مقاله پژوهشی

بکارگیری هوش مصنوعی در امنیت سایبری

Doi: 10.30508/kdip.2025.471226.1109

مصطفی حسینی[۱]
۱- مربی گروه کامپیوتر، موسسه آموزش عالی شاهرود، شاهرود، ایران

تاریخ دریافت: ۱۴۰۳/۰۵/۱۲
تاریخ پذیرش: ۱۴۰۳/۱۰/۱۷
صفحه: ۴۶ - ۳۶

چکیده

در چشم‌انداز در حال تحول سریع امنیت سایبری، ادغام هوش مصنوعی[۱] (AI) بسیار حائز اهمیت شده است. این مقاله به بررسی تأثیر عمیق ابزارهای مبتنی بر هوش مصنوعی در تقویت دفاع‌های سایبری پرداخته و ضرورت آنها را در مقابله با تهدیدات سایبری پیچیده‌تر شده برجسته می‌کند. توانایی هوش مصنوعی در تحلیل مجموعه داده‌های عظیم با سرعت بی‌سابقه، امکان شناسایی آسیب‌پذیری‌ها و تشخیص ناهنجاری‌هایی که ممکن است به حملات احتمالی اشاره کنند را فراهم می‌آورد، اغلب قبل از اینکه این حملات شکل بگیرند. ابزارهای کلیدی هوش مصنوعی مورد بحث شامل سیستم‌های پیشرفته تشخیص تهدید، الگوریتم‌های یادگیری ماشین برای تحلیل پیش‌بینی‌کننده و مکانیسم‌های پاسخگویی خودکار به حوادث هستند. این ابزارها نه تنها کارایی و دقت تشخیص تهدید را افزایش می‌دهند، بلکه زمان پاسخگویی به حوادث سایبری را به‌طور قابل توجهی کاهش می‌دهند و در نتیجه آسیب‌های احتمالی را کاهش می‌دهند. مقاله همچنین به جنبه تاریک‌تر هوش مصنوعی می‌پردازد و بررسی می‌کند که چگونه بازیگران مخرب از هوش مصنوعی برای افزایش پیچیدگی حملات خود استفاده می‌کنند. بدافزارهای مبتنی بر هوش مصنوعی، طرح‌های فیشینگ خودکار و تاکتیک‌های مهندسی اجتماعی مبتنی بر هوش مصنوعی نشان‌دهنده مرز جدیدی از تهدیدات سایبری هستند، جایی که مهاجمان از هوش مصنوعی برای دور زدن اقدامات امنیتی سنتی بهره‌برداری می‌کنند. با برجسته کردن این جنبه‌های دوگانه هوش مصنوعی در امنیت سایبری، مقاله دیدگاهی جامع از وضعیت فعلی ارائه می‌دهد و بر نیاز حیاتی سازمان‌ها به پذیرش دفاع‌های مبتنی بر هوش مصنوعی برای پیشی گرفتن در مسابقه تسلیحات سایبری تأکید می‌کند. در نهایت، ادغام هوش مصنوعی در استراتژی‌های امنیت سایبری نه تنها یک گزینه بلکه ضرورتی برای اطمینان از حفاظت قوی در عصر دیجیتال است.

کلمات کلیدی: هوش مصنوعی (AI)،امنیت سایبری،تشخیص تهدید،یادگیری ماشین

1- Artificial Intelligence

۱- مقدمه

با پیشرفت عصر دیجیتال، پیچیدگی و فراوانی تهدیدات سایبری افزایش یافته است و چالش‌های مهمی را برای اقدامات سنتی امنیت سایبری ایجاد می‌کند. مجرمان سایبری از فناوری های پیشرفته برای اجرای حملات پیچیده تر استفاده می کنند و مکانیسم های دفاعی متعارف را ناکافی می‌کنند. در این محیط پر خطر، هوش مصنوعی (AI) به عنوان یک نیروی دگرگون کننده در حوزه امنیت سایبری ظاهر شده است. ظرفیت هوش مصنوعی برای پردازش و تجزیه و تحلیل مقادیر زیادی از داده‌ها در زمان واقعی، همراه با توانایی آن در شناسایی الگوها و ناهنجاری ها، آن را به عنوان ابزاری حیاتی در مبارزه با جرایم سایبری قرار می‌دهد.

ادغام هوش مصنوعی در استراتژی‌های امنیت سایبری مزایای زیادی را ارائه می‌دهد. الگوریتم‌های یادگیری ماشینی می‌توانند رفتار غیرعادی را که ممکن است نشان‌دهنده نقض امنیتی باشد، اغلب قبل از اینکه آسیب قابل توجهی ایجاد شود، شناسایی کنند. سیستم‌های مجهز به هوش مصنوعی همچنین می‌توانند وظایف معمولی مانند نظارت بر ترافیک شبکه و مدیریت وصله‌ها را خودکار کنند و منابع انسانی را برای تمرکز بر مسائل پیچیده‌تر آزاد کنند. علاوه بر این، هوش مصنوعی با پیش‌بینی آسیب‌پذیری‌های بالقوه و بردارهای حمله، هوش تهدید را تقویت می‌کند و سازمان‌ها را قادر می‌سازد تا رویکردی پیشگیرانه به جای واکنشی برای امنیت اتخاذ کنند.

با این حال، پذیرش هوش مصنوعی در امنیت سایبری بدون چالش نیست. مسائلی مانند نیاز به حجم وسیعی از داده‌های با کیفیت بالا، سوگیری‌های احتمالی در الگوریتم‌های هوش مصنوعی، و خطر همیشه حاضر حملات خصمانه به خود سیستم‌های هوش مصنوعی باید مورد توجه قرار گیرد. همان‌طور که این پیچیدگی‌ها بررسی می‌شود، برای سازمان‌ها ضروری است که چارچوب‌های حاکمیت هوش مصنوعی قوی را پیاده سازی کنند و از نظارت مستمر و به روزرسانی سیستم‌های هوش مصنوعی اطمینان حاصل کنند. با انجام این کار، آنها می توانند از پتانسیل کامل هوش مصنوعی برای ایجاد یک چشم‌انداز دیجیتال امن‌تر استفاده کنند. در ادامه این مقاله؛ توضیحات کامل‌تری در رابطه با هوش مصنوعی، علل استفاده و کاربرد آن در امنیت سایبری، ارائه شده است.

