

कृत्रिम बुद्धिमत्ता के युग में साइबर सुरक्षा

डॉ. कुलदीप सिंह

असि. प्रोफे.रक्षा एवं स्रातेजिक अध्ययन विभाग।

बाबा बरुआ दास पी जी कॉलेज परूडया आश्रम अम्बेडकरनगर।

Email: kuldeep791@gmail.com



Received 05 August 2026

Accepted 10 August 2026

Published 17 August 2026

Corresponding Author:

डॉ. कुलदीप सिंह

Email: kuldeep791@gmail.com

DOI: <https://doi.org/10.64880/pcjaict.v1i4.04>

Funding: This research received no specifics grant from any funding agency in the public, commercial, or not-for-profit sector.

Copyright: © 2026 The Author(s).



This is an open access article under the [Creative Commons Attribution-Noncommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

OPEN ACCESS

शोध सार:

इक्कीसवीं सदी के तीसरे दशक में कृत्रिम बुद्धिमत्ता मानव जीवन, अर्थव्यवस्था, प्रशासन, शिक्षा, स्वास्थ्य, बैंकिंग, उद्योग और संचार व्यवस्था में तीव्र परिवर्तन लाने वाली प्रमुख प्रौद्योगिकी के रूप में उभरी है। मशीन लर्निंग, डीप लर्निंग, प्राकृतिक भाषा प्रसंस्करण, जनरेटिव आर्टिफिशियल इंटेलिजेंस और स्वचालित निर्णय प्रणालियों के विकास ने डिजिटल सेवाओं को अधिक तेज, सक्षम और उपयोगकर्ता-केंद्रित बनाया है। किंतु इसी परिवर्तन के समानांतर साइबर सुरक्षा की प्रकृति भी जटिल हो गई है। आज कृत्रिम बुद्धिमत्ता केवल साइबर सुरक्षा का उपकरण नहीं है, बल्कि साइबर अपराधियों द्वारा उपयोग किया जाने वाला शक्तिशाली माध्यम भी बन चुकी है। AI की सहायता से अत्यंत विश्वसनीय फिशिंग संदेश, डीपफेक वीडियो, नकली आवाज, स्वचालित मालवेयर, पहचान की चोरी तथा बड़े पैमाने पर सोशल इंजीनियरिंग अभियान तैयार किए जा सकते हैं। दूसरी ओर, AI आधारित सुरक्षा प्रणालियाँ विशाल मात्रा में नेटवर्क डेटा का विश्लेषण कर असामान्य गतिविधियों की पहचान, खतरे का पूर्वानुमान, मालवेयर वर्गीकरण और स्वचालित घटना-प्रतिक्रिया में महत्वपूर्ण भूमिका निभा रही हैं।

प्रस्तुत शोध आलेख कृत्रिम बुद्धिमत्ता और साइबर सुरक्षा के अंतर्संबंध का विश्लेषण करता है तथा AI-आधारित साइबर खतरों, AI-सक्षम सुरक्षा उपायों, जनरेटिव AI की जोखिमपूर्ण प्रकृति, भारतीय परिप्रेक्ष्य, डेटा संरक्षण, नीति-निर्माण और भविष्य की सुरक्षा रणनीतियों पर विचार करता है। अध्ययन का निष्कर्ष है कि कृत्रिम बुद्धिमत्ता को साइबर सुरक्षा के लिए पूर्णतः खतरा अथवा पूर्णतः समाधान मानना उचित नहीं है। इसकी उपयोगिता इस बात पर निर्भर करती है कि इसके विकास, प्रयोग और नियंत्रण के लिए किस प्रकार की सुरक्षित, उत्तरदायी और मानव-केंद्रित व्यवस्था स्थापित की जाती है।

मुख्य शब्द: कृत्रिम बुद्धिमत्ता, साइबर सुरक्षा, जनरेटिव AI, साइबर अपराध, फिशिंग, डीपफेक, डेटा सुरक्षा, मशीन लर्निंग, साइबर जोखिम, डिजिटल भारता

1. प्रस्तावना

सूचना और संचार प्रौद्योगिकी के विकास ने आधुनिक समाज की कार्यप्रणाली में व्यापक परिवर्तन किया है। बैंकिंग से लेकर शिक्षा, शासन से लेकर चिकित्सा, परिवहन से लेकर व्यापार और व्यक्तिगत संचार से लेकर राष्ट्रीय सुरक्षा तक लगभग प्रत्येक क्षेत्र डिजिटल प्रणालियों से जुड़ गया है। इस डिजिटल परिवर्तन

के केंद्र में कृत्रिम बुद्धिमत्ता तेजी से अपनी भूमिका का विस्तार कर रही है। आज कंप्यूटर प्रणालियाँ केवल पूर्व निर्धारित निर्देशों का पालन करने तक सीमित नहीं हैं, बल्कि उपलब्ध डेटा से सीखने, पैटर्न पहचानने, भाषा समझने, निर्णय लेने और नवीन सामग्री तैयार करने में सक्षम हो रही हैं।

