh consent for providing subsidies, benefits and services.

The DPDP Rules were notified on 13 November 2025 with a phased commencement. Provisions establishing the Data Protection Board took effect immediately, consent manager registration follows after a year, and most substantive obligations apply from May 2027 (AZB & Partners, 2025). For the next several months, therefore, the substantive protections of the Act are not yet fully in force, while state data collection and AI deployment continue.

2.3 From localised data to sovereign AI

Localisation is no longer only about storage. It is becoming the raw material for national AI. In March 2025, the Government launched AIKosh, the IndiaAI datasets platform, which collects non-personal and anonymised datasets from ministries, public agencies and private contributors for AI model training. A Right to Information application filed in 2026 asked which ministries are contributing data, how "non-personal" is being defined and whether the legal basis for using government data in AI training has been vetted; at the time of filing, the government had not published answers to these questions (MediaNama, 2026).

This matters for tribal communities because much tribal data is collective rather than personal: village-level records, forest maps, language recordings, community resource data. Once anonymised or aggregated, such data falls outside the DPDP Act entirely, since the Act governs only personal data. The policy architecture, in short, moves data inward to Indian servers and then outward into AI pipelines, while the one statute meant to protect people covers only part of what moves. The point is not that localisation is wrong. It has real protective value, which Section 5 acknowledges. The point is that India's localisation framework was designed around two actors, the State and the foreign corporation. Communities do not appear in it at all.

III. Tribal Communities, the Constitution and the Question of Data

Scheduled Tribes made up about 8.6 per cent of India's population at the 2011 Census. They are not a uniform group: the category spans hundreds of communities with distinct languages, land relations and forms of self-governance. What they share, in constitutional terms, is a recognition that their relationship with the State requires special protection.

3.1 A constitutional architecture of collective rights

Article 244 and the Fifth Schedule provide for the administration of Scheduled Areas, with a Tribes Advisory Council and special powers for the Governor to modify the application of laws. The Sixth Schedule creates autonomous district councils in parts of the North-East. PESA extends Panchayati Raj to Fifth Schedule areas while placing the Gram Sabha at the centre, with powers over community resources, consultation before land acquisition and the safeguarding of customs and traditions. The FRA recognises both individual and community forest rights and gives the Gram Sabha the role of initiating the recognition process.

Taken together, these instruments express a principle unusual in Indian law: that a community, not only an individual, can hold rights and make binding decisions about its own affairs. It is exactly this principle that is missing from India's data law.

3.2 Privacy as an individual right

The Supreme Court's decision in *K. S. Puttaswamy v. Union of India* (2017) recognised privacy, including informational privacy, as a fundamental right under Article 21. The Court's reasoning centred on the autonomy and dignity of the individual. In the later Aadhaar judgment, *K. S. Puttaswamy v. Union of India* (2019), the Court upheld the Aadhaar framework for welfare delivery while striking down several provisions and noting concerns about exclusion.

The DPDP Act follows the same individual model. This creates a mismatch for tribal communities. Harms from data about a community, such as stigmatising labels attached to a whole tribe, inaccurate maps of community forests or language recordings used without consent, often do not fall on any one identifiable person. An individual-consent model has no way to recognise them.

3.3 Indigenous data sovereignty

Internationally, Indigenous scholars and organisations have developed the idea of Indigenous data sovereignty to address this mismatch. It holds that Indigenous peoples have the right to govern the collection, ownership and use of data about their communities, lands and knowledge (Kukutai & Taylor, 2016). The CARE Principles for Indigenous Data Governance (Collective benefit, Authority to control, Responsibility and Ethics) were developed to complement data-sharing principles that focus only on technical openness (Carroll *et al.*, 2020). The

© 2026 JAIET All Rights Reserved CC BY 4.0 License

Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 11, pp 1-10, November-2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.311001
Available Online at https://www.jaiet.com

United Nations Declaration on the Rights of Indigenous Peoples (2007) provides the normative backdrop, including the principle of free, prior and informed consent.

A 2025 systematic review found that the literature on Indigenous peoples and AI identifies risks such as erosion of cultural knowledge and data extraction that ignores Indigenous data sovereignty (Perera *et al.*, 2025). Most of this scholarship, however, comes from settler-colonial contexts such as Australia, New Zealand and North America, where Indigenous peoples are numerical minorities whose sovereignty claims rest on treaties or prior occupation.

