



"मराठी भाषेत कृत्रिम बुद्धिमत्तेच्या सहाय्याने भाषा अधिगम आणि अनुवाद प्रणालींचा विकास"

प्रा. डॉ. प्रभाकर जे. जांभुळे

कला, वाणिज्य महाविद्यालय, येवदा,

ता. दर्यापूर, जि. अमरावती, महाराष्ट्र,

ईमेल: pjjambhule38@gmail.com

सारांश (Abstract):

हा संशोधन लेख मराठी भाषेसाठी कृत्रिम बुद्धिमत्तेच्या (AI) आधारे भाषा अधिगम आणि अनुवाद प्रणालींच्या विकासाचा सखोल अभ्यास करतो. मराठी ही समृद्ध सांस्कृतिक आणि भाषिक वारसा असलेल्या भाषांपैकी एक असून तिच्या डिजिटल व शैक्षणिक क्षेत्रातील मर्यादा आणि आव्हाने भेदून निघण्यासाठी AI आणि नैसर्गिक भाषा प्रक्रिया (NLP) सारख्या तंत्रज्ञानांचा उपयोग कसा होऊ शकतो, यावर प्रकाश टाकतो. पारंपरिक भाषा शिकवणी व अनुवाद पद्धतींच्या मर्यादा मांडून AI आधारित संवादात्मक व वैयक्तिकृत भाषा अधिगम प्रणाली, तसेच अचूक आणि वेगवान स्वयंचलित अनुवाद यांचा महत्त्वाचा वापर आणि त्याचे फायदे तपासले आहेत. या तंत्रज्ञानाच्या अंगभूत मर्यादा, जसे की डेटा अभाव, बोलीभाषांचा वैविध्य, सामाजिक-सांस्कृतिक अडथळे आणि नैतिक समस्यांबाबत देखील चर्चा केली आहे. भविष्यातील संशोधनासाठी मराठी भाषेतील संसाधन साठवण, AI संशोधनाला पाठिंबा, आणि शैक्षणिक तसेच सांस्कृतिक क्षेत्रातील AI आधारित तंत्रज्ञानाच्या समावेशाच्या संधींचा आढावा घेतला आहे. AI च्या साहाय्याने मराठी भाषा शिकणाऱ्यांना अधिक सुलभ शिक्षण, प्रभावी संवाद आणि जागतिक पातळीवरील पोहोच मिळवून देऊन, भाषेच्या टिकाऊपणा आणि आधुनिकतेसाठी एक नवा मार्ग मोकळा होईल.

बीज शब्द (Keywords): मराठी भाषा, कृत्रिम बुद्धिमत्ता, नैसर्गिक भाषा प्रक्रिया, भाषा अधिगम, मशीन ट्रान्सलेशन, न्यूरल मॉडेल्स, डिजिटल संरक्षण, भाषा तंत्रज्ञान, AI नैतिकता, बहुभाषिक संवाद

१. प्रस्तावना

मराठी भाषा ही महाराष्ट्र राज्यातील एक प्राचीन आणि समृद्ध सांस्कृतिक व भाषिक परंपरेची धारक आहे. सुमारे ८ कोटी लोकांची मातृभाषा असलेल्या या भाषेचे साहित्यिक, वैचारिक आणि सामाजिक क्षेत्रात विशेष स्थान आहे (Deshpande, 2017). जग पुढे जात असताना, तंत्रज्ञानाच्या प्रगत अध्ययन प्रणाली आणि संवादात्मक साधनांची गरज वाढली आहे. भाषेचा अभ्यास, शिक्षण आणि जागतिक संपर्क यांमध्ये कृत्रिम बुद्धिमत्तेचा (Artificial Intelligence – AI) वापर करण्यात येत आहे, विशेषतः नैसर्गिक भाषा प्रक्रिया (Natural Language Processing – NLP) तंत्रज्ञानामुळे (Jurafsky & Martin, 2020). मराठीसारख्या मोठ्या प्रमाणात बोलल्या जाणाऱ्या स्थानिक भाषेसाठी AI-आधारित भाषा अधिगम आणि अनुवाद प्रणालींचा विकास हा या भाषेच्या सर्वांगीण प्रगतीसाठी एक महत्त्वाचा टप्पा आहे (Patil & Kulkarni, 2021). AI-आधारित तंत्रज्ञानाने भाषिक संवादाला नवी गती दिली आहे, ज्यामुळे भौगोलिक आणि सांस्कृतिक अडथळे कमी होतात (Sundararajan, 2019). मराठीसाठी अशी तंत्रे अद्याप प्राथमिक अवस्थेत असली तरी, येत्या काळात त्यांचा अभ्यास, विकास आणि व्यावहारिक उपयोग मोठ्या प्रमाणात वाढण्याची शक्यता आहे. AI-आधारित

संसाधने आणि साधने मराठी भाषिक समुदायाला केवळ शिक्षणाचे नवे आणि सुलभ माध्यम पुरवत नाहीत, तर स्थानिक व जागतिक पातळीवर संवादाचे नवीन व्यासपीठदेखील उपलब्ध करून देतात. या लेखाचा प्रमुख हेतू मराठी भाषेत AI-च्या साहाय्याने भाषा अधिगम प्रणाली आणि अनुवाद तंत्रज्ञानाचा विकास, त्यांची कार्यक्षमता, शैक्षणिक व सामाजिक परिणाम, तसेच या क्षेत्रातील आव्हाने आणि भविष्यातील संधी यांचा समग्र अभ्यास करणे हा आहे. यामुळे केवळ मराठी भाषेची शैक्षणिक प्रगतीच नव्हे, तर डिजिटल युगात स्थानिक भाषांचा सांस्कृतिक वारसा जतन करणेही शक्य होईल (Deshpande, 2017).

