مطالعات علوم سیاسی، حقوق و فقه
دوره ۱۲، شماره ۱، بهار ۱۴۰۵
صفحات ۱۶۸۶-۱۶۷۱
Online ISSN: ۲۴۷۶-۴۴۶۹
Print ISSN: ۲۵۳۸-۲۰۳۹
www.irijournals.com

تأثیر هوش مصنوعی در ارتکاب جرایم رایانه ای

حجت فلاح خواجکینی[۱]*، حسین اصغری[۲]

[۱]دکتری حقوق جزا و جرم شناسی دانشگاه آزاد اسلامی واحد لاهیجان ایران (نویسنده مسئول)
[۲]دانشجوی کارشناسی ارشد حقوق کیفری و جرم شناسی، دانشگاه ازاد اسلامی، واحد بندرانزلی، بندرانزلی، ایران

چکیده

هوش مصنوعی در سالهای اخیر از یک فناوری تخصصی و محدود به سامانه های پژوهشی به زیرساختی عمومی و در دسترس تبدیل شده است. این تحول در کنار فرصتهای گسترده برای افزایش بهره وری بهداشت آموزش و امنیت ظرفیت های جدیدی برای ارتکاب تسهیل و گسترش جرایم رایانه ای نیز ایجاد کرده است. مجرمان سایبری میتوانند از سامانه های هوش مصنوعی برای تولید پیامهای فیشینگ متقاعد کننده جعل صدا و تصویر، شناسایی آسیب پذیری ها تولید یا تغییر بدافزار خودکارسازی حملات تحلیل داده های سرقت شده و فریب قربانیان استفاده کنند. در نتیجه مرز میان حملات سنتی و حملات هوشمند خودکار و مقیاس پذیر روز به روز کمرنگ تر می شود. پژوهش حاضر با روش توصیفی تحلیلی و با استفاده از مطالعات کتابخانه ای قوانین ایران اسناد بین المللی و گزارشهای نهادهای تخصصی تأثیر هوش مصنوعی بر جرایم رایانه ای را بررسی میکند. یافته ها نشان می دهد که هوش مصنوعی الزاماً ماهیت جدیدی برای همه جرایم ایجاد نمی کند، اما در بسیاری از موارد سرعت دقت مقیاس شخصی سازی و پنهان سازی رفتار مجرمانه را افزایش میدهد. همچنین استفاده از هوش مصنوعی موجب پیچیده تر شدن کشف جرم جمع آوری ادله دیجیتال انتساب رفتار به مرتکب و تعیین مسئولیت کیفری طراحان ارائه دهندگان و کاربران سامانه ها شده است. در حقوق ایران قانون جرایم رایانه ای و مقررات مرتبط ظرفیت هایی برای مقابله با رفتارهای مجرمانه ناشی از کاربرد هوش مصنوعی دارند؛ با این حال ابهام در مفهوم سامانه های خودکار فقدان مقررات جامع درباره دیپ فیک دشواری انتساب تصمیم ماشینی و ضعف هماهنگی نهادی نیاز به بازنگری تقنینی و سیاست گذاری منسجم را آشکار میکند. مقاله در پایان چارچوبی چند لایه شامل پیشگیری فنی حقوقی سازمانی آموزشی و بین المللی پیشنهاد می دهد.

کلیدواژه ها: هوش مصنوعی جرایم رایانه ای فیشینگ دیپ فیک بدافزار ادله دیجیتال مسئولیت کیفری امنیت سایبری

۱. مقدمه

تحول دیجیتال، ساختار زندگی اجتماعی، اقتصادی و اداری را دگرگون کرده است. ارتباطات، خدمات بانکی، تجارت، آموزش، حمل‌ونقل و بخش قابل‌توجهی از تعاملات عمومی به شبکه‌های رایانه‌ای و سامانه‌های اطلاعاتی وابسته شده‌اند. همین وابستگی، فرصت‌های جدیدی برای ارتکاب جرم ایجاد کرده است. جرایم رایانه‌ای در ابتدا عمدتاً به دسترسی غیرمجاز، تخریب داده‌ها، انتشار ویروس و سرقت اطلاعات محدود بود؛ اما امروزه با جرایمی همچون اخاذی سایبری، فیشینگ سازمان‌یافته، کلاهبرداری برخط، سرقت هویت، سوءاستفاده از رمزارزها، حملات زنجیره تأمین و انتشار محتوای جعلی مواجه هستیم(UNODC, 2013) .

هوش مصنوعی این روند را وارد مرحله تازه‌ای کرده است. سامانه‌های مولد می‌توانند متن، تصویر، صدا، ویدئو و کد رایانه‌ای تولید کنند. الگوریتم‌های یادگیری ماشین قادرند حجم عظیمی از داده‌ها را تحلیل کنند و الگوهای رفتاری، ارتباطی و امنیتی را بیابند. چنین قابلیت‌هایی در دست مجرمان، امکان اجرای حملات شخصی‌سازی‌شده، خودکار و کم‌هزینه را فراهم می‌سازد (Europol, 2023; Brundage et al., 2018).

در گذشته، بسیاری از حملات سایبری به دانش فنی قابل‌توجه و زمان طولانی نیاز داشتند. امروزه بخشی از این فرآیند با استفاده از ابزارهای هوش مصنوعی ساده‌تر شده است. مهاجم می‌تواند پیام‌هایی با زبان طبیعی و بدون خطاهای رایج تولید کند، قربانیان را بر اساس علایق و سوابق آنان دسته‌بندی نماید، آسیب‌پذیری‌های احتمالی را اولویت‌بندی کند و واکنش سامانه‌های دفاعی را تحلیل نماید. البته نباید نقش هوش مصنوعی را به‌صورت اغراق‌آمیز و قطعی تفسیر کرد؛ بسیاری از حملات همچنان به زیرساخت، تصمیم انسانی، دسترسی اولیه و سازمان‌دهی نیاز دارند. مسئله اصلی، ترکیب توانایی انسانی با قابلیت‌های خودکار و تحلیلی هوش مصنوعی است(Hadnagy, 2018) .

۲. بیان مسئله

جرم رایانه‌ای پدیده‌ای پویا است و همواره خود را با فناوری‌های جدید سازگار می‌کند. هوش مصنوعی از یک‌سو می‌تواند ابزار دفاعی مهمی برای شناسایی ناهنجاری‌ها، تحلیل رخدادها و پیش‌بینی حملات باشد و از سوی دیگر در خدمت مهاجمان قرار گیرد. این دوگانگی، یکی از ویژگی‌های اصلی فناوری‌های دوکاربردی است(Brundage et al., 2018) .

