भारत सरकार

विज्ञान और प्रौद्योगिकी मंत्रालय

विज्ञान और प्रौद्योगिकी विभाग

लोक सभा

तारांकित प्रश्न संख्या *257

06 अगस्त, 2025 को उत्तर देने के लिए

भारतजेन एआई मॉडल

†*257. श्री जनार्दन सिंह सीग्रीवालः

श्री जनार्दन मिश्राः

क्या **विज्ञान और प्रौद्योगिकी** मंत्री यह बताने की कृपा करेंगे किः

(क) वर्तमान में भारतजेन के एआई मॉडल द्वारा समर्थित भारतीय भाषाओं की संख्या का ब्यौरा क्या है और भाषा संबंधी पूर्ण कवरेज प्राप्त करने के लिए क्या रूपरेखा है;

(ख) स्वास्थ्य देखभाल, कृषि, शिक्षा और शासन के क्षेत्रों में भारतजेन के विशिष्ट अनुप्रयोगों का मध्य प्रदेश के देवास-शाजापुर सहित राज्य-वार और जिला-वार ब्यौरा क्या है;

(ग) वर्तमान में भारतजेन नेटवर्क के अंतर्गत कार्यरत प्रौद्योगिकी नवोन्मेष केन्द्रों और ट्रांसलेशनल रिसर्च पार्कों की राज्य-वार संख्या कितनी है और उनमें किन-किन विषयों पर ध्यान केन्द्रित किया गया है;

(घ) भारतजेन परिसंघ में शामिल संस्थानों का ब्यौरा क्या है और इसके विकास और उपयोग में उनकी क्या भूमिका है;

(ङ) क्या सरकार का इस क्षेत्र में भारतजेन के अनुसंधान, भाषा-विकास अथवा प्रायोगिक अनुप्रयोगों को बढ़ावा देने के लिए दक्षिण कन्नड़ में शैक्षणिक अथवा अनुसंधान संस्थानों के साथ भागीदारी करने का विचार है और यदि हां, तो तत्संबंधी ब्यौरा क्या है;

(च) क्या सरकार ने मध्य प्रदेश के देवास-शाजापुर लोकसभा निर्वाचन क्षेत्र में सार्वजनिक और संस्थागत उपयोग के लिए भारतजेन एआई शुरू किया है या इसे शुरू करने की प्रक्रिया चल रही है; और

(छ) यदि हां, तो भारतजेन एआई के उद्देश्यों व विशेषताओं और अभीष्ट प्रयोक्ता समूहों तथा देवास-शाजापुर के ग्रामीण और अर्ध-शहरी क्षेत्रों को इससे होने वाले लाभों सहित तत्संबंधी ब्यौरा क्या है?

उत्तर

विज्ञान और प्रौद्योगिकी तथा पृथ्वी विज्ञान मंत्रालय के राज्य मंत्री (स्वतंत्र प्रभार)

(डॉ. जितेंद्र सिंह)

(क) से (छ): विवरण सदन के पटल पर रख दिया गया है।

"भारतजेन एआई मॉडल" के संबंध में दिनांक 06.08.2025 को लोक सभा में पूछे गए तारांकित प्रश्न संख्या 257 के भाग (क) से (छ) के उत्तर में उल्लिखित विवरण

(क) से (छ): भारतजेन भारतीय भाषाओं और सामाजिक संदर्भों के अनुरूप आदर्श आधारभूत एआई मॉडलों की एक श्रृंखला विकसित करने वाली प्रथम सरकारी सहायित राष्ट्रीय पहल है। इसमें पाठ (वृहत भाषा मॉडलों के माध्यम से), वाक् (पाठ-से-वाक् और स्वचालित उच्चारण अभिज्ञान), और विज़न-लैंग्वेज सिस्टम सहित कई प्रकार के मॉडल शामिल हैं।

वर्तमान में, भारतजेन मॉडल 9 भारतीय भाषाओं को कवर करता है जिनमें हिंदी, मराठी, तमिल, मलयालम, बंगाली, पंजाबी, गुजराती, तेलुगु और कन्नड़ शामिल हैं।

इन मॉडलों में भाषा कवरेज के रोडमैप में निम्नलिखित महत्वपूर्ण उपलब्धियां शामिल हैं:

- दिसंबर 2025 तक, कुल 15 भारतीय भाषाओं (असमिया, बंगाली, गुजराती, हिंदी, कन्नड़, मैथिली, मलयालम, मराठी, नेपाली, उड़िया, पंजाबी, संस्कृत, सिंधी, तमिल और तेलुगु सहित) को कवर किया जाएगा।
- जून 2026 तक सभी 22 अनुसूचित भारतीय भाषाओं को इसमें शामिल किया जाएगा।

