मराठी लोकसाहित्याचे डिजिटल दस्तऐवजीकरण आणि AI : संकलन, वर्गीकरण व अभ्यास

प्रा. डॉ. अशोक भारत भक्ते
प्राध्यापक व मराठी विभागप्रमुख,
नगरपरिषद शिवाजी महाविद्यालय, मोवाड, ता. नरखेड,
जि. नागपूर

Crossref DOI - https://doi.org/10.63665/rh.v7i2.135

सारांश :

लोकसाहित्य म्हणजे एका पिढीकडून दुसऱ्या पिढीकडे तोंडीरूपात संक्रमित झालेलं, एखाद्या समाजाच्या सांस्कृतिक आठवणींचं, रीती-रिवाजांचं आणि एकत्रित जगण्यातून आलेल्या अनुभवांचं चैतन्यमय जतन! मराठी लोकसाहित्याच्या समृद्ध परंपरेत ओव्या, भारुडं, पोवाडे, लोकगीतं, लोककथा, म्हणी, वाक्प्रचार, व्रतकथा, सोहळ्यांची गाणी आणि वेगवेगळ्या प्रादेशिक बोलींमधील मौखिक साहित्याचा खजिना सामावलेला आहे. पण दुर्दैवाने, आजच्या धावपळीच्या युगात शहरीकरण, माध्यमांचं बदलतं स्वरूप आणि जीवनशैलीतील तीव्र बदलांमुळे लोकसाहित्याची ही तोंडी परंपरा वेगाने कमी होत चालली आहे. त्यामुळे अनेक लोककथा, पारंपरिक गाण्यांचे प्रकार, म्हणी आणि प्रादेशिक भाषेची खास ओळख हळूहळू इतिहास जमा होण्याची भीती निर्माण झाली आहे.

या सगळ्या परिस्थितीत, लोकसाहित्याचं डिजिटल जतन आजच्या युगाची गरज आहे. डिजिटल तंत्रज्ञानामुळे लोकसाहित्याचं व्यवस्थित संकलन, जतन, एकत्रिकरण, लिखाण आणि पुन्हा वाचणं सोपं होतं. आता यात Artificial Intelligence (AI) आणि नैसर्गिक भाषा प्रक्रिया (Natural Language Processing-NLP) सारखी साधने मोलाची भर घालू शकतात. AI च्या मदतीने लोकसाहित्याचा मोठा डेटा सांभाळता येतो, त्याचं वर्गीकरण करता येतं, त्यात काय आहे ते शोधता येतं, भाषेची वैशिष्ट्ये ओळखता येतात आणि महत्वाचं म्हणजे, डिजिटल लायब्ररी तयार करता येते, जी सहज उपलब्ध होईल.

मराठी लोकसाहित्याला डिजिटल रूप देण्याच्या प्रवासात कृत्रिम बुद्धिमत्ता (AI) कशी मदत करते, डेटा कसा जमा केला जातो आणि त्याचे वर्गीकरण कसे केले जाते, तसेच लोकसाहित्याच्या अभ्यासात AI च्या मदतीने काय काय नवीन गोष्टी करता येतात आणि त्याच्या मर्यादा काय आहेत, याचा कसून शोध या शोधनिबंधात घेतला आहे. या अभ्यासात हे स्पष्ट होते की, AI लोकसाहित्याचे जतन करण्यासाठी आणि त्यावर संशोधन करण्यासाठी एक भक्कम आधार आहे. पण, लोकसाहित्याची सांस्कृतिक पार्श्वभूमी, भाषेतील खास लकबी आणि ते सादर करण्याच्या पद्धती यांसारख्या गोष्टी समजून घेण्यासाठी माणसाची मदत गरजेची आहे. त्यामुळे लोकसाहित्य संशोधनात AI चा वापर विचारपूर्वक, नीतिमत्तेच्या आधारावर आणि शास्त्रीय पद्धतीने

करायला हवा.