२. कृत्रिम बुद्धिमत्ता: एक विस्तृत परिचय

कृत्रिम बुद्धिमत्ता (Artificial Intelligence - AI) हे संगणकीय प्रणालींना मानवी मेंदूच्या कार्यक्षमतेप्रमाणे विचार करण्याची, शिकण्याची व निर्णय घेण्याची क्षमता प्रदान करणारे शास्त्र आणि तंत्रज्ञान आहे (Russell & Norvig, 2021). AI-च्या मुख्य घटकांमध्ये मशीन लर्निंग आणि डीप लर्निंग यांचा समावेश होतो. मशीन लर्निंगमध्ये संगणक विशिष्ट डेटावरून शिकून आपली कार्यक्षमता सुधारतो, तर डीप लर्निंगमध्ये हार्डवेअर व सॉफ्टवेअरच्या साहाय्याने अधिक जटिल नेटवर्क तयार करून माहितीचे सखोल विश्लेषण केले जाते (Goodfellow et al., 2016).

नैसर्गिक भाषा प्रक्रिया (NLP) ही AI-ची एक शाखा असून, ती संगणक प्रोग्रॅम्सना मानवी भाषा समजण्यासाठी, तिचा अर्थ लावण्यासाठी व संवाद साधण्यासाठी सक्षम बनवते (Manning et al., 2020). NLP-च्या परिणामस्वरूप संगणक अनुवाद, भाषिक विश्लेषण, भावना ओळख आणि संवादात्मक प्रणाली तयार करण्यास सक्षम होतो (Jurafsky & Martin, 2020). यामुळे भाषा अधिगमातील अडचणी कमी होतात आणि सुव्यवस्थित संवाद साधता येतो.

मराठीसारख्या स्थानिक आणि कमी संसाधने असलेल्या भाषांसाठी NLP विकसित करणे हे अनेक आव्हानांनी भरलेले आहे. मराठीची व्याकरणशास्त्रीय रचना, विविध बोलीभाषा आणि सांस्कृतिक संदर्भ तंत्रज्ञानासाठी विशेष अडचणी निर्माण करतात (Patil & Kulkarni, 2021). त्याचप्रमाणे, सहज उपलब्ध डेटाचा अभाव आणि भाषिक संपदेचे कमी प्रमाणातील डिजिटलीकरण हीदेखील मोठी आव्हाने आहेत. याशिवाय, भाषिक संवादातील विविधता आणि उच्चारांतील फरक AI मॉडेल्सना प्रशिक्षित करताना अडथळा आणतात.

AI-मध्ये ट्रान्सफॉर्मर-आधारित मॉडेल्सनी (उदा. BERT, GPT) भाषेच्या आकलनात क्रांती घडवून आणली आहे (Vaswani et al., 2017). या मॉडेल्सचा वापर करून मराठी भाषेच्या अनुवाद आणि अधिगम क्षेत्रातही सखोल सुधारणा साधता येऊ शकते. विशेषतः, बहुभाषिक संवाद, ज्ञानप्रसार आणि औद्योगिक कामकाजातील संवाद सुलभ करण्यासाठी या तंत्रज्ञानाचे महत्त्व वाढत आहे (Sundararajan, 2019). स्वयंचलित भाषांतर, स्मार्ट ट्युटर्स, ध्वनी ओळख आणि भाषेचा अर्थ समजून संवाद साधण्यासाठी AI-चा वापर झाल्यामुळे मराठीसारख्या भाषांसाठी नव्या प्रकारचे शिक्षण आणि संवादाचे मार्ग खुले होत आहेत. यामुळे शैक्षणिक, सामाजिक आणि आर्थिक विकासाला गती मिळण्याची शक्यता निर्माण होत आहे (Russell & Norvig, 2021).

३. मराठी भाषेतील भाषा अधिगम: पारंपरिक पद्धती आणि मर्यादा

मराठी भाषेला एक समृद्ध सांस्कृतिक आणि सहस्रकांचा इतिहास लाभलेला आहे आणि तिचा शिक्षणविषयक पारंपरिक प्रवासही तितकाच विस्तृत आहे (Deshpande, 2017). पारंपरिकदृष्ट्या, मराठी भाषा गुरुकुल पद्धतीतून तसेच शाळांमध्ये ग्रंथाधारित शिक्षण प्रणालीद्वारे शिकवली गेली आहे, जिथे शिक्षक-केंद्रित शिक्षणावर अधिक भर होता (Kulkarni, 2014). या पद्धतीमध्ये मुख्यतः व्याकरण, पाठांतर, लेखन आणि

पारंपरिक साहित्य समजावून सांगितले जात असे. मात्र, संवादात्मक तंत्रज्ञान किंवा साधनांच्या अभावामुळे भाषेच्या बोली आणि संवादात्मक अंगाच्या अभ्यासाकडे तुलनेने दुर्लक्ष झाले (Bharne & Deshpande, 2018).

पारंपरिक शिक्षण पद्धतीमुळे मराठी भाषा शिकणाऱ्यांना अनेक मर्यादांचा सामना करावा लागला आहे. विशेषतः, जे महाराष्ट्राबाहेरील किंवा मराठी न बोलणाऱ्या समाजातील आहेत, त्यांच्यासाठी या मर्यादा अधिक स्पष्ट होतात (Gokhale, 2015). हे शिक्षण आशयाच्या सैद्धांतिक बाजूस अधिक केंद्रित असल्याने, संवादात्मक आणि व्यावहारिक भाषा कौशल्ये विकसित करणे अतिशय कठीण जाते (Shende, 2019).

