भारत सरकार

इलेक्ट्रॉनिकी और सूचना प्रौद्योगिकी मंत्रालय

लोक सभा

अतारांकित प्रश्न संख्या 1800

जिसका उत्तर 30 जुलाई, 2025 को दिया जाना है ।

8 श्रावण, 1947 (शक)

हाइ-इंपेक्ट एआई मॉडल का मूल्यांकन

1800. श्री यदुवीर वाडियारः

क्या इलेक्ट्रॉनिकी और सूचना प्रौद्योगिकी मंत्री यह बताने की कृपा करेंगे कि:

(क) क्या सरकार कृत्रिम बुद्धिमत्ता (एआई) सुरक्षा, विश्वसनीयता और जिम्मेदारीपूर्ण उपयोग के लिए एक ठोस राष्ट्रीय ढांचा विकसित करने की आवश्यकता को स्वीकार करती है;

(ख) क्या एआई निगरानी में वैश्विक सर्वोत्तम पद्धतियों का अध्ययन करने और भारत-विशिष्ट दिशानिर्देशों की सिफारिश करने के लिए कोई विशेषज्ञ या अंतर-मंत्रालय समूह गठित किया गया है और यदि हाँ, तो तत्संबंधी ब्यौरा क्या है;

(ग) क्या सरकार का सार्वजनिक क्षेत्रों में उपयोग किए जाने वाले हाइ-इंपेक्ट एआई मॉडलों के मूल्यांकन के लिए तकनीकी मानकों और संस्थागत तंत्रों को शुरू करने का प्रस्ताव है; और

(घ) क्या सरकार की व्याख्यात्मकता मॉडल मूल्यांकन और नैतिक आश्वासन के आसपास प्रोटोकॉल को सह-विकसित करने के लिए शैक्षणिक, उद्योग और सभ्य समाज के हितधारकों के साथ जुड़ने की संभावना है?

उत्तर

इलेक्ट्रॉनिकी और सूचना प्रौद्योगिकी राज्य मंत्री (श्री जितिन प्रसाद)

(क) से **(घ):** सरकार ने आर्टिफिशियल इंटेलिजेंस (एआई) शासन के लिए एक समावेशी और नवाचार-अनुकूल दृष्टिकोण अपनाया है। यह नागरिकों के अधिकारों की रक्षा, डेटा गोपनीयता सुनिश्चित करने और नवाचार को बढ़ावा देने के लिए एक संस्थागत व्यवस्था द्वारा समर्थित है।

<u>एआई पर सलाहकार समूह</u>

- सरकार ने प्रधानमंत्री के प्रधान वैज्ञानिक सलाहकार की अध्यक्षता में एआई पर एक सलाहकार समूह का गठन किया है ताकि उत्तरदायी एआई के लिए भारत-विशिष्ट नियामक ढाँचा विकसित किया जा सके। शिक्षा जगत, उद्योग और सरकार के हितधारकों से युक्त इस समूह का उद्देश्य सुरक्षित और विश्वसनीय एआई विकास सुनिश्चित करना है।
- एआई गवर्नेंस दिशानिर्देशों पर रिपोर्ट, भारत के एआई परिदृश्य के निरंतर विकास के साथ प्रभावी अनुपालन और प्रवर्तन सुनिश्चित करने के लिए एक समन्वित, समग्र सरकारी दृष्टिकोण की आवश्यकता पर बल देती है। एआई प्रौद्योगिकियों की विकासशील प्रकृति को देखते हुए, रिपोर्ट एआई विनियमन के लिए एक तकनीकी-कानूनी दृष्टिकोण अपनाने की सिफारिश करती है।
- एआई गवर्नेंस दिशानिर्देश विकास पर रिपोर्ट पर सार्वजनिक परामर्श पूरा हो चुका है और 100 से अधिक सुझाव प्राप्त हुए हैं।

<u>विशेषज्ञ समूह की रिपोर्ट</u>

- भारत के प्रत्येक एआई स्तंभ के लिए दृष्टिकोण, उद्देश्यों, परिणामों और डिजाइन पर सहयोगात्मक विचार-मंथन के लिए सात विशेषज्ञ समूहों का गठन किया गया।

- इंडियाएआई विशेषज्ञ समूह की रिपोर्ट के पहले संस्करण में इंडियाएआई के प्रमुख स्तंभों के उद्देश्यों का विस्तृत विवरण दिया गया है और सामाजिक विकास हेतु एआई की क्षमता का दोहन करने तथा 'सभी के लिए एआई' के लक्ष्य को प्राप्त करने हेतु आगामी कार्ययोजनाओं की अनुशंसा की गई है। रिपोर्ट की एक प्रति https://indiaai.gov.in/news/indiaai-2023-expert-group-report-first-edition पर उपलब्ध है।

