بیلنت
مرکز نوآفرینی بیمه و مالی
Insurtech & Fintech Hub

چشم‌انداز فرصت‌ها و ریسک‌های سیستمی ناشی از هوش مصنوعی

مترجم:
سپیده فخارزاده

تهیه شده در:
واحد ارتباطات و ترویج نوآوری

زمستان
۱۴۰۳

بخش اول: مقدمه

هوش مصنوعی مولد دیگر یک مفهوم آینده‌نگر نیست. این فناوری در حال حاضر وجود دارد و در حال تحول سیستم‌هایی است که محتوای منحصربه‌فرد تولید می‌کنند، از جمله متن، تصاویر و حتی موسیقی. بدون شک، Gen AI در حال تغییر نحوه کار ماست.

سازمان‌ها می‌توانند با استفاده از Gen AI در عملکردهای روتین کسب‌وکار، صنایع خود را به شکل اساسی متحول کنند. به عنوان مثال، در صنعت بیمه، Gen AI می‌تواند به ارزیابی ریسک‌ها توسط کارشناسان کمک کند؛ این فناوری با تحلیل حجم زیادی از داده‌ها از جمله سوابق خسارت‌ها، اطلاعات مشتریان و عوامل امنیت سایبری داخلی/خارجی، می‌تواند پروفایل‌های ریسک را خلاصه کرده و به کارشناسان کمک کند تا پوشش‌های مناسب را سریع‌تر و با اطلاعات کامل‌تری توسعه دهند.

با این حال، فناوری هوش مصنوعی نیز خطرات جدیدی در زمینه امنیت سایبری به همراه دارد.

در حالی که هوش مصنوعی مولد می‌تواند بهره‌وری عملیاتی را بهبود بخشد، در عین حال فرصت‌هایی برای بازیگران مخرب فراهم می‌کند تا از قابلیت‌های آن برای حملات سایبری سوءاستفاده کنند.

در مقاله پژوهشی آگوست ۲۰۲۴ با عنوان "هوش مصنوعی: عامل چندوجهی ریسک تجمیع سایبری" که توسط مرکز برتری سایبری گای کارپنتر[1] و مرکز اطلاعات ریسک سایبری مارش مک‌لنن[2] نوشته شده است، چهار پویایی جدید بررسی می‌شود که از طریق آنها استفاده از هوش مصنوعی می‌تواند منجر به ریسک تجمیع سایبری شود:

۱- هوش مصنوعی به‌عنوان تهدید زنجیره تأمین نرم‌افزار

سازمان‌هایی که از هوش مصنوعی استفاده می‌کنند ممکن است به راه‌حل‌های ارائه‌شده توسط اشخاص ثالث مانند چت جی‌پی‌تی وابسته شوند، که در این صورت به خطر افتادن مدل ارائه‌دهنده می‌تواند به یک نقطه شکست واحد برای تمام مشتریانی که از آن مدل استفاده می‌کنند، تبدیل شود.

۲- هوش مصنوعی یک سطح حمله جدید ایجاد می‌کند

هنگامی که هوش مصنوعی مستقر می‌شود، کاربران می‌توانند با مدل تعامل داشته باشند. چه این مدل یک چت‌بات باشد، یک ابزار پردازش خسارت یا یک مدل تحلیل تصویر سفارشی، فرآیند دریافت ورودی و ارسال خروجی در معرض

[1] Guy Carpenter's Cyber Center of Excellence

[2] Marsh McLennan's Cyber Risk Intelligence Center

دستکاری‌های مخرب یا حتی تصادفی قرار می‌گیرد.

۳- هوش مصنوعی تهدیدی برای حریم خصوصی داده‌ها محسوب می‌شود

یک مدل تنها به اندازه داده‌هایی که بر اساس آنها آموزش دیده خوب عمل می‌کند. برای آموزش این مدل‌ها، باید به مجموعه داده‌های مرتبط، که اغلب بزرگ و حساس هستند، دسترسی داده شود. به خطر افتادن ذخیره‌سازی متمرکز این مجموعه داده‌ها می‌تواند اثرات جدی و گسترده‌ای داشته باشد.

۴- هوش مصنوعی در نقش‌های امنیتی

یکی از موارد استفاده بسیار مطرح‌شده برای هوش مصنوعی، عملیات امنیت سایبری است، به‌ویژه در رویه‌هایی که نیاز به سطح بالایی از دسترسی دارند، مانند به‌روزرسانی نرم‌افزار معیوب اخیر کراود استرایک[۳]. هنگامی که تصمیمات حساس پاسخ‌دهی به هوش مصنوعی واگذار می‌شود، ممکن است احتمال بروز خطاها یا سوءپیکربندی‌ها افزایش یابد و در نتیجه خطرات بیشتری ایجاد شود.

در حالی که آن گزارش این خطرات را از دیدگاهی مفهومی و آینده‌نگر بررسی می‌کند، این گزارش به‌عنوان مکمل عمل کرده و بر جنبه‌های فنی و تحلیلی در حال تحول تأثیرات هوش مصنوعی تمرکز دارد.

با درک خطرات احتمالی ناشی از انباشت ریسک مرتبط با هوش مصنوعی، برای صنعت بیمه و بیمه اتکایی مهم است که به آینده نگاه کرده و مسیری تحلیلی برای اندازه‌گیری این ریسک ایجاد کند، در عین حالی که از جنبه‌های مثبت هوش مصنوعی بهره می‌گیرد. در همکاری با شرکت سایبرکیوب[۴] که پیشرو در مدل‌سازی ریسک سایبری است، این مطالعه چارچوبی برای کمی‌سازی ریسک سیستمیک ارائه می‌دهد و سپس دو نمونه فرضی را به‌عنوان الگوهایی برای یک حمله سایبری تقویت‌شده با هوش مصنوعی بررسی می‌کند.