बीजशब्द : मराठी लोकसाहित्य, डिजिटल दस्तऐवजीकरण, कृत्रिम बुद्धिमत्ता (AI), लिप्यंतरण, डिजिटल तंत्रज्ञान, मेटाडेटा, मशीन लर्निंग आणि नैसर्गिक भाषा प्रक्रिया

प्रस्तावना :

महाराष्ट्राच्या सांस्कृतिक परंपरेतील तोंडी आणि सामुदायिक आविष्कार म्हणजे मराठी लोकसाहित्य! या साहित्य प्रकाराची खासियत म्हणजे हे लिखाणातून नाही, तर पिढ्यान्पिढ्या बोलण्यातून, गाण्यातून, सादरीकरणातून, लोकांच्या दैनंदिन जीवनातून पुढे सरकतं. म्हणूनच, लोकसाहित्य नुसतं एक साहित्य न राहता, समाजाच्या जगण्याचा महत्त्वाचा भाग बनून जातं. लोककथा असोत, ओव्या, भारूड, पोवाडे असोत, म्हणी, वाक्प्रचार, व्रतकथा असोत किंवा लोकनाट्य; यातून लोकसाहित्य लोकांच्या मनातले विचार, श्रद्धा, त्यांची मूल्यं आणि समाजातील सत्य परिस्थिती उघड करतं.

पण २१ व्या शतकात समाजात जे बदल झाले, त्यामुळे लोकसाहित्याच्या अस्तित्वावर संकट आलं आहे. शहरांची वाढ, औद्योगिकीकरण, लोकांचे स्थलांतर, शिक्षणाच्या नव्या पद्धती, आणि डिजिटल साधनांचा वाढता वापर यामुळे लोकांचे पारंपरिक जीवन बदलून गेले आहे. अनेक लोकगीते फक्त सण-उत्सवांपुरतीच उरली आहेत, तर काही परंपरा पूर्णपणे बंद व्हायच्या मार्गावर आहेत. त्यात लोकसाहित्य जपणारे कथाकार, गायक, भारूडकार, पोवाडे सांगणारे कलाकार कमी झाले आहेत. त्यामुळे लोकसाहित्याचा खूप मोठा खजिना नाहीसा होण्याचा धोका आहे.

अशा परिस्थितीत, लोकसाहित्याचं डिजिटल जतन करणं म्हणजे ते सुरक्षित ठेवण्याचा महत्त्वाचा उपाय आहे. डिजिटल जतन म्हणजे काय, तर लोकसाहित्याचे ऑडिओ-व्हिडिओ रेकॉर्डिंग, त्याचं लिप्यंतरण, Text तयार करणं, ते व्यवस्थित संग्रहित करणं आणि डिजटल संग्रह तयार करणं, ज्यामुळे ते सहज शोधता येईल. यात AI (Artificial Intelligence) आणि NLP (Natural Language Processing) सारखी आधुनिक Tools (साधने) वापरली, तर लोकसाहित्याच्या अभ्यासाला आणखी गती मिळेल, ते जास्त लोकांपर्यंत पोहोचेल आणि शास्त्रशुद्ध पद्धतीने करता येईल. AI based Tools (एआय आधारित साधने) खूप मोठ्या प्रमाणात Data (डेटा) हाताळू शकतात, त्यातले Patterns (नमुने) शोधू शकतात, भाषेची वैशिष्ट्ये ओळखू शकतात आणि वर्गीकरण सोपे करू शकतात. त्यामुळे

'मराठी लोकसाहित्याचे Digital जतन आणि AI' हा विषय फक्त तांत्रिक नाही, तर तो साहित्यशास्त्र, समाजशास्त्र आणि संस्कृतीचा अभ्यास करणाऱ्यांसाठी खूप महत्त्वाचा आहे.

लोकसाहित्य : संकल्पना व स्वरूप :

लोकसाहित्य म्हणजे असं साहित्य, जे खेड्यापाड्यातल्या आणि सामान्य माणसांच्या मनात जन्म घेतं.

त्यांच्या रोजच्या जीवनातले सुख-दुःख, त्यांची मेहनत, सण-उत्सव, श्रद्धा आणि काहीवेळा अंधश्रद्धा, त्यांची नीतिमत्ता, समाजाची रचना, त्यांचे झगडे, प्रेम, भक्ती, पराक्रम अशा अनेक गोष्टी अगदी सहजपणे या साहित्यातून बाहेर येतात.