مسئله اساسی این پژوهش آن است که هوش مصنوعی چگونه بر عناصر، شیوه‌ها و پیامدهای جرایم رایانه‌ای تأثیر می‌گذارد. آیا این فناوری صرفاً ابزار جدیدی برای ارتکاب جرایم قدیمی است یا آنکه الگوهای مجرمانه کاملاً نو ایجاد می‌کند؟ استفاده از سامانه هوش مصنوعی تا چه اندازه می‌تواند موجب مسئولیت کیفری کاربر، طراح یا ارائه‌دهنده شود؟ در صورت تولید مستقل یا نیمه‌مستقل رفتار زیان‌بار، مسئولیت به چه شخصی منتسب خواهد شد؟ همچنین چگونه می‌توان اصالت ادله دیجیتال، منشأ یک محتوای جعلی و رابطه میان متهم و خروجی سامانه را اثبات کرد؟

اهمیت این پرسش‌ها زمانی بیشتر می‌شود که هوش مصنوعی مولد در اختیار عموم قرار گرفته است. در چنین شرایطی، مرز میان کاربر عادی، پژوهشگر امنیتی و مجرم حرفه‌ای همیشه روشن نیست. یک ابزار واحد ممکن است برای آزمایش امنیتی، آموزش، تولید محتوای مشروع یا ارتکاب جرم استفاده شود. بنابراین پاسخ حقوقی باید میان استفاده مشروع و سوءاستفاده مجرمانه تفکیک کند و از جرم‌انگاری افراطی یا مسئولیت مطلق پرهیز نماید(Chesney & Citron, 2019) .

۳. اهمیت و ضرورت پژوهش

اهمیت پژوهش در چند سطح قابل‌بررسی است:

۳-۱. اهمیت امنیتی

زیرساخت‌های حیاتی، سامانه‌های بانکی، شبکه‌های ارتباطی و خدمات عمومی هدف حملات پیچیده قرار دارند. استفاده از هوش مصنوعی می‌تواند سرعت شناسایی اهداف و مقیاس عملیات را افزایش دهد(ENISA, 2023; NCSC, 2024) .

۲-۳. اهمیت اقتصادی

جرایم سایبری موجب زیان مستقیم مالی، هزینه بازیابی سامانه‌ها، ازدست‌رفتن اعتماد مشتریان و توقف فعالیت اقتصادی می‌شوند. گزارش‌های بین‌المللی رشد مستمر هزینه‌های ناشی از جرایم سایبری را نشان می‌دهند (FBI, 2024; World Economic Forum, 2024).

۳-۳. اهمیت اجتماعی

جعل تصویر و صدا، انتشار اخبار ساختگی و سوءاستفاده از هویت اشخاص می‌تواند اعتماد عمومی را کاهش دهد و به حیثیت افراد، امنیت روانی جامعه و فرایندهای سیاسی آسیب برساند(Chesney & Citron, 2019) .

۴-۳. اهمیت حقوقی

قواعد سنتی مسئولیت کیفری بر رفتار انسان، قصد مجرمانه و رابطه سببیت استوار است. حضور سامانه‌های خودکار و مولد، مفاهیمی مانند عمد، علم، کنترل و قابلیت پیش‌بینی را با پرسش‌های جدید مواجه می‌کند(Russell & Norvig, 2021) .

۵-۳. اهمیت سیاست‌گذاری

نهادهای قانون‌گذار، پلیس، دادسرا، دادگاه، شرکت‌های فناوری و مراکز آموزشی نیازمند سیاستی هماهنگ هستند. اتخاذ تدابیر پراکنده و صرفاً واکنشی، اثربخشی لازم را نخواهد داشت(OECD, 2019) .

۴. اهداف و پرسش‌های پژوهش

۱-۴. هدف اصلی

تحلیل آثار هوش مصنوعی بر شیوه‌ها، ابعاد، پیامدها و مسئولیت‌های ناشی از ارتکاب جرایم رایانه‌ای و ارائه راهکارهای پیشگیرانه و حقوقی.

۲-۴. اهداف فرعی

- تبیین مفاهیم هوش مصنوعی و جرم رایانه‌ای؛
- بررسی نقش هوش مصنوعی در فیشینگ، دیپ‌فیک، بدافزار و حملات خودکار؛
- تحلیل دشواری‌های کشف جرم و انتساب رفتار؛
- بررسی ظرفیت قوانین ایران؛
- مطالعه رویکردهای بین‌المللی؛
- ارائه چارچوب چندلایه برای مقابله با جرایم هوش مصنوعی-محور.

۳-۴. پرسش‌های پژوهش

۱. هوش مصنوعی چگونه وقوع و مقیاس جرایم رایانه‌ای را تغییر می‌دهد؟

۲. مهم‌ترین کاربردهای هوش مصنوعی در ارتکاب جرم کدام‌اند؟

۳. چه چالش‌هایی در کشف و اثبات جرایم هوش مصنوعی-محور وجود دارد؟

۴. مسئولیت کیفری کاربر، طراح و ارائه‌دهنده سامانه چگونه قابل تحلیل است؟

۵. قوانین ایران و اسناد بین‌المللی تا چه اندازه پاسخ‌گوی این تحولات هستند؟

۶. چه چارچوبی برای پیشگیری و مقابله مؤثر پیشنهاد می‌شود؟

۵. روش‌شناسی پژوهش

این پژوهش از نظر هدف کاربردی و از نظر روش توصیفی-تحلیلی است. داده‌ها از طریق مطالعه قوانین، مقررات، کتاب‌ها، مقالات علمی، گزارش‌های رسمی و اسناد نهادهای تخصصی گردآوری شده‌اند. در تحلیل حقوقی از روش تفسیر نظام‌مند و مقایسه‌ای استفاده شده است؛ بدین معنا که مقررات ایران در ارتباط با اصول کلی حقوق کیفری، اسناد بین‌المللی و تجربه‌های سیاست‌گذاری سایر کشورها بررسی شده‌اند.

جامعه آماری پژوهش شامل کلیه قوانین و مقررات مرتبط با جرایم رایانه‌ای در ایران از جمله قانون جرایم رایانه‌ای (مصوب ۱۳۸۸)، قانون مجازات اسلامی (مصوب ۱۳۹۲)، قانون تجارت الکترونیکی (مصوب ۱۳۸۲)، قانون انتشار و دسترسی آزاد به اطلاعات (مصوب ۱۳۸۷) و قانون حمایت از حقوق پدیدآورندگان نرم‌افزارهای رایانه‌ای (مصوب ۱۳۷۹) می‌باشد.

پژوهش حاضر مدعی ارائه آمار مستقل درباره تعداد جرایم هوش مصنوعی-محور نیست؛ زیرا در بسیاری از نظام‌های آماری، استفاده از هوش مصنوعی به‌عنوان عامل جرم به‌طور جداگانه ثبت نمی‌شود. ازاین‌رو، تمرکز مقاله بر تحلیل روندها، قابلیت‌ها، ریسک‌ها و الزامات حقوقی است.

۶. مبانی مفهومی و نظری

۶-۱. نظریه فرصت جرم

بر اساس نظریه فعالیت روزمره(Cohen & Felson, 1979) ، وقوع جرم معمولاً به همزمانی سه مؤلفه وابسته است: **مجرم بالقوه، هدف مناسب** و **فقدان نگهبان مؤثر**. هوش مصنوعی می‌تواند هر سه مؤلفه را تحت تأثیر قرار دهد: هزینه ورود مجرم را کاهش دهد، اهداف مناسب را دقیق‌تر شناسایی کند و با فریب سامانه‌های دفاعی، نقش نگهبان را تضعیف نماید.

۶-۲. نظریه انتخاب عقلانی

بر اساس نظریه انتخاب عقلانی(Becker, 1968) ، مجرم با ارزیابی سود، هزینه و احتمال کشف تصمیم می‌گیرد. هوش مصنوعی با خودکارسازی و ناشناس‌سازی، هزینه حمله را کاهش داده و در برخی موارد احتمال شناسایی را کم می‌کند. در نتیجه، انگیزه ارتکاب برخی جرایم افزایش می‌یابد.