بخش دوم: بررسی یک چارچوب تحلیلی برای کمی‌سازی ریسک هوش مصنوعی

پیامدهای رویدادهای فاجعه‌آمیز سایبری

برای درک پیامدهایی که فناوری‌های هوش مصنوعی ممکن است بر رویدادهای فاجعه‌آمیز سایبری داشته باشند، اثراتی که هوش مصنوعی احتمالا بر اجزای اصلی مدل‌های فاجعه‌آمیز سایبرکیوب خواهد داشت را از منظر زنجیره حمله[۵] بررسی می‌کنیم.

[3] CrowdStrike
[4] CyberCube

[5] Kill Chain

با اســـتفاده از مراحل زنجیره حمله، حوزه‌های تحقیقاتی با اجزای مدل و مناطق حمله‌ای که هوش مصـــنوعی ممکن اســـت بر رویدادهای فاجعه‌آمیز ســایبری آینده تأثیر بگذارد، هماهنگ می‌شوند. تحقیقات اولیه نمونه‌هایی را در سراسر زنجیره حمله نشـــان داده‌اند که در آنها هوش مصـنوعی می‌تواند در حملات اسـتفاده شـود. در این مقاله، ما بر مناطقی که دارای اثبات مفهومی هســـتند تمرکز خواهیم کرد تـا اهمیـت آنهـا را برجسته کنیم.

استفاده از مدل Kill Chain

رویکرد Kill Chain، که توســـط لاکهید مارتین معرفی شده است، به شناسایی مراحل ممکن در یک حمله کمک می‌کند و امکان تجزیه‌وتحلیل ویژگی‌های حمله را در یک چارچوب مشـــترک فراهم می‌سازد. برای اهداف این تمرین، چارچوب Kill Chain به ما اجازه می‌دهد تا شناسایی کنیم کـه تـأثیر هوش مصـــنوعی در کـدام بخش از چشـم‌انداز تهدیدات سـایبری بیشـتر احسـاس می‌شود. مدل Kill Chain شامل ۷ مرحله است کـه ســـایبرکیوب مرحلـه هشـــتمی بـا عنوان «اقدامات پس از اهداف»[6] به آن اضـــافه کرده است. این مراحل به سه گروه تقسیم می‌شوند: پیـش از نـفـوذ (pre)، پـس از نـفـوذ (post) و «اقدامات در راستای اهداف» (AoO).

حوزه‌های اصلی مدل

سـه مؤلفه اصـلی مدل سـایبرکیوب برای تحلیل هر رویداد عبارتند از:

- فرکانس: رویداد هر چند وقت یک‌بار رخ می‌دهد؟
- دامنه تأثیر: کدام شرکت‌ها به این رویداد متصل هستند و از آن تأثیر می‌پذیرند؟
- شـــدت: میزان خســـارت مـالی برای شرکت‌های تحت تأثیر چقدر است؟

این مؤلفه‌های مدل، با هم یک مجموعه رویداد ایجاد می‌کنند، تعیین می‌کنند کدام شـــرکت‌ها تحت تأثیر قرار گرفته‌اند و میزان خســـارت مالی برای هر شرکت آسیب‌دیده را مطابق با مؤلفه‌های هزینـه مربوط بـه هر رویـداد در مجموعـه رویـداد مشخص می‌نمایند.

تأثیرات فرکانس و دامنه تأثیر

فرکانس و دامنه تأثیر یک رویداد می‌تواند در هر دو بخش پیش از نفوذ[7] و پس از نفوذ[8] از مـدل Kill Chain بررسـی شـود. تحقیقات اولیه نشـان داده اســـت که افزایش احتمال وقوع رویدادهای بزرگ مقیـاس ممکن اســـت بـه دلیـل افزایش ســـرعت و توانایی عاملان تهدید با اســـتفاده از هوش مصنوعی ایجاد شود.

ابزارهـایی مـانـنـد مـدل‌هـای زبـان بزرگ (LLMs) نشان داده‌اند که امکان اجرای مهندسی اجتماعی

[6] Post-AoO
[7] pre-intrusion
[8] post-intrusion

با کیفیت بالاتر در مقیاس وسیع (مانند فیشینگ، دیپ‌فیک‌ها و غیره)، شناسایی سریع‌تر آسیب‌پذیری‌ها و احتمال افزایش دامنه تأثیر اولیه را فراهم می‌کنند. این موضوع ممکن است باعث شود که حملات با سرعت بیشتری به مقیاس قابل توجهی برسند، به این معنا که فرکانس رویدادهای فاجعه‌بار جهانی می‌تواند از طریق افزایش تعداد رویدادهای کوچک‌تری که به سطح اهمیت می‌رسند، افزایش یابد.

رویدادهایی که پیش از این به اندازه کافی مهم تلقی می‌شدند که در مدل‌سازی قرار گیرند، ممکن است از طریق همین روش‌ها دامنه تأثیر خود را افزایش دهند و میانگین تعداد شرکت‌های تحت تأثیر را بالا ببرند.

شواهد و تحقیقات

در سال جاری، نمونه‌های اولیه و گزارش‌های اولیه اطلاعات تهدیدات سایبری از شرکت ریکوردد فیوچر[9] نشان داده‌اند که استفاده از LLMها در حملات فیشینگ و مهندسی اجتماعیُ کارایی و اثربخشی مراحل شناسایی، تسلیح[10] و تحویل[11] حملات را افزایش داده است.