आर. डी. जेम्सन यांच्या नुसार, "लोकसाहित्याच्या सामग्रीमध्ये मिथ्स्, लोककथा, परंपरा, लोकविश्वास, लोकभ्रम, विधी, रूढी, लोकनृत्य यांचा अंतर्भाव होतो. मानवी आकांक्षा आणि मनातील भीतीची अभिव्यक्ती लोकसाहित्यातून होते."[१]

लोकसाहित्य हा 'लोकांचा साहित्यप्रकार' आहे. म्हणजेच जो साहित्यप्रकार ग्रंथरूपात लिहिला गेलेला नसतो, पण तो पिढ्यानपिढ्या तोंडी परंपरेने चालत आलेला असतो, त्याला लोकसाहित्य म्हणतात. लोकजीवनातील अनुभव, श्रद्धा, भावना, संस्कृती, परंपरा, रूढी, आचार-विचार, आनंद-दुःख, संघर्ष, निसर्ग, श्रमजीवन इत्यादींचे प्रतिबिंब लोकसाहित्यात दिसते. "लोकसाहित्यात लोकप्रतिभेचा आविष्कार असतो. तसेच लोकांच्या पारंपरिक जीवनाचा संदर्भ दडलेला असतो. लोकसाहित्य हे लोकमानसाचा आविष्कार असते, त्यातून समाजमानस अभिव्यक्त होते."[२] अशी व्याख्या डॉ. प्रभाकर मांडे यांनी केली आहे.

लोकसाहित्याची मुख्य रूपे :

लोककथा, लोकगीते (ओव्या, लावणी, भजने, अंगाई, श्रमगीते), पोवाडे, म्हणी व वाक्प्रचार, उखाणे, कोडी, लोकनाट्य (तमाशा, वगनाट्य, दशावतार इ.), लोकश्रद्धा व समजुती इ. मुख्य रुपे दिसून येतात.

लोकसाहित्याचे स्वरूप :

"मराठीतील लोकसाहित्याच्या अभ्यासाचा प्रारंभ १९ व्या शतकात झाला. मेरी फ्रियर नावाच्या स्त्रीने १८५८ साली 'Old Deccan Days' या नावाने पुस्तक प्रकाशित केले. या पुस्तकात लोककथा आणि लोकसाहित्याचे संकलन केले आहे."[३] याच अनुषंगाने स्वरूप स्पष्ट करतांना असे लक्षात येते की, लोकसाहित्य लोकांच्या जीवनाशी घट्ट बांधलेले असते. हे साहित्य कुणा एका माणसाने घडवलेले नसून, ते अख्ख्या गावाने, समाजाने मिळून बनवलेले असते. त्यामुळे यात लोकांच्या मनातले भाव आणि त्यांचे अनुभव अगदी स्पष्टपणे दिसतात. हे साहित्य पुस्तकात नाही, तर तोंडी परंपरेतून पिढ्यान्पिढ्या चालत आले आहे. त्यामुळे छापील पुस्तकांसारखं ते नसतं, त्यात सतत बदल होत राहतात. काळ बदलला, प्रदेश बदलला, गोष्ट सांगणारा माणूस बदलला की तेही बदलतं. लोक साहित्याची भाषा एकदम सोपी, सरळ आणि आपल्या बोलीभाषेसारखी असते, ज्यामुळे ते सगळ्या माणसांना चटकन समजतं. यात लोकांची संस्कृती, त्यांचा देवधर्म, रीती-रिवाज, सण-उत्सव, त्यांची मेहनत, निसर्ग आणि रोजच्या जगण्यातले व्यवहार जसेच्या तसे रेखाटलेले असतात. मनोरंजनासोबतच हे आपल्याला जगण्याची शिकवण देते, समाजाची जाणीव करून देते आणि आपल्या संस्कृतीला जपण्याचे काम करते. लोकसाहित्य वेगवेगळ्या रूपात आपल्यासमोर येतं- जसं की लोककथा, लोकगीते, म्हणी, उखाणे, कोडी, पोवाडे, लोकनाट्ये. यामुळेच लोकसाहित्य म्हणजे आपल्या समाजाच्या संस्कृतीची ओळख आणि आपल्या जीवनातील अनमोल ठेवा आहे.

