

## आत्मनिर्भर भारत की दिशा में भाषा प्रौद्योगिकी: एक समीक्षात्मक अध्ययन

**Dr. K. Neeraja**

SKR GDC (Women), Rajamahendravaram.

### प्रस्तावना

आत्मनिर्भर भारत अभियान केवल आर्थिक स्वावलंबन का नारा नहीं, बल्कि सांस्कृतिक, तकनीकी और भाषाई स्वायत्तता का समन्वित कार्यक्रम है, जो भारत की सभ्यतागत चेतना और समकालीन तकनीकी आवश्यकताओं के बीच सेतु का कार्य करता है। भारत में 22 अनुसूचित भाषाएँ और लगभग 19,500 से अधिक बोलियाँ विद्यमान हैं, जो इसे विश्व के सबसे समृद्ध भाषिक परिवेशों में एक विशिष्ट स्थान प्रदान करती हैं। इस बहुलता को डिजिटल युग में संरक्षित, सशक्त और उपयोगी बनाने के लिए उन्नत भाषा प्रौद्योगिकी का विकास अनिवार्य हो जाता है। वैश्विक स्तर पर अधिकांश डिजिटल प्रौद्योगिकियाँ अंग्रेजी-केंद्रित होने के कारण भारत की विशाल गैर-अंग्रेजीभाषी आबादी डिजिटल लाभों से वंचित रह जाती है। जनगणना 2011 के आँकड़ों के अनुसार मात्र लगभग 10.6% भारतीय अंग्रेजी में संवाद करने में सक्षम हैं, जबकि अधिकांश डिजिटल प्लेटफॉर्म अंग्रेजी-प्रमुख हैं, जिससे भाषा-आधारित डिजिटल विषमता और बढ़ जाती है। ऐसी स्थिति में स्वदेशी भाषा प्रौद्योगिकी का संगठित विकास न केवल तकनीकी आवश्यकता, बल्कि आत्मनिर्भरता की मूल पूर्वशर्त के रूप में सामने आता है।

### भाषा प्रौद्योगिकी: अवधारणा और परिभाषा

भाषा प्रौद्योगिकी (Language Technology) एक अंतर्विषयक क्षेत्र है, जो कम्प्यूटर विज्ञान, भाषाविज्ञान और कृत्रिम बुद्धिमत्ता के समन्वय से मानव भाषा के विश्लेषण, संसाधन और उत्पादन के लिए उपकरण एवं प्रणालियाँ विकसित करता है। इसके अंतर्गत प्रमुखतः निम्न क्षेत्र सम्मिलित हैं:

- प्राकृतिक भाषा संसाधन (NLP): मशीनों द्वारा मानव भाषा के वाक्य-विन्यास, अर्थ-विज्ञान और प्रयोग-विज्ञान को समझने, विश्लेषित करने और उत्पन्न करने की क्षमता।
- वाक् पहचान (Speech Recognition): बोली गई भाषा को स्वचालित रूप से डिजिटल पाठ में परिवर्तित करने की तकनीक।
- वाक् संश्लेषण (Text-to-Speech): लिखित पाठ को स्वाभाविक ध्वनि रूप में प्रस्तुत करने की प्रक्रिया।
- मशीन अनुवाद (Machine Translation): एक भाषा से दूसरी भाषा में स्वचालित अनुवाद करने वाली प्रणालियाँ।
- ऑप्टिकल कैरेक्टर रिकग्निशन (OCR): मुद्रित या हस्तलिखित पाठ की मशीन-पठनीय डिजिटल पहचान और रूपांतरण।
- संवाद प्रणालियाँ (Chatbots / Conversational Agents): मानव-मशीन संवाद को स्वाभाविक और संदर्भ-संवेदी बनाने वाली बुद्धिमान प्रणालियाँ।

इन सभी उपक्षेत्रों का साझा उद्देश्य यह है कि भाषा बाधा बने बिना नागरिक डिजिटल संसार से जुड़ सकें और मशीनें मानव भाषा को यथासंभव स्वाभाविक रूप में समझ-समझा सकें।

