

## कृत्रिम बुद्धिमत्ता आणि मराठी भाषा संवर्धन: तांत्रिक प्रगती, आव्हाने आणि भविष्यातील दिशा

सहा. प्रा. रामभाऊ धोंडिराम कांबळे

मराठी विभाग

विश्वासराव रणसिंग महाविद्यालय, कळंब-वालचंदनगर

[Kamblerd20@gmail.com](mailto:Kamblerd20@gmail.com)

### प्रस्तावना :

मराठी भाषा ही केवळ संवादाचे साधन नसून ती महाराष्ट्राची सांस्कृतिक ओळख, ऐतिहासिक वारसा आणि ज्ञानाचा ठेवा आहे. या भाषेला अलिकडेच केंद्र सरकारकडून 'अभिजात भाषा' (Classical Language) म्हणून गौरवण्यात आले आहे. मात्र, २१ व्या शतकातील डिजिटल क्रांतीच्या युगात, केवळ लोकसंख्या किंवा अधिकृत दर्जा भाषेचे अस्तित्व टिकवण्यासाठी पुरेसा नाही. जागतिक स्तरावर ४०% भाषा नामशेष होण्याच्या मार्गावर असताना, मराठीला 'डिजिटल' स्वरूपात सक्षम करणे ही काळाची गरज बनली आहे. कृत्रिम बुद्धिमत्ता (AI) आणि नैसर्गिक भाषा प्रक्रिया (NLP) या क्षेत्रातील संशोधनाने भाषेच्या संवर्धनासाठी आणि विकासासाठी अभूतपूर्व संधी उपलब्ध करून दिल्या आहेत. तरीही, इंग्रजीसारख्या उच्च-संसाधन (High-resource) भाषांच्या तुलनेत मराठी अजूनही तांत्रिक प्रक्रियेत 'कमी-संसाधन' (Low-resource) श्रेणीत मोडते, त्यामुळे संशोधकांसमोर मोठी आव्हाने उभी राहिली आहेत.

## मराठी भाषेची संरचना आणि नैसर्गिक भाषा प्रक्रियेतील आव्हाने :

मराठी भाषेचे संवर्धन करताना कृत्रिम बुद्धिमत्तेचा वापर करण्यापूर्वी तिच्या वैशिष्ट्यपूर्ण भाषिक रचनेचा अभ्यास करणे अनिवार्य आहे. मराठी ही एक 'अॅग्लुटिनेटिव्ह' (Agglutinative) आणि मॉर्फोलॉजिकल दृष्ट्या अत्यंत समृद्ध भाषा आहे. याचा अर्थ असा की, मराठीत एका मूळ धातूपासून किंवा शब्दापासून विभक्ती प्रत्यय, काळ आणि लिंगानुसार शेकडो विविध रूपे तयार होऊ शकतात. उदाहरणार्थ, 'जा' या शब्दापासून 'जातो', 'जाते', 'जावे लागेल', 'जाण्यापूर्वी' अशी अनेक रूपे तयार होतात, जी संगणकीय प्रणालींना समजण्यासाठी इंग्रजीच्या तुलनेत अधिक कठीण असतात.

मराठी व्याकरणामध्ये लिंग (पुल्लिंगी, स्त्रीलिंगी, नपुंसकलिंगी) आणि वचनानुसार (एकवचन, अनेकवचन) क्रियापदांची रूपे बदलतात. जेव्हा एखादे कृत्रिम बुद्धिमत्ता मॉडेल मराठीत मजकूर तयार करते, तेव्हा 'लिंग-वचन-विभक्ती' यांचा ताळमेळ राखणे हे तांत्रिकदृष्ट्या आव्हानात्मक ठरते. याव्यतिरिक्त, मराठीमध्ये 'मुक्त शब्दक्रम' (Free word order) असतो, ज्यामुळे वाक्यातील शब्दांची जागा बदलली तरी अर्थ तोच राहू शकतो, परंतु संगणकाला वाक्याचा नेमका संदर्भ आणि कर्ता-कर्म संबंध समजून घेण्यात अडथळे येतात.