कृत्रिम बुद्धिमत्ता की यही क्षमता साइबर सुरक्षा के लिए अवसर और संकट दोनों प्रस्तुत करती है। परंपरागत साइबर हमलों में अपराधियों को तकनीकी ज्ञान, समय और संसाधनों की आवश्यकता होती थी। AI के प्रयोग ने कई गतिविधियों को स्वचालित और अधिक प्रभावी बना दिया है। अपराधी अब बड़ी संख्या में संभावित लक्ष्यों का विश्लेषण कर सकते हैं, व्यक्तिगत जानकारी के आधार पर विश्वसनीय फिशिंग संदेश तैयार कर सकते हैं और मानव आवाज अथवा चेहरे की नकल करके धोखाधड़ी कर सकते हैं।

इसी समय सुरक्षा विशेषज्ञ भी AI का उपयोग नेटवर्क निगरानी, विसंगति पहचान, खतरे की भविष्यवाणी, स्वचालित विश्लेषण और घटना-प्रतिक्रिया के लिए कर रहे हैं। परिणामस्वरूप साइबर सुरक्षा अब केवल मनुष्य और मशीन के बीच संघर्ष नहीं रह गई है; कई परिस्थितियों में यह आक्रामक AI और रक्षात्मक AI के बीच प्रतिस्पर्धा का रूप ग्रहण कर रही है।

NIST ने AI और साइबर सुरक्षा के संबंध को तीन प्रमुख क्षेत्रों में समझने की दिशा विकसित की है—AI प्रणालियों की स्वयं की साइबर सुरक्षा, AI-सक्षम साइबर हमले तथा साइबर रक्षा के लिए AI का उपयोग। यह वर्गीकरण स्पष्ट करता है कि आधुनिक साइबर सुरक्षा रणनीति में केवल AI का उपयोग करना पर्याप्त नहीं है; स्वयं AI मॉडल, उसके प्रशिक्षण डेटा, सॉफ्टवेयर आपूर्ति शृंखला और उससे जुड़े संसाधनों की सुरक्षा भी आवश्यक है।

2. अध्ययन के उद्देश्य

प्रस्तुत अध्ययन का प्रमुख उद्देश्य कृत्रिम बुद्धिमत्ता के विकास के कारण परिवर्तित हो रहे साइबर सुरक्षा परिदृश्य का समग्र अध्ययन करना है। इसके अंतर्गत AI-सक्षम साइबर अपराधों की प्रकृति का विश्लेषण करना, साइबर सुरक्षा में AI की रक्षात्मक उपयोगिता को समझना, जनरेटिव AI से उत्पन्न नए जोखिमों का अध्ययन करना तथा भारतीय डिजिटल व्यवस्था के संदर्भ में आवश्यक नीतिगत एवं तकनीकी उपायों की पहचान करना शामिल है।

इसके अतिरिक्त अध्ययन यह समझने का प्रयास करता है कि भविष्य की साइबर सुरक्षा रणनीति में तकनीकी सुरक्षा, मानव जागरूकता, डेटा संरक्षण, AI शासन और संस्थागत उत्तरदायित्व को किस प्रकार एकीकृत किया जा सकता है।

3. शोध पद्धति

यह अध्ययन मुख्यतः वर्णनात्मक एवं विश्लेषणात्मक प्रकृति का है। इसमें द्वितीयक स्रोतों का उपयोग किया गया है। साइबर सुरक्षा और कृत्रिम बुद्धिमत्ता से संबंधित सरकारी दस्तावेजों, तकनीकी दिशानिर्देशों, राष्ट्रीय तथा अंतरराष्ट्रीय संस्थाओं की रिपोर्टों और नीति दस्तावेजों का अध्ययन किया गया है। विशेष रूप से भारतीय कंप्यूटर आपातकालीन प्रतिक्रिया दल (CERT-In), इलेक्ट्रॉनिक्स और सूचना प्रौद्योगिकी मंत्रालय (MeitY), IndiaAI, अमेरिकी National Institute of Standards and Technology (NIST) तथा European Union Agency for Cybersecurity (ENISA) के दस्तावेजों का संदर्भ लिया गया है।

अध्ययन का उद्देश्य किसी एक तकनीकी उत्पाद का मूल्यांकन करना नहीं, बल्कि AI और साइबर सुरक्षा के व्यापक संबंध का सैद्धांतिक तथा व्यावहारिक विश्लेषण प्रस्तुत करना है।

4. कृत्रिम बुद्धिमत्ता और साइबर सुरक्षा का बदलता संबंध

साइबर सुरक्षा का मूल उद्देश्य कंप्यूटर प्रणालियों, नेटवर्क, डिजिटल उपकरणों और सूचनाओं की गोपनीयता, अखंडता और उपलब्धता की रक्षा करना है। परंपरागत सुरक्षा प्रणालियाँ पहले से निर्धारित नियमों, ज्ञात मालवेयर हस्ताक्षरों और संदिग्ध गतिविधियों की निश्चित सूचियों के आधार पर कार्य करती थीं। किंतु आधुनिक साइबर हमले तेजी से अपना स्वरूप बदलते हैं। ऐसे में केवल स्थिर नियमों पर आधारित प्रणालियाँ पर्याप्त नहीं रह जातीं।

AI आधारित प्रणाली बड़ी मात्रा में ऐतिहासिक और वास्तविक समय के डेटा से व्यवहार का सामान्य पैटर्न सीख सकती है। यदि किसी उपयोगकर्ता का लॉग-इन व्यवहार अचानक बदलता है, किसी सर्वर से असामान्य मात्रा में डेटा बाहर जाने लगता है या किसी नेटवर्क में सामान्य से अलग संचार दिखाई देता है, तो मशीन लर्निंग मॉडल संभावित खतरे का संकेत दे सकता है।

