TARGET PRELIMS 2024
BOOKLET-4; S&T-4
COMPUTER, IT: AI, ML, CHATGPT ETC.

1. TABLE OF CONTENTS

2. ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

» Intro

- **Artificial Intelligence** is the science and engineering of making intelligent machines, especially intelligent computer programs which can complete tasks that typically require human intelligence.
  - » With the **explosion of available data and expansion of computing capacity**, the world is witnessing rapid advancements in AI, ML, and deep learning.

- **Machine learning** is a science that involves **development of self-learning algorithms**. Machine learning uses **statistics (mostly inferential statistics)** to develop self-learning algorithm. It is a type of artificial intelligence.
  - » **Note:** All Machine Learning is AI, but not all AI is machine learning

  - » For e.g., symbolic logic (rules engines, expert systems, and knowledge graphs) as well as evolutionary algorithms and Bayesian statistics could all be described as AI, and none of them are machine learning.

  - » In Machine Learning the computer program should learn from experience "i.e., given data" such that the overall performance on doing a certain task increase.
    i. Input data
    ii. Model Training
    iii. Output

- **Applications of Artificial Intelligence** and **Machine Learning**
  - Advertisements, Online shopping suggestions etc.
  - Spam filtering
  - Search engines
  - **Fighting Black Money** (e.g., Project Insight of India)
  - **Space Exploration** (e.g., identifying exoplanets from pictures)
  - **Health Sector**:
    - **Diagnosis:** E.g., a Bengaluru based startup has developed a non-invasive, AI-enabled technology to screen for early signs of breast cancer.
    - **Treatment**: AI powered Clinical Decision Support (CDS) tools can aid in developing appropriate and accurate diagnostic and treatment recommendations. E.g. Apollo hospital has launched Apollo Clinical Intelligence Engine, a CDS, open to use by all Indian doctors.
    - **Supply chain resilience**: By accurately predicting the demand and supply for medicines.
    - **Development of new Medicines/Molecules** – For e.g. AI can help in identifying and studying new molecules.
    - **Improvement in Governance: E.g.** For **COVID-19,** AI enabled chatbot was used by MyGov for ensuring communications.
  - **Developing new materials** (E.g. Google Deepmind predicted the structures of 2 million new materials)
  - **Education** (e.g., Personalized learning through adaptive tools; customizing professional development courses etc.)
  - **Agriculture Sector**:

- Tech like image recognition, drones etc can help farmers kill weeds more effectively to increase productivity.
- **Efficient resource utilization –** AI enabled solution for water management crop insurance etc are also being developed.
- **AI Powered decision making**: For e.g: **ICRISAT** has developed an **AI-power sowing app**, which utilises weather models and data on local crop yield and rainfall to predict and advise local farmers on when they should plant their seeds more accurately**.**
- **AI4AI (AI for Agriculture Innovation)** initiative has been launched by the WEF to transform agriculture sector in India. Under this, '**Saagu-Baagu**' initiative has been launched in the state of Telangana.

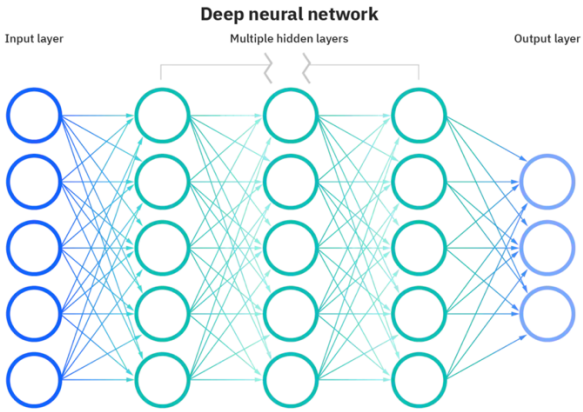
- **Disaster Management:** An AI-based flood forecasting system has been deployed in Bihar and is now being deployed throughout the country. It gives warnings 48 hours earlier about impending floods.
- **Improve Ease of Doing Business**
- Natural Language Processing (NLP)
- Image Processing (Facial Recognition)

1) ADVANCEMENTS IN MACHINE LEARNING

A) NEURAL NETWORKS

- Neural network, also known as Artificial Neural Network (ANNs) or simulated neural networks (SNNs), are a subset of machine learning and are at the heart of deep learning algorithms. Their name and structure are inspired by the human brain, mimicking the way biological neurons signal to each other.
- A neural network can fine tune its output based on the feedback given to it during stages of training.

