

# हिंदी स्क्रिप्ट में भाषिक विविधता का प्रतिनिधित्व: मानक हिंदी और बोलियों के संदर्भ में एक अध्यायन

**Ms. Sudha N<sup>1</sup>, Ms. Dhanyashree B<sup>2</sup>, Ms. Sejal Turlapati<sup>3</sup>,  
Ms. Nafiya Taskeen M<sup>4</sup>**

<sup>1</sup>Hindi Teacher[HOD], Hindi, KRM public school

<sup>2,3,4</sup>Student, Grade 10, KRM public school

## परिचय

भाषाई विविधता का मतलब है कि किसी समाज में अलग-अलग भाषाएँ बोलियाँ उच्चारण और संचार के तरीके मौजूद हों। भाषा संस्कृति और पहचान का एक अहम हिस्सा है। अलग-अलग समुदाय अपने विचारों और भावनाओं को दिखाने के लिए अपने खास बोलने के तरीके बनाते हैं। भारत एक ऐसा देश है जहाँ भाषाई विविधता बहुत अधिक है यहाँ के लोगों ने कई भाषाएँ और क्षेत्रीय भाषाई रूपों का इस्तेमाल किया है। हिंदी भारत में सबसे ज्यादा बोली जाने वाली भाषाओं में से एक है और यह कई अलग-अलग रूपों में बोली और लिखी जाती है। मानक हिंदी का प्रयोग आम तौर पर औपचारिक संवाद, शिक्षा, पुस्तकें, समाचार पत्र और आधिकारिक दस्तावेजों में किया जाता है। मानक हिंदी के साथ ही लोग रोज़मर्रा की बातचीत में बोलचाल की हिंदी का भी प्रयोग करते हैं। इसके अलावा हिंदी की कई क्षेत्रीय बोलियाँ और भाषाई रूप मौजूद हैं जिनकी अपनी शब्दावली, उच्चारण और अभिव्यक्तियाँ हैं।

बड़े भाषा मॉडल ऐसे सिस्टम हैं जो कृत्रिम बुद्धिमत्ता पर आधारित हैं और मानव भाषा को समझने व बनाने के लिए बनाए गए हैं। ये सिस्टम बड़ी मात्रा में उपलब्ध पाठ से भाषा के पैटर्न सीखते हैं और सवालों के जवाब देने, पाठ तैयार करने, जानकारी का सारांश बनाने और भाषाओं का अनुवाद करने जैसे काम कर सकते हैं। चूँकि भाषा ही मनुष्य और इन सिस्टमों के बीच संवाद का मुख्य माध्यम है इसलिए बड़े भाषा मॉडल में उपयोग होने वाले भाषा डेटा की गुणवत्ता और विविधता बहुत ज़रूरी है। भाषाई विविधता को समझने के लिए बड़े भाषा मॉडल के डेटा को विविध बनाना ज़रूरी है।

इंटरनेट, सोशल मीडिया, डिजिटल प्लेटफॉर्म और अन्य तकनीकों पर हिंदी के बढ़ते उपयोग के साथ हिंदी बड़े भाषा मॉडल के लिए भी अधिक महत्वपूर्ण हो गई है। हिंदी को सही ढंग से दिखाने के लिए सिर्फ औपचारिक या मानक हिंदी पर ध्यान देने से बढ़कर इसके सभी रूपों को ध्यान में रखना ज़रूरी है। इसलिए मानक हिंदी, बोलचाल की हिंदी और क्षेत्रीय हिंदी के बड़े भाषा मॉडल में प्रतिनिधित्व का अध्ययन भाषाई विविधता और नई भाषा प्रौद्योगिकी के बीच के रिश्ते को समझने में मदद कर सकता है।

## दुनिया में बड़े भाषा मॉडल क्या हैं?

बड़े भाषा मॉडल जिन्हें आमतौर पर स्क्रिप्ट कहा जाता है वे कृत्रिम बुद्धिमत्ता वाली एक प्रणाली होती है जो मानव भाषा को समझने, इसे देखने और इसे बनाने के लिए बनाई गई है। इन्हें कई बड़े अक्षरों के पाठ के माध्यम से

तैयार किया जाता है। ये पाठ किताबें लेख वेबसाइटों और अन्य डिजिटल सामग्री से एकत्र किए गए होते हैं। इस पाठ के अध्ययन के बाद ये मॉडल शब्दों वाक्यों व्याकरण और अर्थ के पैटर्न सीखते हैं। इससे वे ऐसा उत्तर बना सकते हैं जो मानव द्वारा लिखा गया लगता है।

सब्ज शब्दों और वाक्यों के बीच के संबंध और पैटर्न की पहचान करते हैं। जब कोई व्यक्ति कोई निर्देश देता है या कोई सवाल पूछता है तो मॉडल उस भाषा को समझता है और उसके आधार पर जवाब बनाता है। ये मॉडल मनुष्य की तरह भाषा को नहीं समझते लेकिन अपने प्रशिक्षण डेटा में मौजूद जानकारी के आधार पर एक अर्थपूर्ण और संबंधित उत्तर बना सकते हैं।

इन मॉडलों की शिक्षा के लिए डेटा की मात्रा और विविधता बहुत जरूरी होती है। जब कई भाषाएं विषय क्षेत्र लेखन शैलियां और संचार रूप शामिल किए जाते हैं तो मॉडल मानव भाषा का एक अधिक व्यापक प्रतिनिधित्व विकसित कर सकते हैं। हालांकि जहां डिजिटल सामग्री कम होती है वहां भाषा के प्रतिनिधित्व कम हो सकते हैं।