परंतु यही क्षमता अपराधियों के लिए भी उपलब्ध है। AI का उपयोग संभावित कमजोरियों का विश्लेषण करने, हमले के लक्ष्य चुनने, भाषा-विशिष्ट धोखाधड़ी संदेश बनाने और पहचान छिपाने के लिए किया जा सकता है। इस प्रकार कृत्रिम बुद्धिमत्ता ने साइबर सुरक्षा में गति, पैमाने और स्वचालन तीनों को प्रभावित किया है।

5. AI-सक्षम साइबर अपराधों का विस्तार

5.1 अधिक प्रभावी फिशिंग

फिशिंग लंबे समय से प्रचलित साइबर धोखाधड़ी है। इसमें उपयोगकर्ता को नकली ई-मेल, संदेश अथवा वेबसाइट के माध्यम से पासवर्ड, बैंकिंग जानकारी या अन्य गोपनीय विवरण देने के लिए प्रेरित किया जाता है।

पूर्व में फिशिंग संदेशों की भाषा अथवा व्याकरण की त्रुटियों से उन्हें पहचानना अपेक्षाकृत आसान हो सकता था। जनरेटिव AI के आगमन के बाद अपराधी बेहतर भाषा, स्थानीय संदर्भ और व्यक्तिगत विवरण के आधार पर अधिक विश्वसनीय संदेश तैयार कर सकते हैं। किसी कर्मचारी के पद, संगठन, सहकर्मियों और सार्वजनिक सोशल मीडिया गतिविधियों को ध्यान में रखते हुए बनाया गया संदेश साधारण फिशिंग की तुलना में कहीं अधिक प्रभावी हो सकता है। ENISA के 2025 Threat Landscape ने भी AI के माध्यम से सामाजिक अभियांत्रिकी अभियानों और फिशिंग की क्षमता बढ़ने की प्रवृत्ति को साइबर खतरे के विकसित होते परिदृश्य का हिस्सा माना है।

5.2 डीपफेक और पहचान आधारित धोखाधड़ी

डीपफेक तकनीक AI का ऐसा अनुप्रयोग है जिसके माध्यम से किसी व्यक्ति की आवाज, चेहरा या वीडियो अत्यंत वास्तविक रूप में कृत्रिम रूप से बनाया जा सकता है। वित्तीय धोखाधड़ी में इसका उपयोग विशेष रूप से चिंताजनक है।

कल्पना कीजिए कि किसी संस्था के वित्त अधिकारी को उसके वरिष्ठ अधिकारी की आवाज में फोन प्राप्त हो और तत्काल धन हस्तांतरण का निर्देश दिया जाए। यदि आवाज और वार्तालाप पर्याप्त विश्वसनीय हो, तो कर्मचारी सामान्य सत्यापन प्रक्रियाओं को नजरअंदाज कर सकता है।

डीपफेक की समस्या केवल आर्थिक अपराध तक सीमित नहीं है। इसका उपयोग प्रतिष्ठा को नुकसान पहुँचाने, गलत सूचना फैलाने, किसी व्यक्ति को झूठे वक्तव्य से जोड़ने और सामाजिक तनाव पैदा करने में भी हो सकता है। इससे डिजिटल सामग्री की विश्वसनीयता का प्रश्न गंभीर होता जा रहा है।

5.3 मालवेयर निर्माण में AI

AI किसी साधारण उपयोगकर्ता को स्वचालित रूप से हानिकारक कोड बनाने की पूर्ण क्षमता दे दे, ऐसा मानना अतिशयोक्ति हो सकता है; फिर भी AI आधारित प्रोग्रामिंग सहायता अपराधियों की दक्षता बढ़ा सकती है। कोड की त्रुटियों को पहचानना, हमले की स्क्रिप्ट को अनुकूलित करना, विभिन्न प्लेटफॉर्म के लिए कोड बदलना तथा सार्वजनिक रूप से उपलब्ध कमजोरियों को समझना AI के माध्यम से सरल हो सकता है।

इससे एक गंभीर समस्या उत्पन्न होती है—जिन व्यक्तियों के पास पहले सीमित प्रोग्रामिंग क्षमता थी, वे भी AI आधारित उपकरणों की सहायता से अधिक जटिल साइबर गतिविधियों में शामिल हो सकते हैं।

5.4 स्वचालित सोशल इंजीनियरिंग

साइबर अपराध केवल तकनीकी कमजोरी पर निर्भर नहीं करता। मनुष्य स्वयं सुरक्षा शृंखला की एक महत्वपूर्ण कड़ी है। अपराधी विश्वास, भय, लालच, अधिकार और तात्कालिकता जैसी मानवीय प्रवृत्तियों का उपयोग करते हैं।

AI सार्वजनिक स्रोतों से व्यक्ति की नौकरी, पसंद, संपर्क, संस्था और गतिविधि जैसी जानकारी का विश्लेषण करके अत्यधिक व्यक्तिगत संदेश तैयार कर सकता है। हजारों लोगों के लिए अलग-अलग सामाजिक अभियांत्रिकी संदेश तैयार करना अब पहले की अपेक्षा अधिक सरल हो सकता है।

