Journal of Strategic Futures Studies
Vol. 3, No. 11, Autumn 2024
pp. 7-30

Artificial Intelligence Risks for Critical Infrastructure and Future Directions

Ahmadreza Mirzaei[1]

Received: 24-05-2024 **Accepted: 25-08-2024**

Abstract

Artificial Intelligence (AI), like any human-made artefact, comes with both challenges and benefits that must be considered together to enable optimal utilization. This study examines the challenges and best practices associated with managing AI risks in critical infrastructure. By analyzing significant incidents, case studies, and emerging trends, it provides insights into effective risk management strategies for AI deployment. Moreover, the research highlights the crucial role of robust cybersecurity measures, system constraints, and secure mechanisms for AI-based critical infrastructure. It also explores emerging trends in AI for critical infrastructure, such as edge computing, explainable AI, and federated learning. These trends offer opportunities to enhance performance, transparency, and privacy in AI systems operating within critical environment. By combining these multidimensional components, this study contributes to understanding AI risk management in critical infrastructure, offering valuable insights into challenges, best practices, and future directions for effective risk governance. Data collection is based on a document review, and future-oriented recommendations are validated by expert opinions. Policymakers, researchers, and industry professionals can leverage these findings to strengthen the safe and sustainable deployment of AI in critical infrastructure.

Keywords: Artificial intelligence, Risk management, Critical infrastructure, Emerging trends, Future directions.

1. PhD Graduate in Futures Studies, Faculty of Social Sciences, Imam Khomeini International University, Qazvin, Iran arm.official@chmail.ir

فصلنامه آینده‌پژوهی راهبردی
سال سوم، شماره ۱۱، پاییز ۱۴۰۳
مقاله اول از صفحات ۷-۳۰

مخاطرات هوش مصنوعی برای زیرساخت‌های حیاتی و جهت‌گیری‌های آینده

احمدرضا میرزائی[۱]

تاریخ دریافت: ۱۴۰۳/۰۳/۰۶ تاریخ پذیرش: ۱۴۰۳/۰۶/۰۴

چکیده

هوش مصنوعی مانند هر مصنوع ساخته بشر واجد چالش‌ها و منافعی است که باید در کنار یکدیگر دیده شوند تا بهینه‌ترین شیوه‌های بهره‌برداری از آن امکان‌پذیر شود. پژوهش حاضر چالش‌ها و بهترین شیوه‌های مرتبط با مدیریتِ مخاطراتی هوش مصنوعی در زیرساخت‌های حیاتی را بررسی کرده و با تجزیه و تحلیل حوادث قابل توجه، بررسی مطالعات موردی و بررسی روندهای نوظهور، بینش‌هایی را در مورد راهبردهای مدیریت مخاطره مؤثر برای استقرار هوش مصنوعی ارائه می‌دهد. به‌علاوه، نقش حیاتی اقدامات امنیت سایبری قوی، محدودیت‌های سامانه و سازوکارهای ایمن در مورد زیرساخت‌های حیاتی مبتنی بر هوش مصنوعی در این تحقیق برجسته شده است. در کنار این موارد، همچنین بررسی روندهای نوظهور در هوش مصنوعی برای زیرساخت‌های حیاتی، مانند محاسبات لبه، هوش مصنوعی قابل توضیح و یادگیری مشارکتی موضوعیت پیدا می‌کند؛ چراکه این روندها فرصت‌هایی را برای بهبود عملکرد، شفافیت و حریم خصوصی سامانه‌های هوش مصنوعی در محیط‌های زیرساختی حیاتی ارائه می‌دهند. با ترکیب ستون‌های این مجموعه محتوای چندبعدی، پژوهش حاضر به درک مدیریت مخاطرات هوش مصنوعی در زیرساخت‌های حیاتی کمک کرده و بینش‌های ارزشمندی را در مورد چالش‌ها، بهترین شیوه‌ها و جهت‌گیری‌های آینده برای مدیریت مؤثر مخاطره ارائه می‌دهد. گرداوری داده‌ها براساس مطالعه اسنادی و تجویزهای آینده‌نگر براساس اسناد و اعتبارسنجی خبرگانی صورت گرفته است. سیاستگذاران، محققان و متخصصان صنعت می‌توانند از این یافته‌ها برای تقویت استقرار هوش مصنوعی ایمن و پایدار در زیرساخت‌های حیاتی استفاده کنند.

کلیدواژه‌ها: هوش مصنوعی، مدیریت مخاطره، زیرساخت‌های حیاتی، روندهای نوظهور.

۱. دکتری آینده‌پژوهی، دانشکده علوم اجتماعی، دانشگاه بین‌المللی امام خمینی (رحمت‌الله‌علیه)، قزوین، ایران
arm.official@chmail.ir

مقدمه

هوش مصنوعی[1] به‌عنوان یک فناوری قدرتمند با طیف گسترده‌ای از کاربردها ظهور نموده و صنایع و بخش‌های مختلف را متحول کرده است. در سال‌های اخیر، هوش مصنوعی به‌طور فزاینده‌ای نقش مهمی در زیرساخت‌های حیاتی ایفا کرده است که شامل سامانه‌های ضروری می‌شود و از عملکردهای لجستیکی از جمله شبکه‌های حمل‌ونقل، شبکه‌های انرژی، مخابرات و تأمین آب پشتیبانی می‌کنند.

ادغام هوش مصنوعی در زیرساخت‌های حیاتی، ظرفیت افزایش کارایی عملیاتی، بهینه‌سازی تخصیص منابع و بهبود فرایندهای تصمیم‌گیری را منجر می‌شود. فناوری‌های هوش مصنوعی، مانند الگوریتم‌های یادگیری ماشین و تجزیه و تحلیل داده‌ها، امکان پردازش حجم وسیعی از داده‌ها و استخراج بینش‌های معنادار را فراهم می‌کنند. این امر کاربران زیرساخت‌های حیاتی را قادر می‌سازد تا تصمیم‌های آگاهانه و به‌موقع بیشتری اتخاذ کنند که منجر به بهبود عملکرد و پایداری می‌شود (Perera et al., 2014).

درحالی‌که استفاده از هوش مصنوعی در زیرساخت‌های حیاتی مزایای متعددی را ارائه می‌دهد، مجموعه‌ای از مخاطرات و چالش‌ها را نیز ایجاد می‌کند که باید به‌دقت مورد توجه قرار گیرند. درک و کاهش این مخاطرات برای اطمینان از امنیت[2]، قابلیت اطمینان[3] و ایمنی[4] سامانه‌های زیرساخت حیاتی است (Hahn & Werners, 2018).

یکی از مخاطرات مهم مرتبط با هوش مصنوعی در زیرساخت‌های حیاتی، افزایش آسیب‌پذیری امنیت سایبری است. از آنجایی‌که سامانه‌های هوش مصنوعی در بخش‌های زیرساختی حیاتی رایج‌تر می‌شوند، توجه بازیگران مخربی را که به‌دنبال بهره‌برداری از نقاط ضعف بالقوه هستند، جلب می‌کنند. حملات سایبری که زیرساخت‌های حیاتی را هدف قرار می‌دهند می‌توانند عواقب شدیدی داشته باشند که منجر به اختلال در خدمات، نقض داده‌ها و حتی به‌خطر افتادن امنیت عمومی شود. حفاظت از زیرساخت‌های حیاتی در برابر تهدیدهای سایبری برای محافظت از یکپارچگی و عملکرد این سامانه‌ها ضروری است (Rathore et al., 2016).

1. Artificial Intelligence (AI)
2. Security
3. Reliability
4. Safety

علاوه بر این، استقرار سامانه‌های کنترل‌شده با هوش مصنوعی نگرانی‌های ایمنی را در زیرساخت‌های حیاتی مطرح می‌کند. قابلیت‌های تصمیم‌گیری مستقل سامانه‌های هوش مصنوعی، درحالی‌که در بسیاری از موارد سودمند است، سؤالاتی را در مورد قابلیت اطمینان و مخاطرات احتمالی آنها ایجاب می‌کند (Kshetri, 2017). اطمینان از ایمنی سامانه‌های زیرساختی حیاتی مبتنی بر هوش مصنوعی برای جلوگیری از حوادث، خرابی‌ها یا پیامدهای ناخواسته‌ای که می‌تواند به زندگی انسان‌ها یا محیط‌زیست آسیب برساند، از اهمیت بالایی برخوردار است (Shu & Zhang, 2019).

