भारत सरकार
शिक्षा मंत्रालय
उच्चतर शिक्षा विभाग

**लोक सभा**
**अतारांकित प्रश्न संख्या-1604**
उत्तर देने की तारीख-09/02/2026

**एआई-आधारित असमिया भाषा और अन्य भाषाएं**

†1604. **श्री परिमल शुक्ला बैद्यः**

क्या **शिक्षा** मंत्री यह बताने की कृपा करेंगे कि:

(क) भाषिणी और भारतजेन जैसे आर्टिफ़िशियल इंटेलीजेन्स (एआई)-आधारित भाषा प्लेटफॉर्मों के माध्यम से सभी 22 अनुसूचित भारतीय भाषाओं को बढ़ावा देने के लिए सरकार द्वारा की गई पहलों का ब्यौरा क्या है और इस संबंध में अब तक क्या प्रगति हुई है;

(ख) असम में बोली जाने वाली भाषाओं, जिनमें असमिया, बंगाली, बिष्णुप्रिया मणिपुरी और बराक घाटी तथा कछार जिले की अन्य भाषाएं शामिल हैं, के विशेष संदर्भ में भाषाई डेटा के डिजिटलीकरण, बहुभाषी एआई उपकरणों के विकास और ओपन लैंग्वेज डेटासेट के निर्माण की सीमा क्या है;

(ग) भाषा प्रौद्योगिकियों के विकास और उपयोग में शैक्षणिक संस्थानों, स्टार्टअप्स और सार्वजनिक क्षेत्र की एजेंसियों, विशेषकर पूर्वोत्तर क्षेत्र की संस्थाओं द्वारा निभाई गई भूमिका क्या है; और

(घ) क्या सरकार का भाषाई रूप से विविधतापूर्ण जिलों, जैसे असम के कछार जिले में शिक्षा, शासन और सार्वजनिक सेवा प्रदायगी के लिए इन प्लेटफॉर्मों के उपयोग को बढ़ाने का विचार है और यदि हां, तो तत्संबंधी ब्यौरा क्या है?

**उत्तर**

**शिक्षा मंत्रालय में राज्य मंत्री**
**(डॉ. सुकान्त मजूमदार)**

(क) से (घ): राष्ट्रीय शिक्षा नीति (एनईपी) 2020 बहुभाषिकता के महत्व को दर्शाती है और सभी भारतीय भाषाओं को बढ़ावा देने पर अत्यधिक बल देती है। एनईपी 2020 के उद्देश्यों

के अनुरूप, भारत सरकार ने भारतीय भाषाओं पर शिक्षा, संरक्षण और अनुसंधान को बढ़ावा देने हेतु कई पहलें शुरू की हैं। इन प्रयासों को कृत्रिम मेधा और मशीन लर्निंग (एआई/एमएल) प्रौद्योगिकियों को अपनाने के माध्यम से और बढ़ाया गया है, जिसमें भारतीय भाषाओं में व्यापक भाषा मॉडल, अनुवाद उपकरण और भाषा-सक्षम डिजिटल सेवाओं का विकास शामिल है।

भारत सरकार ने 'भारत जेन' परियोजना शुरू की है, जो एक मल्टीमॉडल व्यापक भाषा मॉडल परियोजना है जो ऐसे कुशल और समावेशी एआई समाधान विकसित करने पर केंद्रित है, जो सभी 22 अनुसूचित भाषाओं को सपोर्ट करें और भारत के अनुपम सामाजिक-सांस्कृतिक संदर्भ और विविध क्षेत्रों के लिए एक मजबूत डिजिटल एआई अवसंरचना निर्माण को सक्षम बनाएँ। भारत जेन आईआईटी बॉम्बे में स्थापित है, जहां भाषा प्रौद्योगिकियों का मुख्य विकास किया जाता है। यह पहल आईआईटी कानपुर, आईआईटी मद्रास, आईआईटी हैदराबाद, आईआईआईटी हैदराबाद, आईआईएम इंदौर और आईआईटी मंडी सहित प्रमुख शैक्षणिक संस्थाओं के एक संघ के माध्यम से कार्यान्वित की गई है।