### भारत में भाषा प्रौद्योगिकी का ऐतिहासिक विकास

#### 1. प्रारंभिक चरण (1980–2000)

भारत में भाषा प्रौद्योगिकी विकास की संस्थागत यात्रा मुख्यतः Technology Development for Indian Languages (TDIL) कार्यक्रम से आरंभ मानी जाती है, जिसे इलेक्ट्रॉनिक्स और सूचना प्रौद्योगिकी मंत्रालय (MeitY) ने 1991 में प्रारंभ किया। इस चरण में प्राथमिक लक्ष्य भारतीय लिपियों के लिए फॉन्ट-विकास, कीबोर्ड लेआउट, एन्कोडिंग मानक और बुनियादी वर्ड प्रोसेसिंग सुविधाओं का निर्माण था। ISCI (Indian Script Code for Information Interchange) का विकास इस दौर की एक महत्वपूर्ण उपलब्धि रहा, जिसने विभिन्न भारतीय लिपियों के एकीकृत डिजिटल प्रतिनिधित्व की नींव रखी और आगे के मानकीकरण प्रयासों का आधार बना।

#### 2. विकास चरण (2000–2015)

दूसरे चरण में भाषा प्रौद्योगिकी शोध और अनुप्रयोग दोनों स्तरों पर विस्तार प्रारंभ हुआ। इस अवधि में मंत्र राजभाषा जैसे सरकारी दस्तावेजों के अंग्रेजी-हिंदी अनुवाद हेतु उपकरण, विभिन्न मशीन अनुवाद प्रणालियाँ, वर्तनी-जाँचक, भारतीय भाषाओं के लिए खोज इंजन और अन्य भाषा-

सहायक सॉफ्टवेयर विकसित किए गए। TDIL कार्यक्रम के अंतर्गत IITs, IIITs और केंद्रीय विश्वविद्यालयों में भाषा प्रौद्योगिकी से जुड़े विशेष अनुसंधान केंद्र स्थापित हुए, जिन्होंने कॉर्पस निर्माण, टूलकिट विकास और मानकीकरण की दिशा में महत्वपूर्ण योगदान दिया।

### 3. आधुनिक चरण (2015-वर्तमान)

कृत्रिम बुद्धिमत्ता, मशीन लर्निंग और विशेष रूप से गहन शिक्षण (Deep Learning) के आगमन ने भाषा प्रौद्योगिकी को गुणात्मक रूप से नई ऊँचाई पर पहुँचा दिया। इस चरण की प्रमुख पहलों में Digital India Bhashini Mission के अंतर्गत भाषिणी, IIT मद्रास की AI4Bharat पहल, वाणी जैसे वाक्-डेटासेट कार्यक्रम तथा IndicNLP Suite एवं अन्य इंडिक टूलकिट विशेष रूप से उल्लेखनीय हैं। भाषिणी को राष्ट्रीय भाषा प्रौद्योगिकी मिशन (National Language Technology Mission) के रूप में 2022 में प्रारंभ किया गया, जिसका उद्देश्य भारतीय भाषाओं के लिए भाषा प्रौद्योगिकी समाधान को डिजिटल सार्वजनिक वस्तु (Digital Public Goods) के रूप में उपलब्ध कराना है। AI4Bharat जैसे प्लेटफॉर्म ने भारतीय भाषाओं के लिए उन्नत मॉडल और ओपन-सोर्स उपकरणों की समृद्ध श्रृंखला विकसित की है।

#### वर्तमान प्रमुख पहलें और उपलब्धियाँ

##### भाषिणी (Bhashini) — राष्ट्रीय भाषा अनुवाद मिशन

भाषिणी भारत सरकार की महत्वाकांक्षी पहल है, जिसका उद्देश्य देश के प्रत्येक नागरिक को उसकी अपनी भाषा में डिजिटल सेवाओं तक सरल और सहज पहुँच प्रदान करना है। यह मिशन AI/ML और NLP पर आधारित ओपन-सोर्स मॉडल, टूल और समाधान विकसित कर उन्हें स्टार्टअप, उद्योग, शैक्षणिक संस्थानों और सरकारी एजेंसियों के लिए उपलब्ध कराता है। भाषिणी के माध्यम से बहुभाषी अनुवाद, वाक्-से-पाठ, पाठ-से-वाक् और वाक्-से-वाक् जैसे अनुप्रयोगों पर विशेष बल दिया जा रहा है, ताकि ई-गवर्नेंस, शिक्षा, स्वास्थ्य और वित्तीय सेवाओं का बहुभाषीकरण संभव हो सके।