تحقیقات همچنین منجر به کشف روش‌هایی مانند تزریق دستورات[12] و دستکاری مدل‌های زبانی توسط مهاجمان شده است که از این روش‌ها برای دسترسی یا انجام اقدامات پس از نفوذ در شبکه‌ها استفاده می‌کنند

استقرار مدل‌های زبانی بزرگ (LLM) برای مشتریان

این موضوع نشان می‌دهد که خطر تنها از مهاجمانی نیست که از LLMها برای اهداف خود استفاده می‌کنند، بلکه خطر نفوذ به LLMها نیز می‌تواند باعث ایجاد رویدادهای «تهدید داخلی» برای شرکت‌هایی شود که LLMهای مشتری‌محور را مستقر می‌کنند.

تأثیرات فرکانس

رواج هوش مصنوعی مولد نسل جدید به‌عنوان ابزاری برای تقویت و گسترش روش‌های حمله مختلف، قطعاً فرکانس حملات را افزایش خواهد داد. هوش مصنوعی دارای مزیت ذاتی اتوماسیون است، حتی اگر از ابتدا در حمله گنجانده نشده باشد. علاوه بر این، هوش مصنوعی می‌تواند در طول حملات تکامل یابد و با یادگیری از تجربیات تلاش‌های قبلی خود، خود را تطبیق دهد.

این مزایا در مقایسه با ابزارهای حمله سنتی، مزایای منحصربه‌فردی به مهاجمان می‌دهند. این‌که آیا این ویژگی‌ها در نهایت منجر به افزایش تناوب حملات موفقیت‌آمیز می‌شوند یا خیر، به موفقیت مدافعان در توسعه و استقرار

[9] Recorded Future
[10] weaponization
[11] delivery
[12] prompt injection

دفاع‌هایی مبتنی بر استراتژی‌های سنتی و رویکردهای دفاعی مبتنی بر هوش مصنوعی بستگی دارد.

مزیت مدافعان در برابر مهاجمان

به‌طور کلی، انتظار می‌رود مدافعان نسبت به عاملان تهدید مزیت مشخصی داشته باشند، به‌ویژه به این دلیل که توسعه‌دهندگان قانونی ابزارهای دفاعی معمولاً به فناوری‌های برتر هوش مصنوعی و داده‌های آموزشی گسترده‌تر دسترسی دارند.

با این حال، همه مدافعان منابع یا انگیزه لازم برای استفاده از این فناوری پیشرفته را ندارند. بنابراین، منطق نشان می‌دهد که احتمالاً افزایش خالصی در فرکانس حملات موفقیت‌آمیز به سازمان‌های دارای منابع و آمادگی کم به وجود خواهد آمد.

عوامل تأثیرگذار و روندهای پیش‌بینی‌شده

پیش‌بینی تأثیر هر روند به‌دلیل متغیرهای متعدد دشوار است، به‌ویژه با توجه به اینکه پیشرفت‌ها در حوزه هوش مصنوعی پویا و نامطمئن هستند.

احتمالاً سازمان‌های بزرگ‌تر و با منابع بیشتر یا آمادگی بهتر، شانس بیشتری برای کاهش خطرات سایبری خود با استفاده از مکانیسم‌های دفاعی مبتنی بر هوش مصنوعی خواهند داشت. در مقابل، شرکت‌های کوچک‌تر و کم‌منابع احتمالاً در برابر این روش‌ها و حملات جدید آسیب‌پذیرتر خواهند بود. این امر همچنین به احتمال زیاد باعث افزایش تفاوت در تأثیرات ممکن میان سازمان‌های مختلف، حتی در یک صنعت یا اندازه مشابه می‌شود.

داده‌ها و روندهای چشم‌انداز تهدید سایبری

داده‌های موجود در مورد تهدیدات سایبری نشان می‌دهند که روندهای مربوط به فرکانس رویدادها اغلب به‌صورت «موجی» است. این بدان معناست که افزایش فرکانس رویدادها معمولاً با کاهش نسبی در فرکانس همراه می‌شود. این روند معمولاً به این دلیل رخ می‌دهد که روش‌ها و تکنیک‌های حمله جدید با پیشرفت در روش‌ها و قابلیت‌های دفاعی مقابله می‌شوند.

نمونه‌ای از این روند

نمونه‌ای از این روند، افزایش حملات باج‌افزاری است که با بستن برخی پورت‌ها و پیشرفت در الزامات و استانداردهای پشتیبان‌گیری با آن‌ها مقابله شد. پیش‌بینی می‌شود این روند با پیشرفت‌های بیشتر در دفاع ادامه پیدا کند

هوش مصنوعی و تأثیر آن بر چرخه پیشرفت حملات و دفاع

اما پیشرفت‌های هوش مصنوعی، با کاهش زمان بین قله‌های موج (افزایش و کاهش فرکانس)، نشان می‌دهند که مهاجمان و مدافعان با سرعت بیشتری از یکدیگر یاد

می‌گیرند و سازگار می‌شوند. این اثرات ممکن است یکدیگر را خنثی کنند، به‌ویژه با بلوغ فناوری.

قانون مور (اصلی که می‌گوید سرعت و توانایی رایانه‌ها هر دو سال دو برابر می‌شود) و شتاب کلی در پیشرفت نیمه‌رساناها، نمونه‌ای از محدودیت‌های فیزیکی است که برای فناوری ترانزیستورها وجود دارد. احتمالاً پیشرفت فناوری هوش مصنوعی نسل جدید نیز از الگوی مشابهی پیروی خواهد کرد.

