Shodhbodhalaya: An International Peer reviewed Multidisciplinary Journal

Volume 3 | Issue 4 | ISSN: 2584-1807 | (https://shodhbodh.com/)

कृत्रिम बुद्धिमत्ता और नैतिकता : एक दार्शनिक विश्लेषण

Artificial Intelligence and Ethics: A Philosophical Analysis

डॉ सच्चिदानंद मिश्र

अतिथि विद्वान, दर्शनशास्त्र

प्रधानमंत्री कॉलेज ऑफ़ एक्सीलेंस

शासकीय शहीद केदारनाथ महाविद्यालय

मऊगंज

सारांश (Abstract)

कृत्रिम बुद्धिमत्ता (Artificial Intelligence–AI) आधुनिक युग की सर्वाधिक क्रांतिकारी तकनीकी उपलब्धि है। यह तकनीक मानव जीवन के प्रत्येक क्षेत्र में गहरे परिवर्तन ला रही है चाहे वह चिकित्सा, शिक्षा, न्याय, अर्थव्यवस्था, सुरक्षा या कला का क्षेत्र हो। किन्तु इस तकनीकी क्रांति के साथ नैतिकता (Ethics) के अनेक जटिल प्रश्न भी उठ खड़े हुए हैं। प्रस्तुत शोध पत्र दार्शनिक दृष्टिकोण से कृत्रिम बुद्धिमत्ता और नैतिकता के अंतर्संबंधों का विश्लेषण करता है। इसमें यह विचार किया गया है कि क्या AI में स्वायत्त नैतिक निर्णय लेने की क्षमता है, मानव मूल्यों तथा यंत्र-आधारित तर्क के बीच कौन से द्वंद्व उत्पन्न होते हैं, और किस प्रकार एक न्यायसंगत एवं उत्तरदायी AI-नैतिकता की संरचना निर्मित की जा सकती है। शोध पत्र में पाश्चात्य एवं भारतीय दार्शनिक परम्पराओं के प्रकाश में AI नैतिकता के विविध आयामों — गोपनीयता, समानता, स्वायत्तता, जवाबदेही, और मानवीय गरिमा का विस्तृत विवेचन प्रस्तुत किया गया है। निष्कर्ष में यह प्रतिपादित किया गया है कि AI नैतिकता केवल तकनीकी समस्या नहीं, अपितु यह एक गहन दार्शनिक और सामाजिक चुनौती है जिसके समाधान के लिए व्यापक मानवीय संवाद, विधिक ढाँचे और दार्शनिक मार्गदर्शन अनिवार्य है।

मुख्य शब्द (Keywords)

कृत्रिम बुद्धिमत्ता, नैतिकता, दर्शनशास्त्र, मशीन लर्निंग, स्वायत्तता, गोपनीयता, जवाबदेही, मानवीय गरिमा, एल्गोरिदम पूर्वाग्रह, AI शासन।

प्रस्तावना (Introduction)

मानव सभ्यता का इतिहास निरन्तर तकनीकी प्रयोगों और नवाचारों का इतिहास रहा है। पहिये के आविष्कार से लेकर औद्योगिक क्रांति तक, मुद्रण यन्त्र से लेकर इंटरनेट तक प्रत्येक तकनीकी परिवर्तन ने मानव जीवन के सामाजिक, आर्थिक और नैतिक ताने-बाने को नए सिरे से परिभाषित किया है। किन्तु इक्कीसवीं शताब्दी में उभरी कृत्रिम बुद्धिमत्ता की तकनीक इन सभी परिवर्तनों से गुणात्मक रूप से भिन्न और अधिक गहरी चुनौती प्रस्तुत करती है।

कृत्रिम बुद्धिमत्ता वह विज्ञान और अभियांत्रिकी है जो मशीनों को मानवीय बुद्धि जैसे कार्य करने योग्य बनाती है जैसे तर्क करना, सीखना, समस्या सुलझाना, भाषा समझना और निर्णय लेना। वर्तमान में AI केवल एक प्रयोगशाला की कल्पना नहीं रही; यह चिकित्सा निदान, न्यायिक निर्णय, शैक्षिक मूल्यांकन, वित्तीय ऋण प्रदान, सैन्य संचालन और सामाजिक निगरानी तक में प्रत्यक्ष भूमिका निभा रही है। जब कोई तकनीक इस स्तर पर मानव जीवन को प्रभावित करने लगे, तो दर्शनशास्त्र का यह दायित्व बनता है कि वह उसके नैतिक आधारों, सीमाओं और निहितार्थों की गहन परीक्षा करे।