۶-۳. نظریه یادگیری اجتماعی

بخشی از دانش مجرمانه در جوامع برخط، انجمن‌ها و بازارهای زیرزمینی منتقل می‌شود. ابزارهای هوش مصنوعی می‌توانند آموزش، ترجمه و انتقال تجربه میان مجرمان را تسهیل کنند؛ هرچند همین ابزارها برای آموزش دفاع سایبری نیز قابل‌استفاده هستند (Bandura, 1977).

۶-۴. رویکرد سامانه‌های اجتماعی-فنی

جرم رایانه‌ای محصول تعامل انسان، فناوری، سازمان و محیط حقوقی است. بنابراین نباید مسئولیت و پیشگیری را فقط به الگوریتم نسبت داد. تصمیم‌های مدیریتی، ضعف کنترل دسترسی، فرهنگ سازمانی و نبود آموزش نیز در وقوع حمله مؤثرند (NIST, 2024a).

۷. تعریف هوش مصنوعی و جرایم رایانه‌ای

۷-۱. تعریف هوش مصنوعی

هوش مصنوعی مجموعه‌ای از روش‌ها و سامانه‌هایی است که برای انجام وظایفی طراحی می‌شوند که معمولاً به توانایی‌هایی مانند ادراک، یادگیری، استدلال، پیش‌بینی، تولید محتوا و تصمیم‌گیری نیاز دارند. یادگیری ماشین، یادگیری عمیق، پردازش زبان طبیعی، بینایی ماشین و سامانه‌های مولد از مهم‌ترین شاخه‌های آن هستند(Russell & Norvig, 2021) .

هوش مصنوعی مولد سامانه‌ای است که با استفاده از الگوهای آموخته‌شده، محتوای جدید مانند متن، تصویر، صدا، ویدئو یا کد تولید می‌کند. این سامانه‌ها معمولاً فاقد قصد و شخصیت مستقل هستند و خروجی آنان به طراحی، داده‌های آموزشی، دستور کاربر و محدودیت‌های فنی وابسته است(Europol, 2023) .

۲-۷. تعریف جرایم رایانه‌ای

جرایم رایانه‌ای به رفتارهای مجرمانه‌ای گفته می‌شود که در آن رایانه، سامانه رایانه‌ای، شبکه یا داده، موضوع، وسیله یا محیط ارتکاب جرم است. این جرایم شامل دسترسی غیرمجاز، شنود غیرمجاز، تخریب یا اخلال در داده‌ها و سامانه‌ها، سرقت و کلاهبرداری رایانه‌ای، جعل رایانه‌ای، نقض محرمانگی و انتشار محتوای مجرمانه می‌شود (قانون جرایم رایانه‌ای، ۱۳۸۸(.

جرایم رایانه‌ای را می‌توان به سه گروه تقسیم کرد:

۱. جرایم علیه محرمانگی، تمامیت و دسترس‌پذیری داده و سامانه؛
۲. جرایم رایانه‌ای سنتی که با ابزار دیجیتال ارتکاب می‌یابند؛
۳. جرایم محتوایی و ارتباطی که در فضای مجازی گسترش می‌یابند(UNODC, 2013) .

۸. تحول نقش هوش مصنوعی در ارتکاب جرم

نقش هوش مصنوعی در جرایم رایانه‌ای را می‌توان در چهار مرحله بررسی کرد:

مرحله نخست: ابزار کمکی

در این مرحله، هوش مصنوعی برای جست‌وجو، ترجمه، اصلاح متن یا تحلیل مقدماتی به‌کار می‌رود. نقش انسان همچنان غالب است.

مرحله دوم: خودکارسازی

برخی فعالیت‌ها مانند تولید پیام، دسته‌بندی اهداف، اسکن شبکه و تحلیل پاسخ قربانی خودکار می‌شوند.

مرحله سوم: شخصی‌سازی و مقیاس‌پذیری

هوش مصنوعی بر اساس داده‌های فردی، پیام‌ها و سناریوهای متفاوت تولید می‌کند. یک حمله می‌تواند همزمان هزاران قربانی را هدف بگیرد و برای هر کدام محتوایی متفاوت داشته باشد.

مرحله چهارم: سازگاری و تصمیم‌گیری نیمه‌خودکار

سامانه می‌تواند از نتایج قبلی یاد بگیرد و روش حمله را تغییر دهد. در این مرحله، نظارت انسانی کاهش می‌یابد، اما معمولاً هنوز طراحی، هدف، تأمین زیرساخت و تصمیم‌های راهبردی به انسان وابسته است(Europol, 2023; Brundage et al., 2018) .

۹. کاربرد هوش مصنوعی در انواع جرایم رایانه‌ای

۱-۹. فیشینگ و مهندسی اجتماعی

فیشینگ به تلاش برای فریب کاربر و دریافت اطلاعات محرمانه، نصب بدافزار یا انتقال وجه گفته می‌شود. هوش مصنوعی در این حوزه چند کارکرد مهم دارد:

- تولید متن‌های روان و بدون خطا؛
- تقلید از لحن مدیر بانک یا سازمان؛
- ترجمه سریع پیام به زبان‌های مختلف؛
- تحلیل اطلاعات عمومی قربانی؛
- انتخاب زمان مناسب برای ارسال پیام؛

- تولید گفت‌وگوی تعاملی با قربانی؛
- ترکیب متن، صدا و تصویر برای افزایش اعتماد.

در مهندسی اجتماعی، هدف اصلی سوءاستفاده از اعتماد، ترس، فوریت یا اقتدار است. هوش مصنوعی می‌تواند پیام را از حالت عمومی خارج و آن را با نام، شغل، روابط سازمانی و علایق قربانی هماهنگ کند. این امر احتمال موفقیت حمله را افزایش می‌دهد (Hadnagy, 2018).

بااین‌حال، فیشینگ هوش مصنوعی-محور همچنان ممکن است نشانه‌هایی مانند درخواست غیرعادی، تغییر مسیر پرداخت، فشار زمانی، استفاده از حساب جدید یا ناسازگاری با رویه‌های سازمانی داشته باشد. بنابراین احراز هویت چندعاملی و آموزش کارکنان اهمیت زیادی دارد.

۹-۲. تولید محتوای جعلی و دیپ‌فیک

دیپ‌فیک به محتوای صوتی، تصویری یا ویدئویی گفته می‌شود که با روش‌های هوش مصنوعی، چهره، صدا یا رفتار فردی را به‌صورت جعلی بازسازی می‌کند. کاربردهای مجرمانه آن عبارت‌اند از:

- اخاذی و تهدید؛
- تخریب حیثیت؛
- فریب مالی؛
- جعل دستور مدیر یا مقام مسئول؛
- انتشار اطلاعات نادرست؛
- مداخله در فرایندهای سیاسی و اجتماعی؛
- جعل شهادت یا انتساب اظهارات دروغین(Chesney & Citron, 2019; Mirsky & Lee, 2021) .

دیپ‌فیک اصل اعتماد به ادراک بصری و شنیداری را تضعیف می‌کند. در گذشته، تصویر یا صدای ضبط‌شده قرینه‌ای قوی محسوب می‌شد؛ اما اکنون اصالت آن باید با روش‌های فنی و بررسی زنجیره نگهداری اثبات شود. تشخیص دیپ‌فیک نیز رقابتی مستمر میان تولیدکننده و آشکارساز است و هیچ ابزار تشخیصی به‌تنهایی تضمین مطلق ارائه نمی‌دهد (Europol, 2022; Mirsky & Lee, 2021).