तसेच, मराठी भाषेतील बोलीभाषांच्या विविधतेमुळे भाषेच्या सुसंगततेत आणि आकलनात अडथळे येतात. उदाहरणार्थ, कोकणी, आगरी (ठाणे) आणि नगरकडील बोलीभाषांमध्ये उच्चार आणि शब्दसंग्रहात मोठा फरक आढळतो, ज्यामुळे भाषेचा प्रभावी अधिगम सोपा होत नाही (Deshpande, 2017). व्याकरणीय नियमांचे आणि संवेदनशील सांस्कृतिक संदर्भांचे अपूर्ण आकलन शिक्षण प्रक्रियेत गुंतागुंत निर्माण करते. हे दिसून येते की, पारंपरिक शिक्षण पद्धती अर्थपूर्ण संवाद आणि अभ्यासात मर्यादित राहिल्या आहेत, ज्यामुळे प्रदेशानुसार भाषा आत्मसात करणे आणि संवाद कौशल्याचा विकास खुंटतो (Bharne & Deshpande, 2018). अधिक महत्वाचे म्हणजे, पारंपरिक शिक्षण पद्धती भाषा संपादनाची गती मंदावतात आणि विशेषकरून ज्यांच्यासाठी मराठी ही दुसरी किंवा तिसरी भाषा आहे, त्यांच्यासाठी त्या अपुऱ्या ठरतात (Gokhale, 2015).

आजच्या तांत्रिक युगात, पारंपरिक शिक्षक-केंद्रित मॉडेल्सच्या जागी संवादात्मक, वैयक्तिकृत आणि तात्काळ अभिप्राय देणाऱ्या प्रणालींची गरज आहे. पारंपरिक शिक्षणात तंत्रज्ञानाच्या अभावामुळे नैसर्गिक साक्षरता आणि संवादात्मक स्वायत्तता निर्माण होण्यात अडथळा येतो (Shende, 2019). त्यामुळे AI आणि नैसर्गिक भाषा प्रक्रिया (NLP) आधारित तंत्रज्ञानाचा वापर करणे अधिक आवश्यक ठरते, जे शैक्षणिक व सामाजिक क्षेत्रात भाषिक कौशल्ये अधिक प्रभावीपणे विकसित करेल (Chen et al., 2021).

४. AI-आधारित मराठी भाषा अधिगम प्रणाली

कृत्रिम बुद्धिमत्तेच्या माध्यमातून भाषा अधिगम हे क्षेत्र जागतिक स्तरावर झपाट्याने प्रगती करत आहे आणि त्यामुळे मराठीसाठीही AI-आधारित नव्या अधिगम प्रणाली तयार होत आहेत (Li et al., 2019). AI-आधारित भाषा अधिगम प्रणाली वापरकर्त्यांच्या गरजा, स्तर आणि भाषिक क्षमतांनुसार वैयक्तिकृत शिक्षण प्रदान करतात, जे पारंपरिक शिक्षण पद्धतींपेक्षा अधिक फलदायी ठरते (Kukafka, 2020).

AI तंत्रज्ञानाने संवादात्मक चॅटबॉट्स, व्हर्च्युअल ट्युटर्स आणि स्पीच-टू-टेक्स्ट प्रणालींद्वारे भाषा अधिगम अधिक आकर्षक आणि परिणामकारक बनवले आहे (Bengio et al., 2021). अशा प्लॅटफॉर्म्सवर भाषिक चुका त्वरित ओळखल्या जातात आणि योग्य सुधारणा करून वापरकर्त्याला अभिप्राय दिला जातो, ज्यामुळे भाषा आत्मसात करण्याची प्रक्रिया अधिक जलद आणि नैसर्गिक होते (Thompson & Neubig, 2020). AI-आधारित उच्चार सुधारणा तंत्रांचा वापर करून मराठी उच्चार आणि बोलींचे प्रमाणित मॉडेल तयार करण्याची प्रक्रिया सुरू आहे, जे परंप्रांतीय किंवा परदेशी मराठी शिकणाऱ्यांसाठी खूप उपयुक्त आहे (Ramesh et al., 2022).

अधिक व्यापक आणि गुंतागुंतीच्या भाषिक पैलूंसाठी, मशीन लर्निंग आणि डीप लर्निंग तंत्रज्ञानाचा वापर करून भाषेतील व्याकरण, वाक्यरचना, भाषिक संदर्भ आणि शैलींचा सखोल अभ्यास करून अधिक अचूक मॉडेल्स विकसित केली जात आहेत (Goodfellow et al., 2016). हे मॉडेल्स शाळा, महाविद्यालये आणि स्व-अध्ययनासाठी डिजिटल ट्युटर्स तयार करण्यात वापरले जात आहेत, जे मराठी शिकण्याचा अनुभव अधिक समृद्ध करतात (Chen et al., 2021).

भारतीय भाषांसाठी बहुभाषिक आणि स्थानिक भाषांसाठी AI-आधारित अधिगम प्लॅटफॉर्मची निर्मिती झाल्यामुळे मराठीही त्याचा भाग बनत आहे (Patil & Kulkarni, 2021). जागतिक स्तरावरील संशोधनानुसार, या प्रणालींमध्ये स्वर, व्याकरण, बोलीभाषा आणि संस्कृती यांचा एकत्रित विचार केला जातो. त्यामुळे त्या अधिक व्यापक, संवेदनशील व प्रभावी ठरतात. अशा प्रणाली मराठी शिक्षक आणि विद्यार्थ्यांना प्रशिक्षण देण्यासाठी सुव्यवस्थित आधार पुरवतात (Ramesh et al., 2022).

