

The Encyclopedia of Comparative Jurisprudence and Law

Feasibility of Redefining the Mental Element and the Problem of Attribution in Criminal Law Related to Artificial Intelligence

Raziyeh Sekhavati^{1*}

1. MA, Department of Criminal Law and Criminology, Payame Noor University, South Tehran Branch, Iran

* Corresponding Author's Email: baroonsharji88@gmail.com

ABSTRACT

The emergence and expansion of artificial intelligence-based technologies with partial or full autonomy have posed unprecedented challenges to the traditional foundations of criminal law. The core of this challenge lies in the difficulty of applying classical concepts of the “mental element” (intent, knowledge, negligence) and the “attribution of criminal responsibility” to a non-human entity lacking consciousness and free will. This article, using a descriptive-analytical method and relying on library and documentary sources, examines the feasibility of reinterpreting and redefining these concepts in light of the nature of artificial intelligence. The findings indicate that the Iranian criminal law system, similar to many other legal systems, is inherently “person-centered” and faces serious theoretical and practical obstacles in recognizing the direct criminal liability of artificial intelligence, such as the inability of such entities to bear traditional punishments and the failure to fulfill the objectives of punishment, including rehabilitation and reform. Consequently, current legal approaches primarily focus on the liability of related human actors (designer, manufacturer, programmer, user) or on the development of civil liability or risk-based frameworks. Nevertheless, this article argues that, given the increasing complexity and autonomy of such systems, a functionalist and teleological interpretation may allow the concept of “algorithmic intent” or “programmed defective performance” to be analyzed as a novel form of the mental element. Furthermore, through mechanisms such as “conditional virtual legal personality” or a “hierarchical hybrid liability system,” a pathway for attributing responsibility may be established. Finally, the article proposes the necessity of enacting specific legislation, establishing specialized regulatory institutions, and introducing mandatory insurance schemes for high-risk technologies as essential infrastructures for this theoretical transformation.

Keywords: Artificial Intelligence, Criminal Law, Mental Element, Crime Attribution, Criminal Liability, Legal Personality, Algorithmic Autonomy.

How to cite: Sekhavati, R. (2027). Feasibility of Redefining the Mental Element and the Problem of Attribution in Criminal Law Related to Artificial Intelligence. *The Encyclopedia of Comparative Jurisprudence and Law*, 5(3), 1-17.



تاریخ ارسال: ۱ دی ۱۴۰۴

تاریخ بازنگری: ۴ اردیبهشت ۱۴۰۵

تاریخ پذیرش: ۱۳ اردیبهشت ۱۴۰۵

تاریخ چاپ اولیه: ۱ مرداد ۱۴۰۵

تاریخ چاپ نهایی: ۱ شهریور ۱۴۰۶

دانشنامه فقه و حقوق تطبیقی

امکان‌سنجی بازتعریف عنصر روانی و مسئله انتساب در حقوق کیفری مرتبط با هوش مصنوعی

راضیه سخاوتی^۱

۱. کارشناسی ارشد، گروه حقوق کیفری و جرم‌شناسی، دانشگاه پیام‌نور، واحد تهران جنوب، ایران