۹-۳. جعل هویت و کلاهبرداری

هوش مصنوعی می‌تواند برای ساخت حساب‌های جعلی، تولید مدارک ساختگی، تقلید صدا و ایجاد شخصیت‌های برخط استفاده شود. در کلاهبرداری موسوم به «رئیس شرکت»، مهاجم با جعل ایمیل، صدا یا تصویر مدیر، کارمند را به انتقال وجه یا افشای اطلاعات وادار می‌کند.

در مقیاس فردی، مهاجم ممکن است هویت قربانی را برای دریافت اعتبار، افتتاح حساب یا انجام معامله جعل کند. در مقیاس سازمانی، جعل هویت تأمین‌کننده، مشتری یا مقام دولتی می‌تواند خسارات گسترده ایجاد کند. این وضعیت ضرورت احراز هویت چندمرحله‌ای و تأیید مستقل درخواست‌های حساس را نشان می‌دهد.

۹-۴. تولید و انتشار بدافزار

بدافزار شامل برنامه‌هایی است که برای دسترسی غیرمجاز، تخریب، جاسوسی، سرقت داده یا اخلال طراحی می‌شوند. هوش مصنوعی می‌تواند در تولید قطعات کد، اصلاح خطا، تغییر ظاهر فایل و تطبیق رفتار بدافزار با محیط هدف به‌کار رود.

خطر اصلی در این حوزه، کاهش موانع مهارتی است؛ فردی که دانش عمیق برنامه‌نویسی ندارد ممکن است از ابزارهای مولد برای تهیه نمونه‌های ساده استفاده کند. بااین‌حال، توسعه بدافزار پیشرفته همچنان به آزمون، زیرساخت، دانش سامانه عامل و مدیریت

عملیات نیاز دارد. از دید حقوقی، صرف تولید کد ذاتاً جرم نیست؛ جرم زمانی شکل می‌گیرد که قصد، محتوا، زمینه و کاربرد آن با عناصر قانونی جرم منطبق باشد(NIST, 2022) .

۹-۵. کشف آسیب‌پذیری و خودکارسازی حملات

هوش مصنوعی در تحلیل کد، شناسایی الگوهای آسیب‌پذیر و اولویت‌بندی نقاط ضعف توانمند است. همین قابلیت می‌تواند برای آزمایش امنیتی مشروع یا حمله غیرمجاز استفاده شود. تفاوت میان این دو به مجوز، هدف، حدود فعالیت و نحوه مدیریت یافته‌ها وابسته است(NIST, 2024a) .

در حملات خودکار، سامانه می‌تواند اهداف را شناسایی، داده‌ها را تحلیل و مراحل مختلف عملیات را هماهنگ کند. این امر زمان واکنش مدافع را کاهش می‌دهد. راهکار مناسب، استفاده از مدیریت آسیب‌پذیری مبتنی بر ریسک، وصله‌گذاری سریع، آزمون نفوذ مجاز و ثبت دقیق فعالیت‌هاست(ENISA, 2023) .

۹-۶. حملات منع خدمت و اخاذی سایبری

حملات منع خدمت توزیع‌شده (DDoS) با ارسال حجم زیادی از درخواست، منابع سامانه را مصرف می‌کنند. هوش مصنوعی می‌تواند در شناسایی الگوهای دفاعی، انتخاب زمان حمله و تغییر شکل ترافیک مؤثر باشد. همچنین در باج‌افزار، هوش مصنوعی برای شناسایی فایل‌های مهم، طبقه‌بندی دارایی‌ها و اولویت‌بندی اهداف قابل‌استفاده است(ENISA, 2022) .

اخاذی سایبری جدید گاهی از الگوی دوگانه فراتر می‌رود؛ مهاجم علاوه بر رمزگذاری داده، آن‌ها را سرقت و تهدید به انتشار می‌کند. در این زمینه، تقسیم‌بندی شبکه، پشتیبان‌گیری آفلاین، کنترل دسترسی و برنامه پاسخ به حادثه اهمیت دارد (World Economic Forum, 2024).

۹-۷. سرقت داده و نقض حریم خصوصی

هوش مصنوعی می‌تواند داده‌های پراکنده را به اطلاعات حساس تبدیل کند. ترکیب داده‌های عمومی، سوابق برخط، تصاویر، موقعیت مکانی و ارتباطات اجتماعی، امکان ایجاد پروفایل دقیق از افراد را فراهم می‌سازد. چنین پروفایل‌هایی برای کلاهبرداری، اخاذی، تبعیض یا نظارت غیرمجاز قابل‌استفاده‌اند(OECD, 2013) .

همچنین مدل‌های هوش مصنوعی ممکن است در اثر طراحی ضعیف، داده‌های محرمانه را در خروجی بازتاب دهند یا اطلاعات واردشده توسط کاربران را در معرض خطر قرار دهند. رعایت اصل حداقل‌سازی داده، رمزنگاری، کنترل دسترسی و ارزیابی اثرات حریم خصوصی ضروری است(NIST, 2024b) .

۱۰. آثار اقتصادی، اجتماعی و امنیتی

۱۰-۱. آثار اقتصادی

آثار اقتصادی شامل سرقت مستقیم، هزینه بررسی حادثه، ازکارافتادگی سامانه، جبران خسارت، ازدست‌رفتن مشتری و کاهش ارزش اعتبار سازمان است. کسب‌وکارهای کوچک به‌دلیل منابع محدود، آسیب‌پذیری بیشتری دارند. هوش مصنوعی با افزایش مقیاس حملات می‌تواند تعداد قربانیان را افزایش دهد، اما از سوی دیگر هزینه دفاع را نیز با خودکارسازی پایش و پاسخ کاهش می‌دهد(FBI, 2024) .

۱۰-۲. آثار اجتماعی

انتشار محتوای جعلی، اعتماد عمومی به رسانه‌ها و نهادها را تضعیف می‌کند. اگر مردم نتوانند میان محتوای واقعی و ساختگی تمایز بگذارند، بی‌اعتمادی عمومی گسترش می‌یابد. این وضعیت ممکن است به آسیب روانی قربانیان، تشدید شایعات، اختلاف اجتماعی و کاهش مشارکت عمومی منجر شود(Chesney & Citron, 2019) .

۱۰-۳. آثار امنیتی

حملات هوشمند می‌توانند زیرساخت‌های حیاتی، زنجیره تأمین و سامانه‌های دولتی را هدف قرار دهند. همچنین استفاده دولت‌ها یا گروه‌های سازمان‌یافته از هوش مصنوعی در عملیات نفوذ، جاسوسی و عملیات اطلاعاتی، امنیت ملی را با چالش مواجه می‌کند (NCSC, 2024; INTERPOL, 2022).

۱۱. چالش‌های کشف جرم و ادله دیجیتال

۱۱-۱. حجم و سرعت داده‌ها

حملات خودکار حجم زیادی از لاگ‌ها، پیام‌ها، فایل‌ها و ارتباطات ایجاد می‌کنند. تحلیل دستی چنین داده‌ای دشوار است و پلیس و مراجع قضایی به ابزارهای تحلیل هوشمند نیاز دارند(ENISA, 2023) .