- ANNs consist of node layers, containing an input layer, one or more hidden layers, and an output layer. Each node, or artificial neurons, connects to another and has an associated weight and threshold. If the output of any individual node is above the specified threshold value, that node is activated, sending data to the next layer of the network. Otherwise, no data is passed along the next layer of the network.



- **Note:** ANN also rely on training data to learn and improve their accuracy over time.

- **Neural Networks** vs. **Deep Learning**:
  - Terms are sometimes used interchangeably. 'Deep' in deep learning is just referring to the depth of layers in a neural network. A neural network that consists of more than three layers – which would be inclusive of the inputs and output – can be considered a deep learning algorithm. A neural network that only has two or three layers is just a basic neural network.

B) DEEP LEARNING

- Deep learning is a machine learning technique that teaches computers to do what comes naturally to humans: learn by example. In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound. It can achieve state of art accuracy, sometimes exceeding human-level performance. **Models are trained by using a large set of labeled data** and **neural network architecture that contain many layers**.
- Most deep learning methods use **neural network architecture**, which is why deep learning models are often referred as **Deep Neural networks**. The term deep usually refers to number of hidden layers in the neural network.

» Some Criticism of AI
  - Idea of intelligent machines is obscene anti human and immoral.
  - Would make life more mechanical.
  - A lot of investment has taken place -> many AI companies going bankrupt
  - Taking away the human jobs

C) GENERATIVE ARTIFICIATION INTELLIGENCE LIKE CHATGPT (CHAT GENERATIVE PRE-TRAINED TRANSFORMER)

ABOUT CHATGPT:

It is an **AI tool** developed by **OpenAI**.

> **OpenAI** is a research institution and company that focuses on developing AI intelligence technology in a responsible and safe way. It was founded in 2015 by a group of entrepreneurs and researchers, including **Elon Musk, Sam Altman**, and **Greg Brockman**.

- ChatGPT is based on **Generative Pre-trained Transformer Architecture**.
  - It is **trained on massive amount of text data from the internet**. It used 570 GB of text data mined from the internet.
  - It is a type of **neural network** and was first introduced in 2017 in a paper titled "Attention is all you need". A neural network can fine tune its output based on the feedback given to it during stages of training. This allows the model to better understand the context and meaning of the input and to generate **conversational response**.
  - Thus, we can say that ChatGPT is fine tuned to provide **conservational responses**, as against essay-type content. It is because the neural network behind it has been **additionally trained on conversational transcripts with human feedback**.

- But it is **more than a chatbot.** It can do tasks like writing software applications, new poems, stories etc.
- ChatGPT can become a **powerful pedagogy** tool on any topic to anyone, because we can instruct it to "explain it to me like I am a six-year-old". It can explain in simple terms anything from **philosophy** to cooking recipes, including **new recipes of its own**.

It is a Language Model (rather than a chatbot) that can produce text that sound like human response in a conversation setting.

What is language model?

It is a software that prints out a sequence of words as output that are related to some words given as input with appropriate semantic relation. In practical terms, it means that it can **perform tasks like answering questions and carrying on a conversation with humans**. It is often used in Natural Language Processing (NLP) applications, such as speech recognition, automatic translation, and text generation.

It is also a **Neural Network**

It can be thought of as a large network of computers that can fine tune its output of words based on the feedback given to it during stages of training; this training process and the technology together are called **Reinforcement training**. The input data is typically huge corpus of text.

Another key idea of "**Word embedding**" has been used. It represents words as a matrix of numbers that can be manipulated inside computers. When a neural network processes these numbers, it can differentiate words according to different contexts: for example, when "shoot" appears with "gun" the neural network knows that the words that will follow may mostly be "bullets" or "victims", whereas when "shoot" appears with "camera", the neural network knows that the following words may be "picture" or "pixel".

With a further refining technique called "**Transformer**", a neural network can accurately understand the context of a sentence or a paragraph. This "comprehension" can be used for multiple purposes like answering a question, summarising a paragraph or an article, translating documents and so on.