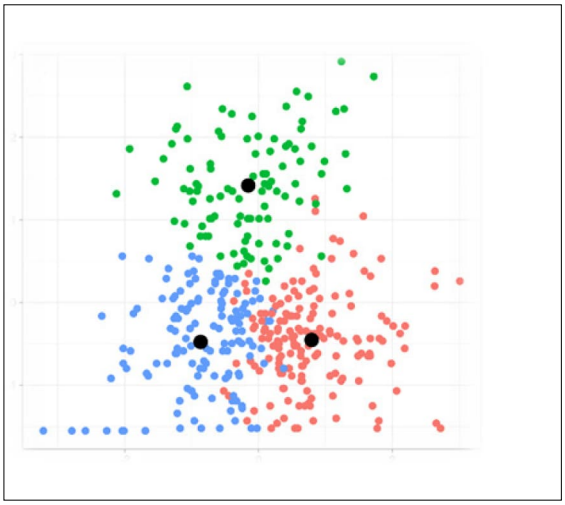
۲- مبانی نظری

یادگیری هوش مصنوعی: پیش از آنکه هوش مصنوعی بتواند در هر زمینه‌ای مورد استفاده واقع شود، بایستی در رابطه با آن موضوع یادگیری انجام دهد و تحت آموزش قرار گیرد. یادگیری هوش مصنوعی انواع مختلفی دارد و از آنجایی که در پژوهش به انواع یادگیری سیستم‌ها اشاره شده است، لازم است به انواع آن اشاره شود (آرپ، و همکاران[1]، ۲۰۲۴؛ گاتسگین[2]، ۲۰۲۳؛ و واگنر[3]، ۲۰۱۷).

الف) یادگیری نظارتی: در این نوع یادگیری، سیستم با استفاده از داده‌های برچسب‌گذاری شده، آموزش‌دیده و پس‌ازآن به‌پیش‌بینی نوع داده‌های جدید و ارزیابی عملکرد پیش‌بینی بر اساس برچسب واقعی داده، می‌پردازد. الگوریتم SVM را می‌توان در این روش استفاده نمود. این الگوریتم عمل طبقه‌بندی[4] را انجام می‌دهد.

ب) یادگیری بدون نظارت: در صورتی‌که داده‌ها جهت آموزش فاقد هرگونه برچسب‌گذاری باشند، در این صورت از مدل‌هایی استفاده می‌شود که تنها بر اساس برآیند ویژگی داده‌ها، میانگین و داده‌های مرکزی را محاسبه نموده و تشخیص دهند. سپس بر طبق نتایج به‌دست‌آمده می‌توان داده‌های خطا و ناهنجاری را از داده‌های حقیقی تعیین و جدا نمود. مدل Kmeans ازجمله الگوریتم‌های بدون نظارت است که به عمل خوشه‌بندی[5]، می‌پردازد. توضیحات را در شکل شماره (۱) مشاهده می‌کنید.

1- Arp,& etal
2- Gottsegen
3- Wagner
4- Classification
5- Clustering



شکل (۱): نمودار مربوط به Kmeans

ج) یادگیری نیمه نظارتی: این روش پلی بین دو نوع نظارت‌شده و بدون نظارت است. که دو کاربرد متفاوت دارد. اولین کاربرد برچسب‌گذاری است. درصورتی‌که مجموعه داده‌ی وسیعی وجود داشته باشد که بدون برچسب است، شما به‌صورت دستی نمی‌توانید آن‌ها را برچسب‌گذاری نمایید. از این‌رو می‌توان قسمت کوچکی از داده را به صورت دستی برچسب‌گذاری کرد و سپس سیستم با استفاده قسمت برچسب‌گذاری شده، سایر داده‌ها را برچسب‌گذاری نمود.

دومین کاربرد ارزیابی مدل‌های خوشه‌بندی است. هنگامی که این تصمیم اتخاذ گردد تا از یک مدل خوشه‌بندی بر داده‌های دارای برچسب اعمال گردد، می‌توان با استفاده از یک مدل، یادگیری بدون نظارت بر داده‌ها انجام شود. در ادامه نتایج به‌دست‌آمده از خوشه‌بندی را با برچسب داده‌ها مقایسه نموده تا بررسی نمود چه مقدار خوشه‌بندی به‌دست‌آمده نسبت به ویژگی‌های حقیقی داده‌ها دقت داشته است.

د) یادگیری تقویتی: در این نوع یادگیری، هوش مصنوعی بدون داشتن هیچ اطلاعاتی از فضا و محیط داده‌ها، با استفاده از آزمون خطا و در نظر گرفتن دو تابع پاداش (در صورت به دست آمدن نتایج مفید) و مجازات (در صورت رخداد اشتباه از سوی کاربر) به یادگیری کلی در رابطه با و داده‌های موجود پرداخته و به‌صورت صفرتا صد بدون هیچ‌گونه ارزیابی، این عمل صورت می‌پذیرد.

علل استفاده در امنیت سایبری

با گسترش روزافزون شبکه‌های شرکتی و سازمانی و افزایش چشمگیر IPها، جمع‌آوری اطلاعات، یکپارچه‌سازی و بررسی آن‌ها، تست نفوذ به شبکه، تشخیص تهدیدات، بررسی و ایمن‌سازی شبکه از توان red team و blue team انسانی خارج‌شده و در صورت امکان زمان‌بر و نیازمند هزینه گزاف است. از طرفی آموزش نیروی انسانی متناسب با میزان تقاضا، تا رسیدن به سطح حرفه‌ای نیازمند پروسه‌ای طولانی است که نیاز بازار کار را رفع نمی‌کند. در حالت کلی می‌توان علل به‌کارگیری هوش مصنوعی را به‌صورت زیر شرح داد.

تشخیص تهدید پیشرفته: هوش مصنوعی می‌تواند حجم وسیعی از داده‌ها را با سرعت بالا تجزیه‌وتحلیل کند و آن را قادر می‌سازد الگوها و ناهنجاری‌هایی را که ممکن است نشان‌دهنده حمله سایبری باشد، شناسایی کند. این می‌تواند به شناسایی فعالیت‌های مخرب، بدافزار یا رفتار شبکه مشکوک کمک کند و امکان شناسایی و پاسخ زودهنگام را فراهم کند (آراجو، کوسیرو، سیفرت، سارمنتو، و دیویدز[1]، ۲۰۲۱؛ پاپاسوراچا، نیکولداکسی، کفالوکس، پالیس و مارکاکیس[2]، ۲۰۲۱؛ لیو و همکاران[3]، ۲۰۲۲).

