कृत्रिम बुद्धिमत्ता और नैतिकता

यह एडिटोरियल 01/08/2024 को 'द हिंदू' में प्रकाशित "AI needs cultural policies, not just regulation" लेख पर आधारित है। इसमें इस बात पर बल दिया गया है कि AI को उत्तरदायी तरीके से आगे बढ़ाने के लिये उच्च-गुणवत्तापूर्ण, नैतिक डेटा को बढ़ावा देने के साथ उनके विनियमन को संतुलित करना अत्यंत महत्त्वपूर्ण है। इसमें सांस्कृतिक धरोहर के डिजिटलीकरण एवं साझेदारी के प्रयास भी शामिल हैं, जहाँ सुनिश्चित किया जाए कि AI का विकास नवोन्मेषी होने के साथ ही सांस्कृतिक मूल्यों एवं नैतिकता के प्रति सम्मानपूर्ण हो। यह दृष्टिकोण पारदर्शिता और पहुँच को बढ़ाकर अधिक समावेशी एवं प्रभावी AI प्रणालियाँ सुनिश्चित कर सकेगा।

प्रिलिम्स के लिये:

कृत्रिम बुद्धिमत्ता (AI), मशीन लर्निंग (ML), डीप लर्निंग (DL), जनरेटिव एआई, विश्व आर्थिक मंच, स्वायत्त हथियार, यूरोपीय संघ, यूके एआई सुरक्षा शिखर सम्मेलन, सूचना प्रौद्योगिकी नियम 2021, लार्ज लैंग्वेज़ मॉडल (LLM), इंडियाएआई मिशन, युवाओं के लिये उत्तरदायी आर्टिफिशियल इंटेलिजेंस (AI), नीति आयोग।

मेन्स के लिये:

भारतीय समाज के लिये प्रौद्योगिकी का महत्त्व और उनसे जुड़ी नैतिक चिंताएँ।

कृत्रिम बुद्धिमत्ता (Artificial Intelligence- AI), जो मशीनों **द्वारा मानवीय बुद्धिमत्ता की नकल** करने की क्षमता है, अपने उन्नत **डेटा प्रोसेसिंग एवं पूर्वानुमान क्षमताओं** के माध्यम से उद्योगों को रूपांतरित करने और मानवीय क्षमताओं को बढ़ाने का वादा करती है। हालाँकि, जैसे-जैसे AI दैनिक जीवन का अभिन्न अंग बनता जा रहा है, इसके नैतिक निहितार्थों पर तत्काल ध्यान देने की आवश्यकता भी अनुभव की जा रही है। पूर्वाग्रहों को बनाए रखने, निजता का उल्लंघन करने और रोज़गार विस्थापन का कारण बनने की इस प्रौद्योगिकी की क्षमता महत्त्वपूर्ण चिंताएँ पैदा करती है। इसके अतिरिक्त **AI विकास की तेज़ गति प्रायः मौजूदा नियामक ढाँचों** से संबोधित नहीं हो पाती, जिससे जवाबदेही तथा ज़िम्मेदार उपयोग की समस्याएँ और जटिल हो जाती हैं।

AI के संबंध में जारी नैतिक बहस में, महत्त्वपूर्ण क्षेत्रों में स्वचालित निर्णयन से लेकर रचनात्मक क्षेत्रों में इसकी भूमिका तक, **विभिन्न अनुप्रयोग शामिल** हैं। इन चिंताओं को संबोधित करने के लिये नीति निर्माताओं, प्रौद्योगिकीविदों और **नैतिकतावादियों को संलग्न करते हुए एक सहयोगी दृष्टिकोण** अपनाने की आवश्यकता है। मज़बूत नैतिक दिशा-निर्देश विकसित करना, पारदर्शिता सुनिश्चित करना और निजता की सुरक्षा करना आवश्यक कदम होंगे। जैसे-जैसे AI विकसित होता जा रहा है, जोखिमों को कम करते हुए इसके लाभों को अधिकतम करने और इसके उपयोग को सामाजिक मूल्यों के साथ संरेखित करने के लिये इसकी नैतिक सीमा पर विचार करना महत्त्वपूर्ण हो गया है।

कृत्रिम बुद्धिमत्ता (AI) क्या है?