توانایی نوآوری و بهبود ممکن است در برخی نقاط کاهش یابد، که به دوره‌هایی با کاهش پیشرفت در حملات و دفاع و آرامش نسبی منجر می‌شود.

پیش‌بینی‌های فرکانس و اثربخشی

- پیش‌بینی فرکانس: احتمال افزایش تفاوت میان نهادهایی که از هوش مصنوعی در دفاع استفاده می‌کنند در مقابل آن‌هایی که استفاده نمی‌کنند.
- پیش‌بینی اثربخشی: افزایش نوسانات در نتیجه مقابله میان دفاع و حمله.

تأثیرات دامنه تأثیر (Footprint)

همان‌طور که پیش‌تر بحث شد، بهبود روش‌های حمله با استفاده از هوش مصنوعی باعث افزایش اثربخشی و کارایی حملات در مراحل پیش از نفوذ در چرخه Kill Chain سایبری می‌شود. مهاجمان تهدید می‌توانند به تعداد بیشتری از اهداف حمله کنند و این کار را به‌طور مقرون‌به‌صرفه‌تر انجام دهند، که انتظار می‌رود نرخ موفقیت افزایش یابد (با تمرکز بیشتر بر سازمان‌های ضعیف‌تر)، و در نتیجه دامنه تأثیر تهدیدهای سایبری گسترش یابد.

تأثیرات در مراحل پس از نفوذ

علاوه بر این، انتظار می‌رود هوش مصنوعی تأثیر قابل‌توجهی بر اثربخشی مراحل پس از نفوذ در چرخه Kill Chain داشته باشد. هوش مصنوعی می‌تواند توانایی مهاجمان را در موارد زیر افزایش دهد:

۱-شناسایی اهداف[۱۳]

۲-حرکت جانبی در شبکه[۱۴]

۳-افزایش سطح دسترسی[۱۵]

۴-تلاش برای فرار از شناسایی نفوذ

این پیشرفت‌ها در مراحل پس از نفوذ احتمالاً به مهاجمان امکان می‌دهند که تعداد بیشتری از دارایی‌ها را با نرخ آلودگی بالاتر به خطر بیندازند، که به خسارات بالقوه بزرگ‌تری منجر می‌شود.

تأثیرات بالقوه

۱) افزایش تعداد دارایی‌های در معرض خطر

[13] Target Enumeration
[14] Lateral Movement
[15] Privilege Escalation

۲) حجم بیشتر داده‌های افشا شده در حملات رخنه داده‌ها
۳) افزایش توان چانه‌زنی مهاجمان در مذاکرات باج‌گیری

این نتایج نشان می‌دهند که پیشرفت‌های هوش مصنوعی می‌توانند خطرات و خسارات بالقوه حملات سایبری را به‌طور چشمگیری افزایش دهند.

بهبودهای هوش مصنوعی در بردارهای حمله، اثربخشی و کارایی حملات را در مراحل پیش از نفوذ در چرخه Kill Chain سایبری افزایش خواهد داد.

افزایش کارایی در فاز اقدامات بر اهداف

انتظار می‌رود که در فاز اقدامات بر اهداف (AoO) نیز بهبودهای هوش مصنوعی باعث افزایش کارایی شود، به‌ویژه از طریق فرآیندهای مؤثرتر استخراج داده یا رمزگذاری داده، که به مهاجمان کمک می‌کند.

هوش مصنوعی همچنین می‌تواند فعالیت‌های پس از نفوذ و AoO را برای ابزارهای سنتی شناسایی و پاسخ بسیار دشوارتر کند، که این امر زمان ماندگاری حمله[16] را افزایش می‌دهد. زمان ماندگاری، یعنی مدت زمانی که مهاجم پیش از شناسایی و حذف در سیستم باقی می‌ماند، ممکن است تعیین‌کننده‌ترین عامل در میزان تأثیر حملات سایبری مانند رخنه داده یا باج‌افزار باشد.

همانند مراحل پیش از نفوذ، این پویایی‌ها ممکن است با به‌کارگیری هوش مصنوعی در دفاع سایبری مهار شوند. با این حال، این پیشرفت‌ها پیچیده و پیش‌بینی‌ناپذیر هستند و بسیاری از متغیرها و روابط علت‌ومعلولی در حال حاضر نامشخص‌اند.

پیش‌بینی تأثیر دامنه

احتمال افزایش تفاوت میان مدافعانی که از هوش مصنوعی استفاده می‌کنند در مقابل آن‌هایی که استفاده نمی‌کنند وجود دارد.

بدافزار

هوش مصنوعی به بهبود بهره‌برداری، فرماندهی و کنترل و اقدامات بر اهداف از طریق ایجاد بدافزارهای جدید کمک کرده است که می‌توانند آسیب‌پذیری‌های شناخته‌شده را با کارایی بیشتری مورد بهره‌برداری قرار دهند. این تغییرات می‌توانند انواع حملات سایبری را تحت تأثیر قرار دهند، از قطعی خدمات تا رخنه داده‌ها، بدافزار و باج‌افزار.

بدافزار چندشکلی[17]

یکی از کاربردهای پیشرفته‌تر هوش مصنوعی در حملات سایبری، جهش بدافزار از طریق چرخه Kill Chain است، که به‌عنوان بدافزار چندشکلی

[16] Dwell Time

[17] Polymorphic Malware

شـــناخته می‌شـــود. این بـدافزار می‌توانـد از فناوری‌های دفاعی رایج که از شناسایی الگوها یا روش‌های اکتشافی استفاده می‌کنند، فرار کند.