5.5 स्वचालित हमले और बड़े पैमाने का खतरा

AI का सबसे महत्वपूर्ण प्रभाव साइबर अपराध की गति और पैमाने पर पड़ता है। जिस कार्य के लिए पहले अनेक लोगों को घंटों या दिनों की आवश्यकता होती थी, स्वचालित प्रणाली उसे अपेक्षाकृत कम समय में कर सकती है।

भविष्य में खतरा केवल इस बात से नहीं होगा कि कोई हमला कितना उन्नत है, बल्कि इस बात से भी होगा कि एक ही समय में कितनी प्रणालियों पर हमला किया जा सकता है। इसलिए सुरक्षा व्यवस्था को भी समान रूप से स्वचालित और त्वरित बनाना आवश्यक है।

6. साइबर सुरक्षा के रक्षक के रूप में कृत्रिम बुद्धिमत्ता

कृत्रिम बुद्धिमत्ता को केवल साइबर खतरे के रूप में देखना अधूरा दृष्टिकोण होगा। वास्तव में आधुनिक साइबर सुरक्षा में AI की उपयोगिता तेजी से बढ़ी है।

6.1 असामान्य व्यवहार की पहचान

किसी बड़े संगठन के नेटवर्क में प्रति दिन लाखों गतिविधियाँ हो सकती हैं। प्रत्येक गतिविधि का मानव द्वारा विश्लेषण करना संभव नहीं है। AI सामान्य व्यवहार का पैटर्न विकसित कर सकता है और उससे असामान्य गतिविधि की पहचान कर सकता है।

उदाहरण के लिए, यदि कोई कर्मचारी सामान्यतः भारत से कार्य करता है और अचानक कुछ ही समय के भीतर किसी दूसरे महाद्वीप से उसके खाते का प्रयोग दिखाई देता है, तो प्रणाली संभावित अनधिकृत प्रवेश का संकेत दे सकती है।

6.2 मालवेयर पहचान

पारंपरिक एंटीवायरस ज्ञात मालवेयर हस्ताक्षरों पर निर्भर रहते थे। समस्या तब आती है जब मालवेयर का नया संस्करण बनाया जाता है। मशीन लर्निंग प्रणालियाँ केवल फाइल के नाम पर निर्भर न रहकर उसके व्यवहार, संरचना और क्रियाओं का विश्लेषण कर सकती हैं।

इस प्रकार पूर्व में अज्ञात मालवेयर की पहचान की संभावना बढ़ जाती है।

6.3 खतरे का पूर्वानुमान

AI आधारित Threat Intelligence प्रणाली विभिन्न स्रोतों से कमजोरियों, दुर्भावनापूर्ण वेबसाइटों, IP पत्तों, नए मालवेयर और हमले की तकनीकों से संबंधित सूचना का विश्लेषण कर सकती है। इसके माध्यम से सुरक्षा दल हमले के बाद प्रतिक्रिया करने के बजाय पहले से तैयारी कर सकते हैं।

6.4 स्वचालित Incident Response

साइबर हमले में प्रतिक्रिया का समय अत्यंत महत्वपूर्ण होता है। यदि संदिग्ध उपकरण को नेटवर्क से अलग करने में कई घंटे लग जाएँ तो अपराधी बड़ी मात्रा में डेटा चुरा सकता है।

AI आधारित प्रणालियाँ कुछ पूर्व निर्धारित परिस्थितियों में संक्रमित उपकरण को अलग कर सकती हैं, संदिग्ध खाते को अस्थायी रूप से रोक सकती हैं और सुरक्षा अधिकारियों को प्राथमिकता के आधार पर चेतावनी भेज सकती हैं।

हालाँकि अत्यधिक स्वचालन भी जोखिमपूर्ण हो सकता है। गलत निर्णय से वैध उपयोगकर्ता अवरुद्ध हो सकते हैं अथवा महत्वपूर्ण सेवा बाधित हो सकती है। इसलिए संवेदनशील निर्णयों में मानव पर्यवेक्षण आवश्यक रहना चाहिए।

7. AI प्रणाली स्वयं भी साइबर हमले का लक्ष्य

AI को सुरक्षा उपकरण के रूप में अपनाने समय यह मान लेना उचित नहीं है कि AI स्वयं सुरक्षित है। AI प्रणाली कई प्रकार के हमलों का लक्ष्य बन सकती है।

7.1 Data Poisoning

AI मॉडल की गुणवत्ता उसके प्रशिक्षण डेटा पर निर्भर करती है। यदि अपराधी प्रशिक्षण डेटा में जानबूझकर गलत अथवा हानिकारक सूचना सम्मिलित कर दे, तो मॉडल गलत व्यवहार सीख सकता है। इसे डेटा पॉइजनिंग कहा जाता है।

7.2 Adversarial Attacks

कुछ AI प्रणालियों को विशेष रूप से तैयार इनपुट देकर भ्रमित किया जा सकता है। उदाहरण के लिए किसी चित्र में ऐसे सूक्ष्म परिवर्तन किए जा सकते हैं जो मनुष्य को लगभग दिखाई न दें, परंतु AI मॉडल का वर्गीकरण बदल जाए।

7.3 Prompt Injection

जनरेटिव AI प्रणालियों में Prompt Injection विशेष चिंता का विषय बन चुका है। आक्रमणकर्ता मॉडल को ऐसे निर्देश दे सकता है जिससे वह अपनी निर्धारित सीमाओं से बाहर जाकर कार्य करे या ऐसी जानकारी प्रस्तुत करे जिसे सार्वजनिक नहीं होना चाहिए।