डिजिटल दस्तऐवजीकरण : अर्थ आणि गरज :

डिजिटल दस्तऐवजीकरण म्हणजे एखाद्या विषयाशी संबंधित माहितीचे संकलन, स्कॅनिंग, टायपिंग, ऑडिओ-व्हिडिओ रेकॉर्डिंग, वर्गीकरण, संग्रहण आणि सुरक्षित जतन करून ती माहिती ऑनलाईन किंवा डिजिटल माध्यमातून उपलब्ध करून देणे. डिजिटल दस्तऐवजीकरण ही लोकसाहित्य जतन करण्याची आधुनिक आणि प्रभावी प्रक्रिया आहे. डिजिटल दस्तऐवजीकरणामध्ये लोकसाहित्याचे संकलन क्षेत्रकार्याच्या माध्यमातून केले जाते. संशोधक ग्रामीण भागात, आदिवासी भागात, धार्मिक परंपरांमध्ये, जत्रा-उत्सवांमध्ये किंवा सांस्कृतिक कार्यक्रमांमध्ये सहभागी होऊन कथाकार, गायक, कीर्तनकार, भारूडकार, पोवाडेकरी यांच्याकडून लोकसाहित्य संकलित करतो. या संकलनासाठी ऑडिओ-व्हिडिओ रेकॉर्डिंग अत्यंत महत्त्वाचे ठरते, कारण लोकसाहित्य हे केवळ शब्दांचे नाही तर सादरीकरणाचेही साहित्य असते.

डिजिटल संग्रहाची रचना : मेटाडेटा व डेटाबेस :

लोकसाहित्याचा डिजिटल संग्रह करताना केवळ सामग्री साठवणे पुरेसे नाही, तर त्यासोबत मेटाडेटा तयार करणे अत्यावश्यक असते. मेटाडेटामध्ये लोककथेचा प्रकार, संकलन स्थळ, सांगणाऱ्याचे नाव/वय/व्यवसाय, बोलीचा प्रकार, संकलनाची तारीख, विषय, संदर्भ, गाण्याचा राग/लय, प्रसंग, धार्मिक/सामाजिक संदर्भ, तसेच संबंधित कीवर्ड यांचा समावेश होतो. अशा पद्धतीने डेटाबेस तयार केल्यास लोकसाहित्य शोधणे, तुलना करणे, वर्गीकरण करणे आणि अभ्यास करणे सुलभ होते. ही पद्धत ग्रंथालयशास्त्रातील वर्गीकरण संकल्पनेशी सुसंगत ठरते.

AI चा वापर : लिप्यंतरण (Speech-to-Text) आणि मजकूरनिर्मिती :

AI आधारित भाषातंत्रज्ञानामुळे लोकसाहित्याच्या ध्वनिफितींचे स्वयंचलित लिप्यंतरण (Speech-to-Text) शक्य होते. पारंपरिक पद्धतीत ध्वनी ऐकून हस्तलिखित किंवा टायपिंग करणे वेळखाऊ असते. AI तंत्रज्ञानामुळे हे काम जलद होते, तसेच मोठ्या प्रमाणावर संग्रह तयार करता येतो. मात्र मराठीतील बोली, ग्रामीण उच्चार, मिश्रभाषा, व लोकगीतांची लय यामुळे AI ला काही त्रुटी येऊ शकतात. त्यामुळे मानवी संपादन (Human Editing) आवश्यक ठरते. तरीही AI चा उपयोग वेळ व श्रम बचतीसाठी मोठ्या प्रमाणावर फायदेशीर ठरतो.