* پست الکترونیک نویسنده مسئول: baroonsharji88@gmail.com

چکیده

ظهور و گسترش فناوری‌های مبتنی بر هوش مصنوعی با قابلیت خودمختاری نسبی یا کامل، بنیان‌های سستی حقوق کیفری را با چالش‌هایی بی‌سابقه مواجه ساخته است. هسته مرکزی این چالش، دشواری تطبیق مفاهیم کلاسیک «عنصر روانی» (قصد، علم، تقصیر) و «انتساب مسئولیت کیفری» بر موجودیتی غیرانسانی، فاقد شعور و اراده آزاد است. این مقاله با روشی توصیفی-تحلیلی و با مطالعه منابع کتابخانه‌ای و اسنادی، به امکان‌سنجی بازخوانی و بازتعریف این مفاهیم در پرتو ماهیت هوش مصنوعی می‌پردازد. یافته‌ها نشان می‌دهد که نظام حقوق کیفری ایران، مشابه بسیاری از نظام‌های حقوقی دیگر، ذاتاً «شخص‌محور» است و برای پذیرش مسئولیت کیفری مستقیم هوش مصنوعی با موانع نظری و عملی جدی مانند فقدان قابلیت تحمل مجازات‌های سستی و عدم تحقق اهداف مجازات مانند اصلاح و بازپروری روبروست. در نتیجه، تمرکز کنونی رویکردهای حقوقی بر مسئولیت انسان‌های مرتبط (طراح، سازنده، برنامه‌نویس، کاربر) و یا توسعه چارچوب‌های مسئولیت مدنی یا مبتنی بر ریسک معطوف شده است. با این حال، مقاله حاضر استدلال می‌کند که با توجه به افزایش پیچیدگی و خودمختاری سامانه‌ها، می‌توان با تفسیری کارکردگرا و غایت‌نگر، مفهوم «قصد الگوریتمی» یا «عملکرد معیوب برنامه‌ریزی شده» را به عنوان صورتی نوین از عنصر روانی تحلیل کرد و از طریق مکانیسم‌هایی مانند «شخصیت حقوقی مجازی مشروط» یا «نظام مسئولیت ترکیبی سلسله‌مراتبی»، راهی برای انتساب مسئولیت هموار ساخت. در پایان، ضرورت تدوین قوانین خاص، ایجاد نهادهای نظارتی تخصصی و پیش‌بینی نظام‌های بیمه اجباری برای فناوری‌های پرخطر به عنوان بسترهای لازم برای این تحول نظری پیشنهاد می‌گردد.

کلیدواژه‌ها: هوش مصنوعی، حقوق کیفری، عنصر روانی، انتساب جرم، مسئولیت کیفری، شخصیت حقوقی، خودمختاری الگوریتمی.

نحوه استناددهی: سخاوتی، راضیه. (۱۴۰۶). امکان‌سنجی بازتعریف عنصر روانی و مسئله انتساب در حقوق کیفری مرتبط با هوش مصنوعی. *دانشنامه فقه و حقوق تطبیقی*، ۵(۳)، ۱۷-۱.

انقلاب فناوری هوش مصنوعی در حال دگرگون‌سازی تمامی عرصه‌های زندگی بشری، از حمل‌ونقل و پزشکی تا امور مالی و قضائی است. در این میان، یکی از پیچیده‌ترین و چالش‌برانگیزترین تقاطعات، مواجهه نظام عدالت کیفری با ماشین‌های هوشمند و خودمختار است (Masoumi Janaqerd & Zarei, 2024). حقوق کیفری، به مثابه ابزار نهایی جامعه برای حفظ نظم و ارزش‌های اساسی، همواره بر دو رکن استوار بوده است: عنصر مادی، رفتار فیزیکی مجرمانه، و عنصر روانی، حالت ذهنی مرتکب شامل قصد، علم، بی‌احتیاطی یا تقصیر. انتساب مسئولیت و تحمیل مجازات مشروط به احراز این دو رکن، آن هم در مورد یک «شخص» حقیقی یا حقوقی است که از اهلیت جزایی و اراده آزاد برخوردار باشد (Basami, 2020). پرسش بنیادین عصر حاضر این است: هنگامی که یک سامانه هوش مصنوعی خودمختار، مانند یک خودروی خودران که در اثر تفسیر نادرست داده‌های حسگری‌اش منجر به قتل غیرعمد می‌شود، یک ربات جراح که خطای محاسباتی فاجعه‌باری مرتکب می‌گردد، یا یک الگوریتم معاملاتی در بازار مالی که با دستکاری خودکار و پیچیده، سبب اختلال گسترده در بازار می‌شود، رفتاری را انجام می‌دهد که از منظر قانون کیفری جرم محسوب می‌شود، چگونه می‌توان مسئولیت آن را تبیین کرد؟ (Hekmatnia et al., 2019). آیا می‌توان به این موجودیت دیجیتال، که فاقد شعور، آگاهی، احساس گناه یا اراده آزاد به معنای فلسفی و حقوقی آن است، «عنصر روانی» نسبت داد؟ اگر پاسخ منفی است، آیا می‌توان آن را به عنوان فاعل جرم شناخت و مستقیماً مجازات کرد؟ و اگر مجازات آن ناممکن یا بی‌معناست، چگونه می‌توان خلأ مسئولیت را پر کرد و از بروز بی‌کیفرمانی در قبال آسیب‌های شدید ناشی از هوش مصنوعی جلوگیری نمود؟ (Farajiha & Alamdari, 2017).

