Gyanvividha, Vol.-2; Issue-4 (Oct.-Dec.) 2025 : PP-374-378

ज्ञानविविधा

कला, मानविकी और सामाजिक विज्ञान की सहकर्मी-समीक्षित, मूल्यांकित, त्रैमासिक शोध पत्रिका

ISSN : 3048-4537(Online)
3049-2327(Print)
IIFS Impact Factor-2.25
Vol.-2; Issue-4 (Oct.-Dec.) 2025
Page No.-374-378
©2025 Gyanvividha
https://journal.gyanvividha.com
Author's :

Dr. Sumit kumar

sr. Assistant professor,
Department of philosophy, S.M. College, Bhagalpur, TMBU.

Corresponding Author :

Dr. Sumit kumar

sr. Assistant professor,
Department of philosophy, S.M. College, Bhagalpur, TMBU.

कृत्रिम बुद्धि और नैतिक उत्तरदायित्व : क्या मशीनें नैतिक कर्ता हो सकती हैं?

सारांश : कृत्रिम बुद्धि (AI) के बढ़ते प्रयोग ने यह दार्शनिक प्रश्न उत्पन्न किया है कि क्या मशीनें नैतिक कर्ता (मोरल एजेंट) बन सकती हैं। नैतिक कर्ता वह होता है जो अपने कर्मों के लिए नैतिक रूप से उत्तरदायी माना जाता है। परंपरागत रूप से केवल मानवों को ही नैतिक एजेंसी प्राप्त है, जबकि औज़ार एवं मशीनें नैतिक दृष्टि से तटस्थ मानी जाती हैं। यह शोध पत्र पश्चिमी नैतिक दर्शन और भारतीय दर्शन दोनों के दृष्टिकोण से इस जटिल प्रश्न का विश्लेषण करता है। पश्चिमी दर्शन में चेतना, स्वतंत्र इच्छा (फ़्री विल), स्वायत्तता एवं इरादा जैसी शर्तों पर ज़ोर दिया गया है, जिनके बिना मशीनों को नैतिक कर्तृत्व प्रदान करना कठिन दिखता है। दूसरी ओर, भारतीय दार्शनिक परंपरा – विशेषकर हिन्दू और बौद्ध दर्शन – में नैतिक उत्तरदायित्व की अवधारणा चेतना (चेतना/आत्मा) और धर्म के सिद्धांत से जुड़ी है। इस पत्र में प्राचीन भारतीय ग्रंथों से उद्धरण लेकर विश्लेषण किया गया है कि कैसे भारतीय दर्शन चेतना के अभाव में मशीनों को *जड़* मानता है, तथा क्यों नैतिक कर्म हेतु अंतःप्रेरणा और आत्मज्ञान आवश्यक माने गए हैं। साथ ही, वर्तमान AI प्रौद्योगिकी से जुड़े नैतिक दायित्व (जैसे स्वचालित वाहन या हथियारों द्वारा होने वाली हानि की ज़िम्मेदारी) पर भी चर्चा की गई है। निष्कर्षतः, आज की परिस्थिति में मशीनों को पूर्ण नैतिक कर्ता नहीं माना जा सकता, परन्तु दोनों ही दर्शन हमें यह समझ प्रदान करते हैं कि मशीनों को नैतिक दायरे में लाने के लिए किन शर्तों की आवश्यकता होगी। ये अंतर्दृष्टियाँ AI के नैतिक रूप से उत्तरदायी विकास और उपयोग के लिए मार्गदर्शन देती हैं।

मुख्य शब्द : कृत्रिम बुद्धिमत्ता, नैतिक उत्तरदायित्व, नैतिक कर्ता, नैतिक एजेंसी, भारतीय दर्शन, पश्चिमी दर्शन, चेतना, धर्म।

1. परिचय : आधुनिक समाज में कृत्रिम बुद्धिमत्ता (AI) तीव्र गति से विकसित हो रही है। AI प्रणालियाँ स्वतः वाहन चलाने, चिकित्सीय निदान करने, वित्तीय निर्णय लेने तथा युद्धक ड्रोन संचालित करने जैसे कार्य करने लगी हैं। जैसे-जैसे AI अधिक स्वायत्त होती जा रही है, एक बुनियादी प्रश्न सामने आता है – यदि कोई स्वचालित प्रणाली कोई नैतिक रूप से गलत कार्य करती है (जैसे दुर्घटना में किसी को चोट पहुँचाती है), तो नैतिक ज़िम्मेदारी किसकी होगी? क्या उस "चतुर" मशीन की कोई नैतिक ज़िम्मेदारी बनती है, या समस्त उत्तरदायित्व इसके निर्माताओं और उपयोगकर्ताओं का है? दूसरे शब्दों में, **क्या मशीनें नैतिक कर्ता हो सकती हैं?**