नैतिकता (Ethics) दर्शनशास्त्र की वह शाखा है जो यह विचार करती है कि क्या उचित है, क्या अनुचित है, मानव को किस प्रकार आचरण करना चाहिए और किस आधार पर। जब AI स्वयं निर्णय लेने लगती है निदान देती है, संदिग्धों की पहचान करती है, नौकरी के उम्मीदवारों को छाँटती है — तो यह प्रश्न उठता है कि इन निर्णयों का नैतिक उत्तरदायित्व किसका है? क्या मशीन नैतिकता की अवधारणाओं को समझ सकती है? क्या वह उन्हें अपने संचालन में आत्मसात कर सकती है? ये प्रश्न केवल तकनीकी नहीं, बल्कि गहन दार्शनिक प्रकृति के हैं और इनका उत्तर तलाशना आज के युग की एक महत्त्वपूर्ण बौद्धिक आवश्यकता है।

प्रस्तुत शोध पत्र इन्हीं दार्शनिक प्रश्नों की खोज में प्रयत्नशील है। इसमें AI और नैतिकता के संबंध को ऐतिहासिक, सैद्धान्तिक और व्यावहारिक — तीनों स्तरों पर विश्लेषित किया गया है। पाश्चात्य नैतिक दर्शन की परम्पराओं उपयोगितावाद, कर्तव्यवाद, सद्गुण नैतिकता — के साथ-साथ भारतीय दार्शनिक चिन्तन में धर्म, अहिंसा और लोकसंग्रह की अवधारणाओं के आलोक में AI नैतिकता के विभिन्न पक्षों का परीक्षण किया गया है।

शोध पत्र का उद्देश्य (Objectives)-

प्रस्तुत शोध पत्र के निम्नलिखित प्रमुख उद्देश्य हैं जो इसके वैचारिक और विश्लेषणात्मक ढाँचे को परिभाषित करते हैं। प्रथम उद्देश्य यह है कि कृत्रिम बुद्धिमत्ता की अवधारणा, उसके विकास के चरणों और उसके वर्तमान स्वरूप को दार्शनिक परिप्रेक्ष्य में स्पष्ट किया जाए, जिससे यह समझा जा सके कि AI क्या है और यह मानवीय बुद्धि से किस प्रकार भिन्न एवं समान है।

द्वितीय उद्देश्य यह है कि पाश्चात्य और भारतीय नैतिक दर्शन की मुख्य धाराओं का अध्ययन करते हुए यह परीक्षित किया जाए कि वे AI के संदर्भ में किस सीमा तक प्रासंगिक हैं और उनकी क्या सीमाएँ हैं। तृतीय उद्देश्य यह है कि AI से उत्पन्न होने वाली प्रमुख नैतिक चुनौतियों—जैसे पूर्वाग्रह, गोपनीयता का उल्लंघन, निर्णय में पारदर्शिता की कमी, रोजगार विस्थापन और सैन्य अनुप्रयोग—का व्यवस्थित विश्लेषण प्रस्तुत किया जाए।

चतुर्थ उद्देश्य यह है कि AI प्रणालियों में नैतिक मूल्यों को किस प्रकार अंतर्निहित किया जा सकता है, इसके संभव दृष्टिकोणों और सीमाओं का परीक्षण किया जाए। पंचम उद्देश्य यह है कि एक न्यायसंगत, समावेशी और मानवाधिकार-सम्मत AI-नैतिकता के लिए नीतिगत एवं दार्शनिक सुझाव प्रस्तुत किए जाएँ जो नीति-निर्माताओं, तकनीकविदों और दार्शनिकों के लिए उपयोगी हों।

महत्व (Significance)

प्रस्तुत शोध पत्र का महत्व बहुआयामी है। वैचारिक स्तर पर यह शोध पत्र उस दार्शनिक शून्य को भरने का प्रयास करता है जो तकनीकी विकास और नैतिक चिन्तन के बीच बन गया है। अधिकांश AI-सम्बंधी विमर्श तकनीकी या नीतिगत

स्तर पर होता है; दार्शनिक गहराई से यह विश्लेषण अपेक्षाकृत कम हुआ है। इस संदर्भ में यह शोध पत्र एक आवश्यक दार्शनिक संवाद की पहल करता है।

सामाजिक स्तर पर AI नैतिकता का प्रश्न अत्यन्त महत्त्वपूर्ण है क्योंकि AI प्रणालियाँ समाज के सर्वाधिक संवेदनशील क्षेत्रों—न्याय, शिक्षा, स्वास्थ्य और रोजगार—में निर्णय ले रही हैं। यदि इन प्रणालियों में नैतिक पूर्वाग्रह हो तो उनके परिणाम समाज में असमानता और अन्याय को और बढ़ा सकते हैं। अतः AI नैतिकता का विश्लेषण सामाजिक न्याय की दृष्टि से भी आवश्यक है।