तथापि, या प्रणालींचा प्रभावी आणि व्यापक वापर सुनिश्चित करण्यासाठी गुणवत्तापूर्ण डेटा संग्रहण आणि स्थानिक भाषिक समुदायाचा सहभाग अत्यंत आवश्यक आहे (Patil & Kulkarni, 2021). या संदर्भात सरकार, शैक्षणिक संस्था आणि संशोधन केंद्रांनी एकत्र येऊन AI-आधारित मराठी भाषा अधिगमासाठी योग्य धोरणे आणि तंत्रज्ञानाची पायाभरणी करणे महत्वाचे ठरेल.

५. मराठी भाषेतील AI-आधारित अनुवाद प्रणाली

जगभरातील भाषिक संवादाच्या गरजांना प्रतिसाद देण्यासाठी आणि डिजिटल युगातील भाषिक अडथळे दूर करण्यासाठी कृत्रिम बुद्धिमत्ता (AI) समर्थित स्वयंचलित अनुवाद (Machine Translation - MT) प्रणाली उदयास आल्या आहेत. यातील ट्रान्सफॉर्मर-आधारित मॉडेल्सनी (Transformer Models), जसे की BERT (Devlin et al., 2019) आणि GPT (Brown et al., 2020), भाषा आकलन आणि अनुवादाच्या क्षेत्रात नव्या संशोधनांना चालना दिली आहे. मराठीसाठीदेखील या तंत्रज्ञानाचा वापर वेगाने वाढत असून, इंग्रजी आणि हिंदीसारख्या भाषांमधून अधिक अचूक अनुवाद करणे शक्य होत आहे (Ramesh et al., 2022).

पारंपरिक नियमाधारित आणि शब्दकोश-आधारित अनुवाद पद्धतींमध्ये भाषांतरातील संदर्भ व शैलीगत सूक्ष्मता जपली जात नव्हती (Koehn, 2020). मात्र, AI-आधारित न्यूरल मशीन ट्रान्सलेशन (NMT) प्रणाली मोठ्या प्रमाणात भाषिक डेटाचा अभ्यास करतात. परिणामी, ही प्रणाली संभाषणातील अर्थ, शब्दशैली आणि सांस्कृतिक संदर्भ यांचा विचार करून अधिक स्वाभाविक अनुवाद करते (Jiang et al., 2020). मराठी भाषेची समृद्ध शब्दसंपदा, व्याकरणीय गुंतागुंत, वाक्यरचना आणि विविध बोलीभाषांमधील भिन्नता ही तिची वैशिष्ट्ये आहेत (Mukherjee & Dasgupta, 2019). उदाहरणार्थ, मराठीमध्ये विभक्तीप्रत्ययांचा वापर मोठ्या प्रमाणात असतो, ज्यामुळे हजारो शब्दांमधून योग्य अर्थ काढणे AI मॉडेल्ससाठी आव्हानात्मक ठरते. त्याचप्रमाणे, शब्दार्थ, शब्दांचे विविध अर्थ आणि सांस्कृतिक संदर्भ समजण्यासाठी अद्यापही परिष्कृत डेटासेटची गरज आहे (Patil & Kulkarni, 2021). योग्य भाषिक डेटासेटचा अभाव ही एक प्रमुख अडचण आहे, ज्यामुळे उच्च दर्जाच्या मराठी AI अनुवाद प्रणाली विकसित करणे कठीण झाले आहे (Sundararajan, 2019). अशा डेटासेटमध्ये लिखित आणि बोली भाषेतील घटक, संदर्भ आणि विविध संभाषणांच्या उदाहरणांचा समावेश असणे आवश्यक आहे. अनेक संशोधक स्थानिक लोकांच्या सहयोगाने क्राउडसोर्सिंग आणि भाषिक कॉर्पस तयार करून या आव्हानांवर मात करण्याचा प्रयत्न करत आहेत (Ramesh et al., 2022).

AI-आधारित अनुवाद प्रणालींचा प्रभाव केवळ तांत्रिकच नव्हे, तर सामाजिक, शैक्षणिक आणि व्यावसायिक क्षेत्रांतही दिसून येतो. विविध सरकारी आणि खासगी क्षेत्रांमध्ये बहुभाषिक यंत्रणा निर्माण करून त्यांचा वापर सरकारी आदेश, शैक्षणिक साहित्य इत्यादींचे वेगाने भाषांतर करण्यासाठी होत आहे (Koehn, 2020). यामुळे भाषिक अडथळे कमी होऊन व्यापक संवाद सुलभ होतो. बहुभाषिक भारतात मराठीसारख्या स्थानिक भाषांसाठी AI-आधारित अनुवाद संशोधनाचे महत्त्व अधिक वाढले आहे.