اگرچه این مفهوم کاملا جدید نیســت، تحقیقات از سـال ۲۰۱۹ نشـان می‌دهد که نمونه‌های اولیه بدافزار چندشـکلی توانایی بازنویسـی خود را برای فرار از آنتی‌بـدافزارهـای مبتنی بر روش‌هـای اکتشـافی بهبود داده‌اند. با کمک مدل‌های زبانی بزرگ (LLMs)، این قابلیت می‌توانـد در مقیاس گســترده برای انجام عملیات مخرب اســتفاده شود.

قابلیت‌های هوش مصنوعی در بدافزارها

مهـاجمـان می‌تواننـد فرآینـد جهش ویروس را خودکار کرده و به موارد زیر دست یابند:

۱) افزایش زمان ماندگاری، که ممکن اسـت شدت حمله را بیشتر کند.
۲) جهش مداوم برای فرار از شـــناســـایی با امضاهای ثابت.
۳) اتوماســیون یادگیری و فرآیندهای فرمان و کنترل برای گســترش ســریع‌تر، چـه به‌صورت خارجی و چه داخلی در شبکه‌ها.

این پیشـــرفت‌ها می‌توانند الگوریتم‌های جهش فعلی بدافزارها را که چندان پویا نیســتند، بهبود دهنـد. مـدل‌هـای زبانی بزرگ در مرکز این فرآینـد می‌توانند یاد بگیرند، همان‌طور که سـیسـتم‌های دفاعی یاد می‌گیرند و تلاش می‌کنند که از آن‌ها پیشی بگیرند.

قابلیت‌های حرکت جانبی و گســترش آلودگی[18] بـه‌ویژه در کمپین‌هـای بـاج‌افزار قـابـل اعمـال خواهند بود، که قصــد دارند برای کســب ســود بیشـتر، تعداد بیشـتری از سـیسـتم‌ها را هدف اخاذی قرار دهند.

کنترل و مذاکره توسط LLMها

اســتفـاده از مـدل‌هـای زبـانی بزرگ می‌توانـد فرآینـدهای فرمانـدهی و کنترل در شـــبکـه‌هـای قربانی و همچنین مقیاس پرداخت‌ها و مذاکرات را تسریع کند.

یـک نمونه اولیه به نام بلک‌ممبا[19] که توســط آزمایشگاه‌های HYAS توسعه یافته است، از یک مـدل زبـانی بزرگ برای «ســنتز عملکرد کی‌لاگر چندشکلی[20] در لحظه» استفاده کرده است.

این مثال‌ها نشـــان می‌دهند که بدافزارهایی که به‌طور خودکار تولید می‌شـــوند و چندشـــکلی هســـتنـد، احتمـالاً در طول زمـان اثربخشـــی و گسترش بیشتری پیدا خواهند کرد.

کاربردهای دفاعی هوش مصنوعی

اسـتفاده دفاعی از هوش مصـنوعی نسـل جدید توسط فروشندگان امنیت سایبری و سرویس‌های اطلاعات تهدید، توانایی مدافعان را در تشـخیص

[18] Infection Propagation
[19] BlackMamba
[20] Polymorphic Keylogger

بدافزارها از عملیات‌های عادی سیستم و شناسایی فعالیت‌های بالقوه مخرب با سرعت و دقت بیشتر افزایش خواهد داد.

این پیشرفت‌ها به‌طور گسترده‌ای مطرح و کمّی‌سازی شده‌اند، اما آزمایش مداوم این قابلیت‌ها در برابر کمپین‌های واقعی حمله برای درک اثربخشی آن‌ها ضروری خواهد بود.

تأثیر هوش مصنوعی بر بدافزار: افزایش احتمال زمان ماندگاری

رخنه داده‌ها

استخراج گسترده داده‌ها همواره برای مهاجمان چالش‌برانگیز بوده است. توانایی استخراج یا انتقال حجم زیادی از داده‌ها با نرخ بالا برای اخاذی و فروش، اغلب مانعی برای افزایش سودآوری حملات بوده است.

تحقیقات نشان داده‌اند که یادگیری ماشین می‌تواند استخراج داده را سریع‌تر و پنهانی‌تر کند، از طریق:

- کاهش اندازه فایل‌های استخراج‌شده
- خودکارسازی تحلیل انبوه داده‌ها برای شناسایی اطلاعات ارزشمند در میان داده‌های بی‌ارزش

این قابلیت‌ها می‌توانند منجر به رخنه داده‌های مؤثرتری شوند که اطلاعات حساس را سریع‌تر شناسایی کرده و تنها داده‌های ارزشمند را با سرعت بیشتری استخراج کنند. این امر به افزایش باج‌ها و افزایش فراوانی و شدت مسئولیت‌های قانونی منجر خواهد شد.

پیشرفت‌های دفاعی

هوش مصنوعی دفاعی نیز توانایی شناسایی فعالیت‌های استخراج و دسترسی غیرمجاز به داده‌ها را در مقیاس بزرگ افزایش داده است.

توسعه مداوم سیستم‌های شناسایی رفتاری و بازرسی بسته‌ها در مقیاس بزرگ (Zero Trust که دسترسی را تنها با احراز هویت مداوم ممکن می‌سازد) برای مقابله با پیشرفت‌های تهاجمی حیاتی است.

بخش سوم: بررسی تأثیرات هوش مصنوعی بر رویدادهای تاریخی

پس از بررسی نظری روش‌هایی که هوش مصنوعی می‌تواند فرکانس، دامنه و تأثیر حملات سایبری را تغییر دهد، اکنون دو مثال فرضی را به‌عنوان الگویی برای یک حمله سایبری مجهز به هوش مصنوعی بررسی می‌کنیم.

