GOVERNMENT OF INDIA
MINISTRY OF SCIENCE AND TECHNOLOGY
DEPARTMENT OF SCIENCE AND TECHNOLOGY
RAJYA SABHA
UNSTARRED QUESTION No. 2844
ANSWERED ON 19/12/2024

LAUNCH OF BHARATGEN

2844. SHRI MADAN RATHORE:
SHRI NARHARI AMIN:
DR. MEDHA VISHRAM KULKARNI:

Will the Minister of SCIENCE AND TECHNOLOGY be pleased to state:
(a)   the manner in which BharatGen will ensure that generative AI models adequately represent the country's diverse linguistic landscape and the steps being taken to include lesser represented languages; and
(b)   the manner in which BharatGen will prioritise making its generative AI models accessible as a public good, particularly for marginalised communities?

ANSWER
MINISTER OF STATE (INDEPENDENT CHARGE) FOR THE
MINISTRY OF SCIENCE AND TECHNOLOGY & EARTH SCIENCES
(DR. JITENDRA SINGH)

(a)   BharatGen is a multimodal multilingual large language model initiative, developing advanced generative AI models tailored to India's linguistic, cultural, and socio-economic diversity. To ensure that generative AI models adequately represent India's diverse linguistic landscape, BharatGen has launched an initiative called "Bharat Data Sagar", focusing on primary data collection. This data collection attempts to meet the requirement that training data is available for Indian languages that are lesser represented in data corpora.

(b)   BharatGen is building partnerships with research groups across the country, to ensure that the generative AI models being developed can be extended by partners and made available to the larger research and non-academic community for further development and usage. BharatGen is also developing partnerships with the government, industry and start-ups for applications geared towards efficient administration and public at large, including marginalized and underrepresented communities in the country.

*****