



هوش مصنوعی در امنیت سایبری: بررسی کاربردها، چالش‌ها و چشم‌اندازهای آینده

علیرضا جعفری

کارشناس ارشد امنیت سایبرنتیک

Aj.intel76@gmail.com

نازنین صادقپور

کارشناس ارشد مدیریت

n.sadeghpour.ns@gmail.com

چکیده

افزایش تعداد، سرعت و پیچیدگی حملات سایبری و محدودیت برخی روش‌های سنتی در شناسایی تهدیدات ناشناخته، نیاز به استفاده از رویکردهای هوشمند و تطبیق‌پذیر در امنیت سایبری را افزایش داده است. هوش مصنوعی با بهره‌گیری از قابلیت‌هایی مانند پردازش حجم گسترده داده‌ها، شناسایی الگوها، تحلیل رفتاری و تشخیص ناهنجاری، ظرفیت قابل توجهی برای ارتقای فرایندهای دفاع سایبری فراهم کرده است. هدف این پژوهش، بررسی مهم‌ترین کاربردهای هوش مصنوعی در امنیت سایبری، تحلیل مزایای آن و شناسایی چالش‌ها و محدودیت‌های مرتبط با استفاده از این فناوری است. پژوهش حاضر از نظر هدف کاربردی و از نظر روش، مروری - تحلیلی است و با بررسی مطالعات علمی و اسناد تخصصی مرتبط با کاربرد یادگیری ماشین، یادگیری عمیق، پردازش زبان طبیعی، یادگیری فدرال و هوش مصنوعی مولد انجام شده است. یافته‌های پژوهش نشان می‌دهد که هوش مصنوعی در حوزه‌هایی از جمله تشخیص نفوذ، شناسایی ناهنجاری، کشف بدافزار، تشخیص فیشینگ، تحلیل اطلاعات تهدید، امنیت اینترنت اشیا و رایانش ابری و خودکارسازی پاسخ به رخدادهای امنیتی کاربرد دارد. با وجود این ظرفیت‌ها، چالش‌هایی مانند کیفیت و عدم توازن داده‌ها، خطاهای تشخیصی، تغییر توزیع داده‌ها، آسیب‌پذیری در برابر حملات خصمانه، حریم خصوصی، هزینه محاسباتی و محدودیت توضیح‌پذیری می‌توانند قابلیت اعتماد و اثربخشی سامانه‌های هوشمند را کاهش دهند. همچنین، ظهور هوش مصنوعی مولد و مدل‌های زبانی بزرگ، ضمن ایجاد فرصت‌های جدید برای تحلیل و خودکارسازی عملیات امنیتی، تهدیدهای تازه‌ای مانند دستکاری ورودی، افشای اطلاعات و سوءاستفاده مهاجمان ایجاد کرده است. در مجموع، نتایج نشان می‌دهد که هوش مصنوعی نباید جایگزین کامل سازوکارهای امنیتی موجود تلقی شود، بلکه استفاده مؤثر از آن مستلزم ادغام با معماری دفاعی چندلایه، نظارت انسانی، ارزیابی مستمر و سازوکارهای مدیریت ریسک است.



واژگان کلیدی: هوش مصنوعی، امنیت سایبری، یادگیری ماشین، یادگیری عمیق، تشخیص تهدیدات

مقدمه

گسترش فناوری‌های دیجیتال، رایانش ابری، اینترنت اشیا، سامانه‌های هوشمند و خدمات برخط، وابستگی سازمان‌ها و جوامع به زیرساخت‌های اطلاعاتی را افزایش داده است. هم‌زمان، محیط تهدیدات سایبری نیز پویاتر و پیچیده‌تر شده و مهاجمان از روش‌های متنوعی برای نفوذ، سرقت اطلاعات، اختلال در خدمات و سوءاستفاده از آسیب‌پذیری‌ها استفاده می‌کنند. افزایش حجم و تنوع داده‌های تولیدشده در شبکه‌ها و سامانه‌های اطلاعاتی، شناسایی رفتارهای مخرب را برای روش‌های مبتنی بر قواعد ثابت و امضاهای از پیش تعریف‌شده دشوارتر کرده است. در چنین شرایطی، توانایی تحلیل مستمر داده‌ها، شناسایی الگوهای پیچیده و سازگاری با تغییرات محیطی به یکی از نیازهای مهم امنیت سایبری تبدیل شده است. هوش مصنوعی در این زمینه ظرفیت قابل‌توجهی برای خودکارسازی فعالیت‌های تکراری، تسریع تشخیص و پاسخ به تهدیدات و پشتیبانی از تصمیم‌گیری امنیتی فراهم کرده است. یک مرور نظام‌مند گسترده نشان داده است که کاربردهای هوش مصنوعی در امنیت سایبری را می‌توان در حوزه‌هایی مانند شناسایی، حفاظت، تشخیص، پاسخ و بازبانی دسته‌بندی کرد (Kaur et al., 2023). همچنین مطالعات جدیدتر نشان می‌دهند که یادگیری ماشین و یادگیری عمیق در سامانه‌های تشخیص نفوذ، تحلیل ترافیک و شناسایی الگوهای غیرعادی نقش فزاینده‌ای پیدا کرده‌اند (Ali et al., 2024).