आंतरराज्यीय आणि आंतरराष्ट्रीय व्यावसायिक संवाद, पर्यटन उद्योग आणि ऑनलाइन शिक्षण क्षेत्रात AI अनुवाद प्रणालींचा वापर वाढवणे ही एक महत्वाची संधी आहे (Sundararajan, 2019). अनुवाद तंत्रज्ञानामुळे

मराठी साहित्य, कला आणि सांस्कृतिक वारशाचा जागतिक स्तरावर प्रसार करणे शक्य होते, ज्यामुळे स्थानिक भाषेचे महत्त्व जागतिक संदर्भातही वाढते. तथापि, मराठी AI-आधारित अनुवाद प्रणालीसमोरील तांत्रिक, सामाजिक व सांस्कृतिक मर्यादा पार करणे आवश्यक आहे. यासाठी तंत्रज्ञानातील संशोधनाची व्याप्ती वाढवणे, स्थानिक भाषिक समुदायाकडून अधिक प्रमाणात डेटा गोळा करणे आणि तंत्रज्ञानाच्या नैतिक व सामाजिक पैलूंवर सखोल चर्चा करण्याची गरज आहे (Patil & Kulkarni, 2021).

६. AI प्रणालींचे मराठी भाषेच्या अभ्यासावर, सांस्कृतिक वारशावर आणि लेखनावर होणारे परिणाम

AI प्रणालींनी मराठी भाषेच्या अभ्यासात आणि सांस्कृतिक क्षेत्रात क्रांतिकारी बदल घडवून आणला आहे (Jockers, 2013). 'डिजिटल ह्यूमॅनिटीज' या शाखेत AI-आधारित टेक्स्ट विश्लेषण, भावना ओळख (sentiment analysis) आणि भाषिक नेटवर्क विश्लेषण (linguistic network analysis) यांमुळे मराठी साहित्यातील संख्याशास्त्रीय तसेच शैलीगत पैलूंचा अभ्यास सुलभ झाला आहे. पूर्वी अनेक महिने लागणारी कामे AI तंत्रज्ञानामुळे काही तासांत पूर्ण होतात (Bamman et al., 2014).

भावना ओळख तंत्रज्ञानाच्या साहाय्याने मराठी काव्य, कथा, नाटके यांतील भावनिक बदल आणि सामाजिक संदर्भ समजून घेता येतात. विविध कालखंडांतील लेखनशैलीतील फरक, विषयांची उत्क्रांती आणि सामाजिक संदेश यांचा अभ्यास अधिक सुलभ झाल्यामुळे भाषा आणि साहित्याचा सखोल अभ्यास शक्य होत आहे (Bamman et al., 2014). ही तंत्रे संशोधकांना भाषिक इतिहास, सामाजिक बदल आणि सांस्कृतिक वारशाचा मागोवा घेण्यास मदत करतात. लेखन प्रक्रियेतही AI तंत्रज्ञानाने मोठे साहाय्य केले आहे. लेखन-सहाय्यक ॲप्लिकेशन्स AI-द्वारे व्याकरण तपासणी, शुद्धलेखन आणि शैली सुधारण्यासारख्या सेवा देऊन लेखकांना अधिक प्रभावी साहित्य निर्मितीची संधी देतात (Kukafka, 2020). विविध शैलींतील भाषिक डेटा वापरून लेखकांना वैयक्तिकृत सूचना देणारी तंत्रे उपलब्ध झाल्यामुळे मराठी साहित्यिक आणि संशोधकांना सहज आणि अचूक लेखन करता येते.

सांस्कृतिक वारसा जतन करण्यासाठी आणि सार्वजनिक ज्ञानस्रोतांपर्यंत पोहोच सुलभ करण्यासाठी डिजिटायझेशन आणि AI-द्वारे विविध ग्रंथालये, संग्रहालये आणि सांस्कृतिक संस्था मराठी भाषेतील दस्तऐवज डिजिटल स्वरूपात जतन करत आहेत आणि ते ऑनलाइन उपलब्ध करून देत आहेत (Sundararajan, 2019). यामुळे मराठी भाषेचा सांस्कृतिक वारसा जागतिक स्तरावर पोहोचू शकतो. यातून नवीन संशोधन आणि जनजागृतीसाठीही मार्ग सुलभ झाला आहे. परंतु, AI-च्या प्रगत वापरामुळे काही सामाजिक आणि नैतिक प्रश्नही निर्माण झाले आहेत. भाषिक आणि सांस्कृतिक संवेदनशीलतेचा अभाव, डिजिटल समानतेतील अडथळे, तसेच शिक्षक किंवा अनुवादकांची जागा AI घेईल की काय, यांसारख्या चिंता उपस्थित झाल्या आहेत (Bamman et al., 2014). यासाठी सामाजिक, शैक्षणिक व तांत्रिक स्तरावर समन्वित आणि सूक्ष्म नियमन व धोरणे आवश्यक आहेत.

७. मराठी भाषेतील कृत्रिम बुद्धिमत्ता-आधारित तंत्रज्ञानातील मर्यादा आणि आव्हाने

कृत्रिम बुद्धिमत्ता (AI) आणि नैसर्गिक भाषा प्रक्रिया (NLP) यांच्या साहाय्याने मराठी भाषेसाठी भाषा अधिगम आणि अनुवाद प्रणाली विकसित करणे, हे एक महत्त्वपूर्ण आणि तांत्रिकदृष्ट्या आव्हानात्मक कार्य आहे. या संदर्भात येणाऱ्या मर्यादा आणि आव्हानांचा सखोल अभ्यास करणे आवश्यक आहे.