नीतिगत स्तर पर यह शोध पत्र विधि-निर्माताओं, सरकारों और अन्तर्राष्ट्रीय संगठनों के लिए उपयोगी वैचारिक आधार प्रदान करता है। भारतीय संदर्भ में जहाँ AI नीति अभी विकास के चरण में है, वहाँ दार्शनिक आधारों पर आधारित नैतिक ढाँचे की तत्काल आवश्यकता है। शैक्षणिक स्तर पर यह शोध पत्र दर्शनशास्त्र, सूचना प्रौद्योगिकी और सामाजिक विज्ञान के विद्यार्थियों एवं शोधकर्ताओं के लिए एक व्यापक सन्दर्भ सामग्री प्रस्तुत करता है।

कृत्रिम बुद्धिमत्ता और नैतिकता: एक दार्शनिक विश्लेषण-

1. कृत्रिम बुद्धिमत्ता: अवधारणा और दार्शनिक परिप्रेक्ष्य-

कृत्रिम बुद्धिमत्ता शब्द सर्वप्रथम 1956 में जॉन मैकार्थी (John McCarthy) द्वारा प्रयुक्त किया गया था। दार्शनिक दृष्टिकोण से AI का सर्वाधिक महत्त्वपूर्ण प्रश्न यह है: क्या मशीन वास्तव में 'सोच' सकती है? एलन ट्यूरिंग (Alan Turing) ने 1950 में अपने प्रसिद्ध निबन्ध 'Computing Machinery and Intelligence' में यह प्रश्न उठाया था। उन्होंने 'ट्यूरिंग टेस्ट' का प्रस्ताव किया, जिसके अनुसार यदि कोई मशीन किसी मानव से इस प्रकार संवाद कर सके कि परीक्षक उसे मानव समझे, तो वह 'बुद्धिमान' कही जा सकती है।

दार्शनिक जॉन सर्ल (John Searle) ने इसके विरोध में 'चाइनीज़ रूम आर्गुमेंट' प्रस्तुत किया। उनका तर्क था कि एक मशीन वाक्यों का अर्थ नहीं समझती, केवल प्रतीकों का यांत्रिक संयोजन करती है। इसी को वे 'सिंटैक्स विदाउट सिमेन्टिक्स' कहते हैं। यह विवाद दर्शनशास्त्र में मन-शरीर समस्या (Mind-Body Problem) से गहरे जुड़ा है। यदि मन

केवल भौतिक प्रक्रियाओं का परिणाम है, तो सम्भवतः मशीन भी 'मन' से युक्त हो सकती है। किन्तु यदि चेतना (Consciousness) में कुछ ऐसा है जो भौतिक यांत्रिकता से परे है, तो मशीन कभी 'बुद्धिमान' नहीं हो सकती।

भारतीय दर्शन की दृष्टि से यह प्रश्न और भी जटिल हो जाता है। न्याय दर्शन के अनुसार 'ज्ञान' की उत्पत्ति के लिए 'जिज्ञासा' और 'प्रयोजन' का होना आवश्यक है। AI में इन दोनों का अभाव है। वेदान्त दर्शन के अनुसार 'चेतना' (Consciousness) किसी भी ज्ञान-व्यापार की पूर्वापेक्षा है और यह केवल ब्रह्म में है। अतः मशीन की 'बुद्धिमत्ता' एक व्यवहारिक अनुकरण मात्र है, वास्तविक बोध नहीं। इन दार्शनिक प्रश्नों का नैतिकता से गहरा संबंध है, क्योंकि यदि AI वास्तव में 'सोचती' नहीं है, तो उसके निर्णयों की नैतिक जवाबदेही सदैव उसके निर्माताओं और उपयोगकर्ताओं की रहेगी।