یادگیری ماشین و یادگیری عمیق می‌توانند با تحلیل داده‌های شبکه، گزارش‌های رخداد، رفتار کاربران، ویژگی‌های فایل‌ها و سایر داده‌های امنیتی، الگوهای مرتبط با فعالیت‌های مشکوک را استخراج کنند. این قابلیت زمینه استفاده از مدل‌های هوشمند را در حوزه‌هایی مانند تشخیص نفوذ، تشخیص ناهنجاری، شناسایی بدافزار، تحلیل فیشینگ، اولویت‌بندی هشدارها و تحلیل اطلاعات تهدید فراهم کرده است. برخلاف رویکردهای صرفاً مبتنی بر امضا، برخی روش‌های یادگیری ماشین قادرند روابط آماری و رفتاری موجود در داده‌ها را یاد بگیرند و برای شناسایی الگوهایی که پیش‌تر به‌صورت صریح تعریف نشده‌اند مورد استفاده قرار گیرند (Buczak and Guven, 2016).

با این حال، نقش هوش مصنوعی در امنیت سایبری تنها به استفاده از آن به‌عنوان ابزار دفاعی محدود نیست. هوش مصنوعی مولد و مدل‌های زبانی بزرگ، هم ظرفیت‌های جدیدی برای تحلیل و خودکارسازی ایجاد کرده‌اند و هم سطح تهدید را تغییر داده‌اند. مهاجمان می‌توانند از قابلیت‌های هوش مصنوعی برای تولید محتوای متقاعدکننده، مهندسی اجتماعی و مقیاس‌دهی برخی فعالیت‌های مخرب استفاده کنند؛ در مقابل، سازمان‌ها نیز می‌توانند از همین فناوری برای تحلیل حجم بالای اطلاعات، خلاصه‌سازی رخدادها و پشتیبانی از تحلیل‌گران بهره بگیرند. چارچوب مدیریت ریسک هوش مصنوعی NIST نیز بر ضرورت توجه هم‌زمان به ویژگی‌هایی مانند



امنیت، تاب‌آوری، شفافیت، توضیح‌پذیری و حفظ حریم خصوصی در چرخه عمر سامانه‌های هوش مصنوعی تأکید دارد (NIST, 2023). از سوی دیگر، استفاده از هوش مصنوعی با مجموعه‌ای از ریسک‌ها و محدودیت‌ها همراه است. کیفیت و نمایندگی داده‌های آموزشی، عدم توازن داده‌ها، هشدارهای کاذب، تغییر توزیع داده‌ها و کاهش قابلیت تعمیم مدل می‌توانند عملکرد سامانه‌های هوشمند را کاهش دهند. پژوهش‌های جدید در زمینه تشخیص بدافزار نیز نشان داده‌اند که تغییرات محیطی و Concept Drift می‌تواند موجب کاهش عملکرد مدل‌های یادگیری ماشین در طول زمان شود (Abusnaina et al., 2024).

آسیب‌پذیری مدل‌ها در برابر حملات خصمانه، حریم خصوصی داده‌ها، پیچیدگی مدل‌های عمیق، هزینه محاسباتی و دشواری تفسیر برخی تصمیم‌ها نیز از دیگر چالش‌های مهم هستند. بنابراین، افزایش دقت یک مدل به تنهایی برای استقرار آن در محیط‌های حساس امنیتی کافی نیست و عواملی مانند مقاومت در برابر حمله، قابلیت تعمیم، توضیح‌پذیری و اعتمادپذیری نیز باید مورد توجه قرار گیرند.

بر این اساس، پژوهش حاضر با رویکردی مروری - تحلیلی به بررسی نقش هوش مصنوعی در امنیت سایبری می‌پردازد. هدف اصلی پژوهش، بررسی کاربردهای مهم روش‌های هوش مصنوعی، به‌ویژه یادگیری ماشین و یادگیری عمیق، در شناسایی و تحلیل تهدیدات سایبری و بررسی ظرفیت این فناوری برای تقویت فرایندهای دفاع و پاسخ به رخداد‌های امنیتی است. همچنین چالش‌های مهم استفاده از هوش مصنوعی و روندهای نوظهور، از جمله هوش مصنوعی مولد، مدل‌های زبانی بزرگ، یادگیری تقویتی و خودکارسازی عملیات امنیتی بررسی می‌شوند.

روش تحقیق

پژوهش حاضر از نظر هدف، کاربردی و از نظر روش، مروری - تحلیلی است. با توجه به ماهیت پژوهش، داده‌های میدانی، جامعه آماری و نمونه‌گیری در آن مورد استفاده قرار نگرفته است و اطلاعات مورد نیاز از طریق بررسی مطالعات علمی و منابع تخصصی مرتبط با هوش مصنوعی و امنیت سایبری گردآوری شده است.

منابع مورد بررسی شامل مقالات پژوهشی و مروری منتشرشده در مجلات و کنفرانس‌های علمی معتبر و همچنین اسناد تخصصی نهادهای معتبر در حوزه هوش مصنوعی و امنیت سایبری بوده است. تمرکز اصلی بر مطالعات مرتبط با یادگیری ماشین، یادگیری عمیق، تشخیص نفوذ، تشخیص ناهنجاری، شناسایی بدافزار، امنیت اینترنت اشیا، امنیت رایانش ابری، یادگیری فدرال، هوش مصنوعی مولد و مدیریت ریسک سامانه‌های هوش مصنوعی قرار داشته است.

معیارهای اصلی انتخاب منابع شامل ارتباط مستقیم با موضوع پژوهش، اعتبار علمی یا تخصصی منبع، قابلیت استناد، پوشش مناسب محور مورد بررسی و روزآمدی مطالب بوده است. در کنار منابع بنیادی، مطالعات جدیدتر نیز بررسی شده‌اند تا تحولات اخیر حوزه هوش مصنوعی و امنیت سایبری در تحلیل نهایی لحاظ شود.



