BHARATGEN – SCIENCE & TECHNOLOGY

News: "BharatGen" world's first government-funded multimodal LLM initiative launched by the Ministry of Science & Technology.

WHAT'S IN THE NEWS?

About BharatGen

- **Aim:** To revolutionize public service delivery and enhance citizen engagement by developing foundational models in language, speech, and computer vision.

- **Implementation:** By IIT Bombay under the National Mission on Interdisciplinary **Cyber-Physical Systems (NM-ICPS)**

- **Key Features of BharatGen:**

  - Multilingual and multimodal foundation models.

  - Building and training based on India-centric datasets.

  - Open-source platform for fostering AI research and innovation.

- The project is expected to be **completed by 2026**, with ongoing research, development, and scaling of AI applications.

Significance

- BharatGen will address both text and speech, ensuring representation across India's diverse linguistic landscape. By using multilingual datasets, it will capture the nuances of Indian languages, which are often underrepresented in global AI models.

- This emphasis on data sovereignty gives India greater control over its digital resources and narrative.

- BharatGen will **democratize AI access** across government, education, and private sectors, ensuring AI benefits all segments of society, particularly underserved Indian languages.

- BharatGen aligns with the vision of Atmanirbhar Bharat by developing AI models specifically for India. By building these technologies domestically

What are Large Language Models?

- Large language models, also known as LLMs, are very **large deep learning models** that are **pre-trained on vast amounts of data.**

- Large Language Models (LLMs) use **machine learning techniques** to recognize, interpret, and generate human languages or other complex data.

- Their capabilities also extend to **handling structured and unstructured data**, including speech, images, and other multimodal inputs, which enhances their utility in fields like customer service, healthcare, and education.

Generative AI (GenAI)

- It is an Artificial Intelligence (AI) technology that automatically generates content in response to prompts written in natural language conversational interfaces.
- Rather than simply curating existing web pages, by drawing on existing content, GenAI actually produces new content.
- The content can appear in formats that comprise all symbolic representations of human thinking: texts written in natural language, images (including photographs to digital paintings and cartoons), videos, music and software code.
- GenAI is trained using data collected from web pages, social media conversations and other online media. It generates its content by statistically analysing the distributions of words, pixels or other elements in the data that it has ingested and identifying and repeating common patterns.
- In November 2022, OpenAI released ChatGPT (Chat Generative Pre-trained Transformer) to the public.