افزون بر امنیت سایبری و نگرانی‌های ایمنی، ملاحظات اخلاقی و قانونی زمانی مطرح می‌شوند که هوش مصنوعی در زیرساخت‌های حیاتی پیاده‌سازی شود. فرایندهای تصمیم‌گیری سامانه‌های هوش مصنوعی می‌تواند معضلات اخلاقی و پرسش‌هایی را در مورد شفافیت، انصاف و مسئولیت‌پذیری ایجاد کند. توسعه دستورالعمل‌های اخلاقی و چارچوب‌های قانونی برای رفع این نگرانی‌ها و اطمینان از استفاده مسئولانه و بی‌طرفانه از فناوری‌های هوش مصنوعی در زیرساخت‌های حیاتی بسیار مهم است.

برای به‌دست آوردن درک عمیق‌تر از مخاطرات بالقوه مرتبط با هوش مصنوعی در زیرساخت‌های حیاتی، این مقاله مروری به بررسی ادبیات موجود، مطالعات موردی و گزارش‌های مربوطه می‌پردازد. با تجزیه و تحلیل این منابع، آسیب‌پذیری‌های امنیت سایبری، نگرانی‌های ایمنی و پیامدهای اخلاقی استقرار هوش مصنوعی در زیرساخت‌های حیاتی را بررسی خواهیم کرد. همچنین، ما راهبردهای کاهش، بهترین شیوه‌ها و توصیه‌های سیاستی برای مدیریت مؤثر این مخاطرات و ارتقای ادغام ایمن و مسئولانه هوش مصنوعی در زیرساخت‌های حیاتی را شناسایی و مورد بحث قرار خواهیم داد.

۱. مبانی نظری

هوش مصنوعی در زیرساخت‌های حیاتی اهمیت قابل‌توجهی پیدا کرده است که باعث افزایش کارایی، بهینه‌سازی منابع و بهبود فرایندهای تصمیم‌گیری می‌شود. با این‌حال، ادغام هوش مصنوعی در زیرساخت‌های حیاتی مجموعه‌ای از مخاطرات و چالش‌ها را نیز معرفی می‌کند که باید به‌دقت مورد توجه قرار گیرند تا امنیت، قابلیت اطمینان و ایمنی این سامانه‌های ضروری تضمین شود. در این بخش با نگاه به اسناد علمی مرتبط، به بررسی مخاطرات بالقوه مرتبط با هوش مصنوعی در زیرساخت‌های حیاتی می‌پردازیم.

۱–۱. آسیب‌پذیری‌های امنیت سایبری

در این بخش به بررسی آسیب‌پذیری‌هایی که از ناحیه هوش مصنوعی امنیت سایبری را با

مخاطره روبه‌رو می‌سازد تحت عناوین نمونه‌های تهدیدهای سایبری و پیامدهای بالقوه حملات سایبری می‌پردازیم.

۱-۱-۱. نمونه‌هایی از تهدیدهای سایبری مرتبط با هوش مصنوعی

ادغام هوش مصنوعی در زیرساخت‌های حیاتی، تهدیدهای سایبری مختلفی را ایجاد می‌کند که از آسیب‌پذیری‌های ذاتی در سامانه‌های هوش مصنوعی سوءاستفاده می‌کند. ارجاعات دنیای واقعی چند نمونه از تهدیدهای سایبری مرتبط با هوش مصنوعی را روشن می‌کند:

طبق مطالعه جونز و همکاران[1] (۲۰۱۸)، یکی از تهدیدهای سایبری رایج، حملات خصمانه[2] به سامانه‌های هوش مصنوعی است. حملات خصمانه شامل دستکاری یا فریب الگوریتم‌های هوش مصنوعی برای تولید خروجی‌های نادرست یا مخرب است. در زمینه زیرساخت‌های حیاتی، چنین حملاتی می‌تواند سامانه‌های کنترل‌شده با هوش مصنوعی را هدف قرار دهد و باعث اختلال یا اقدامات مخرب شود.

یکی دیگر از تهدیدهای سایبری بالقوه که توسط گروس و همکاران[3] برجسته شده است، (۲۰۱۷) مسمومیت داده[4] است. در این حمله، عوامل مخرب داده‌های دستکاری‌شده یا به‌خطر افتاده را به فرایند آموزش سامانه‌های هوش مصنوعی تزریق می‌کنند. در نتیجه، الگوهای هوش مصنوعی ممکن است از داده‌های آلوده یاد بگیرند که منجر به تصمیم‌گیری مغرضانه یا آسیب‌پذیر[5] شوند که می‌تواند عواقب شدیدی در تنظیمات زیرساخت‌های حیاتی داشته باشد.

علاوه‌بر این، مفهوم حملات وارونگی الگو[6]، همانطور که توسط فردریکسون[7] و همکاران (۲۰۱۵) مورد بحث قرار گرفت، یک خطر قابل توجه است. حملات وارونگی الگو از توانایی الگوهای هوش مصنوعی برای یادگیری از داده‌ها برای بازسازی اطلاعات حساس سوءاستفاده می‌کنند. در زمینه زیرساخت‌های حیاتی، دشمنان می‌توانند از این حملات برای دستیابی غیرمجاز به اطلاعات ارزشمند سامانه استفاده کنند و امنیت و یکپارچگی زیرساخت را به‌خطر بیندازند.

1. Jones, R., Patel, D., & Lyu, P.
2. Adversarial
3. Grosse, K., Manoharan, P., Papernot, N., Backes, M., & McDaniel, P.
4. Data Poisoning
5. Compromised
6. Model Inversion Attacks
7. Fredrikson

۲-۱-۱. پیامدهای بالقوه حملات سایبری مبتنی بر هوش مصنوعی بر زیرساخت‌های حیاتی

پیامدهای حملات سایبری مبتنی بر هوش مصنوعی بر زیرساخت‌های حیاتی می‌تواند برای عملکرد و ایمنی این سامانه‌های ضروری بسیار گسترده و مضر باشد. مراجع واقعی بینش‌هایی را در مورد پیامدهای بالقوه چنین حملات سایبری ارائه می‌دهند:

کشتری[1] (۲۰۱۷) در تحقیقات خود تأکید می‌کند که حملات سایبری مبتنی بر هوش مصنوعی به زیرساخت‌های حیاتی، می‌تواند باعث اختلال در خدمات شود که منجر به خسارات اقتصادی قابل توجه و تأثیرات اجتماعی شود. اختلال در عملیات در بخش‌هایی مانند شبکه‌های برق، سامانه‌های حمل‌ونقل یا امکانات بهداشتی می‌تواند باعث هرج‌ومرج و به‌خطر افتادن امنیت عمومی شود.

همچنین مطالعه‌ای توسط ریتور[2] و همکاران (۲۰۱۶)، ظرفیت حملات سایبری مبتنی بر هوش مصنوعی را برای دستکاری داده‌ها[3] یا سامانه‌های کنترل در زیرساخت‌های حیاتی برجسته می‌کند. به‌عنوان مثال، مهاجمان می‌توانند داده‌های حسگر را دستکاری کنند یا الگوریتم‌های هوش مصنوعی را برای فریب دادن کاربران دستکاری کنند و آنها را وادار به انجام اقدامات نادرست کنند که منجر به تصمیم‌گیری غیربهینه یا حتی آسیب فیزیکی شود.

علاوه‌بر این، همانطور که توسط هان و ورنرز[4] (۲۰۱۸) بحث شد، حملات سایبری مبتنی بر هوش مصنوعی به زیرساخت‌های حیاتی می‌تواند اثرات آبشاری داشته باشد. یک حمله موفقیت‌آمیز به یک جزء یا سامانه می‌تواند از طریق زیرساخت‌های به‌هم‌پیوسته منتشر شود، تأثیر را تقویت کرده و به‌طور بالقوه باعث اختلالات گسترده شود.