2. नैतिकता के दार्शनिक सिद्धान्त और AI का संदर्भ-

नैतिक दर्शन की तीन प्रमुख परम्पराएँ हैं जो AI नैतिकता के विमर्श में केन्द्रीय भूमिका निभाती हैं। प्रथम परम्परा उपयोगितावाद (Utilitarianism) है, जिसके प्रमुख प्रतिनिधि जेरेमी बेंथम और जॉन स्टुअर्ट मिल हैं। इस सिद्धान्त के अनुसार वह कार्य नैतिक है जो अधिकतम लोगों के लिए अधिकतम सुख का सृजन करे। AI की दृष्टि से यह सिद्धान्त आकर्षक लगता है क्योंकि एल्गोरिदम परिणामों की गणना कर सकते हैं। किन्तु यह दृष्टिकोण समस्याग्रस्त भी है यदि किसी व्यक्ति के अधिकारों का बलिदान करने से बहुसंख्यकों को लाभ हो, तो क्या वह नैतिक होगा? AI प्रणालियाँ यदि केवल उपयोगितावादी सिद्धान्त पर चलें, तो अल्पसंख्यकों और वंचितों के साथ अन्याय की संभावना बढ़ जाती है।

द्वितीय परम्परा कान्ट का कर्तव्यवाद (Deontology) है। इमेनुएल कान्ट के अनुसार नैतिकता परिणामों पर नहीं, कर्तव्य पर आधारित होनी चाहिए। उनका 'श्रेणीबद्ध आदेश' (Categorical Imperative) यह कहता है कि हमें उसी सिद्धान्त पर कार्य करना चाहिए जिसे हम सार्वभौमिक नियम बनाना चाहें। AI के संदर्भ में यह अत्यन्त प्रासंगिक है। AI निर्णयों में पारदर्शिता, व्यक्ति की स्वायत्तता का सम्मान और मानव की गरिमा की रक्षा कान्टीय नैतिकता की अनिवार्यताएँ बन जाती हैं। किन्तु जटिल वास्तविक स्थितियों में कर्तव्यों का टकराव भी होता है जैसे गोपनीयता बनाम सार्वजनिक सुरक्षा जिसे कर्तव्यवाद सहजता से हल नहीं कर पाता।

तृतीय परम्परा अरस्तू की सद्गुण नैतिकता (Virtue Ethics) है। अरस्तू के अनुसार नैतिकता का लक्ष्य 'यूडेमोनिया' (Eudaimonia) अर्थात् मानव विकास और सुफलन है। सद्गुण जैसे न्याय, साहस, विवेक और संयम नैतिक जीवन के आधार हैं। AI के संदर्भ में यह प्रश्न उठता है कि क्या एक मशीन में सद्गुण हो सकते हैं? क्या AI 'विवेक' या 'न्याय' का प्रदर्शन कर सकती है? यह दृष्टिकोण AI डिजाइन में उन मूल्यों को अंतर्निहित करने की आवश्यकता पर बल देता है जो मानव विकास और कल्याण को बढ़ावा दें। भारतीय परम्परा में धर्म, अहिंसा, सत्य और करुणा के मूल्य इसी सद्गुण नैतिकता के समकक्ष हैं और इन्हें AI डिजाइन में शामिल करने का विचार भारतीय संदर्भ में अत्यन्त प्रासंगिक है।

3. AI से उत्पन्न प्रमुख नैतिक चुनौतियाँ-

एल्गोरिदम पूर्वाग्रह (Algorithmic Bias) AI नैतिकता की सबसे गम्भीर चुनौतियों में से एक है। AI प्रणालियाँ ऐतिहासिक डेटा पर प्रशिक्षित होती हैं। यदि वह डेटा स्वयं पूर्वाग्रहग्रस्त हो जैसा कि जाति, लिंग, नस्ल और वर्ग के आधार पर ऐतिहासिक भेदभाव से दूषित डेटा तो AI उस पूर्वाग्रह को न केवल पुनः उत्पादित करती है, बल्कि उसे एक वैज्ञानिक वैधता भी प्रदान कर देती है। संयुक्त राज्य अमेरिका में COMPAS एल्गोरिदम का उदाहरण प्रसिद्ध है, जो न्यायिक जमानत निर्णयों में अश्वेत आरोपियों के विरुद्ध अधिक पूर्वाग्रह दिखाता था। भारत में भी जहाँ जाति और वर्ग आधारित असमानताएँ ऐतिहासिक रूप से गहरी हैं, AI प्रणालियों में इस प्रकार का पूर्वाग्रह सामाजिक न्याय के लिए गम्भीर खतरा है।

गोपनीयता और निगरानी (Privacy and Surveillance) का प्रश्न AI नैतिकता का दूसरा महत्त्वपूर्ण आयाम है। AI-आधारित चेहरे की पहचान तकनीक (Facial Recognition), स्थान-अनुशीलन (Location Tracking), और डिजिटल व्यवहार की निगरानी राज्य और कॉर्पोरेशन दोनों द्वारा तेजी से प्रयुक्त हो रही है। दार्शनिक दृष्टिकोण से गोपनीयता केवल कानूनी अधिकार नहीं, यह मानव स्वायत्तता, पहचान और गरिमा की रक्षा के लिए एक आवश्यक नैतिक शर्त है। जब AI प्रणालियाँ नागरिकों की निगरानी करती हैं, तो यह मिशेल फूको के 'पैनोपटिकन' (Panopticon) के विचार को एक नई तकनीकी वास्तविकता में रूपांतरित कर देती है, जहाँ व्यक्ति निरन्तर अदृश्य दृष्टि के अधीन रहता है।