पारंपरिक रूप से दार्शनिकों ने माना है कि नैतिकता केवल उन प्राणियों पर लागू होती है जो चेतनावान हैं और जिनमें सदिच्छापूर्ण इरादे व स्वतंत्र निर्णय करने की

क्षमता है। इंसान को एक नैतिक कर्ता माना जाता है क्योंकि आमतौर पर यह समझा जाता है कि मनुष्य अपने कर्म समझ-बूझकर और स्वेच्छा से करता है, तथा सही-गलत का भेद जानता है। इसके विपरीत, एक चाकू, बंदूक या कंप्यूटर प्रोग्राम को नैतिक कर्ता नहीं माना जाता – वे तो उपकरण हैं जो मानव के आदेश या डिज़ाइन पर कार्य करते हैं। यदि चाकू से किसी की हत्या होती है तो हम चाकू को दोषी नहीं मानते बल्कि उस व्यक्ति को दोषी मानते हैं जिसने उसे चलाया। इसी तरह, परंपरागत सोच में कंप्यूटर या रोबोट अपने आप में नैतिक निर्णय नहीं लेते, बल्कि मानव निर्देशों का पालन करते हैं।

फिर भी यह विषय उतना सरल नहीं है। उदाहरण के लिए, छोटे बच्चे या मानसिक रूप से असमर्थ व्यक्ति इंसान होने पर भी पूर्ण नैतिक उत्तरदायित्व नहीं निभाते, क्योंकि उनमें निर्णय की परिपक्व क्षमता का अभाव माना जाता है। वहीं कुछ कृत्रिम संस्थाओं को हम कानूनी या नैतिक दायित्व दे देते हैं – जैसे बड़ी कॉरपोरेशन को एक "कानूनी व्यक्ति" का दर्जा मिला है, जो अपने कार्यों के लिए कानूनी रूप से उत्तरदायी हो सकती है। इससे पता चलता है कि नैतिक कर्तृत्व का दर्जा प्राकृतिक या कृत्रिम होने पर निर्भर नहीं, बल्कि और कारकों पर निर्भर है (जॉनसन, 2006)। इसी के साथ, **मशीन नैतिकता** एक उभरता शोध क्षेत्र बन गया है, जो यह अध्ययन करता है कि मशीनों को नैतिक निर्णय लेने योग्य कैसे बनाया जाए और उनकी स्वायत्त क्रियाओं को कैसे नैतिक सीमाओं के भीतर रखा जाए (मूर, 2006)।

इस शोध पत्र में हम पहले नैतिक कर्तृत्व की अवधारणा को स्पष्ट करेंगे और जानेंगे कि विभिन्न नैतिक सिद्धांत (पश्चिमी और भारतीय) किन शर्तों पर किसी इकाई को नैतिक एजेंट मानते हैं। इसके बाद हम पश्चिमी दर्शन की दृष्टि से विश्लेषण करेंगे कि मौजूदा AI इन शर्तों पर कितना खरा उतरती है, तथा किन बिंदुओं पर मशीनें नैतिक कर्ता बनने से वंचित हैं। फिर हम भारतीय दर्शन – विशेषकर हिंदू और बौद्ध दृष्टिकोण – से विचार करेंगे कि आत्मा, चेतना और कर्म के सिद्धांत AI पर कैसे लागू होते हैं। भारतीय शास्त्रों के मूल उद्धरणों द्वारा हम देखेंगे कि चेतना के अभाव में किसी भी "यंत्र" (मशीन) के कार्य को वहां नैतिक कर्म का दर्जा नहीं मिलता। साथ ही आधुनिक प्रसंग में यदि भविष्य में AI वास्तव में मानवीय-सदृश्य चेतना या नैतिक बुद्धि प्राप्त करती है, तो भारतीय दर्शन उसे किस रूप में देखेगा। अंत में, हम वर्तमान व्यावहारिक संदर्भ (जैसे स्वचालित प्रणालियों के गलत निर्णयों की ज़िम्मेदारी) पर चर्चा करके निष्कर्ष प्रस्तुत करेंगे।

2. नैतिक कर्तृत्व: अर्थ एवं आवश्यकताएँ : दार्शनिक दृष्टि से नैतिक कर्ता वह होता है जिसकी क्रियाएँ नैतिक मूल्यांकन (भला या बुरा) के योग्य होती हैं, और जिसे उसके कर्मों के लिए प्रशंसा या दोष दिया जा सकता है। नैतिक कर्तृत्व के लिए आम तौर पर निम्नलिखित शर्तें आवश्यक मानी जाती हैं:

1. **स्वायत्तता:** नैतिक निर्णय लेने वाली इकाई को कुछ हद तक स्वायत्त होना चाहिए, अर्थात् उसके पास स्वतंत्र इच्छाशक्ति या चुनाव की क्षमता हो। यदि किसी पर पूरी तरह बाहर से नियंत्रण हो या वह जबरदस्ती करवाई गई क्रिया कर रहा हो, तो नैतिक दायित्व उस पर नहीं डाला जा सकता। मानव संदर्भ में, अगर कोई व्यक्ति बंदूक की नोक पर कोई काम करने को मजबूर हो, तो उसे नैतिक दोष कम दिया जाता है। मशीनों के संदर्भ में, प्रश्न उठता है कि क्या वे स्वायत्त हैं या केवल पूर्वनिर्धारित प्रोग्रामिंग का पालन कर रही हैं (जॉनसन, 2006)।
2. **इरादतन क्रिया:** नैतिक कर्ता अपनी क्रिया किसी इरादे (उद्देश्यपूर्ण मानसिक स्थिति) के साथ करता है। आकस्मिक या अनैच्छिक क्रियाएँ (जैसे छींकना या रीफ़लेक्स) नैतिक मूल्यांकन से बाहर रखी जाती हैं, क्योंकि उनमें जानबूझकर अच्छाई या बुराई करने का इरादा नहीं होता। दार्शनिकों ने तर्क दिया है कि एक मशीन की क्रिया को यदि नैतिक रूप से परखा जाना है तो उसमें किसी प्रकार का आंतरिक इरादा होना चाहिए (नादेउ, 2006)। यह सिद्ध करना अत्यंत कठिन है कि AI प्रणाली के पास वास्तव में स्वयं का इरादा है – अक्सर जो दिखता है वह या तो प्रोग्रामर का निहित इरादा होता है या उपयोगकर्ता का आदेश, न कि मशीन का अपना स्वतंत्र संकल्प (जॉनसन, 2006)।
3. **ज्ञान एवं समझ:** नैतिक निर्णय के लिए उस कार्य से जुड़े नैतिक सिद्धांतों या परिणामों की समझ होना ज़रूरी है। एक नैतिक एजेंट को यह बोध होना चाहिए कि किस कार्य से हानि होगी, किससे लाभ, या सही-गलत के मानदंड क्या हैं। मनुष्यों में यह नैतिक समझ नैतिक शिक्षा, सहानुभूति, सामाजिक अनुभव आदि से विकसित होती है। सवाल उठता है कि क्या AI को नैतिक सिद्धांतों की वास्तविक समझ हो सकती है, या क्या वह सिर्फ़ निर्धारित नियमों का पालन कर रहा होगा बिना किसी नैतिक संवेदना के (सुलिन्स, 2011)।
4. **जवाबदेही:** नैतिक कर्ता को अपने कर्मों की जवाबदेही स्वीकार करने योग्य होना चाहिए। यदि कोई इकाई ऐसी है जिसे दंड या पुरस्कार की अवधारणा से कोई फर्क ही नहीं पड़ता, या जिसे सुधारना अथवा पश्चाताप कराना संभव नहीं, तो व्यवहारतः उसे नैतिक समुदाय का सदस्य मानना कठिन है। उदाहरण के लिए, हम जानवरों या छोटे बच्चों को नैतिक अपराध हेतु दंडित नहीं करते, क्योंकि वे जवाबदेही की आवश्यक समझ नहीं रखते। इसी प्रकार, यदि एक रोबोट गलती करता है, तो उसे समझ नहीं आएगा कि सज़ा क्यों दी गई – न ही वह अपने "कर्मों" पर लज्जा या अपराधबोध महसूस करेगा। इसलिए उसकी नैतिक जिम्मेदारी निर्धारित करना दुविधापूर्ण है (मथायस, 2004)।

इन बुनियादी शर्तों के अतिरिक्त भी दार्शनिक विमर्श में अन्य पहलू आते हैं, जैसे *मुक्त इच्छा*– क्या निर्णय वास्तव में स्वतंत्र था या पूर्व-नियोजित कारणों द्वारा निर्धारित? अधिकांश पश्चिमी विचारकों का मत रहा है कि नैतिक जिम्मेदारी के लिए कुछ स्तर की मुक्त इच्छा