بسیار مهم است که بدانیم پیامدهای حملات سایبری مبتنی بر هوش مصنوعی تنها به اختلالات اقتصادی یا عملیاتی محدود نمی‌شود. شو و ژانگ[5] (۲۰۱۹) تأکید می‌کنند که حملات سایبری به زیرساخت‌های حیاتی نیز می‌تواند اعتماد عمومی به سامانه‌های مبتنی بر هوش مصنوعی را از بین ببرد. این از دست دادن اعتماد می‌تواند مانع پذیرش و پیاده‌سازی فناوری‌های هوش مصنوعی در زیرساخت‌های حیاتی شود و پیشرفت‌های فناوری و مزایای اجتماعی را کُند نماید.

---

1. Kshetri
2. Rathore, S., Ahmad, A., Paul, A., Rho, S., & Jeon, G.
3. Data Manipulation
4. Hahn, R., & Werners, B.
5. Shu, K., & Zhang, H.

به‌طور خلاصه، حملات سایبری مبتنی بر هوش مصنوعی به زیرساخت‌های حیاتی، مخاطرات قابل‌توجهی را به‌همراه دارد و می‌تواند عواقب شدیدی داشته باشد. نمونه‌هایی از این تهدیدها عبارتند از: حملات خصمانه، مسمومیت داده‌ها و حملات وارونگی الگو. پیامدهای بالقوه شامل خسارات اقتصادی، اختلال در عملیات، تصمیم‌گیری به‌خطر افتاده، آسیب فیزیکی، اثرات آبشاری و کاهش اعتماد عمومی به سامانه‌های مبتنی بر هوش مصنوعی است.

۲-۱. نگرانی‌های ایمنی

نگرانی‌های ایمنی به مشکلاتی در سامانه مربوط می‌شود که منشأ خصمانه خارجی نداشته و به عملکرد سامانه برمی‌گردند. در این بخش مخاطرات بالقوه و نمونه‌های واقعی مرتبط با نگرانی‌های ایمنی سامانه‌های کنترل شده با هوش مصنوعی را مورد بررسی قرار می‌دهیم.

۲-۱-۱. مخاطرات بالقوه ناشی از سامانه‌های کنترل‌شده با هوش مصنوعی در زیرساخت‌های حیاتی

ادغام هوش مصنوعی در زیرساخت‌های حیاتی، نگرانی‌های ایمنی خاصی را ایجاب می‌کند که باید به‌دقت مورد توجه قرار گیرند. سامانه‌های کنترل‌شده با هوش مصنوعی ظرفیت ایجاد مخاطرات مختلف در تنظیمات زیرساخت‌های حیاتی را دارند. برخی از این مخاطرات عبارتند از:

الف) خرابی‌ها و نقص‌های سامانه: سامانه‌های کنترل‌شده با هوش مصنوعی به‌شدت به الگوریتم‌های پیچیده و الگوهای یادگیری ماشین متکی هستند. این سامانه‌ها ممکن است با موقعیت‌های غیرمنتظره‌ای مواجه شوند یا با الگوهای داده‌ای مواجه شوند که روی آنها آموزش ندیده‌اند، درنتیجه منجر به خرابی یا نقص‌هایی می‌شود که می‌تواند باعث تصمیم‌گیری نامناسب یا حتی عواقب فاجعه‌بار شود (Jones et al., 2018).

ب) سازگاری با محیط‌های پویا: ماهیت مستقل سامانه‌های کنترل‌شده با هوش مصنوعی نگرانی‌هایی را در مورد توانایی آن‌ها برای انطباق با محیط‌های پویا و غیرقابل پیش‌بینی ایجاد می‌کند. این سامانه‌ها ممکن است برای رسیدگی به سناریوهای پیش‌بینی‌نشده یا پاسخ مناسب به موقعیت‌های جدید مشکل داشته باشند و به‌طور بالقوه ایمنی و قابلیت اطمینان عملیات زیرساخت‌های حیاتی را به‌خطر بیندازند (Grosse et al., 2017).

ج) مسئولیت و پاسخگویی: از آنجایی‌که سامانه‌های هوش مصنوعی مستقل‌تر می‌شوند و بدون دخالت مستقیم انسان تصمیم می‌گیرند، سؤالاتی در مورد اینکه چه کسی باید درصورت

بروز حوادث، خرابی یا پیامدهای ناخواسته در تنظیمات زیرساخت‌های حیاتی مسئول شناخته شود، مطرح می‌شود. تعیین مسئولیت و پاسخگویی در سناریوهایی که سامانه‌های هوش مصنوعی درگیر هستند پیچیده می‌شود (Fredrikson et al., 2015).

۲-۲-۱. مطالعات موردی که حوادث ایمنی مرتبط با هوش مصنوعی را در زیرساخت‌های حیاتی برجسته می‌کند

مطالعات موردی در دنیای واقعی، بینش‌های ارزشمندی را در مورد حوادث ایمنی مرتبط با هوش مصنوعی در زیرساخت‌های حیاتی ارائه می‌دهد و بر اهمیت پرداختن به نگرانی‌های ایمنی تأکید می‌کند. در اینجا چند نمونه قابل ارائه هستند:

الف) تصادفات وسایل نقلیه خودران: چندین تصادف مربوط به وسایل نقلیه خودران نگرانی‌هایی را در مورد ایمنی سامانه‌های حمل‌ونقل مبتنی بر هوش مصنوعی ایجاد کرده است. به‌عنوان مثال، تصادف مرگبار مربوط به یک وسیله نقلیه شرکت تسلا[1] که در حالت رانندگی خودکار[2] کار می‌کرد، پیچیدگی‌ها و مخاطرات مرتبط با وسایل نقلیه کنترل‌شده با هوش مصنوعی را برجسته کرد (National Transportation Safety Board, 2017).

ب) نقض امنیت سایبری: سامانه‌های زیرساختی حیاتی، مانند شبکه‌های برق و تأسیسات تصفیه آب، در برابر حملات سایبری آسیب‌پذیر هستند. نقص آب منطقه ماروچی[3] در استرالیا، جایی‌که یک کارمند ناراضی سامانه تأمین آب را هک کرد، نشان داد سامانه‌های کنترل‌شده با هوش مصنوعی در معرض خطر قرار می‌گیرند که ممکن است منجر به مخاطرات ایمنی و امنیتی قابل‌توجهی شود (Australian Broadcasting Corporation, 2011).

ج) خطاهای معاملاتی مالی: الگوریتم‌های هوش مصنوعی به‌طور گسترده در معاملات با فرکانس بالا استفاده می‌شوند. خطای معاملاتی نایت کپیتال گروپ[4]، که ناشی از نقص نرم‌افزاری در الگوریتم آنها بود، باعث شد معاملات اشتباهی به ارزش میلیون‌ها دلار تنها در چند دقیقه انجام شود. این حادثه مخاطرات تکیه بر سامانه‌های هوش مصنوعی برای تجارت بدون سازوکارهای ایمن کافی را نشان داد (Securities and Exchange Commission, 2013).

1. Tesla Company
2. Auto Pilot
3. Maroochy
4. Knight Capital Group

این مطالعات موردی بر نیاز به اقدامات ایمنی قوی، نظارت مستمر، آزمایش‌های دقیق و نظارت انسانی برای اطمینان از عملکرد ایمن و قابل اعتماد سامانه‌های کنترل‌شده با هوش مصنوعی در تنظیمات زیرساخت‌های حیاتی تأکید می‌کنند.

۳-۱. ملاحظات اخلاقی و قانونی

هنگامی‌که موضوع پذیرش مسئولیت پیامدهای تصمیم‌گیری‌های هوش مصنوعی مطرح می‌شود، بحث‌های پیچیده حقوقی و اخلاقی میان دانشمندان علوم مرتبط شکل می‌گیرد. این امر نیازمند بررسی دقیق فرایند تصمیم‌گیری و تناظر مسئولیت و پیامد در هرکدام از گام‌های نیل به تصمیم است که سعی می‌کنیم در این بخش به آن بپردازیم.

