BRIDGING MANUSCRIPTS AND MACHINE LEARNING: EMBEDDING INDIAN KNOWLEDGE SYSTEMS IN DIGITAL LIS EDUCATION

Supriya Abhijit Bejalwar
Librarian
Prof Ram Meghe College of Engineering & Management, Badnera-Amravati.
Email id: prmceamlibrary@gmail.com

Abstract :

Indian Knowledge Systems (IKS), embodied in manuscripts, oral traditions, and community practices, are fragile and under-documented, facing risks of loss and inaccessibility. At a global level, Indigenous Knowledge Systems face similar challenges, underscoring the need for sustainable preservation strategies. This paper proposes a dual framework for Library and Information Science (LIS) education: (i) an AI/ML-driven process model for digitizing and disseminating IKS, and (ii) a curricular framework aligned with India's National Education Policy (NEP) 2020. Drawing on literature, case studies, and expert consultations, the study highlights both technological opportunities and ethical imperatives. The findings demonstrate how LIS can evolve into a transformative discipline that bridges tradition and technology, cultivating professionals who are technologically skilled, culturally rooted, and globally competitive.

Keywords: Indian Knowledge Systems, Indigenous Knowledge Systems, Library and Information Science Education, Artificial Intelligence, Machine Learning, Digital Preservation, NEP 2020, Cultural Informatics

---

Introduction & Rationale :

This study adopts the term **Indian Knowledge Systems (IKS)** in line with the National Education Policy (NEP) 2020. While IKS is India-specific, it is situated within the broader global discourse on **Indigenous Knowledge Systems**, enabling comparative analysis with international heritage projects. Indigenous Knowledge Systems (IKS) embody centuries of accumulated wisdom expressed through manuscripts, oral traditions, and community practices. In India, these traditions span diverse fields such as Ayurveda, astronomy, agriculture, linguistics, and philosophy. Despite their richness, much of this heritage remains fragile and under-documented, with palm-leaf manuscripts and oral traditions at particular risk of deterioration or disappearance (Sen, 2005; Mishra, 2018). While initiatives like the National Mission for Manuscripts (NMM) and the Indira Gandhi National Centre for the Arts (IGNCA) have advanced digitization, vast repositories remain inaccessible and vulnerable (NMM, 2019).At the same time, Artificial Intelligence (AI) and Machine Learning (ML) offer transformative possibilities for heritage preservation. OCR for Indic scripts, NLP for

multilingual access, and computer vision for manuscript restoration are redefining how fragile resources are made accessible (Goyal & Jha, 2020; Zhou & Wu, 2023). International initiatives such as Digital Dunhuang and Europeana demonstrate how AI/ML can scale digitization and enhance discovery (Europeana Foundation, 2020; Wong & Zhao, 2020). In India, however, integration of these technologies remains limited, hindered by resource gaps, script diversity, and insufficient expertise (Singh & Gupta, 2023).

LIS professionals are well-positioned to bridge tradition and technology, yet Indian LIS curricula remain rooted in bibliographic control and ICT basics, with limited engagement in digital humanities, cultural informatics, or AI/ML (Satija, 2021; Chakraborty & Banerjee, 2024). This gap contradicts NEP 2020's vision of interdisciplinarity, inclusivity, and integration of Indian knowledge traditions (Kumar, 2021; Chatterjee, 2022).

Objectives :

- To examine how AI/ML technologies can support digitization and preservation of manuscripts, oral traditions, and IKS.
- To propose a curricular framework that integrates AI/ML, digital humanities, and indigenous metadata into LIS education.
- To demonstrate alignment with NEP 2020's vision of interdisciplinarity, inclusivity, and holistic learning.

Methodology :

This study employs a **conceptual and framework-building approach**, synthesizing literature, mapping best practices, and proposing new models.

Stages :

Literature Review : Establish theoretical foundations on IKS, AI/ML, LIS, and NEP 2020.

Situational Analysis : Comparative mapping of Indian and global digitization projects (NMM, IGNCA, Europeana, Digital Dunhuang).

Framework Development : Design of a process model for AI/ML-based IKS digitization and a curricular framework for LIS education.

Expert Validation : Refinement of frameworks via consultations with LIS faculty, technologists, and cultural custodians.