7.4 संवेदनशील डेटा का रिसाव

कर्मचारी सुविधा के लिए ऑनलाइन AI उपकरण में कंपनी का गोपनीय दस्तावेज, ग्राहक डेटा, प्रोग्राम कोड या व्यावसायिक योजना डाल सकता है। यदि संगठन की स्पष्ट नीति नहीं है तो यह गंभीर सूचना सुरक्षा जोखिम बन सकता है।

CERT-In ने जनरेटिव AI के उपयोग से संबंधित अपनी सलाह में स्पष्ट रूप से संवेदनशील व्यक्तिगत विवरण, ग्राहक सूचना, बौद्धिक संपदा अथवा अन्य गोपनीय डेटा को अनियंत्रित जनरेटिव AI सेवाओं में साझा करने से बचने की सलाह दी है। CERT-In ने AI उपकरणों के access rights को भी सावधानीपूर्वक नियंत्रित करने पर बल दिया है।

8. जनरेटिव AI और साइबर सुरक्षा

Chatbot, text generator, code assistant, image generator और AI agents जैसे जनरेटिव AI अनुप्रयोगों ने कार्यकुशलता में उल्लेखनीय वृद्धि की है। लेकिन इनकी विशेषताओं के कारण नए सुरक्षा प्रश्न उत्पन्न हुए हैं।

सबसे पहली चुनौती सत्यता की है। जनरेटिव AI कभी-कभी ऐसी सूचना उत्पन्न कर सकता है जो भाषा की दृष्टि से विश्वसनीय प्रतीत होती है, परंतु तथ्यात्मक रूप से गलत होती है। यदि सुरक्षा विशेषज्ञ बिना सत्यापन AI द्वारा दिए गए समाधान पर निर्भर करे, तो गलत सुरक्षा कॉन्फिगरेशन लागू हो सकता है।

दूसरी समस्या गोपनीयता की है। कर्मचारी अनजाने में संवेदनशील दस्तावेज AI प्रणाली में डाल सकते हैं।

तीसरी समस्या एजेंट आधारित AI से संबंधित है। यदि AI प्रणाली केवल उत्तर देने के बजाय ई-मेल भेजने, फाइल खोलने, डेटाबेस से जुड़ने, भुगतान प्रणाली तक पहुँचने अथवा स्वचालित निर्णय लेने की अनुमति रखती है, तो उसके compromise होने का प्रभाव अधिक गंभीर हो सकता है।

NIST के Generative AI Profile ने जनरेटिव AI से संबंधित जोखिमों को व्यवस्थित रूप से पहचानने, मापने और प्रबंधित करने के लिए AI Risk Management Framework को विस्तारित किया है। NIST का दृष्टिकोण Governance, Mapping, Measurement और Management जैसे आयामों पर आधारित जोखिम प्रबंधन को प्रोत्साहित करता है।

9. भारतीय परिप्रेक्ष्य

भारत विश्व के सबसे तेजी से डिजिटल होते समाजों में से एक है। ऑनलाइन बैंकिंग, UPI, आधार आधारित सेवाएँ, डिजिटल लॉकर, ई-कॉमर्स, ऑनलाइन शिक्षा, डिजिटल स्वास्थ्य सेवाएँ और ई-गवर्नेंस के विस्तार ने नागरिकों के लिए अनेक सुविधाएँ प्रदान की हैं। किंतु डिजिटल गतिविधियों के विस्तार के साथ आक्रमण की संभावित सतह भी विस्तृत हुई है।

CERT-In के अनुसार 2024 में 20,41,360 साइबर सुरक्षा घटनाओं का प्रबंधन किया गया। 2023 में यह संख्या 15,92,917 थी। उपलब्ध आँकड़ों के आधार पर यह लगभग 28 प्रतिशत की वृद्धि है। घटनाओं में वेबसाइट घुसपैठ, मालवेयर, फिशिंग, दुर्भावनापूर्ण कोड तथा Distributed Denial of Service जैसी श्रेणियाँ शामिल रही हैं।

भारत सरकार ने AI विकास को बढ़ावा देने के लिए IndiaAI Mission को आगे बढ़ाया है। इसके साथ सुरक्षित और उत्तरदायी AI शासन को भी महत्व दिया गया है। नवंबर 2025 में जारी India AI Governance Guidelines ने सुरक्षित, उत्तरदायी और समावेशी AI विकास के लिए नीतिगत दिशा को मजबूत किया।

साइबर सुरक्षा और AI के इस संयुक्त संदर्भ में भारत के सामने दोहरा लक्ष्य है। एक ओर AI नवाचार को बाधित किए बिना प्रोत्साहित करना आवश्यक है, दूसरी ओर नागरिकों, सरकारी प्रणालियों, वित्तीय संस्थानों और महत्वपूर्ण डिजिटल अवसंरचना की सुरक्षा भी सुनिश्चित करनी है।

10. डेटा संरक्षण और गोपनीयता

AI प्रणालियों को विकसित और संचालित करने के लिए बड़ी मात्रा में डेटा की आवश्यकता हो सकती है। यदि इस डेटा में नाम, मोबाइल नंबर, बायोमेट्रिक विवरण, वित्तीय जानकारी अथवा अन्य व्यक्तिगत विवरण शामिल हों, तो गोपनीयता महत्वपूर्ण विषय बन जाती है।