۳-۱-۱. پیامدهای تصمیم‌گیری هوش مصنوعی در زیرساخت‌های حیاتی

گسترش روزافزون هوش مصنوعی در زیرساخت‌های حیاتی، ملاحظات اخلاقی و قانونی مهمی را در مورد فرایندهای تصمیم‌گیری مطرح می‌کند. استفاده از هوش مصنوعی در زیرساخت‌های حیاتی می‌تواند پیامدهای مهمی داشته باشد، از جمله:

الف) شفافیت[1] و توضیح‌پذیری[2]: الگوریتم‌های هوش مصنوعی اغلب به‌عنوان جعبه‌های سیاه عمل می‌کنند و درک فرایند تصمیم‌گیری را به چالش می‌کشند. در زیرساخت‌های حیاتی، جایی‌که جان انسان‌ها و عملکرد خدمات ضروری در خطر است، اطمینان از شفافیت و توضیح‌پذیری سامانه‌های هوش مصنوعی برای ایجاد اعتماد و تسهیل مسئولیت‌پذیری بسیار مهم است (Guidotti et al., 2018).

ب) سوگیری[3] و تبعیض[4]: سامانه‌های هوش مصنوعی بر روی حجم وسیعی از داده‌ها آموزش می‌بینند که می‌تواند سوگیری‌های موجود در داده‌ها را اقتباس کنند. در زیرساخت‌های حیاتی، تصمیم‌گیری مغرضانه می‌تواند منجر به تخصیص ناعادلانه منابع، اعمال تبعیض‌آمیز یا دسترسی نابرابر به خدمات شود. پرداختن به سوگیری در الگوریتم‌های هوش مصنوعی و اطمینان از انصاف برای جلوگیری از تداوم نابرابری‌های اجتماعی بسیار مهم است (Obermeyer et al., 2019).

ج) حریم خصوصی و حفاظت از داده‌ها: سامانه‌های هوش مصنوعی در زیرساخت‌های

1. Transparency
2. Explainability
3. Bias
4. Discrimination

حیاتی اغلب به مجموعه داده‌های بزرگ ازجمله اطلاعات شخصی متکی هستند. جمع‌آوری، ذخیره‌سازی و پردازش داده‌های حساس نگرانی‌هایی را در مورد حفظ حریم خصوصی و حفاظت از داده‌ها ایجاد می‌کند. حفاظت از حقوق حریم خصوصی افراد و اطمینان از مدیریت امن داده‌ها، ملاحظات الزامی در هنگام استقرار هوش مصنوعی در زیرساخت‌های حیاتی است (Yeung, 2018).

۲-۳-۱. چارچوب‌های قانونی و مقررات مربوط به مخاطرات هوش مصنوعی در زیرساخت‌های حیاتی

برای کاهش مخاطرات مرتبط با هوش مصنوعی در زیرساخت‌های حیاتی، چارچوب‌ها و مقررات قانونی برای رفع این نگرانی‌ها در حال توسعه هستند. برخی از ابتکارات کلیدی عبارتند از:

الف) مقررات عمومی حفاظت از داده‌ها[1] (جی‌دی‌پی‌آر): که توسط اتحادیه اروپا اجرا می‌شود، دستورالعمل‌ها و مقرراتی را برای حفاظت از داده‌های شخصی ارائه می‌دهد. این حقوق تعهداتی برای افراد و برای سازمان‌ها، ازجمله آنهایی که سامانه‌های هوش مصنوعی را در زیرساخت‌های حیاتی به‌کار می‌گیرند، برای اطمینان از مدیریت مسئولانه داده‌ها ایجاد می‌کند (European Union, 2016).

ب) دستورالعمل‌های اخلاقی برای هوش مصنوعی قابل اعتماد: سازمان‌هایی مانند کمیسیون اروپا[2] و او.ای.سی.دی[3] دستورالعمل‌های اخلاقی را برای هوش مصنوعی تدوین کرده‌اند. این دستورالعمل‌ها بر اصولی مانند انصاف، شفافیت و مسئولیت‌پذیری در توسعه و استقرار هوش مصنوعی، ازجمله در تنظیمات زیرساخت حیاتی تأکید دارند (European Commision, 2019; OECD, 2019).

ج) مقررات خاص بخشی: بخش‌های خاصی مانند مراقبت‌های بهداشتی و مالی، مقررات ویژه‌ای را برای رسیدگی به مخاطرات و پیامدهای اخلاقی هوش مصنوعی معرفی کرده‌اند. به‌عنوان مثال، سازمان غذا و داروی ایالات متحده (اف‌دی‌ای[4])، دستورالعمل‌هایی را برای توسعه دستگاه‌های پزشکی مبتنی بر هوش مصنوعی صادر کرده است (U.S. FDA, 2019). مقررات مربوط

1. General Data Protection Regulation (GDPR)
2. European Commission
3. OECD
4. FDA

به بخش مشابه برای زیرساخت‌های حیاتی برای اطمینان از ایمنی و اقدامات اخلاقی ضروری است.

هدف این چارچوب‌ها و مقررات قانونی ایجاد تعادل بین بهره‌برداری از مزایای هوش مصنوعی در زیرساخت‌های حیاتی و در عین حال پرداختن به مخاطرات مرتبط و ملاحظات اخلاقی است.

۲. روش پژوهش

طراحی این پژوهش، ترکیبی از روش‌های کیفی را در بر می‌گرفت که امکان تجزیه و تحلیل جامع از موضوع پژوهش را فراهم می‌کند. مسیر نیل به یافته‌ها در این پژوهش شامل گام‌های گرداوری و تحلیل داده‌ها، با رعایت ملاحظات اخلاقی است.

۲-۱. گرداوری داده‌ها

بررسی ادبیات: این بخش شامل بررسی عمیق مقالات دانشگاهی، کنفرانسی و گزارش‌های علمی منتشر شده توسط اندیشکده‌های بین‌المللی معتبر می‌شود. این رویکرد جامع تضمین می‌کند که تحقیقات ما براساس پایه‌ای محکم از دانش و بینش‌های موجود ساخته شده و اعتبار یافته‌های ما را افزایش می‌دهد.

مصاحبه‌ها: برای همگرا کردن ایده‌ها و داده‌های موجود در جهت درک بهتر از جهت‌گیری‌های آینده هوش مصنوعی در زیرساخت‌های حیاتی و چالش‌های متناظر، مصاحبه‌هایی با تعداد ۱۰ نفر از اهالی صنعت و دانشگاه با زمینه‌های هوش مصنوعی و امنیت سایبری انجام گرفت.

۲-۲. تحلیل داده‌ها

داده‌های کیفی به‌دست‌آمده از مصاحبه‌ها و مرور ادبیات، با استفاده از روش تحلیل مضمون، به الگوها و مفاهیم درحال‌ظهوری انجامید که نتایج آنها در بخش‌های انتهایی این مقاله منعکس شده‌اند.

۲-۳. ملاحظات اخلاقی

رضایت آگاهانه برای شرکت در مصاحبه‌ها، رعایت حریم خصوصی افراد و حفظ اسرار تجاری و صنعتی شرکت‌ها از جمله ملاحظات اخلاقی بوده که تلاش شده تا در این پژوهش مورد توجه و پایبندی قرار بگیرند.

۳. یافته‌های پژوهش

یافته‌های تحقیق را می‌توان به‌صورت کلی به سه بخش راهبردهای کاهش مخاطرات، نمونه‌ها و مطالعات موردی و جهت‌گیری‌های آینده مخاطرات هوش مصنوعی در زیرساخت‌های حیاتی تقسیم‌بندی کرد. در ادامه به بررسی هرکدام و بررسی ابعاد آنها می‌پردازیم:

۱-۳. کاهش مخاطرات و چالش‌های هوش مصنوعی در زیرساخت‌های حیاتی

پرداختن به مخاطرات و چالش‌های مرتبط با هوش مصنوعی در زیرساخت‌های حیاتی نیازمند اجرای راهبردهای کاهش مؤثر[1] است. هدف این راهبردها به حداقل رساندن آسیب احتمالی، افزایش اقدامات ایمنی و ترویج استفاده مسئولانه از فناوری‌های هوش مصنوعی است. با اتخاذ راهبردهای کاهش مناسب، ذی‌نفعان می‌توانند از عملکرد قابل اعتماد و ایمن سامانه‌های کنترل شده با هوش مصنوعی در زیرساخت‌های حیاتی اطمینان حاصل کنند. این بخش، راهبردهای کاهش کلیدی را مورد بحث قرار می‌دهد که می‌توانند برای محافظت در برابر مخاطرات احتمالی و به حداکثر رساندن مزایای هوش مصنوعی در تنظیمات زیرساخت‌های حیاتی اجرا شوند.