این مثال‌ها بر کاربرد هوش مصنوعی در بدافزارها و رخنه داده‌ها، همان‌طور که در بخش دوم معرفی شد، تمرکز خواهند داشت

مورد فرضی ۱: باج‌افزار ریوک

در سال‌های ۲۰۱۸ تا ۲۰۱۹، ریوک[۲۱] نوعی باج‌افزار بود که در بسیاری از کمپین‌ها علیه نهادهای بزرگ و عمومی با هدف کسب سود مالی از طریق رمزگذاری داده‌ها و پرداخت باج استفاده شد. در آن دوره، ریوک مسئول ۳ مورد از ۱۰ تقاضاهای بزرگ باجگیری بود: ۵/۳ میلیون دلار، ۹/۹ میلیون دلار، و ۱۲/۵ میلیون دلار.

روش‌های انتشار ریوک

ریوک از طریق روش‌های بسیار هدفمند منتشر می‌شد، که شامل موارد زیر بود:

۱-استفاده از ایمیل‌های سوءاستفاده از مهندسی اجتماعی[۲۲]

۲-بهره‌گیری از اعتبارنامه‌های به خطر افتاده برای دسترسی از راه دور به سیستم‌ها از طریق پروتکل دسکتاپ از راه دور (RDP)

روش ارسال این باج‌افزار اغلب از طریق ایمیل‌های اسپم بود که از آدرس‌های جعل شده استفاده می‌کردند تا شک و تردید ایجاد نکنند. در بسیاری از موارد، بدافزار ایموتت[۲۳]، که یک اسب تروجان بانکی بود، همراه با ریوک استفاده می‌شد.

با استفاده از RDP، مجرمان سایبری می‌توانستند ریوک را مستقیماً روی سیستم هدف نصب و اجرا کنند یا از دسترسی خود برای رسیدن به سیستم‌های ارزشمندتر در شبکه بهره بگیرند. بارگذار ایموتت حاوی مقادیر زیادی کد بی‌خطر به‌عنوان بخشی از تکنیک‌های فرار بود و می‌توانست سیستم‌های امنیتی را برای جلوگیری از شناسایی دستکاری کند.

تأثیر هوش مصنوعی بر باج‌افزارها

با استفاده از قابلیت‌های یادگیری ماشین، یک بدافزار چندشکلی می‌تواند بدون نیاز به مداخله انسانی، نسخه‌های جدیدی از کد خود را به‌صورت بازگشتی تولید کند. این کار با استفاده از مدل‌های هوش مصنوعی نسل جدید (مانند چت‌جی‌پی‌تی یا ابزارهای خاص‌تر) انجام می‌شود. این بدافزار می‌تواند به‌طور دوره‌ای نسخه تکامل‌یافته‌ای از کد مخرب خود را ایجاد کند که شناسایی آن سخت‌تر و فرّارتر باشد و از تکنیک‌هایی استفاده کند که ابزارهای امنیتی غالباً قادر به مقابله با آن نیستند.

این قابلیت می‌تواند مدت زمان آلودگی و آسیب‌های ناشی از آن را به‌شدت افزایش دهد. با این حال، بدافزاری که برای به‌روزرسانی کد خود نیاز به ارتباط با یک مدل هوش مصنوعی خارجی دارد، ممکن است توسط تیم‌های امنیتی راحت‌تر شناسایی شود.

[21] Ryuk
[22] Spear-Phishing
[23] Emotet

پیشرفت منطقی این استراتژی

یک توسعه منطقی در چنین استراتژی حمله‌ای ممکن است شامل تغییر به یک تکنیک LOTL[۲۴] باشد. در این روش، بدافزار به‌جای ارتباط با یک مدل خارجی، از یک مدل هوش مصنوعی داخلی برای انجام فعالیت‌های چندشکلی استفاده می‌کند.

به همین دلیل، مدافعان باید مدل‌های داخلی هوش مصنوعی نسل جدید و داده‌هایی که این مدل‌ها بر اساس آن‌ها آموزش می‌بینند را ایمن‌سازی کنند، به‌ویژه هر مدلی که در عملیات دفاع سایبری استفاده می‌شود.

علاوه بر این، هوش مصنوعی می‌تواند طراحی نسخه‌های جدید بدافزارها را برای مهاجمان آسان‌تر کند.

به‌جای صرف ماه‌ها برای ارتقاء بدافزار، آن‌ها می‌توانند از هوش مصنوعی برای آموزش مدل‌ها بر روی مجموعه‌داده‌های عظیم از نمونه‌های بدافزار استفاده کنند تا الگوها را یاد بگیرند و استراتژی‌های جدیدی برای جهش در بازه زمانی کوتاه‌تری ارائه دهند.

این مدل‌ها سپس می‌توانند به‌طور خودکار نسخه‌های جدیدی از بدافزار با ساختار کد تغییر یافته تولید کنند که به‌طور مؤثری یک قدم جلوتر از دفاع‌های امنیتی باقی می‌ماند.

نکته کلیدی

هوش مصنوعی می‌تواند کارایی بدافزارها را افزایش دهد و احتمال وقوع حوادث سایبری را بیشتر کند.

پیامدهای بدافزارهای چندشکلی مبتنی بر هوش مصنوعی عمیق است و می‌تواند در صورت عدم مدیریت ریسک، تأثیرات سیستماتیک بزرگی برای صنعت بیمه و بیمه اتکایی داشته باشد.