भारत का Digital Personal Data Protection Act, 2023 डिजिटल व्यक्तिगत डेटा के वैध प्रसंस्करण, सहमति, डेटा प्रिंसिपल के अधिकार और डेटा फिड्यूशियरी की जिम्मेदारियों के लिए कानूनी ढाँचा प्रस्तुत करता है। इसके बाद Digital Personal Data Protection Rules, 2025 भी अधिसूचित किए गए हैं।

AI के संदर्भ में privacy by design के सिद्धांत को अपनाना आवश्यक है। संस्थाओं को केवल उतना ही डेटा एकत्र करना चाहिए जितना आवश्यक हो। डेटा का सुरक्षित भंडारण, encryption, नियंत्रित पहुँच, audit logs और समयबद्ध deletion जैसी प्रक्रियाएँ महत्वपूर्ण हैं।

AI को प्रभावी बनाने के नाम पर अनियंत्रित डेटा संग्रह भविष्य में बड़े साइबर जोखिम में बदल सकता है। जितना अधिक संवेदनशील डेटा किसी एक प्रणाली के पास होगा, उसके उल्लंघन का प्रभाव उतना ही गंभीर होगा।

11. मानव कारक और साइबर जागरूकता

अत्याधुनिक सुरक्षा तकनीक के बावजूद साइबर सुरक्षा की सबसे महत्वपूर्ण कड़ी मनुष्य है। अपराधी अक्सर तकनीकी सुरक्षा प्रणाली को तोड़ने की अपेक्षा कर्मचारी को भ्रमित करना अधिक आसान समझते हैं।

AI के कारण यह समस्या बढ़ जाती है क्योंकि नकली ई-मेल, आवाज और वीडियो अधिक विश्वसनीय बनाए जा सकते हैं। इसलिए केवल यह प्रशिक्षण देना पर्याप्त नहीं है कि “अज्ञात लिंक पर क्लिक न करें।” प्रशिक्षण में डीपफेक, AI जनित संदेश, QR कोड धोखाधड़ी, पहचान प्रतिरूपण तथा असामान्य भुगतान निर्देशों की पहचान भी शामिल करनी होगी।

संस्थानों को महत्वपूर्ण आर्थिक और प्रशासनिक आदेश के लिए बहुस्तरीय सत्यापन प्रणाली विकसित करनी चाहिए। उदाहरणतः बड़ी राशि के हस्तांतरण के लिए केवल ई-मेल अथवा आवाज आधारित निर्देश पर्याप्त नहीं होना चाहिए। दूसरे स्वतंत्र माध्यम से पुष्टि आवश्यक होनी चाहिए।

12. कृत्रिम बुद्धिमत्ता आधारित साइबर सुरक्षा की सीमाएँ

AI अत्यंत उपयोगी है, किंतु इसे जादुई समाधान मानना गंभीर गलती होगी। पहली सीमा प्रशिक्षण डेटा से जुड़ी है। यदि डेटा अपूर्ण अथवा पक्षपाती है, तो सुरक्षा प्रणाली का निर्णय भी प्रभावित हो सकता है।

दूसरी सीमा false positive की है। AI कभी-कभी सामान्य गतिविधि को खतरे के रूप में पहचान सकती है। इससे सुरक्षा दल पर अनावश्यक चेतावनियों का दबाव बढ़ सकता है।

तीसरी सीमा explainability की है। कुछ जटिल मॉडल यह स्पष्ट नहीं कर पाते कि उन्होंने किसी गतिविधि को खतरनाक क्यों माना। उच्च जोखिम वाले निर्णयों में यह पारदर्शिता आवश्यक है।

चौथी समस्या AI पर अत्यधिक निर्भरता है। यदि संगठन मानव सुरक्षा विशेषज्ञों की भूमिका कम करके पूर्णतः स्वचालित प्रणाली पर निर्भर हो जाए तो किसी नए प्रकार के हमले के सामने प्रणाली विफल हो सकती है।

इसलिए भविष्य का प्रभावी मॉडल “AI बनाम मानव” नहीं बल्कि “AI के साथ प्रशिक्षित मानव विशेषज्ञ” होना चाहिए।

13. भविष्य की प्रमुख चुनौतियाँ

भविष्य में AI और साइबर सुरक्षा का संबंध और अधिक जटिल होने की संभावना है। स्वायत्त AI एजेंट स्वतंत्र रूप से अनेक डिजिटल कार्य कर सकेंगे। ऐसे एजेंट के पास यदि ई-मेल, दस्तावेज, वित्तीय सेवाओं और डेटाबेस तक पहुँच हो तो सुरक्षा उल्लंघन का प्रभाव अत्यधिक व्यापक हो सकता है।

दूसरी चुनौती AI मॉडल की आपूर्ति शृंखला है। मॉडल निर्माण में प्रशिक्षण डेटा, open-source libraries, APIs, cloud services और अनेक बाहरी घटक उपयोग होते हैं। इनमें से किसी एक घटक में कमजोरी पूरे AI सिस्टम को प्रभावित कर सकती है।