پاسخ بی‌درنگ به رخدادها: هوش مصنوعی می‌تواند فرآیندهای واکنش به حادثه را خودکار انجام دهد و واکنش‌های سریع‌تر و کارآمدتر به نقض‌های امنیتی را ممکن می‌سازد. می‌تواند داده‌ها را از منابع مختلف تجزیه‌وتحلیل و مرتبط کند، تهدیدها را اولویت‌بندی کند و اقدامات فوری برای کاهش خطرات را آغاز کند (آراجو و همکاران، ۲۰۲۱؛ فرجی، شیکدر، حسان، اسلام، و اختر[4]، ۲۰۲۴).

دفاع تطبیقی: هوش مصنوعی می‌تواند از حملات قبلی درس بگیرد و به‌طور مداوم مکانیسم‌های دفاعی خود را تکامل دهد. بامطالعه الگوها و تکنیک‌های حمله، سیستم‌های هوش مصنوعی می‌توانند توانایی خود را

1- Araújo, Couceiro, Seifert, Sarmento, & Davids
2- Papatsaroucha, Nikoloudakis, Kefaloukos, Pallis, & Markakis
3- Liu & etal
4- Faraji, Shikder, Hasan, Islam, & Akter

برای شناسایی و پیشگیری از تهدیدات آینده تطبیق داده و بهبود بخشند و دور زدن اقدامات امنیتی را برای مهاجمان دشوارتر می‌کنند (روباعباس، پیت، وگل، زفیروکوپولوس[1]، ۲۰۲۳؛ لیو و همکاران، ۲۰۲۲؛).

کاهش: سیستم‌های امنیتی سنتی اغلب هشدارهای خطر اشتباه تولید می‌کنند که منجر به اتلاف زمان و منابع می‌شود. الگوریتم‌های هوش مصنوعی می‌توانند از بازخوردهای انسانی بیاموزند و قابلیت‌های تشخیص خود را اصلاح کنند، میزان کاهش را به حداقل برسانند و به تیم‌های امنیتی اجازه دهند روی تهدیدات واقعی تمرکز کنند (فرجی، و همکاران، ۲۰۲۴).

آنالیز پیش‌بینی کننده: هوش مصنوعی می‌تواند داده‌های تاریخی و روندهای فعلی را برای پیش‌بینی آسیب‌پذیری‌های احتمالی و بردارهای حمله آینده تجزیه‌وتحلیل کند. این رویکرد پیشگیرانه به سازمان‌ها کمک می‌کند تا با رفع آسیب‌پذیری‌ها قبل از سوءاستفاده، موقعیت امنیتی خود را تقویت کنند (روباعباس و همکاران، ۲۰۲۳؛ فرجی و همکاران، ۲۰۲۴).

خطاهای انسانی کمتر: برخلاف کارمندان انسانی، یادگیری ماشینی بدون اینکه خسته شود، محافظت کاملی را در ۷ روز هفته و ۲۴ ساعته ارائه می‌کند. به‌علاوه، می‌تواند از مشاهدات و نتایج به‌دست‌آمده، بیاموزد تا به‌سرعت و دقت عملکرد خود بیفزاید. این بدان معناست که تیم‌های امنیتی می‌توانند عملیات بیشتری را به الگوریتم‌های بسیار آموزش‌دیده محول کنند و خطاهای انسانی را که به‌راحتی می‌توان از آن اجتناب کرد، کاهش داد.

به این نکته نیز باید توجه داشت که مداخله انسانی هنوز مورد نیاز است. سیستم‌های امنیت سایبری مبتنی بر هوش مصنوعی نمی‌توانند به طور کامل جایگزین متخصصان امنیتی شوند. استفاده از هوش مصنوعی در امنیت سایبری زمانی بهترین نتیجه را دارد که عنصر مداخله انسانی وجود داشته باشد. سیستم‌های هوش مصنوعی مانند یادگیری ماشینی به کارشناسان امنیتی کمک می‌کنند تا رفتارهای مخرب را شناسایی کرده و از وقوع آن جلوگیری کنند، که خطرات امنیتی و نقض داده‌ها را کاهش می‌دهد. اساساً، ابزارهای هوش مصنوعی و تیم‌های امنیت سایبری باید مکمل یکدیگر باشند. تیم امنیتی باید از فناوری هوش مصنوعی با به روز نگه داشتن آن در مورد هرگونه تهدید شناخته شده و بررسی منظم عملکرد بهینه سیستم پشتیبانی کند. به نوبه خود، سیستم‌های امنیتی هوش مصنوعی تیم‌های امنیت سایبری را در مورد هرگونه رفتار مشکوک یا مخرب، الگوهای ناشناخته یا دیده نشده قبلی در داده‌ها و هر گونه آسیب‌پذیری امنیتی احتمالی هشدار می‌دهند (بوساری[2]، ۲۰۲۵؛ ساميا[3]، ۲۰۲۳؛ کوواسی[4]، ۲۰۲۳).

کاربردها در حملات سایبری

علاوه بر تیم‌های مدافع[5]، حمله کنندگان و عاملین تهدید نیز در رابطه با بهبود سرعت، افزایش دقت در بررسی داده‌ها، افزایش کیفیت حمله و مدت زمان بقا و ماندگاری در شبکه از هوش مصنوعی در روند کارهای خود استفاده می‌کنند. پیش از شرح بیشتر و ارائه توضیحات تکمیلی در زمینه حوزه‌های بکارگیری، نیاز است مقدمه‌ای در رابطه با روند حملات بیان گردد که در ادامه به آن‌ها پرداخته شده است.

روند حملات

چندین متدولوژی و استاندارد معروف وجود دارد که می‌توان از آن‌ها برای انجام تست‌های نفوذ استفاده کرد، مانند: ISSAF، PTES، NIST، OWASP، OSSTMM. همه آن‌ها روندی از مراحل پیشبرد حملات سایبری را ارائه کردند که مراحل آن به ترتیب شرح داده می‌شود.