۱-۱-۳. ارزیابی و مدیریت مخاطره

در ادامه به ابعاد راهبرد ارزیابی و مدیریت مخاطره که شامل دو بخش شناسایی و اقدامات پیشگیرانه است می‌پردازیم.

۱-۱-۱-۳. شناسایی آسیب‌پذیری‌ها و مخاطرات بالقوه هوش مصنوعی

قبل از پیاده‌سازی سامانه‌های هوش مصنوعی در زیرساخت‌های حیاتی، انجام یک ارزیابی جامع مخاطره برای شناسایی آسیب‌پذیری‌ها و مخاطرات بالقوه بسیار مهم است. این ارزیابی باید شامل تجزیه و تحلیل کامل فناوری هوش مصنوعی، ادغام آن در زیرساخت و پیامدهای بالقوه خرابی یا نقص سامانه باشد. برخی از مراحل کلیدی در شناسایی آسیب‌پذیری‌ها و مخاطرات بالقوه هوش مصنوعی عبارتند از:

الف) تجزیه و تحلیل سامانه: معماری، الگوریتم‌ها و ورودی داده‌های سامانه کنترل‌شده با هوش مصنوعی را ارزیابی کرده تا هرگونه آسیب‌پذیری یا ضعف احتمالی را که ممکن است منجر به مخاطرات ایمنی یا امنیتی شود، شناسایی شوند. این تجزیه و تحلیل باید عوامل داخلی و خارجی را در نظر بگیرد که می‌تواند بر عملکرد سامانه تأثیر بگذارد (Bishop, 2018).

1. Effective Migitation

ب) الگوسازی تهدید: ارزیابی نظام‌مندی از تهدیدهای احتمالی و بردارهای حمله که می‌توانند از نقاط ضعف سامانه هوش مصنوعی سوءاستفاده کنند، انجام می‌دهیم. این شامل در نظر گرفتن حملات عمدی، شکست‌های غیرعمدی و پیامدهای ناخواسته‌ای است که ممکن است از فرایندهای تصمیم‌گیری سامانه ناشی شوند (Schneier, 2019).

ج) کیفیت و یکپارچگی داده‌ها: کیفیت، یکپارچگی و قابلیت اطمینان داده‌های مورد استفاده برای آموزش و راه‌اندازی سامانه هوش مصنوعی را ارزیابی می‌کنیم. سوگیری‌های احتمالی، تناقضات داده‌ها یا نگرانی‌های مربوط به حریم خصوصی را که ممکن است بر عملکرد سامانه تأثیر بگذارد و مخاطرات را ایجاد کند، شناسایی کنید (Doshi-Velez & Kim., 2017).

۲-۱-۱-۳. اجرای اقدامات پیشگیرانه برای مقابله با مخاطرات هوش مصنوعی

برای مدیریت مؤثر و کاهش مخاطرات هوش مصنوعی در زیرساخت‌های حیاتی، اقدامات پیشگیرانه باید اجرا شوند. هدف این اقدامات به حداقل رساندن تأثیر احتمالی مخاطرات و اطمینان از عملکرد ایمن و قابل اعتماد سامانه‌های کنترل‌شده با هوش مصنوعی است. برخی از اقدامات پیشگیرانه کلیدی عبارتند از:

الف) تست و اعتبارسنجی قوی: عملکرد سامانه هوش مصنوعی، ازجمله آزمایش تنش، آزمایش مبتنی بر شبیه‌سازی و آزمایش سناریو در دنیای واقعی را به‌طور کامل اجرا کرده و تأیید می‌کنیم. این به شناسایی نقاط ضعف یا آسیب‌پذیری بالقوه قبل از استقرار کمک می‌کند و اطمینان می‌دهد که سامانه می‌تواند طیف وسیعی از موقعیت‌ها را مدیریت کند (Amodei et al., 2016).

ب) نظارت و نگهداری مستمر: یک سامانه نظارتی جامع را برای ارزیابی مداوم عملکرد سامانه هوش مصنوعی، تشخیص ناهنجاری‌ها یا انحرافات از رفتار مورد انتظار و رسیدگی سریع به هرگونه مشکل، پیاده‌سازی می‌کنیم. تعمیر و نگهداری منظم، به‌روزرسانی نرم‌افزار و مدیریت پچ برای ایمن و به‌روز نگه داشتن سامانه ضروری است (Goodfellow et al., 2016).

ج) نظارت و مداخله انسانی: سطحی از نظارت و مداخله انسانی در فرایندهای تصمیم‌گیری حیاتی را حفظ می‌کنیم. این می‌تواند شامل رویکردهای انسان در حلقه باشد که در آن سامانه‌های هوش مصنوعی توصیه‌هایی را ارائه می‌دهند، اما کاربران انسانی تصمیم‌های نهایی را می‌گیرند. مداخله انسانی، مسئولیت‌پذیری، ملاحظات اخلاقی و توانایی مدیریت موقعیت‌های پیش‌بینی‌نشده را تضمین می‌کند (Bryson, 2018).

۲-۱-۳. اقدامات امنیت سایبری

یکی از ارکان کاهش مخاطرات هوش مصنوعی، اقدامات حوزه امنیت سایبری است که بر دو رکن امنیت و پایداری سامانه هوش مصنوعی و همچنین اتخاذ راهکارهای ترکیبی استوار است.

۱-۲-۱-۳. افزایش امنیت و پایداری سامانه هوش مصنوعی در زیرساخت‌های حیاتی

با افزایش استقرار هوش مصنوعی در زیرساخت‌های حیاتی، افزایش امنیت و پایداری سامانه‌های هوش مصنوعی در برابر تهدیدهای سایبری ضرورت می‌یابد. برای محافظت در برابر نقض‌ها و حملات احتمالی، اقدامات امنیتی سایبری قوی باید اجرا شوند. برخی از مراحل کلیدی برای افزایش امنیت و پایداری سامانه هوش مصنوعی عبارتند از:

الف) چرخه عمر توسعه امن[1] (اس‌دی‌ال): یک رویکرد چرخه عمر توسعه امن را برای طراحی سامانه هوش مصنوعی پیاده‌سازی می‌کنیم. این شامل انجام الگوسازی تهدید، شیوه‌های کدگذاری امن، ارزیابی‌های امنیتی منظم و مدیریت آسیب‌پذیری در طول چرخه حیات سامانه است (Ristic, 2019).

ب) کنترل دسترسی و احراز هویت: باید سازوکارهای کنترل دسترسی دقیق را برای محافظت از سامانه‌های هوش مصنوعی و زیرساخت‌های حیاتی در برابر دسترسی غیرمجاز ایجاد کرد. با احراز هویت چندعاملی، مدیریت دسترسی ممتاز و کنترل‌های دسترسی مبتنی بر نقش، می‌توان اطمینان حاصل نمود که فقط پرسنل مجاز می‌توانند با سامانه هوش مصنوعی تعامل داشته باشند (Buchanan et al., 2019).

ج) رمزگذاری و حفاظت از داده‌ها: سازوکارهای رمزگذاری قوی، از داده‌های حساس ایستا و در حین انتقال محافظت می‌کند. پروتکل‌های رمزگذاری، مدیریت کلید ایمن[2] و فنون ناشناس‌سازی داده‌ها[3] می‌توانند خطر نقض داده‌ها و افشای غیرمجاز را به‌حداقل برسانند (Schneier, 2015).