## खालील तक्ता मराठी भाषेच्या संगणकीय आव्हानांचे तुलनात्मक विश्लेषण दर्शवतो:

भाषिक वैशिष्ट्ये	संगणकीय प्रक्रिया (AI) वरील परिणाम	प्रमुख अडथळा
मॉर्फोलॉजिकल समृद्धता	एकाच शब्दाच्या अनेक रूपांमुळे 'टोकॅंब्युलरी'चा आकार वाढतो.	टोकनायझेशन (Tokenization) मध्ये गुंतागुंत.
मुक्त शब्दक्रम	वाक्याचा अन्वयार्थ लावणे कठीण होते.	सिंटॅक्स विश्लेषण (Syntactic Parsing) मध्ये त्रुटी.

लिपी वैविध्य	देवनागरी व्यतिरिक्त मोडी आणि रोमन लिपीचा वापर (उदा. Hinglish/Marathish).	स्क्रिप्ट नॉर्मलायझेशन आणि ट्रान्सलिटरेशन.
कोड-मिक्सिंग	संभाषणात इंग्रजी-हिंदी शब्दांचा सर्रास वापर.	शुद्ध भाषेची मॉडेल्स संवादात अपयशी ठरतात.

मराठीतील 'जोडाक्षरे' आणि 'अनुस्वार' यांच्या वापराबाबतही संगणकीय प्रणालींमध्ये संदिग्धता आढळते. अनेकदा एकाच शब्दाचे विविध लेखन प्रकार (उदा. 'मराठीत' विरुद्ध 'मराठी मध्ये') एआय मॉडेल्सना संभ्रमात टाकू शकतात. या सर्व आव्हानांमुळे मराठीसाठी तयार करण्यात आलेले एनएलपी अल्गोरिदम हे अत्यंत लवचिक आणि भाषिक बारकावे समजून घेणारे असणे आवश्यक आहे.

## ऐतिहासिक वारसा आणि तांत्रिक पायाभरणी: सी-डॅकचे योगदान

मराठी भाषेच्या तांत्रिक संवर्धनाची मुहूर्तमेढ १९८८ मध्ये 'सेंटर फॉर डेव्हलपमेंट ऑफ अॅडव्हान्स्ड कम्प्युटिंग' (C-DAC) च्या स्थापनेनंतर रोवली गेली. सी-डॅकने केवळ सुपरकम्प्युटिंगमध्येच नव्हे, तर 'मल्टिलिंग्युअल कम्प्युटिंग' (Multilingual Computing) क्षेत्रातही क्रांती घडवून आणली. त्यांच्या 'जिस्ट' (GIST) तंत्रज्ञानाने भारतीय भाषांना संगणकाच्या पडद्यावर स्थान दिले.

सी-डॅकच्या प्रमुख प्रकल्पांमध्ये 'इनस्क्रिप्ट' (InScript) कीबोर्डचा समावेश आहे, जो आजही भारतीय भाषांच्या टायपिंगसाठी प्रमाणित मानला जातो. त्यांच्या 'अनुवादक्ष' (Anuvadaksh) या इंग्रजी ते भारतीय भाषा अनुवाद प्रणालीने मराठीला डिजिटल प्रशासनात समाविष्ट करण्यात मोठी भूमिका बजावली. ही प्रणाली विशेषतः आरोग्य आणि पर्यटन क्षेत्रातील माहितीच्या अनुवादासाठी विकसित करण्यात आली होती, ज्यामध्ये हायब्रीड आर्किटेक्चरचा (Statistical, Example-based आणि Rule-based) वापर केला गेला आहे.

सी-डॅकने विकसित केलेले 'श्रुतलेखन' (Shrutlekhan) हे स्पीच-टू-टेक्स्ट सॉफ्टवेअर मराठी भाषणाचे मजकुरात रूपांतर करण्यासाठी मैलाचा दगड ठरले आहे. याचा वापर प्रशासकीय कामकाजात वेगाने माहिती नोंदवण्यासाठी केला जातो. तसेच, 'आयएसएम' (ISM V6) सारख्या सॉफ्टवेअरने प्रकाशन आणि ग्राफिक डिझाइन क्षेत्रात मराठीचा वापर सुलभ केला असून त्यात १९ भारतीय भाषांसह युनिकोड सपोर्ट आणि विविध फॉन्ट फॉर्मॅट्स उपलब्ध आहेत. सी-डॅकने केवळ मजकूर नव्हे, तर 'श्रुती दृष्टी' सारख्या प्रकल्पांतून अंध व्यक्तींसाठी स्क्रीन रीडिंग आणि टेक्स्ट-टू-स्पीच सुविधाही मराठीत उपलब्ध करून दिली आहे.