India's position is different, and in one respect stronger. India has historically avoided the term "Indigenous" in international forums, but its own Constitution already recognises tribal collective authority. Indian tribal communities therefore do not need to import a foreign theory to claim collective control over their data. They can ground it in the Fifth and Sixth Schedules, PESA and the FRA. The difficulty is that these instruments were written for land and forests, not for databases.

3.4 The territorial limit

There is a further problem that existing writing on tribal data governance tends to pass over. PESA and the Fifth Schedule apply only in notified Scheduled Areas. An estimate placed before a national Adivasi convention suggested that around 59 per cent of India's Scheduled Tribes live outside the protections of Article 244 (The Wire, n.d.). Even if that figure is debated, the structural point holds: many tribal people live in districts that were never notified, have migrated to cities for work, or appear in state records far from their villages, as accused persons, prisoners, welfare applicants or litigants.

A governance model that depends entirely on Gram Sabha consent will not reach them. The data that most shapes their lives, such as police records, prison registers and court files, is generated in police stations, jails and courtrooms, not in village assemblies. Any serious proposal for tribal data governance has to address both places: the village and the institution.

IV. Where AI Meets Tribal Lives: Evidence from Four Sites

A common objection to research on AI and marginalised communities is that AI in Indian governance is still developing and that no harm has yet been proved. This section shows that the objection is only partly correct. Some harms are already documented in Indian courts and administration. Others are documented in the data systems on which AI is now being built. The four sites below were chosen because each involves state-held, domestically stored data and each touches tribal communities directly.

4.1 Welfare: from biometric exclusion to algorithmic exclusion

The link between data systems and tribal welfare is not hypothetical. Between September 2017 and July 2018, the Right to Food Campaign recorded 14 suspected starvation deaths in Jharkhand and found that Aadhaar-related failures clearly contributed in at least seven of them (Saha, 2018). The failures included ration cards cancelled for want of an Aadhaar link and fingerprints that point-of-sale machines would not read. One of those who died, Rajendra Birhor, belonged to a Particularly Vulnerable Tribal Group and had received neither a ration card nor a PVTG pension (Saha, 2018). The State government disputed that these were starvation deaths, and that dispute should be acknowledged; but the pattern of data-driven exclusion was documented repeatedly. Scholarly work on Aadhaar in the Public Distribution System has described this as a form of data injustice, in which entitlement becomes conditional on being correctly recorded (Masiero & Das, 2019; Khera, 2019).

The next stage is algorithmic. Since 2016, Telangana has used Samagra Vedika, a system that links records from multiple government databases to build profiles of residents and test welfare eligibility. A 2024 investigation found that faulty data and flawed matching had wrongly excluded thousands of eligible people from subsidised food, including a widow whose late husband was confused with a car owner of a similar name (Sambhav & Tapasya, 2024). Amnesty International's technical review raised concerns about the rights to social security, privacy, equality and remedy, and noted that neither the government nor the vendor had released source code or accuracy data (Amnesty International, 2024).

Samagra Vedika is not a tribal-specific system, and the investigations did not disaggregate by tribe. But the mechanism, linking locally held state databases and treating mismatches as ineligibility, is exactly the mechanism that failed tribal families in Jharkhand. Tribal households, with irregular documentation, remote locations and names often transliterated inconsistently, are structurally more likely to produce the mismatches such systems penalise.

4.2 Land and forest rights: when the portal overrules the Gram Sabha

The FRA places the Gram Sabha at the centre of recognising forest rights. Digitisation has quietly shifted that centre. In 2019, Madhya Pradesh introduced the Van Mitra portal and mobile application to review rejected claims. Of 6,27,513

© 2026 JAIET All Rights Reserved CC BY 4.0 License

Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 11, pp 1-10, November-2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.311001
Available Online at https://www.jaiet.com

claims received on the portal, only 2,94,585 were approved (101Reporters, n.d.). Claimants reported not knowing whether their applications had been accepted or rejected years after filing online (Vats, 2023). The Union Ministry of Tribal Affairs itself wrote to the State expressing concern that the application was undermining the rights of the Gram Sabha.

In Gujarat, satellite imagery has been used to verify forest rights claims. Activists reported that the imagery maps were not shared with claimants and that, in one phase, an agency evaluated about 1,82,000 claims and rejected 1,13,000 on the basis of imagery while other records were set aside (Newslaundry, 2024). Satellite imagery is meant to supplement other evidence, not replace it; it cannot capture traditional or seasonal use, and it cannot identify who occupies a plot.