پس از گردآوری منابع، مطالب در سه محور اصلی شامل کاربردهای هوش مصنوعی، چالش‌ها و محدودیت‌ها و چشم‌اندازهای آینده دسته‌بندی و تحلیل شدند. در مرحله نخست، کاربردهای هوش مصنوعی در تشخیص و مقابله با تهدیدات بررسی شد. در مرحله دوم، محدودیت‌هایی مانند کیفیت داده، خطاهای تشخیصی، حملات خصمانه، حریم خصوصی، توضیح‌پذیری و هزینه محاسباتی تحلیل شد. در مرحله سوم نیز روندهای نوظهور شامل هوش مصنوعی مولد، مدل‌های زبانی بزرگ، عامل‌های هوشمند، یادگیری تقویتی و خودکارسازی پاسخ مورد بررسی قرار گرفتند. در نهایت، یافته‌های مطالعات مختلف به صورت مقایسه‌ای و تحلیلی بررسی شدند.

کاربردهای هوش مصنوعی در امنیت سایبری

هوش مصنوعی به یکی از رویکردهای مهم برای تقویت قابلیت‌های تشخیص، تحلیل و پاسخ در امنیت سایبری تبدیل شده است. مرورهای علمی نشان می‌دهند که کاربردهای این فناوری از شناسایی و تشخیص تهدید تا پاسخ و بازیابی را دربرمی‌گیرد (Kaur et al., ۲۰۲۳; Ofusori et al., ۲۰۲۴).

تشخیص تهدیدات و نفوذ سایبری

سامانه‌های تشخیص نفوذ با هدف شناسایی فعالیت‌های غیرمجاز یا مشکوک در شبکه‌ها و سامانه‌های اطلاعاتی طراحی شده‌اند. در روش‌های مبتنی بر یادگیری ماشین، مدل با استفاده از داده‌های ترافیک شبکه و رخدادهای امنیتی، الگوهای مرتبط با رفتار عادی و مخرب را فرا می‌گیرد و از آن‌ها برای طبقه‌بندی رخدادهای جدید استفاده می‌کند. الگوریتم‌هایی مانند درخت تصمیم، ماشین بردار پشتیبان، جنگل تصادفی و شبکه‌های عصبی در مطالعات مختلف برای توسعه سامانه‌های تشخیص نفوذ به کار گرفته شده‌اند (Buczak and Guven, ۲۰۱۶).

مطالعات جدید نیز نشان می‌دهند که استفاده از یادگیری ماشین و یادگیری عمیق در IDS می‌تواند قابلیت تشخیص تهدیدات را افزایش دهد، هرچند عملکرد این مدل‌ها به کیفیت داده، مجموعه داده مورد استفاده، معیارهای ارزیابی و شرایط محیط عملیاتی وابسته است (Ali et al., ۲۰۲۴). از این رو، استفاده عملی از این مدل‌ها نیازمند ارزیابی مستمر و بررسی قابلیت تعمیم آن‌ها است.

• تشخیص ناهنجاری و تحلیل رفتاری



تشخیص ناهنجاری یکی از کاربردهای مهم هوش مصنوعی در امنیت سایبری است. در این رویکرد، به جای جست و جوی صرف الگوهای شناخته شده حمله، رفتارهایی که از الگوی معمول شبکه، سامانه یا کاربر فاصله دارند شناسایی می شوند. این قابلیت به ویژه در مواجهه با تهدیدات جدید که ممکن است با امضاهای شناخته شده مطابقت نداشته باشند اهمیت دارد. روش های بدون نظارت و نیمه نظارتی می توانند در شرایطی که داده های برچسب خورده محدود هستند، برای یادگیری رفتار عادی و شناسایی انحراف از آن مورد استفاده قرار گیرند. با این حال، تغییر دائمی رفتار کاربران، سرویس ها و ساختار شبکه می تواند باعث تغییر توزیع داده شود و عملکرد مدل را در طول زمان کاهش دهد. بنابراین، پایش مستمر و به روزرسانی مدل از الزامات مهم است (Kaur et al., 2023).

• شناسایی بدافزار و حملات فیشینگ

هوش مصنوعی می تواند در شناسایی بدافزارها با تحلیل ویژگی های ایستا و پویا، رفتار فایل ها و برنامه ها و الگوهای مرتبط با اجرای آن ها مورد استفاده قرار گیرد. یادگیری عمیق نیز برای استخراج الگوهای پیچیده از داده های مربوط به فایل ها، کد و رفتار اجرای برنامه ها مورد توجه قرار گرفته است (Redhu et al., 2024).

با این حال، عملکرد مدل های تشخیص بدافزار می تواند در اثر تفاوت مجموعه داده ها، تغییر رفتار بدافزارها و تغییر محیط عملیاتی کاهش یابد. مفهوم Concept Drift در این زمینه اهمیت ویژه ای دارد، زیرا رابطه میان ویژگی های ورودی و کلاس تهدید در طول زمان ممکن است تغییر کند (Abusnaina et al., 2024).

در حوزه فیشینگ نیز مدل های یادگیری ماشین می توانند ویژگی های مربوط به نشانی وب، ساختار صفحات، محتوای پیام و سایر نشانه های رفتاری را تحلیل کنند. پردازش زبان طبیعی نیز امکان تحلیل محتوای متنی ایمیل ها و پیام ها و شناسایی برخی الگوهای زبانی مرتبط با تلاش های فریبکارانه را فراهم می کند. با توجه به تغییر مداوم روش های فیشینگ، به روزرسانی مدل های تشخیص اهمیت زیادی دارد.