- **परिचय:**
  - AI (AI) से आशय **किसी कंप्यूटर या कंप्यूटर द्वारा नियंत्रित रोबोट** की उस क्षमता से है, जिसके तहत वे ऐसे कार्य कर सकते हैं, जिनके लिये आमतौर पर मानवीय बुद्धिमत्ता एवं निर्णय क्षमता की आवश्यकता होती है।
    - हालाँकि कोई भी **AI उन सभी कार्यों को कर सकने में सक्षम** नहीं है जिन्हें एक औसत मानव कर सकता है, फिर भी कुछ AI प्रणालियाँ कुछ विशिष्ट कार्यों को करने में उत्कृष्टता रखती हैं।
- **विशेषताएँ एवं घटक:**
  - AI की प्रमुख विशेषता इसकी **युक्तिसंगतता और ऐसी कार्रवाई करने की क्षमता** है जो किसी विशेष लक्ष्य की प्राप्ति की संभावना को अधिकतम करती है।
  - मशीन लर्निंग (ML) AI का एक उपसमूह है और डीप लर्निंग (DL) तकनीकें **असंरचित डेटा (जैसे टेक्स्ट, इमेज या वीडियो)** की बड़ी मात्रा को संसाधित कर स्वचालित अधिगम की सुविधा प्रदान करती हैं।

एथिकल AI (Ethical AI) क्या है?

- **परिचय:**

- ◦ **एथकिल AI, जसि नैतकि या ज़म्मिेदार AI (Moral or Responsible AI)** के रूप में भी जाना जाता है, AI प्रणालयिों के इस तरह के वकिास एवं उपयोग को संदर्भति करता है जो नैतकि सद्धिांतों, सामाजकि मूल्यों और मानव अधकिारों के साथ संरेखति हो।
  - ◦ यह **AI प्रौद्योगकिी के ज़म्मिेदार उपयोग पर बल देता** है ताकि सुनश्चिति कयिा जा सके कि यह व्यक्तयिों, समुदायों और समग्र रूप से समाज को लाभान्वति करे तथा संभावति हानयिों और पूर्वाग्रहों को न्यूनतम करे।
- एथकिल AI के प्रमुख पहलू:

सद्धिांत	**ववरिण**
पारदर्शति और व्याख्यात्मकता	AI प्रणालयिों को इस तरह से डज़िाइन एवं कार्यान्वति कयिा जाना चाहयि कि उसके संचालन तथा नर्णिय लेने की प्रक्रयिाएँ उपयोगकर्ताओं और हतिधारकों के लयि समझने एवं समझाने योग्य हों। इससे वश्विास तथा जवाबदेही को बढ़ावा मलिेगा।
नष्पिक्षता और पूर्वाग्रह शमन	एथकिल AI का उद्देश्य नस्ल, लगि, जातीयता या सामाजकि- आर्थकि स्थति जैसे कारकों के आधार पर कुछ व्यक्तयिों या समूहों के वरिुद्ध भेदभाव को रोकने के लयि पूर्वाग्रहों को कम करना और AI एल्गोरदिम एवं मॉडल में नष्पिक्षता सुनश्चिति करना है।
नजिता और डेटा संरक्षण	एथकिल AI व्यक्तयिों की नजिता के अधकिार को अक्षुण्ण रखता है और व्यक्तगित डेटा के सुरक्षति एवं उत्तरदायी प्रबंधन का समर्थन करता है; यह प्रासंगकि नजिता/गोपनीयता कानूनों और वनियिमों के साथ सहमत एवं अनुपालन सुनश्चिति करता है।
जवाबदेही और ज़म्मिेदारी	AI प्रणालयिाँ तैनात करने वाले डेवलपर्स और वभिन्नि संगठनों को अपनी AI प्रौद्योगकियिों के परणिामों के लयि जवाबदेह होना चाहयि। त्रुटयिों या हानकिारक प्रभावों को संबोधति करने तथा सुधार हेतु ज़म्मिेदारी सुनश्चिति कयिा जाना आवश्यक है।
सुदृढ़ता और वश्विसनीयता	AI प्रणाली सुदृढ़, वश्विसनीय और वभिन्नि स्थतियिों एवं परस्थितियिों में लगातार प्रदर्शन सक्षम होनी चाहयि। AI प्रणाली में हेर-फेर या उसे नष्ट करने जैसे प्रतकिूल प्रयासों से नपिटने हेतु उपाय कयि जाने चाहयि।
मानवता के लयि लाभ	AI का वकिास और उपयोग मानव कल्याण को बढ़ावा देने, सामाजकि चुनौतयिों को हल करने तथा समाज, अर्थव्यवस्था एवं पर्यावरण में सकारात्मक योगदान देने हेतु कयिा जाना चाहयि।