आवश्यक है। यदि सभी मानवीय क्रियाएँ भी प्राकृतिक या जैविक नियत कारणों का परिणाम मात्र हों, तो कठोर निर्धारणवाद (डिटर्मिनिज़्म) की स्थिति में नैतिक जिम्मेदारी का प्रश्न जटिल हो जाता है। इसी तर्क का अनुप्रयोग AI पर करने से प्रश्न उठता है – मशीन तो मानव-निर्मित प्रोग्राम और डाटा के अनुसार कार्य करती है, तो क्या वह पूर्णतः निर्धारित या सांख्यिकीय यादृच्छ प्रक्रियाओं पर नहीं चलती? अगर हाँ, तो क्या उसमें स्वतंत्र नैतिक निर्णय की कोई जगह है? कुछ विचारकों जैसे ब्रिंग्ज़यॉर्ड (ब्रिंग्ज़यॉर्ड, 2007) का तर्क है कि आज की मशीनों में तथाकथित "अनियमितता" (जैसे मशीन लर्निंग के यादृच्छ आउटपुट) भी असल में वास्तविक स्वतंत्र निर्णय नहीं, बल्कि एक तरह की पूर्वनिर्धारित या यादृच्छिक प्रक्रिया है। परंतु यही आपत्ति मानवों पर भी लगाई जा सकती है – मानव भी अपने जैविक, सामाजिक और आनुवंशिक कारकों से प्रभावित होकर निर्णय लेते हैं, तो पूरी तरह कौन स्वायत्त है (नादेउ, 2006)? इस बिंदु पर दार्शनिकों में मतभेद है कि किस हद तक मुक्त इच्छा की मांग की जाए।

3. मशीनें और नैतिक कर्तव्य: पश्चिमी दर्शन का परिप्रेक्ष्य : पश्चिमी नैतिक दर्शन में व्यक्तित्व, चेतना और तर्कसंगतता को नैतिक एजेंसी का केंद्रीय आधार माना गया है। इमैनुएल कांट के अनुसार नैतिकता का आधार **कर्तव्य** है और नैतिक एजेंट वह है जो अपने तर्कसंगत विवेक से स्वयं पर नैतिक कानून लागू करता है। कांटियन दृष्टि में नैतिक अधिकार-उत्तरदायित्व उसी को मिलते हैं जो **स्वायत्त तर्कसंगत प्राणी** हो। जो केवल कर्तव्यभाव से नैतिक नियमों का सम्मान कर सके, न कि बाहरी पुरस्कार/दंड के कारण। स्पष्ट है कि वर्तमान AI प्रणालियाँ इस कसौटी पर खरी नहीं उतरतीं, क्योंकि उनमें न स्वयं-निर्धारित इच्छा है, न "कर्तव्य" के प्रति कोई चेतन निष्ठा। कांट के नज़रिये से एक अध्ययन ने भी निष्कर्ष निकाला कि AI वास्तविक अर्थ में "नैतिक एजेंट" नहीं हैं, भले ही वे व्यवहार में नैतिकता की नकल कर लें (सानवो ओलु, 2025)। उदाहरणतः, AI ईमानदारी सिखा सकता है और ईमानदार व्यवहार दिखा सकता है, पर यह उसके नैतिक चयन से नहीं, बल्कि सिखाए गए पैटर्न से होगा (सानवो ओलु, 2025)।

उपयोगितावादी दृष्टिकोण में नैतिकता का मानदंड परिणाम है। अधिकतम सुख/हित। इस नजरिये से AI "नैतिक निर्णय" ले सकता है यदि वह आँकड़ों के आधार पर कम हानि/अधिक भलाई का विकल्प चुन सके। पर यहाँ समस्या AI की **आंतरिक नैतिकता** नहीं, उसकी **उपयोगिता** भर रह जाती है। स्वचालित कार के दुविधा-निर्णय में AI गणना कर सकता है, लेकिन वह नैतिक दर्द, पश्चाताप या सही-गलत का निजी बोध अनुभव नहीं करता। इसलिए कई विचारक AI को नैतिक उपकरण/"नैतिक इकाई" मानते हैं, पर स्वतंत्र नैतिक कर्ता नहीं (जॉनसन, 2006)। डेबोरा जी. जॉनसन के अनुसार कंप्यूटर सिस्टम्स के कार्यों के नैतिक परिणाम होते हैं, किंतु नैतिक मंशा और अंतिम ज़िम्मेदारी मनुष्य की रहती है (जॉनसन, 2006)।