<u>सुरक्षित और विश्वसनीय एआई</u>

- एआई सुरक्षा, संरक्षा और विविध हितधारकों के साथ विश्वास विकसित करने के लिए इंडियाएआई सेफ्टी इंस्टीट्यूट
- शिक्षा जगत, स्टार्टअप, उद्योग और सरकारी निकायों सहित विविध हितधारक हब-एंड-स्पोक मॉडल के माध्यम से जुड़ते हैं
- उत्तरदायी एआई के इर्द-गिर्द उपकरण और ढांचे बनाए जा रहे हैं, जिनका ध्यान 10 विषयों पर केंद्रित है
- मशीन अनलर्निंग, एआई पूर्वाग्रह शमन, गोपनीयता बढ़ाने वाले उपकरण, एआई गवर्नेंस परीक्षण आदि को कवर करने वाली 8 जिम्मेदार एआई परियोजनाएं चल रही हैं।
- वॉटरमार्किंग, नैतिक एआई, जोखिम मूल्यांकन, तनाव परीक्षण और डीपफेक डिटेक्शन टूल जैसे प्रमुख विषयों पर अधिक परियोजनाओं का भी मूल्यांकन किया जा रहा है

भारत फरवरी 2026 में एआई इम्पैक्ट शिखर सम्मेलन की मेजबानी भी कर रहा है, जो एआई एक्शन शिखर सम्मेलन के सह-अध्यक्ष के रूप में अपनी भूमिका को आगे बढ़ाएगा और वैश्विक एआई चर्चाओं को आकार देने में अपना नेतृत्व जारी रखेगा।

<u>एआई को विनियमित करने के लिए तकनीकी-कानूनी दृष्टिकोण</u>

भारत की एआई रणनीति का एक प्रमुख स्तंभ विनियमन के प्रति इसका संतुलित और व्यावहारिक तकनीकी-कानूनी दृष्टिकोण है:

- भारत कानूनी सुरक्षा उपायों को तकनीकी समाधानों के साथ जोड़ रहा है।
- सरकार डीपफेक का पता लगाने, गोपनीयता बढ़ाने और साइबर सुरक्षा के लिए एआई उपकरण बनाने हेतु आईआईटी जैसे प्रमुख संस्थानों में अनुसंधान एवं विकास परियोजनाओं को वित्तपोषित कर रही है।
- यह भारत के इस विश्वास को दर्शाता है कि प्रभावी एआई शासन को केवल कानून बनाने से आगे जाना चाहिए तथा नवीन तकनीकी हस्तक्षेपों द्वारा समर्थित होना चाहिए।
- यह विशिष्ट मॉडल सुनिश्चित करता है कि नवाचार में कोई बाधा न आए

<u>इंडियाएआई इम्पैक्ट समिट 2026</u>

- इंडियाएआई इम्पैक्ट समिट 2026 की तैयारी के एक हिस्से के रूप में , सरकार ने शिक्षा जगत, अंतर्राष्ट्रीय संगठनों , उद्योग, स्टार्टअप्स, थिंक-टैंक और नागरिक समाज के हितधारकों से इनपुट एकत्र करने के लिए सार्वजनिक परामर्श आयोजित किया है। इस अंतर्दृष्टि का उपयोग शिखर सम्मेलन के एजेंडे को आकार देने और समावेशी प्रतिनिधित्व सुनिश्चित करने के लिए किया जाता है।

<u>नैतिक एआई मानकों के संबंध में पहल</u>

- 'कृत्रिम बुद्धिमत्ता प्रणालियों के निष्पक्षता मूल्यांकन और रेटिंग के लिए मानक ' विकसित किया गया (**टीईसी** मानक संख्या: **टीईसी** 57050:2023)। यह मानक, विशेष रूप से दूरसंचार और संबंधित आईसीटी अनुप्रयोगों के संदर्भ में, कृत्रिम बुद्धिमत्ता प्रणालियों की निष्पक्षता के मूल्यांकन हेतु एक ढाँचा और प्रक्रियाएँ प्रदान करता है।
- भारतीय मानक IS/ISO/IEC/TR 24368:2022 प्रकाशित किया गया। यह एआई से जुड़ी नैतिक और सामाजिक चिंताओं का एक उच्च-स्तरीय अवलोकन प्रदान करता है। इसमें एआई से जुड़ी नैतिक और सामाजिक चिंताओं से उत्पन्न होने वाले मुद्दों को संबोधित करने वाले अंतर्राष्ट्रीय मानकों का अवलोकन भी शामिल है।