भावविश्लेषण व सामाजिक वास्तवाचा अभ्यास :

ऐतिहासिक लोकसाहित्यातील भावना आणि सामाजिक संबंधांचा अभ्यास करण्यासाठी आता AI (Artificial Intelligence) आधारित 'भावविश्लेषण' (Sentiment Analysis) एक प्रभावी साधन आहे. लोकगीतांमधील विरह आणि आनंद, तसेच दुःख, उपहास, श्रद्धा आणि संघर्षाच्या विविध भावना AI च्या साहाय्याने ओळखता येतात. यासोबतच, लोककथांमधील नायक आणि खलनायकांच्या कल्पना, सामाजिक न्याय आणि अन्याय, जातीय संदर्भ, स्त्री-पुरुषांच्या भूमिका, आणि आर्थिक विषमता यांसारख्या घटकांचा अभ्यास मोठ्या प्रमाणात करता येणे शक्य आहे. यामुळे लोकसाहित्याचा समाजशास्त्रीय आणि सांस्कृतिक

दृष्टिकोन अधिक वैज्ञानिक आणि विस्तृत होऊ शकतो.

AI च्या साहाय्याने वर्गीकरण : प्रकार, विषय आणि बांधणी :

लोकसाहित्याच्या विविध प्रकारांना विभागणे म्हणजे संशोधनात टाकलेले महत्वाचे पाऊल असते. आर्टिफिशियल इंटेलिजन्सच्या मदतीने लोककथा, म्हणी, गाणी, पोवाडे, भारुडे, उखाणे अशा अनेक प्रकारांना अचूक ओळखता येते आणि त्यांचे आपोआप वर्गीकरण करता येते. त्याचप्रमाणे, विषय लक्षात घेऊन- उदाहरणार्थ, महिलांचे जीवन, शेती, निसर्ग, धार्मिक श्रद्धा, विवाह, कष्टकरी जीवनशैली, समाजातील दोष, पराक्रम, नैतिकता- अशा प्रकारे विषयानुसार वर्गीकरण करता येते. 'मशीन लर्निंग' आणि ‘नैसर्गिक भाषा प्रक्रिया’ वापरून, मजकुरातील शब्द, वाक्यांची रचना आणि अर्थ समजून घेऊन, AI त्या साहित्याचा प्रकार आणि विषय निश्चित करू शकते.

शब्दसंग्रह, बोलीभाषा व भाषिक वैशिष्ट्यांचे विश्लेषण :

मराठी लोकसाहित्यातील बोलीभाषा, स्थानिक शब्द, म्हणी, विशिष्ट वाक्प्रचार आणि उच्चारांसंबंधीची वैशिष्ट्ये भाषावैज्ञानिक दृष्ट्या अत्यंत महत्त्वपूर्ण आहेत. कृत्रिम बुद्धिमत्तेच्या (AI) साहाय्याने शब्दांची वारंवारता, बोली शब्दांचे भौगोलिक नकाशावर स्थान निश्चितीकरण, समानार्थी शब्दांचे जाळे (Word Network), त्याचप्रमाणे काळानुसार भाषेमध्ये होणारे बदल अभ्यासता येतात. उदाहरणार्थ, कोकणी मराठीतील शब्दसंपदा आणि विदर्भातील शब्दसंपदा यांची तुलना AI च्या साहाय्याने सुलभपणे करता येणे शक्य आहे. यामुळे मराठी भाषेच्या विविध स्वरूपांचे दस्तऐवजीकरण व जतन होऊ शकते.

डिजिटल लोकसाहित्य आणि शिक्षण-प्रसार :

डिजिटल लोकसाहित्याचा महत्त्वपूर्ण फायदा म्हणजे शिक्षण आणि प्रसार होय. शिक्षणसंस्थांमध्ये लोकसाहित्याचे अध्यापन करताना डिजिटल ऑडिओ-व्हिडिओ माध्यमांचा वापर केल्यास विद्यार्थ्यांना ते अधिक प्रभावीपणे आकलन होते. त्याचप्रमाणे, ऑनलाइन वस्तुसंग्रहालये, लोकसाहित्य पोर्टल्स, ई-लायब्ररी, पॉडकास्ट आणि डिजिटल प्रदर्शने यांच्याद्वारे लोकसाहित्य समाजात प्रसारित होऊ शकते. कृत्रिम बुद्धिमत्ता (AI) आधारित शोध प्रणालीमुळे एखादा वापरकर्ता शब्द, वाक्प्रचार, विषय किंवा ठिकाण टाकून संबंधित माहिती सुलभपणे मिळवू शकतो.