۱۱-۲. اصالت و تمامیت محتوا

در پرونده‌های دیپ‌فیک، صرف ارائه فایل صوتی یا تصویری کافی نیست. باید منشأ، زمان تولید، مسیر انتقال، ابزار ایجاد و احتمال تغییر آن بررسی شود. استانداردهای ثبت منشأ محتوا و امضای دیجیتال می‌توانند به افزایش اعتماد کمک کنند (C2PA, 2024).

۱۱-۳. زنجیره نگهداری ادله

ادله دیجیتال باید به‌گونه‌ای جمع‌آوری و نگهداری شود که تغییر نکردن آن قابل‌اثبات باشد. ثبت زمان، ابزار، شخص مسئول، نسخه داده و عملیات انجام‌شده از الزامات زنجیره نگهداری است.

۱۱-۴. استفاده از ابزارهای ناشناس‌ساز

شبکه‌های خصوصی مجازی(VPN) ، ابزارهای ناشناس‌سازی، حساب‌های جعلی و رمزارزها می‌توانند شناسایی مرتکب را دشوار کنند. بااین‌حال، ناشناس‌سازی مطلق نیست و خطاهای عملیاتی، الگوهای ارتباطی و داده‌های جانبی ممکن است هویت مرتکب را آشکار سازند.

۱۱-۵. قابلیت توضیح‌پذیری

اگر الگوریتمی رفتار را مشکوک اعلام کند، این پرسش مطرح است که دادگاه تا چه اندازه می‌تواند به آن اعتماد کند. خروجی مدل باید قرینه‌ای قابل‌ارزیابی باشد، نه جایگزین کامل استدلال قضایی. شفافیت، آزمون‌پذیری و امکان بازبینی انسانی اهمیت دارد (European Commission, 2019; NIST, 2024a).

۱۲. مسئله انتساب رفتار مجرمانه

انتساب در جرایم هوش مصنوعی-محور در سه سطح مطرح است:

۱. انتساب فنی

آیا خروجی یا حمله واقعاً از سامانه موردادعا ناشی شده است؟ این مسئله نیازمند بررسی لاگ‌ها، حساب کاربری، کلیدهای دسترسی، داده‌های دستگاه و ارتباطات شبکه است.

۲. انتساب انسانی

چه شخصی سامانه را به کار گرفته، دستور داده یا از نتیجه آن بهره برده است؟ وجود حساب کاربری به‌تنهایی همیشه برای اثبات هویت کافی نیست؛ زیرا ممکن است حساب سرقت یا مشترک باشد.

۳. انتساب حقوقی

حتی اگر مشخص شود شخصی خروجی را تولید کرده، باید ثابت شود که عناصر قانونی جرم از جمله رفتار، نتیجه، رابطه سببیت و در موارد لازم سوءنیت وجود داشته است.

سامانه هوش مصنوعی در وضعیت فعلی، شخص حقوقی یا طبیعی مستقل محسوب نمی‌شود و فاقد مسئولیت کیفری به معنای سنتی است. بنابراین مسئولیت اصولاً باید به شخصی منتسب شود که کنترل مؤثر، قصد مجرمانه یا تقصیر قابل‌انتساب داشته باشد (قانون مجازات اسلامی، ۱۳۹۲).

۱۳. مسئولیت کیفری مرتکب، طراح و ارائه‌دهنده سامانه

۱-۱۳. مسئولیت کاربر یا مرتکب مستقیم

اگر شخصی آگاهانه از هوش مصنوعی برای فریب، دسترسی غیرمجاز، تخریب، جعل یا سرقت استفاده کند، استفاده از سامانه مانع مسئولیت او نیست. هوش مصنوعی در این حالت، وسیله ارتکاب است؛ همانند رایانه، تلفن همراه یا ابزار ارتباطی.

۲-۱۳. مسئولیت طراح

طراح تنها در صورتی مسئولیت کیفری می‌یابد که مشارکت آگاهانه و مؤثر او در جرم، تسهیل عمدی رفتار مجرمانه یا نقض تکالیف قانونی اثبات شود. صرف طراحی یک فناوری دوکاربردی بدون علم و رابطه مؤثر با جرم، برای مسئولیت کیفری کافی نیست (Brundage et al., 2018).

۳-۱۳. مسئولیت ارائه‌دهنده

ارائه‌دهنده ممکن است در صورت بی‌احتیاطی شدید، فقدان کنترل‌های متعارف، بی‌توجهی به هشدارهای روشن یا عدم رعایت تکالیف قانونی با مسئولیت مدنی، اداری یا در موارد استثنایی کیفری مواجه شود. مسئولیت باید مبتنی بر تقصیر و قابلیت پیش‌بینی باشد، نه صرف وقوع جرم توسط کاربر(European Commission, 2024) .

۴-۱۳. مسئولیت اشخاص حقوقی

چنانچه جرم به نام یا در راستای منافع شخص حقوقی و توسط مدیر یا نماینده آن انجام شود، مقررات مربوط به مسئولیت کیفری اشخاص حقوقی قابل‌بررسی است. تعیین این مسئولیت به ساختار سازمانی، تصمیم‌گیری، نظارت و منفعت حاصل‌شده وابسته است (قانون مجازات اسلامی، ۱۳۹۲، مواد ۱۲۶-۱۴۴)

۱۴. بررسی قوانین ایران

۱-۱۴. قانون جرایم رایانه‌ای (مصوب ۱۳۸۸)

قانون جرایم رایانه‌ای چارچوب اصلی مقابله با جرایم سایبری در ایران است و مقرراتی درباره دسترسی غیرمجاز، شنود، تخریب و اخلال در داده‌ها و سامانه‌ها، سرقت و کلاهبرداری مرتبط با رایانه، جعل رایانه‌ای، هتک حیثیت و انتشار محتوای مجرمانه پیش‌بینی کرده است.

این قانون در بسیاری از رفتارهای هوش مصنوعی-محور قابلیت اعمال دارد. برای نمونه، اگر دیپ‌فیک با هدف هتک حیثیت یا انتشار محتوای مجرمانه تولید شود، مقررات مرتبط با هتک حیثیت و نشر اکاذیب می‌تواند مطرح گردد. اگر سامانه برای فریب مالی و انتقال وجه به کار رود، مقررات کلاهبرداری و کلاهبرداری رایانه‌ای قابل‌بررسی است.

۱۴-۲. قانون مجازات اسلامی (مصوب ۱۳۹۲)

قواعد عمومی قانون مجازات اسلامی درباره عمد، تقصیر، معاونت، مشارکت، شروع به جرم، تعدد و مسئولیت اشخاص حقوقی در تحلیل جرایم هوش مصنوعی-محور اهمیت دارند. عنصر فناورانه، اصول بنیادین مسئولیت کیفری را حذف نمی‌کند؛ بلکه اثبات رابطه میان رفتار انسان و نتیجه را دشوارتر می‌سازد.

۱۴-۳. قانون تجارت الکترونیکی (مصوب ۱۳۸۲)

قانون تجارت الکترونیکی ایران مقرراتی درباره داده‌پیام، امضای الکترونیکی، حمایت از داده‌ها و برخی رفتارهای مجرمانه در بستر تجارت الکترونیکی دارد. این قانون در ارزیابی اصالت ارتباطات، اسناد الکترونیکی و معاملات برخط قابل‌استفاده است.