GOOGLE BARD

Google's Generative AI model

ABOUT GOOGLE GEMINI (DEC 2023)

- Google GEMINI is a new multimodal general AI model, which the tech giant calls its most powerful yet.
- It is now available to users through Bard, some developer platforms, and even the new Google Pixel 8 Pro phones.
- The flexible AI model comes in **three sizes** – Ultra (yet to be released), Pro, and Nano – is being seen as Google's answer to ChatGPT, which has been ahead of the game so far when it comes to generative AI.

- Google claims that GEMINI Ultra is the first model to outperform human experts on massive multitask language understanding (MMLU), which uses a combination of 57 subjects such as math, physics, history, law, medicine, and ethics for testing both world knowledge and problem-solving abilities.
- **So**, IS GEMINI better than ChatGPT 4?
  - **Hard to say now**. But it does seem to be more flexible. Its ability to work with videos and on devices without internet, gives it some edge.

- **Some Concerns** about Generative AI:
  - **Teachers are unhappy about it** as they feel that it can be used to turn in plagiarized essays which could be hard to detect for invigilators. Recently, New York City's Education department banned ChatGPT in its public schools.
  - **Skilled white color jobs** like that of computer programmers in the IT sector is at threat.
  - **India's IT services-based** exports may get impacted.

D) MULTIMODAL AI

- **Definition**: Multimodal AI is a type of AI that can **process and understand** information from multiple types of sources like text, images, audio, and video. By integrating information from different sources, multimodal AI aims to enhance the system's ability to perceive and comprehend the world in a more holistic and human like manner. It is like brain that can see, hear, and read all at the same time.

- **Advantages: Multimodal AI can do several things which traditional AI can't**:
  - **Understand the meaning of a video**: By combining audio and video, the multimodal AI will be able to tell you what is happening in the video, who the people are etc.
  - **Generate more realistic images**: This is because this AI will consider of things like lighting, shadows, reflections etc.
  - **Create more natural sounding speeches** – It is because the AI will be able to take into consideration the emotions and context of the conversation.
  - **Important areas where they can be used?**
    - **Processing CT scans or identifying rare genetic diseases** all need AI systems that can analyze complex datasets of images, and then respond in plain words.

- **E.g.** Gemini is Google's multimodal large language model.
- **OpenAI** is also reportedly working on a new project called **Gobi** which is expected to be a multimodal AI system from scratch, unlike GPT models.

E) GOOGLE DEEPMIND AI BREAKTHROUGH (NOV 2023)

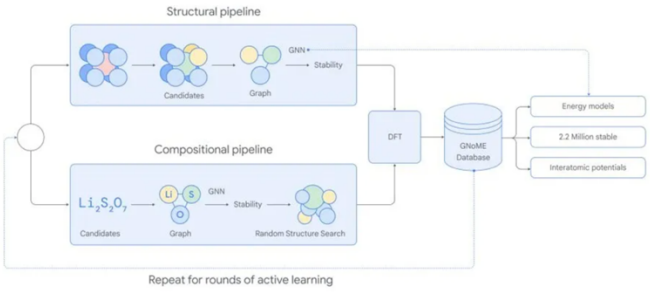
» **How are new materials discovered in Chemistry -> Trial and Errors -> Expensive and time-consuming process.**
» **In last decades,** experimentation by humans has resulted in the discovery of the structures of some 28,000 stable materials, which are listed in the Inorganic Crystal Structure Database, the largest database of identified materials.

» **What is DeepMind AI breakthrough?**
  » Google DeepMind AI Tool known as **Graph Networks for Material Exploration (GNoME)** has successfully predicted the structures of more than **2 million new materials**. This was done with the help of AI.
  » While these materials will still need to undergo the process of synthesis and testing, DeepMind has published a list of **381,000 of the 2.2 million crystal structure that it predicts to be most stable**.

» **How does GNoME actually work?**
  » GNoME is a state of art **graph neural network model or GNN**, where the input data for the model takes the form of a graph that can be likened to connections between atoms.

  » GNoME was **trained using active learning**, a technique to scale up a model first trained on a small, specialized dataset. Developers can then introduce new targets allowing machine learning to label new data with human assistance. This makes the algorithm well suited to the science of discovering new materials, which requires searching for patterns not found in original dataset.

  » **GNoME** uses two pipelines to discover low energy (stable materials).
    - The **structure pipeline** creates candidates with structures similar to known crystals.
    - The **composition pipeline** follows a more randomized approach based on chemical formulas.
    - The output of both the pipelines are evaluated using established **Density Function Theory (DFT) calculations** and those results are added to the GNoME database, informing the next round of active learning.