These systems are not yet AI in the strict sense. They are, however, the data layer that AI-based land and forest analytics will inherit. The lesson is that once tribal land data is held in state digital systems, decisions migrate from the community to whoever operates the system.

4.3 Policing and prisons: inherited labels in digital records

The colonial Criminal Tribes Act, 1871 branded entire communities as hereditary criminals. It was repealed in 1952, and the communities concerned were "denotified", but many States later enacted habitual offender laws, and police continued to maintain registers of so-called habitual offenders. Researchers have documented how these registers are being digitised through the Crime and Criminal Tracking Network and Systems (CCTNS), creating a digital record that can feed crime-mapping and analytics (Bokil et al., 2021).

In *Sukanya Shantha v. Union of India* (2024), the Supreme Court held that prison manuals cannot use the category of "habitual offenders" to target members of Denotified Tribes, struck down such provisions as unconstitutional, and directed police to follow existing arrest guidelines so that Denotified Tribe members are not arbitrarily arrested. The judgment addresses manuals and practices. It does not address the digitised records that carry the same classifications forward.

The prison population shows why this matters. According to *Prison Statistics India 2023*, undertrials formed about three-quarters of India's prisoners, and Scheduled Tribes made up about 10 per cent of undertrials, above their share of the population (National Crime Records Bureau [NCRB], 2025). Legal aid reached only about half of all undertrials. Any analytical or AI tool trained on these records will learn the over-representation as a pattern, unless someone with the authority to object can examine the data first.

4.4 Courts: fabricated authority and missing languages

The Indian judiciary is adopting AI for translation, transcription, legal research and case management under Phase III of the e-Courts Project (Centre for Research and Planning, 2025). Two developments show that judicial AI risks are no longer speculative.

First, in *Pooja Ramesh Singh v. Jammu and Kashmir Bank Ltd.* (2026), the Supreme Court set aside orders of the National Company Law Tribunal and its appellate tribunal after finding that they relied on non-existent or misattributed precedents traceable to AI tools. The case involved a well-resourced corporate litigant represented by senior counsel who could detect the error. An Adivasi undertrial depending on overstretched legal aid would be far less likely to notice a fabricated authority in an order against them.

Second, AI translation in the courts has largely bypassed tribal languages. As of 27 September 2024, 73,963 Supreme Court judgments had been made available in vernacular languages on the e-SCR portal. Of these, 36,302 were in Hindi and 21,183 in Punjabi; only 31 were in Santali, 7 in Garo and 3 in Khasi (Ministry of Law and Justice, 2024). That is, fewer than one in a thousand translated judgments were in a tribal language. Languages such as Gondi, Bhili and Halbi, which are not in the Eighth Schedule, do not appear at all. Translated judgments are also issued for information only, since English remains authoritative under Article 348.

The gap is not academic. Lawyers and researchers working in Bastar and Jharkhand have long reported that many Adivasi undertrials speak only languages such as Gondi or Halbi and that courts rarely provide interpreters, so proceedings take place in a language the accused does not understand (Dogra, 2014; Bagaicha, 2016). The Ministry of Tribal Affairs launched Adi Vaani, an AI translation tool, in September 2025, covering Santali, Bhili, Mundari and Gondi in its beta version, against 461 tribal languages recorded in the 2011 Census (Free Press Journal, 2025). The tool is a welcome start, but its existence shows how far judicial AI is from reaching the languages tribal litigants actually speak.

V. The Localisation Paradox: Protection and Exposure

It would be easy, and wrong, to conclude from Section 4 that localisation is the problem. Every case discussed there involved data already held by the Indian State; none would have been prevented by storing it abroad. The more accurate conclusion is that localisation answers one question, *where* data sits, while the harms in Section 4 arise from two different questions: *who decides* how data is used, and *who can check* that

© 2026 JAIET All Rights Reserved CC BY 4.0 License

Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 11, pp 1-10, November-2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.311001
Available Online at https://www.jaiet.com

it is used fairly. This section sets out what localisation does protect, what it does not, and why the gap is especially wide for tribal communities.

