

## साहित्य आणि ऐतिहासिक मजकुराचे विश्लेषण करण्यासाठी AI एक साधन

प्रा. डॉ. योगेश या. किनकर

अरविंदबाबू देशमुख महाविद्यालय भारसिंगी

प्रा. राजेंद्र वा. काटोले

संशोधक विद्यार्थी

Crossref DOI - <https://doi.org/10.63665/rh.v7i2.122>

### प्रस्तावना :

आर्टिफिशियल इंटेलिजन्स हा खूप विशाल आणि तांत्रिक विषय आहे. त्यामुळे तो ऐतिहासिक मजकुराचे विश्लेषण करण्यासाठी कसा उपयोगी येऊ शकतो व आर्टिफिशियल इंटेलिजन्स म्हणजे नक्की काय आणि ते कसे कार्य करते हा सध्याच्या काळातील अतिशय महत्त्वाचा विषय आहे. AI ची संकल्पना अमेरिकन संगणक शास्त्रज्ञ जॉन मकार्टी यांनी इ.स. 1956 साली मांडली. या संकल्पनेत क्लॉड शानोन, मार्विन मिन्स्की, रे सॉलोमनोफ, आलान नेवेल, आर्थर सॅम्युअल इत्यादी संशोधक वैज्ञानिकांनी आपले योगदान दिले. या सर्व वैज्ञानिकांनी AI म्हणजे विज्ञान आणि तंत्रज्ञान यांची सांगड घालून एक अधिक बुद्धिमान मशीन म्हणजेच सॉफ्टवेअर तयार करून ते भविष्यातील तंत्रज्ञानात किती आमुलाग्र बदल घडवून आणू शकते हे पटवून दिले. आजच्या काळात जवळपास सर्वजण दैनंदिन जीवनामध्ये AI चा उपयोग करत आहे.

त्याचप्रमाणे साहित्य आणि ऐतिहासिक मजकुराचे विश्लेषण करण्यासाठी AI चा उपयोग कसा होऊ शकतो याची बरीचशी उदाहरणे देता येतील. जसे AI मजकूर समजून घेण्यास मदत करू शकतो, थीम्स ओळखू शकतो, साहित्यिक शैलीचे विश्लेषण करू शकतो, ऐतिहासिक विश्लेषण व आत दडलेले पॅटर्न आणि ट्रेंड्स सुद्धा शोधता येतो. AI च्या मदतीने शिक्षणाचा दर्जा सुधारला जाऊ शकतो आणि लोकांचा शिक्षणापर्यंतचा प्रवेश वाढू शकतो. AI च्या मदतीने प्रशासनातील कार्यक्षमता व व्यापार उद्योगात त्याचा फायदा करून घेता येतो. त्याचप्रमाणे शेतीशी संबंधित समस्यांचे निराकरण करता येऊ शकते. जवळपास सर्व क्षेत्रात AI मानवाची मदत करू शकतो.

### AI चे कार्य :

AI हे असे एक तंत्रज्ञान आहे जे यांत्रिक किंवा प्रक्रिया प्रणाली म्हणजेच सॉफ्टवेअरला त्याच्या कार्याच्या संबंधित मिळणाऱ्या माहितीचा उपयोग करून निरंतर विकसित आणि वापरण्यायोग्य करत राहतो. अगदी सोप्या भाषेत समजवायचे झाल्यास आपणा सर्वांच्या मोबाईल मध्ये गुगल फोटोज हे आपलिकेशन आहे. त्यामध्ये पीपल अँड पेट्स हे ऑप्शन आहे. त्यावर क्लिक केल्यावर तुम्हाला तिथे ओळखीच्या व्यक्तीच्या चेहऱ्याचे आयकॉन दिसतात. ते आयकॉन दुसरे तिसरे काही नसून प्रत्येक विशिष्ट व्यक्तीसाठी स्वतःचे असे स्वतंत्र फोल्डर आहे. ज्यामध्ये तुमच्या मोबाईल मधील केवळ असे छायाचित्र दिसतील विशिष्ट व्यक्ती उपस्थित असेल. इथे गुगल फोटोज हे एक सॉफ्टवेअर आहे आणि मोबाईल मधील सर्व छायाचित्र त्याला मिळणारी माहिती व हेच