इन मॉडलों का उपयोग अक्सर सवालों के जवाब देने पाठ बनाने सारांश बनाने अनुवाद करने भाषा सीखने और जानकारी तैयार करने में होता है। ये हिंदी सहित कई भाषाओं में काम कर सकते हैं। इसलिए हिंदी के अलग-अलग रूपों को कैसे संसाधित किया जाता है इसका अध्ययन आधुनिक भाषा प्रौद्योगिकी में इनकी भूमिका को समझने में मदद करता है।

## हिंदी डिजिटल दुनिया में

हिंदी डिजिटल दुनिया में एक महत्वपूर्ण भाषा बन गई है। इंटरनेट स्मार्टफोन कंप्यूटर और अन्य डिजिटल तकनीकों के तेजी से विकास के साथ संचार और जानकारी प्राप्त करने के लिए हिंदी का उपयोग लगातार बढ़ रहा है। पहले इंटरनेट पर उपलब्ध अधिकांश सामग्री मुख्य रूप से अंग्रेजी और कुछ अन्य व्यापक रूप से प्रयोग की जाने वाली भाषाओं में होती थी। आज उपयोगकर्ता कई अलग-अलग डिजिटल प्लेटफॉर्म पर हिंदी में सामग्री प्राप्त कर सकते हैं। लोग जानकारी खोजने समाचार पढ़ने वीडियो देखने दूसरों के साथ संवाद करने नए विषय सीखने और अपने विचार व्यक्त करने के लिए हिंदी का उपयोग करते हैं। इंटरनेट पर हिंदी की बढ़ती उपस्थिति ने उन लोगों के लिए डिजिटल तकनीक को अधिक सुलभ बना दिया है जो हिंदी में संवाद करना पसंद करते हैं।

वेबसाइटों और ऑनलाइन प्लेटफॉर्मों पर हिंदी का विभिन्न उद्देश्यों के लिए व्यापक रूप से उपयोग किया जाता है। समाचार वेबसाइटें हिंदी में लेख प्रकाशित करती हैं शैक्षिक प्लेटफॉर्म अध्ययन सामग्री उपलब्ध कराते हैं और सरकारी वेबसाइटें हिंदी में महत्वपूर्ण जानकारी तथा सेवाएँ प्रदान कर सकती हैं। हिंदी का उपयोग ब्लॉग ऑनलाइन पत्रिकाओं डिजिटल विज्ञापनों उत्पाद विवरणों और मनोरंजन से जुड़ी वेबसाइटों में भी किया जाता है। सर्च इंजन उपयोगकर्ताओं को हिंदी में प्रश्न लिखकर हिंदी में उपलब्ध जानकारी खोजने की सुविधा देते हैं। इससे हिंदी केवल संचार की भाषा ही नहीं बल्कि इंटरनेट के माध्यम से ज्ञान और सेवाओं तक पहुँच प्राप्त करने की एक उपयोगी भाषा भी बन गई है।

सोशल मीडिया ने डिजिटल दुनिया में हिंदी के विकास में महत्वपूर्ण योगदान दिया है। लोग उन प्लेटफॉर्मों पर हिंदी का उपयोग करते हैं जहाँ वे पोस्ट तस्वीरें वीडियो विचार और अन्य प्रकार की सामग्री साझा करते हैं। कैप्शन टिप्पणियों संदेशों और चर्चाओं में हिंदी का सामान्य रूप से प्रयोग किया जाता है। वीडियो प्लेटफॉर्मों पर भी शिक्षा मनोरंजन समाचार खाना पकाने तकनीक यात्रा और दैनिक जीवन जैसे विषयों से संबंधित बड़ी मात्रा में हिंदी सामग्री उपलब्ध है। इससे लोगों को ऐसी भाषा में सामग्री बनाने और देखने की सुविधा मिलती

है जिससे वे परिचित हैं। सोशल मीडिया ने अलग-अलग क्षेत्रों के लोगों के लिए अपने अनुभव साझा करना और व्यापक दर्शकों के साथ संवाद करना भी आसान बना दिया है।

डिजिटल हिंदी की एक महत्वपूर्ण विशेषता यह है कि इसे हमेशा औपचारिक या मानक हिंदी की तरह नहीं लिखा जाता। ऑनलाइन बातचीत में लोग अक्सर अनौपचारिक शब्दों, छोटे रूपों और हिंदी तथा अंग्रेजी के मिश्रण का प्रयोग करते हैं। हिंदी को देवनागरी लिपि में या कुछ मामलों में रोमन वर्णमाला का उपयोग करके भी लिखा जा सकता है। उदाहरण के लिए लोग ऑनलाइन चैट या टिप्पणियों में अंग्रेजी अक्षरों का उपयोग करके हिंदी के शब्द लिख सकते हैं। इस प्रकार का संचार विशेष रूप से अनौपचारिक डिजिटल बातचीत में सामान्य है। इस तरह के बदलाव दिखाते हैं कि इंटरनेट पर उपयोग की जाने वाली हिंदी केवल पाठ्यपुस्तकों या आधिकारिक दस्तावेजों में प्रयोग होने वाली औपचारिक भाषा तक सीमित नहीं है बल्कि संचार की विभिन्न शैलियों को दर्शाती है।

हिंदी डिजिटल सामग्री के विकास का संबंध भारतीय भाषाओं में इंटरनेट का उपयोग करने वाले लोगों की बढ़ती संख्या से भी है। जैसे-जैसे अधिक लोग ऑनलाइन सक्रिय हो रहे हैं वैसे-वैसे वेबसाइटों, एप्लिकेशन और डिजिटल सेवाओं की ऐसी जानकारी उपलब्ध कराने की आवश्यकता बढ़ रही है जिसे उपयोगकर्ता आसानी से समझ सकें। इस प्रक्रिया में हिंदी महत्वपूर्ण भूमिका निभाती है क्योंकि भारत के विभिन्न क्षेत्रों में इसका व्यापक रूप से उपयोग किया जाता है। हिंदी में उपलब्ध डिजिटल सामग्री उपयोगकर्ताओं को अंग्रेजी पर पूरी तरह निर्भर हुए बिना शिक्षा, मनोरंजन, समाचार, सार्वजनिक जानकारी और अन्य संसाधनों तक पहुँचने में सहायता कर सकती है।