इसके विपरीत कुछ विद्वान मानते हैं कि अत्यंत स्वायत्त और बुद्धिमान AI/एंड्रॉयड मानव से बेहतर नैतिक एजेंट हो सकते हैं, क्योंकि उनमें भावनात्मक पक्षपात जैसी मानवीय कमजोरियाँ नहीं होंगी। नादेउ ने उग्र तर्क दिया कि यदि शुद्ध स्वतंत्रता और तर्क ही पूर्ण नैतिकता का आधार है, तो आदर्श रोबोट ही "प्रथम वास्तविक नैतिक प्राणी" बन सकता है (नादेउ, 2006)। पर आलोचक कहते हैं कि नैतिकता में सहानुभूति, संवेदना और अनुभव का योगदान भी है।

यहाँ "दिखाई देने वाला व्यवहार" बनाम "वास्तविक समझ" का प्रश्न भी निर्णायक है। जॉन सियरल के "चीनी कमरे" तर्क के अनुसार कंप्यूटर नियम लागू कर सकता है, पर जरूरी नहीं कि उसे अर्थ की समझ हो (सियरल, 1980)। इसी तरह AI का नैतिक व्यवहार भीतर से "यांत्रिक" भी हो सकता है; ब्रायसन ने इसे "ज़ोम्बी नैतिकता" कहा अभिनय नैतिक, पर भावना-शून्य (ब्रायसन, 2010)।

व्यावहारिक स्तर पर स्वायत्त प्रणालियों की गलतियों से "ज़िम्मेदारी का अंतराल" उभरता है (मथायस, 2004): दुर्घटना में दोष निर्माता, प्रोग्रामर, मालिक या AI किसे? इसी संदर्भ में कभी यूरोपीय संघ में "इलेक्ट्रॉनिक व्यक्ति" जैसे कानूनी दर्जे पर विचार हुआ, पर विरोध हुआ क्योंकि इससे मानव निर्माता ज़िम्मेदारी से बच सकते हैं।

भविष्य की सुपर-बुद्धिमान AI पर निक बोस्ट्रोम की चेतावनी (बोस्ट्रोम, 2014) और रोबोटों में नैतिक निर्णय-क्षमता पर वैलाच व एलन की चर्चा (वैलाच और एलन, 2009) दिखाती है कि प्रश्न केवल एजेंसी का नहीं, **मूल्य-संरेखण** और शासन का भी है। गंकेल ने "मशीनों के प्रति नैतिक दायित्व" और "मशीनों की नैतिक ज़िम्मेदारी" दोनों को नए सिरे से सोचने की जरूरत बताई है (गंकेल, 2012)। कुल मिलाकर, मुख्यधारा पश्चिमी दृष्टि अभी मशीनों को पूर्ण नैतिक कर्ता मानने से हिचकती है, पर विमर्श तीव्र गति से विकसित हो रहा है।

4. भारतीय दर्शन में नैतिक एजेंसी और मशीन का प्रश्न : भारतीय दर्शन हिंदू, बौद्ध, जैन आदि विविध परंपराओं सहित नैतिकता को मानव जीवन के आध्यात्मिक तथा व्यावहारिक पक्षों से जोड़कर देखता है। "नैतिक कर्तव्य" को समझने हेतु यहाँ कुछ केंद्रीय अवधारणाएँ हैं: धर्म, कर्म, आत्मा (या बुद्धि), और चेतना। मशीनों के नैतिक कर्ता होने की संभावना पर विचार करते समय भारतीय दार्शनिक परिप्रेक्ष्य में तीन प्रश्न उभरते हैं: (1) क्या मशीन में आत्मा या चेतना होती है? (2) क्या मशीन अपना कर्म स्वयं निर्धारित कर सकती है? (3) नैतिक निर्णय में "धर्म" की भूमिका क्या होगी और क्या मशीन उस धर्म का पालन कर सकती है? इन प्रश्नों की पृष्ठभूमि में भारतीय चिंतन का पहला बल "चेतन" और "जड़" के भेद पर है।

अद्वैत वेदांत से लेकर सांख्य दर्शन तक आत्मा को **चैतन्यस्वरूप** कहा गया है। वह शुद्ध चेतना है, स्वयं प्रकाशमान है और मन-बुद्धि को चेतना प्रदान करती है। इसके विपरीत शरीर, इंद्रियाँ, मन, बुद्धि आदि प्रकृति के जड़ तत्व हैं; आत्मा के बिना इनमें अपनी चेतना नहीं।