## राज्य मराठी विकास संस्था (RMVS) आणि सांस्कृतिक डिजिटायझेशन :

मराठी भाषेच्या संवर्धनासाठी महाराष्ट्र शासनाची 'राज्य मराठी विकास संस्था' (RMVS) ही एक कणा म्हणून काम करत आहे. या संस्थेने केवळ भाषेचे नियमन करण्याऐवजी तंत्रज्ञानाच्या माध्यमातून भाषेला आधुनिक स्वरूप देण्यावर भर दिला आहे. संस्थेच्या सर्वात महत्त्वपूर्ण प्रकल्पांपैकी एक म्हणजे 'मराठी दुर्मिळ ग्रंथांचे संगणकीकरण'. या प्रकल्पांतर्गत अशा ग्रंथांचे जतन केले जात आहे ज्यांचे कॉपीराइट संपले आहेत. आतापर्यंत १२९ पुस्तके आणि ५५५ नियतकालिकांचे संगणकीकरण पूर्ण झाले असून ती इंटरनेटवर मोफत उपलब्ध आहेत. यामुळे जगभरातील मराठी अभ्यासकांना आणि संशोधकांना महाराष्ट्राचा साहित्यिक ठेवा एका क्लिकवर उपलब्ध झाला आहे. याशिवाय, 'गिम्प' (GIMP) सारख्या मुक्तस्रोत सॉफ्टवेअरचे मराठीकरण करून संस्थेने भाषिक स्थानिकीकरणाचा (Localization) आदर्श घालून दिला आहे.

## राज्य मराठी विकास संस्थेच्या डिजिटल उपक्रमांची व्याप्ती खालील तक्त्यावरून स्पष्ट होते:

उपक्रम	स्वरूप	उद्देश
दुर्मिळ ग्रंथ संगणकीकरण	पीडीएफ आणि ई-बुक्स	ऐतिहासिक आणि साहित्यिक ठेवा जतन करणे.
श्राव्य पुस्तके (Audiobooks)	ऑडिओ फाइल्स (ज्ञानेश्वरी,	माहिती तंत्रज्ञानाचा वापर करून

	दासबोध इ.)	साहित्य सुलभ करणे.
युनिकोड टंक प्रोत्साहन	तांत्रिक प्रशिक्षण	संगणकीय मराठीत सुसूत्रता आणणे.
विकिपीडिया कार्यशाळा	संपादन प्रशिक्षण (RMVS + CIS-A2K)	मराठी विकिपीडियावर ज्ञानकोशीय माहिती वाढवणे.
बोलीभाषा सर्वेक्षण	भाषाशास्त्रीय अभ्यास	प्रादेशिक बोलींचे डिजिटल आलेखन करणे.

संस्थेने विकिमीडिया फाउंडेशनच्या सहकार्याने महाराष्ट्रातील विविध विद्यापीठांमध्ये कार्यशाळा आयोजित केल्या आहेत, ज्याद्वारे ५०० हून अधिक नवीन संपादकांना मराठी विकिपीडियावर माहिती लिहिण्याचे प्रशिक्षण दिले गेले आहे. यामुळे मराठीतील ऑनलाइन ज्ञानकोश समृद्ध होत असून एआय मॉडेल्सच्या प्रशिक्षणासाठी आवश्यक असलेला शुद्ध डेटा उपलब्ध होत आहे.

### नैसर्गिक भाषा प्रक्रिया (NLP) आणि आधुनिक संशोधनाचा प्रवास :

मराठी एनएलपी संशोधनाने गेल्या दहा वर्षांत मोठी झेप घेतली आहे. सुरुवातीच्या काळात हे संशोधन केवळ पार्ट-ऑफ-स्पीच (POS) टॅगिंग आणि नेमड एन्टीटी रिकग्निशन (NER) पुरते मर्यादित होते. मात्र, 'डीप लर्निंग' (Deep Learning) आणि 'ट्रान्सफॉर्मर' (Transformer) मॉडेल्सच्या आगमनाने यात आमूलाग्र बदल झाला आहे.