समारोप :

लोकसाहित्य हे कोणत्याही समाजाच्या इतिहासाचं जिवंत रूप असतं. ते फक्त मागच्या काळातल्या गोष्टी नाहीत, तर आजच्या जगाचं प्रतिबिंब दाखवतं आणि येणाऱ्या पिढीला योग्य मार्ग दाखवतं, असं ते सांस्कृतिक धन आहे. आजच्या डिजिटल जगात लोकसाहित्याची नोंद करणं फक्त अभ्यासकांचं काम नाही, तर ती समाजाची सांस्कृतिक गरज आहे. एआयसारख्या (AI) आधुनिक टेक्नॉलॉजीमुळे हे काम आणखी सोपं, लवकर आणि व्यवस्थित होऊ शकतं. पण टेक्नॉलॉजीच्या मदतीने लोकसाहित्याचं 'जतन' करताना, त्यातल्या माणसांच्या

भावना, सांस्कृतिक संदर्भ आणि त्या समाजाचा हक्क यांचा आदर करणं खूप महत्त्वाचं आहे. त्यामुळेच 'डिजिटल नोंदणी + एआय (AI) + सांस्कृतिक समज' या तीन गोष्टींना एकत्र करून काम केल्यास, मराठी लोकसाहित्याचा खजिना पुढच्या पिढीसाठी सुरक्षित आणि चांगला राहील.

निष्कर्ष :

१) मराठी लोकसाहित्य हे महाराष्ट्राच्या लोकजीवनाचे, संस्कृतीचे आणि परंपरेचे मौलिक दस्तऐवज आहे.

२) बदलत्या काळात मौखिक परंपरा कमी होत असल्यामुळे लोकसाहित्य विस्मरणाच्या उंबरठ्यावर आहे. त्यामुळे लोकसाहित्याचे डिजिटल दस्तऐवजीकरण करणे ही काळाची अत्यंत गरज ठरते.

३) डिजिटल स्वरूपात लोककथा, लोकगीते, म्हणी, उखाणे, पोवाडे, भारुडे इत्यादींचे दीर्घकालीन जतन शक्य होते.

४) डिजिटल दस्तऐवजीकरणामुळे लोकसाहित्याचे संकलन अधिक शिस्तबद्ध, व्यवस्थित आणि सुरक्षित पद्धतीने करता येते.

५) AI चा उपयोग करताना बोलीभाषा, संदर्भ, सांस्कृतिक अर्थ आणि स्थानिक वैशिष्ट्ये यांचा विचार करणे आवश्यक आहे.

६) या प्रक्रियेमुळे मराठी लोकसाहित्याचा प्रसार नव्या पिढीपर्यंत आणि जागतिक स्तरावरही होऊ शकतो.

संदर्भ टीपा :

- डॉ. विश्वनाथ शिंदे : लोकसाहित्यमीमांसा (भाग पहिला), स्नेहवर्धन पब्लिशिंग हाऊस, पुणे, १९९८, पृष्ठ क्र. २२४
- तत्रैव : पृष्ठ क्र. २२६
- तत्रैव : पृष्ठ क्र. १९-२०

संदर्भ ग्रंथसूची :

- ग. त्र्यं. माडखोलकर – *लोकसाहित्य : स्वरूप आणि परंपरा*
- द. मा. मिरासदार – *लोककथा आणि लोकजीवन*
- प्रा. रा. चिं. ढेरे – *लोकसंस्कृतीचे अभ्यासक्षेत्र*
- भालचंद्र नेमाडे – *साहित्य आणि समाज*
- डॉ. श्री. ना. जोशी – *लोकसाहित्य : संकलन व संशोधन पद्धती*
- डॉ. सदानंद मोरे – *महाराष्ट्राची लोकसंस्कृती*
- विजय तेंडुलकर (संपा.) – *मराठी समाजजीवन आणि सांस्कृतिक संदर्भ*

- Artificial Intelligence: A Modern Approach — *Stuart Russell & Peter Norvig*
- Daniel Jurafsky & James H. Martin – *Speech and Language Processing*
- Christopher Manning & Hinrich Schütze – *Foundations of Statistical Natural Language Processing & AI*
- Google Web Search
- YouTube Channel : *Deep Learning AI, Data School*