सॉफ्टवेअर A I चा उपयोग करून मोबाईल मधून मिळणाऱ्या माहितीवर म्हणजेच छायाचित्रांवर प्रक्रिया करून त्या फोटोमध्ये असणाऱ्या प्रत्येक व्यक्तीसाठी स्वतंत्र असे संग्रह तयार करतो. आणि वाखाण्याजोगी गोष्ट म्हणजे या सॉफ्टवेअर मध्ये A I च्या मदतीने निरंतर सुधारणा होत राहते. AI स्वतःमध्ये सुधारणा करता करता इतके विकसित होऊन जाते की त्याची निर्णय क्षमता ही मानवापेक्षा सुद्धा जास्त अचूक असते व ती दोन चित्रांमधील व्यक्ती एकच आहे हे मान्य करण्यासाठी ते छायाचित्र निरखून बघावे लागते. तसेच दैनंदिन जीवनामध्ये गुगल, फेसबुक, ट्विटर वापरत असताना त्यामध्ये सुद्धा AI च योगदान आहे. गुगलवर सर्च करताना तुम्ही थोडेच शब्द टाईप केल्यावर आपल्याला अपेक्षित असणाऱ्या शब्दाच्या सूचना येतात जो शब्द आपण अजून पर्यंत टाईप सुद्धा केलेला नसतो.

### AI मुळे निर्णयक्षमता खुंटत आहे :

एखादा वापरकर्ता जेव्हा AI चॅटबॉटला प्रश्न माहिती विचारतो किंवा सल्ला मागतो तेव्हा चॅटबॉट खूप सहजतेने वापरकर्त्याच्या भूमिकेशी सहमत होतात आणि वापरकर्त्याला जे ऐकायचे आहे तेच सांगतात मग ते चुकीची असो किंवा त्या माहितीवर निर्णय घेण्यास धोका असो. A I च्या प्रवृत्तीकडे संकट म्हणून बघायला पाहिजे कारण माणूस A I वर विसंबून राहू लागला आणि त्याचे प्रमाण दिवसेंदिवस वाढतच जाणार आहे. प्रत्येक प्रश्नाचे मानवाच्या सोयीनुसार उत्तर मिळाल्यामुळे मानवाची निर्णय क्षमता खुंटत चालली आहे. स्वतः निर्णय न घेऊ शकल्यामुळे त्याला A I वर विसंबून राहावे लागत आहे. याचा फायदा घेऊन एक दिवस मानव पूर्णपणे मानसिक गुलाम बनणार आहे तसेच अनेक लोकप्रिय A I चॅटबॉट वापरकर्त्याला आवडणारी त्याला हवी तशी उत्तरे देऊ लागला आहे. हुजरेगिरी किंवा चापलुसित A I ने मानवाला ही मागे सोडले आहे.

### साहित्याचे विश्लेषण करण्यासाठी AI एक उपयोगी साधन :

प्राचीन भाषा अवगत नसल्यामुळे अनेक समस्या समोर येत होत्या जुने साहित्य समजून घेण्यात कितीतरी महिने व वर्षे लागत होते. संशोधनाकरिता मोडी लिपी किंवा संस्कृत येत नसल्यामुळे अनेक साहित्य समजू शकत नव्हते परंतु आता साहित्याचा अभ्यास ऐतिहासिक कागदपत्रे जतन करण्याच्या पद्धती बदलल्या आहेत. डिजिटल मानवीकीमुळे विशेषतः मराठी सारख्या जुनी आणि समृद्ध परंपरा असलेल्या भाषांसाठी कृत्रिम बुद्धिमत्ता A I खूप महत्त्वाची ठरली आहे. A I केवळ तांत्रिक मदत करत नाही तर ज्ञान जपणे आणि त्याचा अभ्यास करणे यासाठी उपयोगी ठरते. आज उपलब्ध असलेली आधुनिक साधने हजारो पानाचे साहित्य जुनी कागदपत्रे आणि हस्तलिखिते एकत्रितपणे अभ्यासू शकतात या पद्धतीला 'दुरस्त वाचन' असे म्हणतात. यात संगणकाच्या मदतीने भाषेतील बदल लेखनातील पद्धती आणि समाजाशी संबंधित भावी ओळखता येतात ज्या बारकाईने वाचून शोधणे कठीण असते.