» **Significance**:
  - **Drastic increase in the number of 'stable materials' known to mankind by ten-fold**.
    - DeepMind claims its current research is equivalent to nearly 800 years of knowledge, given that 3,80,000 of its stable predictions are now publicly available to help researchers make further breakthrough in materials discovery teams.
  - **The breakthrough** has huge implications for sectors such as renewable energy, battery research, semiconductors, and computing efficiency which have been looking for new material to improve the efficiency in the sector.

F) PREDICTING PROTEIN STRUCTURE WITH AI

- The AI based program, **AlphaFold2**, from the company **DeepMind**, has stunned the world by **accurately and quickly predicting the structure of proteins**, starting from the sequence of amino acids that constitute them.

2) FACIAL RECOGNITION TECHNOLOGY (FRT)

- **FRT** is a type of biometric technology that identifies and verifies individuals by analysing and comparing patterns in their facial features.

- **How does FRT Work?**
  - **Data Acquisition**: It involves capturing a facial image or video of the person through cameras.
  - **Feature Extraction**: In this phase, various features of the face is extracted (e.g. the distance between the two eyes, shape of the nose, width of the jaw etc.)
  - **Feature Matching**: The extracted features are then matched with the database of existing pictures.
  - **Identification or verification**: Based on feature matching, the FT technology identifies a person as someone in the database or verifies that the person is who he claims to be.

- **Applications**
  - **Security and Law Enforcement**: Criminals could be identified from the crowd.
  - **Border Control**: FRT can be used to identify travelers at airports and border crossing.
  - **Biometric Authentication**: For e.g. FRT can be used for unlocking of phones.
  - **Marketing and Advertising**: FRT can be used to track users and user choices which can lead to better marketing
  - **Social Media and Tagging**: Social media platforms use facial recognition for photo tagging and to enhance user experience.

- **Concerns**
  - **Excessive surveillance and violation of Privacy**: Widespread use of facial recognition could lead to mass surveillance and a loss of individual privacy. It may lead to unauthorized tracking, profiling, and potential misuse of personal data.
  - **Technology challenges**:
  - FRT is **prone to digital attacks** or the use of physical or digital portraits, 3-D Models, such as deep-fakes etc.
  - **Accuracy** concerns: Sometimes poor accuracy can lead to wrong authentication.

A) ASTR TOOL OF DOT

- **why in news?**
  - Department of Telecommunication has developed an Artificial Intelligence-based **facial recognition tool** called **ASTR** (May 2023)

- **About ASTR**:
  - **Artificial Intelligence and Facial Recognition power Solution for Telecom Sim Subscriber Verification (ASTR)** can potentially bring down cyber frauds by detecting and blocking possible fraudulent mobile connections.

  - **How does it function**?

- In 2021, DoT had ordered all telecom operators that they would have to **share their subscriber database** including users' pictures with the department. These images constitute the core database on which **authorities are running their facial recognition algorithm using ASTR**.

- **How ASTR Functions?**
  - Human faces in subscribers' images are encoded using **Convolution neural network (CNN) models** in order to account for the tilt and angle of the face, opaqueness and dark color or the images.
  - After that, a face comparison is carried out for each face against all faces in the database, and similar faces are grouped under one directory.
  - Two faces are concluded to be identical by ASTR if they match to the extent of at least 97.5%.
- The DoT **allows an individual to take nine legitimate mobile phone connections** using a single identity proof. In essence, what the ASTR does is -1) it looks up if there are more than nine connections against a single individual's photographs; 2) it runs a search through the database to see if the same person has taken SIMs under different names.
- **Results**:
  - According to the Ministry of Communication, an analysis of more than 87 crore mobile connections was carried out using ASTR in the first phase, where more than 40 lakh cases of people using a single photograph to obtain connections were detected. After "**due verification**', more than **36 lakh connections** were discontinued.

B) DIGIYATRA: AIRPORTS USING FRT IN INDIA

- **What is DigiYatra?**
  - It is an initiative by GoI to make air travel and seamless and hassle free experience using digital technology. It envisages that travelers pass through various checkpoints at the airport through paperless and contactless processing, using facial features to establish their identity, which would be linked to the boarding pass.