• تحلیل اطلاعات تهدید و پیش بینی تهدیدات سایبری

سازمان ها با حجم زیادی از گزارش های رخداد، هشدارهای امنیتی، داده های شبکه، گزارش های آسیب پذیری و اطلاعات مرتبط با فعالیت های مشکوک مواجه هستند. تحلیل دستی این داده ها زمان بر است و می تواند موجب از دست رفتن ارتباط میان رخداد های مختلف شود. روش های هوش مصنوعی می توانند با پردازش داده های ناهمگون، در اولویت بندی هشدارها، استخراج ارتباط میان رخدادها و تحلیل اطلاعات تهدید به تحلیل گران کمک کنند (Kaur et al., 2023; Ofusori et al., 2024).



پیش‌بینی تهدید به معنای استفاده از داده‌ها و شواهد موجود برای شناسایی روندها و برآورد احتمال برخی رخدادهای آینده است و نباید به‌عنوان پیش‌بینی قطعی نوع یا زمان حمله تلقی شود. تغییر رفتار مهاجمان و کمبود داده‌های معتبر، پیش‌بینی دقیق تهدیدات آینده را دشوار می‌کند.

• امنیت در اینترنت اشیا و محیط‌های ابری

گسترش اینترنت اشیا موجب افزایش تعداد تجهیزات، تنوع دستگاه‌ها و حجم داده‌های قابل پردازش شده است. محدودیت منابع محاسباتی برخی تجهیزات IoT و تفاوت الگوهای ارتباطی آن‌ها، طراحی سامانه‌های امنیتی مؤثر را دشوار می‌کند. مطالعات مروری نشان می‌دهند که یادگیری ماشین و یادگیری عمیق در تشخیص نفوذ، شناسایی ناهنجاری و طبقه‌بندی رفتارهای مخرب در محیط‌های IoT کاربرد گسترده‌ای دارند (Ghaffari et al., 2024).

یادگیری فدرال نیز برای آموزش مدل بدون انتقال مستقیم داده‌های خام مورد توجه قرار گرفته است. این رویکرد می‌تواند نیاز به متمرکزسازی داده‌های حساس را کاهش دهد، اگرچه خود فرایند یادگیری فدرال نیز با چالش‌های امنیتی و حریم خصوصی همراه است (Li et al., 2024).

در محیط‌های ابری نیز روش‌های یادگیری ماشین و یادگیری عمیق برای تشخیص تهدیدات، تحلیل رفتار کاربران و سرویس‌ها و شناسایی فعالیت‌های غیرعادی به کار گرفته شده‌اند. مطالعات این حوزه علاوه بر مزایای روش‌های هوشمند، چالش‌هایی مانند حریم خصوصی، مقیاس‌پذیری، توضیح‌پذیری و قابلیت تعمیم را مطرح کرده‌اند (Alzoubi et al., 2024).

• خودکارسازی پاسخ به رخدادهای امنیتی

یکی از جهت‌گیری‌های مهم کاربرد هوش مصنوعی، کاهش فاصله زمانی میان شناسایی تهدید و واکنش به آن است. سامانه‌های هوشمند می‌توانند با اولویت‌بندی هشدارها، ارتباط دادن رخدادهای مرتبط، خلاصه‌سازی اطلاعات و ارائه پیشنهاد برای اقدامات دفاعی، بخشی از فرایند پاسخ به رخداد را تسریع کنند (Kaur et al., 2023; Ofusori et al., 2024).

با وجود ظرفیت بالای خودکارسازی، واگذاری کامل تصمیم‌های امنیتی به مدل‌های هوش مصنوعی می‌تواند ریسک ایجاد کند؛ زیرا خطای مدل ممکن است به مسدودسازی سرویس مجاز یا اجرای اقدام دفاعی نامناسب منجر شود. بنابراین، در سناریوهای حساس، رویکرد «انسان در حلقه» می‌تواند گزینه مناسب‌تری باشد



چالش‌ها و محدودیت‌های استفاده از هوش مصنوعی در امنیت سایبری

با وجود ظرفیت قابل توجه هوش مصنوعی، استفاده از آن با محدودیت‌های فنی، داده‌ای، امنیتی و مدیریتی همراه است. بنابراین، ارزیابی سامانه‌های هوشمند نباید صرفاً بر دقت مدل متمرکز باشد و باید معیارهایی مانند پایداری، قابلیت تعمیم، مقاومت در برابر حمله، توضیح‌پذیری، حریم خصوصی و هزینه عملیاتی نیز مورد توجه قرار گیرد (NIST, 2023).

• کیفیت، کمبود و عدم توازن داده‌ها

داده یکی از مهم‌ترین مؤلفه‌های سامانه‌های هوش مصنوعی است. داده‌های امنیتی واقعاً معمولاً حجم زیاد، ناهمگونی، نویز، عدم توازن کلاس‌ها و تغییرات زمانی دارند. در بسیاری از مجموعه‌داده‌ها، نمونه‌های رفتار عادی بسیار بیشتر از نمونه‌های حملات نادر هستند؛ بنابراین، اتکا به Accuracy می‌تواند تصویر گمراه‌کننده‌ای از عملکرد مدل ایجاد کند.

تفاوت میان داده‌های آزمایشگاهی و شرایط واقعی شبکه نیز از چالش‌های مهم است. مدل ممکن است در یک مجموعه‌داده عملکرد مطلوبی داشته باشد، اما در محیطی با ویژگی‌های متفاوت عملکرد آن کاهش یابد. پژوهش‌های مرتبط با امنیت ابری نیز بر چالش‌هایی مانند قابلیت تعمیم و سوگیری داده‌ها تأکید کرده‌اند (Alzoubi et al., 2024).