5.1 What localisation protects

Localisation has genuine protective value. Data stored in India is within reach of Indian courts and Indian constitutional remedies under Articles 32 and 226. Data held by foreign providers can be subject to foreign legal compulsion, an issue that courts and judicial councils elsewhere have taken seriously when deciding where court records should be hosted. India's own judiciary has recognised this. The Supreme Court's draft Regulations for Use of Artificial Intelligence in Courts, 2026 require contracts with private AI providers to include on-premise or sovereign cloud deployment for systems processing sensitive judicial data, and they prohibit transfers of such data to external systems without express authorisation (Supreme Court of India, 2026, Regs. 46(4)(j), 48(1)).

Localisation also offers tribal communities some protection against a specific modern form of extraction: the harvesting of Indigenous language and cultural data by global technology companies. In New Zealand, the Māori broadcaster Te Hiku Media refused to sell its Māori language recordings to large technology firms, and Māori experts objected when a widely used speech recognition model was found to have been trained on more than 1,300 hours of Māori audio gathered without community consent (Business Day, 2023). As India builds tribal language corpora through initiatives such as Adi Vaani, keeping those corpora within Indian jurisdiction is a reasonable first safeguard.

5.2 What localisation does not protect

Localisation does not, however, address three risks that matter most to tribal communities.

The first is **domestic concentration**. Localisation was justified in large part by the State's own need for access (Committee of Experts, 2018). Combined with Section 17 of the DPDP Act, which allows notified state bodies to be exempted from core data protection duties, localisation increases the State's practical control over data while leaving the people described by that data with few means of contesting its use. For welfare, land and policing data, the State is not a neutral custodian; it is the decision-maker whose decisions are in question.

The second is **security**. Localised data is not automatically safe data. In 2022, a ransomware attack disrupted the All India Institute of Medical Sciences in Delhi, and in 2023 personal data of about 81.5 crore Indians, apparently linked to COVID-19 testing records, was offered for sale online (Business Today, 2023). Concentrating sensitive data in a few domestic systems creates high-value targets. The people least able to seek redress after a breach, including tribal households far from courts and lawyers, bear the greatest risk.

The third is **opacity**. Sovereignty can easily become closure. The draft court AI Regulations illustrate this tension well. They contain strong principles: AI must not introduce bias on grounds including caste and language, with special care for marginalised communities (Reg. 6(2)); it must not widen digital divides for linguistically diverse communities (Reg. 13); systems trained on demonstrably biased datasets are prohibited (Reg. 14(2)); and risk scoring for bail and recidivism is banned outright (Reg. 20(d)). Yet audits must be conducted in-house, and source code, algorithms and datasets may not be shared with any third party (Reg. 38(2)). A grievance remedy is available only for harm caused by *prohibited* uses (Reg. 52), not by permitted tools such as translation or docket prioritisation. Tribal communities, civil society and independent researchers therefore have no route to test whether the non-discrimination promise in Regulation 6 is being kept.

5.3 Why the gap is widest for tribal communities

These three risks fall unevenly. A corporate litigant who suffers from an AI-fabricated precedent can afford counsel to detect and challenge it. A professional whose name stays on the e-Courts portal after a case ends can petition a High Court to have it removed; in *Rohan Palaha v. State of Haryana* (2025), the Punjab and Haryana High Court ordered such a redaction after the petitioner's name on the portal caused failed background checks for jobs. An acquitted Adivasi from a Denotified Tribe, whose community has already been stigmatised by official records, is unlikely to bring such a petition.

This is the localisation paradox in its full form. Localisation solves the problem tribal communities are least likely to face, foreign access to state records, and leaves untouched the problems they are most likely to face: domestic concentration, insecure storage and closed systems that no one outside the State can audit. Keeping tribal data at home is necessary. It is not sufficient, because home is also where the decisions that harm tribal people are made.

VI. Lessons from Other Jurisdictions

India is not the first country to confront these questions. Four experiences, from the Global South and the Global North, offer lessons that are directly relevant.

© 2026 JAIET All Rights Reserved CC BY 4.0 License

Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 11, pp 1-10, November-2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.311001
Available Online at https://www.jaiet.com

Kenya: assess before rolling out: Kenya's national digital identity scheme, Huduma Namba, was challenged first by the Nubian Rights Forum, which warned that marginalised communities risked exclusion from enrolment. In October 2021, the High Court held the rollout of Huduma cards unlawful because the government had not carried out the data protection impact assessment required by the Data Protection Act, 2019, and ordered one before further processing (*R v. Joe Mucheru & 2 others, ex parte Katiba Institute*, 2021). The lesson for India is procedural: a mandatory, published impact assessment before a state data system is deployed gives affected communities a document they can examine and contest.