۱. تشخیص و جمع‌آوری اطلاعات: در طول این مرحله از حملات، عاملین تهدید سعی می‌کند با جمع‌آوری اطلاعات از منابع قابل‌دسترس برای کشف پورت‌ها و سرویس‌هایی که باز هستند، تا حد امکان اطلاعات بیشتری درباره اهداف خود جمع‌آوری کنند. در پایان این

1- Roba Abbas, Pitt, Vogel, & Zaferirakopoulos
2- Busari
3- Samia
4- Kovací
5- Blue Team

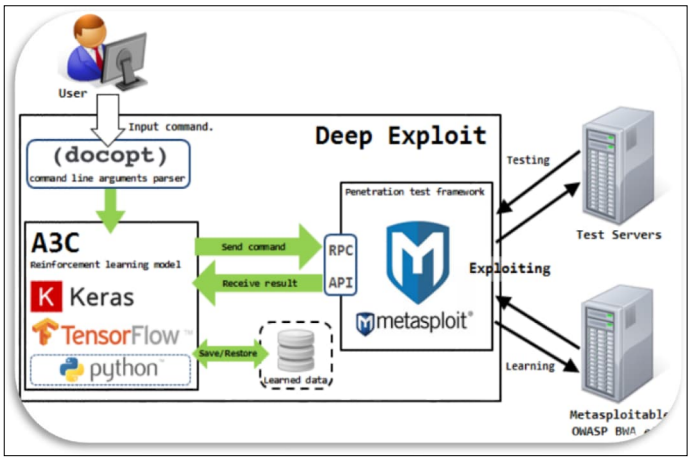
مرحله، عاملین تهدید دانشی از اهداف خود خواهند داشت که شامل اطلاعاتی مانند نام Domain، Hostهای هدف، سرویس‌های فعال، فناوری‌های موجود، نام کارمندان، ایمیل‌های کارکنان، مکان‌های فیزیکی، تصاویر مکان‌های فیزیکی، نام کاربری و رمز عبور و غیره است. هوش مصنوعی و یادگیری ماشین می‌توانند به عاملین تهدید کمک کنند تا نه تنها تمام اطلاعات را به‌صورت خودکار جمع‌آوری کنند، بلکه آن‌ها را تجزیه‌وتحلیل کنند و دوره‌های مختلف عملیات را تعیین کنند. برای مثال، می‌توانند بر اساس اطلاعات جمع‌آوری‌شده بهترین حمله مهندسی اجتماعی را برای اجرا تعیین کنند. یا می‌توان از آن برای شناسایی میزبان‌های هدفی که باید ابتدا مورد حمله قرار گیرند استفاده کرد زیرا احتمال موفقیت بیشتر است (بارسیا، ۲۰۲۵؛ سامیا، ۲۰۲۳؛ کوواسی، ۲۰۲۳).

بررسی و ارزیابی آسیب‌پذیری: در طول این مرحله، عاملین تهدید اسکن‌های آسیب‌پذیری عمیق‌تری را انجام می‌دهند تا تمام آسیب‌پذیری‌های بالقوه‌ای را که اهداف می‌توانند داشته باشد، مشخص کنند. در اینجا، هوش مصنوعی و یادگیری ماشین می‌توانند کمک کنند تا نتایج اسکن را با تجزیه‌وتحلیل آن و حذف هر چیزی که قابل‌اجرا نیست یا نویز ایجاد می‌کند، با در نظر گرفتن اطلاعات جمع‌آوری‌شده از مرحله قبل همراه با اطلاعات تهدید از منابعی مانند رسانه‌های اجتماعی، سوابق، دیپ و دارک وب درک کنند. همچنین، هوش مصنوعی و یادگیری ماشین می‌توانند با همبستگی همه اطلاعات و دانش جمع‌آوری‌شده، بهترین حالت و امکان حمله را تعیین نماید (بارسیا، ۲۰۲۵؛ روباعباس و همکاران، ۲۰۲۳؛ کوواسی، ۲۰۲۳).

بهره‌برداری: در این بخش سعی می‌شود عاملین تهدید بتواند به هدف دسترسی یابند، اختیارات کاربری را افزایش دهند، اطلاعات بیشتری جمع‌آوری کنند و دسترسی دائمی را حفظ کنند. هوش مصنوعی و یادگیری ماشین می‌توانند با تعیین بهترین مسیر عمل برای نفوذ به یک هدف کمک کنند، اما همچنین می‌توانند بهره برداری را به‌طور هم‌زمان انجام دهند. نتایج این بهره‌برداری‌ها می‌تواند به مدل هوش مصنوعی بازخورد داده شود و به آن اجازه می‌دهد جایگزین‌های بهره‌برداری یا مسیرهای بهره‌برداری جدیدی که قبلاً در نظر گرفته نشده‌اند، تولید کند (بارسیا، ۲۰۲۵؛ کوواسی، ۲۰۲۳؛ فرجی و همکاران، ۲۰۲۴).

در حال حاضر ابزارهای open source در بازار وجود دارد که اجرای سه مرحله اول این روش را ترکیب می کند، مانند Deep Exploit در نشانی زیر قابل مشاهده است:
https://github.com/13o-bbr-bbq/machine_learning_security/wiki#deep-exploit

این یک ابزار تست نفوذ کاملاً خودکار است که از یادگیری ماشین نه تنها برای بهبود مرحله جمع‌آوری اطلاعات بلکه برای سوء استفاده از آسیب‌پذیری‌ها استفاده می‌کند. در شکل شماره (۲) مشاهده می‌شود.



شکل (۲): مدل روند کار Deep Exploit

حوزه‌های بکارگیری

علاوه بر مواردی که در بخش روند حملات بیان و شرح داده شد، در این بخش حوزه‌ها و موارد دیگری را در این رابطه بیان خواهیم نمود. موارد جدید حالات پیشرفته و عمیق‌تری را در خود جای داده است.