مورد فرضی ۲: نقض داده‌های اکوییفکس

در سال ۲۰۱۷، در زمان حمله، رخنه اطلاعات اکوییفکس[۲۵] دومین رخنه بزرگ تاریخ بود که بر ۱۶۳ میلیون رکورد در سراسر جهان تأثیر گذاشت، از جمله تقریباً نیمی از جمعیت ایالات متحده.

این رویداد تنها توسط نقض اطلاعات Yahoo در سال ۲۰۱۶ که همچنان بزرگ‌ترین حادثه نقض اطلاعات محسوب می‌شود، تحت‌الشعاع قرار گرفت.

با این حال، جامعیت و حساسیت داده‌های استخراج‌شده از اکوییفکس وزن بیشتری به شدت این رویداد می‌دهد. از سال ۲۰۱۷، چندین رخنه اطلاعات قابل‌توجه دیگر رخ داده است که از نظر تعداد رکوردهای تحت تأثیر، نقض

[24] Living Off the Land

[25] Equifax

اکوییفکس را پشـــت ســـر گذاشـــته‌اند (مانند مایکروسافت اکسچنج[۲۶] و فیسبوک).

تـداوم وقوع این حوادث نشـــان می‌دهـد کـه حملات اسـتخراج داده همچنان یک نگرانی بزرگ اســـت، به‌ویژه زمانی که با قابلیت‌های اضـــافی هوش مصـــنوعی و مدل‌های زبانی بزرگ ترکیب شوند.

نقش هوش مصنوعی در کشف آسیب‌پذیری‌ها

موفقیت‌های به‌دست‌آمده در رویدادهای دفاع از شـبکه DARPA و DEF CON در سـال ۲۰۱۶ ثابت کرده اســـت که هوش مصـــنوعی می‌تواند برای اسکن آسیب‌پذیری‌های ناشناخته استفاده شود.

(مانند حمله به مایکروسـافت اکسـچنج در سـال ۲۰۲۱)

حمله:

از هـوش مصـــنوعـی مـی‌تـوان بـرای یـافـتـن آســـیب‌پذیری‌های جدید و بهره‌برداری از آن‌ها در چندین هدف یا دفاتر اعتباری نیز اســـتفاده کرد، برخلاف حملـه هـدفمند به اکوییفکس. اضـــافـه شـدن قابلیت‌های هوش مصـنوعی برای یافتن و ســـوءاســـتفاده از آســـیب‌پذیری‌ها می‌تواند تأثیر رخنه‌هایی ماننـد اکوییفکس را تشـــدیـد کرده و مقیاس‌پذیری رویداد را افزایش دهد.

نقش هوش مصنوعی در افزایش تأثیر حملات

یک عامل کلیدی در حمله به اکوییفکس توانایی مهاجمان برای یافتن و اسـتفاده از اعتبارنامه‌های ناامن جهت دسترسی به ۴۸ پایگاه داده بود.

مدل‌های زبانی بزرگ می‌توانند فایل‌های حاوی اعتبارنامه‌های ناامن را با سرعت و دقت بیشـتری نسبت به مهاجمان انسانی شناسایی کنند.

در طول رخنه اکوییفکس، هکرها ۹,۰۰۰ کوئری به پایگاه داده‌ها ارسـال کردند که تنها ۲۶۵ مورد آن شامل اطلاعات شناسایی شخصی (PII) بود.

در صـورت اسـتفاده از هوش مصـنوعی، این ابزار می‌توانست با دقت بیشتری به دنبال PII بگردد و نمونه‌هـای احتمـالی داده‌هـای ارزشـــمنـد را برای مهاجمان برجســـته کند. این کار باعث افزایش گسـتردگی داده‌های اسـتخراج‌شـده از پایگاه‌های داده و انجام این فرآیند با کارایی بیشتری می‌شد.

نحوه خاتمه حمله

حمله به اکوییفکس با به‌روزرســـانی یک گواهی SSL منقضی‌شده به پایان رسید. این به‌روزرسانی بـه تیم امنیـت اطلاعـات اکوییفکس اجـازه داد ترافیک سـیسـتم را مشـاهده کرده و فعالیت‌های مشکوک را شناسایی کند.

تقریباً بلافاصـــله، تیم اکوییفکس متوجه ترافیک مشـکوک به آدرس‌های IP چینی و تحت مدیریت

[26] Microsoft Exchange

چین شـــدند و منجر بـه تعطیلی ســـرویس آسیب‌دیده شد.

نقش هوش مصنوعی در مخفی‌سازی فعالیت‌ها

هوش مصـنوعی می‌تواند ترافیک شـبکه را شـبیه به ترافیک قانونی جلوه دهد و از افزایش مشکوک فعالیت‌های شـــبکه که می‌تواند به‌عنوان یک پرچم هشـــدار توســـط ابزارهای امنیت داخلی شناسایی شود، جلوگیری کند.

مخفی کردن ترافیک خروجی شبکه تحت عملیات تجاری قانونی می‌توانســـت به راحتی مدت زمان حضـــور مهـاجم در شـــبکـه را در طول رخنـه اکوییفکس افزایش دهد.

نکته کلیدی:

افزودن ابزارهای هوش مصـنوعی می‌تواند به‌طور قابل‌توجهی اثربخشی یک گروه هکری را از طریق تسهیل حرکت جانبی مؤثرتر افزایش دهد و تأثیر آن را از نظر میزان و ســـطح داده‌های حســـاس استخراج‌شده گسترش دهد.