इंटरनेट पर हिंदी सामग्री की बढ़ती मात्रा भाषा प्रौद्योगिकी और बड़े भाषा मॉडल, स्क्रिप्ट के लिए भी महत्वपूर्ण है। स्क्रिप्ट बड़े संग्रहों में उपलब्ध पाठ और अन्य प्रकार के भाषाई डेटा से भाषा के पैटर्न सीखते हैं। जब हिंदी का प्रतिनिधित्व वेबसाइटों, लेखों, डिजिटल प्रकाशनों, सोशल मीडिया पोस्ट और अन्य ऑनलाइन सामग्री के माध्यम से होता है तो यह तकनीकी प्रणालियों के लिए उपलब्ध हिंदी भाषा संबंधी डेटा की मात्रा और विविधता में योगदान देता है। हिंदी सामग्री का एक विविध संग्रह अलग-अलग शब्दावली, वाक्य संरचनाओं, लेखन शैलियों और संचार के रूपों के उदाहरण प्रदान कर

## मानक हिंदी

मानक हिंदी वह रूप है जिसे लोग औपचारिक रूप से स्वीकार करते हैं। यह मुख्यतः खड़ी बोली पर आधारित है और देवनागरी लिखावट में लिखा जाता है। मानक हिंदी में व्याकरण, शब्द, वर्तनी और वाक्य रचना के सामान्य मानकों का पालन होता है। इसे उन सभी स्थानों में उपयोग किया जाता है जहाँ हिंदी बोली जाती है। क्योंकि लोग अलग-अलग क्षेत्रीय बोलियाँ बोलते हैं मानक हिंदी एक समान रूप देता है जिसे बहुत सारे हिंदी बोलने वाले समझ सकते हैं। मुझे लगता है कि मानक हिंदी सभी के लिए एक सामान्य आधार बनाती है।

मानक हिंदी का इस्तेमाल मुख्यतः औपचारिक स्थानों पर होता है। स्कूल, कॉलेज, पाठ्यपुस्तक, परीक्षा, सरकारी पत्र, आधिकारिक दस्तावेज, समाचार पत्र, पत्रिका और औपचारिक लेखन में इसे प्रयोग किया जाता है। टेलीविजन समाचार, शैक्षिक कार्यक्रम, सार्वजनिक घोषणाएँ और अन्य औपचारिक संवादों में भी मानक हिंदी का उपयोग व्यापक है। क्योंकि मानक हिंदी का व्याकरण और शब्द सरल और स्पष्ट हैं जब हमें बड़ी और विविध जनता तक जानकारी पहुँचानी होती है तो हम इसे चुनते हैं।

शिक्षा ने मानक हिंदी के विकास और फैलाव में अहम भूमिका निभाई है। हिंदी की पाठ्यपुस्तकें और अन्य शैक्षिक सामग्री सामान्यतः मानक व्याकरण और वर्तनी के नियमों का पालन करती हैं। विद्यार्थी इन नियमों के

अनुसार हिंदी पढ़ना और लिखना सीखते हैं जिससे मानक हिंदी का एक साफ़ लिखित रूप बनता है। समाचार पत्र और प्रकाशन भी शब्द व्याकरण और लेखन शैली में समानता रखकर मानक हिंदी के उपयोग को बढ़ावा देते हैं।

मानक हिंदी डिजिटल सामग्री में भी बहुत है। कई हिंदी समाचार वेबसाइट शैक्षिक प्लेटफॉर्म ऑनलाइन लेख सरकारी वेबसाइट और डिजिटल प्रकाशन मानक हिंदी का उपयोग करते हैं। सर्च इंजन और अन्य डिजिटल सेवाएँ भी इस रूप में लिखे बड़े पैमाने पर हिंदी पाठ को प्रोसेस करती हैं। इस कारण ऑनलाइन उपलब्ध लिखित भाषा डेटा में मानक हिंदी की बड़ी उपस्थिति है और यह भाषा प्रौद्योगिकियों में हिंदी के प्रतिनिधित्व के समय एक महत्वपूर्ण रूप बनता है।

मानक हिंदी की खासियतें इसे रोज़मर्रा की अनौपचारिक हिंदी से अलग करती हैं। औपचारिक मानक हिंदी स्थापित व्याकरण और उपयुक्त शब्दों का पालन करती है। परन्तु रोज़मर्रा की बातचीत में लोग वाक्य छोटा करते हैं अनौपचारिक शब्द प्रयोग करते हैं अंग्रेज़ी या अन्य भाषाओं के शब्द लेते हैं या क्षेत्रीय शब्दों का उपयोग करते हैं। इस कारण मानक हिंदी सिर्फ़ एक रूप है जो हिंदी बोलने और लिखने के कई तरीकों में से एक है।

मानक हिंदी बड़े भाषा मॉडल स्केल के लिए खास है क्योंकि पुस्तकों समाचार पत्र वेबसाइट शैक्षिक संसाधन और अन्य औपचारिक स्रोतों में उपलब्ध हिंदी का बड़ा हिस्सा मानक नियमों का पालन करता है। जब स्केल इस पाठ को प्रोसेस करते हैं तो उन्हें मानक हिंदी की शब्दावली व्याकरण वाक्य रचना और लेखन शैली के उदाहरण मिलते हैं। मानक हिंदी से स्केल को हिंदी पाठ उत्पन्न करने और संभालने में मदद मिलती है। परन्तु मानक हिंदी अकेले ही पूरी हिंदी भाषा को नहीं दर्शा सकता।