این مسئله تنها یک بحث انتزاعی آکادمیک نیست، بلکه ضرورتی فوری برای قانون‌گذاری و سیاست‌گذاری است. نظام‌های حقوقی در سراسر جهان، از جمله نظام حقوقی ایران، با معمایی حل‌نشده مواجه‌اند. از یک سو، قوانین موجود مانند قانون مجازات اسلامی (2013 "Code of Criminal Procedure") و قانون جرایم رایانه‌ای (Shirzad, 2009)، ذیل مفهوم سنتی «شخص» اعم از حقیقی و حقوقی عمل می‌کنند و هوش مصنوعی را به عنوان فاعل مستقل جرم به رسمیت نمی‌شناسند. از سوی دیگر، افزایش روزافزون خودمختاری سامانه‌ها به این معناست که نمی‌توان همواره آن‌ها را صرفاً به عنوان «ابزار» در دست کاربر انسانی تلقی کرد (Alipour, 2011). در مواردی، تصمیم‌گیری توسط الگوریتم‌های پیچیده‌ای انجام می‌شود که حتی برای طراحان اولیه نیز به طور کامل قابل پیش‌بینی یا توضیح نیست؛ پدیده‌ای که به «مشکل جعبه سیاه» معروف است (Moazenadegan et al., 2015). این امر توزیع سنتی مسئولیت میان طراح، سازنده و کاربر را با مشکل مواجه می‌سازد؛ چرا که ممکن است هیچ‌یک از این افراد به صورت مستقیم «قصد» یا «تقصیر» خاصی در وقوع نتیجه مجرمانه نداشته باشند، حال آنکه سامانه هوشمند به طور مستقل و بر پایه یادگیری و استنتاج خود، مرتکب عملی زیان‌بار شده است (Tavakoli Rad & Miri, 2021).

مقاله حاضر با پذیرش این چالش بنیادین، در پی امکان‌سنجی «بازتعریف عنصر روانی و مسئله انتساب» در حقوق کیفری است. سؤال اصلی این است: آیا می‌توان در پرتو ویژگی‌های منحصر به فرد هوش مصنوعی، خودمختاری، قابلیت یادگیری، و تصمیم‌گیری بر پایه الگوریتم، مفاهیم حقوقی عنصر روانی و انتساب را به گونه‌ای تفسیر مجدد یا توسعه داد که پاسخگویی مؤثر و عادلانه به رفتارهای مجرمانه ناشی از این فناوری را ممکن سازد؟ یا اینکه راه‌حل، نه در بازتعریف این مفاهیم برای هوش مصنوعی، بلکه در طراحی کامل چارچوب‌های مسئولیت جدید، مانند مسئولیت مدنی

ب. خطای محض. حالتی است که فعل، غیر عمدی است و مرتکب نه قصد ارتکاب جرم را دارد و نه تقصیری از ناحیه او سر زده است، مانند موارد قوه قاهره (Katouzian, 1995).

ج. شبه عمد و بی احتیاطی. در این حالات که ذیل عنوان کلی تقصیر می گنجد، مرتکب قصد نتیجه مجرمانه را ندارد، ولی به دلیل عدم رعایت موازین احتیاط یا مهارتی که از او انتظار می رود، مرتکب تقصیر شده و نتیجه مجرمانه محقق می شود (مواد ۱۴۵ و ۱۴۶ ق.م.ا) (Basami, 2020).