تقاطع هوش مصنوعی و آسیب‌پذیری‌های مداوم مـحـیـط نـرم‌افـزاری فـرصـــتـی بـرای رخ دادن نقض‌هایی از نوع اکوییفکس با مقیاس‌پذیری و شدت بیشتر ایجاد می‌کند

بخش ۴: نتیجه‌گیری و نگاه به آینده

این گزارش به‌طور انحصاری بر شناسایی خطرات ســـایبری ســـنتی متمرکز بود که می‌توانند با اســـتـفـاده از ابزارهـای هوش مصـــنوعی در کمپین‌های حمله تقویت شـــوند. جنبه دیگری از تأثیر هوش مصـنوعی بر ریسـک تجمعی پرتفوی بیـمـه، هـدف قـرار دادن خـود فـنـاوری هـوش مصـنوعی اسـت (یعنی هوش مصـنوعی به‌عنوان یک نقطه واحد شـــکســـت - SPoF). توجه به هوش مصنوعی به‌عنوان یک نقطه واحد شکست چالش‌برانگیز اســـت به دلیل ترکیب پیچیدگی هوش مصنوعی، طبیعت غیرقابل پیش‌بینی و در حـال تحول آن، نوپـایی نســـبی پـذیرش آن، وابســـتگی آن به داده‌ها و حیاتی بودن هوش مصـــنوعی و ســـیســـتم‌هایی که با آن یکپارچه شده‌اند برای عملیات‌های اصلی کسب‌وکار. درک تعـامـل این عوامـل، بـه‌علاوه عوامـل دیگر، تـأثیر زیادی بر امکان‌پذیری کمی‌ســـازی ریســـک قابل اعتمـاد از هوش مصـــنوعی به‌عنوان یک نقطه واحد شکست خواهد داشت.

این یـک موضـــوع پیچیـده اســـت کـه نیـاز بـه تحقیقات بیشـتر و رویکرد تحلیلی دقیق دارد. در حالی که هوش مصـــنوعی نســـل جدید به‌طور خاص در به‌روزرســـانی‌های مدل خســـارت‌های طبیعی سایبرکیوب (ALM) گنجانده شده است، مدل‌سازی حوادث فاجعه‌بار سایبری یک ملاحظه دیگر است. ارزیابی‌های دقیق انجام شده و ادامه خواهد داشـت تا مشـخص شـود که چه زمانی حوادث مربوط به نقاط واحد شـکسـت مبتنی بر هوش مصـــنوعی به ســـطح حوادث فاجعه‌بار ســـایبری می‌رســـند و بـه ایـن ترتیـب در

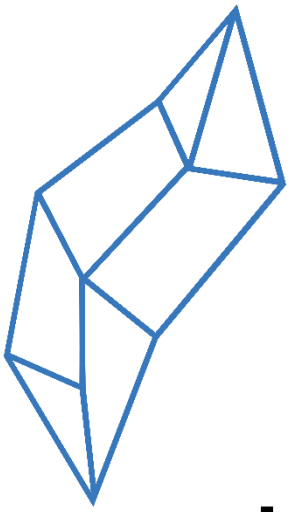
به‌روزرسانی‌های مدل ریسک فاجعه‌های سایبری گنجانده خواهند شد.

اگرچه بیشتر گفت‌وگوها در زمینه هوش مصنوعی در امنیت سایبری بر پیامدهای منفی آن متمرکز بوده است، باید از ذکر تأثیرات مثبت آن در این زمینه غافل نشویم. به‌طور خاص، تحقیقات اولیه نشان داده است که پیشرفت‌های امیدوارکننده‌ای در شناسایی و کنترل بدافزار، برچسب‌گذاری داده‌ها، نظارت و پیشگیری از خسارت به‌طور کلی وجود دارد.

هسته هر سیستم هوش مصنوعی، مانند مدل‌های زبان بزرگ یا مدل‌های یادگیری ماشین، داده‌هایی است که برای آموزش آن استفاده می‌شود. از این نظر، مدافعان دست برتری دارند. آن‌ها در زمین خانگی بازی می‌کنند، مدل‌های خود را بر اساس محیط خود آموزش می‌دهند و با فروشندگان همکاری می‌کنند تا تصویر کاملی از چشم‌انداز تهدید ایجاد کنند و راه‌حل‌های دفاعی سفارشی ایجاد کنند.

مهاجم تنها می‌تواند آنچه را که رو به بیرون است ببیند و بقیه را استنباط کند. همان‌طور که هوش مصنوعی به‌طور فزاینده‌ای در پلتفرم‌های امنیتی در سطح نقطه پایانی و پاسخ‌دهی و کشف تهدید (EDR/DR) و پلتفرم‌های امنیتی ابری یکپارچه می‌شود، این مدل‌ها به‌طور مداوم بر اساس داده‌های شبکه‌های مدافع و همچنین داده‌های تهدید فعلی آموزش می‌بینند. تأثیرات استفاده از هوش مصنوعی در مکانیزم‌های دفاعی نیز باید در چارچوب‌های مدل‌سازی سایبری آینده منعکس شود تا از ریسک بزرگ‌نمایی پیامدهای تهدیدات هوش مصنوعی جلوگیری شود.

همان‌طور که فناوری هوش مصنوعی به‌طور فزاینده‌ای در زندگی ما یکپارچه می‌شود، صنعت (باز) بیمه فرصت منحصر به فردی برای کمک به بیمه‌گذاران برای آماده شدن در برابر تهدیدات ناشی از هوش مصنوعی دارد.



پلنت

مرکز نوآفرینی بیمه و مالی
Insurtech & Fintech Hub

۰۹۹۹۹۱۹۰۲۲۵

www.plannet.ir

info@plannet.ir

شعبه ۱: کارخانه نوآوری آزادی

شعبه ۲: نمایشگاه بین‌المللی تهران