##### AI4Bharat

AI4Bharat, IIT मद्रास के नेतृत्व में संचालित एक शोध-पहल है, जिसका लक्ष्य भारतीय भाषाओं के लिए उच्च गुणवत्ता वाले AI मॉडल, डेटासेट और टूलकिट विकसित करके उन्हें ओपन-सोर्स रूप में उपलब्ध कराना है। IndicBERT और अन्य बहुभाषी भाषा मॉडल, IndicTrans जैसी मशीन अनुवाद प्रणाली, Samanantar जैसे समानांतर कॉर्पस तथा विविध वाक् एवं पाठ डेटासेट इस पहल की प्रमुख उपलब्धियाँ हैं। इन संसाधनों ने शोधकर्ताओं, स्टार्टअप और डेवलपर्स के लिए भारतीय भाषा प्रौद्योगिकी में प्रयोग एवं नवोन्मेष की नई संभावनाएँ खोली हैं।

##### स्वदेशी बड़े भाषा मॉडल (LLMs)

हाल के वर्षों में भारतीय संस्थानों और उद्योगों ने स्वदेशी बड़े भाषा मॉडल विकसित करने की दिशा में गंभीर प्रयास किए हैं, जिनका फोकस भारतीय भाषाओं की बहुलता और सांस्कृतिक संदर्भ पर है। ऐसे मॉडल अंग्रेजी-आधारित वैश्विक मॉडलों पर निर्भरता को कम कर, स्थानीय जरूरतों के अनुरूप समाधान विकसित करने में सहायक हो सकते हैं। ये प्रयास दीर्घकाल में तकनीकी आत्मनिर्भरता और डेटा-संप्रभुता दोनों के लिए महत्वपूर्ण हैं।

#### आत्मनिर्भरता के संदर्भ में भाषा प्रौद्योगिकी का महत्व

##### 1. डिजिटल समावेशन

भारत की विशाल जनसंख्या का बड़ा हिस्सा अंग्रेजी में सहज नहीं है, जबकि रोजमर्रा की डिजिटल सेवाएँ प्रायः अंग्रेजी-प्रधान प्लेटफॉर्मों के माध्यम से उपलब्ध हैं। भाषा प्रौद्योगिकी नागरिकों को उनकी मातृभाषा में सरकारी योजनाओं, बैंकिंग, स्वास्थ्य, कृषि और अन्य सेवाओं तक पहुँच प्रदान कर डिजिटल अंतराल को कम करती है। भाषिणी जैसे मिशन का घोष लक्ष्य ही यह है कि भाषा किसी भी नागरिक के लिए डिजिटल विश्व तक पहुँच में बाधा न बने।

##### 2. शिक्षा में परिवर्तनकारी भूमिका

अनुसंधान से स्पष्ट है कि मातृभाषा में शिक्षण, संज्ञानात्मक विकास और अवधारणात्मक समझ दोनों स्तरों पर अधिक प्रभावी होता है। राष्ट्रीय शिक्षा नीति 2020 ने प्रारंभिक स्तर पर मातृभाषा/स्थानीय भाषा में शिक्षण पर विशेष बल दिया है, जिसके सफल क्रियान्वयन में भाषा प्रौद्योगिकी की केंद्रीय भूमिका है। भाषा प्रौद्योगिकी के माध्यम से क्षेत्रीय भाषाओं में उच्च गुणवत्ता वाली डिजिटल सामग्री, पाठ्यपुस्तकों का स्वचालित/अर्ध-स्वचालित अनुवाद, वॉइस-आधारित शिक्षण प्लेटफॉर्म और निरक्षर या अर्ध-साक्षर समूहों के लिए ऑडियो-विजुअल शिक्षण संसाधन विकसित किए जा सकते हैं।

##### 3. आर्थिक सशक्तिकरण

स्वदेशी भाषा प्रौद्योगिकी ग्रामीण और अर्धशहरी उद्यमियों, कारीगरों और लघु उद्यमों को ई-कॉमर्स, डिजिटल भुगतान और विपणन प्लेटफॉर्म तक उनकी भाषा में पहुँच प्रदान करती है। इससे स्थानीय उत्पादों को डिजिटल बाज़ार में स्थान मिलता है और नए रोजगार अवसर—जैसे अनुवादक, कंटेंट क्रिएटर, डेटा एनोटेटर, भाषा-विशेषज्ञ, आदि—उत्पन्न होते हैं।