مبانی فلسفی مستتر در پشت این تفکیک، مبتنی بر اصول اختیار، مسئولیت اخلاقی و قابلیت سرزنش انسان است. نظام کیفری بر این پایه استوار است که فرد به دلیل برخورداری از اراده آزاد و قدرت تمیز، می تواند میان نیک و بد انتخاب کند و در صورت انتخاب مسیر مجرمانه، شایسته سرزنش و مجازات است (Rahimi Nejad, 2010). اهداف مجازات، اعم از سزا، اصل استحقاق، اصلاح و بازپروری، بازدارندگی و ارعاب نیز بر همین پیش فرض انسان محور بنا شده اند (Rosenhan et al., 2010).

گستره شخصیت حقوقی و اهلیت جزایی

حقوق برای انتساب مسئولیت، نیاز به یک «شخص» دارد. در حقوق مدرن، علاوه بر شخص حقیقی، انسان، مفهوم شخص حقوقی، مانند شرکت ها، مؤسسات و دولت، نیز به رسمیت شناخته شده است. ماده ۵۸۸ قانون تجارت ایران مسئولیت کیفری اشخاص حقوقی را در مواردی که نماینده قانونی شخص حقوقی به نام یا در راستای منافع آن مرتکب جرمی شود، پیش بینی کرده است. با این حال، حتی در مسئولیت شخص حقوقی نیز، پایه و اساس انتساب جرم، فعل انسانی، نماینده، است که دارای عنصر روانی انسانی می باشد. شخص حقوقی به عنوان یک انتزاع حقوقی، از طریق ارگان های انسانی خود عمل می کند و اراده و قصد آن، جمع بندی یا تجميع اراده افراد انسانی است (Sadeghi Neshat, 1997).

شدید، مسئولیت مبتنی بر ریسک، یا مسئولیت جمعی، برای انسان های پشت صحنه آن است؟ (Katouzian, 2013). این نوشتار با واکاوی مبانی نظری عنصر روانی در حقوق ایران، تحلیل چالش های عملی انتساب جرم به هوش مصنوعی، بررسی رویکردهای مختلف جهانی، و در نهایت، ارائه سنتز و پیشنهادهایی برای نظام حقوقی ایران، مسیری را برای عبور از این چالش تاریخی ترسیم می نماید.

مبانی نظری: عنصر روانی، شخصیت حقوقی و مسئولیت کیفری در حقوق ایران

ماهیت و اقسام عنصر روانی (معنوی) در حقوق کیفری ایران در حقوق کیفری ایران، همانند اکثر نظام های حقوق نوشته، جرم محصول ترکیب دو عنصر مادی و روانی است. عنصر روانی، که از آن به «قصد مجرمانه» یا «سوءنیت» نیز تعبیر می شود، بیانگر رابطه ذهنی مرتکب با عمل مجرمانه و نتایج آن است. ماده ۱۴۴ قانون مجازات اسلامی ("Code of Criminal Procedure," 2013) به شکلی کلاسیک این مفهوم را بازتاب می دهد و مجازات را منوط به وجود «عمد» یا «قصد» در انجام فعل یا ترک فعل می داند، مگر در موارد خاص که قانون به طور صریح ترتیب دیگری مقرر کرده باشد.

عنصر روانی دارای سطوح و درجات مختلفی است که هر یک دلالت اخلاقی و حقوقی متفاوتی داشته و بر نوع و میزان مجازات تأثیر مستقیم می گذارند:

الف. عمد (سوءنیت خاص و عام). عالی ترین درجه عنصر روانی است که مستلزم علم و اراده تحقق فعل و نتیجه مجرمانه است. در جرایم عمدی، مرتکب هم فعل را اراده می کند (سوءنیت عام) و هم نتیجه مجرمانه خاص را (سوءنیت خاص) (Najafi Abrandabadi & Ebrahimi, 2004).

روانی که مبتنی بر حالات ذهنی آگاهانه است به موجودی فاقد آگاهی، یک مغالطه مقوله‌ای محسوب می‌شود (Khodagholi, 2004).