This study is positioned as a **conceptual paper**, rather than an empirical survey. While no primary quantitative data were collected, the emphasis is on synthesizing existing practices and proposing theoretically grounded frameworks, a recognized contribution in LIS research (Dasgupta, 2019; Joseph & Abraham, 2023).

Literature Review :

IKS transmit wisdom across generations through manuscripts and oral traditions but

]remain at risk of disappearance (Sen, 2005; Mishra, 2018). Indian efforts (NMM, IGNCA) have contributed, but accessibility remains limited (NMM, 2019). Globally, UNESCO (2015) emphasizes safeguarding fragile heritage, while IFLA (2022) stresses Indigenous Data Sovereignty and ethical governance.

AI/ML tools are transforming preservation: OCR for Indic scripts (Goyal & Jha, 2020), NLP for multilingual access (Reddy et al., 2021), and computer vision for restoration (Wong & Zhao, 2020). Projects like Digital Dunhuang and Europeana highlight advanced integration (Europeana Foundation, 2020). Yet challenges remain in India: dataset scarcity, script complexity, and resource constraints (Singh & Gupta, 2023; Zhou & Wu, 2023).

Indian LIS curricula emphasize bibliographic control and ICT but lag in cultural informatics and AI integration (Dasgupta, 2019; Satija, 2021; Ramesh & Nagar, 2022). Global programs (UCL, Michigan, Humboldt) embed digital heritage, computational archival studies, and AI applications, producing graduates who are globally competitive (Wong & Zhao, 2020). The gap between India and global LIS education is widening (Joseph & Abraham, 2023).

NEP 2020 calls for interdisciplinarity, multilingualism, and integration of India's cultural traditions into higher education (Kumar, 2021; Chatterjee, 2022). For LIS, this means embedding AI/ML skills, indigenous metadata, and ethical frameworks into curricula (Satija, 2021; Chakraborty & Banerjee, 2024).

The literature highlights three imperatives—preserve IKS, adopt AI/ML, reform LIS education—which collectively justify the proposed frameworks.

Focus Area	India (Current Trends)	Global (Europe/USA Examples)
Core Curriculum Orientation	Traditional focus on cataloguing, bibliographic control, ICT basics (Satija, 2021)	Balanced focus on library science,information management, and digital heritage (Wong & Zhao, 2020)
Digital Humanities & Cultural Informatics	Limited or absent; very few universities offer courses (Dasgupta, 2019)	Well-developed tracks in digital curation, archival informatics, cultural heritage studies
AI/ML Integration	Minimal inclusion; only isolated electives or research projects (Ramesh & Nagar, 2022)	Integrated modules on OCR, NLP, ML for text mining, digital archives (University College London, Humboldt University)
Metadata & Digital Preservation	Introduced in digital library modules, but not contextualized for IKS (Joseph & Abraham, 2023)	Emphasis on metadata standards (Dublin Core, RDF, TEI/XML) and open linked data (Europeana Foundation, 2020)

Community Engagement & Indigenous Knowledge	Rarely emphasized; community participation largely missing	Community-driven projects, participatory archives, indigenous metadata models (IFLA, 2022)
Interdisciplinarity	Emerging but fragmented; some pilot programs after NEP 2020	Strongly interdisciplinary; LIS intersects with computer science, anthropology, history
Funding & Infrastructure	Often underfunded; lack of AI labs, repositories, and skilled faculty	Substantial investments in labs, repositories, collaborative grants (EU Horizon, NSF)

1. Situational Analysis :

To contextualize how Artificial Intelligence (AI) and Machine Learning (ML) are being applied to preserve Indigenous and Indian Knowledge Systems (IKS), this section reviews selected national and international digitization initiatives. Rather than retelling technical details already covered in the Literature Review, the focus here is on **comparative insights**—how projects differ in scope, tools, and governance, and what lessons they offer for Library and Information Science (LIS) education in India.