### साहित्यिक विश्लेषणाची मूलभूत साधने आणि विश्लेषण प्रतिमाने :

साहित्यिक संशोधनामध्ये कृत्रिम बुद्धिमत्तेच्या अनुप्रयोगांना समजून घेण्यासाठी, सर्वप्रथम त्या साधनांचे विश्लेषण करणे आवश्यक आहे जे पाठ्य (टेक्स्ट) डेटा प्रक्रिया करण्यासाठी प्राथमिक व्यासपीठ उपलब्ध करून



देतात. Voyant Tools आणि AntConc यांसारखी साधने या क्षेत्रातील अग्रगण्य मानली जातात. Voyant Tools हे वेब-आधारित, ओपन-सोर्स व्यासपीठ असून ते पाठ्य विश्लेषण आणि दृश्यीकरणासाठी (व्हिज्युअलायझेशन) लवचिक वातावरण प्रदान करते. हे विशेषतः अशा संशोधकांसाठी डिझाइन केलेले आहे ज्यांना कोडिंगचे तांत्रिक ज्ञान नसते, तरीही ते जटिल डेटासेटचा अभ्यास आणि अन्वेषण करू इच्छितात. Voyant Tools ची सर्वात महत्त्वाची वैशिष्ट्ये म्हणजे त्याचा बहु-पॅनल इंटरफेस. यामध्ये 'Cirrus' सारखा वर्ड क्लाउड जनरेटर, 'Trends' ग्राफ आणि 'Contexts' यांसारखी साधने समाविष्ट आहेत, जी एखादा विशिष्ट शब्द पाठ्यात कुठे आणि कशा प्रकारे वापरला गेला आहे हे दर्शवतात. याच्या विपरीत, AntConc हे डेस्कटॉप-आधारित फ्रीवेअर टूलकिट असून ते कॉर्पस भाषाविज्ञानावर (Corpus Linguistics) अधिक लक्ष केंद्रित करते. संशोधनातून असे सूचित होते की Voyant Tools चे इंटरफेस शोधपरक विश्लेषणासाठी (Exploratory Analysis) अधिक उपयुक्त आहे, तर AntConc हे त्या संशोधकांसाठी अधिक योग्य आहे जे आधीच आपल्या पाठ्यसामग्रीशी परिचित आहेत आणि विशिष्ट प्रश्नांची उत्तरे शोधत आहेत.