مسموم سازی داده‌ها: هوش مصنوعی می‌تواند داده‌های آموزشی مورداستفاده توسط سیستم‌های یادگیری ماشین را برای تصمیم‌گیری دست‌کاری کند که با دو روش امکان‌پذیر است:

تزریق داده: عاملین تهدید می‌توانند داده‌های مخرب یا گمراه‌کننده را به مجموعه داده‌های حقیقی مورداستفاده برای آموزش مدل‌های هوش مصنوعی تزریق کنند. با افزودن نمونه‌های نادرست یا دست‌کاری شده، فرآیند یادگیری الگوریتم می‌تواند خراب شود. این عمل باعث به وجود آمدن مدل‌های مغرضانه یا نادرست می‌شود که پیش‌بینی‌ها یا تصمیم‌گیری‌های نادرست انجام می‌دهند. شبکه‌های متخاصم مولد GAN:(GANs)ها نوعی الگوریتم هوش مصنوعی هستند که جهت تولید داده‌های مصنوعی از یک مجموعه داده حقیقی از آن‌ها استفاده می‌شود. عاملین تهدید می‌توانند یک GAN را برای تولید داده‌های مصنوعی، اما با تغییرات ظریف در برچسب داده‌ها(به هرکدام از داده‌ها برچسب نامرتبط تخصیص داده شود) آموزش دهند. سپس این داده‌های مصنوعی مسموم را می‌توان با داده‌های حقیقی مخلوط کرد و باعث به خطر افتادن مدل هوش مصنوعی حاصل شد.

دستکاری یادگیری تقویتی: عاملین تهدید می‌توانند سیگنال‌های مربوط به پاداش، مجازات و محیط را دست‌کاری کنند تا سیستم را هنگام یادگیری به‌وسیله رفتارهای نامطلوب فریب دهند. با هدایت سیستم به سمت اقدامات مخرب، عاملین تهدید می‌توانند از سیستم هوش مصنوعی سوءاستفاده کنند (آگاروال[1]، ۲۰۲۵؛ پتل[2]، ۲۰۲۰).

بدافزار: بدافزار مجهز به هوش مصنوعی می‌تواند شناسایی و دفاع در برابر حملات را برای سیستم‌های امنیتی سنتی چالش برانگیزتر کند. بدافزار می‌تواند تکامل یابد و از محیط خود یاد بگیرد و ریشه‌کن کردن آن را دشوارتر کند. این نوع بدافزارها، از رفتار تطبیقی هدف استفاده می‌کند. این رفتار می‌تواند بدافزار را قادر به تطبیق و تکامل بر اساس محیط خود کند. می‌تواند از تعاملات خود با سیستم‌ها درس بگیرد، اقدامات امنیتی را تجزیه‌وتحلیل کند و منبع کدینگ خود را برای جلوگیری از شناسایی تغییر دهد. این امر، شناسایی و ریشه‌کن کردن بدافزار را برای آنتی ویروس یا دفاع شبکه چالش‌برانگیزتر می‌کند. الگوریتم‌های هوش مصنوعی می‌توانند حجم وسیعی از داده‌ها، ازجمله نمایه‌های رسانه‌های اجتماعی، عادت‌های جستجو و فعالیت‌های آنلاین هدف را برای ایجاد حملات شخصی‌سازی‌شده تجزیه‌وتحلیل کنند. با تنظیم بدافزار به‌طور خاص بر اساس رفتار و علایق هدف، شانس نفوذ موفقیت‌آمیز و سازش را افزایش می‌دهد. همچنین هوش مصنوعی به بدافزار کمک می‌کند تا آسیب‌پذیری‌ها را در سیستم‌ها شناسایی کرده و به طور مؤثر از آن‌ها سوءاستفاده کند، همچنین الگوهای ترافیک شبکه، تنظیمات سیستم و رفتار کاربر را تجزیه و تحلیل کند، تا بهترین زمان و روش برای شروع حمله را تعیین کند. این بدافزار را قادر می‌سازد تا اقدامات امنیتی را دور بزند و به اطلاعات حساس دسترسی غیرمجاز پیدا کند (آگاروال، ۲۰۲۵؛ فرجی و همکاران، ۲۰۲۴).

مهندسی اجتماعی: این امکان وجود دارد که هوش مصنوعی را با تکنیک‌های مهندسی اجتماعی برای دست‌کاری و فریب افراد ترکیب نمود. هوش مصنوعی می‌تواند حجم وسیعی از داده‌ها را برای ایجاد ایمیل‌ها، پیام‌ها یا تماس‌های تلفنی برای حملات فیشینگ به صورت شخصی‌سازی‌شده و باقابلیت قانع‌کنندگی بالا تجزیه‌وتحلیل کند. پیام ایجادشده توسط این سیستم قانع‌کننده‌تر به نظر برسند و احتمال گرفتار شدن قربانیان را افزایش دهند. هوش مصنوعی حتی می‌تواند آدرس‌های ایمیل واقعی یا شخصیت و هویت جعلی را برای فریب دادن افراد ایجاد کند. همچنین این قابلیت را دارد تا صداها و الگوهای گفتار انسان را تقلید کند، و امکان ایجاد

1- Aggarwal
2- Patel

حملات فیشینگ صوتی پیچیده را فراهم کند. با استفاده از ضبط‌های صوتی تولیدشده توسط هوش مصنوعی، عاملین تهدید می‌توانند هویت افراد یا سازمان‌های مورد اعتماد را جعل کنند و تلاش‌های مهندسی اجتماعی آن را متقاعدکننده‌تر و باورپذیرتر کنند. این نوع سیستم می‌توانند پست‌ها، نظرات و تعاملات رسانه‌های اجتماعی را برای شناسایی اهداف بالقوه برای مهندسی اجتماعی بررسی نمایند. با درک علایق، باورها و نقاط ضعف افراد، عاملین تهدید می‌توانند پیام‌های مناسبی برای سوءاستفاده از احساسات یا دست‌کاری اهداف ایجاد کنند. ازاین‌رو می‌توانند میزان موفقیت حملات را افزایش دهند و به جمع‌آوری اطلاعات حساس کمک کنند (آگاروال، ۲۰۲۵؛ لیو و همکاران، ۲۰۲۲؛ فرجی و همکاران، ۲۰۲۴).

