ROLE OF ARTIFICIAL INTELLIGENCE AND BIG DATA IN ENHANCING ACCESS TO TRADITIONAL KNOWLEDGE SYSTEMS

Dipak Kharole
Research Scholar
RTM Nagpur University, Nagpur

Dhananjay Gabhane
Librarian
Samarth Mahavidyalaya, Lakhani

Abstract :

Traditional knowledge systems (TKS) — local, indigenous, and community-held knowledge, practices, and beliefs — are critical for biodiversity conservation, healthcare, agriculture, cultural heritage, and long-term livelihoods. However, TKS must contend with issues like loss due to generational shifts, restricted accessibility, language hurdles, fragmentation, and misappropriation risk. This study looks at how Big Data and Artificial Intelligence (AI) technology can improve TKS accessibility, preservation, and ethical use. We examine pertinent technologies (such as distributed ledgers, knowledge graphs, computer vision, and natural language processing), give examples of applications, talk about moral, legal, and cultural issues, and suggest a technology framework that respects people's rights and is focused on the community. In order to guarantee that AI and Big Data enhance rather than compromise the independence and integrity of conventional knowledge holders, we conclude with research and policy proposals.

Keyword : traditional knowledge systems, indigenous knowledge, artificial intelligence, big data, knowledge graphs, NLP, ethics, biodiversity, cultural heritage

---

Introduction :

Traditional knowledge systems (TKS) are centuries of accumulated learning, practices, and cultural manifestations created by indigenous and local populations. They are holistic in nature, frequently combining spiritual beliefs, agricultural methods, medical knowledge, and ecological observations into a single, cohesive worldview. TKS is usually passed down orally through apprenticeship, storytelling, rituals, and lived practice, as contrast to official scientific information, which is written and codified in standardised formats. Despite being rich and flexible, this mechanism of transfer leaves such knowledge susceptible to extinction due to factors like climatic change, cultural homogenisation, rapid globalisation, and dwindling intergenerational transmission.

At the same time, the importance of old knowledge in tackling modern issues is becoming increasingly acknowledged on a global scale. For instance, local ecological knowledge promotes biodiversity conservation and climate adaption methods; traditional medicine aids in pharmaceutical discoveries and healthcare solutions; and indigenous agricultural traditions offer sustainable alternatives to industrial farming. The Convention on

Biological Diversity (CBD), the World Intellectual Property Organisation (WIPO), and UNESCO all emphasise how important TKS is to maintaining cultural diversity and advancing sustainable development. However, language diversity, a lack of documentation, and the lack of digital infrastructure in many communities continue to restrict access to and preservation of TKS despite its importance. The advent of **Artificial Intelligence (AI)** and **Big Data technologies** presents a transformative opportunity to address these challenges. AI encompasses computational methods that enable machines to perform tasks such as speech recognition, natural language processing, computer vision, and pattern recognition — all of which are highly relevant for digitizing and analyzing traditional knowledge. For instance, AI-driven speech-to-text systems can transcribe oral histories, while machine translation models can render indigenous languages into globally accessible formats without erasing their cultural nuances. Similarly, Big Data platforms allow for the storage, management, and analysis of vast, heterogeneous datasets, enabling researchers and communities to link knowledge across regions, languages, and disciplines. There are advantages and disadvantages to incorporating AI and Big Data into the preservation and distribution of TKS. Positively, by improving discoverability, removing language barriers, and offering decision-support tools for sustainability and community development, these technologies can democratise access. In order to help younger generations stay linked to their cultural history while adjusting to contemporary situations, they can also aid in the transfer of knowledge across generations. Misappropriation, loss of community control, intellectual property rights, and the possible decontextualisation of delicate cultural resources are urgent issues, nevertheless.

This study makes the case that AI and big data can significantly improve access to TKS if they are used in a framework of community governance and ethical responsibility. The study aims to show how digital innovation might promote the responsible use and preservation of traditional knowledge by examining pertinent technologies, application fields, ethical issues, and case studies. Crucially, the objective is to empower communities and guarantee that their knowledge contributes to global well-being on their own terms, not to replace or commercialise TKS.