### ऐतिहासिक हस्तलिखितांचे डिजिटायझेशन आणि एआय-आधारित ट्रान्सक्रिप्शन :

ऐतिहासिक मजकूर विश्लेषणातील सर्वात मोठे आव्हान म्हणजे हस्तलिखितांची भौतिक अवस्था आणि त्यांची हस्तलेखन शैली. ऐतिहासिक दस्तऐवजांचे एआय-आधारित विश्लेषण तेव्हाच शक्य होते, जेव्हा त्यांना मशीन-वाचनीय स्वरूपात रूपांतरित केले जाते. ट्रान्सक्रिबस (Transkribus) हे या क्षेत्रातील एक क्रांतिकारी साधन म्हणून पुढे आले आहे, जे हस्तलिखित मजकूर ओळख (Handwritten Text Recognition – HTR) साठी कृत्रिम बुद्धिमत्तेचा वापर करते. हे संशोधकांना त्यांच्या विशिष्ट हस्तलेखन शैलीसाठी सानुकूल एआय मॉडेल्स प्रशिक्षित करण्याची परवानगी देते, जी शतकानुशतके जुनी लिपी वाचण्यास सक्षम असतात. ट्रान्सक्रिबसचा कार्यप्रवाह अत्यंत सुव्यवस्थित आहे. यात प्रतिमांची स्पष्टता आणि वाचनीयता वाढवण्यासाठी प्रतिमा अनुकूलन, मेटाडेटा व्यवस्थापन आणि डीप लर्निंगद्वारे मजकूर ओळख यांचा समावेश होतो. अॅरोल्सेन अर्काइव्हज यांसारख्या संस्थांनी या तंत्रज्ञानाचा वापर करून लाखो दस्तऐवजांचे अनुक्रमण (इंडेक्सिंग) केले आहे. तथापि, एआय-आधारित प्रणालींमध्ये सध्या त्रुटी दर सुमारे ५% आहे, तर मानवी तज्ज्ञांसाठी तो केवळ १% आहे. तरीही, एआयची गती संशोधन प्रक्रियेला सुमारे ९९% पर्यंत वेगवान करते, ज्यामुळे मोठ्या प्रमाणावर डिजिटायझेशन शक्य होते. ऐतिहासिक दस्तऐवजांच्या प्रक्रियेत जेटस्ट्रीम एआय सारख्या प्रणाली 'डॉक्युमेंट इंटेलिजन्स'चा एक नवा स्तर सादर करतात. ही प्रणाली केवळ मुद्रित आणि हस्तलिखित मजकूर वाचत नाही, तर जुन्या दस्तऐवजांना त्यांच्या प्रकार, कालखंड किंवा विषयाच्या आधारे आपोआप वर्गीकृतही करते. याचे 'जेटस्ट्रीम अंडरस्टॅंडिंग' मॉड्यूल मोठ्या भाषा मॉडेल्स चा वापर करून मजकुरात स्पष्टपणे न मांडलेल्या अंतर्दृष्टी आणि सारांशही काढते. यामुळे इतिहासकारांना वैयक्तिक पत्रे आणि न्यायालयीन नोंदींमधील लपलेले सामाजिक संबंध समजून घेण्यास मदत होते.

### मराठी भाषा आणि मोडी लिपी विश्लेषणासाठी विशेष AI साधने :

मराठी ही भारतातील तिसरी सर्वाधिक बोलली जाणारी भाषा असून, तिची गुंतागुंतीची व्याकरणरचना



आणि श्लेष (Agglutinative) प्रकृतीमुळे संगणकीय विश्लेषणासाठी ती एक वेगळीच आव्हाने निर्माण करते. मराठीत शब्दांना अनेक प्रत्यय जोडले जातात (Suffix Stacking), ज्यामुळे एका मूलशब्दावरून लिंग, वचन आणि कारक यांचे अर्थ व्यक्त होतात. या गुंतागुंतीवर मात करण्यासाठी आयआयटी बॉम्बे येथील संगणकीय भारतीय भाषा तंत्रज्ञान केंद्राने (CFILT) प्रगत असे 'मराठी रूपात्मक विश्लेषक' (Marathi Morphological Analyzer) विकसित केले आहेत. ऐतिहासिक दृष्टीने पाहता, मराठी भाषेसाठी मोडी लिपीचे महत्त्व अत्यंत मोठे आहे. १२व्या शतकापासून ते २०व्या शतकाच्या मध्यापर्यंत महाराष्ट्रातील प्रशासकीय व अधिकृत कामकाजासाठी मोडी लिपीचा वापर केला जात होता. आजच्या घडीला सुमारे ४ कोटी मोडी दस्तऐवज अप्रकाशित आणि अपठित अवस्थेत आहेत, कारण ही लिपी वाचू शकणाऱ्या तज्ज्ञांची संख्या फारच मर्यादित आहे. मोडी लिपीतील शिरोरेषेचा अभाव आणि वळणदार अक्षररचना तिला देवनागरीपेक्षा वेगळी बनवते; लेखनाचा वेग वाढवण्यासाठीच ही लिपी विकसित करण्यात आली होती. या क्षेत्रात आयआयटी रुडकीने विकसित केलेले 'MoScNet' (Modi Script Network) हे एक महत्त्वाचे टप्पे मानले जाते. हे एक व्हिजन-लॅंग्वेज मॉडेल (VLM) असून, हस्तलिखित मोडी लिपीतील प्रतिमा थेट देवनागरी मजकुरात लिप्यंतरित (Transliterate) करण्यासाठी 'ज्ञान आसवन' (Knowledge Distillation) तंत्राचा वापर करते. या संशोधनासाठी 'MoDeTrans' नावाचा व्यापक डेटासेट तयार करण्यात आला आहे, ज्यामध्ये शिवाजी महाराजांच्या काळापासून ते ब्रिटिश कालखंडापर्यंतचे २,००० हून अधिक सत्यापित प्रतिमा आणि त्यांची भाषांतरे समाविष्ट आहेत.