****

अनुबंध ।

सुरक्षित और विश्वसनीय एआई स्तंभ के अंतर्गत जारी पहली अभिरुचि अभिव्यक्ति (ईओआई) के अंतर्गत चयनित परियोजनाओं का विवरण

विषय का नाम	**चयनित आवेदक**	**परियोजना का शीर्षक**	**परियोजना के विषय में**
मशीन अनलर्निंग	आईआईटी जोधपुर	जनरेटिव फाउंडेशन मॉडल में मशीन अनलर्निंग	समग्र मॉडल प्रदर्शन पर नकारात्मक प्रभाव को कम करते हुए ओपन-सोर्स जनरेटिव फाउंडेशन मॉडल में लक्षित अनलर्निंग के लिए नवीन विधि विकसित करना।
सिंथेटिक डाटा जनरेशन	आईआईटी रुड़की	डेटासेट में पूर्वाग्रह को कम करने के लिए सिंथेटिक डेटा उत्पन्न करने के लिए विधि का डिजाइन और विकास; और उत्तरदायी एआई के लिए मशीन लर्निंग पाइपलाइन में पूर्वाग्रह को कम करने के लिए फ्रेमवर्क	एमएल मॉडल विकास के मॉडल प्रशिक्षण और इन-प्रोसेसिंग चरण में पूर्वाग्रह से निपटने के लिए एल्गोरिथ्म और विधि विकसित करना
एआई पूर्वाग्रह शमन रणनीति	राष्ट्रीय प्रौद्योगिकी संस्थान रायपुर	स्वास्थ्य देखभाल प्रणालियों में पूर्वाग्रह शमन के लिए जिम्मेदार कृत्रिम बुद्धिमत्ता का विकास	चिकित्सा प्रणाली अनुप्रयोगों, छवि विश्लेषण और नैदानिक निर्णयों में पूर्वाग्रहों को कम करने वाले उत्तरदायी एआई एल्गोरिदम विकसित करना।
व्याख्यात्मक एआई ढाँचा	डीआईएटी पुणे और माइंडग्राफ टेक्नोलॉजी प्राइवेट लिमिटेड	सुरक्षा के लिए व्याख्यात्मक और गोपनीयता संरक्षण एआई को सक्षम करना।	ऐसे एआई मॉडल बनाना जो भीड़-भाड़ वाले वातावरण में प्रभावी सुरक्षा के लिए मानव गतिविधि विश्लेषण के लिए सटीक और व्याख्यात्मक परिणाम प्रदान करते हैं।
गोपनीयता संवर्धन रणनीति	आईआईटी दिल्ली, आईआईआईटी दिल्ली, आईआईटी धारवाड़ और दूरसंचार इंजीनियरिंग केंद्र (टीईसी)	सुदृढ़ गोपनीयता-संरक्षण मशीन लर्निंग मॉडल	सुदृढ़ वितरित/फ़ेडरेटेड लर्निंग एल्गोरिदम विकसित करना जो हमलों के लिए अतिसंवेदनशील प्रतिकूल वातावरण में अच्छा प्रदर्शन करते हैं
एआई नैतिक प्रमाणन ढाँचा	आईआईआईटी दिल्ली और दूरसंचार इंजीनियरिंग केंद्र (टीईसी)	एआई मॉडल की निष्पक्षता का आकलन करने के लिए उपकरण	भारतीय संदर्भ में एआई सिस्टम के लिए निष्पक्षता सुनिश्चित करने के लिए पूर्वाग्रह जोखिम मूल्यांकन, मैट्रिक्स के लिए प्रसंस्करण और पूर्वाग्रह परीक्षण को शामिल करते हुए त्रि-चरणीय प्रमाणन प्रक्रिया विकसित करना
एआई एल्गोरिथम ऑडिटिंग टूल	सिविक डेटा लैब्स	परखएआई – सहभागी एल्गोरिथम ऑडिटिंग के लिए एक ओपन-सोर्स फ्रेमवर्क और टूलकिट	प्रस्तावित ढांचा और टूलकिट नागरिकों को एल्गोरिथम निर्णय लेने वाली प्रणालियों के जिम्मेदार डिजाइन, विकास और परिनियोजन में शामिल करने में सक्षम करेगा।
एआई गवर्नेंस टेस्टिंग फ्रेमवर्क	अमृता विश्व विद्यापीठम और दूरसंचार इंजीनियरिंग केंद्र (टीईसी)	ट्रैक-एलएलएम, पारदर्शिता, जोखिम मूल्यांकन, संदर्भ और लार्ज़ भाषा मॉडल के लिए ज्ञान	एलएलएम के डाउनस्ट्रीम उपयोग-मामले और नियोजन से संबंधित मौजूदा शासन परीक्षण ढांचे में विशिष्ट गैप की पहचान करना और उनका समाधान करना।

****