Objectives and Scope :

This paper aims to:

1. Map AI and Big Data technologies relevant to documenting and enhancing access to TKS.
2. Describe concrete application patterns and illustrative case examples.
3. Identify main risks, ethical issues, and legal constraints.
4. Propose a community-centred technical and governance framework for responsible use.
5. Outline directions for future research and policy.

Scope :

We focus on digitally mediated access (discovery, retrieval, translation, visualisation, and provenance) and do not go into detail about commercialisation. Applications that promote benefit-sharing and community agency are highlighted.

Literature Review :

The importance of community archives, metadata standards, and digitisation has been emphasised in knowledge preservation research. In order to extract insights from a variety of sources, big data research highlights the significance of scale, heterogeneity, and analytics. Tasks like transcription, semantic search, and image-based plant identification are made possible by the maturation of AI research (NLP, computer vision, and knowledge representation) to handle various languages, lower-resource environments, and multimodal inputs.

Data sovereignty, Indigenous data governance principles (such as CARE— Collective benefit, Authority to control, Responsibility, Ethics), and the necessity of participatory design to prevent extractive activities are other political aspects that scholars and practitioners emphasise. This study combines these governance viewpoints with technical literature to suggest workable solutions.

Technologies and Methods Relevant to TKS :

1. Big Data platforms and practices :

- **Data lakes and federated repositories :** store heterogeneous digital objects (audio, video, text, geospatial data).
- **Metadata standards & ontologies :** essential for interoperability (e.g., Dublin Core, domain-specific ontologies).
- **Federated search & APIs :** enable queries across distributed community repositories without centralizing control.

2. Artificial Intelligence techniques :

- **Natural Language Processing (NLP) :**
  - Automatic transcription of oral narratives (speech-to-text, including low-resource language models).
  - Machine translation and cross-lingual semantic mapping.
  - Named-entity recognition and topic modelling to index concepts, species, practices, locations.
- **Computer Vision :**

- Visual identification of plants, artifacts, tools, and techniques from images and video (fine-grained classification).
- Optical character recognition (OCR) for scanned manuscripts and inscriptions.

- **Knowledge Graphs & Semantic Web :**
  - Structured representation of entities (people, practices, species, places) and relationships — supports semantic search and reasoning.
- **Multimodal AI :**
  - Linking audio, text, images, and geospatial data for richer context and retrieval.
- **Provenance, Blockchain & Distributed Ledger Technologies (DLT) :**
  - Immutable audit trails for origin, access, consents, and benefit-sharing agreements.
- **Privacy-enhancing technologies (PETs) :**
  - Differential privacy, homomorphic encryption, and secure multiparty computation to allow analytics without exposing sensitive raw data.
- **Explainability & Human-in-the-loop systems :**
  - Systems that provide transparent outputs and enable community curation and correction.

Application Patterns and Illustrative Examples :

1. Documentation & archiving :

- **Automated transcription pipelines** convert oral histories into searchable text; combined with NLP, these transcripts can be indexed by themes and local terms.
- **Image-based catalogs** use computer vision to index artifacts, crafts, and plant species.

2. Discovery & semantic search :

- **Knowledge graphs** permit queries like: "Which communities have traditional methods for pest control for maize?" — returning contextualized results with linked sources and permissions.

3. Translation & accessibility :

- **Cross-lingual models** enable translations between local languages and national/global lingua francas, increasing access by researchers and younger

- community members.

4. Decision support & sustainable practice dissemination :

- **AI-driven recommender systems** suggest indigenous crop varieties or water-harvesting techniques to communities facing climate stress, with provenance and suitability filters.

5. Rights management & provenance :

- **DLT-backed ledgers** record consent events, access logs, and benefit-sharing agreements to deter misappropriation and enable transparency.

Ethical, Legal, and Social Considerations :

1. Data sovereignty and consent :

TKS are not neutral data; they belong to living communities. Projects must implement Free, Prior and Informed Consent (FPIC), enabling communities to define access levels, permitted uses, and benefit-sharing.

2. Cultural sensitivity and contextual integrity :

Automated systems risk decontextualizing rituals or sacred knowledge. Metadata must capture cultural restrictions (e.g., genders, seasons, ceremonial access) and systems must enforce them.

3. Misappropriation and IP issues :

Digitization may inadvertently enable bioprospecting and cultural exploitation. Legal frameworks (e.g., Nagoya Protocol) and technological safeguards (access controls, provenance) are necessary but not sufficient.