लोग अपने क्षेत्र उम्र समाज और बातचीत के उद्देश्य के अनुसार हिंदी का अलग-अलग ढंग से उपयोग करते हैं। उदाहरण के लिए कोई व्यक्ति शैक्षिक लेख लिखते समय मानक हिंदी का प्रयोग कर सकता है लेकिन दोस्तों के साथ बातें करते समय अनौपचारिक हिंदी इस्तेमाल कर सकता है। इसी तरह क्षेत्रीय लोग ऐसे शब्द और वाक्य प्रयोग करते हैं जो मानक हिंदी में नहीं मिलते। ये अंतर दर्शाते हैं कि स्केल में हिंदी का अध्ययन करते समय सिर्फ़ मानक हिंदी नहीं बल्कि सभी रूपों पर ध्यान देना ज़रूरी है।

इसी तरह स्केल में मानक हिंदी का प्रतिनिधित्व हिंदी और भाषा प्रौद्योगिकी को समझने का एक अहम हिस्सा है। इसकी व्यवस्थित व्याकरण और औपचारिक लिखित सामग्री में व्यापक उपस्थिति इसे हिंदी डेटा का एक बड़ा स्रोत बनाती है। साथ ही हिंदी की बड़ी भाषा विविधता को समझने के लिए मानक हिंदी के साथ बोलचाल की हिंदी और क्षेत्रीय रूपों का भी अध्ययन करना ज़रूरी है। इन सभी रूपों का मिलकर अध्ययन करने से हमें पता चलता है कि परम्परागत और डिजिटल दोनों माहौल में लोग हिंदी का कैसे उपयोग करते हैं

## बोलचाल की हिंदी

बोलचाल की हिंदी एक अनौपचारिक और सहज रूप है जिसे लोग रोज़मर्रा में बोलते हैं। यह वही भाषा है जो परिवार दोस्तों सहपाठियों पड़ोसियों और अन्य परिचितों के साथ बातचीत में प्रयोग होती है। मानक हिंदी जो व्याकरण और शब्दावली के नियमों पर आधारित होती है के विपरीत बोलचाल की हिंदी अधिक लचीली और बातचीत पर आधारित होती है। इसका प्रयोग किसी व्यक्ति की उम्र क्षेत्र सामाजिक समूह और जिस परिस्थिति में वह बात कर रहा है उसके अनुसार बदलता रहता है।

इस शैली में सरल शब्द अनौपचारिक अभिव्यक्तियाँ छोटे वाक्य और रोज़मर्रा के संवाद में सामान्य वाक्यांश शामिल होते हैं। लोग बोलते समय अंग्रेज़ी या अन्य भारतीय भाषाओं से लिए गए शब्दों का भी इस्तेमाल कर

सकते हैं। यह मिश्रण खासकर युवाओं और शहरी इलाकों में आम है। बोलचाल की हिंदी टेक्स्ट संदेशों, सोशल मीडिया पोस्ट, ऑनलाइन टिप्पणियों, वीडियो सामग्री और अनौपचारिक बातचीत में भी दिखाई देती है। इन संचार के रूपों में औपचारिक व्याकरण के नियमों से अधिक विचारों को स्वाभाविक रूप से व्यक्त करने पर जोर दिया जाता है।

मानक हिंदी और बोलचाल की हिंदी के बीच मुख्य अंतर उनके उद्देश्य और प्रयोग की शैली में दिखता है। मानक हिंदी का उपयोग आम तौर पर पाठ्यपुस्तकों, परीक्षाओं, आधिकारिक दस्तावेजों, समाचार पत्रों और अन्य औपचारिक संचार में होता है। दूसरी तरफ, बोलचाल की हिंदी का प्रयोग तब अधिक होता है जब लोग अनौपचारिक रूप से बात करते हैं या अपने परिचित लोगों के साथ संवाद करते हैं। कोई व्यक्ति शैक्षिक कार्य लिखते समय औपचारिक हिंदी का प्रयोग कर सकता है पर दोस्तों के साथ चैट करते समय स्वाभाविक रूप से अधिक अनौपचारिक हिंदी का उपयोग कर सकता है। इसलिए दोनों रूप हिंदी भाषा के महत्वपूर्ण हिस्से हैं भले ही उनका प्रयोग अलग-अलग परिस्थितियों में किया जाता हो।

बोलचाल की हिंदी डिजिटल दुनिया में खास तौर पर महत्वपूर्ण है क्योंकि ऑनलाइन संचार का एक बड़ा हिस्सा अनौपचारिक होता है। सोशल मीडिया पोस्ट, टिप्पणियाँ, संदेश, ऑनलाइन चर्चाएँ और वीडियो सामग्री में अक्सर सख्त मानक हिंदी के बजाय रोज़मर्रा में प्रयोग होने वाली हिंदी दिखाई देती है। ऑनलाइन संवाद करते समय लोग अलग-अलग वर्तनी, संक्षिप्त रूप, रोमन लिपि, अंग्रेज़ी शब्द या क्षेत्रीय अभिव्यक्तियों का इस्तेमाल कर सकते हैं। इससे हिंदी भाषा के डेटा की एक विस्तृत विविधता बनती है जो दिखाती है कि लोग अपने दैनिक जीवन में वास्तव में किस तरह संवाद करते हैं।