فقدان اراده آزاد؛ جبر علی‌الگوریتمی

بسیاری از پژوهش‌ها بر ماهیت جبرگرایانه یا احتمالی تصمیم‌گیری هوش مصنوعی تأکید می‌کند. «اراده آزاد» در حقوق کیفری، نه به معنای فلسفی محض، بلکه به معنای قدرت انتخاب مبتنی بر شعور و قوه تمیز است؛ توانایی عمل کردن برخلاف انگیزه‌های قوی یا برنامه‌های از پیش تعیین‌شده. تصمیم یک سیستم هوش مصنوعی، حتی اگر پیچیده و غیرقابل پیش‌بینی برای ناظر بیرونی باشد، همواره محصول ضروری و اجتناب‌ناپذیر داده‌های ورودی، وزن‌های از پیش تنظیم‌شده شبکه عصبی، و تابع هدف الگوریتم است. این سیستم نمی‌تواند به معنای واقعی «انتخاب» کند؛ بلکه تنها می‌تواند محاسبه کند. از این رو، مفهوم قصد که در جرایم عمدی سنگ بنای مسئولیت است، قابل انطباق بر فرآیندی صرفاً محاسباتی نیست. سامانه‌ای که یک خطای مرگبار مرتکب می‌شود، نه «قصد» آسیب داشته و نه از روی «بی‌مبالاتی» عمل کرده است؛ بلکه صرفاً مطابق با منطق درونی خود که ممکن است معیوب یا ناکامل باشد، عملی را انجام داده است (Keshtkar & Hajilou, 2024).

فقدان قابلیت سرزنش اخلاقی؛ مجازات بدون گناه

یکی از اصول اخلاقی و زیربنای مجازات، سزادهی مستحقانه است. مجازات، پیش و بیش از هر چیز، واکنشی اخلاقی به عمل سرزنش‌بار مرتکب است. این سرزنش‌پذیری مشروط به این است که فاعل، عاملیت اخلاقی داشته باشد؛ یعنی بتوان او را به خاطر انتخاب نادرستش ملامت کرد. هوش مصنوعی فاقد چنین عاملیتی است. یک الگوریتم نمی‌تواند احساس پشیمانی کند، از عمل خود شرم‌منده شود یا قصد جبران داشته باشد. مجازات یک سامانه، مثلاً حذف آن یا اعمال جریمه نقدی به حسابش، فاقد هرگونه محتوای

مسئله اهلیت جزایی نیز از ملزومات مسئولیت کیفری است. برای تحمیل مجازات، فرد باید دارای اهلیت استیفاء باشد؛ یعنی عاقل، بالغ و مختار باشد (ماده ۱۴۰ ق.م.ا). فقدان هر یک از این شرایط، مانند صغر سن، جنون، اجبار، می‌تواند به سلب یا کاهش مسئولیت بینجامد. این ملاحظات کاملاً مرتبط با ویژگی‌های زیست‌شناختی و روان‌شناختی انسان است (Najafi Abrandabadi & Ebrahimi, 2004).

چالش ذاتی هوش مصنوعی با پارادایم سنتی

این بخش از مقاله به هسته مرکزی تنش میان هوش مصنوعی خودمختار و چارچوب نظری حقوق کیفری کلاسیک می‌پردازد. پارادایم سنتی حقوق کیفری، که در نظام‌هایی مانند ایران ریشه دوانده، بر پایه‌ای انسان‌محور و سزاگرا بنا شده است. این چارچوب، مسئولیت و مجازات را مشروط به وجود مجموعه‌ای از ویژگی‌های ذهنی و اخلاقی می‌داند که منحصراً در انسان یافت می‌شود. هوش مصنوعی، صرف‌نظر از پیچیدگی، فاقد این کیفیات ذاتی است و بنابراین در کانون این پارادایم قرار نمی‌گیرد. در ادامه، هر یک از چالش‌های چهارگانه به تفصیل تحلیل می‌شود.