इसी कारण CERT-In ने Software Bill of Materials के साथ AI Bill of Materials जैसे तकनीकी दृष्टिकोणों पर भी दिशानिर्देश जारी किए हैं, ताकि जटिल डिजिटल प्रणालियों के घटकों की पहचान और जोखिम प्रबंधन बेहतर हो सके।

तीसरी चुनौती साइबर हमलों की अंतरराष्ट्रीय प्रकृति है। हमलावर किसी देश में, सर्वर दूसरे देश में और पीड़ित तीसरे देश में हो सकता है। इसके लिए अंतरराष्ट्रीय सहयोग, साझा threat intelligence और प्रभावी कानूनी तंत्र की आवश्यकता होगी।

14. साइबर सुरक्षा को मजबूत करने के लिए सुझाव

AI युग की साइबर सुरक्षा के लिए बहुस्तरीय रणनीति अपनाने की आवश्यकता है। सबसे पहले प्रत्येक संगठन को AI उपयोग नीति बनानी चाहिए जिसमें स्पष्ट किया जाए कि कर्मचारियों को कौन से AI उपकरण प्रयोग करने की अनुमति है और किस प्रकार की सूचना उनमें दर्ज नहीं की जा सकती।

दूसरा, Zero Trust Architecture को बढ़ावा देना चाहिए। किसी उपयोगकर्ता अथवा उपकरण को केवल इसलिए स्थायी रूप से विश्वसनीय नहीं माना जाना चाहिए क्योंकि वह संगठन के नेटवर्क के भीतर है। प्रत्येक महत्वपूर्ण अनुरोध का सत्यापन आवश्यक होना चाहिए।

तीसरा, Multi-Factor Authentication को व्यापक रूप से लागू करना चाहिए। केवल पासवर्ड पर आधारित व्यवस्था आधुनिक साइबर वातावरण के लिए पर्याप्त नहीं है।

चौथा, AI प्रणालियों का नियमित सुरक्षा परीक्षण आवश्यक है। इसमें prompt injection testing, adversarial testing, data leakage assessment और access-control audit शामिल होना चाहिए।

पाँचवाँ, संगठन को AI द्वारा उत्पन्न सुरक्षा चेतावनियों के लिए Human-in-the-loop व्यवस्था विकसित करनी चाहिए ताकि महत्वपूर्ण निर्णय केवल स्वचालित प्रणाली पर निर्भर न हों।

छठा, कर्मचारियों को निरंतर प्रशिक्षण दिया जाना चाहिए। एक बार आयोजित साइबर सुरक्षा कार्यशाला पर्याप्त नहीं है क्योंकि साइबर अपराध की तकनीक तेजी से बदलती रहती है।

सातवाँ, नियमित backup, network segmentation और incident response plan बनाए रखना आवश्यक है। Ransomware के संदर्भ में यह विशेष रूप से उपयोगी है। CERT-In ने ransomware जोखिमों और प्रतिक्रिया रणनीतियों पर भी विस्तृत रिपोर्ट प्रकाशित की है।

आठवाँ, AI मॉडल की पूरी जीवन-चक्र सुरक्षा सुनिश्चित होनी चाहिए—डेटा संग्रह से प्रशिक्षण, deployment, monitoring और retirement तक।

नौवाँ, महत्वपूर्ण सरकारी, स्वास्थ्य, वित्तीय, रक्षा और ऊर्जा प्रणालियों में AI आधारित निर्णय के लिए अतिरिक्त audit और accountability व्यवस्था आवश्यक होनी चाहिए।

अंततः नागरिक स्तर पर डिजिटल साक्षरता को साइबर सुरक्षा नीति का अनिवार्य भाग बनाया जाना चाहिए। सुरक्षित डिजिटल समाज केवल तकनीकी विशेषज्ञों से निर्मित नहीं होता; इसके लिए जागरूक उपयोगकर्ताओं की भी आवश्यकता होती है।

15. चर्चा

कृत्रिम बुद्धिमत्ता और साइबर सुरक्षा के संबंध का अध्ययन यह स्पष्ट करता है कि आधुनिक डिजिटल वातावरण में खतरा और सुरक्षा दोनों अधिक बुद्धिमान हो रहे हैं। अपराधी AI का उपयोग भाषा निर्माण, पहचान की नकल, स्वचालन और हमले की दक्षता बढ़ाने के लिए कर सकते हैं। सुरक्षा संस्थाएँ भी AI का उपयोग असामान्य व्यवहार पहचानने, खतरे का विश्लेषण करने और त्वरित प्रतिक्रिया के लिए कर रही हैं।

यह एक प्रकार की तकनीकी प्रतिस्पर्धा है जिसमें स्थिर सुरक्षा नीति लंबे समय तक पर्याप्त नहीं रह सकती। सुरक्षा प्रणाली को निरंतर सीखना और अद्यतन होना पड़ेगा।

महत्वपूर्ण प्रश्न यह नहीं है कि AI साइबर सुरक्षा के लिए अच्छा है या बुरा। वास्तविक प्रश्न यह है कि AI का नियंत्रण किसके पास है, उसे किस उद्देश्य से उपयोग किया जा रहा है, उसके डेटा का स्रोत क्या है और निर्णयों के लिए कौन उत्तरदायी है।

NIST का AI Risk Management Framework इस बात पर बल देता है कि AI जोखिम को साइबर सुरक्षा, गोपनीयता और अन्य संगठनात्मक जोखिमों से अलग करके नहीं देखा जाना चाहिए। एकीकृत जोखिम प्रबंधन अधिक प्रभावी दृष्टिकोण प्रदान करता है।