- **How does it work?**
  - **Passenger Enrollment**: Passengers download the Digi Yatra app and link it to their Aadhaar card (a 12-digit unique ID). They can create a travel profile with their boarding pass and a self-image capture. These credentials are shared with airport authority.
  - **Facial Recognition**: At the airport, the passengers proceed to Digi Yatra Kiosk where their faced as scanned using a secure Facial Recognition tool. The system verifies the passenger's identity against their Aadhaar details stored in the app.
  - **Seamless Travel**: Once verified, passengers can simply walk through designated e-gates at various checkpoints **without needing to show any physical document**. The facial recognition system automatically grants them access.

- **Advantages**:
  - **Faster and smoother travel**; **Paperless travel**.
  - **Enhanced security**
  - **Data Privacy**

- **Who is running DigiYatra?**

- **DigiYatra Foundation**: It is a joint-venture company whose shareholders are the AAI (26%) and Bengaluru Airport, Delhi Airport, Hyderabad Airport, Mumbai Airport, and Cochin International Airport. These five shareholders equally hold the remaining 74% of the shares.

3) DEEPFAKES

- **Why in news?**
  - The Ministry of Electronics and Information Technology (MEITy) has sent an advisory to social media platforms on deepfakes (Dec 2023)
    - Earlier PM Modi had warned against Deepfakes calling on media to educate people on misinformation.
  - Following the controversy created by Deepfake videos of actress Rashmika Mandana and Katrina Kaif's deepfakes being circulated online, the GoI has asked social media companies to remove deepfake within 36 hours of a complaint being registered (Nov 2023)

- **Basics:** Deepfakes refer to manipulated media (audio, video, images etc) created using a form of Artificial intelligence called Deep Learning (or Deep Neural Network). This manipulated content use lip syncing, swapping of face etc. – mostly without consent.

- **How does the Deepfake technology work?**
  - The technology involves modifying or creating images or videos using a machine learning technique called **Generative Adversarial Network (GAN)**. The AI driven software detects and learns the subjects' movements and facial expressions from the source material and then duplicates this in another video or image.
  - **Larger the source material used**, better will be the quality of deepfake. Therefore, highest number of deepfakes are made of public figures like politicians and film stars.
  - Through a **collaborative work of two softwares**, the fake video is rendered until the second software package can no longer detect the forgery. This is known as **"unsupervised learning"** when machine language models teach themselves. The method makes it difficult for other software to identify deepfakes.

- **Advantages:**
  - Synthetic Media/ Deepfakes can create **possibilities and opportunities for all** people, regardless of how people listen, speak, or communicate. It can give people voice, purpose, and ability to make an impact at scale and with speed.
  - It has been used by the ALS association in collaboration with a company to **use voice cloning technology** to **help people with ALS digitally recreate their voices in future**.

- **Concerns:**
  - Like most new technologies, it can also be **weaponized to inflict harm** to individuals, institutions, businesses or a country.
  - **Crime against women** can increase with malicious use of Deepfakes in pornography and can inflict emotional, reputational and in some cases violent outcome for some individuals. (for e.g. viral deepfake video of actress Rashmika Mandana incident)
  - **Endanger Social Harmony** – Communal/caste-based statements.

- **Decrease trust towards institutions like government/media** – by propagating false propaganda against them.
- **Undermine democracy and impair diplomacy** – false information about institutions, public policy, and politicians powered by a Deepfakes can be exploited to spin the story and manipulate belief.

- **How to spot/identify a deepfake?**
  - Look for **unnatural blinking or lack of it**.
  - **Lighting** that just don't sit right.
  - Sometimes, voice could be too robotic.
  - It the video sounds too sensational to be true, trust your gut.
  - Voices that miss the mark on lip synchronization

- **Meity has sent another advisory to social media firms to comply with Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021** (Dec 2023)
  - The advisory was aimed at getting social media firms to **crack down more forcefully on 'deepfake' clips** of people.
  - It mandates that intermediaries **communicate prohibited content**, particularly those specified under Rule 3(1)(b) of the IT Rules, clearly and precisely to users.

- **Recent Advisory released by Ministry of electronics and Information Technology** (Nov 2023)
  - **IT Rules, 2021** require that all content reported to be fake or produced using deepfake be taken down by intermediary platforms within 36 hours.
  - An advisory was sent to social media platforms in Nov 2023, reminding them that they may lose "**safe harbour immunity"** under the IT Act, if they fail to remove within 36 hours deepfake content that has been reported.