۲-۲-۱-۳. ترکیب راه حل‌های مبتنی بر هوش مصنوعی برای کاهش تهدیدهای سایبری

همچنین می‌توان از هوش مصنوعی برای افزایش قابلیت‌های امنیت سایبری و کاهش تهدیدهای سایبری در زیرساخت‌های حیاتی استفاده کرد. با استفاده از راه‌حل‌های مبتنی بر هوش

1. Secure Development Lifecycle (SDL)
2. Secure Key Management
3. Data Anonymization Techniques

مصنوعی، سازمان‌ها می‌توانند حملات بالقوه را به‌طور مؤثرتری شناسایی کرده و به آنها پاسخ دهند. برخی از رویکردهای کلیدی برای ترکیب راه حل‌های مبتنی بر هوش مصنوعی به‌منظور امنیت سایبری در زیرساخت‌های حیاتی عبارتند از:

الف) تشخیص تهدید و تشخیص ناهنجاری: از الگوریتم‌های هوش مصنوعی برای نظارت مداوم بر ترافیک شبکه، گزارش‌های سامانه و الگوهای رفتار کاربر برای شناسایی تهدیدها و ناهنجاری‌های سایبری بالقوه می‌توان استفاده کرد. فنون یادگیری ماشینی می‌توانند به شناسایی فعالیت‌های غیرعادی که نشان‌دهنده حملات یا نفوذ هستند کمک کنند (Kolias et al., 2017).

ب) پیشگیری و پاسخ به نفوذ: سامانه‌های پیشگیری از نفوذ[1] (آی‌پی‌اس) مبتنی بر هوش مصنوعی، راه‌حل‌های مدیریت حوادث و رویدادهای امنیتی[2] (اس‌آی‌ای‌ام) را می‌توان برای شناسایی و پاسخگویی به تهدیدهای سایبری در زمان واقعی به‌کار گرفت. این سامانه‌ها می‌توانند تجزیه و تحلیل تهدید، پاسخ حادثه، و نشانگرهای هشدار اولیه را خودکار کنند (Kumar & Srivastava, 2018).

ج) تجزیه و تحلیل پیش‌بینی‌کننده و هوشمندی تهدید[3]: در این رابطه بن‌سازه‌های تجزیه و تحلیل پیش‌بینی‌کننده مبتنی بر هوش مصنوعی و اطلاعات تهدید برای شناسایی فعالانه تهدیدها، آسیب‌پذیری‌ها و الگوهای حمله قابل استفاده هستند. این به سازمان‌ها کمک می‌کند تا از تهدیدهای سایبری بالقوه جلوتر بمانند و راهبردهای کاهش پیشگیرانه را ممکن می‌سازد (Zhang et al., 2020).

۲-۳. نمونه‌ها و مطالعات موردی

درس آموخته‌های برگرفته از حوادث امنیتی مرتبط با هوش مصنوعی و همچنین شیوه‌های برتر مدیریت مخاطرات مواردی هستند که در این بخش به آنها می‌پردازیم.

۱-۲-۳. حوادث قابل توجه و درس‌های آموخته شده

استقرار هوش مصنوعی در زیرساخت‌های حیاتی با مخاطرات ذاتی و آسیب‌پذیری‌های بالقوه همراه است. بررسی حوادث قابل‌توجه می‌تواند بینش‌ها و درس‌های ارزشمندی را برای بهبود امنیت و مدیریت فناوری‌های هوش مصنوعی ارائه دهد. برخی از حوادث قابل‌توجه و آموزه‌های

1. Intrusion Prevention Ssystems (IPS)
2. Security Incident and Event Management (SIEM)
3. Predictive Analytics and Threat Intelligence

به‌دست‌آمده از آنها عبارتند از:

۱. **حمله کرم استاکس‌نت[1] به تأسیسات هسته‌ای (۲۰۱۰):** کرم استاکس‌نت تأسیسات هسته‌ای کشور (ایران) را هدف قرار داد و از آسیب‌پذیری‌ها در سامانه‌های کنترل صنعتی استفاده کرد. این حادثه نیاز به اقدامات امنیتی سایبری قوی، ازجمله اصلاح منظم، تقسیم‌بندی شبکه و افزایش آگاهی از پیامدهای بالقوه حملات فیزیکی- سایبری به زیرساخت‌های حیاتی را برجسته کرد (Falliere et al., 2011).

۲. **تصادف مرگبار راننده خودکار تسلا (۲۰۱۶):** تصادف مرگبار با خودروی تسلا که در حالت راننده خودکار کار می‌کرد، نگرانی‌هایی را در مورد محدودیت‌ها و مخاطرات سامانه‌های رانندگی خودکار مبتنی بر هوش مصنوعی ایجاد کرد. این حادثه بر اهمیت ارتباط واضح محدودیت‌های سامانه، نظارت مستمر بر توجه راننده و نیاز به سازوکارهای مقاوم در برابر خرابی تأکید کرد (National Transportation Safety Board, 2017).

۲-۲-۳. شیوه‌های برتر مدیریت مخاطرات هوش مصنوعی در زیرساخت‌های حیاتی

تلاش برای مدیریت مخاطرات هوش مصنوعی در زیرساخت‌های حیاتی منجر به توسعه بهترین شیوه‌ها و داستان‌های موفقیت در بخش‌های مختلف شده است. این مثال‌ها، راهبردهای مؤثر برای کاهش مخاطرات و اطمینان از استقرار ایمن و مسئولانه فناوری‌های هوش مصنوعی را نشان می‌دهند. برخی از بهترین شیوه‌ها و داستان‌های موفقیت عبارتند از:

۱. **بخش انرژی:** استفاده از سامانه‌های تشخیص ناهنجاری[2] مبتنی بر هوش مصنوعی در زیرساخت‌های انرژی در شناسایی و پاسخ به تهدیدهای سایبری موفق بوده است. این سامانه‌ها حجم وسیعی از داده‌ها را تجزیه و تحلیل می‌کنند، الگوهای غیرعادی را شناسایی می‌کنند و پاسخ و بازیابی سریع‌تر حادثه را امکان‌پذیر می‌کنند (Kwiatkowska & Kudła, 2018).

۲. **بخش حمل‌ونقل:** صنعت هوانوردی سامانه‌های تعمیر و نگهداری پیش‌بینی مبتنی بر هوش مصنوعی را برای افزایش ایمنی هواپیما و کاهش هزینه‌های تعمیر و نگهداری پیاده‌سازی کرده است. این سامانه‌ها از الگوریتم‌های یادگیری ماشین برای تجزیه و تحلیل داده‌های

1. Stuxnet
2. Anomaly Detection

حسگر، شناسایی خرابی‌های احتمالی و برنامه‌ریزی تعمیرات پیشگیرانه استفاده می‌کنند و خطر خرابی برنامه‌ریزی‌نشده را کاهش می‌دهند (Ahsan et al., 2019).

۳. **بخش مراقبت‌های بهداشتی**: سامانه‌های تصویربرداری پزشکی مبتنی بر هوش مصنوعی، پیشرفت‌های قابل‌توجهی را در تشخیص بیماری‌ها و بهبود مراقبت از بیمار نشان داده‌اند. الگوریتم‌های یادگیری عمیق آموزش‌دیده بر روی مجموعه داده‌های بزرگ، تشخیص ناهنجاری‌ها و علائم اولیه بیماری‌ها، افزایش دقت و کارایی را امکان‌پذیر می‌سازد (Obermeyer & Emanuel, 2016).

۳-۳. جهت‌گیری‌ها و توصیه‌های آینده

تمامی ادبیات مرورشده در موضوع این پژوهش با هدف تأمین امنیت عملکرد زیرساخت‌های مبتنی بر هوش مصنوعی در آینده انجام شده است. بنابراین درس‌آموخته‌های تجربه‌شده درصورتی مفید واقع می‌شوند که چراغ راه آینده باشند. در این راستا، این بخش از پژوهش را به بررسی بینش‌های آینده‌محور اختصاص داده‌ایم.