जवाबदेही और पारदर्शिता (Accountability and Transparency) की समस्या भी अत्यन्त गम्भीर है। AI प्रणालियाँ, विशेषकर 'डीप लर्निंग' मॉडल, प्रायः 'ब्लैक बॉक्स' की तरह काम करती हैं अर्थात् उनके निर्णयों की प्रक्रिया मानव द्वारा

सुगमता से समझी और समझाई नहीं जा सकती। इसे 'एक्सप्लेनेबिलिटी की समस्या' कहते हैं। यदि एक AI चिकित्सक कैंसर का निदान करे और वह निदान गलत निकले, तो दोषी कौन होगा डेवलपर, अस्पताल, या AI निर्माता? यह नैतिक जवाबदेही का प्रश्न अत्यन्त जटिल है। कान्टीय दर्शन में नैतिक उत्तरदायित्व के लिए 'इच्छाशक्ति' (Will) की आवश्यकता है; चूँकि AI में इच्छाशक्ति नहीं होती, अतः नैतिक जवाबदेही सदैव मानव पर रहती है।

स्वायत्त शस्त्र प्रणालियाँ (Autonomous Weapon Systems) AI नैतिकता की सबसे तीव्र और भयावह चुनौती हैं। जब एक AI प्रणाली युद्ध में शत्रु की पहचान और उसे नष्ट करने का निर्णय स्वयं ले सकती हो, तो यह मानव नैतिकता की मूलभूत समझ को ही चुनौती देती है। 'जस्ट वॉर थ्योरी' (न्यायसंगत युद्ध का सिद्धान्त) के अनुसार युद्ध में नागरिकों की रक्षा, आनुपातिकता और विवेक के सिद्धान्त आवश्यक हैं—जिन्हें एक मशीन के लिए सुनिश्चित करना असाधारण रूप से कठिन है। गाँधी की अहिंसा दर्शन की दृष्टि से ऐसे शस्त्रों का विकास ही नैतिक रूप से अस्वीकार्य है।

रोजगार विस्थापन (Employment Displacement) और आर्थिक असमानता का प्रश्न AI नैतिकता का एक और महत्त्वपूर्ण आयाम है। AI-आधारित स्वचालन (Automation) लाखों श्रमिकों के रोजगार को प्रभावित कर रहा है। मार्क्सवादी दृष्टिकोण से यह प्रश्न उठता है कि इस तकनीकी अधिशेष का लाभ किसे मिलेगा—मुट्ठी भर तकनीकी पूँजीपतियों को, या समग्र समाज को? भारतीय संदर्भ में जहाँ करोड़ों लोग अर्ध-कुशल और अकुशल रोजगार पर निर्भर हैं, AI का तीव्र प्रसार एक गहरी सामाजिक और आर्थिक समस्या उत्पन्न कर सकता है जिसका समाधान नैतिक और नीतिगत स्तर पर किया जाना आवश्यक है।

4. AI में नैतिक मूल्यों को अंतर्निहित करने की चुनौती-

'मूल्य संरेखण' (Value Alignment) का प्रश्न AI नैतिकता की केन्द्रीय समस्या है। इसका अभिप्राय यह है कि AI प्रणालियों को मानव मूल्यों और उद्देश्यों के अनुरूप बनाना। किन्तु यह कार्य अत्यन्त जटिल है क्योंकि मानव मूल्य स्वयं सांस्कृतिक, ऐतिहासिक और परिस्थितिगत दृष्टि से विविध और परिवर्तनशील हैं। किस संस्कृति के मूल्यों को AI में अंतर्निहित किया जाए? क्या पश्चिमी उदारवादी मूल्य सार्वभौमिक हैं? भारतीय, अफ्रीकी या पूर्व-एशियाई सांस्कृतिक मूल्यों को कैसे शामिल किया जाए?