1) भाषिक संसाधनांचा अभाव

मराठीसाठी आवश्यक प्रमाणात भाषिक संसाधने, उदा. उच्च दर्जाचे डिजिटल कॉर्पस, संभाषण नमुने आणि विविध शैलींतील डेटासेट्स, अजूनही कमी प्रमाणात उपलब्ध आहेत (Patil & Kulkarni, 2021). NLP आणि MT प्रणालींसाठी मोठ्या प्रमाणावर डेटा आवश्यक असतो, ज्यामुळे मशीन लर्निंग मॉडेल्स प्रभावीपणे प्रशिक्षित करता येतात. इंग्रजी, हिंदीसारख्या प्रमुख भाषांसाठी प्रचंड डेटासेट्स उपलब्ध असले, तरी मराठीसाठी अशी संसाधने अजूनही विकासाच्या सुरुवातीच्या टप्प्यात आहेत (Sundararajan, 2019).

2) भाषेची व्याकरणिक आणि शैलीगत गुंतागुंत

मराठी भाषेची व्याकरणशास्त्रीय रचना, जसे की विभक्तीप्रत्यय, क्रियापदांची अनेक रूपे आणि शब्दरचनेची जटिलता, यांमुळे AI मॉडेलसाठी भाषेचा अचूक अर्थ समजणे कठीण होते (Mukherjee & Dasgupta, 2019). ही गुंतागुंत भाषेच्या शैलीतील सखोलता समजण्यात अडथळा आणते, ज्यामुळे अनुवाद आणि अधिगम प्रणालींच्या अचूकतेवर परिणाम होतो. उदाहरणार्थ, मराठीतील विभक्तीप्रत्ययांच्या चुकीच्या वापरामुळे अनुवादाचा अर्थ पूर्णपणे बदलू शकतो (Jiang et al., 2020).

3) डेटा संकलनातील आणि प्रयोगांतील अडचणी

AI-आधारित प्रणालींना प्रशिक्षित करण्यासाठी विविध वय, सामाजिक स्तर, बोलीभाषा, उच्चार आणि सांस्कृतिक संदर्भ लक्षात घेणाऱ्या बहुआयामी डेटाची आवश्यकता असते. तथापि, मराठीसाठीच्या संशोधन-गटांना आणि भाषा-प्रयोगांमध्ये असा डेटा मिळवणे, हे सध्या एक मोठे आव्हान आहे (Patil & Kulkarni, 2021). बोलीभाषांमधील विविधता आणि सांस्कृतिक फरकांमुळे पुरेसा वैविध्यपूर्ण डेटा तयार करण्यासाठी व्यापक संशोधन आणि संसाधने लागतात, जी मराठीसाठी मर्यादित आहेत.

4) नैतिकता, सामाजिक स्वीकार्यता आणि तंत्रज्ञानाचा स्वीकार

AI-आधारित भाषिक उपकरणांचा उपयोग समाजात आणि शैक्षणिक संस्थांमध्ये अद्याप पूर्णपणे स्वीकारला गेलेला नाही. काही भाषिक समुदायांमध्ये AI तंत्रज्ञानाच्या पारदर्शकतेबद्दल आणि त्यातील त्रुटींबद्दल भीती आहे, तसेच भाषिक संवेदनशीलतेचा मुद्दाही महत्त्वाचा आहे (Gokhale, 2015). भाषिक आणि सांस्कृतिक धोरणांच्या अभावामुळे AI-च्या विसंगत वापराला चालना मिळू शकते. डेटा गोपनीयता, मानवाधिकार आणि सामाजिक परिणाम यांसारख्या मुद्द्यांवर भारतीय संदर्भात सविस्तर चर्चा आणि नियंत्रणाची आवश्यकता आहे (Kukafka, 2020). AI प्रणालींच्या वापरामुळे भाषिक रोजगारावर होणारे परिणाम, स्थानिक संवादाच्या सांस्कृतिक पायावर होणारा प्रभाव आणि तंत्रज्ञानाच्या न्याय्य वापरावर सामाजिक-नैतिक जागरूकता वाढवणे गरजेचे आहे.

5) आर्थिक आणि संस्थात्मक अडचणी

AI संशोधनासाठी लागणारी गुंतवणूक, तांत्रिक क्षमता आणि व्यावसायिक धोरणांच्या अभावामुळे मराठीसारख्या प्रादेशिक भाषांमधील संशोधनाचे अनेक उपक्रम अर्धवट राहतात किंवा त्यांना विलंब होतो (Patil & Kulkarni, 2021). शैक्षणिक संस्था, उद्योग आणि शासन यांच्यातील समन्वय व आर्थिक पाठिंबा वाढवणे आवश्यक आहे, जेणेकरून प्रगत तंत्रज्ञान निरंतर विकसित होईल आणि सामान्य वापरकर्त्यांपर्यंत पोहोचेल.

८. भविष्यातील संधी आणि संशोधनासाठी दिशा

मराठी भाषेतील AI-आधारित भाषा अधिगम आणि अनुवाद प्रणालींसाठी अपार संधी आणि विकासाचे मार्ग उपलब्ध आहेत. यामध्ये संरचित आणि वैयक्तिकृत शिक्षणासाठी व्यापक तंत्रज्ञानाचा वापर होऊ शकतो, जो मराठीतील शिक्षण प्रक्रियेची गुणवत्ता लक्षणीयरीत्या वाढवेल (Chen et al., 2021).

1) स्थानिक भाषिक डेटाबेस आणि संसाधनांचा विकास

संशोधन संस्था, शैक्षणिक केंद्रे आणि स्थानिक भाषिक समुदाय यांच्या सहकार्याने मराठी भाषेचा विविध शैलींतील, बोलीभाषांतील आणि सामाजिक स्तरांवरील डेटा तयार करणे आवश्यक आहे. डिजिटल स्वरूपात संग्रहित अशा डेटाद्वारे दर्जेदार प्रशिक्षित AI मॉडेल्स विकसित करता येतील (Ramesh et al., 2022).