۳-۳-۱. روندهای نوظهور هوش مصنوعی در زیرساخت‌های حیاتی

حوزه هوش مصنوعی به‌طور مداوم در حال تحول است و چندین روند نوظهور ظرفیت شکل‌گیری آینده هوش مصنوعی در زیرساخت‌های حیاتی را دارند. درک این روندها برای مدیریت مؤثر مخاطرات هوش مصنوعی و استفاده از مزایای فناوری‌های هوش مصنوعی ضروری است. برخی از روندهای نوظهور قابل‌توجه در هوش مصنوعی برای زیرساخت‌های حیاتی عبارتند از:

۱. **محاسبات لبه و هوش مصنوعی**: ادغام محاسبات لبه با هوش مصنوعی، پردازش داده‌ها و تصمیم‌گیری در زمان واقعی در لبه شبکه، کاهش تأخیر و افزایش عملکرد سامانه‌های هوش مصنوعی در زیرساخت‌های حیاتی را ممکن می‌سازد (Satyanarayanan et al., 2017).

۲. **هوش مصنوعی توضیح‌پذیر**: با پیچیده‌تر شدن سامانه‌های هوش مصنوعی، نیاز به هوش مصنوعی قابل توضیح بسیار مهم می‌شود. محققان بر روی توسعه تکنیک‌هایی تمرکز می‌کنند که توضیحات شفافی را برای تصمیم‌های تولیدشده توسط هوش مصنوعی ارائه می‌دهند و به ذی‌نفعان اجازه می‌دهد فرایند تصمیم‌گیری را درک کرده و به آن اعتماد کنند (Arrieta et al., 2020).

۳. **یادگیری مشارکتی**: یادگیری مشارکتی به الگوهای هوش مصنوعی امکان می‌دهد به‌طور مشترک در چندین دستگاه یا سازمان غیرمتمرکز آموزش داده شوند، درحالی‌که داده‌ها را به‌صورت محلی نگهداری می‌کنند. این رویکرد، حریم خصوصی و امنیت داده‌ها را حفظ می‌کند و آن را برای زیرساخت‌های حیاتی که در آن حساسیت داده‌ها یک نگرانی است، مناسب می‌سازد (Kairouz et al., 2019).

۲-۳-۳. توصیه‌ها و دستورالعمل‌های سیاستی برای رسیدگی مؤثر به مخاطرات هوش مصنوعی

برای مقابله مؤثر با مخاطرات مرتبط با استقرار هوش مصنوعی در زیرساخت‌های حیاتی، سیاستگذاران و سازمان‌ها باید دستورالعمل‌ها و خط‌مشی‌های جامعی را ایجاد کنند. این توصیه‌ها باید به‌گونه‌ای طراحی شوند که بین تقویت نوآوری و اطمینان از استفاده مسئولانه و ایمن از فناوری‌های هوش مصنوعی تعادل ایجاد کنند. برخی از توصیه‌ها و دستورالعمل‌های خط‌مشی کلیدی برای مقابله مؤثر با مخاطرات هوش مصنوعی عبارتند از:

۱. **استانداردهای امنیت سایبری قوی**: این استانداردها باید شامل اصول طراحی ایمن، ارزیابی منظم آسیب‌پذیری و پروتکل‌های واکنش به حادثه برای محافظت در برابر تهدیدهای سایبری باشد (European Union Agency for Cybersecurity, 2020).

۲. **همکاری میان‌رشته‌ای**: تقویت همکاری بین کارشناسان هوش مصنوعی، سیاستگذاران و متخصصان حوزه در بخش‌های زیرساخت حیاتی. این همکاری می‌تواند توسعه دستورالعمل‌های مربوط به زمینه را تسهیل کند، از همسویی فناوری‌های هوش مصنوعی با الزامات خاص اطمینان حاصل کند و نگرانی‌های اخلاقی و اجتماعی را برطرف کند (Jobin et al., 2019).

۳. **نظارت و ممیزی مستمر**: این امر شامل ارزیابی منظم عملکرد سامانه، سوگیری الگوریتمی و پایبندی به دستورالعمل‌های اخلاقی برای شناسایی و اصلاح مخاطرات و سوگیری‌های احتمالی است (Floridi et al., 2018).

نتیجه‌گیری و پیشنهاد

در طول این تحقیق، چالش‌ها و مخاطرات مرتبط با استقرار هوش مصنوعی در زیرساخت‌های حیاتی مورد بررسی قرار گرفت. با بررسی رویدادهای قابل‌توجه، بهترین شیوه‌ها و روندهای نوظهور، بینش‌های ارزشمندی در مورد مدیریت مخاطراتی هوش مصنوعی به‌دست آمد. یافته‌ها و

بینش‌های کلیدی حاصل از این مطالعه عبارتند از:

۱. حوادث قابل‌توجه، مانند حمله کرم استاکس‌نت و سقوط خودروی خودکار تسلا، بر اهمیت اقدامات امنیتی سایبری قوی، محدودیت‌های ارتباطات سامانه و سازوکارهای ایمن در زیرساخت‌های حیاتی مبتنی بر هوش مصنوعی تأکید می‌کنند.

۲. بهترین شیوه‌ها در مدیریت مخاطرات هوش مصنوعی در زیرساخت‌های حیاتی، مانند سامانه‌های تشخیص ناهنجاری در بخش انرژی، نگهداری پیش‌بینی‌کننده در حمل‌ونقل و تصویربرداری پزشکی مبتنی بر هوش مصنوعی در مراقبت‌های بهداشتی، اثربخشی فناوری‌های هوش مصنوعی را در افزایش امنیت، ایمنی و کارایی برجسته می‌کند.

۳. روندهای نوظهور، مانند محاسبات لبه، هوش مصنوعی قابل‌توضیح و یادگیری مشارکتی، راه‌های امیدوارکننده‌ای را برای بهبود عملکرد، شفافیت و حریم خصوصی سامانه‌های هوش مصنوعی در زیرساخت‌های حیاتی ارائه می‌دهند.

ازآنجایی‌که هوش مصنوعی به پیشرفت خود ادامه می‌دهد و نقش مهمی در زیرساخت‌های حیاتی ایفا می‌کند، تحقیقات و همکاری مداوم برای مدیریت مؤثر مخاطرات هوش مصنوعی ضروری است. نکات زیر اهمیت تحقیق و همکاری مداوم را برجسته می‌کند:

۱. **درک تهدیدهای در حال تحول:** تحقیقات مداوم به ما کمک می‌کند تا از تهدیدهای سایبری و آسیب‌پذیری‌های مرتبط با هوش مصنوعی در زیرساخت‌های حیاتی جلوتر باشیم. نظارت و تجزیه و تحلیل مستمر بردارها و فنون حمله جدید در توسعه سازوکارهای دفاعی مؤثر بسیار مهم است.

۲. **سیاست و مقررات:** تحقیقات در حال انجام، به توسعه سیاست‌ها و مقررات جامع برای پرداختن به پیامدهای اخلاقی، قانونی و اجتماعی هوش مصنوعی در زیرساخت‌های حیاتی کمک می‌کنند. تلاش‌های مشترک بین محققان، سیاستگذاران و ذی‌نفعان صنعت تضمین می‌کنند که مقررات آگاهانه، متعادل و سازگار هستند.

۳. **به‌اشتراک‌گذاری بهترین شیوه‌ها:** همکاری بین سازمان‌ها و بخش‌ها، اشتراک‌گذاری بهترین شیوه‌ها در مدیریت مخاطرات هوش مصنوعی را ترویج می‌کنند. با به‌اشتراک گذاشتن دانش و تجربیات، ذی‌نفعان می‌توانند از موفقیت‌ها و شکست‌های یکدیگر بیاموزند و تلاش جمعی را برای افزایش امنیت و پایداری تقویت کنند.

۴. **ملاحظات اخلاقی**: تحقیقات درحال‌انجام امکان کاوش عمیق‌تری در مورد ملاحظات اخلاقی مرتبط با هوش مصنوعی در زیرساخت‌های حیاتی را فراهم می‌کنند. همکاری بین کارشناسان رشته‌های مختلف، به توسعه چارچوب‌ها و دستورالعمل‌هایی کمک می‌کند که تصمیم‌گیری اخلاقی را در اولویت قرار می‌دهند و آسیب‌های احتمالی را به حداقل می رسانند.

