भारतजेन: भारत का पहला AI आधारित मल्टीमॉडल लार्ज लैंग्वेज मॉडल

स्रोत: पी.आई.बी

केंद्रीय विज्ञान एवं प्रौद्योगिकी राज्य मंत्री (स्वतंत्र प्रभार) ने "**भारतजेन (BharatGen)**" – **भारतीय भाषाओं के लिये देश का पहला स्वदेशी रूप से विकसित, सरकारी वित्तपोषित, आर्टिफिशियल इंटेलिजेंस आधारित मल्टीमॉडल लार्ज लैंग्वेज मॉडल** का शुभारंभ किया।

भारतजेन (BharatGen):

- **परिचय:** यह **22 भारतीय भाषाओं** में **भारत का पहला स्वदेशी रूप से विकसित** , **सरकार द्वारा वित्तपोषित मल्टीमॉडल** लार्ज लैंग्वेज मॉडल है।
    - **मल्टीमोडल LLMs (लार्ज लैंग्वेज मॉडल्स)** वे बड़े भाषा मॉडल हैं जिन्हें विभिन्न प्रकार के डेटा **(जैसे- टेक्स्ट, इमेज, ऑडियो और वीडियो)** पर प्रशिक्षित किया जाता है। जो उन्हें जटिल मानव भाषा तथा मल्टीमीडिया को समझने एवं व्याख्या करने में सक्षम बनाता है।
        - अनेक डेटा प्रकारों में सुसंगत प्रतिक्रिया प्रदान करने की उनकी क्षमता, उन्हें यूनिमॉडल की कमियों को दूर करने में मदद करती है, जैसे कि ChatGPT के पिछले संस्करण।
- **विकास:** नेशनल मिशन ऑन इंटरडिसिप्लिनरी साइबर-फिजिकल सिस्टिम्स (NM-ICPS), **IIT बॉम्बे में IoT और IoE के लिये TIH फाउंडेशन** द्वारा कार्यान्वित।
    - **NM-ICPS को वर्ष 2018** में **विज्ञान और प्रौद्योगिकी मंत्रालय** द्वारा साइबर-भौतिक प्रणालियों (CPS) तथा **नए युग की प्रौद्योगिकियों में नवाचार तथा अनुसंधान एवं विकास** को बढ़ावा देने के लिये लॉन्च किया गया था।
- **उद्देश्य:** भारतीय मूल्यों पर आधारित **नैतिक, समावेशी, बहुभाषी AI को बढ़ावा देना, स्वास्थ्य सेवा, कृषि, शिक्षा और शासन में क्षेत्र-विशिष्ट समाधान** प्रदान करना तथा स्थानीय भाषा बोलने वाले AI संचालित **ग्रामीण टेलीमेडिसिन** को बढ़ावा देना।
- यह पहल **विज्ञान और प्रौद्योगिकी विभाग (DST)** द्वारा समर्थित है तथा इसमें प्रमुख शैक्षणिक संस्थानों, विशेषज्ञों एवं नवोन्मेषकों का एक मज़बूत गठबंधन शामिल है।

विशेषता/पहलू	लार्ज लैंग्वेज मॉडल (LLM)	जनरेटिव एडवर्सरियल नेटवर्क (GANs)	ऑटोरिग्रैसिव मॉडल (ARMs)
परिभाषा	मानव जैसी भाषा उत्पन्न करने के लिये बड़े टेक्स्ट डेटा पर प्रशिक्षित AI मॉडल	दो नेटवर्क (जेनरेटर और डिस्क्रिमिनेटर) वाले AI मॉडल जो यथार्थवादी सामग्री उत्पन्न करते हैं	मॉडल जो पिछले अनुक्रम के आधार पर अगले मूल्य/टोकन की पूर्वानुमान करते हैं
मुख्य उद्देश्य	टेक्स्ट जनरेशन, ट्रांसलेशन, सारांशीकरण	इमेज जनरेशन, डीपफेक, डेटा संवर्द्धन	सीक्वेंस मॉडलिंग (टेक्स्ट, स्पीच, टाइम सीरीज़)
सामग्री प्रकार	**प्राइमरी टेक्स्ट**	प्राइमरी इमेज, वीडियो या ऑडियो	कोई भी सीक्वेंसींग डेटा (टेक्स्ट, स्पीच, ऑडियो)
जनरेटिव AI से संबंध	टेक्स्ट के लिये जनरेटिव AI का एक **उपसमूह**	मीडिया सामग्री के लिये एक **प्रकार का जनरेटिव AI**	एलएलएम और टाइम सीरीज़ मॉडल दोनों में उपयोग की जाने **वाली** तकनीक
उदाहरण	GPT-4, PaLM2, LLaMA	StyleGAN, CycleGAN	GPT, WaveNet, PixelRNN

और पढ़ें: लार्ज लैंग्वेज मॉडल, नेशनल मिशन ऑन इंटरडिसिप्लिनरी साइबर-फिजिकल सिस्टिम्स