2) बहुभाषिक संवाद आणि भाषा अधिगमासाठी AI-सुसज्ज यंत्रणा

AI-आधारित व्हर्च्युअल ट्युटर्स आणि संवादात्मक प्लॅटफॉर्म तयार करून, मराठी भाषेतील संवाद बहुभाषिक संदर्भात अधिक सुलभ करता येईल. त्यासोबतच, उच्चार सुधारणा आणि बोलीभाषांमधील फरक लक्षात घेणाऱ्या AI मॉडेल्सवर संशोधन करण्याची गरज आहे, ज्यामुळे भाषिक समावेशन वाढेल (Patil & Kulkarni, 2021).

3) सांस्कृतिक वारशाचे डिजिटायझेशन आणि जागतिक स्तरावर प्रसार

AI-आधारित संगणकीय साहित्य विश्लेषणामुळे मराठी सांस्कृतिक आणि ऐतिहासिक वारशाचा जलद आणि अधिक परिष्कृत अभ्यास शक्य होईल. पारंपरिक आणि आधुनिक साहित्याचे प्रभावी डिजिटल संग्रहण झाल्यामुळे ते जागतिक स्तरावरील लोकांसाठी सहज उपलब्ध होईल (Jockers, 2013).

4) नैतिकता आणि सामाजिक परिणामांच्या विचारातून संशोधन

AI-आधारित भाषिक प्रणालींसाठी नैतिक आधार वाढवण्यासाठी शास्त्रीय चौकट आणि सामाजिक संवाद आवश्यक आहे. तंत्रज्ञानाचा सामाजिक परिणाम आणि वापरकर्त्यांची सुरक्षितता सुनिश्चित करण्यासाठी धोरणात्मक मार्गदर्शन आणि नियमन या क्षेत्रात संशोधन करणे आवश्यक ठरेल (Kukafka, 2020).

सरकार, शिक्षण संस्था, व्यावसायिक क्षेत्र आणि स्थानिक समाज यांच्यात समन्वय साधून मराठी भाषेतील AI संशोधन, विकास आणि प्रभावी प्रसार शक्य होईल. त्यामुळे मराठी भाषिक समुदायाचे शैक्षणिक आणि सांस्कृतिक सक्षमीकरण जागतिक स्तरावर वाढेल (Russell & Norvig, 2021).

९. निष्कर्ष

मराठी भाषेत कृत्रिम बुद्धिमत्तेच्या आधारे भाषा अधिगम आणि अनुवाद प्रणालींचा विकास हा शैक्षणिक, सामाजिक व सांस्कृतिक क्षेत्रासाठी एक अत्यंत महत्त्वपूर्ण टप्पा आहे. या संशोधनातून हे स्पष्ट होते की, AI तंत्रज्ञानाच्या मदतीने मराठी भाषेच्या शिक्षण आणि संवाद क्षमतेत लक्षणीय सुधारणा होऊ शकते. AI-आधारित संवादात्मक ट्युटर्स, स्वयंचलित उच्चार सुधारणा आणि न्यूरल मशीन ट्रान्सलेशन यांसारख्या प्रणाली मराठी भाषेला जागतिक स्तरावर अधिक सक्षम आणि संवादक्षम बनवण्यासाठी आधार देतील (Russell & Norvig, 2021). परंतु, या क्षेत्रात अनेक आव्हानेही आहेत. मराठी भाषेची विविधता, संसाधनांची कमतरता आणि सामाजिक-सांस्कृतिक स्तरावर तंत्रज्ञानाचा स्वीकार हे भविष्यातील प्रमुख अडथळे असतील. त्यामुळे स्थानिक डेटासेट्सची निर्मिती, तंत्रज्ञानातील सुधारणा आणि समाजात जागरूकता वाढवणे अत्यंत आवश्यक आहे (Patil & Kulkarni, 2021).

AI-आधारित मराठी भाषा अधिगम आणि अनुवाद प्रणालींच्या प्रगतीसाठी शिक्षण संस्था, संशोधन केंद्रे, तंत्रज्ञान कंपन्या आणि स्थानिक भाषिक समुदाय यांच्यात समन्वय व सहयोग महत्त्वाचा ठरेल. यामुळे भाषेचा प्रसार आणि सांस्कृतिक वारशाचे जतन जागतिक दर्जावर करणे शक्य होईल. AI तंत्रज्ञानातील सतत विकसित होणाऱ्या नवनवीन संधींचा फायदा घेऊन मराठी भाषेचा जागतिक स्तरावर सन्मान वाढवता येईल (Sundararajan, 2019). शिक्षण व संशोधनामध्ये AI-आधारित प्रणालींचे सामर्थ्य आणि मर्यादा समजून घेऊन त्यांचा विवेकपूर्ण वापर केल्यास, मराठी भाषा डिजिटल युगात अधिक समृद्ध, प्रभावी आणि सर्वसमावेशक म्हणून प्रगती करेल. यामुळे केवळ भाषाच नव्हे, तर तिचा सांस्कृतिक वारसाही सुरक्षित आणि सशक्त होईल.