The Netherlands: transparency as a condition of legality: In February 2020, the District Court of The Hague halted SyRI, an algorithmic system used to flag welfare fraud risks, which had been deployed mainly in low-income and immigrant neighbourhoods. The court found that the legislation lacked sufficient safeguards for privacy under Article 8 of the European Convention on Human Rights and stressed that the system's opacity prevented people from understanding or challenging how they were flagged (*NJCM et al. v. The State of the Netherlands*, 2020). The relevance to India's draft court AI Regulations is plain: closed, in-house audits may protect institutional autonomy, but they make it difficult for affected people to test whether a system discriminates.

Australia: regulate before harm spreads: Australia's Robodebt scheme used automated income averaging to raise welfare debts, many of which were unlawful. A Royal Commission reported in July 2023 and recommended, among other things, a statutory body to monitor and audit automated decision-making by government (Royal Commission into the Robodebt Scheme, 2023). Robodebt is often cited as a warning about speed: by the time the law caught up, hundreds of thousands of people had been affected. India's draft court AI Regulations and the phased DPDP Rules create a window to build safeguards before scale.

New Zealand: Indigenous data inside the sovereignty conversation: New Zealand offers the most useful positive model. Te Hiku Media's refusal to sell Māori language data to large technology firms showed that a community organisation can act as a guardian of collective data (Business Day, 2023). At the institutional level, when the Ministry of Justice procured Te Au Reka, a new digital case management system for the courts, its procurement documents treated both jurisdictional risk and Māori data sovereignty as relevant considerations (Ministry of Justice [New Zealand], 2023). Court data sovereignty and Indigenous data sovereignty were discussed together, not separately.

Taken together, these experiences point in one direction. Kenya shows the value of prior assessment; the Netherlands, of transparency; Australia, of independent oversight; New Zealand, of placing Indigenous interests inside the sovereignty decision itself. India's current framework has elements of the first three in draft form, but it lacks the fourth entirely.

VII. Towards Community-Centred Data Governance: A Two-Track Model

The argument so far suggests that tribal data governance needs to answer three questions together: where data sits, who decides how it is used, and who can check its use. Localisation answers the first. The proposals below address the second and third, and they are organised in two tracks because tribal people encounter state data systems in two different settings.

Territorial track

- Applies in notified Scheduled Areas (Fifth Schedule) and Sixth Schedule areas.
- Authority rests with the Gram Sabha and autonomous district councils.
- Covers community land, forest, resource and language data collected in or about the area.

Non-territorial track

- Applies wherever tribal people appear in state records, including outside Scheduled Areas.
- Authority rests with the National Commission for Scheduled Tribes, State Tribal Research Institutes and legal services authorities.
- Covers welfare, police, prison and court data, and AI tools trained on it.

7.1 The territorial track

Within Scheduled Areas, the existing proposal for Gram Sabha consent is sound and should be adopted, with three refinements. First, consent should be required before community-level data is used to *train* or *deploy* state AI, not merely before it is collected, because the risks in Section 4 arise at the point of use. Second, the Gram Sabha should receive, in an accessible language, a plain summary of what the system does and what data it uses. Third, consent should be revocable for future uses. These refinements can be achieved by guidelines under PESA and the FRA without constitutional amendment, and Governors could use their Fifth Schedule powers to give them effect in individual States.

© 2026 JAIET All Rights Reserved CC BY 4.0 License

Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 11, pp 1-10, November-2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.311001
Available Online at https://www.jaiet.com

7.2 The non-territorial track

Outside Scheduled Areas, and in institutional settings such as police stations, prisons and courts, a Gram Sabha has no role. Here the aim should be oversight rather than consent. Four measures are suggested.

1. **Independent bias audits with tribal participation:** The draft court AI Regulations should allow audits by an independent panel bound by confidentiality, rather than in-house audits only, and that panel should include a nominee of the National Commission for Scheduled Tribes. The same model could apply to welfare algorithms like Samagra Vedika.

2. **Dataset certification consistent with *Sukanya Shantha*:** Regulation 14(2) already prohibits AI trained on demonstrably biased datasets. It should be given operational content: before police, prison or court records are used in any AI tool, they should be checked for classifications the Supreme Court has held unconstitutional, such as habitual-offender labels targeting Denotified Tribes.