पुण्यातील एल३क्यूब (L3Cube) या संस्थेने विकसित केलेली 'महाएनएलपी' (mahaNLP) लायब्ररी ही मराठीसाठीची पहिली व्यापक ओपन-सोर्स लायब्ररी आहे. ही लायब्ररी 'महाबर्ट' (MahaBERT) सारख्या प्रगत मॉडेल्सवर आधारित असून त्यात सेंटिमेंट अॅनालिसिस, द्वेषपूर्ण भाषण ओळख (Hate Speech Detection) आणि वाक्य पूर्ण करणे यांसारखी वैशिष्ट्ये आहेत. महाएनएलपीने मराठीसाठी महाहॅट

(MahaHate), महासेंट (MahaSent) आणि महानेर (MahaNER) सारखे महत्त्वपूर्ण डेटासेट्स उपलब्ध करून दिले आहेत.

**मराठी एनएलपी अजूनही काही मूलभूत समस्यांशी झुंजत आहे:**

- **डेटाची कमतरता:** इंग्रजीसाठी अब्जावधी शब्दांचा डेटा उपलब्ध असताना मराठीसाठी उच्च-गुणवत्तेच्या समांतर वाक्यांची (Parallel Sentences) संख्या मर्यादित आहे.
- **कोड-मिक्स मजकूर:** सोशल मीडियावरील मराठी मजकूर अनेकदा रोमन लिपीत आणि इंग्रजी शब्दांच्या मिश्रणाने लिहिला जातो, ज्यामुळे शुद्ध मराठीवर आधारित मॉडेल्स गोंधळतात.
- **डोमेन-स्पेसिफिक डेटा:** वैद्यकीय किंवा कायदेशीर क्षेत्रातील मराठी शब्दावलीचे पुरेसे डिजिटल भांडार उपलब्ध नाही.या उणीवा भरून काढण्यासाठी 'एआय४भारत' (AI4Bharat) आणि 'भाषिणी' (Bhashini) सारख्या राष्ट्रीय प्रकल्पांनी मराठीसह सर्व २२ अधिकृत भाषांसाठी पायाभूत डेटा संकलित करण्याचे काम हाती घेतले आहे.

**भाषा मॉडेल्समध्ये (LLMs) मराठीचे स्थान आणि मूल्यांकन :**

कृत्रिम बुद्धिमत्तेच्या जगात सध्या चॅटजीपीटी (ChatGPT), लामा (Llama) आणि जेमिनी (Gemini) यांसारख्या मोठ्या भाषा मॉडेल्सचा बोलबाला आहे. या मॉडेल्सची मराठी भाषेतील क्षमता तपासण्यासाठी अनेक चाचण्या घेण्यात आल्या आहेत. संशोधनानुसार, मराठी ही जागतिक एलएलएममध्ये 'कमी-संसाधन' भाषा म्हणून गणली जात असली तरी, तिच्या कामगिरीत लक्षणीय सुधारणा होत आहे.

'इंडिकपॅराम' (IndicParam) या ११ भारतीय भाषांच्या मानकावर आधारित केलेल्या मूल्यांकनात असे दिसून आले की, गुगलचे जेमिनी-२.५ (Gemini-2.5) मॉडेल मराठीत ५८% अचूकता दर्शवते, तर जीपीटी-५

(GPT-5) आणि लामा-३.३ (Llama 3.3) अनुक्रमे ४५% आणि ३९% पर्यंत पोहोचतात. हे आकडे दर्शवतात की अजूनही जागतिक स्तरावर मराठी भाषेतील संदर्भांचे आकलन पूर्णपणे झालेले नाही.

**मराठी व्याकरण सुधारणेच्या बाबतीत विविध एआय टूल्सची कामगिरी खालीलप्रमाणे आढळली आहे :**

एआय टूल	व्याकरणातील अचूकता (Accuracy)	मुख्य वैशिष्ट्ये
ChatGPT (GPT-4)	९५.७२%	उच्च सुसंगतता आणि संवाद शैली.
Copilot (Microsoft)	९०.१८%	प्रशासकीय कामात उपयुक्त.
Gemini (Google)	९०.९०%	जलद प्रतिसाद आणि बहुभाषिक समन्वय.
Perplexity	९०.९०%	माहिती शोधण्यासाठी प्रभावी.