فقدان آگاهی و شعور؛ ماشین بدون ذهن

عملیات هوش مصنوعی، هرچند پیشرفته، در نهایت محاسبات ریاضی و پردازش الگوریتمی داده‌هاست. این سیستم‌ها فاقد آگاهی ذهنی هستند؛ یعنی تجربه از بودن، احساس یا درک. یک سیستم تشخیص تصویر ممکن است یک گربه را با دقتی بالا شناسایی کند، اما هرگز «حس» دیدن یک گربه یا لذت از نگاه به آن را ندارد. این فقدان درک اخلاقی ذاتی حیاتی است. درک انسان از مفاهیمی چون «عدالت»، «انصاف» یا «آسیب»، ریشه در ترکیب پیچیده‌ای از عقل، احساس، همدلی و تجربه زیسته دارد. هوش مصنوعی می‌تواند قواعد اخلاقی را به صورت دستورالعمل‌هایی فرمال بیاموزد و اجرا کند، اما فاقد حکمت عملی برای قضاوت اخلاقی در موقعیت‌های مبهم و بی‌سابقه است. بنابراین، انتساب عنصر

ارتباطی - اخلاقی با خود سامانه است. این اقدامات ممکن است جنبه‌های کارکردی، مانند حذف خطر، جبران خسارت، داشته باشند، اما هرگز معادل مجازات یک انسان نخواهند بود. این امر یکی از موانع اصلی تحقق هدف سزادهی در حقوق کیفری در قبال هوش مصنوعی است (Rahimi Nejad, 2010).

عدم امکان تحقق اهداف مجازات؛ اهدافی بی‌مخاطب

برخی پژوهش‌ها به بحران در تحقق سایر اهداف مجازات اشاره می‌کند که مهم‌ترین آن‌ها عبارتند از:

الف. اصلاح و بازپروری. این هدف مبتنی بر تغییر نگرش، باورها و انگیزه‌های مجرم برای بازگشت به جامعه است. یک الگوریتم را نمی‌توان با «درمان»، «آموزش اخلاق» یا «ارائه مشاوره» اصلاح کرد. تنها می‌توان آن را به‌روزرسانی، اصلاح کد یا جایگزین کرد که عملیاتی کاملاً فنی و فاقد بعد تربیتی است.

ب. ارعاب عام و خاص. ارعاب، بر ترس مجرم بالقوه از پیامدهای ناخوشایند استوار است. یک سامانه هوش مصنوعی فاقد ترس، عقلانیت حسابگرانه انسانی یا حتی آگاهی از وجود مجازات است. بنابراین، تهدید به مجازات، هیچ اثر بازدارنده‌ای بر آن ندارد. اثر بازدارندگی ممکن است به‌طور غیرمستقیم متوجه سازندگان و بهره‌برداران انسانی شود.

ج. سلب توانایی. این هدف از طریق مجازات‌هایی مانند حبس یا انحلال شخص حقوقی محقق می‌شود. در مورد هوش مصنوعی، سلب توانایی می‌تواند به معنای خاموش کردن، محدود کردن دسترسی یا حذف فیزیکی سرورها باشد. این شاید تنها هدفی باشد که به‌صورت فیزیکی قابل تحقق است، اما باز هم فاقد آن بار اخلاقی و اجتماعی است که در سلب آزادی یک انسان نهفته است (Hekmatnia et al., 2019).

بنابراین، این چهار چالش در نهایت نشان می‌دهند که هوش مصنوعی خودمختار در پارادایم سستی مسئولیت کیفری «جا نمی‌گیرد». پذیرش آن به‌عنوان فاعل مستقل مسئول، مستلزم تغییر

در مبانی انسان‌محورانه حقوق کیفری است. با این حال، این بن‌بست، پژوهشگر را به طرح این پرسش کلیدی سوق می‌دهد: «آیا می‌توان مفاهیم را به‌گونه‌ای توسعه داد که این شکاف پر شود؟» پاسخ به این پرسش، مسیر اصلی مقاله را شکل می‌دهد و به بخش‌های آتی مانند بازتعریف عنصر روانی، تبدیل «قصد» به «عملکرد معیوب برنامه‌ریزی‌شده»، طراحی مکانیسم‌های جدید انتساب، مانند مسئولیت سلسله‌مراتبی یا شخصیت حقوقی الکترونیک، و بازنگری در اهداف مجازات در عصر دیجیتال منتهی می‌شود. این بخش (۱-۳)، بنابراین، نه یک پایان، بلکه پایه‌گذاری ضرورت تحول نظری در حقوق کیفری است.