इस संदर्भ में 'AI संरेखण अनुसंधान' (AI Alignment Research) एक उभरता हुआ क्षेत्र है। स्टुअर्ट रसेल जैसे विशेषज्ञों का मत है कि AI को मानव की अनिश्चित प्राथमिकताओं के प्रति 'अनिश्चित' रहना चाहिए और मानव अनुमोदन पर निर्भर रहना चाहिए। किन्तु दार्शनिक दृष्टि से यह प्रश्न बना रहता है कि 'कौन सा मानव' किसकी प्राथमिकताएँ, किसके मूल्य? क्या AI डिजाइन में बहुलवादी और लोकतान्त्रिक प्रक्रियाएँ अपनाई जा सकती हैं? यह एक दार्शनिक और राजनीतिक प्रश्न है जिसे केवल तकनीकी उपायों से हल नहीं किया जा सकता।

भारतीय दर्शन में 'लोकसंग्रह' की अवधारणा—अर्थात् समस्त मानवता और सृष्टि के कल्याण की भावना — AI नैतिकता के लिए एक उपयोगी दार्शनिक आधार प्रस्तुत करती है। गाँधी जी का 'ट्रस्टीशिप' का सिद्धान्त भी इस दृष्टि से महत्त्वपूर्ण है कि तकनीक का स्वामित्व एक ट्रस्ट की भाँति है जिसका उपयोग समाज के कल्याण के लिए होना चाहिए। अम्बेडकर की सामाजिक न्याय की अवधारणा यह माँग करती है कि AI प्रणालियाँ हाशिये पर स्थित समूहों के अधिकारों की रक्षा करें और उनके विरुद्ध भेदभाव न करें।

5. AI नैतिकता का वैश्विक और भारतीय संदर्भ-

वैश्विक स्तर पर AI नैतिकता के लिए विभिन्न ढाँचे विकसित हो रहे हैं। यूरोपीय संघ का 'AI Act' (2024) विश्व का प्रथम व्यापक AI विनियमन है जो जोखिम-आधारित दृष्टिकोण अपनाता है। UNESCO ने 2021 में AI नैतिकता पर एक वैश्विक अनुशंसा जारी की है जो मानव अधिकारों, पर्यावरणीय स्थिरता, पारदर्शिता और जवाबदेही के सिद्धान्तों पर आधारित है। OECD के AI सिद्धान्त भी इसी दिशा में महत्त्वपूर्ण पहल हैं। किन्तु इन सभी ढाँचों में पश्चिमी उदारवादी दृष्टिकोण का प्रभुत्व स्पष्ट है और विकासशील देशों की विशिष्ट आवश्यकताओं और परिस्थितियों को पर्याप्त स्थान नहीं मिला है।

भारतीय संदर्भ में AI नैतिकता के प्रश्न विशेष महत्त्व रखते हैं। भारत में AI के विकास के साथ कई विशिष्ट नैतिक चुनौतियाँ हैं। प्रथम, भाषाई विविधता: भारत में सैकड़ों भाषाएँ हैं और अधिकांश AI प्रणालियाँ हिन्दी और अंग्रेजी के लिए अधिकतम प्रशिक्षित हैं, जो भाषाई अल्पसंख्यकों के साथ नैतिक असमानता उत्पन्न करती है। द्वितीय, जाति और वर्ग आधारित डेटा पूर्वाग्रह: ऐतिहासिक जाति भेदभाव से दूषित डेटा AI में एम्बेड हो जाता है। तृतीय, डिजिटल विभाजन: ग्रामीण और शहरी, धनी और निर्धन के बीच AI तक पहुँच की असमानता एक गम्भीर नैतिक प्रश्न है।

भारत सरकार ने 'IndiaAI' मिशन और 'Responsible AI for Youth' कार्यक्रम जैसी पहलें की हैं। किन्तु एक व्यापक विधिक और दार्शनिक ढाँचे की अभी भी आवश्यकता है जो भारतीय संवैधानिक मूल्यों समानता, स्वतंत्रता, सामाजिक न्याय, और धर्मनिरपेक्षता को AI नैतिकता के केन्द्र में रखे। भारतीय संविधान का अनुच्छेद 21 जो 'प्राण और दैहिक स्वाधीनता' का अधिकार प्रदान करता है, और अनुच्छेद 14 जो समानता का अधिकार देता है — ये संवैधानिक मूल्य AI नैतिकता के लिए एक ठोस भारतीय आधार प्रस्तुत करते हैं।