DDoS: این امکان وجود دارد که با استفاده از الگوریتم هوش مصنوعی، حملات DDoS صورت پذیرد. سیستم می‌تواند الگوهای ترافیک شبکه را مشاهده و بررسی نماید، با سیستم دفاعی هدف سازگار شود، و موج عظیمی از ترافیک مخرب را برای غلبه بر سیستم‌های هدف هماهنگ کند. این می‌تواند منجر به اختلال در سرویس، زیان مالی و آسیب به اعتبار سازمان‌ها شود. همچنین می‌تواند با مدیریت و کنترل بات‌نت‌ها، حملات DDoS را راه‌اندازی نماید. این الگوریتم می‌تواند هماهنگی و توزیع ترافیک حمله را بهینه کنند با کاهش اثر حمله، دفاع را چالش‌برانگیزتر نماید (آگاروال، ۲۰۲۵؛ لیو و همکاران، ۲۰۲۲).

شکستن گذرواژه: در صورتی که قدرت محاسباتی هوش مصنوعی با الگوریتم‌های پیشرفته ترکیب شود، به شکل مؤثرتری می‌توان انواع مختلف گذرواژه‌ها را شکست و یافت نمود. هوش مصنوعی می‌تواند الگوها، ساختارهای گذرواژه رایج را بیاموزد و حتی رفتار انسان را برای افزایش شانس شکستن موفقیت‌آمیز گذرواژه پیش‌بینی کند. از طرفی چنین سیستمی می‌تواند مقادیر زیادی از داده‌های مربوط به گذرواژه‌های نقل‌شده را برای شناسایی الگوها و روش‌های رایج مورداستفاده توسط افراد هنگام ایجاد گذرواژه را بررسی کند. با یادگیری موارد فوق، هوش مصنوعی می‌تواند استراتژی‌های هدفمند و مؤثرتری برای شکست گذرواژه، ازجمله حملات Dictionary، brute force یا حملات ترکیبی ایجاد کند. از دیگر قابلیت‌های این سیستم می‌توان به این نکته اشاره نمود که می‌تواند رفتار کاربر، اطلاعات شخصی و سایر داده‌های زمینه‌ای را برای پیش‌بینی دقیق گذرواژه‌های انتخاب‌شده توسط هدف را تحلیل و بررسی کند. الگوریتم‌های هوش مصنوعی با درک اینکه افراد چگونه فکر می‌کنند و گذرواژه ایجاد می‌کنند، می‌توانند حدس‌های دقیق‌تری ایجاد کنند و زمان موردنیاز برای شکستن گذرواژه را به میزان قابل‌توجهی کاهش دهند. از طرفی می‌توانند از سخت‌افزارهای تخصصی مانند GPU برای تسریع شکستن گذرواژه استفاده کنند. با استفاده از قدرت محاسباتی پردازنده‌های گرافیکی، هوش مصنوعی می‌تواند پردازش موازی را انجام دهد، چندین ترکیب گذرواژه را به‌طور هم‌زمان انجام دهد و روند شکستن گذرواژه را تا حد زیادی سرعت بخشد (پتل، ۲۰۲۰؛ سامیا، ۲۰۲۳؛ ورما، و ناوایا، ۲۰۲۵).

DeepFake: از فناوری هوش مصنوعی برای ایجاد ویدیوها یا ضبط‌های صوتی Deepfake متقاعدکننده استفاده می‌شود. با ترکیب الگوریتم‌های هوش مصنوعی و یادگیری ماشینی، می‌توان محتوای رسانه‌ای را برای فریب دادن افراد یا انتشار اطلاعات نادرست دست‌کاری و تولید نمود. در برخی موارد مشاهده‌شده است که با استفاده از Deepfake صدا و تصویر مدیران ارشد برخی شرکت‌ها یا مسئولان دولتی را تولید نموده و از آن به عنوان یک سخنرانی قانع‌کننده در حملات Phishing استفاده‌شده است. ازاین‌رو می‌توان برای باج‌خواهی، افترا یا حتی برای دست‌کاری افکار عمومی استفاده نمود (پتل، ۲۰۲۰؛ فرجی و همکاران، ۲۰۲۴).

مقایسه حملات دستی، اتوماتیک و مبتنی بر هوش مصنوعی

می‌توان تفاوت‌هایی را میان حالات دستی، اتوماتیک و مبتنی بر هوش مصنوعی در حملات سایبری و انجام تست‌های نفوذ مقایسه نمود. به همین جهت جدول شماره (۱) در رابطه با این زمینه تهیه شده است که مطالعه دقیق آن مفید است (کوواسی، ۲۰۲۳).

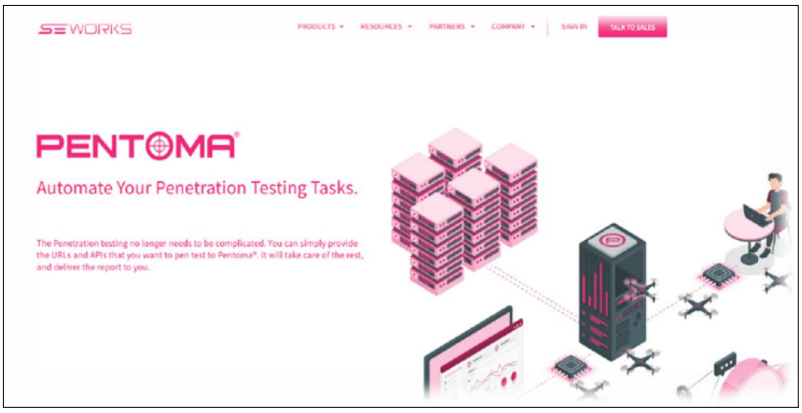
1- Verma, & Navya

جدول (۱): مقایسه پیشبرد حملات سایبری		
هوش مصنوعی	اتوماتیک	دستی
تست نفوذ فعال با هوش مصنوعی در مقایسه با تست خودکار دقیق است.	احتمال بازگشتFalse positive بیشتر است.	تست دستی همیشه به دلیل خطای انسانی دقیق نیست.
آزمایش با هوش مصنوعی زمان زیادی را صرف نمی‌کند. الگوریتم‌ها را می‌توان برای هزاران سیستم در یک لحظه مستقر کرد.	تست خودکار توسط ابزارهای نرم‌افزاری اجرا می‌شود، بنابراین به‌طور قابل توجهی سریع‌تر از روش دستی است.	تست دستی زمان‌بر است و منابع انسانی را به خود اختصاص می‌دهد.
هوش مصنوعی باعث صرفه‌جویی در سرمایه‌گذاری برای منابع انسانی در تست نفوذ می‌شود. نسبتاً می‌توان از همان کارمندان برای انجام کارهای کمتر تکراری و کارآمدتر استفاده کرد.	سرمایه‌گذاری برای تست ابزار موردنیاز است.	سرمایه‌گذاری برای نیروی انسانی موردنیاز است.
تست نفوذ مبتنی بر هوش مصنوعی در سازمان‌هایی با هزاران سیستم که برای صرفه‌جویی در زمان و منابع نیاز به آزمایش هم‌زمان دارند، عملی است.	تست خودکار زمانی عملی است که ابزارها، آسیب‌پذیری‌های تست را خارج از محدوده قابل برنامه‌ریزی پیدا کنند.	تست دستی تنها زمانی عملی است که موارد تست یک یا دو بار اجرا شوند و نیازی به تکرار مکرر نباشد.