//

AI से जुड़ी नैतिक चिताएँ:

- **‘डीपफेक’ और गलत सूचना संबंधी चिता:** AI-जनित डीपफेक की बढ़ती परिष्कृतता गलत **सूचना और भ्रामक सूचना** (misinformation and disinformation) के प्रसार के लिये एक महत्त्वपूर्ण खतरा बन गई है।
    - उदाहरण के लिये, एक अभिनेत्री का चेहरा दिखाने वाले वायरल डीपफेक वीडियो के निर्माता ने पुलिस के सामने स्वीकार किया कि यह वीडियो इंस्टाग्राम चैनल पर फॉलोअर्स की संख्या बढ़ाने के लिये बनाया गया था। यह ऐसी तकनीकों के उपयोग में नैतिकता की कमी को दर्शाता है।
- **एल्गोरिदम संबंधी पूर्वाग्रह:** यदि AI प्रणालियों को पक्षपातपूर्ण डेटा पर प्रशिक्षित किया जाए तो वे मौजूदा सामाजिक पूर्वाग्रहों को कायम रख सकती हैं या बढ़ा सकती हैं, जिसके भेदभावपूर्ण परिणाम सामने आ सकते हैं।
    - उदाहरण के लिये, जब शोधकर्त्ताओं ने एक जनरेटिव AI मॉडल ‘**स्टेबल डिफ्यूजन**’ (Stable Diffusion) से एक गरीब व्यक्ति की छवि सृजित करने के लिये कहा तो उसने प्रायः अश्वेत लोगों की छवि दिखाई।
    - इसके अलावा, संयुक्त राष्ट्र शैक्षिक, वैज्ञानिक और सांस्कृतिक संगठन (UNESCO के एक अध्ययन में पाया गया कि LLMs (Large Language Models) में गंभीर लैंगिक पूर्वाग्रह, समलैंगिकता-विरोध और नस्लीय रूढ़िवादिता प्रदर्शित होती है।
        - उदाहरण के लिये, महिलाएँ असमान रूप से घरेलू भूमिकाओं और ‘घर’ एवं ‘परिवार’ जैसे शब्दों से संबद्ध की गई थीं, जबकि पुरुष प्रायः ‘व्यवसाय’ एवं ‘करियर’ जैसे शब्दों से संबद्ध थे।
- **प्राथमिक स्रोत प्रतिनिधित्व की चुनौतियाँ:** AI प्रणालियाँ प्रायः द्वितीयक स्रोतों पर निर्भर रहती हैं, जो मुख्यतः अंग्रेज़ी में होते हैं, और अभिलेखीय **दस्तावेजों एवं मौखिक परंपराओं** जैसे प्राथमिक स्रोतों को शामिल नहीं करतीं।
    - प्राथमिक स्रोतों की अनदेखी करने से कुछ समाजों और संस्कृतियों का कम प्रतिनिधित्व या गलत प्रतिनिधित्व हो सकता है **इस अनदेखी के कारण प्रायः AI** मॉडल को प्रशिक्षित करने के लिये इस्तेमाल किये जाने वाले डेटा में पूर्वाग्रहों का समावेश हो जाता है।
    - प्राथमिक साक्षरता स्रोतों तक पहुँच और उनका डिजिटलीकरण विविध संस्कृतियों एवं इतिहासों के बारे में AI की समझ को बढ़ा सकता है, लेकिन अभी भी इसका काफी हद तक दोहन नहीं हुआ है।
- **डेटा गोपनीयता/निजता:** AI विकास के लिये व्यक्तिगत डेटा का संग्रह एवं उपयोग निजता के **उल्लंघन और दुरुपयोग के बारे में चिताएँ पैदा** करता है। इसके साथ ही, निगरानी उद्देश्यों के लिये AI पर बढ़ती निर्भरता बड़े पैमाने पर निगरानी और नागरिक स्वतंत्रता के क्षरण का कारण बन सकती है।
    - उदाहरण के लिये, इंटरनेट से प्राप्त डेटा पर प्रशिक्षित जनरेटिव AI टूल्स व्यक्तियों के निजी विवरण को बनाए रख सकते हैं, जिसमें उनके