जरी अचूकतेचे आकडे ९० टक्क्यांच्या पुढे असले, तरी क्लिष्ट व्याकरण, लिंग-वचन समन्वय आणि मराठीतील विशिष्ट वाक्प्रचार हाताळताना ही मॉडेल्स अनेकदा चुका करतात. उदाहरणार्थ, उपहासात्मक किंवा उपरोधिक विधाने (Sarcasm) ओळखण्यात एआयला अद्यापही मर्यादा आहेत. तरीही, 'एल३क्युब-महाइमोशन्स' (L3Cube-MahaEmotions) सारख्या डेटासेटच्या मदतीने मराठीतील भावनिक बारकावे ओळखण्याचे प्रयत्न सुरू आहेत, ज्यामध्ये जीपीटी-४ सारख्या मॉडेल्सने स्थानिक मॉडेल्सपेक्षा चांगली कामगिरी केली आहे.

**ऐतिहासिक वारसा जतन: ओसीआर आणि मोडी लिपीचे डिजिटायझेशन :**

मराठी भाषेचा इतिहास केवळ साहित्यातच नव्हे, तर हजारो शिलालेख आणि हस्तलिखितांमध्ये सामावलेला आहे. यामध्ये 'मोडी' लिपीचा वाटा मोठा आहे. मोडी लिपी ही ऐतिहासिक काळात प्रशासकीय कामासाठी वापरली जात असे, मात्र काळाच्या ओघात ती वाचणारे लोक कमी झाले आहेत. कृत्रिम

बुद्धिमतेचा वापर करून ही जुनी हस्तलिखिते वाचणे आणि त्यांचे आधुनिक देवनागरी मराठीत रूपांतर करणे हा संशोधनाचा एक महत्त्वाचा भाग बनला आहे.

## ऐतिहासिक कागदपत्रांच्या डिजिटायझेशनमधील प्रमुख तांत्रिक आव्हाने:

- **कागदाची झीज:** जुन्या हस्तलिखितांमधील शाई पसरलेली असते किंवा कागद फाटलेले असतात, ज्यामुळे मजकूर स्पष्ट दिसत नाही.
- **अक्षरांचे गुंफण:** मोडी लिपीमध्ये अक्षरे एकमेकांना जोडून लिहिली जातात (Overlapping), ज्यामुळे ओसीआर सिस्टिमला अक्षरे वेगळी करणे कठीण जाते.
- **हस्ताक्षरातील विविधता:** प्रत्येक व्यक्तीची मोडी लिहिण्याची शैली वेगळी असल्याने युनिव्हर्सल मॉडेल तयार करणे कठीण असते.

यावर उपाय म्हणून संशोधकांनी 'कन्नाव्होल्युशनल न्यूरल नेटवर्क्स' (CNN) आणि 'वेवलेट ट्रान्सफॉर्म' (Wavelet Transform) या तंत्रज्ञानाचा वापर केला आहे. याद्वारे ऐतिहासिक शिलालेखांमधील अक्षरे ओळखण्यात ८८.९५% पर्यंत अचूकता गाठता आली आहे. 'वाक्यसार' (VakyaSaar) आणि इतर सारख्या साधनांनी प्राचीन मराठी साहित्याचे संक्षेपीकरण आणि आधुनिक मराठीत भाषांतर करण्याचे काम सुलभ केले आहे.

## बोलीभाषांचे संवर्धन आणि तांत्रिक समावेशकता :

मराठी भाषेची खरी समृद्धी तिच्या बोलीभाषांमध्ये आहे. वऱ्हाडी, अहिराणी, कोकणी, मालवणी, झाडीबोली अशा अनेक बोलीभाषा महाराष्ट्राच्या कानाकोपऱ्यात बोलल्या जातात. मात्र, बहुतांश तांत्रिक विकास हा 'प्रमाण मराठी'वर (Standard Marathi) केंद्रित आहे. यामुळे ग्रामीण भागातील वक्ते डिजिटल क्रांतीपासून दूर राहण्याचा धोका निर्माण झाला आहे.