ابزارها

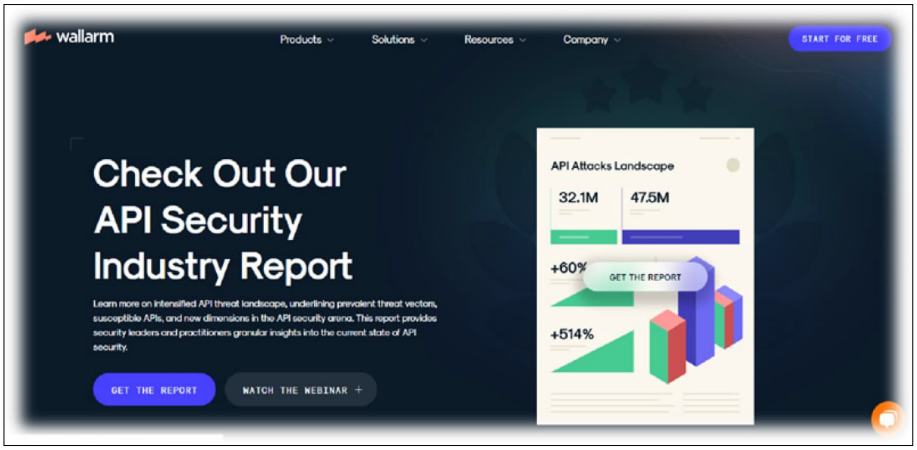
علاوه بر ابزار DeepExploit که پیش از این در بخش روند حملات به آن اشاره شد، ابزارها و APIهای دیگری نیز وجود دارند که به آن‌ها خواهیم پرداخت که هرکدام مزیت‌ها و معایب خود را دارند.

Pentoma: این ابزار با دریافت URL و APIها می‌تواند به‌صورت خودکار و مبتنی بر هوش مصنوعی تست نفوذ را بر برنامه‌ها و وب‌سرورها پیاده‌سازی نموده و نتایج به دست آمده به همراه نفوذپذیری‌های موجود را به کاربر تحویل دهد. نقاط حمله احتمالی را از دیدگاه مهاجم تجزیه و تحلیل می‌کند و تست‌های نفوذ را با شبیه‌سازی اکسپلویت‌ها انجام می‌دهد. تهیه این ابزار از طریق نشانی https://se.works به ثبت‌نام در سایت و انجام مذاکره از طریق مشاوران فروش قابل تهیه است.



شکل (۳): صفحه مربوط به وبگاه se.works مربوط به Pentoma

Wallarm: یکی دیگر از ابزارهای تست مبتنی بر هوش مصنوعی Wallarm است که منابع شبکه را کشف می‌کند، آسیب‌پذیری‌های رایج را اسکن می‌کند و واکنش برنامه‌ها را برای الگوهای رفتاری غیرعادی نظارت می‌کند. آسیب‌پذیری‌های خاص برنامه را به‌وسیله تائید خودکار تهدیدات کشف می‌کند. محتوای یک درخواست مخرب مسدود شده برای ایجاد یک تست پاک‌سازی‌شده با همان بردار حمله استفاده می‌شود تا ببینند برنامه یا کپی آن در جعبه ایمنی چگونه پاسخ می‌دهد. این ابزار به‌صورت رایگان در نشانی https://www.wallarm.com قابل تهیه است اما نیازمند به ثبت‌نام در سایت به‌وسیله ایمیل تجاری است.



شکل(۴): صفحه اصلی wallarm.com

Snyk DeepCode AI: از دیگر مواردی که می‌توان به آن اشاره نمود، پلتفرم DeepCode AI است. از طریق نشانی https://snyk.io قابل تهیه است. به‌وسیله Deepcode در حملات سایبری، می‌توان از قابلیت‌های آن برای یافتن آسیب‌پذیری‌ها در سیستم‌های نرم‌افزاری استفاده کنید. این پلتفرم می‌تواند نقاط ضعف و حفره‌های امنیتی را در برنامه‌ها، وب‌سایت‌ها یا شبکه‌های مختلف شناسایی کند. این به شما امکان می‌دهد از این نقاط ضعف سوءاستفاده کنید و دسترسی غیرمجاز به دست آورید، اطلاعات حساس را سرقت کنید یا به سیستم‌های هدف آسیب وارد کنید. علاوه بر یافتن آسیب‌پذیری‌ها، می‌توان از هوش مصنوعی Deepcode برای ساخت بدافزارهای پیچیده و اکسپلویت نیز استفاده نمود. با استفاده از این پلتفرم قدرتمند، می‌توان نرم‌افزار مخربی ایجاد کرد که می‌تواند به سیستم‌ها نفوذ کند، داده‌ها را بدزدد یا حتی کنترل کل شبکه‌ها را در دست بگیرد. این پلتفرم شامل سه پلن Free، Team و Enterprise است.

Shennina: یک پلتفرم خودکار Host exploitation است. مأموریت این پروژه خودکارسازی کامل اسکن، اسکن و تحلیل آسیب‌پذیری و سوءاستفاده از آن‌ها با استفاده از هوش مصنوعی است. Shennina به همراه Metasploit و Nmap برای انجام حملات یکپارچه‌شده است و همچنین با یک سرور C&C داخلی برای استخراج خودکار داده‌ها از ماشین‌های در معرض خطر یکپارچه‌شده است. این پروژه بر اساس DeepExploit توسعه‌یافته است. این ابزار در نشانی https://github.com/mazen۱۶۰/shennina در دسترس عموم قرار دارد و با سیستم‌عامل Kali سازگار است (چن، کانگ، اکسیونگ، و شوو[1]، ۲۰۲۴؛ ورما، و ناوایا[2]، ۲۰۲۵).