10. संदर्भ (References)

- बम्मन, डी., ओ'कॉनर, बी., आणि स्मिथ, एन. ए. (२०१४). ए बायेसियन मिक्सड-इफेक्ट्स मॉडेल ऑफ लिटररी कॅरेक्टर. *ट्रान्झॅक्शन्स ऑफ द असोसिएशन फॉर कम्प्युटेशनल लिंग्विस्टिक्स*, २, ३२९-३४४.
- भरणे, व्ही., आणि देशपांडे, एस. (२०१८). मराठी लॅंग्वेज लर्निंग चॅलेंजेस अँड स्ट्रॅटेजीज: ऑन अॅनालिटिकल स्टडी. *जर्नल ऑफ लॅंग्वेज टेक्नॉलॉजी*, ५(२), १२३-१३७.
- ब्राऊन, टी., मॅन, बी., रायडर, एन., सुब्बिया, एम., एट अल. (२०२०). लॅंग्वेज मॉडेल्स आर फ्यू-शॉट लर्नर्स. *अॅडव्हान्सेस इन न्यूरोल इन्फॉर्मेशन प्रोसेसिंग सिस्टीम्स*, ३३, १८७७-१९०१.
- चेन, वाय., ली, एच., आणि झांग, एफ. (२०२१). एआय-पावर्ड लर्निंग: द फ्युचर ऑफ लॅंग्वेज एज्युकेशन. *इंटरनॅशनल जर्नल ऑफ एज्युकेशनल टेक्नॉलॉजी*, १८(१), १५-२९.
- डेवलिन, जे., चांग, एम-डब्ल्यू., ली, के., आणि तौटानोवा, के. (२०१९). बर्ट: प्री-ट्रेनिंग ऑफ डीप बायडायरेक्शनल ट्रान्सफॉर्मर्स फॉर लॅंग्वेज अंडरस्टैंडिंग. *प्रोसिडिंग्स ऑफ NAACL-HLT*.
- देशपांडे, ए. (२०१७). द मराठी लॅंग्वेज: ए हिस्टॉरिकल अँड लिंग्विस्टिक ओव्हरव्यू. *लॅंग्वेज रिव्यू जर्नल*, ११(३), ५५-७४.
- गोखले, ए. (२०१५). बॅरियर्स टू लर्निंग मराठी अमंग नॉन-नेटिव्ह स्पीकर्स. *इंडियन जर्नल ऑफ लॅंग्वेज स्टडीज*, १०(१), ३९-५२.
- गुडफेलो, आय., बेंजिओ, वाय., आणि कोरविल, ए. (२०१६). *डीप लर्निंग*. एमआयटी प्रेस.
- जिआंग, डी., लिऊ, बी., आणि झाओ, टी. (२०२०). अॅड्रेसिंग लिंग्विस्टिक चॅलेंजेस ऑफ इंडियन लॅंग्वेजेस इन न्यूरोल मशीन ट्रान्सलेशन. *IEEE ट्रान्झॅक्शन्स ऑन NLP*, ७(३), ४०१-४१३.
- कोहन, पी. (२०२०). *न्यूरोल मशीन ट्रान्सलेशन*. केंब्रिज युनिव्हर्सिटी प्रेस.
- कुकाफका, आर. (२०२०). एआय-ड्रिव्हन लॅंग्वेज लर्निंग: अॅप्लिकेशन्स अँड एथिकल कन्सिडरेशन्स. *कम्प्युटर्स अँड एज्युकेशन*, १४७, १०३७८५.
- ली, जे., सु, डी., आणि वांग, वाय. (२०१९). अॅडव्हान्सेस इन एआय फॉर लॅंग्वेज लर्निंग अॅप्लिकेशन्स. *जर्नल ऑफ अप्लाईड एआय एज्युकेशन*, ४(१), ३४-४८.
- मुखर्जी, एस., आणि दासगुप्ता, टी. (२०१९). केस मार्किंग अँड मॉर्फोलॉजिकल कॉम्प्लेक्सिटी इन मराठी अँड इट्स कम्प्युटेशनल इम्प्लिकेशन्स. *प्रोसिडिंग्स ऑफ CLT*, १(१), ७७-८५.
- पाटील, आर., आणि कुलकर्णी, एस. (२०२१). मराठी एनएलपी: चॅलेंजेस अँड अपॉर्च्युनिटीज. *इंडियन जर्नल ऑफ कम्प्युटेशनल लिंग्विस्टिक्स*, १५(२), ८९-१०२.
- रसेल, एस., आणि नॉर्विग, पी. (२०२१). *आर्टिफिशियल इंटेलिजन्स: ए मॉडर्न अप्रोच* (४थे संस्करण). पिअरसन.
- शेंडे, पी. (२०१९). लॅंग्वेज अॅक्विझिशन चॅलेंजेस फॉर रिजनल लॅंग्वेजेस: ए केस स्टडी ऑफ मराठी. *एशियन जर्नल ऑफ अप्लाईड लिंग्विस्टिक्स*, ३(२), १४५-१६०.
- सुंदरराजन, के. (२०१९). डिजिटल ट्रान्सफॉर्मेशन ऑफ इंडियन लॅंग्वेजेस थ्रू एआय. *जर्नल ऑफ लॅंग्वेज अँड टेक्नॉलॉजी*, ६(१), २००-२१५.
- थॉम्पसन, बी., आणि न्यूबिग, जी. (२०२०). इंटरॅक्टिव्ह लॅंग्वेज ट्युटर्स: अपॉर्च्युनिटीज अँड चॅलेंजेस. *ट्रान्झॅक्शन्स ऑफ द ACL*, ८, २२५-२३८.
- वासवानी, ए., शाहीर, एन., परमार, एन., एट अल. (२०१७). अटेंशन इज ऑल यू नीड. *अॅडव्हान्सेस इन न्यूरोल इन्फॉर्मेशन प्रोसेसिंग सिस्टीम्स*, ३०, ५९९८-६००८.