4. Biases and epistemic injustice :

AI models trained on external corpora may misinterpret local naming, taxonomies, or practices. Co-design and model fine-tuning with community-curated data reduce these harms.

5. Equity of access :

Technological solutions must address digital divides (connectivity, literacy, language) so benefits reach intended communities.

Proposed Community-Centred Framework :

Principles :

1. **Community agency :** communities set goals, control data, and govern access.

2. **Contextual metadata first :** record rights, restrictions, provenance, and meanings.
3. **Interoperability & federation :** enable search and collaboration while avoiding forced centralization.
4. **Privacy and consent by design :** implement PETs and enforceable access policies.
5. **Transparency & explainability :** AI outputs should be understandable and correctable by custodians.
6. **Benefit-sharing :** economic, educational, and cultural benefits flow back to knowledge holders.

Technical architecture (high-level) :

1. **Local Node (community repository) :** stores raw objects (audio, video, images) and local metadata; offline-first capabilities.
2. **Federated Index & Knowledge Graph :** pointers and semantically rich metadata published with controlled visibility; supports discovery without exposing sensitive content.
3. **AI Services Layer :** modular services (ASR, OCR, vision, translation) deployed either locally or as privacy-respecting cloud functions; human-in-the-loop validation is mandatory before publication.
4. **Governance & Consent Layer :** records consents, access policies, and benefit-sharing terms (optionally anchored on DLT).
5. **Interface Layer :** multilingual search portals, mobile apps, and APIs tailored to both community users and authorized researchers.

Methodological Considerations for Researchers and Practitioners :

- **Participatory design :** involve elders, knowledge custodians, and youth in system design and evaluation.
- **Mixed methods evaluation :** combine technical metrics (e.g., transcription accuracy) with social measures (community satisfaction, sense of control).
- **Iterative deployment :** start with pilots, progressively scale with community feedback.
- **Model documentation :** publish model cards and data sheets explaining training data, limitations, and recommended uses.
- **Capacity building :** invest in local training for curation, metadata creation, and simple AI tool operation.

Conclusion :

AI and Big Data hold significant promise for enhancing access to traditional knowledge systems — improving documentation, discoverability, education, and sustainable use. However, technology alone cannot solve the ethical, legal, and cultural complexities. A rights-respecting, community-centred approach — combining appropriate technologies, strong governance, capacity building, and legal safeguards — is essential to ensure that AI amplifies rather than exploits traditional knowledge holders. When implemented collaboratively and transparently, AI can be a powerful ally in sustaining the world's rich tapestry of local knowledges for generations to come.

Reference :

- 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: Collective benefit, authority to control, responsibility, and ethics. *Data Science Journal, 19*(1), 43. https://doi.org/10.5334/dsj-2020-043
- Convention on Biological Diversity. (2011). *Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization.* United Nations. https://www.cbd.int/abs/
- Das, A., & Dayal, U. (2016). Big data analytics: A primer. *Communications of the ACM, 59*(11), 86–94. https://doi.org/10.1145/3003667
- Government of India. (2001). *Traditional Knowledge Digital Library (TKDL): An Indian initiative.* Council of Scientific and Industrial Research. http://www.tkdl.res.in
- Khurana, S. (2025). Decolonizing artificial intelligence: Indigenous knowledge systems, epistemic pluralism, and the ethics of technology. *Journal of Computer Allied Intelligence, 1*(1), 1–14. https://doi.org/10.69996/jcai.2025013
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think.* Houghton Mifflin Harcourt.
- Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology, 33*(4), 659–684. https://doi.org/10.1007/s13347-020-00405-8
- Pinhanez, C., Cavalin, P., Storto, L., Finbow, T., Cobbinah, A., Nogima, J., Vasconcelos, M., Domingues, P., Mizukami, P. de S., Grell, N., Gongora, M., & Gonçalves, I. (2024). Harnessing the power of artificial intelligence to vitalize endangered Indigenous languages: Technologies and experiences. *arXiv preprint*. https://arxiv.org/abs/2407.12620
- oberts, J. S., & Montoya, L. N. (2023). In consideration of Indigenous data sovereignty: Data mining as a colonial practice. *arXiv preprint*. https://arxiv.org/abs/2309.10215