भगवद्गीता में श्रीकृष्ण कहते हैं: "ईश्वरः सर्वभूतानां हृद्देशेऽर्जुन तिष्ठति, भ्रामयन् सर्वभूतानि यन्त्रारुढानि मायया" यहाँ शरीर को "यंत्र" कहा गया है, जिसमें आत्मा सवार है और प्रकृति/माया के गुण उसे संचालन में लगाते हैं। कठोपनिषद् की रथ उपमा इसी को स्पष्ट करती है: "आत्मानं रथिनं विद्धि, शरीरं रथमेव तु । बुद्धिं तु सारथिं विद्धि, मनः प्रग्रहमेव च ॥" (कठोपनिषद् 1.3.3)। अर्थात शरीर-रूपी रथ का स्वामी आत्मा है; बुद्धि सारथी, मन लगाम, और इंद्रियाँ घोड़े हैं। संकेत यह है कि स्वामी (आत्मा) के बिना रथ दिशाहीन है; और सारथी (बुद्धि) भी तभी सही मार्ग पर ले जा सकता है जब आत्मा का संकेत अंतरात्मा की प्रेरणा हो।

इसी बिंदु पर AI के संदर्भ में मूल निष्कर्ष उभरता है: जब तक किसी मशीन में आत्मज्ञानी चेतन सत्ता का निवास न हो, वह मात्र जड़ उपकरण है; भारतीय दृष्टि में केवल चैतन्ययुक्त प्राणी ही वास्तविक "कर्ता" कहलाने योग्य है। गीता कहती है: "प्रकृतेः क्रियमाणानि गुणैः कर्माणि सर्वशः। अहङ्कारविमूढ़ात्मा कर्ताहमिति मन्यते ॥" (गीता 3.27) प्रकृति के गुण ही कर्म कराते हैं; अहंकार से मोहित आत्मा "मैं कर्ता हूँ" मान बैठती है। शंकराचार्य की व्याख्या के अनुसार आत्मा निष्क्रिय साक्षी है और क्रियाएँ शरीर-मन-बुद्धि रूपी यंत्र में घटती हैं; जीवात्मा अहंकार से जुड़कर कर्तृत्व का दावा करता है। आधुनिक संदर्भ में AI जैसी मशीन जड़ तत्वों से बनी है; इसलिए वह अपने आप में कर्ता नहीं नैतिक फल उन चेतन प्राणियों से जुड़े होंगे जिन्होंने उसे बनाया/चलाया या जिनकी इच्छाओं से वह कार्य कर रही है (निर्माता, प्रोग्रामर, उपयोगकर्ता)।

बौद्ध दर्शन अनात्मवाद मानता है, फिर भी कर्म के केंद्र में **चेतना/इरादा** रखता है: "चेतनाहं भिक्खवे कम्मं वदामि" "मैं इरादे को ही कर्म कहता हूँ; इरादा करके ही जीव शरीर, वाणी व मन से कर्म करता है" (अंगुत्तरा निकया 6.63)। अतः यदि कोई स्वचालित मशीन किसी को हानि पहुँचाए, पर उसमें स्वयं का चेतन इरादा न हो, तो बौद्ध दृष्टि से वह "कर्म" नैतिक अर्थ में नहीं होगा हालाँकि परिणाम भौतिक रूप से गंभीर हो सकते हैं। इसलिए जब तक मशीन में अपना मन/चेतना नहीं, वह नैतिक स्तर पर निर्दोष उपकरण ही रहेगी और जिम्मेदारी मानव पक्षों में ही विभाजित होगी।

5. विमर्श: वर्तमान परिदृश्य और संभावित भविष्य : आज की वास्तविकता में AI को नैतिक कर्ता न सही, पर नैतिक समीकरण का एक महत्वपूर्ण घटक अवश्य माना जा रहा है। **आचरण बनाम उत्तरदायित्व** के स्तर पर हम AI को नैतिक आचरण सिखाने का प्रयास कर सकते हैं। जैसे स्वचालित कार को इस तरह प्रोग्राम करना कि दुर्घटना की स्थिति में कम से कम जनहानि हो। पर यह "AI का नैतिक आचरण" होते हुए भी श्रेय/आरोप AI को नहीं दिया जा सकता, क्योंकि यह मनुष्य द्वारा डाले गए मूल्यों का प्रतिबिंब है। व्यावहारिक दृष्टि से अभी के लिए यही मार्ग है। **एआई संरेखण**: AI को मानव मूल्यों के अनुरूप चलाना (रसेल, 2019)। बड़े भाषा मॉडल (जैसे जीपीटी-श्रृंखला) को अनुचित सामग्री न बनाने हेतु प्रशिक्षण दिया जाता है, पर यदि वे चूक जाएँ तो दोष मॉडल का या डवलपर का? आम सहमति यही है कि जिम्मेदारी अंततः डवलपरों/नियंत्रकों की है, क्योंकि AI की अपनी "समझ" नहीं।