निष्कर्ष (Conclusion)-

प्रस्तुत शोध पत्र में किए गए दार्शनिक विश्लेषण से निम्नलिखित महत्त्वपूर्ण निष्कर्ष उभरकर आते हैं। प्रथमतः, कृत्रिम बुद्धिमत्ता एक शक्तिशाली तकनीकी उपकरण है, किन्तु दार्शनिक दृष्टि से यह मानव चेतना, स्वायत्तता और नैतिक क्षमता से सर्वथा भिन्न है। AI निर्णय ले सकती है, किन्तु वह नैतिक निर्णय नहीं ले सकती — अर्थात् वह परिणामों की गणना कर सकती है, किन्तु 'उचित' और 'अनुचित' का बोध उसमें नहीं है। अतः AI के सभी निर्णयों का नैतिक उत्तरदायित्व मानव पर ही रहता है।

द्वितीयतः, पाश्चात्य नैतिक दर्शन की तीनों प्रमुख परम्पराएँ उपयोगितावाद, कर्तव्यवाद और सद्गुण नैतिकता — AI के संदर्भ में आंशिक रूप से प्रासंगिक हैं, किन्तु कोई भी एकल दृष्टिकोण सम्पूर्ण नहीं है। AI नैतिकता के लिए एक बहुलवादी और समन्वित दार्शनिक दृष्टिकोण की आवश्यकता है जो विभिन्न परम्पराओं से उपयुक्त तत्त्वों का संग्रह करे। भारतीय दर्शन का धर्म, अहिंसा, सत्य, लोकसंग्रह और सामाजिक न्याय का समन्वित दृष्टिकोण इस संदर्भ में एक मूल्यवान योगदान दे सकता है।

तृतीयतः, एल्गोरिदम पूर्वाग्रह, गोपनीयता का अतिक्रमण, जवाबदेही का संकट, स्वायत्त शस्त्र और रोजगार विस्थापन ये सभी समस्याएँ तकनीकी होने के साथ-साथ गहन नैतिक और दार्शनिक समस्याएँ हैं। इनका समाधान केवल तकनीकी नवाचार से नहीं, बल्कि सामाजिक संवाद, नीतिगत हस्तक्षेप और दार्शनिक मार्गदर्शन के समन्वय से ही संभव है। चतुर्थतः, AI नैतिकता के लिए एक वैश्विक सहमति-निर्माण की आवश्यकता है जो विभिन्न संस्कृतियों, समुदायों और देशों की विविधता का सम्मान करे। केवल पश्चिमी मूल्यों को सार्वभौमिक नहीं माना जा सकता।

पंचमतः, भारतीय संदर्भ में AI नैतिकता के लिए एक ऐसे ढाँचे की आवश्यकता है जो संवैधानिक मूल्यों, सामाजिक न्याय की अनिवार्यताओं और भारतीय दार्शनिक परम्पराओं के प्रकाश में निर्मित हो। इस ढाँचे में हाशिये पर स्थित समूहों — दलित, आदिवासी, महिलाएँ, ग्रामीण निर्धन — के हितों और अधिकारों की रक्षा को केन्द्रीय स्थान मिलना चाहिए। अन्ततः यह कहा जा सकता है कि AI का भविष्य तकनीक नहीं, मानवीय नैतिकता तय करेगी — और इस नैतिकता को परिभाषित करने में दर्शनशास्त्र की भूमिका अनिवार्य और अपरिहार्य है।

सुझाव (Recommendations)-

प्रस्तुत शोध पत्र के विश्लेषण के आधार पर निम्नलिखित महत्त्वपूर्ण सुझाव प्रस्तुत किए जाते हैं। प्रथम सुझाव यह है कि भारत में एक व्यापक 'AI नैतिकता आयोग' की स्थापना की जाए जिसमें दार्शनिक, सामाजिक वैज्ञानिक, कानूनविद्, तकनीकविद्, नागरिक समाज प्रतिनिधि और विशेषकर हाशिये के समुदायों के प्रतिनिधि सम्मिलित हों। यह आयोग AI नैतिकता के मानक निर्धारित करे, उल्लंघनों की जाँच करे और नीति-निर्माण में मार्गदर्शन दे।

द्वितीय सुझाव यह है कि AI प्रणालियों में पारदर्शिता और व्याख्यात्मकता (Explainability) को कानूनी रूप से अनिवार्य बनाया जाए। नागरिकों को यह जानने का अधिकार होना चाहिए कि उनके विषय में कोई भी महत्त्वपूर्ण AI-आधारित निर्णय किस आधार पर लिया गया है और उसे चुनौती देने का अधिकार भी होना चाहिए। तृतीय सुझाव यह है कि AI डेटा में विविधता और समावेशिता सुनिश्चित करने के लिए कठोर मानक बनाए जाएँ। विशेषकर सरकारी AI प्रणालियों में जाति, धर्म, लिंग या भाषा के आधार पर भेदभावपूर्ण पूर्वाग्रह की जाँच अनिवार्य की जाए।