A) HOW VOICE CLONING THROUGH ARTIFICIAL INTELLIGENCE IS BEING USED FOR SCAMS (JAN 2024)

- **Famous Examples**:
  - » In April 2023, a family living in Arizona, USA, was threatened to pay ransom for a **fake kidnapping** pulled off by an AI cloned voice.
  - » In Dec 2023, a Lucknow resident was duped to transfer a substantial amount through UPI.

- **India**:
  - » A **report**, titled '**The Artificial Imposter'** published in May 2023, revealed that 47% of surveyed Indians have either been a victim or knew someone who had fallen prey to an AI generated voice scam. Thus, numbers are almost twice the global average of 25%.
  - » In fact, India topped the list with the maximum number of victims to AI voice scams.

- **How are voice clones done?**
  - » Once a scammer finds an audio clip of an individual, there are host of online sites / applications like Murf, Resemble, and Speechify which can be used to generate voice clones.

- **Various real time translation tools are also available:**

» For e.g. recently Meta released **SeamlessM4T**, an open-source multilingual foundational model that can understand nearly 100 languages from speech or text and generate translation in real-time.
» Apple introduced a voice cloning feature in **iOS 7** intended to help people who may be in danger of losing their voice say to degenerative diseases.
» On 2nd of Jan 2024, MIT and Tsinghua University in Beijing, China, and members of AI Startup MyShell released **OpenVoice,** an open-source voice cloning tool that is almost instant and offers granular controls to modify one's voice that isn't found on other such platforms.

4) GPAI (THE GLOBAL PARTNERSHIP ON ARTIFICIAL INTELLIGENCE)

- **Why in news?**
  » Global Partnership on AI (GPAI) members unanimously adopt New Delhi Declaration on AI (Dec 2023)

- GPAI is an **international and multi-stakeholder initiative** to guide the **responsible development and use of AI**, grounded in human rights, inclusion, diversity, innovation, and economic growth.
  » This is also a first initiative of its type **for evolving better understanding of the challenges and opportunities around AI** using the experience and diversity of participating countries.
  » GPAI was first proposed by Canada and France in 2018 G7 summit, and was **officially launched in June 2020 with 15 members** (including India)
  » **Currently** (as of Dec 2023), it consist of 29 members (28 countries and EU).
    - **China**, a major techpower is not a part of the grouping.

  » It is supported by a Secretariat hosted by OECD, Paris.

- **Dec 2023 Meeting:**
  » India hosted the summit and will also chair GPAI in 2024.
  » This summit was important as it was the first summit after the explosive release of ChatGPT.
  » The GPAI has unanimously adopted '**New Delhi Declaration**'.

  » **Key Highlights of the New Delhi Declaration:**
    - It underscores the **need to mitigate risks** arising from the development and deployment of AI systems. It **flagged concerns emanating from such systems** including misinformation, unemployment, lack of transparency, and fairness, protection of IP and personal data and threat to human rights and democratic values.
    - It also **promotes equitable access** to critical resources for AI innovation including computing and high quality diverse data sets.
    - It also **fosters inclusivity** so that countries outside the purview of GPAI can also reap AI benefits.
    - It also says that global framework for the use of AI should be rooted in democratic values and human rights; safeguarding dignity and well-being; ensuring personal data protection; the protection of IPR etc.
    - Members also agreed to **support AI innovation** in the **agriculture sector** as a new 'thematic priority". Earlier GPAI themes include healthcare, climate action and building resilient society.

A) AI SAFETY SUMMIT AND BLETCHLEY DECLARATION (NOV 2023)

- AI **Safety Summit, 2023**
  - » AI Safety summit was an international conference discussing the safety and regulation of AI. It was held in the UK at **Bletchley Park on 1st and 2nd Nov 2023**.
  - » It was the **first ever global summit on AI** which is planned to become a recurring event.
  - » **27 countries** from across the globe including the US, the UK, China, Australia, and **India**, as well as **EU**, agreed on **Bletchley Declaration on AI Safety**.