भारतजेन ने कृषि, शासन और रक्षा क्षेत्रों में अनुप्रयोग विकसित किए हैं जिनका प्रायोगिक कार्य भी शुरू किया जा चुका है। पूर्णत: परिनियोजित होने के बाद, इन अनुप्रयोगों को सभी राज्यों और जिलों में उपलब्ध कराने की योजना है।

भारतजेन को विज्ञान और प्रौद्योगिकी विभाग (डीएसटी) के राष्ट्रीय अंतर्विषयक साइबर-भौतिक प्रणाली मिशन (एनएम-आईसीपीएस) के अंतर्गत एक परियोजना के रूप में क्रियान्वित किया गया है। भारतजेन नेटवर्क के भाग के रूप में, निम्नलिखित प्रौद्योगिकी नवोन्मेष केंद्र (टीआईएच) वर्तमान में सक्रिय हैं:

1. **टीआईएच आईओटी और आईओई फाउंडेशन**, **आईआईटी बॉम्बे (महाराष्ट्र)**
   यह टीआईएच भारतजेन परियोजना का आयोजन कर रहा है और केंद्रीय कार्यक्रम कार्यान्वयन एवं समन्वय केंद्र के रूप में कार्य करता है। यह संपूर्ण कार्यान्वयन के लिए उत्तरदायी है जिसमें राष्ट्रीय शैक्षणिक संघ का प्रबंधन, पाठ, वाक् और दृष्टि में आदर्श आधारभूत एआई मॉडलों के विकास का पर्यवेक्षण करना और कंप्यूट, डेटा और प्रतिभा के लिए पारिस्थितिकी तंत्र साझेदारी को आगे बढ़ाना शामिल है। टीआईएच भारतजेन परियोजना के लिए शासन और कार्यनीतिक योजना भी चलाता है जिससे सभी हितधारकों के बीच समेकित प्रगति सुनिश्चित होती है।

2. **आईआईटीएम प्रवर्तक टेक्नोलॉजीज फाउंडेशन**, **आईआईटी मद्रास (तमिलनाडु)**

   यह टीआईएच एक कार्यान्वयन भागीदार के रूप में कार्य करता है जो भारतजेन एआई प्रौद्योगिकियों के समाधान और वास्तविक विश्व में उनके परिनियोजन पर ध्यान देता है। ध्यान दिए जाने वाले विषयगत क्षेत्रों में शासन, सुरक्षा और मीडिया-संबंधी उपयोग के मामले शामिल हैं।

निम्नलिखित संस्थान भारतजेन संकाय (कंसोर्टियम) के भाग हैं:

संस्थान का नाम	**भारतजेन में भूमिका**
भारतीय प्रौद्योगिकी संस्थान, बॉम्बे	संस्थान का संचालन, अनुसंधान का मार्गदर्शन और संकाय भागीदारों के बीच एकीकरण
अंतर्राष्ट्रीय सूचना प्रौद्योगिकी संस्थान, हैदराबाद	दृष्टि-भाषा दस्तावेज़ मॉडलिंग
भारतीय प्रौद्योगिकी संस्थान, मद्रास	वाक् आधारित (फाउंडेशन) मॉडल विकास और मूल्यांकन
भारतीय प्रौद्योगिकी संस्थान, कानपुर	विधिक एआई अनुसंधान, क्षेत्र-विशिष्ट डेटासेट, और बहुभाषी मॉडलों हेतु टोकनीकरण कार्यनीतियों का विकास करना
भारतीय प्रौद्योगिकी संस्थान, हैदराबाद	वृहत बहुभाषी एलएलएम के लिए उन्नत टोकनीकरण और शब्दावली अनुकूलन
भारतीय प्रौद्योगिकी संस्थान, मंडी	एलएलएम के लिए कुशल प्रशिक्षण कार्यनीतियों पर समावेशी बहुभाषी मॉडल विकास और अनुसंधान
भारतीय प्रबंधन संस्थान, इंदौर	एलएलएम का भारत-केंद्रित मूल्यांकन और बेंचमार्किंग, बहुभाषी और बहुविध डेटा संग्रह

यद्धपि भारतजेन संकाय सदस्यों के साथ कार्यरत है तथापि यह कर्नाटक में अनुसंधान संस्थानों के साथ साझेदारी की संभावना तलाशता है।

चूँकि भारतजेन एआई अभी प्रायोगिक परिनियोजन चरण में है अत: इसे सार्वजनिक और संस्थागत उपयोग के लिए जारी नहीं किया गया है। पूर्णत: परिनियोजित होने के उपरांत, सभी राज्यों और ज़िलों में इसकी अनुप्रयोज्यता को विस्तार करने की योजना है।

*****