3. **Verified translation in liberty matters:** Where an accused person speaks a tribal language, orders affecting liberty, such as bail rejections, charges and convictions, should be provided with human-verified translation or interpretation in that language. AI tools such as Adi Vaani can assist, but should not be relied on alone while their accuracy remains limited.

4. **A wider grievance remedy:** Regulation 52 should be extended to harm from permitted AI uses, not only prohibited ones, and legal services authorities should be tasked with helping tribal litigants use it.

7.3 Safeguards common to both tracks

Two measures apply across both tracks. The first is a published impact assessment, on the Kenyan model, for any state AI system processing tribal data, with a summary in the relevant tribal languages. The second concerns collective data that falls outside the DPDP Act. Rules or guidelines under the IndiaAI Mission should require that anonymised or aggregated datasets describing identifiable tribal communities, including language corpora contributed to AIKosh, be used only with documented community authorisation and with benefits returned to the community, consistent with the CARE principles.

None of these proposals requires abandoning localisation. They build on it. Localisation keeps tribal data within reach of Indian law; the two-track model makes sure that tribal communities, and not only the State, can use that reach.

VIII. Conclusion

India's debate on data localisation has been conducted largely between the State and foreign technology companies. This article has argued that a third party has been left out of that debate: the communities whose data is being kept at home. For Scheduled Tribes, localisation is a necessary safeguard against foreign extraction, particularly of language and cultural data. But it does nothing about the risks they actually face most often, which arise inside the country, in welfare offices, forest rights portals, police stations, prisons and courtrooms. The evidence reviewed here shows that these risks are not speculative. Data-linking failures have excluded tribal families from food entitlements; digital portals and satellite imagery have displaced the Gram Sabha's role in recognising forest rights; police and prison records still carry the imprint of colonial criminalisation; and Indian courts have already set aside decisions resting on AI-fabricated authority, while translation of judgments into tribal languages remains almost absent. The Supreme Court's draft Regulations for Use of Artificial Intelligence in Courts, 2026 contain strong principles of non-discrimination and inclusivity, but their closed audit and narrow remedy provisions leave communities unable to test whether those principles are honored.

Two contributions are offered. The first is the idea of the *localisation paradox*: localisation protects against the risk tribal communities are least likely to face and leaves untouched the risks they are most likely to face. The second is the *territorial gap*: Gram Sabha consent, the remedy most often proposed in existing writing, cannot reach tribal people outside Scheduled Areas or inside state institutions. The two-track model proposed in Section 7 is an attempt to close both gaps without abandoning localisation. This is a doctrinal and exploratory study, and it has limits. It relies in part on journalistic and civil society evidence, and it does not measure how often AI-driven decisions harm tribal people. Future research could use Right to Information requests, court records and field interviews to build that evidence, for example by tracking translation of orders for Adivasi undertrials or auditing welfare exclusions by tribe. What can be said now is that the window for building safeguards is open. The DPDP Rules are still being phased in, and the court AI Regulations remain in draft. If tribal communities are to have a voice in how their data is used, this is the time to give it to them.

© 2026 JAIET All Rights Reserved

CC BY 4.0 License

Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 11, pp 1-10, November-2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.311001
Available Online at https://www.jaiet.com