Comparative Overview of Digitization Initiatives

Project/Initiative	AI/ML Tools/Techniques	Key Lessons Learned
National Mission for Manuscripts (India)	OCR for Indic scripts, metadata digitization	Need for scalable OCR models tailored to diverse Indian scripts
Indira Gandhi National Centre for the Arts (India)	Digitization + metadata creation; limited AI integration	Community collaboration and layered access models are essential
Digital Library of India (India)	Bulk scanning, OCR pipelines, metadata standardization	Scalability and interoperability require standardized workflows
Bhandarkar Oriental Research Institute (India)	High-resolution imaging, palaeographic metadata	Specialist expertise is critical for accuracy in cultural contexts
Muktabodha Digital Library (India)	Text encoding (TEI/XML), facsimile integration	Integration of encoding standards supports scholarly research
Digital Dunhuang (China)	Computer vision (3D imaging, multispectral enhancement), semantic indexing	Interdisciplinary teams and advanced AI can enrich heritage preservation
Europeana (Europe)	Ontology & knowledge graph construction, linked open data, multilingual NLP	Standardized metadata and open access foster large-scale reuse
Native American Digital Archives (USA)	Community-driven metadata, access control, participatory digitization	Ethics, indigenous governance, and cultural sensitivity must guide access

Cross-Cutting Lessons :

From this comparative mapping, several key insights emerge:

- **Interdisciplinary collaboration** (technologists, conservators, linguists, community custodians) is critical for sustainable digitization.
- **AI/ML applications** such as OCR, NLP, and computer vision are widely used globally but remain underutilized in India due to resource and skill constraints.
- **Metadata interoperability** (Dublin Core, RDF, TEI/XML) enables aggregation and reuse, yet Indian projects often rely on localized, non-standard schemas.
- **Ethics and governance** are central: community-led access models prevent cultural exploitation and build trust (IFLA, 2022).
- **Capacity building** in LIS curricula is essential so future professionals can manage and critically evaluate AI-driven digitization pipelines.

The situational analysis reveals that India possesses rich institutional experience in digitization, but its adoption of AI/ML is **uneven and constrained** by script diversity, data scarcity, and limited interdisciplinary expertise. In contrast, global initiatives demonstrate how **advanced AI/ML pipelines, standardized metadata, and community governance** can scale preservation efforts. For India, the path forward lies in leveraging global technical models while embedding **ethical, community-led practices and targeted capacity building** into LIS education.

2. Framework for AI/ML-based Digitization and Preservation of Indigenous Knowledge Systems (IKS)

Framework for AI/ML-based Digitization and Preservation of Indian Knowledge Systems (IKS)

Based on the literature and situational analysis, this study proposes a dual framework:

- A process model for AI/ML-driven digitization and preservation of IKS.
- A curricular framework to embed these practices in LIS education, aligned with NEP 2020.

1. AI/ML-Enabled Digitization Pipeline :

The digitization pipeline is designed as a sequential process, balancing technological innovation with cultural sensitivity. It consists of five interconnected components:

- **Collection & Documentation :** Identification and documentation of manuscripts, oral traditions, and folk practices in collaboration with cultural custodians and repositories.
- **Digitization & Pre-processing :** High-resolution scanning, audio-video capture, OCR for Indic scripts, Unicode/TEI-XML encoding, and metadata creation adapted to cultural contexts.

- **AI/ML Integration :** NLP for multilingual retrieval, semantic indexing, and translation. Computer Vision for restoration and fragment reconstruction.
- Ontology & Knowledge Graphs blending indigenous classification systems with linked open data.
- **Dissemination & Access :** Secure repositories, open-access portals, multilingual interfaces, and AI-powered discovery tools to maximize scholarly and community engagement.
- **Ethical Protocols :** Consent-driven documentation, community ownership of digitized materials, culturally appropriate access tiers, and benefit-sharing mechanisms.
- This model ensures that fragile IKS resources are not only digitized but also enriched with AI/ML capabilities, while remaining ethically governed.

2. LIS Curricular Framework :

The curricular framework integrates IKS and AI/ML into LIS education in alignment with NEP 2020's vision of interdisciplinarity, inclusivity, and skill-based learning.

Core Principles:

Interdisciplinarity : Bridging LIS with computer science, anthropology, and digital humanities.

Skill-Based & Experiential : Emphasis on labs, practicums, and community immersion**.**

Cultural Sensitivity : Embedding indigenous epistemologies, ethics, and community voices.

Proposed Modules (illustrative) :

Digital Humanities & IKS → Manuscript traditions, oral heritage documentation, TEI/XML encoding.