### डिजिटल मानविकीतील व्यावहारिक उपयोग आणि केस स्टडी :

साहित्यिक विश्लेषणाच्या क्षेत्रात कृत्रिम बुद्धिमत्तेचे (एआय) व्यावहारिक मूल्य विविध केस स्टडीजमधून स्पष्टपणे दिसून येते. उदाहरणार्थ, संत तुकाराम आणि संत ज्ञानेश्वर यांच्या भक्तिकाव्याचे एआयच्या साहाय्याने केलेले विश्लेषण त्यांच्या तात्त्विक साम्यस्थळांबरोबरच भाषिक वैशिष्ट्येही अधोरेखित करते. ज्ञानेश्वरी ही तिच्या बौद्धिक स्पष्टतेसाठी आणि शिस्तबद्ध व्याकरणासाठी ओळखली जाते, तर तुकारामांचे अभंग हे त्यांच्या लोकशैलीसाठी आणि सामाजिक समतेवरील भरासाठी प्रसिद्ध आहेत. एआय साधने या कवितांमधील शब्दप्रयोगांची वारंवारता आणि भावनिक चढउतार यांचे आलेख तयार करू शकतात, ज्यातून या संतकवींनी मराठी भाषेला आध्यात्मिक अभिव्यक्तीचे प्रभावी माध्यम कसे बनवले हे स्पष्ट होते.

ऐतिहासिक बखरींच्या (Maratha Bakhars) संदर्भात, एआय-आधारित विश्लेषण संशोधकांना वस्तुनिष्ठ तथ्ये आणि अतिरंजित वर्णने यांमधील फरक ओळखण्यास मदत करते. सभासद बखर (१६९४) यांसारख्या प्राथमिक स्रोतांचे विश्लेषण करताना, विशिष्ट शब्दसंपदेचा कालानुक्रमिक प्रसार मागोवा घेण्यासाठी एआयचा वापर करता येतो. यामुळे मराठा प्रशासनात फारसी शब्दांची जागा हळूहळू मराठी शब्दांनी कशी घेतली, हे समजून घेणे शक्य होते. भाषिणी (Bhashini) यांसारख्या सरकारी उपक्रमांमुळे या सर्व तंत्रज्ञानांना एकत्रित मंच उपलब्ध होत आहे. भाषिणी अंतर्गत 'लेखानुवाद' ही सेवा दस्तऐवजांचे भाषांतर आणि डिजिटलीकरण सुलभ करते, तर 'अभिलेख' ही सेवा प्रत्यक्ष वेळेत भाषण-ते-मजकूर (Speech-to-Text) रूपांतरण उपलब्ध करून



देते. ही साधने केवळ संशोधकांसाठीच नव्हे, तर सर्वसामान्य नागरिकांसाठीही ऐतिहासिक ज्ञानाचे दरवाजे खुले करत आहेत.

संगणकीय विश्लेषणाचा आणखी एक महत्त्वाचा पैलू म्हणजे रूपात्मक विश्लेषण (Morphological Analysis). मराठीसारख्या योगात्मक भाषांमध्ये शोध परिणाम अधिक अचूक करण्यासाठी लेमटायझेशन (Lemmatization) अत्यावश्यक ठरते. आयआयटी बॉम्बेच्या संशोधनानुसार, लेमटायझेशनमुळे मराठी एकभाषिक माहिती पुनर्प्राप्तीची कार्यक्षमता 0.3366 वरून 0.4003 (MAP स्कोअर) इतकी वाढते. यावरून साहित्यिक डेटाबेसमध्ये शोधासाठी एआय अल्गोरिदमचे महत्त्व स्पष्ट होते.