چالش‌های نوین در انتساب جرم به هوش مصنوعی

طیف خودمختاری و پیدایش «شکاف مسئولیت»

همه سامانه‌های هوش مصنوعی یکسان نیستند. طیفی از کنترل انسانی کامل تا خودمختاری کامل وجود دارد. در یک سوی طیف، ربات‌های صنعتی ساده قرار دارند که عملیات کاملاً از پیش تعریف‌شده‌ای انجام می‌دهند. در این موارد، نظریه سستی «مسئولیت کاربر یا اپراتور» یا «مسئولیت سازنده به دلیل عیب تولید» به‌خوبی قابل اعمال است (Katouzian, 2013). مشکل اصلی از جایی آغاز می‌شود که سامانه‌ها از خودمختاری عملیاتی برخوردار می‌گردند؛ یعنی می‌توانند در محیط‌های پیچیده و غیرقابل پیش‌بینی، بدون دخالت مستقیم و لحظه‌به‌لحظه انسان، برای تحقق اهداف از پیش تعیین‌شده، تصمیم بگیرند و عمل کنند (Alipour, 2011).

خودمختاری بالا دو پیامد دارد: نخست، غیرقابل پیش‌بینی بودن برخی تصمیمات، به‌ویژه در سامانه‌های مبتنی بر یادگیری عمیق که از داده‌ها الگوهای پیچیده استخراج می‌کنند. دوم، عدم امکان ردیابی ساده و خطی علّت از طراحی اولیه تا نتیجه نهایی (Moazenadegan et al., 2015). اینجاست که مفهوم «شکاف مسئولیت» مطرح می‌شود. شکاف مسئولیت زمانی رخ

عادلانه از جمله حق دفاع و استماع طرفین را تهدید می‌کند (Tabrizi et al., 2022).

تحلیل موردی؛ خودروهای خودران و سلاح‌های خودمختار

دو حوزه‌ای که چالش مسئولیت در آن‌ها به اوج خود می‌رسد، خودروهای خودران و سلاح‌های خودمختار مرگبار هستند.

خودروهای خودران

در یک تصادف مرگبار، پرسش‌های متعددی مطرح است: آیا خطا از حسگرهای فیزیکی بود؟ آیا الگوریتم بینایی ماشین دچار سوگیری در تشخیص شده بود؟ آیا داده‌های آموزشی کافی یا متنوع نبوده؟ آیا کاربر، سرنشین، می‌بایست در لحظه مداخله می‌کرد؟ توزیع مسئولیت میان سازنده قطعات، توسعه‌دهنده نرم‌افزار، شرکت مونتاژکننده، مالک خودرو و احتمالاً شهرداری به‌عنوان مسئول نگهداری راه، بسیار پیچیده می‌شود (Keshtkar & Hajilou, 2024).

سلاح‌های خودمختار مرگبار

این سلاح‌ها می‌توانند بدون دخالت انسانی در میدان جنگ، اهداف را انتخاب و مورد حمله قرار دهند. این امر مستقیماً با اصول حقوق بین‌الملل بشردوستانه، از جمله اصل تمایز بین نظامی و غیرنظامی و اصل تناسب در اصابت، در تعارض است (Salehi, 2019). در صورت وقوع جنایت جنگی توسط چنین سلاحی، آیا می‌توان فرمانده نظامی را طبق دکترین مسئولیت فرماندهی مسئول دانست، اگر که سلاح از کنترل و پیش‌بینی او خارج شده باشد؟ (Calerick et al., 2010).