1- Chen, Kang, Xiong, & Shu
2- Verma, & Navya

۳- نتیجه‌گیری

با پیشرفت عصر اطلاعات، پیچیدگی و افزایش تهدیدات سایبری افزایش یافته است و چالش‌های مهمی را برای اقدامات سنتی امنیت سایبری ایجاد می‌کند. مجرمان سایبری از فناوری های پیشرفته برای اجرای حملات پیچیده تر استفاده می کنند و مکانیسم های دفاعی متعارف را ناکافی می‌کنند. در این محیط پر خطر، هوش مصنوعی (AI) به عنوان یک نیروی دگرگون کننده در حوزه امنیت سایبری ظاهر شده است. ظرفیت هوش مصنوعی برای پردازش و تجزیه و تحلیل مقادیر زیادی از داده‌ها در زمان واقعی، همراه با توانایی آن در شناسایی الگوها و ناهنجاری ها، آن را به عنوان ابزاری حیاتی در مبارزه با جرایم سایبری قرار می‌دهد. الگوریتم‌های یادگیری ماشینی می‌توانند رفتار غیرعادی را که ممکن است نشان‌دهنده نقض امنیتی باشد، اغلب قبل از اینکه آسیب قابل توجهی ایجاد شود، شناسایی کنند. سیستم‌های مجهز به هوش مصنوعی همچنین می‌توانند وظایف معمولی مانند نظارت بر ترافیک شبکه و مدیریت وصله‌ها را خودکار کنند و منابع انسانی را برای تمرکز بر مسائل پیچیده‌تر آزاد کنند. علاوه بر این، هوش مصنوعی با پیش‌بینی آسیب‌پذیری‌های بالقوه و بردارهای حمله، هوش تهدید را تقویت می‌کند و سازمان‌ها را قادر می‌سازد تا رویکردی پیشگیرانه به جای واکنشی برای امنیت اتخاذ کنند. با توجه به تحقیقات صورت گرفته، با پیشرفت تکنولوژی در زمینهی هوش مصنوعی و ML، می‌توان چه در زمینه‌ی دفاع و امنیت سایبری و چه در زمینه حملات، تولید باج افزار و انجام مهندسی اجتماعی و Phishing از هوش مصنوعی استفاده نمود. بااین‌حال در عمل به دلیل حساسیت حوزه‌ی حملات سایبری، ابزاری به دست نیامد؛ اما پلتفرم DeepCode AI را می‌توان به‌عنوان یک نمونه ابزار مفید در این زمینه معرفی کرد. از طرفی دیگر نیز ابزارها، پلتفرم‌ها و APIهای مختلفی در رابطه با حفظ امنیت در قبال حملات، رفتارهای غیرعادی، کلاه‌برداری و Phish وجود دارد که به کمک ادمین شبکه‌ها و تیم‌های امنیتی سازمان‌ها و شرکت‌ها می‌آیند. این نکته نیز قابل‌ذکر است که در صورت استفاده از ابزارهای مبتنی بر هوش مصنوعی، همچنان به تیم‌های امنیتی انسانی نیاز است و ابزارهای هوش مصنوعی تنها برای کارآمدتر شدن عملکرد تیم انسانی مورداستفاده قرار می‌گیرد.

منابع:

1-Aggarwal, G. (2025). Integrated Technologies in Electrical, Electronics and Biotechnology Engineering.
2-Araújo, D., Couceiro, M., Seifert, L., Sarmento, H., & Davids, K. (2021). *Artificial intelligence in sport performance analysis*. Routledge.
3-Arp, D., Quiring, E., Pendlebury, F., Warnecke, A., Pierazzi, F., Wressnegger, C., ... & Rieck, K. (2024). Pitfalls in Machine Learning for Computer Security. *Communications of the ACM*, *67*(11), 104-112.
4-Busari, M. (2025). Unlocking Emerging Technologies: The Role of Artificial Intelligence in Cyber security.
5-Chen, Z., Kang, F., Xiong, X., & Shu, H. (2024). A Survey on Penetration Path Planning in Automated Penetration Testing. *Applied Sciences*, *14*(18), 8355
6-Faraji, M. R., Shikder, F., Hasan, M. H., Islam, M. M., & Akter, U. K. (2024). Examining the role of artificial intelligence in cyber security (CS): A systematic review for preventing prospective solutions in financial transactions. *International Journal*, *5*(10), 4766-4782.
7-Gottsegen, G. (2023). *Machine Learning Cybersecurity: How It Works and Companies to Know*.
8-Kovací, P. D. (2023). THREAT ACTORS SEEKING TO EXPLOIT AI CAPABILITIES. TYPES AND THEIR GOALS. *Strategic Impact*, *89*(4), 53-63.
9-Liu, H., Wang, Y., Fan, W., Liu, X., Li, Y., Jain, S., ... & Tang, J. (2022). Trustworthy ai: A computational perspective. *ACM Transactions on Intelligent Systems and Technology*, *14*(1), 1-59.
10-Papatsaroucha, D., Nikoloudakis, Y., Kefaloukos, I., Pallis, E., & Markakis, E. K. (2021). A survey on human and personality vulnerability assessment in cyber-security: Challenges, approaches, and open issues. *arXiv preprint arXiv:2106.09986*.
11-Patel, R. (2020). Internet of Things (IoT): Cybersecurity Risks in Healthcare.
12-Pearlson, K. E., Saunders, C. S., & Galletta, D. F. (2024). *Managing and using information systems: A strategic approach*. John Wiley & Sons.
13-Roba Abbas, K. M., Pitt, J., Vogel, K. M., & Zaferirakopoulos, M. (2023). Artificial Intelligence (AI) in Cybersecurity: a Socio-Techni
14-Samia, N. K. (2023). *Global Cyber Attack Forecast using AI Techniques* (Master's thesis, The University of Western Ontario (Canada)).
15-Verma, S. S., & Navya, K. (2025). EYES EVERYWHERE: ETHICAL CONSIDERATION OF ARTIFICIAL INTELLIGENCE IN GOVERNANCE THROUGH FOUCAULT'S PANOPTICON LENS.
16-Wagner, D. (2017, October). Security and machine learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1-1).