बोलचाल की हिंदी बड़े भाषा मॉडल, स्क्रिप्ट के लिए भी महत्वपूर्ण है। स्क्रिप्ट डिजिटल और लिखित डेटा के बड़े संग्रहों से भाषा के पैटर्न सीखते हैं जिनमें औपचारिक और अनौपचारिक दोनों प्रकार की हिंदी शामिल हो सकती है। बोलचाल की हिंदी दैनिक जीवन की शब्दावली, बातचीत में प्रयोग होने वाली अभिव्यक्तियों, वाक्य संरचनाओं और हिंदी-अंग्रेज़ी मिश्रित संचार के उदाहरण प्रदान करती है। हालांकि इसकी अनौपचारिक प्रकृति के कारण इसे एक समान रूप से प्रस्तुत करना अक्सर मुश्किल होता है क्योंकि एक ही शब्द या अभिव्यक्ति अलग-अलग तरीकों से लिखा जा सकता है। इसलिए बोलचाल की हिंदी का अध्ययन लोगों द्वारा उपयोग की जाने वाली हिंदी की व्यापक विविधता और स्क्रिप्ट में दैनिक भाषा के प्रतिनिधित्व से जुड़ी चुनौतियों को समझने में मदद करता है।

## क्षेत्रीय हिंदी बोलियाँ

भारत में अलग-अलग जगहों पर भाषा और बोलने का तरीका बहुत भिन्न होता है। हिंदी भारत की मुख्य भाषाओं में से एक है पर यह हर जगह एक जैसी नहीं बोली जाती। हर जगह पर हिंदी के अलग-अलग रूप और बोलियाँ चलती हैं। इन्हें क्षेत्रीय बोलियाँ कहते हैं। इन बोलियों में शब्द, उच्चारण, वाक्य बनाना और बोलने का तरीका अलग होता है। क्षेत्रीय बोलियाँ किसी क्षेत्र की संस्कृति, इतिहास और लोगों के जीवन से जुड़ी होती हैं। उदाहरण के लिए अवधी, भोजपुरी, ब्रज और हरियाणवी हिंदी से संबंधित प्रमुख क्षेत्रीय बोलियाँ हैं। इनके अलावा भी भारत के विभिन्न हिस्सों में हिंदी के कई स्थानीय रूप देखने को मिलते हैं।

## अवधी

अवधी मुख्य रूप से उत्तर प्रदेश के कुछ हिस्सों में बोली जाती है। इसका साहित्यिक और सांस्कृतिक महत्व बहुत अधिक है। प्रसिद्ध कवि गोस्वामी तुलसीदास की रचना रामचरितमानस अवधी में लिखी गई थी जिससे पता चलता है कि अवधी का भारतीय साहित्य और संस्कृति में बड़ा स्थान है।

अवधी के शब्द और उच्चारण मानक हिंदी से कुछ अलग हो सकते हैं। आज भी कई लोग अपने रोज़मर्रा के जीवन में अवधी का इस्तेमाल करते हैं। भोजपुरी भोजपुरी मुख्य रूप से पूर्वी उत्तर प्रदेश और बिहार के कुछ हिस्सों में बोली जाती है।

यह भारत के बाहर भी कुछ प्रवासी समुदायों द्वारा बोली जाती है। भोजपुरी में लोकगीतएँ कथाएँ और फिल्मों का समृद्ध संसार है। भोजपुरी बोलने वाले लोग रोज़मर्रा में ऐसे शब्द और वाक्य-रचना का उपयोग करते हैं जो मानक हिंदी से अलग होते हैं।

इसलिए यदि किसी भाषा मॉडल को केवल मानक हिंदी का प्रशिक्षण मिला होए तो उसे भोजपुरी के कुछ शब्दों या वाक्यों को समझने में कठिनाई हो सकती है।

### **भोजपुरी**

भोजपुरी मुख्य रूप से पूर्वी उत्तर प्रदेश और बिहार के कुछ हिस्सों में बोली जाती है। यह भारत के बाहर भी कुछ प्रवासी समुदायों द्वारा बोली जाती है। भोजपुरी में लोकगीतएँ कथाएँ और फिल्मों का समृद्ध संसार है।

भोजपुरी बोलने वाले लोग रोज़मर्रा में ऐसे शब्द और वाक्य-रचना का उपयोग करते हैं जो मानक हिंदी से अलग होते हैं।

इसलिए यदि किसी भाषा मॉडल को केवल मानक हिंदी का प्रशिक्षण मिला होए तो उसे भोजपुरी के कुछ शब्दों या वाक्यों को समझने में कठिनाई हो सकती है।

### **हरियाणवी**

हरियाणवी मुख्य रूप से हरियाणा और उसके आसपास के क्षेत्रों में बोली जाती है।

इसके उच्चारण और शब्दावली मानक हिंदी से अलग होती है। हरियाणवी में बातचीत का तरीका भी अपनी अलग पहचान रखता है।

आज सोशल मीडिया और इंटरनेट के कारण हरियाणवी गीतएँ वीडियो और अन्य सामग्री बड़े स्तर पर उपलब्ध हो रही है। इससे इस बोली की डिजिटल उपस्थिति भी बढ़ रही है।

### **स्ड के लिए क्षेत्रीय बोलियों का महत्व**

स्तहम स्दहनहम डवकमसे या स्ड को मानव भाषा समझने और उत्पन्न करने के लिए बहुत बड़ी मात्रा में भाषा-संबंधी डेटा की आवश्यकता होती है।

उदाहरण के लिए यदि कोई व्यक्ति किसी क्षेत्रीय बोली में प्रश्न पूछता है तो एक स्ड को उसका अर्थ समझकर उचित उत्तर देने में सक्षम होना चाहिए।

इसके लिए प्रशिक्षण डेटा में अलग-अलग बोलियों की पर्याप्त और विविध सामग्री होना ज़रूरी है। इस प्रकार क्षेत्रीय हिंदी बोलियाँ केवल भाषा के अलग-अलग रूप नहीं हैं। वे भारत की सांस्कृतिक और भाषाई विविधता का महत्वपूर्ण हिस्सा हैं।

### **हिंदी के विभिन्न रूपों का प्रतिनिधित्व**

हिंदी में कई तरह की बोलियाँ हैं। मानक हिंदीएँ रोज़मर्रा की बोलचाल की हिंदी और क्षेत्रीय बोलियाँ प्रमुख हैं। हर एक का अपना महत्व है और अलग-अलग परिस्थितियों में उपयोग होता है।