परिवार और मित्रों के बारे में जानकारी शामिल है, जिससे पहचान की चोरी या धोखाधड़ी की संभावना उत्पन्न हो सकती है।

- '**ब्लैक बॉक्स' समस्या:** कई AI मॉडल जटिल और समझने में कठिन होते हैं, जिससे उनकी निर्णय लेने की प्रक्रिया को समझाना चुनौतीपूर्ण हो जाता है। **पारदर्शिता की यह कमी** जवाबदेही में बाधा डाल सकती है।
  - उदाहरण के लिये, स्वचालित कारों को जटिल नैतिक दुविधाओं का सामना करना पड़ता है, जैसे कि यह तय करना कि दुर्घटना की स्थिति में किसे प्राथमिकता दी जाए।
- **उत्तरदायित्व का मुद्दा:** जब कोई **AI प्रणाली क्षति पहुँचाती** है तो इसका निर्धारण करना एक जटिल कानूनी एवं नैतिक चुनौती है कि ज़िम्मेदार किसे ठहराया जाए।
  - उदाहरण के लिये, एयर कनाडा को अपने एक चैटबॉट द्वारा ग्राहक को दी गई लापरवाहीपूर्ण गलत जानकारी के लिये उत्तरदायी ठहराया गया था, जो इस बात को उजागर करता है कि AI टूल्स अपनाते समय व्यवसायों को किन व्यापक जोखिमों पर विचार करना चाहिये।
- **स्वचालन और बेरोज़गारी:** AI द्वारा नौकरियों को स्वचालित करने की क्षमता से नौकरी या रोज़गार विस्थापन और आर्थिक असमानता के बारे में चिंताएँ पैदा होती हैं। **AI विकास की तेज़** गति से आर्थिक व्यवधान और उद्योगों एवं श्रमिकों के लिये चुनौतियाँ पैदा हो सकती हैं।
  - उदाहरण के लिये, [विश्व आर्थिक मंच (WEF)]() के अनुसार, **वर्ष 2025 तक AI के** कारण लगभग 85 मिलियन नौकरियाँ समाप्त हो सकती हैं। इस परिदृश्य में फिर आर्थिक असमानता बढ़ सकती है।
- **डेटा स्वामित्व:** व्यक्तियों द्वारा उत्पन्न डेटा का स्वामित्व एक जटिल कानूनी और नैतिक मुद्दा है। जैसे-जैसे AI सिस्टम उपयोगकर्ता-जनित सामग्री पर अधिक से अधिक निर्भर होते जा रहे हैं, सवाल खड़ा हो रहा है कि डेटा का स्वामित्व किसके पास है और इसका उपयोग कैसे किया जा सकता है। यह कॉपीराइट और बौद्धिक संपदा अधिकारों (IPR) के बारे में भी चिंताएँ पैदा करता है।
  - उदाहरण के लिये, **AI के उपयोग से कला (art) का सृजन कॉपीराइट स्वामित्व** और साहित्यिक चोरी (plagiarism) एवं कॉपीराइट उल्लंघन की संभावना के बारे में सवाल उठाता है।
- **स्वायत्त हथियार**: स्वायत्त हथियारों का विकास निर्णय प्रक्रियाओं लेने में मानवों की भूमिका और अनपेक्षित परिणामों की संभावना के बारे में सवाल उठाता है। स्वायत्त प्रणालियों द्वारा घातक बल का उपयोग जटिल नैतिक एवं सुरक्षा संबंधी दुविधाएँ पेश करता है।
- [**'डिजिटल डिवाइड':**]() AI प्रौद्योगिकी तक असमान पहुँच मौजूदा सामाजिक असमानताओं को बढ़ा सकती है।
  - उदाहरण के लिये, **भारत में कुल जनसंख्या के लगभग 52% तक इंटरनेट** की पहुँच है, ऐसे में AI के भेदभावपूर्ण उपयोग से डिजिटल डिवाइड और AI के लाभों में अंतराल और बढ़ सकता है।
- **पर्यावरणीय नैतिकता:** AI प्रौद्योगिकियों के विकास और परिनियोजन का पर्यावरणीय प्रभाव उत्पन्न होता , जिससे संवहनीयता और नैतिक ज़िम्मेदारी के बारे में सवाल खड़े होते हैं।
  - अपनी नवीनतम वार्षिक प**र्यावरण रिपोर्ट में गूगल (Google) ने वर्ष 2023 में डेटा** केंद्रों द्वारा बिजली के **उपयोग में 17% की वृद्धि दर्ज** की। यह प्रवृत्ति अभी जारी रहने की उम्मीद है क्योंकि AI टूल्स अधिक व्यापक रूप से तैनात और उपयोग किये जा रहे हैं।