चतुर्थ सुझाव यह है कि उच्च शिक्षा के पाठ्यक्रमों में विशेषकर तकनीकी और प्रबन्धन शिक्षा में 'AI नैतिकता' को एक अनिवार्य विषय के रूप में सम्मिलित किया जाए। भावी AI विकासकर्ताओं को नैतिक संवेदनशीलता और दार्शनिक समझ से युक्त बनाना समय की आवश्यकता है। पंचम सुझाव यह है कि स्वायत्त शस्त्र प्रणालियों पर अन्तर्राष्ट्रीय स्तर पर एक बाध्यकारी संधि की पहल की जाए। भारत को इस दिशा में नेतृत्वकारी भूमिका निभानी चाहिए और गाँधीवादी अहिंसा के सिद्धान्त को वैश्विक AI शासन में प्रतिष्ठित करने का प्रयास करना चाहिए।

षष्ठ सुझाव यह है कि AI से प्रभावित रोजगार विस्थापन से निपटने के लिए एक व्यापक 'डिजिटल न्याय' नीति बनाई जाए। इसमें पुनःप्रशिक्षण (Re-skilling), सार्वभौमिक मूल आय (Universal Basic Income) पर विचार, और AI के

आर्थिक लाभों का न्यायपूर्ण वितरण शामिल होना चाहिए। सप्तम सुझाव यह है कि भारतीय भाषाओं और सांस्कृतिक परम्पराओं को AI प्रणालियों में समुचित प्रतिनिधित्व देने के लिए विशेष राष्ट्रीय कार्यक्रम चलाए जाएँ, जिससे डिजिटल युग में भाषाई और सांस्कृतिक विविधता का संरक्षण हो सके।

संदर्भ ग्रन्थ (References)-

1. अम्बेडकर, भीमराव रामजी (1936). जाति का विनाश. नई दिल्ली: सम्यक प्रकाशन.
2. अरस्तू (अनुवाद: 2010). निकोमेशियन एथिक्स. नई दिल्ली: राजकमल प्रकाशन.
3. गाँधी, मोहनदास करमचन्द (1927). मेरे सपनों का भारत. अहमदाबाद: नवजीवन प्रकाशन मंदिर.
4. प्रभु, आर.के. एवं राव, यू.आर. (सं.) (1960). महात्मा गाँधी के विचार. अहमदाबाद: नवजीवन प्रकाशन.
5. शर्मा, चन्द्रधर (2000). भारतीय दर्शन: एक आलोचनात्मक सर्वेक्षण. वाराणसी: चौखम्बा विद्याभवन.
6. उपाध्याय, बालदेव (1998). भारतीय दर्शन का इतिहास. वाराणसी: शारदा मन्दिर.
7. Bostrom, Nick (2014). Superintelligence: Paths, Dangers, Strategies. Oxford: Oxford University Press.
8. Floridi, Luciano (2019). The Logic of Information: A Theory of Philosophy as Conceptual Design. Oxford: Oxford University Press.
9. Kant, Immanuel (1785). Groundwork of the Metaphysics of Morals. Cambridge: Cambridge University Press (2012 edition).
10. Mill, John Stuart (1863). Utilitarianism. London: Parker, Son, and Bourn.
11. Russell, Stuart (2019). Human Compatible: Artificial Intelligence and the Problem of Control. New York: Viking Press.
12. Searle, John R. (1980). Minds, Brains, and Programs. Behavioral and Brain Sciences, 3(3), 417-424.
13. Turing, Alan M. (1950). Computing Machinery and Intelligence. Mind, 59(236), 433-460.
14. UNESCO (2021). Recommendation on the Ethics of Artificial Intelligence. Paris: UNESCO.
15. Vallor, Shannon (2016). Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting. New York: Oxford University Press.
16. Wachter, Sandra; Mittelstadt, Brent & Russell, Chris (2017). Counterfactual Explanations Without Opening the Black Box. Harvard Journal of Law & Technology, 31(2), 841-887.
17. Zuboff, Shoshana (2019). The Age of Surveillance Capitalism. New York: PublicAffairs.

18. European Parliament (2024). AI Act (Artificial Intelligence Act). Official Journal of the European Union. Regulation (EU) 2024/1689.
19. NITI Aayog (2021). Responsible AI for All: Adopting the Framework. New Delhi: Government of India.
20. Jobin, Anna; Ienca, Marcello & Vayena, Effy (2019). The Global Landscape of AI Ethics Guidelines. Nature Machine Intelligence, 1(9), 389-399.