۱۴-۴. قانون انتشار و دسترسی آزاد به اطلاعات (مصوب ۱۳۸۷)

شفافیت و دسترسی قانونی به اطلاعات باید با حمایت از داده‌های شخصی و محرمانگی متوازن شود. استفاده مجرمانه از داده‌های عمومی یا نیمه‌عمومی، ضرورت بازتعریف حدود انتشار و پردازش داده را نشان می‌دهد.

۱۴-۵. قانون حمایت از حقوق پدیدآورندگان نرم‌افزارهای رایانه‌ای (مصوب ۱۳۷۹)

در مواردی که کد یا نرم‌افزار بدون اجازه تکثیر یا مورد بهره‌برداری قرار گیرد، قواعد حمایت از حقوق نرم‌افزار نیز ممکن است مطرح شود. البته باید میان نقض مالکیت فکری و تولید ابزار مجرمانه تفکیک کرد.

۱۴-۶. خلأهای موجود

مهم‌ترین خلأها عبارت‌اند از:

- نبود تعریف جامع از محتوای جعلی آسیب‌زا؛
- فقدان مقررات خاص درباره جعل صوتی و تصویری؛
- ابهام در مسئولیت ارائه‌دهندگان سامانه‌های مولد؛
- نبود استاندارد جامع برای اصالت‌سنجی محتوای دیجیتال؛
- فقدان نظام آماری مستقل برای جرایم هوش مصنوعی-محور؛
- پراکندگی مقررات مرتبط با حریم خصوصی؛
- نیاز به قواعد روشن‌تر درباره همکاری پلتفرم‌ها و نگهداری داده‌ها.

۱۵. بررسی اسناد و رویکردهای بین‌المللی

۱۵-۱. کنوانسیون بوداپست (۲۰۰۱)

کنوانسیون بوداپست درباره جرایم سایبری، همکاری بین‌المللی، صلاحیت، اقدامات آیین دادرسی و ادله الکترونیکی را تقویت می‌کند. اهمیت این سند در جرایم هوش مصنوعی-محور از آن جهت است که بسیاری از حملات از زیرساخت‌ها و حساب‌های واقع در چند کشور استفاده می‌کنند(Council of Europe, 2001) .

۱۵-۲. رویکرد سازمان ملل متحد

گزارش‌ها و اسناد سازمان ملل بر همکاری بین‌المللی، ظرفیت‌سازی، حمایت از حقوق بشر و توازن میان امنیت و آزادی‌های بنیادین تأکید دارند(UNODC, 2013; United Nations, 2021) .

۱۵-۳. رویکردOECD

اصول OECD درباره هوش مصنوعی بر رشد فراگیر، شفافیت، توضیح‌پذیری، ایمنی، امنیت و پاسخ‌گویی تأکید می‌کند. این اصول الزام کیفری مستقیم ایجاد نمی‌کنند، اما برای طراحی سیاست عمومی و استانداردهای حکمرانی اهمیت دارند(OECD, 2019) .

۱۵-۴. چارچوبNIST

چارچوب مدیریت ریسک هوش مصنوعی NIST (AI RMF 1.0) مدیریت ریسک را بر مبنای حکمرانی، شناسایی، سنجش و مدیریت پیشنهاد می‌کند. این چارچوب می‌تواند در سازمان‌ها برای ارزیابی خطر سوءاستفاده از مدل‌ها و کاهش آسیب‌پذیری‌ها به کار رود(NIST, 2023; NIST, 2024b) .

۱۵-۵. رویکرد اتحادیه اروپا

قانون هوش مصنوعی اتحادیه اروپا (AI Act - 2024/1689) رویکردی مبتنی بر سطح خطر اتخاذ می‌کند و برای کاربردهای پرخطر، الزامات شفافیت، مدیریت ریسک و نظارت انسانی در نظر می‌گیرد. همچنین مقررات حفاظت از داده‌های عمومی اروپا (GDPR)پردازش داده و حقوق اشخاص را تنظیم می‌کند(European Union, 2016; European Union, 2024) .

۱۵-۶. گزارش‌های Europol وENISA

گزارش‌های Europol درباره استفاده مجرمان از هوش مصنوعی مولد بر احتمال افزایش فریب، تولید محتوای جعلی و دسترسی افراد کم‌مهارت به ابزارهای مجرمانه تأکید دارند .(Europol, 2022; Europol, 2023) گزارش‌های ENISA نیز بر تهدیدهای نوظهور، امنیت زنجیره تأمین و ضرورت آمادگی سازمانی تمرکز می‌کنند(ENISA, 2022; ENISA, 2023) .

۱۵-۷. رویکردINTERPOL

گزارش روندهای جرم جهانی اینترپل بر افزایش استفاده از هوش مصنوعی در جرایم سازمان‌یافته و نیاز به همکاری پلیس‌های ملی تأکید دارد(INTERPOL, 2022) .

۱۶. چالش‌های حقوقی و اخلاقی

۱۶-۱. اصل قانونی بودن جرم و مجازات

توسعه فناوری نباید به تفسیر موسع و پیش‌بینی‌ناپذیر مقررات کیفری منجر شود. هر رفتار باید با نص قانونی روشن و قابل‌دسترس منطبق باشد (قانون مجازات اسلامی، ماده ۲).

۱۶-۲. حریم خصوصی

تحلیل گسترده داده‌ها برای کشف جرم می‌تواند به نظارت فراگیر منجر شود. استفاده از هوش مصنوعی در پلیس و نهادهای امنیتی باید با ضرورت، تناسب، محدودیت هدف و نظارت مستقل همراه باشد(OECD, 2013; European Union, 2016) .

۱۶-۳. تبعیض الگوریتمی

مدل‌های آموزشی ممکن است خطاهای داده‌های گذشته را بازتولید کنند. اگر ابزار ارزیابی خطر جرم بدون کنترل استفاده شود، ممکن است گروه‌هایی را ناعادلانه در معرض سوءظن قرار دهد.

۱۶-۴. آزادی بیان

مبارزه با محتوای جعلی نباید به حذف بی‌ضابطه محتوای انتقادی، طنز یا هنری منجر شود. باید میان جعل زیان‌بار، محتوای سیاسی، اثر هنری و بیان مشروع تفکیک شود(Chesney & Citron, 2019) .

۱۶-۵. شفافیت و توضیح‌پذیری

افراد باید تا حد امکان بدانند تصمیم مهم درباره آنان چگونه اتخاذ شده است. در حوزه کیفری، تصمیم نهایی نباید صرفاً بر خروجی غیرقابل‌توضیح یک مدل مبتنی باشد(European Commission, 2019; NIST, 2024a) .

۱۶-۶. اصل انسانی بودن مسئولیت کیفری

در شرایط کنونی، اعطای شخصیت مستقل به سامانه هوش مصنوعی و انتقال مسئولیت به ماشین با اصول حقوق کیفری سازگار نیست. مسئولیت باید به انسان یا شخص حقوقی قابل‌شناسایی و بر اساس تقصیر یا قصد منتسب شود.

۱۷. راهکارهای پیشگیری فنی، حقوقی و سازمانی

۱-۱۷. پیشگیری فنی

- استفاده از احراز هویت چندعاملی؛
- رمزنگاری داده‌های حساس؛
- تفکیک شبکه و اصل حداقل دسترسی؛
- ثبت و پایش مداوم رویدادها؛
- استفاده از سامانه‌های تشخیص ناهنجاری؛
- پشتیبان‌گیری منظم و آزمایش بازیابی؛
- امضای دیجیتال و ثبت منشأ محتوا(C2PA, 2024) ؛
- به‌روزرسانی و وصله‌گذاری سریع؛
- آزمون امنیتی پیش از استقرار مدل؛
- ارزیابی حملات فریب‌دهنده علیه مدل‌ها.