स्डए यानी बड़े भाषा मॉडलएँ को मानवीय भाषा समझने और जवाब देने के लिए अलग-अलग प्रकार के डेटा पर प्रशिक्षित किया जाता है। इसलिए प्रशिक्षण में हिंदी के सभी रूपों का सही प्रतिनिधित्व ज़रूरी है।

### मानक हिंदी का प्रतिनिधित्व

मानक हिंदी का इस्तेमाल स्कूलों में किताबों में समाचार-पत्रों में सरकारी कागजात और औपचारिक लिखावट में ज्यादा होता है। इसलिए इंटरनेट और डिजिटल स्रोतों पर मानक हिंदी की सामग्री अधिक मिलती है। इन सामग्री से स्लै को व्याकरण शब्दावली और औपचारिक शैली सीखने में मदद मिलती है।

### बोलचाल की हिंदी का प्रतिनिधित्व

लोग रोज़मर्रा में हमेशा औपचारिक हिंदी नहीं बोलते। बातचीत में छोटे वाक्यों अनौपचारिक शब्द स्थानीय शब्द और कभी-कभी अन्य भाषाओं के शब्द भी मिलते हैं। सोशल मीडिया चैट टिप्पणियाँ और ऑनलाइन बातों में ऐसी बोलचाल बहुत देखी जाती है। इसलिए स्लै को बोलचाल की हिंदी समझना भी ज़रूरी है।

### क्षेत्रीय हिंदी का प्रतिनिधित्व

क्षेत्रीय बोलियों का डिजिटल डेटा मानक हिंदी से कम हो सकता है। इससे मॉडल के लिए कुछ बोलियों को समझना मुश्किल हो जाता है। अगर किसी मॉडल के प्रशिक्षण में अवधी भोजपुरी ब्रज या हरियाणवी जैसी बोलियों का पर्याप्त डेटा हो तो वह इन रूपों को बेहतर समझ सकता है।

### डिजिटल सामग्री में असमानता

सभी भाषाई रूप इंटरनेट पर बराबर उपलब्ध नहीं हैं। कुछ में किताबें वेबसाइटें समाचार और सोशल मीडिया बहुत हैं जबकि कुछ में डिजिटल सामग्री कम हो सकती है। इसे डेटा असंतुलन कहा जा सकता है। अगर किसी बोली का डेटा कम हो तो स्लै उस बोली के शब्द और वाक्य सही ढंग से नहीं समझ पाएगा।

### भाषा की विविधता और स्लै

भारत में लोग अलग-अलग हिस्सों और समुदायों से आते हैं और भाषा का इस्तेमाल भी अलग-अलग तरीकों से होता है। इसलिए एक अच्छा स्लै सिर्फ एक ही प्रकार की हिंदी तक सीमित नहीं होना चाहिए। सभी रूपों का प्रतिनिधित्व मॉडल को यह सिखाता है कि एक ही विचार अलग-अलग तरीकों से कहा जा सकता है। उदाहरण के लिए कोई व्यक्ति मानक हिंदी में सवाल पूछ सकता है और दूसरा उसी सवाल को बोलचाल या क्षेत्रीय शैली में पूछ सकता है। स्लै को दोनों प्रश्नों का मतलब समझना चाहिए।

### उचित प्रतिनिधित्व क्यों आवश्यक है

हिंदी के विभिन्न रूपों का उचित प्रतिनिधित्व

- अलग-अलग क्षेत्रों के लोगों को तकनीक से जोड़ता है। स्लै की भाषा समझने की क्षमता बेहतर बनाता है।
- कम प्रतिनिधित्व वाली बोलियों को डिजिटल दुनिया में स्थान देता है। भारत की भाषाई विविधता को संरक्षित रखने में मदद करता है।
- तकनीक को अधिक समावेशी बनाता है।

इसलिए स्लै विकसित करते समय सिर्फ ज्यादा डेटा इकट्ठा करना पर्याप्त नहीं है। डेटा को विविध और संतुलित बनाना भी ज़रूरी है।

हिंदी के सभी रूपों का सही प्रतिनिधित्व यह सुनिश्चित करता है कि तकनीक सिर्फ एक विशेष प्रकार की भाषा बोलने वाले लोगों के लिए नहीं बल्कि अलग-अलग भाषाई पृष्ठभूमि वाले सभी लोगों के लिए उपयोगी हो।

### **चुनौतियां हिंदी को स्क्रिप्ट में शामिल करते समय**

हिंदी के अलग-अलग रूपों को स्क्रिप्ट में शामिल करना एक जरूरी काम है लेकिन इसमें कई मुश्किलें भी होती हैं। हिंदी की विविधता के कारण बस मानक हिंदी तक ही सीमित रहना ठीक नहीं है। मॉडल को बोलचाल की भाषाएँ क्षेत्रीय बोलियाँ विभिन्न तरह की वर्तनी और उच्चारण भी समझने होते हैं।

कम डिजिटल सामग्री कुछ क्षेत्रीय बोलियों के लिए इंटरनेट पर डेटा बहुत कम होता है। किताबों या मौखिक परंपराओं में उपलब्ध सामग्री हमेशा डिजिटल रूप में उपलब्ध नहीं होती। स्क्रिप्ट को प्रशिक्षित करने के लिए डिजिटल डेटा की आवश्यकता होती है। इसलिए जिन बोलियों के डेटा कम हैं उनका प्रतिनिधित्व भी कम हो सकता है।