REFERENCES

[1] Thomson Reuters Foundation. (2017, October).
[2] 101Reporters. (n.d.). Online or offline, accessing forest rights is an insurmountable task for tribals in Madhya Pradesh. https://101reporters.com/article/development/Online_or_offline_accessing_forest_rights_is_an_insurmountable_task_for_tribals_in_Madhya_Pradesh
[3] Amnesty International. (2024, April). *Entity resolution in India's welfare digitalization*. https://www.amnesty.org/en/latest/research/2024/04/entity-resolution-in-indias-welfare-digitalization/
[4] AZB & Partners. (2025, November 14). *India's Digital Personal Data Protection Act: Phased rollout and key compliance milestones*. https://www.azbpartners.com/bank/indias-digital-personal-data-protection-act-phased-rollout-and-key-compliance-milestones/
[5] Bagaicha. (2016). Impoverished Adivasis hunted as criminals: A study of Jharkhandis alleged as Naxalites. *Bagaicha*. https://www.adivasi-koordination.de/wpdev/wp-content/uploads/2023/01/bagaicha_study-of-adivasi_Undertrials_2016.pdf
[6] Bokil, A., Khare, A., Sonavane, N., Bej, S., & Janarthanan, V. (2021). Settled habits, new tricks: Casteist policing meets big tech in India. *In State of power 2021. Transnational Institute*. https://www.tni.org/en/article/settled-habits-new-tricks
[7] Burman, A., & Sharma, U. (2021). How would data localization benefit India? *Carnegie Endowment for International Peace*. https://carnegieendowment.org/research/2021/04/how-would-data-localization-benefit-india
[8] Business Day. (2023, April 3). *Indigenous groups in New Zealand and US take issue with some AI*. https://www.businessday.co.za/bd/world/2023-04-03-indigenous-groups-in-new-zealand-and-us-take-issue-with-some-ai/
[9] Business Today. (2023, October 30). *81.5 crore Indians' personal data leaked, claims hacker*. https://www.businesstoday.in/technology/story/815-crore-indians-personal-data-leaked-claims-hacker-403861-2023-10-30
[10] Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE principles for Indigenous data governance. *Data Science Journal*, 19(1), 43. https://doi.org/10.5334/dsj-2020-043
[11] Centre for Research and Planning, Supreme Court of India. (2025). *White paper on artificial intelligence and judiciary. Supreme Court of India*. https://cdnbbsr.s3waas.gov.in/s3ec0490f1f4972d133619a60c30f3559e/uploads/2025/11/2025112244.pdf
[12] Committee of Experts under the Chairmanship of Justice B. N. Srikrishna. (2018). A free and fair digital economy: Protecting privacy, empowering Indians. *Ministry of Electronics and Information Technology, Government of India*.
[13] Designing human-centric AI with tribal communities: A shared governance framework for power, participation, and data sovereignty in India. (2026). *AI & Society*. https://doi.org/10.1007/s00146-026-03339-5
[14] Digital Personal Data Protection Act, 2023, No. 22, *Acts of Parliament*, 2023 (India).
[15] Digital Personal Data Protection Rules, 2025 (India).
[16] Dogra, B. (2014). Of trials and undertrials. *Frontier*, 47(21). https://frontierweekly.com/archive/vol-number/vol/vol-47-2014-15/47-21/47-21-Of%20Trials%20and%20Undertrials.html
[17] Dutta, S., & Mazumdar, S. (2025). Our data, ourselves: Participation, justice, and alternative futures of data sovereignty in India. *Big Data & Society*, 12(3). https://doi.org/10.1177/20539517251365235
[18] *Free Press Journal*. (2025, September). Govt launches beta version of Adi Vaani, India's first AI translator for tribal languages. https://www.freepressjournal.in/business/govt-launches-beta-version-of-adi-vaani-indias-first-ai-translator-for-tribal-languages
[19] From territorial to digital control: Tribal self-governance under the Indian constitutional framework. (2025). *Indian Journal of Law and Legal Research*, 8(2). https://www.ijllr.com/post/from-territorial-to-digital-control-tribal-self-governance-under-the-indian-constitutional-framewor
[20] K. S. Puttaswamy v. *Union of India*, (2017) 10 SCC 1 (India).
[21] K. S. Puttaswamy v. *Union of India*, (2019) 1 SCC 1 (India).
[22] Khera, R. (Ed.). (2019). Dissent on Aadhaar: Big data meets big brother. *Orient BlackSwan*.
[23] Kukutai, T., & Taylor, J. (Eds.). (2016). Indigenous data sovereignty: Toward an agenda. *ANU Press*.
[24] Masiero, S., & Das, S. (2019). Datafying anti-poverty programmes: Implications for data justice. *Information, Communication & Society*, 22(7), 916–933. https://doi.org/10.1080/1369118X.2019.1575448
[25] MediaNama. (2026, March). *RTI filed: Which ministries are part of MeitY's AI curation units, and how is "non-personal" data being defined?*

Journal of Artificial Intelligence and Emerging Technologies
Volume 3, Issue 11, pp 1-10, November-2026
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2026.311001
Available Online at https://www.jaiet.com

https://www.medianama.com/2026/03/223-rti-filed-ministries-part-meitys-ai-curation-units-non-personal-data-being-defined/

[26] Ministry of Justice (New Zealand). (2023). *Response to Official Information Act request* OIA-107110. https://www.justice.govt.nz/assets/Documents/Publications/OIA-107110_FINAL.pdf