AI/ML Applications in Libraries → OCR, NLP, computer vision, machine translation for low-resource languages.

Indigenous Metadata & Knowledge Organization → Mapping indigenous classifications (Paninian grammar, Vedic divisions) to RDF/Dublin Core.

Ethics & Community Engagement → Intellectual property, consent, benefit-sharing, cultural protocols.

Digital Curation & Repository Management → Repository design, persistent identifiers, sustainability planning.

Pedagogy : Project-based learning, labs, hackathons, and community immersion.

Assessment : Continuous evaluation through projects, reflective journals, and practical demonstrations.

Infrastructure Needs : Digital humanities labs, OCR/NLP toolkits, curated datasets, and interdisciplinary faculty teams.

The digitization pipeline provides a technological roadmap for preserving IKS, while the curricular framework prepares LIS graduates to implement and sustain it. Together, they position LIS professionals as mediators of tradition and technology, ensuring that India's heritage is safeguarded while aligning with NEP 2020's educational vision.

3. Expert Validation :

To ensure the feasibility, cultural sensitivity, and curricular relevance of the proposed dual frameworks, an expert validation process was conducted. Expert validation is a widely recognized method for evaluating conceptual frameworks and curricular innovations in LIS and education research (Okoli & Pawlowski, 2004; Hasson, Keeney, & McKenna, 2000).

Method :

A purposive sampling strategy was adopted to identify experts with diverse backgrounds in LIS education, AI/ML applications in cultural heritage, and community custodianship. In total, seven experts were consulted:

Four LIS educators from central and state universities, Two technologists with expertise in manuscript digitization and AI/ML, One community archivist representing indigenous knowledge custodians.

The validation was carried out through semi-structured interviews and consultations (online and in-person) conducted between January–March 2024. Prior informed consent was obtained, and discussions were guided by themes of curricular adequacy, technical feasibility, and ethical safeguards.

Expert Feedback :

Key insights are summarized below:

Expert Type	Affiliation/Role	Mode of Consultation	Key Feedback
LIS Educator (1)	Central University	Online Interview	Highlighted curricular gaps; welcomed integration of AI/ML and IKS.
LIS Educator (2)	State University	Online Interview	Emphasized need for interdisciplinary teaching and digital humanities.

LIS Educator (3)	National Institute	Virtual Consultation	Supported alignment with NEP 2020; suggested phased implementation.
LIS Educator (4)	Private University	Online Consultation	Raised concern about faculty reskilling and infrastructure deficits.
Technologist (1)	Research Institute on AI & Heritage	In-person Meeting	Confirmed feasibility of OCR/NLP for Indic scripts but noted dataset gaps.
Technologist (2)	Cultural Informatics Start-up	Virtual Consultation	Stressed importance of curated labs and repositories for LIS training
Community Archivist	Regional Manuscript Repository	Field Visit	Advocated embedding community consent, cultural protocols, and access tiers.

Synthesis of Validation

Curricular Gaps : Five LIS faculty underscored that current syllabi remain focused on cataloguing and bibliographic control, with insufficient exposure to AI/ML and cultural informatics (Joseph & Abraham, 2023; Satija, 2021).

Technical Feasibility : Technologists affirmed that OCR, NLP, and computer vision tools are maturing for Indic scripts (Goyal & Jha, 2020; Singh & Gupta, 2023), but cautioned that usable datasets and sustained infrastructure are essential.

Ethical Safeguards : The archivist's input aligned with international discourses on Indigenous Data Sovereignty, highlighting the need for community-led governance in digitization (IFLA, 2022).

Consensus :

Experts broadly endorsed the frameworks as timely, innovative, and aligned with NEP 2020, while recommending:

- Faculty capacity-building through training and interdisciplinary collaboration (Virkus, 2021),
- Government and institutional investment in AI/ML labs and curated datasets, Stronger participatory mechanisms with community custodians.

This validation provides documented evidence that the proposed models are not only theoretically robust but also practically viable and ethically grounded.

Discussion :

The proposed frameworks strongly align with the principles of the National Education Policy (NEP) 2020, particularly its emphasis on interdisciplinarity, inclusivity, and holistic learning. However, embedding Indigenous Knowledge Systems (IKS) and Artificial Intelligence (AI)/Machine Learning (ML) into Library and Information Science (LIS) education is not without challenges.