कानूनी पहलू में भी यही रुझान दिखता है। यूरोपीय आयोग की 2021 की मसौदा AI नियमन आदि में स्पष्ट है कि AI प्रणाली की ज़िम्मेदारी निर्माताओं और उपयोगकर्ताओं पर आएगी। फिर भी जैसे-जैसे AI के फैसले स्वायत्त होते जाते हैं। न्याय-क्षेत्र में एल्गोरिथम निर्णय, भर्ती/छँटाई एल्गोरिथम आदि जिम्मेदारी बाँटना कठिन हो रहा है। कई मामलों में एलगोरिदम के भेदभावपूर्ण परिणाम सामने आए हैं (भर्ती में महिला उम्मीदवारों के साथ पक्षपात, आपराधिक पुनरावृत्ति अनुमान में नस्लीय पक्षपात आदि)। ऐसे में AI को अपराधी नहीं ठहराया जा सकता, पर हानि वास्तविक है। इसलिए समाधानस्वरुप **मानवीय निगरानी** पर जोर दिया जाता है। अंतिम निर्णय में एक मानव नैतिक एजेंट शामिल रहे ताकि जवाबदेही तय हो सके (फ्लोरिडी और काउल्स, 2020)।

भावी दर्शन की दृष्टि से, यदि कभी AI बौद्धिक ही नहीं, "आत्मिक" मामलों में भी होड़ करे। जैसे चेतना को कृत्रिम माध्यम में उत्पन्न करने की परिकल्पनाएँ तो नैतिक चिंतन को विस्तृत करना होगा। पश्चिम में इसे पोस्ट-ह्युमनिज़्म/टेक्नो-सेंट्रिक दर्शन कहा जाता है (गंकेल, 2012)। भारतीय परंपरा पहले से व्यापक रही है। अहिंसा, समता, और वसुधैव कुटुंबकम् जैसी दृष्टि के कारण यदि बुद्धिमान मशीनें उभरें, तो संभवतः उन्हें भी नैतिक "परिवार" में स्थान मिल सकता है, पर यह सदस्यता चेतना एवं संवेदना की शर्त से बंधी होगी।

मशीनी भावनाएँ यहाँ निर्णायक प्रश्न बनती हैं। नैतिक कर्ता बनने के लिए केवल बुद्धि नहीं, सहानुभूति/करुणा जैसी भावनात्मक क्षमता भी आवश्यक है। अभी AI भावनाओं का "अनुकरण" कर सकती है, पर स्वयं महसूस नहीं करती। भारतीय दर्शन में "दया धर्म का मूल" और बौद्ध आदर्श में करुणा के बिना निर्वाण अधूरा मानना इसी बिंदु को पुष्ट करता है। अंततः मशीनों को नैतिक कर्तृत्व देने/न देने का प्रश्न तकनीकी भविष्य के गठन से जुड़ा है: पश्चिमी परंपरा जवाबदेही और इरादे/समझ के मानकों पर जोर देती है, जबकि भारतीय परंपरा चेतना की केंद्रीयता और धर्म के सूक्ष्म विवेक की याद दिलाती है। दोनों मिलकर चेतावनी देते हैं कि विवेक व चेतना के बिना शक्ति खतरनाक हो सकती है।

6. निष्कर्ष : उपरोक्त विश्लेषण के आधार पर निष्कर्ष रूप में कहा जा सकता है कि वर्तमान परिदृश्य में मशीनों को नैतिक कर्ता (मोरल एजेंट) कहना तर्कसंगत नहीं है। पश्चिमी नैतिक दर्शन की मान्य शर्तें – आत्मचेतना, स्वतंत्र इच्छा, इरादा, और नैतिक समझ – इनमें से किसी पर भी आज की ए.आई. खरी नहीं उतरती। भारतीय दर्शन भी यही संकेत देता है कि जहाँ आत्मा/चेतना नहीं, वहाँ नैतिक कर्तृत्व नहीं; और मशीनें तो स्पष्टतः जड़ प्रकृति की रचनाएँ हैं जिनमें चेतना का संचार नहीं हुआ है। इसलिए उनके द्वारा किए गए कार्य नैतिक प्रभाव रखने पर भी उनकी

अपनी नैतिक जिम्मेदारी नहीं बनती। नैतिक दायित्व अंततः उन मनुष्यों का है जिन्होंने ए.आई. को बनाया व चलाया।

