AI वेब क्रॉलर पर प्रतिबंध

[स्रोत: इंडियन एक्सप्रेस]

चर्चा में क्यों?

एक ऐतिहासिक कदम के तहत, प्रमुख अमेरिकी और ब्रिटिश प्रकाशकों ने अपनी सामग्री/विषयवस्तु के अनधिकृत उपयोग को रोकने के लिये **कृत्रिम बुद्धिमत्ता (AI) वेब क्रॉलरों** को प्रतिबंधित करना शुरू कर दिया है।

- इसने **भारत में सहमति-आधारित कॉपीराइट संरक्षण और न्यायसंगत राजस्व साझा** व्यवस्था की मांग को पुनः तेज़ कर दिया है, जिससे डिजिटल शासन, कॉपीराइट प्रवर्तन और नैतिक AI उपयोग से जुड़े प्रमुख प्रश्नों पर चिंताएँ उठी हैं।

AI वेब क्रॉलर क्या है?

- **परिचय:** AI वेब क्रॉलर एक प्रकार का **स्वचालित सॉफ्टवेयर या बॉट** होता है, जो विशेष रूप से कृत्रिम बुद्धिमत्ता मॉडलों, जैसे कि **लार्ज लैंग्वेज मॉडल (LLM)**, के प्रशिक्षण में सहायक सामग्री एकत्र करने या AI सहायक के लिये तात्कालिक सूचना प्राप्ति हेतु इंटरनेट से विषयवस्तु स्कैन और संग्रह करता है।
- **प्रकार:**
  - **मॉडल प्रशिक्षण क्रॉलर: जनरेटिव AI मॉडलों के प्रशिक्षण हेतु वेबसाइट डेटा एकत्र करता है।**
    - उदाहरण: GPTBot (OpenAI), Amazonbot (Amazon), GoogleOther (Google)।
  - **लाइव रिट्रीवल क्रॉलर:** ये बॉट वेबसाइटों से वास्तविक समय में डेटा प्राप्त करते हैं ताकि उपयोगकर्त्ता की पूछताछ के दौरान पूर्व-प्रशिक्षित मॉडलों को पूरक जानकारी प्रदान की जा सके, जिससे AI खोज उपकरणों में अद्यतन और संदर्भित उत्तर सुनिश्चित हो सकें।
    - इसे Bing, ChatGPT आदि जैसे AI प्लेटफॉर्म द्वारा अद्यतन जानकारी प्राप्त करने हेतु उपयोग किया जाता है।
- **चिंताएँ:**
  - **नियामक ढाँचे का अभाव:** वर्तमान में भारत में ऐसा कोई नियामक ढाँचा नहीं है जो AI कंपनियों द्वारा वेब सामग्री तक सुलभता और उसके उपयोग की निगरानी कर सके।
    - **इससे ऐसी स्थिति उत्पन्न हो गई है जहाँ बड़ी तकनीकी कंपनियाँ भारतीय विषयवस्तु का बिना किसी सहमति या निगरानी के मुक्त रूप से लाभ उठा रही हैं, जबकि छोटे प्रकाशकों के पास ऐसी पहुँच की निगरानी या उसे सीमित करने के लिये कोई प्रभावी साधन नहीं है।**
  - **कॉपीराइट प्रवर्तन:** समाचार लेखों, ब्लॉगों और शैक्षणिक विषयवस्तु का उपयोग लार्ज लैंग्वेज मॉडल (LLM) के प्रशिक्षण के लिये बिना अनुमति या पारिश्रमिक के किया जा रहा है।
    - भारत का **कॉपीराइट अधिनियम, 1957** कृत्रिम बुद्धिमत्ता (AI) से संबंधित विशेष मामलों, जैसे कि AI-जनित व्युत्पन्न (Derivative) कृतियों या प्रशिक्षण डेटा अधिकारों को संबोधित करने के लिये सक्षम नहीं है।
    - भारतीय संदर्भ में **'उचित उपयोग' और 'बिना लाइसेंस प्रशिक्षण'** के बीच कोई स्पष्ट व्याख्या उपलब्ध नहीं है।
      - भारत में ऐसा कोई डेटा संरक्षण कानून नहीं है जो गैर-व्यक्तिगत डेटा पर केंद्रित हो, जबकि AI प्रशिक्षण के लिये अधिकांशतः इसी प्रकार के डेटा पर LLM निर्भर रहते हैं।
  - **कृत्रिम बुद्धिमत्ता (AI) का नैतिक उपयोग:** AI डेवलपर्स प्रायः यह खुलासा नहीं करते कि वे किस डेटा का उपयोग करते हैं, जिससे मूल रचनाकारों को न तो श्रेय मिलता है और न ही कोई प्रतिफल।
    - इसके अतिरिक्त, बिना जाँचे-परखे या पुराने सामग्री पर AI को प्रशिक्षित करने से **पूर्वाग्रह उत्पन्न हो सकते हैं और यह गलत या हानिकारक परिणाम दे सकता है,** जिससे AI प्रणालियों में जनसामान्य का विश्वास कमज़ोर पड़ता है।
    - ये चुनौतियाँ इस तर्क पर बल देती हैं कि भारत में सहमति-आधारित और अधिकार-सम्मानित डिजिटल इकोसिस्टम की स्थापना अत्यंत आवश्यक है।
- **वैश्विक रूपरेखा और भारत की आगामी दिशा:** यूरोपीय संघ का **AI अधिनियम, 2024** अब कॉपीराइटयुक्त डेटा पर AI प्रशिक्षण से संबंधित मुद्दों को संबोधित करने लगा है।
  - अमेरिका में प्रकाशक या तो AI कंपनियों के साथ लाइसेंसिंग समझौते कर रहे हैं या फिर उनके खिलाफ कानूनी कार्रवाई कर रहे हैं।
  - भारत इन उदाहरणों का अध्ययन कर एक ऐसा भारतीय मॉडल विकसित कर सकता है जो नवाचार और रचनाकारों के अधिकारों के बीच संतुलन स्थापित करे।