इसके अतिरिक्त, भारतीय भाषाओं में सामग्री के अनुवाद को सुगम बनाने के लिए, प्रौद्योगिकीय प्रगति में अखिल भारतीय तकनीकी शिक्षा परिषद (एआईसीटीई) द्वारा अनुवादिनी और इलेक्ट्रॉनिकी और सूचना प्रौद्योगिकी मंत्रालय द्वारा डिजिटल इंडिया कार्यक्रम के तहत एक पहल भाषिणी जैसे एआई-आधारित अनुवाद उपकरणों का विकास शामिल है।

भाषिणी ने 70 से अधिक अनुसंधान भागीदार संस्थाओं के सहयोग से भारतीय भाषाओं हेतु अत्याधुनिक एआई मॉडल विकसित किए हैं। भाषिणी प्लेटफॉर्म पर तीन सौ पचास से अधिक एआई-आधारित भाषा मॉडल का भंडार है और यह 22 से अधिक विशिष्ट भाषा सेवाएँ प्रदान करता है। इन सेवाओं में स्वचालित वाक् पहचान (एएसआर), मशीन अनुवाद (एमटी), टेक्स्ट-टू-स्पीच (टीटीएस), ऑप्टिकल कैरेक्टर रिकग्निशन (ओसीआर) और लिप्यंतरण शामिल हैं। डेटासेट कॉर्पस में 246 मिलियन समानांतर वाक्य पेयर और 3.7 मिलियन एकभाषी टेक्स्ट प्रविष्टियाँ शामिल हैं। सभी डेटासेट और मॉडल सार्वजनिक रूप से एआईकोश प्लेटफॉर्म पर भाषिणी प्लेटफॉर्म या डिजिटल इंडिया भाषिणी डिवीजन अकाउन्ट के माध्यम से उपलब्ध हैं। इसके अतिरिक्त, राष्ट्रीय भाषा अनुवाद मिशन भाषिणी मंच के माध्यम से सभी 22 अनुसूचित भाषाओं में बड़ी संख्या में पाठ और भाषण डेटा का डिजिटलीकरण कर रहा है।

भाषिणी मौलिक अनुसंधान, डेटासेट तैयार करने और आधारभूत एआई मॉडल के विकास हेतु पूर्वोत्तर क्षेत्र सहित शैक्षणिक और अनुसंधान संस्थाओं को शामिल करके बहु-हितधारक भागीदारी मॉडल का अनुसरण करता है। भाषिणी ने असम सरकार की सहकार्यता से

असमिया, बंगाली और अन्य भारतीय भाषाओं के लिए कदम उठाए हैं, ताकि इन्हें डिजिटल शासन प्लेटफार्मों और नागरिक सेवाओं में उपयोग किया जा सके, जिसमें कछार जैसे भाषायी रूप से विविध जिलों को भी शामिल किया गया है। इनमें अनुवाद प्लगइन्स का एकीकरण, एपीआई आदि पहुंच की सुविधा आदि शामिल हैं।

इसके अतिरिक्त, भारतीय भाषाओं के लिए भाषाई आंकड़ा संघ (एलडीसी-आईएल) योजना के अंतर्गत शिक्षा मंत्रालय के केन्द्रीय भारतीय भाषा संस्थान (सीआईआईएल) द्वारा अनुसूचित भारतीय भाषाओं के लिए व्यापक भाषाई संसाधन विकसित किए गए हैं। वर्ष 2019 से, एलडीसी-आईएल ने सरकारी एजेंसियों, सरकार द्वारा प्रोत्साहित पहलों, शोधकर्ताओं और भाषा प्रौद्योगिकियों के विकास में लगे वाणिज्यिक और औद्योगिक उपयोगकर्ताओं को संसाधन प्रदान किए हैं। असमिया भाषा के लिए जारी किए गए संसाधनों में एक रॉ टेक्स्ट कॉर्पस, रॉ स्पीच कॉर्पस, सेंटेंस-अलाइंड स्पीच कॉर्पस, टीटीएस कॉर्पस और मदर टंग पैरेलल टेक्स्ट कॉर्पस शामिल हैं। डेटासेट के अलावा, एलडीसी-आईएल ने भारतीय भाषाओं में अनुसंधान और विकास में सहायता देते हुए वेब-आधारित एप्लिकेशन और उपकरण भी विकसित किए हैं। इनमें से कई उपकरण, जो सीआईआईएल डेटा केंद्रों पर होस्ट किए गए हैं, मेधा भाषिका वेबसाइट https://medha.ciil.org के माध्यम से उपलब्ध हैं।

******