Perspective - AI: The Road Ahead

Why in News?

Recently, India, along with **27** other nations and the **European Union,** reached a historic agreement at **Bletchley Park, United Kingdom**, emphasizing a shared understanding of Artificial Intelligence's **potential** and **risks.**

- This landmark agreement, known as the **Bletchley Declaration**, marks a crucial step towards addressing the multifaceted aspects of **AI.**

What is the Bletchley Declaration?

- The **Bletchley Park Declaration** is the first global pact on tackling **frontier AI risks,** and it reflects a high-level political consensus and commitment among the major AI players in the world.
- It acknowledges the potential of AI to enhance human well-being but also recognizes the risks posed by AI, especially frontier AI, which may cause serious harm, either deliberate or unintentional, particularly in domains like cybersecurity, biotechnology, and **disinformation.**
- It emphasizes the need for international cooperation to address AI-related risks, as they are inherently global, and calls for collaboration among all actors, including **companies, civil society,** and **academia.**
- The declaration also announces the establishment of a regular **AI Safety Summit,** which will provide a platform for dialogue and collaboration on frontier AI safety.

What is Artificial Intelligence (AI)?

- **About:**
  - AI is the ability of a computer, or a robot controlled by a computer to do tasks that are usually done by humans because they require **human intelligence** and discernment.
    - Although there is no AI that can perform the wide variety of tasks an ordinary human can do, some AI can match humans in specific tasks.
- **Characteristics & Components:**
  - The ideal characteristic of artificial intelligence is its ability to rationalize and take actions that have the best chance of achieving a specific goal. A subset of AI is **Machine Learning (ML).**
    - **Deep Learning (DL) techniques** enable this automatic learning through the absorption of huge amounts of unstructured data such as **text, images**, or **video**.

What are the Important Initiatives and Summits Related to Artificial Intelligence?

- **Global Initiative**:
    - **Global Partnership on Artificial Intelligence (GPAI):** India is the lead chair of **GPAI** in **2024**. It is an alliance of **28 countries**; the European Union adopted the 'New Delhi Declaration' of the GPAI.
    - **UK AI Safety Summit:** It was held in 2023 at **Bletchley Park, England.** The summit focused on addressing safety and security aspects of AI, emphasizing the need for international cooperation.
        - The 28 major countries, including the **United States, China, India**, and the **European Union,** signed the **Bletchley Park Declaration at this first-ever AI Safety Summit.**
    - **G7 AI Summit**: The **G7 nations** have been engaging in meetings and discussions to collaboratively tackle the global challenges presented by AI. This summit took place in Hiroshima, Japan, and concluded in December 2023. Its summit **marked the beginning of the Hiroshima AI Process (HAP),** representing a major advancement in the regulation of AI.
        - The **G7 Leaders'** recognised the importance of **inclusive AI governance** and set forth a **vision of trustworthy AI** aligned with shared democratic values.
- **National Initiative:**
    - **National Program on Artificial Intelligence (India):** India's domestic initiative focusing on responsible AI, innovation, and balancing the benefits of AI with potential risks.
    - **Future Skills Prime:** A program in India that promotes reskilling and upskilling in AI-related skills to prepare the workforce for the changing job landscape.
    - **US India Artificial Intelligence (USIAI):** This initiative has been launched to scale up the science and technology relationship between India and the United States.
    - **Responsible AI for Youth:** The government has launched a National Program for the youth namely, 'Responsible Artificial Intelligence (AI) for Youth'.
        - The government has also launched India's **National Artificial Intelligence Portal.**

What are the Challenges Associated with Artificial Intelligence?

- **Dealing with Bias in AI:** Biases in AI, originating from biased data and human annotations, pose challenges. Ongoing efforts involve researching, implementing auditing mechanisms, and establishing evaluation structures to study and address biases in AI models.
- **AI in Services Sector**: AI may lead to job losses, particularly in customer relationship management and creative writing in the services sector. The focus is on retraining workers to cope with the evolving job market.
- **Cybersecurity:** AI can be used to create sophisticated cyberattacks that are harder to detect and defend against. Malicious actors could use AI to launch targeted attacks against critical infrastructure, steal sensitive data, or disrupt essential services.
- **Misinformation:** AI can be used to create and spread misinformation at an alarming rate. Fake news articles, social media bots, and deepfakes can all be used to manipulate public opinion and sow discord. This can have serious consequences for elections, public health, and social cohesion.
- **Deepfakes:** Deepfakes are realistic video or audio recordings that have been manipulated to make it appear as if someone is saying or doing something they never did. These can be used to damage reputations, spread misinformation, or even commit fraud.
- **Influencing Public Opinion**: AI can be used to target individuals with personalized propaganda and advertising. This can be used to manipulate people's voting behavior, sway public opinion on important issues, and even radicalized individuals.

Way Forward

- **India's National Strategy on AI**: India's strategy for AI emphasizes responsible and innovative deployment, seeking a balance between fostering innovation and managing associated risks. The national AI program strives to utilize AI for enhancing social empowerment and the delivery of

public services.

- **Ethical AI Frameworks:** International collaboration is crucial for establishing ethical frameworks for AI development and deployment. Shared values and guidelines can help mitigate the negative impacts of AI, such as biases and discrimination, ensuring fair and responsible AI practices.
- **Democratization of AI:** Collaborative initiatives can help bridge the AI divide by ensuring that the benefits of AI are not concentrated among a select few nations or corporations. By promoting inclusivity and sharing resources, international collaboration supports a more equitable distribution of AI advancements.
- **Ensuring Inclusive Benefits:** There is a concern about the potential concentration of AI benefits in the hands of a select group or large corporations. The conversation emphasizes the need to ensure that the benefits of AI reach all segments of society, including marginalized communities and rural areas.
- **AI Centric Future:** The vision is of a future where humans and AI work together, each complementing the strengths of the other. A collaborative approach needed to increase overall productivity and creativity globally.

UPSC Civil Services Examination, Previous Year Questions (PYQs)

Q. With the present state of development, Artificial Intelligence can effectively do which of the following? (2020)

1. Bring down electricity consumption in industrial units
2. Create meaningful short stories and songs
3. Disease diagnosis
4. Text-to-Speech Conversion
5. Wireless transmission of electrical energy

Select the correct answer using the code given below:

(a) 1, 2, 3 and 5 only
(b) 1, 3 and 4 only
(c) 2, 4 and 5 only
(d) 1, 2, 3, 4 and 5

Ans: (b)