Key Challenges :

- **Resistance from Traditional LIS Departments :** Many LIS programs in India remain rooted in conventional paradigms of cataloguing, classification, and bibliographic control. Faculty and administrators may view digital humanities or AI/ML as peripheral, leading to reluctance in revising syllabi (Satija, 2021). This inertia risks widening the gap between global LIS practices and Indian curricula.

- **Faculty Expertise and Training Deficits :** Most LIS educators are trained in traditional library science, with limited exposure to advanced computational tools, natural language processing, or cultural informatics. Without significant reskilling, faculty may struggle to teach AI/ML-based modules (Joseph & Abraham, 2023).

- **Funding and Infrastructure Constraints** : Establishing digital humanities labs, procuring high-resolution scanners, developing OCR/NLP datasets, and maintaining digital repositories require substantial investment. Given that LIS schools are often underfunded compared to STEM disciplines, sustained infrastructure development poses a real challenge (Ramesh & Nagar, 2022).

- **Community Participation and Ethical Complexities :** While NEP 2020 calls for inclusivity, operationalizing community consent, cultural protocols, and benefit-sharing in digitization projects is complex. There is a risk of tokenism if community custodians are not genuinely integrated into decision-making (IFLA, 2022).

A further concern raised by experts is the potential risk of **over-technologization**. Excessive reliance on AI/ML tools without adequate cultural contextualization may flatten the richness of indigenous epistemologies, reducing them to datasets rather than lived traditions. Addressing this requires embedding humanistic interpretation, linguistic expertise, and community oversight into every stage of the digitization pipeline.

Strategies for Overcoming Challenges :

- **Capacity Building for Faculty**: Intensive training programs, faculty development workshops, and interdisciplinary teaching collaborations with computer science and digital humanities departments can bridge the skills gap.
- **Policy-Level Mandates** : University Grants Commission (UGC) and professional bodies can institutionalize requirements for digital heritage and AI/ML modules in LIS curricula, ensuring consistency and compliance across institutions.
- **Funding and Partnerships** : Strategic partnerships with national initiatives such as the National Mission for Manuscripts, IGNCA, and global platforms like Europeana can secure technical expertise and shared resources. Additionally, government schemes under NEP 2020 could earmark funding for LIS technology infrastructure.
- **Community-Centric Models** : Embedding indigenous custodians as co-teachers, metadata contributors, and advisory partners ensures genuine participation, aligning with Indigenous Data Sovereignty principles (IFLA, 2022).
- **Phased Implementation** : Institutions may begin with pilot modules or electives, gradually scaling up to full curricular reform. This reduces institutional resistance and allows iterative refinement.

Alignment with NEP 2020 :

Despite these challenges, the frameworks are firmly aligned with NEP 2020's vision. They advance interdisciplinarity by blending LIS with computer science, anthropology, and cultural studies; they enhance inclusivity by foregrounding regional languages and indigenous traditions; and they embody holistic learning through project-based labs, community immersion, and capstone projects. Importantly, by cultivating technologically skilled yet culturally sensitive professionals, the frameworks contribute to India's aspiration for *Atmanirbhar Bharat* (self-reliant India) while also positioning Indian LIS education within global digital heritage networks.

Conclusion and Future Directions :

This study proposed a dual framework that integrates **AI/ML-driven digitization and preservation of Indigenous Knowledge Systems (IKS)** with a **curricular model for LIS education**, aligned to the transformative vision of the National Education Policy (NEP) 2020. The frameworks collectively address three pressing imperatives: the urgent need to safeguard fragile indigenous resources, the necessity of embedding advanced technological tools into LIS practices, and the responsibility to cultivate culturally sensitive, interdisciplinary, and future-ready information professionals.

The findings underscore that effective preservation of IKS requires not only digitization

and technological intervention but also robust **ethical safeguards** and **community participation**. Similarly, LIS education must evolve beyond conventional bibliographic control to encompass **digital humanities, AI/ML applications, indigenous metadata, and cultural informatics**, thereby producing graduates capable of bridging tradition and technology. The expert validation process further demonstrated the feasibility, relevance, and adaptability of the proposed frameworks in real-world contexts.