- **इलेक्ट्रॉनिक्स और सूचना प्रौद्योगिकी मंत्रालय (MeitY)** तथा **सूचना एवं प्रसारण मंत्रालय (I&B)** को मिलकर '**अनधिकृत डेटा स्क्रैपिंग**' की कानूनी परिभाषा तय करनी चाहिये और रचनाकारों के अधिकारों की रक्षा हेतु एक **सहमति-आधारित AI लाइसेंसिंग ढाँचा** स्थापित करना चाहिये।
    - साथ ही, इन मंत्रालयों को **क्लाउडफ्लेयर** जैसे प्लेटफॉर्म्स के साथ मिलकर भारतीय प्रकाशकों को डिजिटल सामग्री की सुरक्षा के लिये **AI बॉट-ब्लॉकिंग टूल्स** उपलब्ध कराने चाहिये, ताकि तकनीकी सुरक्षा उपाय भी सुनिश्चित किये जा सकें।

??????? ?????? ???????:

प्रश्न: भारत की कॉपीराइट व्यवस्था के लिये आर्टिफिशियल इंटेलिजेंस (AI) वेब क्रॉलर्स द्वारा उत्पन्न चुनौतियों की जाँच कीजिये।

UPSC सिविल सेवा परीक्षा, विगत वर्ष के प्रश्न (PYQ)

?????

प्रश्न. वैश्वीकृत संसार में, बौद्धिक सम्पदा अधिकारों का महत्त्व हो जाता है और वे मुकद्दमेबाज़ी का एक स्रोत हो जाते हैं। कॉपीराइट, पेटेंट और व्यापार गुप्तियों के बीच मोटे तौर पर विभेदन कीजिये। (2014)

PDF Refernece URL: https://www.drishtiias.com/hindi/printpdf/blocking-of-ai-web-crawler