#### 4. सुशासन और पारदर्शिता

ई-गवर्नेंस की सफलता नागरिकों की सक्रिय भागीदारी पर निर्भर करती है, जो तभी संभव है जब सरकारी संचार नागरिक की अपनी भाषा में उपलब्ध हो। भाषा प्रौद्योगिकी के माध्यम से सरकारी दस्तावेजों का स्वचालित अनुवाद, बहुभाषी पोर्टल, वॉइस-आधारित हेल्पलाइन, शिकायत निवारण प्रणालियाँ और न्यायिक प्रक्रियाओं में बहुभाषी सहायता विकसित की जा सकती है। इससे न केवल पारदर्शिता बढ़ती है, बल्कि शासन प्रणाली के प्रति नागरिकों का विश्वास भी सुदृढ़ होता है।

#### वैश्विक नेतृत्व की संभावना

भारत की बहुभाषी विशेषज्ञता और विशाल भाषिक संसाधन उसे वैश्विक भाषा प्रौद्योगिकी बाज़ार में एक विशिष्ट स्थान दिला सकते हैं। यदि भारत अपने लिए विकसित समाधान को मानकीकृत और स्केलेबल रूप में तैयार करे, तो उन्हें अफ्रीका, दक्षिण-पूर्व एशिया और अन्य बहुभाषी क्षेत्रों में निर्यात किया जा सकता है। इस प्रकार भाषा प्रौद्योगिकी न केवल आर्थिक अवसर, बल्कि तकनीकी कूटनीति और सॉफ्ट पॉवर का प्रभावी माध्यम भी बन सकती है।

#### निष्कर्ष

आत्मनिर्भर भारत की परिकल्पना तब तक पूर्ण नहीं मानी जा सकती, जब तक देश का प्रत्येक नागरिक अपनी मातृभाषा या पसंदीदा भाषा में डिजिटल विश्व से सार्थक रूप से जुड़ने में सक्षम न हो। भाषा प्रौद्योगिकी केवल तकनीकी नवाचार का क्षेत्र नहीं, बल्कि सामाजिक न्याय, समावेशी विकास और सांस्कृतिक संरक्षण का शक्तिशाली उपकरण है। पिछले दशक में भारत ने भाषिणी, TDIL और AI4Bharat जैसी पहलों के माध्यम से उल्लेखनीय प्रगति की है, जो स्वदेशी क्षमता-निर्माण की दिशा में मील के पत्थर हैं। तथापि, डेटा की कमी, अवसंरचना और निवेश की सीमाएँ, तथा मानकीकरण और मानव संसाधन विकास की चुनौतियाँ अभी भी निर्णायक समाधान की प्रतीक्षा कर रही हैं। इन चुनौतियों का सामना बहु-हितधारक, बहु-स्तरीय और दीर्घकालिक दृष्टिकोण से ही संभव है, जिसमें केंद्र और राज्य सरकारें, शैक्षणिक संस्थान, उद्योग जगत और नागरिक समाज सभी समान रूप से सहभागी हों। अंततः, भाषा प्रौद्योगिकी में आत्मनिर्भरता भारत को डिजिटल रूप से सशक्त बनाने के साथ-साथ उसकी सांस्कृतिक अस्मिता और भाषिक विरासत को सुरक्षित और समृद्ध बनाए रखने का मार्ग प्रशस्त करती है; यही वह समन्वित पथ है जो भारत को सच्चे अर्थों में आत्मनिर्भर और वैश्विक स्तर पर मार्गदर्शक बना सकता है।

#### संदर्भ

1. भारत सरकार, इलेक्ट्रॉनिक्स और सूचना प्रौद्योगिकी मंत्रालय — भाषिणी पोर्टल
2. AI4Bharat, IIT मद्रास — शोध प्रकाशन और ओपन-सोर्स उपकरण
3. TDIL कार्यक्रम — प्रौद्योगिकी विकास हेतु भारतीय भाषाएँ
4. राष्ट्रीय शिक्षा नीति 2020
5. भारतीय जनगणना 2011 — भाषा संबंधी आँकड़े