भारतीय संदर्भ में यह विशेष रूप से प्रासंगिक है। भारत की डिजिटल अर्थव्यवस्था और AI पारिस्थितिकी तंत्र तेजी से विस्तृत हो रहे हैं। इसलिए सुरक्षित और विश्वसनीय AI का विकास डिजिटल विकास की पूर्व शर्त बनता जा रहा है।

16. निष्कर्ष

कृत्रिम बुद्धिमत्ता ने साइबर सुरक्षा की पारंपरिक अवधारणा में मौलिक परिवर्तन उत्पन्न किया है। आज साइबर अपराध केवल वायरस, पासवर्ड चोरी या वेबसाइट हैकिंग तक सीमित नहीं है। AI ने धोखाधड़ी को अधिक व्यक्तिगत, स्वचालित और विश्वसनीय बनाने की क्षमता प्रदान की है। डीपफेक, AI-सक्षम फिशिंग, स्वचालित सोशल इंजीनियरिंग, डेटा पॉइजनिंग, prompt injection और AI प्रणाली पर प्रतिकूल हमले आधुनिक साइबर सुरक्षा के प्रमुख प्रश्न बन रहे हैं।

साथ ही, यही कृत्रिम बुद्धिमत्ता सुरक्षा विशेषज्ञों को विशाल डेटा के विश्लेषण, असामान्य गतिविधि की पहचान, मालवेयर वर्गीकरण और खतरे की त्वरित प्रतिक्रिया में अभूतपूर्व क्षमता प्रदान करती है। इसलिए AI को प्रतिबंधित करना अथवा उस पर पूर्ण निर्भरता, दोनों ही चरम दृष्टिकोण उचित नहीं हैं।

भविष्य की साइबर सुरक्षा व्यवस्था तकनीक, मानव विशेषज्ञता, नीति, कानून और नैतिक उत्तरदायित्व के संतुलित संयोजन पर आधारित होनी चाहिए। AI प्रणाली को प्रारंभ से ही security-by-design और privacy-by-design सिद्धांतों के अनुसार विकसित करना आवश्यक है। इसके साथ डेटा न्यूनतमकरण, बहुस्तरीय प्रमाणीकरण, Zero Trust, नियमित सुरक्षा परीक्षण, स्वतंत्र audit और incident response planning को संस्थागत रूप देना होगा।

भारत में डिजिटल सेवाओं का तीव्र विस्तार यह अवसर प्रदान करता है कि देश AI आधारित सुरक्षित डिजिटल अवसंरचना के निर्माण में अग्रणी भूमिका निभाए। इसके लिए IndiaAI जैसी पहलों, CERT-In की तकनीकी विशेषज्ञता, डेटा संरक्षण व्यवस्था, शैक्षणिक अनुसंधान तथा निजी क्षेत्र के नवाचार के बीच प्रभावी समन्वय आवश्यक होगा।

अंततः साइबर सुरक्षा को केवल कंप्यूटर प्रणाली की सुरक्षा नहीं बल्कि नागरिकों के विश्वास, व्यक्तिगत गोपनीयता, आर्थिक स्थिरता और राष्ट्रीय सुरक्षा से जुड़ा विषय मानना होगा। कृत्रिम बुद्धिमत्ता के युग में वही समाज अधिक सुरक्षित होगा जो AI की शक्ति का उपयोग तो करे, किंतु उसके जोखिमों के प्रति

भी समान रूप से सजग रहे। मानव विवेक, तकनीकी नवाचार और उत्तरदायी शासन का संतुलन ही भविष्य के सुरक्षित डिजिटल समाज का वास्तविक आधार बन सकता है।

संदर्भ

1. Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. National Institute of Standards and Technology.
2. CERT-In. (2024). *Annual Report 2024*. Indian Computer Emergency Response Team, Ministry of Electronics and Information Technology, Government of India.
3. CERT-In. (2025). *Best Practices against Vulnerabilities while Using Generative AI Solutions*. Indian Computer Emergency Response Team.
4. CERT-In. (2025). *Digital Threat Report 2024 for the BFSI Sector*. Indian Computer Emergency Response Team.
5. CERT-In. (2024). *Ransomware Report 2024*. Indian Computer Emergency Response Team.
6. ENISA. (2025). *ENISA Threat Landscape 2025*. European Union Agency for Cybersecurity.
7. Government of India. (2023). *The Digital Personal Data Protection Act, 2023*. Ministry of Electronics and Information Technology.
8. Ministry of Electronics and Information Technology. (2025). *Digital Personal Data Protection Rules, 2025*. Government of India.
9. IndiaAI. (2025). *India AI Governance Guidelines*. Ministry of Electronics and Information Technology, Government of India.
10. National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. U.S. Department of Commerce.
11. National Institute of Standards and Technology. (2025). *Cybersecurity and AI: Integrating and Building on Existing NIST Guidelines*. NIST.
12. CERT-In. (2025). *Technical Guidelines on SBOM, QBOM & CBOM, AIBOM and HBOM*. Indian Computer Emergency Response Team, Government of India.

Disclaimer/Publisher's Note: The views, findings, and opinions expressed in this publication are solely those of the author(s). The publisher and editorial team are not responsible for any errors, claims, or consequences arising from the use of the published content.