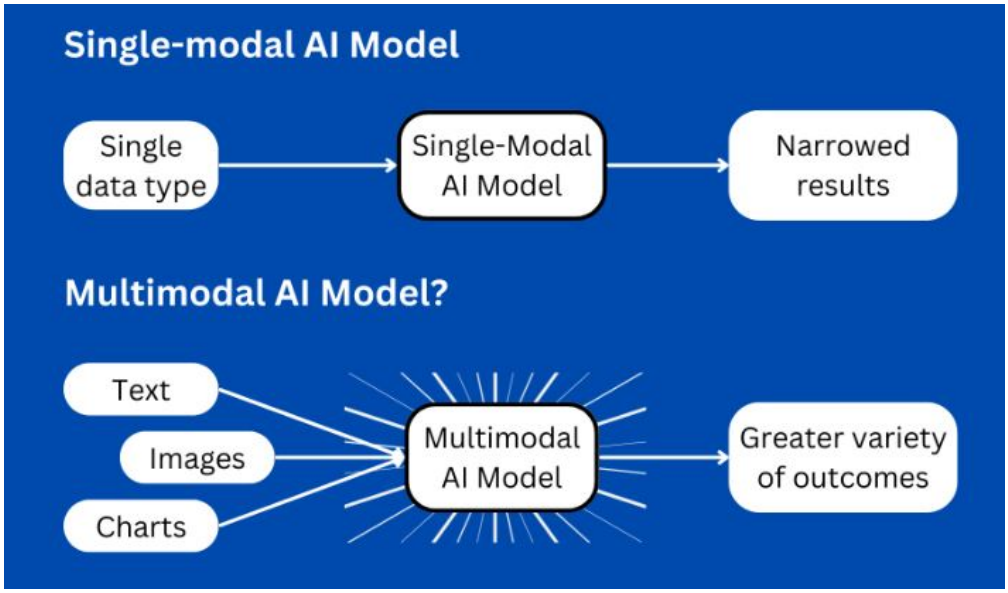
Multimodal Artificial Intelligence

Why in news?

Recently Microsoft-backed OpenAI made ChatGPT multimodal giving the bot the ability to analyse images and speak to users via its mobile app.

What is Multimodal AI?

- **Multimodal AI** - It is artificial intelligence that combines *multiple types or modes of data* to create more accurate determinations, draw insightful conclusions or make more precise predictions about real-world problems.
- Multiple modalities include video, audio, speech, images, text and a range of traditional numerical data sets.
    - Example - OpenAI's text-to-image model, ***DALL.E***, upon which ChatGPT's vision capabilities are based, is a multimodal AI model that was released in 2021.
- **Coventional AI** - Most AI systems today are *unimodal* that are designed and built to work with one type of data exclusively.
    - Example- ***ChatGPT*** uses *natural language processing* ***(NLP)*** algorithms to extract meaning from text content, and the only type of output the chatbot can produce is text.



How does multimodality work?

- Multimodal AI architecture typically consists of the following components.
- **Input module** – It consists of unimodal neural networks which receive and pre-process different types of data separately.
- This module may use different techniques, such as natural language processing or computer vision, depending on the specific modality.

- **Fusion module** – It is meant for integrating information from multiple modalities, such as text, images, audio, and video.
- Its goal is to capture the relevant information from each modality and combine it in a way that leverages the strengths of each modality.
- **Output module** – It is responsible for generating the final output or prediction based on the information processed and fused by the earlier stages of the architecture.

The Artificial Intelligence Race
• ChatGPT-maker *OpenAI* has announced that it had enabled its GPT-3.5 and GPT-4 models to study images and analyse them in words, while its mobile apps will have speech synthesis so that people can have full-fledged conversations with the chatbot. • The *Microsoft*-backed company had promised multimodality during the release of GPT-4. • *Google's* new yet-to-be-released multimodal large language model called ***Gemini,*** is already being tested in a bunch of companies. • *OpenAI* is also reportedly working on a new project called ***Gobi*** which is expected to be a multimodal AI system from scratch, unlike the GPT models.

What are the advantages of multimodal AI over the current AI?

- **Versatility-** It can handle multiple types of data, making it more adaptable to different situations and use cases.
- **Natural interaction-** By integrating multiple modalities, multimodal AI can interact with users in a more natural and intuitive way, similar to how humans communicate.
- **Improved accuracy**- Multimodal AI can also improve the accuracy of its predictions and classifications.
- **Enhanced user experience-** It can enhance the user experience by providing multiple ways for users to interact with the system.
- **Robustness against noise-** Multimodal AI can be more robust against noise and variability in the input data.
    - For example, in a speech recognition system, the system can continue to recognize speech even if the audio signal is degraded or the speaker's mouth is partially obscured.
- **Efficient usage of resources**- It can help to make more efficient use of computational and data resources by enabling the system to focus on the most relevant information from each modality.
- **Better interpretability-** It can help to improve interpretability by providing multiple sources of information that can be used to explain the system's output.

What are the applications of multimodal AI?

- **Healthcare-** It can help doctors and patients communicate more effectively, especially for those who have limited mobility or are non-native speakers of a language.
    - According to a report, the healthcare industry is expected to be the largest user of multimodal AI technology, with a CAGR of 40.5% from 2020 to 2027.
- **Education-** It can improve learning outcomes by providing more personalized and interactive instruction that adapts to a student's individual needs and learning style.
- **Entertainment-** It can create more immersive and engaging experiences in video

games, movies, and other forms of media.

- **Agriculture-** It can help monitor crop health, predict yields, and optimize farming practices.
    - By integrating satellite imagery, weather data, and soil sensor data, farmers can gain a richer understanding of crop health and optimize irrigation and fertilizer application, resulting in improved crop yields and reduced costs.
- **Manufacturing**-It can be leveraged to improve quality control, predictive maintenance, and supply chain optimization.
    - By incorporating audio visual data, manufacturers can identify defects in products and optimize manufacturing processes, leading to improved efficiency and reduced waste.
- **Voice assistants-** It can enable more sophisticated and personalized voice assistants that can interact with users through speech, text, and visual displays.
- **Smart homes**- It can create more intelligent and responsive homes that can understand and adapt to a user's preferences and behaviours.
- **Virtual shopping assistants**- It can help customers navigate and personalize their shopping experience through voice and visual interactions.
- **Law and order**- Microblogging platform X has updated policies to fight a stream of misleading videos and hate speech on the platform since the renewed conflict between Israel and Palestine.
- In 2020, Meta was working on a multimodal system to automatically detect hateful memes on Facebook.

What lies ahead?

- Businesses can unlock insights that were previously hidden, enabling them to make better decisions and improve outcomes, by combining the strength of different modalities.
- As technology continues to evolve and become more advanced, we can expect even greater innovation and impact from multimodal AI in the years ahead.

References

1. The Hindu- What is multimodal AI and why it is significant
2. The Hindu- What is multimodal AI