• خطاهای تشخیصی و هشدارهای کاذب

دو خطای مهم در سامانه‌های تشخیص امنیتی، شناسایی نادرست فعالیت سالم به عنوان تهدید و ناتوانی در شناسایی فعالیت مخرب واقعی هستند که به ترتیب False Positive و False Negative نامیده می‌شوند. افزایش هشدارهای کاذب موجب افزایش بار کاری تحلیل‌گران و کاهش اعتماد به سامانه می‌شود، در حالی که افزایش هشدارهای از دست‌رفته می‌تواند به عبور تهدید واقعی منجر شود. از این‌رو، ارزیابی مدل‌های امنیتی نباید تنها بر Accuracy متکی باشد. معیارهایی مانند Precision, Recall و F1-score و همچنین زمان تشخیص، نرخ هشدار کاذب و هزینه عملیاتی خطاها باید مورد توجه قرار گیرند (Ali et al., 2024).

• آسیب‌پذیری در برابر حملات خصمانه



مدل‌های یادگیری ماشین ممکن است در معرض حملات خصمانه قرار گیرند. در این حملات، مهاجم با ایجاد تغییرات هدفمند در ورودی یا فرایند یادگیری تلاش می‌کند مدل را به اتخاذ تصمیم نادرست وادار کند. حملات Evasion با هدف فریب مدل در مرحله استفاده عملیاتی و حملات Poisoning با هدف دستکاری داده‌های آموزشی انجام می‌شوند.

این مسئله نشان می‌دهد که استفاده از هوش مصنوعی برای دفاع سایبری باید هم‌زمان با حفاظت از خود مدل، داده‌های آموزشی و زیرساخت اجرای آن همراه باشد. چارچوب مدیریت ریسک هوش مصنوعی NIST نیز امنیت و تاب‌آوری را از ویژگی‌های مهم سامانه‌های قابل اعتماد می‌داند (NIST, ۲۰۲۳).

• تغییرپذیری محیط و کاهش قابلیت تعمیم مدل

محیط امنیت سایبری دائماً در حال تغییر است. مهاجمان روش‌های جدیدی توسعه می‌دهند، زیرساخت‌های سازمانی تغییر می‌کنند و رفتار کاربران ثابت باقی نمی‌ماند. در نتیجه، توزیع آماری داده‌ها ممکن است تغییر کند و رابطه‌ای که مدل در مرحله آموزش آموخته است، در آینده دیگر معتبر نباشد.

این مسئله در قالب مفاهیمی مانند Concept Drift و Distribution Shift مطرح می‌شود. پژوهش Abusnaina و همکاران (۲۰۲۴) نشان داده است که Concept Drift می‌تواند محدودیت مهمی برای مدل‌های یادگیری ماشین در تشخیص بدافزار ایجاد کند. بنابراین، پایش عملکرد، تشخیص تغییر توزیع داده و به‌روزرسانی مدل از الزامات استقرار بلندمدت سامانه‌های هوشمند است.

• توضیح‌پذیری و اعتمادپذیری تصمیم‌های هوش مصنوعی

مدل‌های پیچیده یادگیری عمیق در بسیاری از موارد توضیح‌پذیری محدودی دارند. در محیط امنیت سایبری، تحلیل‌گر باید بتواند دلایل طبقه‌بندی یک فایل، کاربر، اتصال یا رخداد به‌عنوان تهدید را بررسی کند. نبود توضیح مناسب می‌تواند اعتماد به سامانه را کاهش دهد.

روش‌های Explainable AI می‌توانند به تفسیر بهتر خروجی مدل و افزایش نظارت انسانی کمک کنند. با این حال، توضیح ارائه‌شده باید تا حد امکان با فرایند واقعی تصمیم‌گیری مدل ارتباط داشته باشد و صرفاً یک توضیح ظاهری نباشد. موضوع توضیح‌پذیری و اعتمادپذیری از جمله چالش‌های مطرح در کاربرد یادگیری ماشین و یادگیری عمیق در امنیت ابری نیز محسوب می‌شود (Alzoubi et al., 2024).



• حریم خصوصی و امنیت داده‌ها

استفاده از هوش مصنوعی در امنیت سایبری معمولاً به جمع‌آوری و تحلیل حجم زیادی از داده‌ها نیاز دارد. این داده‌ها ممکن است شامل اطلاعات کاربران، ارتباطات، دستگاه‌ها و رفتارهای سازمانی باشند. در نتیجه، پردازش و ذخیره‌سازی آن‌ها می‌تواند نگرانی‌های حریم خصوصی ایجاد کند.

یادگیری فدرال یکی از رویکردهای مورد توجه برای کاهش نیاز به انتقال مستقیم داده‌های خام است. با این حال، این روش نیز در برابر تهدیدهای خاص خود قرار دارد و تبادل پارامترها و مدل‌ها می‌تواند هدف حملات قرار گیرد (Li et al., ۲۰۲۴). بنابراین، حفاظت از حریم خصوصی باید در تمام چرخه عمر داده و مدل مورد توجه باشد.

• هزینه محاسباتی و پیچیدگی پیاده‌سازی

مدل‌های پیشرفته، به‌ویژه مدل‌های عمیق و مدل‌های بزرگ، ممکن است به منابع محاسباتی و زیرساخت تخصصی قابل توجهی نیاز داشته باشند. علاوه بر هزینه سخت‌افزار، توسعه، آموزش، ارزیابی، به‌روزرسانی و نگهداری مدل نیز به نیروی انسانی متخصص نیاز دارد.