या भाषिक दरीला सांधण्यासाठी कृत्रिम बुद्धिमत्तेचा वापर करून बोलीभाषा ते प्रमाण मराठी अनुवाद प्रणाली विकसित केल्या जात आहेत. 'रिट्रायव्हल-ऑगमेंटेड जनरेशन' (RAG) आधारित सिस्टिम्स अशा वापरकर्त्यांना मदत करत आहेत जे त्यांच्या स्थानिक भाषेत प्रश्न विचारतात. उदाहरणार्थ, एखाद्या शेतकऱ्याने अहिराणीमध्ये विचारलेला प्रश्न एआय प्रथम प्रमाण मराठीत अनुवादित करते, सरकारी कागदपत्रांमधून त्याचे उत्तर शोधते आणि पुन्हा ते सोप्या मराठीत किंवा संबंधित बोलीभाषेत देते.

### बोलीभाषांच्या संशोधनातील प्रगती खालीलप्रमाणे आहे:

- **व्हाडी आणि नागपुरी:** यूट्यूब आणि इंस्टाग्राम सारख्या प्लॅटफॉर्मवर या बोलींमधील आशयाचे प्रमाण वाढले असून एआय मॉडेल्स या डेटाचा वापर बोली ओळखीसाठी (Dialect Identification) करत आहेत.
- **कोकणी (देवनागरी):** जेमिनी (Gemini API) सारख्या मॉडेल्सचा वापर करून कोकणी प्रश्नांचे अचूक प्रमाण मराठीत रूपांतर करण्यात यश आले आहे.
- **मल्टी-टास्क लर्निंग:** एकाच वेळी अनेक बोलीभाषांचे व्याकरण शिकणारी मॉडेल्स विकसित केली जात आहेत, ज्यामुळे कमी डेटा असलेल्या बोलीभाषांसाठीही एआय तयार करता येईल.

### ई-प्रशासन आणि एआय चॅटबॉट्स: नागरिक सेवांचे मराठीकरण :

महाराष्ट्र सरकारने प्रशासनात पारदर्शकता आणि वेग आणण्यासाठी एआय चॅटबॉट्सचा प्रभावी वापर सुरू केला आहे. 'आपले सरकार' (Aaple Sarkar) हा चॅटबॉट याचे उत्तम उदाहरण आहे. हॅप्टिक (Haptik) या कंपनीच्या सहकार्याने विकसित केलेला हा बॉट १४०० हून अधिक सरकारी सेवांची माहिती मराठीत देतो.

अलिकडेच, महाराष्ट्र सरकारने 'मेटा' (Meta) सोबत भागीदारी करून व्हाॅट्सअॅपवर आधारित नागरिक सेवा पोर्टल लॉन्च केले आहे. हे पोर्टल 'लामा' (Llama) या ओपन-सोर्स जनरेटिव्ह एआयवर

आधारित असून ते १२५ दशलक्ष नागरिकांना मराठी, हिंदी आणि इंग्रजी भाषेत मदत करते. नागरिक आता व्हॉट्सअॅपवर मेसेज करून किंवा बोलून आपली तक्रार नोंदवू शकतात किंवा बस तिकीट बुकिंग सारख्या सेवा मिळवू शकतात.

याशिवाय, नाशिक विभागीय आयुक्त कार्यालयाने 'जीआर' (GR) डेटावर आधारित चॅटबॉट सुरू केला आहे, जो गेल्या ७ वर्षांतील सरकारी निर्णयांची माहिती नागरिकांना मराठीत उपलब्ध करून देतो. सायबर सुरक्षेच्या क्षेत्रात 'महाक्राइम ओएस एआय' (Maha CrimeOS AI) हे मायक्रोसॉफ्टच्या मदतीने विकसित केलेले व्यासपीठ पोलीस अधिकाऱ्यांना मराठीतील एफआयआर आणि पुराव्यांचे विश्लेषण करण्यास मदत करत आहे.