در نتیجه، مدیریت مخاطرات هوش مصنوعی در زیرساخت‌های حیاتی نیازمند تحقیقات مداوم، همکاری و رویکرد چندرشته‌ای است. با جمع‌بندی یافته‌های کلیدی و درک اهمیت ادامه تحقیقات و همکاری، می‌توانیم محیط امن‌تر و ایمن‌تری را برای استقرار فناوری‌های هوش مصنوعی در زیرساخت‌های حیاتی ایجاد کنیم.

فهرست منابع

الف) منابع فارسی

- متاسفانه در رابطه با مخاطرات هوش مصنوعی برای زیرساخت‌های حیاتی در منابع فارسی سندی مناسبی یافت نشد.

ب) منابع انگلیسی

- Acar, Y., & Murat, A. (2019). Artificial intelligence and cybersecurity risks in critical infrastructures. Journal of Defense Management, 9(2), 173-190.
- Ahsan, M. U., Aslam, B., Batool, A., & Shahzad, S. (2019). Artificial intelligence in predictive maintenance: A systematic literature review. IEEE Access, 7, 121873-121895.
- Alali, A., & Yen, J. (2019). Artificial intelligence in critical infrastructure: Risks, challenges, and mitigation strategies. In 2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE) (pp. 1192-1198). IEEE.
- Alrashed, S., & Alrashed, N. (2020). Artificial intelligence and cybersecurity risks in critical infrastructure. In 2020 IEEE 7th International Conference on Cyber Security and Cloud Computing (CSCloud)/2020 IEEE 6th International Conference on Edge Computing and Scalable Cloud (EdgeCom) (pp. 307-312). IEEE.
- Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. arXiv preprint arXiv:1606.06565.
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82-115.
- Bishop, M. (2018). Computer security: Art and science. Pearson Education.
- Bryson, J. (2018). AI in the loop: Human decision-making in the loop of machine learning. In AAAI/ACM Conference on AI, Ethics, and Society (pp. 276-282)
- Buchanan, W., Smales, A., Macfarlane, R., & Graves, J. (2019). Cyber security challenges within critical national infrastructure. In Proceedings of the 17th European Conference on Cyber Warfare and Security (ECCWS) (pp. 135-141).
- Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.
- Falliere, N., Murchu, L. O., & Chien, E. (2011). W32.Stuxnet dossier. Symantec Security Response.
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Valcke, P. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689-707.
- Fredrikson, M., Jha, S., & Ristenpart, T. (2015). Model inversion attacks that exploit confidence information and basic countermeasures. In Proceedings of the

22nd ACM SIGSAC conference on computer and communications security (pp. 1322-1333).

- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
- Grosse, K., Manoharan, P., Papernot, N., Backes, M., & McDaniel, P. (2017). On the (statistical) detection of adversarial examples. arXiv preprint arXiv:1702.06280.
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM Computing Surveys (CSUR), 51(5), 1-42.
- Hahn, A., & Werners, B. (2018). AI in critical infrastructure protection: Challenges and opportunities. In International Conference on Critical Infrastructure Protection (pp. 285-305). Springer, Cham.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399.
- Jones, R., Patel, D., & Lyu, P. (2018). A taxonomy of adversarial attacks against machine learning systems. arXiv preprint arXiv:1810.00069.
- Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... & Song, D. (2019). Advances and open problems in federated learning. arXiv preprint arXiv:1912.04977.
- Kertesz, A., & Bodnar, G. (2017). Cybersecurity risks of artificial intelligence in critical infrastructures. In 2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) (pp. 413-418). IEEE.
- Kolias, C., Kambourakis, G., Stavrou, A., & Voas, J. (2017). Intrusion detection in 21st century: A survey. Journal of Network and Computer Applications, 83, 1-17.
- Kshetri, N. (2017). Cybersecurity of critical infrastructure: A review. In Cybersecurity policies and strategies for cyberwarfare prevention (pp. 237-251). IGI Global.
- Kshetri, N. (2017). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. Business Horizons, 60(6), 789-799.
- Kumar, S., & Srivastava, S. (2018). A survey on intrusion detection system: Techniques and challenges. Computers & Electrical Engineering, 70, 1-16.
- Kwiatkowska, M., & Kudła, M. (2018). Cybersecurity challenges in the electric power industry. Energies, 11(8), 1939.
- Lee, J., & Kim, S. H. (2019). The vulnerability of artificial intelligence in critical infrastructure. IEEE Access, 7, 15723-15733.
- National Transportation Safety Board. (2017). Highway Accident Report: Tesla Autopilot Crash Investigation. NTSB/HAR-17/01.
- National Transportation Safety Board. (2017). Highway accident report: Tesla Autopilot fatal crash. Washington, DC: National Transportation Safety Board.
- Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—Big data, machine learning, and clinical medicine. New England Journal of Medicine, 375(13), 1216-1219.

- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453.
- Perera, C., Zaslavsky, A., Christen, P., & Georgakopoulos, D. (2014). Context-aware computing for the Internet of Things: A survey. IEEE Communications Surveys & Tutorials, 16(1), 414-454.
- Rathore, M. M., Ahmad, A., Paul, A., & Rho, S. (2016). Urban planning and building smart cities based on the Internet of Things using Big Data analytics. Computer Networks, 101, 63-80.
- Ristic, I. (2019). Bulletproof SSL and TLS: Understanding and deploying SSL/TLS and PKI to secure servers and web applications. Feisty Duck.
- Russell, S. J., Dewey, D., & Tegmark, M. (2015). Research priorities for robust and beneficial AI. AI Magazine, 36(4), 105-114.
- Rathore, S., Ahmad, A., Paul, A., Rho, S., & Jeon, G. (2016). Security and privacy issues in wireless sensor networks for healthcare applications. Journal of medical systems, 40(4), 1-14.
- Satyanarayanan, M., Bahl, P., Caceres, R., & Davies, N. (2017). The case for edge computing. IEEE Computer, 50(1), 30-39.
- Schneier, B. (2015). Data and Goliath: The hidden battles to collect your data and control your world. W. W. Norton & Company.
- Schneier, B. (2019). Click here to kill everybody: Security and survival in a hyper-connected world. W. W. Norton & Company.
- Securities and Exchange Commission. (2013). Report of investigation: In the Matter of Knight Capital Americas LLC. Washington, DC: Securities and Exchange Commission.
- Shu, K., & Zhang, H. (2019). Cybersecurity challenges of AI in critical infrastructure: A survey. IEEE Transactions on Big Data, 5(2), 159-173.
- U.S. Food and Drug Administration. (2019). Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD).
- Yeung, K. (2018). From principles to practice: An interdisciplinary framework to operationalize AI ethics. IEEE Technology and Society Magazine, 37(4), 46-59.
- Zhang, Y., Qin, W., & Liu, M. (2020). Predictive analytics for cyber security: A survey. Journal of Network and Computer Applications, 163, 102607.
- Zhang, B. H., Lemoine, B., & Mitchell, M. (2018). Mitigating unwanted biases with adversarial learning. Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, 335-340.

ج) منابع اینترنتی

- Australian Broadcasting Corporation. (2011). Maroochy Water breach: The inside story. Retrieved from https://www.abc.net.au/news/2011-06-24/maroochy-water-breach-the-inside-story/2760836
- European Commission. (2019). Ethics guidelines for trustworthy AI. Retrieved from https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

- European Union. (2016). General Data Protection Regulation (GDPR). Retrieved from https://eur-lex.europa.eu/eli/reg/2016/679/oj
- European Commission. (2019). Ethics guidelines for trustworthy AI. Retrieved from https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai
- European Union Agency for Cybersecurity. (2020). Baseline security recommendations for IoT in the context of critical information infrastructures. Retrieved from https://www.enisa.europa.eu/publications/baseline-security-recommendations-for-iot
- OECD. (2019). OECD principles on artificial intelligence. Retrieved from http://www.oecd.org/going-digital/ai/principles/