۲-۱۷. پیشگیری حقوقی

قانون‌گذار باید:

- مفهوم جعل دیجیتال و دیپ‌فیک زیان‌بار را تعریف کند؛
- تفاوت میان تولید، انتشار و بهره‌برداری مجرمانه را روشن سازد؛
- قواعد نگهداری و ارائه ادله دیجیتال را به‌روز کند؛
- مسئولیت ارائه‌دهندگان را بر اساس سطح خطر و تقصیر تنظیم نماید؛
- همکاری بین‌المللی و تبادل ادله را تسهیل کند؛
- از جرم‌انگاری کلی و مبهم پرهیز نماید.

۳-۱۷. پیشگیری سازمانی

هر سازمان باید برنامه مدیریت ریسک هوش مصنوعی داشته باشد. این برنامه باید شامل فهرست سامانه‌ها، نوع داده‌های پردازش‌شده، سطح دسترسی، مسئول نظارت، سناریوهای سوءاستفاده و برنامه واکنش به حادثه باشد(NIST, 2024a) .

۴-۱۷. حاکمیت داده

کیفیت و امنیت داده آموزشی و عملیاتی اهمیت اساسی دارد. داده‌های اضافی، نامعتبر یا بدون رضایت می‌توانند هم خطر نقض حریم خصوصی و هم خطر خطای تصمیم‌گیری را افزایش دهند.

۱۸. نقش آموزش و سواد رسانه‌ای

فناوری‌های تشخیص بدون آموزش کاربران، اثربخشی محدودی دارند. آموزش باید از سطح توصیه‌های کلی فراتر رود و سناریوهای واقعی را تمرین دهد. کاربران باید بدانند:

- هر پیام فوری الزاماً معتبر نیست؛
- صدای آشنا دلیل قطعی هویت نیست؛
- تصویر و ویدئو قابل جعل است؛
- درخواست تغییر شماره حساب باید مستقلاً تأیید شود؛

- رمز و کد یک‌بارمصرف نباید در اختیار دیگران قرار گیرد؛
- گزارش سریع حادثه، احتمال کاهش خسارت را افزایش می‌دهد.

سواد رسانه‌ای شامل توانایی ارزیابی منبع، تشخیص تناقض، بررسی تاریخ انتشار، مقایسه چند منبع و شناخت انگیزه تولیدکننده محتواست. آموزش مدیران ارشد، کارکنان مالی و مسئولان فناوری باید متناسب با سطح خطر آنان طراحی شود (Hadnagy, 2018).

۱۹. بحث و تحلیل

بررسی مطالب نشان می‌دهد که هوش مصنوعی سه اثر اصلی بر جرایم رایانه‌ای دارد: **کاهش هزینه، افزایش مقیاس** و **افزایش باورپذیری**. این سه اثر در کنار یکدیگر، مدل اقتصادی جرم را تغییر می‌دهند. حمله‌ای که پیش‌تر نیازمند گروهی متخصص بود، ممکن است اکنون با نیروی انسانی کمتر و ابزارهای آماده اجرا شود(Europol, 2023; Brundage et al., 2018) .

بااین‌حال، تحلیل علمی باید از فناوری‌هراسی دور باشد. هوش مصنوعی به‌خودی‌خود مجرم نیست و همه کاربردهای مولد خطرناک نیستند. بسیاری از سامانه‌های دفاعی نیز از هوش مصنوعی استفاده می‌کنند و می‌توانند زمان کشف و پاسخ را کاهش دهند. بنابراین مسئله اصلی، دوکاربردی بودن و نحوه حکمرانی بر فناوری است(NIST, 2024a; NCSC, 2024) .

از دید حقوقی، پاسخ مؤثر نیازمند ترکیب سه سطح است: **نخست**، اجرای مقررات موجود در مواردی که رفتار با جرم شناخته‌شده منطبق است؛ **دوم**، اصلاح مقررات در حوزه‌هایی مانند دیپ‌فیک و ادله منشأ محتوا؛ **سوم**، ایجاد استانداردهای پیشگیرانه برای ارائه‌دهندگان و سازمان‌های استفاده‌کننده. مسئولیت کیفری باید فرد-محور و مبتنی بر قصد یا تقصیر باقی بماند، اما مسئولیت‌های مدنی، اداری و حرفه‌ای می‌توانند برای طراحی ایمن‌تر و مدیریت خطر به کار روند.

۲۰. چارچوب پیشنهادی مقابله

چارچوب پیشنهادی این پژوهش، مدل **پنج‌لایه حکمرانی و مقابله** نام دارد:

لایه اول: شناسایی و ارزیابی خطر

- شناسایی دارایی‌ها و داده‌های حساس؛
- سنجش سطح خودکارشدن سامانه؛
- بررسی قابلیت سوءاستفاده؛
- اولویت‌بندی سناریوهای پرخطر.

لایه دوم: طراحی ایمن

- محدودسازی دسترسی؛
- ثبت رخدادها؛
- کنترل خروجی‌های حساس؛
- آزمایش مدل در برابر فریب و سوءاستفاده؛
- طراحی سازوکار توقف اضطراری.

لایه سوم: نظارت انسانی

- تعیین مسئول انسانی؛
- بازبینی تصمیم‌های پرخطر؛
- ممنوعیت تصمیم‌گیری کاملاً خودکار در حوزه‌های حساس؛

- ثبت دلایل و سوابق تصمیم.

لایه چهارم: پاسخ و کشف

- ایجاد تیم واکنش به حادثه؛
- حفظ فوری ادله؛
- همکاری میان پلیس، دادستانی، ارائه‌دهنده و قربانی؛
- استفاده از تحلیل الگوریتمی همراه با ارزیابی کارشناسی.

لایه پنجم: پاسخ‌گویی و اصلاح

- گزارش‌دهی رخداد؛
- ارزیابی پس از حادثه؛
- جبران خسارت؛
- اصلاح مدل و رویه‌های سازمانی؛
- انتشار هشدارهای عمومی و شاخص‌های تهدید.

این چارچوب باید با همکاری دولت، بخش خصوصی، دانشگاه، جامعه مدنی و نهادهای بین‌المللی اجرا شود (NIST, 2024a; OECD, 2019).

۲۱. محدودیت‌های پژوهش

- بسیاری از جرایم در آمار رسمی با عنوان مستقل «جرم ناشی از هوش مصنوعی» ثبت نمی‌شوند.
- فناوری‌های مولد با سرعت زیادی تغییر می‌کنند و برخی یافته‌ها ممکن است نیازمند به‌روزرسانی باشند.
- دسترسی محدود به پرونده‌های قضایی و داده‌های عملیاتی ایران، امکان تحلیل آماری داخلی را کاهش داده است.
- بخشی از اطلاعات مربوط به ابزارهای مجرمانه در گزارش‌های عمومی منتشر نمی‌شود.
- تفاوت نظام‌های حقوقی کشورها، مقایسه مستقیم مقررات را دشوار می‌سازد.
- این مقاله از ارائه دستورالعمل عملی برای ساخت بدافزار یا اجرای حمله خودداری کرده و بر تحلیل حقوقی و پیشگیرانه تمرکز دارد.