### सरकारी एआय उपक्रमांचा प्रभाव खालील तक्त्यात पाहता येईल:

उपक्रम	तंत्रज्ञान	मुख्य लाभ
आपले सरकार बॉट	हॅटिक एआय (Conversational AI)	सरकारी सेवांची माहिती २४/७ उपलब्ध.
व्हॉट्सअॅप नागरिक सेवा	मेटा लामा (Llama-based Gen AI)	घरबसल्या कागदपत्रे आणि तक्रार निवारण.
महाक्राइम ओएस	अझर ओपनएआय (Azure OpenAI)	सायबर गुन्हांचा जलद तपास आणि बहुभाषिक पुरावे विश्लेषण.
महावेध (Mahavedh)	एआय आणि मशीन लर्निंग	शेतकऱ्यांना हवामान आणि पीक विमा सल्ला.

## मराठी साहित्याचे संक्षेपीकरण आणि माध्यम साधने :

माहितीचा स्फोट होत असलेल्या युगात प्रदीर्घ बातम्या किंवा लेख वाचणे कठीण असते. अशा वेळी एआय आधारित 'टेक्स्ट समरायझेशन' (Text Summarization) टूल्स मराठी वाचकांसाठी अत्यंत उपयुक्त ठरत आहेत. 'वाक्यसार' (VakyaSaar) हे अशा प्रकारचे एक प्रगत टूल आहे जे 'इंडिकबार्ट' (IndicBART) आणि 'टेक्स्ट रँक' (TextRank) अल्गोरिदमचा वापर करून मोठ्या मजकुराचे अर्थपूर्ण संक्षेप करते. वाक्यसारांशाच्या संशोधनात असे दिसून आले की:

- **एक्स्ट्रॅक्टिव्ह पद्धत:** मजकुरातील सर्वात महत्त्वाचे शब्द निवडून सारांश तयार करणे (Accuracy: 3.9/5).
- **अॅबस्ट्रॅक्टिव्ह पद्धत:** मजकुराचा अर्थ समजून घेऊन नवीन शब्दांत सारांश लिहिणे (Accuracy: 4.4/5).
- **यूजर फीडबॅक:** वापरकर्त्यांच्या प्रतिक्रिया घेऊन हे मॉडेल सतत स्वतःला सुधारत आहे, ज्यामुळे मराठीतील बातम्यांचे अचूक संक्षेपीकरण करणे शक्य झाले आहे.

याशिवाय, सोशल मीडियावरील मराठी कमेंट्समधील शिवीगाळ किंवा आक्षेपार्ह मजकूर ओळखण्यासाठी 'ट्री अल्गोरिदम' (Trie Algorithm) आणि 'मल्टीलिंग्युअल बर्ट' (MBERT) यांचा संकर करून ९२.४% अचूकता प्राप्त करण्यात आली आहे. यामुळे मराठी डिजिटल अवकाश अधिक सुरक्षित आणि शुद्ध राखण्यास मदत होत आहे.

## कृत्रिम बुद्धिमत्ता आणि भाषिक अस्मिता: सामाजिक परिणाम :

एआयचा वाढता वापर मराठी भाषेच्या शुद्धतेवर आणि अस्मितेवर कसा परिणाम करेल, हा चिंतेचा आणि चर्चेचा विषय आहे. 'जेन-झेड' (Gen Z) पिढी डिजिटल प्लॅटफॉर्मवर मराठी आणि इंग्रजीचे मिश्रण

(Code-mixing) मोठ्या प्रमाणावर करते. एआय मॉडेल्स हे कोड-मिक्सिंग स्वीकारतात, ज्यामुळे भविष्यात शुद्ध मराठीच्या वापरात घट होईल की काय, अशी भीती व्यक्त केली जाते.

मात्र, दुसऱ्या बाजूला एआय हे मराठी भाषेला 'ज्ञानाची भाषा' (Language of Knowledge) बनवण्याचे साधन म्हणून समोर आले आहे. मुख्यमंत्री देवेंद्र फडणवीस यांनी 'विश्व मराठी संमेलन' प्रसंगी नमूद केल्याप्रमाणे, जर आपण मराठी साहित्याचा खजिना 'स्मॉल लॅंग्वेज मॉडेल' (SLM) मध्ये बंदिस्त केला, तर भावी पिढी चॅटजीपीटी सारख्या साधनांमधून ज्ञानेश्वर, तुकाराम किंवा सावरकरांचे साहित्य त्यांच्या भाषेत अनुभवू शकेल.