हालांकि, यह परिणाम मानवता को अपने उत्तरदायित्व से मुक्त नहीं करता – बल्कि और गहरा बना देता है। अगर मशीन स्वयं नैतिक निर्णय के योग्य नहीं, तो हमें उन्हें ऐसे ढांचे में काम में लाना होगा जहाँ वे मानव नैतिकता का पालन करें। यह मनुष्यों की ज़िम्मेदारी है कि वे ए.आई. के प्रयोग से होने वाले संभावित लाभ और हानि दोनों के प्रति सजग रहें। "ज़िम्मेदारी का अंतराल" भरना मानव समाज का कार्य है – उपयुक्त क़ानून, नीतियाँ और तकनीकी व्यवस्थाएँ विकसित करके। भारतीय अवधारणा में कहें तो, धर्म का पालन अंततः कर्ता पर है – और यहाँ कर्ता मानव है, मशीन नहीं।

भविष्य को पूरी तरह नकारा भी नहीं जा सकता। यदि विज्ञान ऐसी दिशा में बढ़ता है जहाँ कभी मशीनों में मानवीय-स्तर या उससे अधिक की बुद्धि के साथ-साथ चेतना के चिह्न भी दिखाई दें, तो हमें नैतिकता की परिभाषा और दायरे का विस्तार करना होगा। उस काल्पनिक स्थिति में, संभव है हमें कुछ मशीनों को नैतिक समुदाय में आंशिक सदस्यता देनी पड़े – जैसे हम उन्नत प्राइमेट जानवरों या मानव-जैसे एलियंस को नैतिक अधिकार देने में संकोच नहीं करेंगे। लेकिन जब तक ऐसा नहीं होता, हमें दोनों मोर्चों पर काम करना है: एक, ए.आई. को नैतिक व्यवहार हेतु अनुशासित और नियंत्रित रखना (जैसा पश्चिमी "एआई नैतिकता" ज़ोर देता है); और दो, स्वयं मानव अपनी नैतिकता को दृढ़ता से संजोए रहें ताकि प्रौद्योगिकी के उपयोग में धर्म (नैतिक सिद्धांत) की अवहेलना न हो (जिस पर भारतीय चिंतन बल देता है)।

संदर्भ सूची :

1. Adhikary, Ramesh Prasad. "Ethical and Philosophical Parallels of Artificial Intelligence (AI) in the Hindu Mythology The Mahabharat." Mongolian Journal of Arts and Culture, vol. 25, no. 49, 2024, pp. 112–120.
2. Allen, Colin, Gary Varner, and Jason Zinser. "Prolegomena to Any Future Artificial Moral Agent." Journal of Experimental & Theoretical Artificial Intelligence, vol. 12, no. 3, 2000, pp. 251–261.
3. Bhagavad Gita. Translated by Eknath Easwaran, Nilgiri Press, 2007. (Original work approx 500 BCE, attributed to Vyasa)
4. Bostrom, Nick. Superintelligence: Paths, Dangers, Strategies. Oxford University Press, 2014.
5. Davis, Donald R., Jr. The Spirit of Hindu Law. Cambridge University Press, 2010.
6. Gunkel, David J. The Machine Question: Critical Perspectives on AI, Robots, and Ethics. MIT Press, 2012.
7. Johnson, Deborah G. "Computer Systems: Moral Entities but Not Moral Agents." Ethics and Information Technology, vol. 8, no. 4, 2006, pp. 195–204.
8. Katha Upanishad. Translated by S. Radhakrishnan, in The Principal Upanishads, HarperCollins India, 1994. (Original text ~5th century BCE)
9. Mahabharata. Translated by Kisari Mohan Ganguli, 1883–1896.
10. Matthias, Andreas. "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata." Ethics and Information Technology, vol. 6, no. 3, 2004, pp. 175–183.
11. Moor, James H. "The Nature, Importance, and Difficulty of Machine Ethics." IEEE Intelligent Systems, vol. 21, no. 4, 2006, pp. 18–21.
12. Nadeau, John. "Only Androids Can Be Ethical." In Thinking about Android Epistemology, edited by K. Ford et al., MIT Press, 2006, pp. 241–248.
13. Sanwoolu, Oluwaseun D. "Kantian Deontology for AI: Alignment without Moral Agency." AI and Ethics, vol. 5, no. 3, 2025, pp. 223–235.
14. Sullins, John P. "When Is a Robot a Moral Agent?" In Machine Ethics, edited by Michael Anderson and Susan Leigh Anderson, Cambridge University Press, 2011, pp. 151–161.
15. Wallach, Wendell, and Colin Allen. Moral Machines: Teaching Robots Right from Wrong. Oxford University Press, 2009.

•