- **Key Highlights: Bletchley Declaration**
  - » It aims to enhance global cooperation on (AI) safety.
  - » It has a **twofold focus**:
    1. **Identifying** shared AI-related risks and enhancing scientific understanding of these risks
    2. **Creating cross country policies** to address these risks.
  - » **Definition of Frontier AI**: Frontier AI refers to highly advanced generative AI models with potentially dangerous capabilities that can pose significant risk to public safety.

- **About Bletchley Park**: This is a site of historic importance in computing.
  - During WW-II, it played an important role in breaking the 'unbreakable' **Enigma Code** which was used by Nazis.
  - It also contributed to the development of the **Colossus** – often considered the world's first programmable electronic computer.

5) REGULATING ARTIFICIAL INTELLIGENCE

- **Why in news?**
  - » EU has reached a landmark agreement to regulate **AI** (Dec 2023)

- **Need of Regulating AI:**
  - » **Controlling Big-Techs**: Most of the advanced development in AI is taking place in the Big-Technology companies like Microsoft, Google, Meta etc who have access to immense data and computing power.
  - » **Controlling Misuse**: Frontier AI has led to increase in the risk of deepfakes, harmful information, and cyber frauds.
  - » **Negative impact on** economy: AI may pose a threat to jobs and inclusive development in future.
  - » **Preventing** violations of Privacy, IPR etc.
  - » **Model Collapse Scenario**: ML models train on Data sets. But AI generated Data sets may create discrepancies and incorporate mistakes of previous AI models.

- **EU has adopted the world's first law on regulating AI** in Dec 2023.
  - » The EU Parliament will now vote on the proposed act early next year (i.e. in 2024), but with the deal done, it's just a formality.

- **What does the EU law propose?**
  - » The law regulates the use of Artificial Intelligence (AI).
  - » It classifies AI systems in four categories based on the associated risks and provides for different level of regulation for each category.

» It includes safeguards on the use of AI within the EU, including **clear guardrails** on its adoption by law enforcement agencies.
    1. The deal includes **strong restrictions** on facial recognition technology, and on Using AI to manipulate human behaviour.
    2. Government can only use **real-time biometric surveillance in public areas** only when there are serious threats involved, such as terrorist attacks.
» **Provision for strong penalties**: The deal threatens stiff financial penalties for violations of up to 35 million euros or 7% of a company's global turnover.
» **Consumers** have been **empowered to launch complaints** against any perceived violations.

» The legislation also proposes to be "**a launch pad for EU start-ups and researchers to lead the global AI race**".
    1. The act works as a unique legal framework for the development of AI you can trust. It will help in development of technology which doesn't threaten people's safety and rights.

- **Significance**:
    » Strong and Comprehensive rules in EU can set a **powerful example** for many governments considering regulations.
    » **AI Companies** who follow these regulations in EU are also expected to extend some of these protections in other jurisdictions.

- **Comparing EU's approach with other regulations**:
    » **EU** has taken a **tougher stance** which segregates AI as per use case scenario based primarily on the degree of invasiveness and risk;
    » **UK** has seen regulation on the other end of the spectrum with a '**light-touch**' approach that aims to foster innovation in this nascent field.
    » **USA**'s approach lies in between that of EU and UK.

- **Leadership in tech regulation**:
    » Over the last decade, Europe has taken **decisive lead** over the US on tech regulation.
        1. EU has enforced the landmark **GDPR (General Data Protection Regulation) since May 2018.** It is an overarching law focused on privacy and requires individuals to give **explicit consent before** their data can be processed and is now a template being used by over **100 countries**.
        2. EU has also passed a pair of sub-legislations – the **Digital Services Act** (DSA) and the **Digital Markets Act** (DMA). These take off from GDPR's overarching focus on the individual's right over her data.
            a. DSA focuses on issues like hate speech, counterfeit goods etc.
            b. DMA has defined a new category of "**dominant gatekeeper**" platforms and is focused on non-competitive practices and abuse of dominance by these players.

    » **On AI**, though, the **US has made an attempt to take a lead** by way of the new **White House Executive Order on AI**, which is being offered as an elaborate template that could act as a blueprint for every other country looking to regulate AI. In Oct 2022, USA released a blueprint on an **AI Bill of Rights** – seen as a building block for the subsequent executive order.

A) AI REGULATION EFFORTS IN INDIA

- GoI plans to bring a Digital India Act to regulate AI.

- NITI Aayog has already released National Strategy on Artificial Intelligence which focuses on Responsible AI for all.