### एआयमुळे भाषिक संवर्धनाचे काही सकारात्मक पैलू:

- **लोकशाहीकरण:** दुर्गम गावातील व्यक्ती एआयच्या मदतीने जागतिक ज्ञान मराठीत मिळवू शकते.
- **रोजगार संधी:** मराठी डेटा लेबलिंग, एआय ट्रेनिंग आणि कंटेंट क्युरेशन क्षेत्रात नवीन नोकऱ्या निर्माण होत आहेत.
- **सांस्कृतिक पुनरुज्जीवन:** जुन्या लोककथा आणि पोवाडे एआयच्या मदतीने ऑडिओ-व्हिड्युअल स्वरूपात नवीन पिढीसमोर येत आहेत.

### भविष्यातील दिशा आणि शिफारसी :

मराठी भाषा संवर्धनासाठी कृत्रिम बुद्धिमत्तेचा वापर हा केवळ एक प्रयोग नसून ती एक चळवळ असणे आवश्यक आहे. संशोधनाचे पुढील टप्पे हे अधिक समावेशक आणि तांत्रिकदृष्ट्या प्रगत असावे लागतील.

१. **स्मॉल लॅंग्वेज मॉडेल्सचा विकास (SLM):** मोठ्या जागतिक मॉडेल्सवर अवलंबून राहण्यापेक्षा मराठी साहित्यावर आधारित स्थानिक मॉडेल्स विकसित करणे, जे सांस्कृतिकदृष्ट्या अधिक अचूक असतील.

२. **बोलीभाषा डेटा बँक:** महाराष्ट्रातील प्रत्येक बोलीभाषेचा ऑडिओ आणि टेक्स्ट डेटा संकलित करणे, जेणेकरून 'व्हॉइस असिस्टंट' प्रत्येक जिल्ह्याच्या भाषेत उत्तर देऊ शकतील.
३. **शिक्षण प्रणालीत समावेश:** शाळा-कॉलेजांमध्ये एआय साधनांचा वापर करून मराठी व्याकरण आणि लेखन शिकवण्यावर भर देणे.
४. **डिजिटल लायब्ररीचे विस्तारीकरण:** केवळ पुस्तकेच नव्हे, तर ऐतिहासिक कागदपत्रे आणि शिलालेख यांचे ओसीआर आधारित विस्तीर्ण भांडार तयार करणे.

मराठी भाषेने गेल्या २ हजार वर्षांत अनेक आक्रमणे झेलली आणि स्वतःला समृद्ध केले. आता तंत्रज्ञानाचे हे 'आक्रमण' नसून ते भाषेच्या 'विस्ताराची' संधी आहे. कृत्रिम बुद्धिमत्तेचा योग्य वापर केल्यास मराठी केवळ महाराष्ट्राचीच नव्हे, तर डिजिटल जगातील एक अग्रगण्य ज्ञानभाषा म्हणून प्रस्थापित होईल. मराठी भाषेचे संवर्धन हे केवळ सरकारचे काम नसून ते प्रत्येक मराठी भाषिक तंत्रज्ञ आणि अभ्यासकाचे सामूहिक कर्तव्य आहे. कृत्रिम बुद्धिमत्तेच्या सहाय्याने आपण आपल्या भाषेचा वारसा भावी पिढ्यांसाठी अधिक प्रखर आणि प्रगत स्वरूपात हस्तांतरित करू शकतो.

#### आधार ग्रंथ :

- १) गोडबोले अच्युत - *आर्टिफिशियल इंटेलिजन्स* - मधुश्री पब्लिकेशन, पुणे - प्र.आ. २०२१
- २) निलांबरी जोशी - *आर्टिफिशियल इंटेलिजन्स स्विकार, नकार का सहकार* - मनोविकास प्रकाशन, पुणे प्र.आ. २०२४
- ३) डॉ. दीपक शिकारपूर - *आर्टिफिशियल इंटेलिजन्स (AI)* - दिलीपराज प्रकाशन, पुणे दु. आ. २०२४
- ४) <https://www.ijcaonline.org>



- ५) <https://www.cadc.in>
- ६) <https://www.ijirset.com>
- ७) <https://www.ijserm.net>
- ८) <https://bhashini.gov.in>
- ९) <https://ictpost.com>
- १०) <https://indiaai.gov.in>