Future directions :

- **Pilot Implementations:** LIS schools and research institutes can adopt the proposed curricular modules on a trial basis, integrating digitization projects into student practicum and research.
- **Capacity Building:** Continuous training workshops for LIS educators, technologists, and community archivists will be critical to sustain and scale these initiatives.
- **Policy Advocacy:** Collaboration with policymakers and professional bodies can ensure that LIS curricular reforms are recognized within national educational standards and funding priorities.
- **Technological Advancements:** Future research should focus on developing AI/ML tools tailored to low-resource Indian languages and scripts, ensuring that technological innovation aligns with local needs.
- **Community-Centered Models:** Long-term sustainability will depend on strengthening mechanisms of community ownership, consent, and benefit-sharing in all digitization and dissemination activities.

In conclusion, the proposed models demonstrate that **LIS education, when reimagined through the lens of IKS and AI/ML, can serve as a catalyst for inclusive, interdisciplinary, and holistic learning**. By aligning with NEP 2020, this approach not only modernizes LIS pedagogy but also repositions libraries and information professionals as vital agents in the preservation of India's cultural heritage and its integration into global knowledge systems. Beyond the Indian context, this dual framework positions LIS education at the intersection of cultural heritage and AI/ML, contributing to global debates on heritage informatics, digital humanities, and sustainable knowledge preservation.

References :

- Chakraborty, A., & Banerjee, S. (2024). "Integrating AI and Digital Humanities in LIS Education: An Indian perspective." *Annals of Library and Information Studies, 71*(2).
- Chatterjee, S. (2022). NEP 2020 and the future of Indian higher education: A critical appraisal. Higher Education for the Future, 9(1), 1–15. https://doi.org/10.1177/23476311221075473
- Dasgupta, A. (2019). Digital curation and LIS education in India: Challenges and opportunities. Annals of Library and Information Studies, 66(3), 145–152.

- Europeana Foundation. (2020). Europeana strategy 2020–2025: Empowering digital heritage in Europe. The Hague: Europeana.
- Goyal, P., & Jha, A. (2020). OCR systems for Indic scripts: A review. Journal of Cultural Heritage, 45, 230–240. https://doi.org/10.1016/j.culher.2020.01.015
- IFLA (2022). *Indigenous Data Sovereignty and Libraries: Global perspectives.* IFLA Journal, 48(3).
- Kumar, R. (2021). Interdisciplinary learning and NEP 2020: Implications for social sciences. Indian Journal of Educational Studies, 8(2), 33–42.
- Mishra, S. (2018). Indigenous knowledge systems and libraries in India: A critical study. International Information & Library Review, 50(4), 287–297. https://doi.org/10.1080/10572317.2018.1536041
- National Mission for Manuscripts. (2019). Annual report 2018–2019. New Delhi: Ministry of Culture, Government of India.
- Reddy, M., Sharma, P., & Singh, R. (2021). Natural language processing for ancient Indian texts: Progress and challenges. International Journal of Information Processing, 15(2), 67–79.
- Satija, M. P. (2021). LIS education in India: Present scenario and future directions. DESIDOC Journal of Library & Information Technology, 41(2), 85–92. https://doi.org/10.14429/djlit.41.2.16335
- Singh, R., & Gupta, A. (2023). "Digitizing Sanskrit manuscripts using deep learning: A case study." *Library Hi Tech, 41*(4), 955–972. https://doi.org/10.1108/LHT-02-2023-0058
- Sen, S. (2005). Indigenous knowledge and libraries: Preserving oral traditions in India. New Delhi: Ess Ess Publications.
- UNESCO. (2015). Memory of the World: General guidelines to safeguard documentary heritage. Paris: UNESCO.
- Wong, W., & Zhao, L. (2020). AI in cultural heritage preservation: A review of applications and challenges. Journal of Documentation, 76(6), 1349–1368. https://doi.org/10.1108/JD-12-2019-0231
- Zhou, H., & Wu, Y. (2023). "AI-driven cultural heritage preservation: Opportunities and challenges." *Information Processing & Management, 60*(2), 103220. https://doi.org/10.1016/j.ipm.2023.103220