अलग-अलग तरीके से लिखना हिंदी लिखते समय लोग अक्सर एक ही तरह की वर्तनी नहीं लिखते। खासकर सोशल मीडिया और ऑनलाइन बातचीत में लोग शब्दों को अलग-अलग तरीके से लिखते हैं। जैसे किसी शब्द को औपचारिक तरीके से नहीं बल्कि बोलचाल के अनुसार लिखा जा सकता है। इन अलग-अलग रूपों को एक ही अर्थ से जोड़ना स्क्रिप्ट के लिए मुश्किल हो सकता है।

बोली और लिखित भाषा में अंतर मनुष्य जिस तरह बोलते हैं वह जरूरी नहीं कि उसी तरह लिखते हों। कई क्षेत्रीय बोलियाँ मुख्य रूप से मौखिक रूप से प्रयोग की जाती हैं और उनका लिखित अंग बहुत कम होता है। इसलिए स्क्रिप्ट को केवल पुस्तकों और लिखित विषयों पर निर्भर नहीं रहना चाहिए। उसे बोली के रूप को भी समझना होगा।

अलग-अलग उच्चारण एक ही शब्द अलग-अलग क्षेत्रों में अलग-अलग तरह से बोला जा सकता है। अगर स्क्रिप्ट को स्पीच डेटा के साथ काम करना है तो उसे इन उच्चारणों को समझना आवश्यक है।

बोली में मिश्रित शब्दों का प्रयोग आज के डिजिटल दुनिया में लोग हिंदी के साथ अंग्रेजी या दूसरी भाषाओं के शब्दों का भी प्रयोग करते हैं। इसे कोडमिक्सिंग कहा जाता है। जैसे एक वाक्य में हिंदी और अंग्रेजी दोनों के शब्द हो सकते हैं। स्क्रिप्ट को इस तरह के मिश्रित वाक्यों को समझने की आवश्यकता होती है।

कम डेटा वाली बोलियाँ कुछ बोलियों में डेटा बहुत कम होता है। ऐसे में मॉडल उस बोली के शब्दों और वाक्य रचना के बारे में कम जानता है। इससे उस बोली बोलने वाले लोगों को कम सटीक जवाब मिल सकते हैं।

### **स्क्रिप्ट में भाषाई विविधता का महत्व**

भारत अपनी भाषाई विविधता के लिए मशहूर है। यहाँ पर अलग-अलग भाषाएँ बोलियाँ और भाषा के कई रूप मिलते हैं। हिंदी भी एक बहुवर्णी भाषा है जिसमें मानक हिंदी के अलावा कई क्षेत्रीय और बोलचाल के रूप हैं। बड़े भाषा मॉडल में इस भाषाई विविधता को शामिल करना जरूरी है क्योंकि तकनीक का इस्तेमाल अलग-अलग भाषाई पृष्ठभूमि वाले लोग करते हैं।

### **अलग-अलग क्षेत्रों के लोगों के लिए उपयोगिता**

यदि स्तंभ संदहनं हम डवकमस सिर्फ मानक हिंदी को ही अच्छी तरह समझे तो क्षेत्रीय बोलियों के लोग अपने सवालों को बताने में दिक्कत महसूस कर सकते हैं। ऐसे मॉडल जो कई बोलियों को समझता है अधिक लोगों के लिए उपयोगी हो जाता है। इससे लोग अपनी रोजमर्रा की भाषा में तकनीक से बात कर सकते हैं।

### **सांस्कृतिक विविधता को समझना**

भाषा और संस्कृति एक-दूसरे से गहराई से जुड़ी हैं। किसी क्षेत्र की बोली में उसकी परंपराएँ इतिहास और जीवनशैली की झलक मिलती है। जब स्तंभ स्तंभ-स्तंभ डबकमसे अलग अलग भाषाई रूपों से सीखते हैं तब वे भारत की सांस्कृतिक विविधता को और बेहतर समझ सकते हैं।

### कम प्रतिनिधित्व वाली बोलियों का संरक्षण

कम प्रतिनिधित्व वाली बोलियों का संरक्षण ज़रूरी है क्योंकि इन बोलियों का डिजिटल उपयोग कम है। यदि इन बोलियों की सामग्री को डिजिटल रूप में इकट्ठा करके सुरक्षित रखा जाए तो भविष्य में उनका अध्ययन और संरक्षण आसान हो सकता है। स्तंभ स्तंभ-स्तंभ डबकमसे इस काम में मददगार साबित हो सकते हैं क्योंकि वे बड़ी मात्रा में भाषा डेटा को संभाल सकते हैं।

### शिक्षा में लाभ

शिक्षा में भी भाषाई विविधता का फायदा है। यदि छात्र अपनी परिचित भाषा या बोली में सवाल पूछ सकें और उन्हें स्पष्ट उत्तर मिलें तो तकनीक का इस्तेमाल उनके लिए आसान हो जाएगा। यह खास तौर पर उन छात्रों के लिए लाभदायक है जिनके लिए औपचारिक भाषा हमेशा आसान नहीं होती।

### डिजिटल समावेशन

डिजिटल समावेशन में भाषाई विविधता का ध्यान रखने से और लोग डिजिटल तकनीक से जुड़ सकते हैं। जब तकनीक लोगों की भाषा समझती है तो वे उसे और आसानी से इस्तेमाल कर सकते हैं। इस तरह भाषाई विविधता सिर्फ भाषा का मामला नहीं बल्कि डिजिटल समावेशन का भी हिस्सा है।

### स्टे को अधिक प्रभावी बनाना

स्टे को अधिक प्रभावी बनाने के लिए इन्हें विभिन्न भाषाई रूपों के डेटा से प्रशिक्षित किया जाना चाहिए। एक ही विचार को लोग अलग-अलग शब्दों में कह सकते हैं। विविध डेटा मॉडल को इन अलग-अलग अभिव्यक्तियों के बीच का संबंध समझने में मदद करता है।