AI से संबद्ध नैतिक चिंताओं को दूर करने के लिये कौन-से कदम उठाए गए हैं?

- **अंतर्राष्ट्रीय स्तर पर:**
  - **सामाजिक उद्यमिता के लिये वैश्विक गठबंधन (Global Alliance for Social Entrepreneurship):** दावोस में आयोजित विश्व आर्थिक मंच की वार्षिक बैठक (2024) में श्वाब फाउंडेशन (Schwab Foundation) के सामाजिक उद्यमिता के लिये वैश्विक गठबंधन द्वारा माइक्रोसॉफ्ट के सहयोग से सामाजिक नवाचार के लिये AI (AI for Social Innovation) पर एक नई पहल शुरू की गई।
    - प्रमुख प्रौद्योगिकी और पारिस्थितिकी तंत्र कंपनियों को शामिल करते हुए लॉन्च की गई इस पहल का उद्देश्य सकारात्मक सामाजिक प्रभाव के लिये AI के उपयोग को बढ़ावा देना, सफल अनुप्रयोगों को प्रदर्शित करना और ज़िम्मेदार कार्यान्वयन दिशानिर्देश विकसित करना है।
  - **यूरोपीय संघ का AI अधिनियम:** [यूरोपीय संघ (EU)]() **पहला व्यापक AI विनियमन** (EU AI Act) लेकर आया है, जिसका उद्देश्य AI प्रणालियों के जोखिमों को नियंत्रित करना और यूरोपीय संघ के नागरिकों के मूल अधिकारों की रक्षा करना है।
    - चीन, कनाडा और सिंगापुर जैसे देशों ने अपने स्वयं के AI विनियमन या दिशानिर्देश पेश किये हैं।
  - **कैलिफोर्निया का उदाहरण:** कैलिफोर्निया के विधि निर्माताओं ने एक विधेयक पेश किया है, जिसमें AI कंपनियों को अपने सिस्टम का परीक्षण करने और दुरुपयोग को रोकने के लिये (जैसे कि इलेक्ट्रिक ग्रिड पर हमले या रासायनिक हथियार निर्माण में सहायता करना) सुरक्षा उपायों को लागू करने की आवश्यकता होगी।
  - **दिग्गज टेक कंपनियों की भूमिका:** [माइक्रोसॉफ्ट, मेटा, गूगल, अमेज़न और ट्विटर उन]() कंपनियों में शामिल हैं जिन्होंने ज़िम्मेदार AI टीमों का गठन किया है, जो **कृत्रिम बुद्धिमत्ता** का उपयोग करने वाले उपभोक्ता उत्पादों की सुरक्षा पर सलाह देते हैं और नैतिक मानकों के साथ उनके संरेखण की देखरेख करते हैं तथा जवाबदेही को बढ़ावा देते हैं।
  - [**यूके AI सुरक्षा शिखर सम्मेलन:**]() वर्ष 2023 में आयोजित **यूके AI सुरक्षा शिखर सम्मेलन (UK AI Safety Summit**) में AI के सुरक्षा एवं संरक्षा पहलुओं पर ध्यान केंद्रित किया गया और अंतर्राष्ट्रीय सहयोग की आवश्यकता पर बल दिया गया।
- **राष्ट्रीय स्तर पर:**
  - **AI मॉडल पर सलाह:** [इलेक्ट्रॉनिकी एवं सूचना प्रौद्योगिकी मंत्रालय (MeiTY)]() ने **सूचना प्रौद्योगिकी नियम 2021 के तहत वर्ष 2024 में AI मॉडल** और डीपफेक पर एक सलाह जारी की।
  - [इंडियाAI मिशन](): **इंडियाAI मिशन (IndiaAI mission) का उद्देश्य रणनीतिक सार्वजनिक-निजी भागीदारी** के माध्यम से एक सुदृढ़ पारिस्थितिकी तंत्र का निर्माण कर AI नवाचार को बढ़ावा देना है।
    - यह कंप्यूटिंग पहुँच, **डेटा गुणवत्ता और स्वदेशी AI क्षमताओं को बढ़ाएगा,** शीर्ष प्रतिभाओं को आकर्षित करेगा, स्टार्टअप्स को समर्थन देगा तथा भारत के AI क्षेत्र में ज़िम्मेदार, समावेशी विकास के लिये नैतिक, प्रभावशाली AI को बढ़ावा देगा।
  - **युवाओं के लिये उत्तरदायी AI:** सरकार ने **'युवाओं के लिये उत्तरदायी कृत्रिम बुद्धिमत्ता' (Responsible Artificial Intelligence for Youth)** नामक कार्यक्रम शुरू किया है।
  - **AI पर राष्ट्रीय रणनीति:** वर्ष 2018 में [नीति आयोग]() ने कृत्रिम बुद्धिमत्ता पर राष्ट्रीय रणनीति (National Strategy on Artificial