در کاربردهای بلندمدت، تأخیر نیز اهمیت دارد. مدلی که دقت بالایی دارد اما زمان پردازش زیادی نیاز دارد ممکن است برای برخی سناریوهای عملیاتی مناسب نباشد. بنابراین، میان دقت، سرعت، مصرف منابع و هزینه عملیاتی باید تعادل برقرار شود (Alzoubi et al., ۲۰۲۴).

• استفاده دوگانه و سوءاستفاده از هوش مصنوعی

هوش مصنوعی ماهیتی دوگانه دارد و همان قابلیت‌هایی که برای تقویت دفاع سایبری استفاده می‌شوند، می‌توانند توسط مهاجمان نیز مورد بهره‌برداری قرار گیرند. تولید محتوای متقاعدکننده، تحلیل اطلاعات و افزایش مقیاس برخی فعالیت‌های مخرب از جمله زمینه‌هایی هستند که ظهور هوش مصنوعی مولد اهمیت آن‌ها را افزایش داده است. از این رو، ارزیابی اثر هوش مصنوعی بر امنیت سایبری باید دو جنبه را هم‌زمان در نظر بگیرد: استفاده از AI برای شناسایی، پیشگیری و پاسخ به تهدیدات و استفاده مهاجمان از AI برای افزایش توانایی و مقیاس فعالیت‌های مخرب. NIST نیز در پروفایل هوش مصنوعی مولد بر ضرورت شناسایی و مدیریت ریسک‌های اختصاصی این فناوری تأکید کرده است (Autio et al., 2024).



چشم‌انداز آینده هوش مصنوعی در امنیت سایبری

تحولات سریع هوش مصنوعی نشان می‌دهد که نقش این فناوری در امنیت سایبری از تشخیص تهدید فراتر خواهد رفت و به سمت تحلیل پیشرفته، تصمیم‌یارهای هوشمند، خودکارسازی فرایندهای دفاعی و تعامل مستمر میان انسان و سامانه‌های هوشمند حرکت خواهد کرد. مرورهای جدید نیز توسعه کاربردهای نوظهور و روش‌های پیشرفته هوش مصنوعی را از مسیرهای مهم آینده این حوزه معرفی کرده‌اند (Kaur et al., ۲۰۲۳; Ofusori et al., ۲۰۲۴).

• هوش مصنوعی مولد و مدل‌های زبانی بزرگ

مدل‌های زبانی بزرگ می‌توانند حجم زیادی از گزارش‌های امنیتی، هشدارها، مستندات فنی و داده‌های متنی را تحلیل و خلاصه‌سازی کنند. این قابلیت می‌تواند به تحلیل‌گران در درک سریع‌تر رخدادها، استخراج اطلاعات مهم و ارتباط دادن هشدارهای مختلف کمک کند. با وجود این، استفاده از مدل‌های زبانی در کاربردهای امنیتی حساس با خطر تولید اطلاعات نادرست، افشای اطلاعات و دستکاری ورودی‌ها همراه است. NIST در پروفایل هوش مصنوعی مولد بر مدیریت ریسک‌های اختصاصی این فناوری در چرخه عمر سامانه تأکید کرده است (Autio et al., ۲۰۲۴).

• عامل‌های هوشمند و خودکارسازی عملیات امنیتی

یکی از مسیرهای آینده، حرکت از مدل‌های صرفاً تحلیل‌گر به سمت عامل‌های هوشمند است. این عامل‌ها می‌توانند با استفاده از مدل‌های هوش مصنوعی، ابزارهای نرم‌افزاری و منابع اطلاعاتی، وظایفی مانند جمع‌آوری شواهد، تحلیل هشدارها و پیشنهاد اقدامات دفاعی را انجام دهند.

با افزایش سطح اختیار عامل، پیامد خطا نیز افزایش می‌یابد. بنابراین، تعیین سطح دسترسی، محدود کردن دامنه اقدامات، ثبت فعالیت‌ها و استفاده از تأیید انسانی برای اقدامات پرریسک از الزامات مهم این حوزه خواهد بود.



• توسعه یادگیری تقویتی در دفاع سایبری

یادگیری تقویتی می تواند در آینده برای تصمیم گیری پویا در محیط های امنیتی مورد استفاده قرار گیرد. عامل هوشمند در این روش با تعامل با محیط و دریافت پاداش یا جریمه، سیاست هایی را برای انتخاب اقدامات مناسب یاد می گیرد.

با این حال، استفاده عملی از یادگیری تقویتی با چالش هایی مانند هزینه آموزش، طراحی تابع پاداش، ایمنی فرایند یادگیری و تفاوت میان محیط شبیه سازی و محیط واقعی همراه است. بنابراین، استفاده از این روش نیازمند محیط های آزمایشی کنترل شده و محدودیت های ایمنی مناسب است.

• تکامل خودکارسازی پاسخ به تهدیدات

در آینده، سامانه های امنیتی می توانند بیش از گذشته از هوش مصنوعی برای هماهنگ سازی مراحل تشخیص، تحلیل و پاسخ استفاده کنند. هدف، ایجاد سامانه هایی است که بتوانند پس از شناسایی تهدید، شواهد مرتبط را جمع آوری، شدت رخداد را ارزیابی و واکنش مناسب را پیشنهاد دهند. با این حال، خودکارسازی کامل برای همه رخدادها مناسب نیست. آینده سامانه های دفاعی احتمالاً به سمت معماری های تطبیقی حرکت خواهد کرد که در آن میزان استقلال سیستم بر اساس سطح اطمینان مدل، نوع تهدید و اهمیت دارایی تعیین می شود.