### भविष्यातील दिशा आणि नैतिक विचार :

जसजशी कृत्रिम बुद्धिमत्ता साहित्यिक आणि ऐतिहासिक विश्लेषणात अधिक खोलवर प्रवेश करत आहे, तसतसे डेटा सार्वभौमत्व (Data Sovereignty) आणि नैतिक प्रतिनिधित्व यासंबंधीचे प्रश्न अधिक महत्त्वाचे ठरत आहेत. भारतजेन आणि भाषिणी यांसारखे उपक्रम एआय मॉडेल्स केवळ तांत्रिकदृष्ट्या सक्षम नसून, भारतीय मूल्ये, संस्कृती आणि इतिहास यांमध्येही खोलवर रुजलेले असावेत, याची खात्री करण्याचा प्रयत्न करत आहेत. ऐतिहासिक दस्तऐवजांच्या बाबतीत, एआयमधील “हॅल्युसिनेशन्स” टाळणे हे सर्वात मोठे आव्हान आहे, कारण चुकीच्या ऐतिहासिक मांडणीतून गंभीर वाद निर्माण होऊ शकतात.

भविष्यात, एआय केवळ विद्यमान मजकुराचे विश्लेषण करेल इतकेच नव्हे, तर साहित्यिक प्रवृत्तींची भाकीत करणे आणि विशिष्ट लेखकांच्या शैलींचे अचूक मॉडेल तयार करणेही शक्य होईल, अशी अपेक्षा आहे. ‘लिपिकर’ यांसारख्या बहुभाषिक ओसीआर प्रणाली आणि IIF-आधारित डिजिटल हस्तलिखित संग्रह भारताची सांस्कृतिक वारसा जागतिक स्तरावर सुलभ करून देतील. अखेरीस, साहित्यिक आणि ऐतिहासिक मजकूर विश्लेषणासाठी एआय हे असे प्रभावी साधन ठरत आहे, जे इतिहासातील मूक आवाजांना शब्द देऊन आजच्या संशोधकांना त्यांच्या पूर्वजांच्या ज्ञानाशी जोडणारे एक नवे दृष्टीकोन प्रदान करते.

### सारांश :

गेल्या अनेक दशकापासून आर्टिफिशियल इंटेलिजन्स हा चर्चेचा विषय आहे. शास्त्रज्ञ वेळोवेळी त्याच्या चांगल्या वाईट परिणामाची चर्चा करत असतात. आज तंत्रज्ञानामुळे जग झपाट्याने बदलत आहे. विकासाला गती देण्यासाठी आणि लोकांना चांगल्या सुविधा देण्यासाठी अत्याधुनिक तंत्रज्ञानाचा प्रत्येक क्षेत्रात मोठ्या प्रमाणात वापर केला जात आहे. वाढत्या औद्योगीकरण, शहरीकरण आणि जागतिकीकरणामुळे विकासाचा वेग वाढला असतानाच अनेक नव्या समस्याही निर्माण झाल्या आहे, ज्यासाठी रोज नवनवीन उपाय योजले जात आहेत. शास्त्रज्ञ कृत्रिम बुद्धिमत्तेचे अनेक फायदे सांगतात परंतु ते असेही मानतात की त्याच्या आगमनामुळे मानवांचे सर्वात मोठे नुकसान होईल, त्यांचे काम यंत्रांच्या ताब्यात जाईल, जे त्यांचे स्वतःचे निर्णय घेतील आणि जर ते नियंत्रित केले नाही तर, ते मानवी सभ्यतेसाठी हानिकारक असू शकतात. अशा परिस्थितीत त्यांचा वापर करण्यापूर्वी नफा आणि तोटा दोन्ही संतुलित करणे आवश्यक आहे.



**संदर्भ :**

- गोडबोले, अच्युत - “आर्टिफिशियल इंटेलिजन्स”, मधुश्री पब्लिकेशन, मुंबई.
- Russell, S., & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach* (3rd ed.). Pearson.
- दैनिक लोकमत दिनांक 12 डिसेंबर 2025 पेज नंबर 05
- दैनिक लोकमत दिनांक 05 डिसेंबर 2025 शुक्रवार पेज नंबर 06
- दैनिक लोकमत दिनांक 01 डिसेंबर 2025 पेज नंबर 02
- दैनिक लोकमत दिनांक 04 डिसेंबर 2025 पेज नंबर 04
- <https://Digital-humanities.phil.muni.cz>