[27] Ministry of Law and Justice. (2024, November 28). *Answer to Rajya Sabha unstarred question No*. 266 [Annexure I: Supreme Court vernacular judgments available on e-SCR portal]. Rajya Sabha. https://rsdebate.nic.in/bitstream/123456789/754106/1/PQ_266_28112024_U444_p313_p317.pdf

[28] National Crime Records Bureau. (2025). *Prison statistics India 2023. Ministry of Home Affairs, Government of India*. https://www.ncrb.gov.in/uploads/files/PSI-2023.pdf

[29] Newslaundry. (2024, July 15). *Why Gujarat's use of tech to verify forest rights claims is concerning*. https://www.newslaundry.com/2024/07/15/why-gujarats-use-of-tech-to-verify-forest-rights-claims-is-concerning

[30] NJCM *et al.* v. The State of the Netherlands, ECLI:NL:RBDHA:2020:865 (District Court of The Hague, Feb. 5, 2020).

[31] Panchayats (Extension to the Scheduled Areas) Act, 1996, No. 40, *Acts of Parliament, 1996 (India).*

[32] Perera, M., Vidanaarachchi, R., Chandrashekeran, S., Kennedy, M., Kennedy, B., & Halgamuge, S. (2025). Indigenous peoples and artificial intelligence: A systematic review and future directions. *Big Data & Society,* 12(2). https://doi.org/10.1177/20539517251349170

[33] Pooja Ramesh Singh v. Jammu and Kashmir Bank Ltd., 2026 *INSC 668 (India*).

[34] R v. Joe Mucheru & 2 others, ex parte Katiba Institute & another, [2021] KEHC 122 (KLR) (Kenya).

[35] Raj, I. (2026, August 6). Two faces of same algorithm: Algorithmic disparate impact on India's Adivasi communities. *LiveLaw.* https://www.livelaw.in/lawschool/articles/algorithmic-disparate-impact-india-adivasi-communities-544518

[36] Reserve Bank of India. (2018). *Storage of payment system data* (RBI/2017-18/153).

[37] Rohan Palaha v. State of Haryana, CRM-M-7673-2025 (*Punjab & Haryana High Court*, Feb. 25, 2025) (India).

[38] Royal Commission into the Robodebt Scheme. (2023). *Report. Commonwealth of Australia*.

[39] Saha, D. (2018). Did Aadhaar glitches cause half of 14 recent Jharkhand starvation deaths? *IndiaSpend*. https://sabrangindia.in/article/did-aadhaar-glitches-cause-half-14-recent-jharkhand-starvation-deaths

[40] Sambhav, K., & Tapasya. (2024, January 24). How an algorithm denied food to thousands of poor in India's Telangana. *Al Jazeera*. https://www.aljazeera.com/economy/2024/1/24/how-an-algorithm-denied-food-to-thousands-of-poor-in-indias-telangana

[41] Scheduled Tribes and Other Traditional Forest Dwellers (Recognition of Forest Rights) Act, 2006, No. 2, *Acts of Parliament*, 2007 (India).

[42] Sukanya Shantha v. *Union of India*, 2024 INSC 753 (India).

[43] Supreme Court of India. (2026). *Draft regulations for use of artificial intelligence (AI) in courts*, 2026. https://cdnbbsr.s3waas.gov.in/s3ec0490f1f4972d133619a60c30f3559e/uploads/2026/06/2026060342.pdf

[44] The Wire. (n.d.). Defending jal jangal jameen: National Adivasi convention exposes bureaucratic war on tribal rights. https://thewire.in/rights/defending-jal-jangal-jameen-national-adivasi-convention-exposes-bureaucratic-war-on-tribal-rights

[45] United Nations. (2007). *United Nations Declaration on the Rights of Indigenous Peoples* (A/RES/61/295).

[46] Vats, S. (2023). How MP govt's VanMitra portal threatens tribal rights over forest land. *IndiaSpend.* https://www.indiaspend.com/land-rights/how-mp-govts-vanmitra-portal-threatens-tribal-rights-over-forest-land-880848

Citation of this Article:

S.Sathia. (2026). Localised Data, Unlocalised Justice: Data Localisation, AI and Tribal Communities in India. *Journal of Artificial Intelligence and Emerging Technologies (JAIET).* 3(11), 1-10. Article DOI: https://doi.org/10.47001/JAIET/2026.311001

***** End of the Article *****