۲۲. نتیجه‌گیری

هوش مصنوعی یکی از مهم‌ترین عوامل تحول در جرایم رایانه‌ای است. این فناوری با تولید محتوای طبیعی، تحلیل داده‌های گسترده، شخصی‌سازی پیام‌ها، خودکارسازی عملیات و تقلید هویت، توان مهاجمان را افزایش می‌دهد. فیشینگ، مهندسی اجتماعی، دیپ‌فیک، جعل هویت، بدافزار، حملات منع خدمت، اخاذی و نقض حریم خصوصی از مهم‌ترین حوزه‌های اثرگذاری آن هستند.

باوجوداین، هوش مصنوعی ماهیت مسئولیت کیفری را به‌طور کامل تغییر نمی‌دهد. در اغلب موارد، سامانه ابزار یا واسطه‌ای است که در چارچوب تصمیم و رفتار انسان عمل می‌کند. بنابراین انتساب جرم باید با بررسی کنترل، علم، قصد، تقصیر، رابطه سببیت و منفعت حاصل‌شده انجام شود. طراح یا ارائه‌دهنده تنها به‌دلیل وجود جرم مسئول نیست، بلکه باید نقش مؤثر و تقصیر یا مشارکت قابل‌اثبات داشته باشد.

قوانین ایران ظرفیت‌هایی برای برخورد با بسیاری از رفتارهای مجرمانه دارند، اما درباره دیپ‌فیک، حریم خصوصی داده، مسئولیت ارائه‌دهندگان، اصالت محتوای مولد و ادله الگوریتمی، نیاز به تکمیل و شفاف‌سازی وجود دارد. مقابله مؤثر نیز صرفاً با جرم‌انگاری

ممکن نیست. امنیت فنی، آموزش، استانداردسازی، پاسخ سریع، همکاری بین‌المللی و نظارت انسانی باید در کنار مقررات کیفری قرار گیرند.

در نهایت، سیاست مطلوب باید میان نوآوری و امنیت، آزادی بیان و مقابله با جعل، بهره‌وری و حریم خصوصی، و مسئولیت‌پذیری و حمایت از توسعه فناوری تعادل برقرار کند.

۲۳. پیشنهادها

- تدوین مقررات جامع درباره محتوای جعلی آسیب‌زا و دیپ‌فیک؛
- ایجاد نظام ملی ثبت و گزارش جرایم هوش مصنوعی-محور؛
- تدوین استاندارد فنی برای منشأ، اصالت و تمامیت محتوای دیجیتال(C2PA, 2024) ؛
- آموزش تخصصی ضابطان، کارشناسان رسمی، قضات و دادستان‌ها؛
- الزام سازمان‌های حساس به ارزیابی ریسک هوش مصنوعی؛
- توسعه احراز هویت چندعاملی و تأیید مستقل تراکنش‌های حساس؛
- حمایت از پژوهش‌های دانشگاهی میان‌رشته‌ای؛
- ایجاد سازوکار همکاری سریع میان پلتفرم‌ها و مراجع قضایی؛
- استفاده از هوش مصنوعی دفاعی با نظارت انسانی و کنترل تبعیض؛
- تقویت سواد رسانه‌ای و آموزش عمومی؛
- توسعه همکاری منطقه‌ای و بین‌المللی برای تبادل ادله؛
- تعیین مسئولیت ارائه‌دهندگان بر اساس سطح خطر، قابلیت پیش‌بینی و میزان تقصیر؛
- حمایت از افشای مسئولانه آسیب‌پذیری‌ها؛
- ایجاد آزمایشگاه‌های تخصصی تشخیص محتوای جعلی؛
- تضمین حمایت از حقوق بنیادین در استفاده دولتی از سامانه‌های هوش مصنوعی.

منابع

۱. جمهوری اسلامی ایران. (۱۳۸۸). *قانون جرایم رایانه‌ای*.

۲. جمهوری اسلامی ایران. (۱۳۹۲). *قانون مجازات اسلامی*.

۳. جمهوری اسلامی ایران. (۱۳۸۲). *قانون تجارت الکترونیکی*.

۴. جمهوری اسلامی ایران. (۱۳۸۷). *قانون انتشار و دسترسی آزاد به اطلاعات*.

۵. جمهوری اسلامی ایران. (۱۳۷۹). *قانون حمایت از حقوق پدیدآورندگان نرم‌افزارهای رایانه‌ای*.

6. Bandura, A. (1977). *Social Learning Theory*. Prentice Hall.

7. Becker, G. S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy, 76*(2), 169-217.

8. Brundage, M., Avin, S., Clark, J., et al. (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. *arXiv:1802.07228*.

9. C2PA. (2024). *C2PA Technical Specification Version 2.1*. Coalition for Content Provenance and Authenticity.

10. Chesney, B., & Citron, D. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review, 107*(6), 1753-1820.

11. Cohen, L. E., & Felson, M. (1979). Social Change and Crime Rate Trends: A Routine Activity Approach. *American Sociological Review, 44*(4), 588-608.
12. Council of Europe. (2001). *Convention on Cybercrime*. Budapest.
13. ENISA. (2022). *ENISA Threat Landscape for Ransomware Attacks*. European Union Agency for Cybersecurity.
14. ENISA. (2023). *ENISA Threat Landscape 2023*. European Union Agency for Cybersecurity.
15. European Commission. (2019). *Ethics Guidelines for Trustworthy AI*. High-Level Expert Group on Artificial Intelligence.
16. European Union. (2016). *Regulation (EU) 2016/679: General Data Protection Regulation (GDPR)*.
17. European Union. (2024). *Regulation (EU) 2024/1689 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*.
18. Europol. (2022). *Facing Reality? Law Enforcement and the Challenge of Deepfakes*. Europol Innovation Lab.
19. Europol. (2023). *ChatGPT: The Impact of Large Language Models on Law Enforcement*. Europol Innovation Lab.
20. FBI. (2024). *Internet Crime Report 2023*. Federal Bureau of Investigation.
21. Hadnagy, C. (2018). *Social Engineering: The Science of Human Hacking* (2nd ed.). Wiley.
22. INTERPOL. (2022). *INTERPOL Global Crime Trend Report 2022*. International Criminal Police Organization.
23. Mirsky, Y., & Lee, W. (2021). The Creation and Detection of Deepfakes: A Survey. *ACM Computing Surveys, 54*(1), 1-41.
24. NCSC. (2024). *The Near-Term Impact of AI on the Cyber Threat*. National Cyber Security Centre (UK).
25. NIST. (2022). *Secure Software Development Framework (SSDF) Version 1.1*. National Institute of Standards and Technology.
26. NIST. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. National Institute of Standards and Technology.
27. NIST. (2024a). *Cybersecurity Framework 2.0*. National Institute of Standards and Technology.
28. NIST. (2024b). *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. National Institute of Standards and Technology.
29. OECD. (2013). *OECD Privacy Guidelines*. Organisation for Economic Co-operation and Development.
30. OECD. (2019). *Recommendation of the Council on Artificial Intelligence*. Organisation for Economic Co-operation and Development.
31. Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
32. United Nations. (2021). *Countering the Use of Information and Communications Technologies for Criminal Purposes*. United Nations.
33. UNODC. (2013). *Comprehensive Study on Cybercrime*. United Nations Office on Drugs and Crime.
34. World Economic Forum. (2024). *Global Risks Report 2024*. World Economic Forum.