• امنیت و اعتمادپذیری مدل های هوش مصنوعی

با افزایش استفاده از هوش مصنوعی در زیرساخت های امنیتی، حفاظت از خود مدل ها به یکی از محورهای اصلی آینده تبدیل خواهد شد. مدل های امنیتی باید در برابر حملات خصمانه، مسموم سازی داده ها، استخراج مدل، نشت اطلاعات و سایر تهدیدهای چرخه عمر هوش مصنوعی مقاوم باشند. در نتیجه، امنیت مدل نباید پس از توسعه آن به عنوان مرحله ای جانبی در نظر گرفته شود، بلکه باید از مرحله طراحی و جمع آوری داده تا آموزش، استقرار، پایش و به روزرسانی در چرخه عمر سامانه لحاظ شود (NIST, ۲۰۲۳).

• نقش هوش مصنوعی در تحلیل زبان و اطلاعات تهدید



گزارش‌های فنی، بولتن‌های امنیتی، گزارش رخدادها و سایر منابع متنی حاوی اطلاعات ارزشمندی درباره تهدیدات هستند. مدل‌های زبانی می‌توانند در استخراج موجودیت‌ها، ارتباط میان رخدادها، خلاصه‌سازی گزارش‌ها و دسته‌بندی اطلاعات تهدید به تحلیل‌گران کمک کنند. ترکیب پردازش زبان طبیعی با داده‌های شبکه، سامانه‌های تشخیص نفوذ و پایگاه‌های اطلاعات تهدید می‌تواند زمینه ایجاد سامانه‌های تحلیل چندوجهی را فراهم کند و دید جامع‌تری از رخدادها و امنیتی ارائه دهد.

• هوش مصنوعی و رایانش کوانتومی

رایانش کوانتومی یکی از فناوری‌هایی است که ممکن است در بلندمدت بر امنیت سایبری و برخی روش‌های محاسباتی مرتبط با هوش مصنوعی تأثیر بگذارد. افزایش توان محاسباتی کوانتومی می‌تواند فرصت‌هایی برای حل برخی مسائل پیچیده ایجاد کند، اما هم‌زمان تهدیدهایی جدی برای برخی سازوکارهای رمزنگاری متداول به وجود می‌آورد. از این رو، ارتباط میان هوش مصنوعی، رایانش کوانتومی و امنیت سایبری موضوعی مهم برای پژوهش‌های آینده محسوب می‌شود.

با این حال، نباید تأثیر رایانش کوانتومی بر هوش مصنوعی را به صورت قطعی و نزدیک‌الوقوع در نظر گرفت. بسیاری از کاربردهای عملی این فناوری همچنان در مرحله توسعه قرار دارند و میزان اثرگذاری آن‌ها بر سامانه‌های امنیتی به پیشرفت سخت‌افزار، الگوریتم‌ها و زیرساخت‌های مرتبط وابسته است. بنابراین، در شرایط کنونی، تمرکز اصلی پژوهش‌ها باید بر کاربردهای عملی و قابل استقرار هوش مصنوعی در امنیت سایبری و هم‌زمان پایش تحولات رایانش کوانتومی باشد.

در مجموع، آینده هوش مصنوعی در امنیت سایبری به سمت سامانه‌هایی حرکت خواهد کرد که توانایی بیشتری در تحلیل خودکار، سازگاری با تغییرات محیط، ترکیب منابع مختلف اطلاعاتی و پشتیبانی از تصمیم‌گیری امنیتی دارند. با این حال، افزایش سطح خودکارسازی نباید به حذف نظارت انسانی منجر شود. ایجاد تعادل میان خودکارسازی، قابلیت اعتماد، امنیت مدل، حریم خصوصی و نظارت انسانی، یکی از مهم‌ترین الزامات توسعه نسل آینده سامانه‌های امنیت سایبری مبتنی بر هوش مصنوعی خواهد بود.



بررسی مطالعات نشان می‌دهد که هوش مصنوعی ظرفیت قابل توجهی برای ارتقای امنیت سایبری، به‌ویژه در زمینه تشخیص نفوذ، شناسایی ناهنجاری، تحلیل بدافزار، تشخیص فیشینگ، تحلیل اطلاعات تهدید و پشتیبانی از پاسخ به رخدادها دارد. این نتیجه با یافته‌های Buczak and Guven (۲۰۱۶) درباره ظرفیت روش‌های یادگیری ماشین و داده‌کاوی در تشخیص نفوذ هم‌راستا است. همچنین مرور نظام‌مند Kaur et al. (۲۰۲۳) نشان می‌دهد که کاربردهای هوش مصنوعی در امنیت سایبری طیف گسترده‌ای از فعالیت‌های شناسایی، حفاظت، تشخیص، پاسخ و بازبایی را پوشش می‌دهد. مطالعات جدیدتر نیز نشان می‌دهند که یادگیری ماشین و یادگیری عمیق در حوزه‌هایی مانند IDS، بدافزار، IoT و رایانش ابری در حال گسترش هستند (Ali et al., ۲۰۲۴; Redhu et al., ۲۰۲۴; Alzoubi et al., ۲۰۲۴; Ghaffari et al., ۲۰۲۴).

در مقابل، نتایج بررسی نشان می‌دهد که افزایش پیچیدگی مدل لزوماً به معنای افزایش قابلیت اعتماد آن نیست. کیفیت داده، تغییر توزیع داده‌ها، حملات خصمانه، خطاهای تشخیصی، حریم خصوصی و توضیح‌پذیری از عوامل مهمی هستند که می‌توانند عملکرد واقعی سامانه‌های هوشمند را محدود کنند. برای نمونه، پژوهش Abusnaina et al. (۲۰۲۴) اهمیت Concept Drift را در تشخیص بدافزار نشان می‌دهد و مطالعات حوزه رایانش ابری نیز چالش‌هایی مانند قابلیت تعمیم، حریم خصوصی و توضیح‌پذیری را برجسته کرده‌اند (Alzoubi et al., ۲۰۲۴). بنابراین، شواهد موجود بیشتر از رویکرد ترکیبی حمایت می‌کنند که در آن هوش مصنوعی در کنار قواعد امنیتی، سامانه‌های سنتی، تخصص تحلیل گران و نظارت انسانی قرار می‌گیرد.

