



## ChatGPT and Associated Ethical Concerns

**For Prelims:** Artificial Intelligence, ChatGPT, Reinforcement Learning from Human Feedback (RLHF), Phishing Email, Plagiarism.

**For Mains:** Significance of ChatGPT as per and ethical concerns associated with it.

### Why in News?

Recently, **OpenAI** (AI research and deployment company) opened its most recent and powerful **AI chatbot, ChatGPT**, to users to test its capability.

### What is ChatGPT?

- **About:**
  - **ChatGPT** is a variant of **GPT (Generative Pre-trained Transformer)** which is a **large-scale neural network-based language model developed by OpenAI**.
  - **GPT** models are trained on vast amounts of **text data** to generate **human-like text**.
  - It can generate responses to a wide range of topics, such as **answering questions, providing explanations, and engaging in conversations**.
  - In addition to being able to **"admit its mistakes, challenge false premises, and refuse unsuitable requests,"** the **ChatGPT** can also **"answer follow-up questions."**
  - The chatbot was also trained using **Reinforcement Learning from Human Feedback (RLHF)**.
- **Limitations:**
  - The chatbot displayed clear **racial and sexist biases**, which remains a problem with almost all AI models.
  - The chatbot gives answers which are grammatically correct and read well- though some have pointed out that these **lack context and substance**, which is largely true.
  - **ChatGPT** occasionally produces **inaccurate information** and that **its knowledge is restricted to global events that occurred before 2021**.

### What are the Ethical Concerns Associated with It?

- **Malicious and Dangerous Coding:** Some users have been experimenting with the **chatbot's potential** to carry out **malicious actions**. It has been claimed by several users that **malicious and dangerous** coding is generated by the **Chatbot** despite their **claims to be amateurs**.
- **Generation of Phishing Email:** **ChatGPT** is set up to reject requests to write **phishing emails** or **malicious code** but in actual sense **ChatGPT** is producing an outstanding **phishing email**.
- **Creation of Biased Data/information:** One concern is the **potential for bias in the generated code**, as the training data used to create the **code generator** may contain biases that are reflected in the generated code.
- **Curtailing the Job Opportunity:** There is a concern that the use of code generators could lead to the **loss of jobs for human programmers**.
- **Plagiarism Chokepoint:** Plagiarism in academics is not a **new issue**, but **ChatGPT** has altered

how **AI** is utilized to produce **original writing**. It is difficult to identify **plagiarized** information as a result. Teachers and academicians have also expressed concerns over ChatGPT's impact on written assignments.

## Way forward

- **Improving its ability to understand and respond to context**, such as being able to maintain a conversation on a specific topic or understanding the tone and intent of a message.
- **Increasing its knowledge base and ability** to generate more accurate and informative responses.
- **Developing more advanced language generation capabilities**, such as being able to write coherent and persuasive essays or generate creative and engaging fiction.
- Improving its ability to understand and respond to **different languages and dialects**.
- Developing more advanced interactive capabilities such as **handling multiple tasks and engaging in more complex conversations**.

## UPSC Civil Services Examination, Previous Year Question (PYQ)

**Q1. 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)**

**Source: TH**

PDF Reference URL: <https://www.drishtias.com/printpdf/chatgpt-and-associated-ethical-concerns>