### भाषाई समानता

भाषाई समानता के लिए स्टे में विविधता को दिखाना एक कदम है जिससे केवल ज्यादा डिजिटल सामग्री वाली भाषाएँ ही महत्व पाएँ। कम प्रतिनिधित्व वाली भाषाएँ और बोलियाँ भी डिजिटल तकनीक में जगह पाना ज़रूरी है।

### निष्कर्षात्मक विचार

निष्कर्ष में भाषाई विविधता को स्टे में शामिल करना भारत जैसे बहुभाषी और बहुसांस्कृतिक देश में खासा महत्वपूर्ण है। मानक हिंदी बोलचाल की हिंदी और क्षेत्रीय हिंदी के बीच का अंतर समझना मॉडल को और व्यापक भाषा ज्ञान देता है। इसलिए स्टे का विकास सिर्फ तकनीकी प्रगति नहीं है यह भाषा संस्कृति और समाज से भी जुड़ा है।

### निष्कर्ष

इस अध्ययन से हमें पता चला कि हिंदी भाषा में कई अलग-अलग रूप हैं और स्टे में इनका कैसे प्रतिनिधित्व

होता है। सिर्फ एक ही प्रकार की हिंदी नहीं है। मानक हिंदीए बोलचाल की हिंदीए और अलग-अलग क्षेत्रों में बोली जाने वाली कई बोलियाँ मौजूद हैं।

हमने देखा कि अवधी, भोजपुरी, ब्रज और हरियाणवी जैसी क्षेत्रीय बोलियाँ हिंदी की भाषाई विविधता का महत्वपूर्ण हिस्सा हैं। इन बोलियों में शब्दावली, उच्चारण, वाक्य रचना और सांस्कृतिक विशेषताएँ पाई जाती हैं।

संस्कृत मानव भाषा को समझने और उत्पन्न करने के लिए बहुत बड़े मात्रा में भाषा डेटा से सीखते हैं। इसलिए प्रशिक्षण डेटा में हिंदी के अलग-अलग रूपों का अच्छी तरह प्रतिनिधित्व होना आवश्यक है। यदि डेटा में केवल मानक हिंदी अधिक हो और क्षेत्रीय या बोलचाल की हिंदी कम हो तो संस्कृत को कुछ रूपों को समझने में कठिनाई हो सकती है।

हमने यह भी जाना कि हिंदी को स्क्रिप्ट में प्रस्तुत करने में कई चुनौतियाँ हैं। इनमें सीमित डिजिटल सामग्री अलग-अलग वर्तनीएँ उच्चारण में अंतरए बोली और लिखित भाषा के बीच अंतरए बवकम-उपगपदह और कम प्रतिनिधित्व वाली बोलियाँ शामिल हैं।

इन चुनौतियों के बावजूद हिंदी की भाषाई विविधता को रूढ़ में शामिल करना ज़रूरी है। इससे अलग-अलग क्षेत्रों के लोग परिचित भाषा में तकनीक का उपयोग कर सकते हैं। इससे डिजिटल समावेशन बढ़ेगा और कम प्रतिनिधित्व वाली बोलियों को डिजिटल दुनिया में जगह मिलेगी।

भाषा सिर्फ संचार का माध्यम नहीं है। यह समाज की संस्कृति, इतिहास और पहचान से भी जुड़ी होती है। इसलिए किसी भाषा के विभिन्न रूपों का प्रतिनिधित्व करना उसकी विविधता को सम्मान देना है।

स्क्रिप्ट के विकास में भाषाई विविधता पर ध्यान देने से तकनीक अधिक लोगों के लिए उपयोगी हो सकती है। भविष्य में यदि इन मॉडलों में मानक बोलचाल और क्षेत्रीय हिंदी के अधिक विविध और गुणवत्तापूर्ण डेटा को शामिल किया जाए तो स्क्रिप्ट हिंदी के अलग-अलग रूपों को बेहतर ढंग से समझ सकते हैं।

अंततः हिंदी की विविधता को स्स्ट में दर्शाना केवल तकनीकी आवश्यकता नहीं है। यह भारत की समृद्ध भाषाई और सांस्कृतिक विविधता को डिजिटल दुनिया में बनाए रखने का एक महत्वपूर्ण कदम है।

इस प्रकार भाषाई विविधता और तकनीक का साथ मिलकर विकास यह सुनिश्चित कर सकता है कि भविष्य की तजपबिबंस प्दजमससपहमदबम तकनीक अधिक समावेशी उपयोगी और लोगों की वास्तविक भाषा के करीब हो।

## हिंदी भाषा का इस्तेमाल करने वाले स्सडरू

|पतंअंजं

ल्यमदभंजीप

॥ तअंउ उवकमसे

चत्वारम्वज पदकने

## संदर्भ

- 1<sup>१</sup> [॥जजचेरुध्दमदेनेपदकपंणवअण्पदद्धंदंकंधपदकमगण्वीचध्वंसवहध॥0191घनजउत्तरेवनतबमत्रबीजहचजणभवउ](#)
- 2<sup>२</sup> [॥जजचेरुध्दमदेनेपदकपंणवअण्पदद्धंदंकंधपदकमगण्वीचध्वंसवहध॥45240धक्वूदसवंकध॥8950धैयन्त्रज्जत्तत्त्रच्छ॥ वैष्णवचक्रदिनजउत्तरेवनतबम  
त्रबीजहचजणभवउ](#)
- 3<sup>३</sup> [॥जजचेरुध्दमचवेपजवतलण्पपजकण्मकनण्पदध्वाउसनपर्धिदकसमध॥123456789ध॥2098घनजउत्तरेवनतबमत्रबीजहचजणभवउ](#)
- 4<sup>४</sup> [॥जजचेरुध्दमदवकवण्वतहध्तमभवतकेध॥7312811घनजउत्तरेवनतबमत्रबीजहचजणभवउ](#)
- 5<sup>५</sup> [ब्जीमत अंतपवने मडेपजमे मजबय](#)