Intelligence- NSAI) जारी की, जिसमें भारत के पाँच सार्वजनिक क्षेत्रों में सुरक्षित एवं समावेशी AI के अंगीकरण के लिये एक रूपरेखा तैयार की गई।

- इस रणनीति में भविष्य में AI विकास के लिये एक मानक के रूप में **'AI for All' के मंत्र को अपनाया गया और AI के ज़िम्मेदार** उपयोग को सुनिश्चित करने पर बल दिया गया।

AI से संबंधित अन्य पहलें

- वैश्विक INDIAai शिखर सम्मेलन (Global INDIAai Summit)
- कृत्रिम बुद्धिमत्ता पर वैश्विक भागीदारी (Global Partnership on Artificial Intelligence- GPAI) शिखर सम्मेलन
- कृत्रिम बुद्धिमत्ता सुरक्षा शिखर सम्मेलन (Artificial Intelligence Safety Summit) 2023
- कृत्रिम बुद्धिमत्ता मिशन (Artificial Intelligence Mission)
- **अमेरिका भारत कृत्रिम बुद्धिमत्ता पहल (US India Artificial Intelligence Initiative)**

आगे की राह :

AI से संबंद्ध नैतिक चुनौतियों का समाधान करने के लिये नीति निर्माताओं, प्रौद्योगिकीविदों, नैतिकतावादियों और नागरिक समाज को शामिल करते हुए एक बहु-विषयक दृष्टिकोण अपनाने की आवश्यकता है। प्रमुख कदम में शामिल होंगे:

- **नैतिक ढाँचे का विकास एवं कार्यान्वयन:** AI विकास एवं परिनियोजन को नियंत्रित करने के लिये राष्ट्रीय और अंतर्राष्ट्रीय स्तर पर व्यापक नैतिक दिशानिर्देश एवं विनियमों का निर्माण करना।
- **विविधता और समावेशिता को बढ़ावा:** सुनिश्चित किया जाए कि पूर्वाग्रहों को कम करने और समावेशी डिज़ाइन को बढ़ावा देने के लिये AI विकास टीमें विविधतापूर्ण हों। उदाहरण के लिये, प्राथमिक साहित्यिक स्रोतों तक पहुँच और उनका डिजिटलीकरण विविध संस्कृतियों और इतिहासों के बारे में AI की समझ को बढ़ा सकता है, लेकिन अभी भी इसका काफी हद तक दोहन नहीं किया गया है।
- **सांस्कृतिक विरासत का डिजिटलीकरण:** यह AI को समृद्ध और विविध डेटासेट प्रदान कर सकता है, जिससे इतिहास की हमारी समझ में बदलाव आएगा और सांस्कृतिक कलाकृतियों की सुरक्षा होगी। यह प्रयास डेटा तक पहुँच को लोकतांत्रिक बनाकर और वैश्विक नवाचार को बढ़ावा देकर छोटी कंपनियों तथा ओपन-सोर्स AI समुदाय को लाभ पहुँचा सकता है।
- **सर्वोत्तम अभ्यासों को अपनाना:** AI प्रणालियों में पारदर्शिता, निष्पक्षता और जवाबदेही के लिये सर्वोत्तम अभ्यासों का पालन किया जाए।
- **पारदर्शिता और व्याख्यात्मकता को बढ़ावा देना :** ऐसी AI प्रणालियाँ डिज़ाइन की जाएँ जो अपने निर्णयों एवं कार्यों के लिये स्पष्ट और समझने योग्य व्याख्या प्रदान करें।
- **एल्गोरिदम संबंधी ऑडिट लागू करना:** जवाबदेही बनाए रखने के लिये ऑडिट के माध्यम से निष्पक्षता और पूर्वाग्रह की जाँच के लिये नियमित रूप से AI प्रणालियों का आकलन किया जाए।
- **निजता और डेटा सुरक्षा को मज़बूत करना:** डेटा गोपनीयता के लिये मज़बूत उपाय अपनाएँ और व्यक्तिगत एवं संवेदनशील सूचना का सुरक्षित संचालन सुनिश्चित करें। इसके साथ ही, व्यक्तियों का डेटा एकत्र करने या उसका उपयोग करने से पहले उनकी स्पष्ट सहमति ली जाए।
- **AI नैतिकता शिक्षा और प्रशिक्षण में निवेश करना:** AI और कंप्यूटर विज्ञान पाठ्यक्रम में नैतिकता पाठ्यक्रमों को एकीकृत किया जाए तथा AI पेशेवरों के लिये निरंतर नैतिकता प्रशिक्षण प्रदान किया जाए।
- **जन जागरूकता बढ़ाना:** आम लोगों को AI प्रौद्योगिकियों, उनके लाभों, जोखिमों और नैतिक निहितार्थों के बारे में शिक्षित किया जाए।
- **जवाबदेही और निरीक्षण तंत्र स्थापित करना:** AI प्रणालियों की निगरानी और नैतिक मानकों के अनुपालन को सुनिश्चित करने के लिये नियामक निकायों का विकास किया जाए।
- **जवाबदेही उपायों को लागू करना:** AI निर्णयों और परिणामों के लिये ज़िम्मेदारी की स्पष्ट सीमाएँ परिभाषित की जाएँ और कानूनी एवं नियामक ढाँचे के माध्यम से उल्लंघनों को संबोधित किया जाए।

अभ्यास प्रश्न: सार्वजनिक क्षेत्र के संस्थानों में निर्णय लेने की प्रक्रियाओं में कृत्रिम बुद्धिमत्ता (AI) के उपयोग से जुड़ी नैतिक चुनौतियों की चर्चा कीजिये। पारदर्शिता और जवाबदेही सुनिश्चित करने के लिये इन चुनौतियों का किस प्रकार प्रभावी ढंग से प्रबंधन किया जा सकता है?

UPSC सिविल सेवा परीक्षा पिछले वर्ष प्रश्न

??????????:

Q. विकास की वर्तमान स्थिति में, कृत्रिम बुद्धिमत्ता (Artificial Intelligence), निम्नलिखित में से किस कार्य को प्रभावी रूप से कर सकती है? (2020)

1. औद्योगिक इकाइयों में विद्युत् की खपत कम करना
2. सार्थक लघु कहानियों और गीतों की रचना
3. रोगों का निदान
4. टेक्स्ट से स्पीच (Text-to-Speech) में परिवर्तन
5. विद्युत् ऊर्जा का बेतार संचरण

नीचे दिये गए कूट का प्रयोग करके सही उत्तर चुनिये:

(a) केवल 1, 2, 3 और 5
(b) केवल 1, 3 और 4
(c) केवल 2, 4 और 5
(d) 1, 2, 3, 4 और 5

उत्तर: (b)

?????:

Q. भारत के प्रमुख शहरों में IT उद्योगों के विकास से उत्पन्न होने वाले मुख्य सामाजिक-आर्थिक निहितार्थ क्या हैं? (2022)

Q. "चौथी औद्योगिक क्रांति (डिजिटल क्रांति) के प्रादुर्भाव ने ई-गवर्नेंस को सरकार का अविभाज्य अंग बनाने में पहल की है"। विवेचन कीजिये। (2020)

PDF Refernece URL: https://www.drishtiias.com/hindi/printpdf/navigating-the-ethical-frontier-of-artificial-intelligence