भारत सरकार
शिक्षा मंत्रालय
उच्चतर शिक्षा विभाग

राज्य सभा
अतारांकित प्रश्न संख्या-2769
उत्तर देने की तारीख-12/08/2026

जनजातीय एवं बहुभाषी क्षेत्रों के लिए मातृभाषा-आधारित एआई शिक्षण प्रणाली

2769 # **श्रीमती महुआ माजी:**

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

(क) क्या सरकार झारखंड जैसे बहुभाषी जनजातीय राज्यों के लिए संथाली, मुंडारी, हो, कुडुख तथा अन्य स्थानीय भाषाओं में कृत्रिम बुद्धिमत्ता (एआई) सक्षम अधिगम सामग्री विकसित करने का विचार रखती है;
(ख) क्या ऐसे क्षेत्रों के लिए मातृभाषा-आधारित डिजिटल अधिगम मॉडल हेतु कोई राष्ट्रीय प्रायोगिक परियोजना प्रस्तावित की गई है; और
(ग) यदि हां, तो क्या इस प्रायोगिक परियोजना हेतु झारखंड को प्राथमिकता दी जाएगी?

उत्तर

शिक्षा मंत्रालय में राज्य मंत्री
(डॉ. सुकान्त मजूमदार)

(क) से (ग): राष्ट्रीय शिक्षा नीति (एनईपी), 2020 बहुभाषावाद के महत्व को रेखांकित करती है और सभी भारतीय भाषाओं के संवर्धन पर बल देती है। एनईपी 2020 के उद्देश्यों के अनुरूप, भारत सरकार ने भारतीय भाषाओं पर शिक्षा, संरक्षण और अनुसंधान को बढ़ावा देने के लिए कई पहल 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 'भारतजेन' परियोजना शुरू की है, जो एक मल्टीमॉडल व्यापक भाषा मॉडल (एलएलएम) है जो कुशल और समावेशी एआई समाधान विकसित करने पर केंद्रित है जो सभी 22 अनुसूचित भाषाओं का समर्थन करता है और एक मजबूत डिजिटल एआई अवसंरचना के निर्माण को सक्षम बनाता है। भारतजेन आईआईटी बॉम्बे में चलाया जा रहा है, जहां भाषा प्रौद्योगिकियों का मुख्य विकास कार्य 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 अलावा, भारतीय भाषाओं में सामग्री के अनुवाद की सुविधा के लिए, तकनीकी प्रगति में अखिल भारतीय तकनीकी शिक्षा परिषद (एआईसीटीई) द्वारा 'अनुवादिनी' जैसे एआई-आधारित अनुवाद उपकरणों का विकास और डिजिटल इंडिया कार्यक्रम के तहत एक पहल 'भाषिणी' शामिल है।

भाषिणी ने 70 से अधिक अनुसंधान भागीदार संस्थाओं के सहयोग से भारतीय भाषाओं के लिए अत्याधुनिक एआई मॉडल विकसित किए हैं। भाषिणी मंच पर तीन सौ साठ से अधिक एआई-आधारित भाषा मॉडल का भंडार मौजूद है और 22 से अधिक विशिष्ट भाषा सेवाएं प्रदान करता है। इन सेवाओं में ऑटोमैटिक स्पीच रिकॉग्निशन (एएसआर), मशीन ट्रांसलेशन (एमटी), टेक्स्ट-टू-स्पीच (टीटीएस), ऑप्टिकल कैरेक्टर रिकॉग्निशन (ओसीआर) और लिप्यंतरण शामिल हैं। डेटासेट कॉर्पस में 246 मिलियन समानांतर वाक्य जोड़े और 3.7 मिलियन मोनोलिंगुअल टेक्स्ट प्रविष्टियां शामिल हैं। सभी डेटासेट और मॉडल भाषिनी मंच के माध्यम से या एआईकोश मंच पर डिजिटल इंडिया भाषिणी डिवीजन खाते के माध्यम से सार्वजनिक रूप से उपलब्ध हैं। इसके अलावा, भाषिणी मंच के माध्यम से राष्ट्रीय भाषा अनुवाद मिशन सभी 22 अनुसूचित भाषाओं में बड़ी मात्रा में पाठ और भाषण डेटा का डिजिटलीकरण कर रहा है। यह मंच भीली और संथाली जैसी आदिवासी भाषाओं में भी सहायता कर रहा है।

इसके अतिरिक्त, राष्ट्रीय शैक्षिक अनुसंधान एवं प्रशिक्षण परिषद के सहयोग से शिक्षा मंत्रालय के अधीनस्थ कार्यालय केन्द्रीय भारतीय भाषा संस्थान (सीआईआईएल) ने संथाली सहित विभिन्न जनजातीय भाषाओं में मूलभूत प्राइमर और भाषाई डेटासेट विकसित किए हैं। इसके अलावा, इसने मातृभाषा आधारित शिक्षा को बढ़ावा देने के लिए स्वयं मंच पर 12 सप्ताह का संथाली भाषा पाठ्यक्रम शुरू किया है।

भारतीय भाषाओं के लिए भाषाई डेटा संघ (एलडीसी-आईएल), जो सीआईआईएल की एक परियोजना है, के तहत, संथाली और मुंडारी सहित जनजातीय भाषाओं हेतु भाषाई संसाधन विकसित किए गए हैं। संथाली के लिए उपलब्ध संसाधनों में एक भाषण डेटासेट शामिल है जिसमें 60 वक्ताओं की 50 घंटे की ऑडियो रिकॉर्डिंग शामिल है। इसके अलावा, भाषाई विशेषताओं और संरचनाओं को शामिल करने वाला समानांतर पाठ कॉर्पोरा विकसित किया गया है, जिसमें संथाली और मुंडारी में प्रत्येक में 5,200 वाक्य शामिल हैं। एलडीसी-आईएल ने भारतीय भाषाओं में अनुसंधान और विकास की सहायता करने के लिए वेब-आधारित एप्लिकेशन और उपकरण भी विकसित किए हैं। इनमें से कई टूल, सीआईआईएल डेटा केंद्रों पर विकसित किए गए हैं, जिन्हें 'मेधा भाषिका' वेबसाइट https://medha.ciil.org पर देखा जा सकता है।

सीआईआईएल ने मुंडारी, खरिया, हो-हिंदी, कुरुख, कोरवा (झारखंड और छत्तीसगढ़), भूमिज, संथाली-हिंदी और माल्टो नामक कई भाषाओं में मूलभूत प्राइमर भी तैयार किए हैं। इसके अलावा, सीआईआईएल का भारतवाणी पोर्टल विविध संथाली संसाधनों जैसे कि भाषाकोश (भाषा सीखने का भंडार), ज्ञानकोश (विश्वकोश), पाठ्यपुस्तक कोश (पाठ्यपुस्तकें) और शब्दकोश (शब्दकोश) का परिचालन करता है।

*****