بر اساس نتایج این پژوهش، پیشنهاد می‌شود سازمان‌ها پیش از استقرار سامانه‌های امنیتی مبتنی بر هوش مصنوعی، کیفیت و نمایندگی داده‌های مورد استفاده را ارزیابی کنند و سازوکارهایی برای پایش مستمر عملکرد و شناسایی تغییرات محیطی ایجاد نمایند. همچنین لازم است مدل‌ها در برابر حملات خصمانه و سایر تهدیدهای چرخه عمر هوش مصنوعی ارزیابی و مقاوم‌سازی شوند. استفاده از روش‌های توضیح‌پذیری، ثبت تصمیم‌های مدل، کنترل سطح دسترسی و نظارت انسانی نیز می‌تواند قابلیت اعتماد سامانه‌های هوشمند را افزایش دهد. در حوزه پژوهش، توسعه مجموعه‌داده‌های واقعی‌تر و متنوع‌تر، طراحی روش‌های مقاوم در برابر تغییرات داده، توسعه مدل‌های کم‌هزینه و قابل استقرار و مطالعه کاربرد هوش مصنوعی مولد و عامل‌های هوشمند در عملیات امنیتی از موضوعات مهم آینده به شمار می‌روند.

در مجموع، هوش مصنوعی را می‌توان یکی از فناوری‌های اثرگذار بر آینده امنیت سایبری دانست که توانایی تغییر شیوه شناسایی، تحلیل و پاسخ به تهدیدات را دارد. با این حال، این فناوری راهکاری مستقل و بدون خطا نیست. وابستگی به داده، امکان دستکاری مدل، مشکلات توضیح‌پذیری، مسائل حریم خصوصی، هزینه محاسباتی و تغییرپذیری تهدیدات از جمله عواملی هستند که باید در زمان طراحی و استقرار سامانه‌های هوشمند مورد توجه قرار گیرند. در نهایت، رویکردی که هوش مصنوعی را در قالب یک معماری دفاعی چندلایه و همراه با نظارت انسانی، کنترل‌های امنیتی و ارزیابی مستمر به کار گیرد، می‌تواند زمینه بهره‌برداری ایمن‌تر و مؤثرتر از ظرفیت‌های این فناوری را در امنیت سایبری فراهم کند.



Abusnaina, A., Anwar, A., Saad, M., Alabduljabbar, A., Jang, R., Salem, S., & Mohaisen, D. (۲۰۲۴). Exposing the limitations of machine learning for malware detection under concept drift. In Web Information Systems Engineering – WISE ۲۰۲۴ (pp. ۲۷۳–۲۸۹). Springer.

Ali, A. H., Charfeddine, M., Ammar, B., Hamed, B. B., Albalwy, F., Alqarafi, A., & Hussain, A. (۲۰۲۴). Unveiling machine learning strategies and considerations in intrusion detection systems: A comprehensive survey. Frontiers in Computer Science, ۶, ۱۳۸۷۳۵۴.

Alzoubi, Y. I., Mishra, A., & Topcu, A. E. (۲۰۲۴). Research trends in deep learning and machine learning for cloud computing security. Artificial Intelligence Review, ۵۷, ۱۳۲.

Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (۲۰۲۴). Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. NIST AI ۶۰۰-۱. National Institute of Standards and Technology.

Buczak, A. L., & Guven, E. (۲۰۱۶). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, ۱۸(۲), ۱۱۵۳–۱۱۷۶.

Ghaffari, A., Jelodari, N., Pouralish, S., Derakhshanfard, N., & Arasteh, B. (۲۰۲۴). Securing Internet of Things using machine and deep learning methods: A survey. Cluster Computing, ۲۷, ۹۰۶۵–۹۰۸۹.

Kaur, R., Gabrijelčič, D., & Klobučar, T. (۲۰۲۳). Artificial intelligence for cybersecurity: Literature review and future research directions. Information Fusion, ۹۷, ۱۰۱۸۰۴.

Li, H., Ge, L., & Tian, L. (۲۰۲۴). Survey: Federated learning data security and privacy-preserving in edge-Internet of Things. Artificial Intelligence Review, ۵۷, ۱۳۰.



بیست و چهارمین کنفرانس بین المللی نوآوری و تحقیق در علوم مهندسی

آکادمی بین المللی و مطالعات گرجستان - ۲۷ شهریور ۱۴۰۵

www.icires.ir
info@icires.ir

18 September 2026 - TBILISI GEORGIA

National Institute of Standards and Technology. (۲۰۲۳). Artificial Intelligence Risk Management Framework (AI RMF ۱/۰). NIST AI ۱۰۰-۱. U.S. Department of Commerce.

Ofusori, L., Bokaba, T., & Mhlongo, S. (۲۰۲۴). Artificial intelligence in cybersecurity: A comprehensive review and future direction. Applied Artificial Intelligence, ۳۸(۱), ۲۴۳۹۶۰۹

Redhu, A., Choudhary, P., Srinivasan, K., & Das, T. K. (۲۰۲۴). Deep learning-powered malware detection in cyberspace: A contemporary review. Frontiers in Physics, ۱۲, ۱۳۴۹۴۶۳