भारत सरकार
विज्ञान और प्रौद्योगिकी मंत्रालय
विज्ञान और प्रौद्योगिकी विभाग
लोक सभा
अतारांकित प्रश्न संख्या 3827
18 दिसंबर, 2024 को उत्तर देने के लिए

भारतजेन का शुभारंभ

†3827. श्रीमती स्मिता उदय वाघ:

क्या **विज्ञान और प्रौद्योगिकी मंत्री** यह बताने की कृपा करेंगे कि:

(क) भारतजेन किस प्रकार यह सुनिश्चित करता है कि उसके जेनरेटिव एआई मॉडल देश के विविध भाषाई परिदृश्य का पर्याप्त रूप से प्रतिनिधित्व करे और कम प्रतिनिधित्व वाली भाषाओं को शामिल करने के लिए क्या विशिष्ट उपाय किए जा रहे हैं;

(ख) भारतजेन अपने जेनरेटिव एआई मॉडलों को विशेषकर देश में हाशिए पर पड़े और कम प्रतिनिधित्व वाले समुदायों के लिए सार्वजनिक वस्तु के रूप में सुलभ बनाने को प्राथमिकता देने के लिए क्या कदम उठाएगा;

(ग) सांस्कृतिक पहचान और क्षेत्रीय विकास को बढ़ावा देने के लिए स्थानीय भाषाओं और बोलियों का प्रभावी उपयोग सुनिश्चित करते हुए भारतजेन क्षेत्र-विशिष्ट विषयवस्तु के विकास में किस प्रकार योगदान देता है; और

(घ) भारतजेन के मॉडलों की समावेशिता को बढ़ाने और देश के विभिन्न सामाजिक-आर्थिक समूहों में समान प्रौद्योगिकीय पहुंच सुनिश्चित करने के लिए किन साझेदारियों अथवा सहयोगों पर विचार किया जा रहा है?

उत्तर
विज्ञान और प्रौद्योगिकी तथा पृथ्वी विज्ञान मंत्रालय के राज्य मंत्री (स्वतंत्र प्रभार)
(डॉ. जितेंद्र सिंह)

(क) भारतजेन एक बहुविधीय बहुभाषी वृहद भाषा मॉडल पहल है, जो भारत की भाषाई, सांस्कृतिक और सामाजिक-आर्थिक विविधता के अनुरूप उन्नत जनरेटिव एआई मॉडल विकसित कर रही है। यह सुनिश्चित करने के लिए कि जनरेटिव एआई मॉडल भारत के विविध भाषाई परिदृश्य का पर्याप्त रूप से प्रतिनिधित्व करते हैं, भारतजेन ने प्राथमिक डेटा संग्रह पर ध्यान केंद्रित करते हुए "भारत डेटा सागर" नामक पहल शुरू की है। यह डेटा संग्रह इस अपेक्षा को पूरा करने का प्रयास करता है कि उन भारतीय भाषाओं के लिए प्रशिक्षण डेटा उपलब्ध हो जिनका डेटा कॉर्पोरा में कम प्रतिनिधित्व है।

(ख) भारतजेन देश भर के अनुसंधान समूहों के साथ साझेदारी कर रहा है, ताकि यह सुनिश्चित किया जा सके कि विकसित किए जा रहे जनरेटिव एआई मॉडल को भागीदारों द्वारा विस्तारित किया जा सके और आगे के विकास और उपयोग के लिए बड़े अनुसंधान और गैर-शैक्षणिक समुदाय को उपलब्ध कराया जा सके। भारतजेन कुशल प्रशासन और आमजन के लिए अनुकूलित अनुप्रयोगों हेतु सरकार उद्योग और स्टार्ट-अप के साथ साझेदारी भी कर रहा है।

(ग) सांस्कृतिक पहचान और क्षेत्रीय विकास को बढ़ावा देने के लिए, भारतजेन उन प्रौद्योगिकियों और उपकरणों को उपलब्ध कराएगा जो स्थानीय भाषाओं और बोलियों में निर्बाध रूप से अनुवाद करके क्षेत्र-विशिष्ट सामग्री के विकास को सहायित करेंगे।

(घ) भारतजेन में भारत के प्रमुख शैक्षणिक संस्थानों के शीर्ष एआई शोधकर्ताओं का समूह शामिल है। इसमें आईआईटी बॉम्बे, आईआईआईटी हैदराबाद, आईआईटी मंडी, आईआईटी कानपुर, आईआईटी हैदराबाद, आईआईएम इंदौर और आईआईटी मद्रास भी शामिल हैं। ये अनुसंधान समूह भारत की भाषिक और सांस्कृतिक विविधता और नागरिकों के लिए समावेशिता को ध्यान में रखते हुए मॉडल विकसित करने और देश के विभिन्न सामाजिक-आर्थिक समूहों में औचित्यपूर्ण तकनीकी पहुंच सुनिश्चित करने के लिए सरकार, उद्योग और स्टार्टअप के साथ साझेदारी कर रहे हैं।

*****