این سناریوها نشان می‌دهند که چارچوب‌های سنتی انتساب جرم به‌شدت تحت فشار قرار دارند.

رویکردهای حقوقی به مسئله مسئولیت کیفری هوش مصنوعی در پاسخ به این چالش‌ها، چهار رویکرد اصلی در ادبیات حقوقی و رویه‌های تقنینی اولیه در جهان شکل گرفته است:

می‌دهد که آسیب یا نتیجه مجرمانه مستقیماً ناشی از عمل خودمختار سامانه هوش مصنوعی است. هیچ فرد انسانی معین، طراح، برنامه‌نویس، کاربر، را نمی‌توان به‌صورت متعارف، با معیارهای قصد یا تقصیر، مقصر دانست، زیرا عمل خارج از سناریوهای پیش‌بینی‌شده یا آزمون‌شده رخ داده است (Zakerinia, 2023). سامانه هوش مصنوعی نیز فاقد صلاحیت برای پذیرش مسئولیت کیفری در نظام حقوقی موجود است. نتیجه این شکاف، بی‌کیفرمانی و عدم جبران خسارت قربانیان است (Farajiha & Alamdari, 2017).

معضل «جعبه سیاه» و دشواری احراز علت‌شناسی

یکی از ملموس‌ترین چالش‌ها در مسیر انتساب جرم، مشکل جعبه سیاه است. در بسیاری از مدل‌های پیشرفته هوش مصنوعی، فرآیند تبدیل داده‌های ورودی به خروجی، شامل میلیون‌ها پارامتر و محاسبات لایه‌بندی‌شده است که حتی برای خالقان سامانه نیز به‌طور کامل قابل تفسیر و دنبال کردن نیست (Masoumi & Zarei, 2024). این امر چند مشکل حقوقی ایجاد می‌کند:

الف. اثبات رابطه سببی. چگونه می‌توان در دادگاه ثابت کرد که تصمیم خاص الگوریتم، مثلاً شناسایی نادرست یک عابر پیاده به‌عنوان کیسه زباله توسط خودروی خودران، علت مستقیم حادثه بوده است؟ (Katouzian, 2013).

ب. اثبات تقصیر. چگونه می‌توان تقصیر طراح یا سازنده را احراز کرد، وقتی نمی‌توان به‌طور مشخص نشان داد که کدام بخش از طراحی، آموزش یا پیاده‌سازی منجر به خطا شده است؟ (Katouzian, 2013).

ج. حق دفاع متهم. اگر تصمیم‌گیری یک سامانه هوش مصنوعی، مثلاً در ارزیابی ریسک تکرار جرم یک متهم، مبنای حکم قاضی قرار گیرد، وکیل چگونه می‌تواند از موکلش دفاع کند و منطق پشت این تصمیم را به چالش بکشد؟ این عدم شفافیت، اصول دادرسی

رویکرد انکار مسئولیت مستقل (ابزارانگاری)

این رویکرد محافظه کارانه، که همچنان رویکرد غالب در بسیاری از نظام‌ها از جمله حقوق ایران و اتحادیه اروپا است، هوش مصنوعی را فاقد شخصیت حقوقی و اهلیت جزایی می‌داند. هوش مصنوعی صرفاً یک «ابزار» یا «محصول» پیچیده است (Khodagholi, 2004). بنابراین، مسئولیت باید متوجه انسان‌های مرتبط باشد:

الف. مسئولیت طراح و سازنده. بر مبنای تقصیر، عدم رعایت استانداردهای ایمنی، طراحی معیوب، آزمون ناکافی، یا بر مبنای مسئولیت محض، بدون تقصیر، برای محصولات معیوب. این رویکرد شبیه مسئولیت مدنی ناشی از عیب تولید است (Katouzian, 2013).

ب. مسئولیت کاربر یا اپراتور. در صورتی که کاربر به‌صورت تقصیرآمیز از سامانه استفاده کرده یا در صورت لزوم مداخله نکرده باشد (Alipour, 2011).

مزیت این رویکرد، انطباق با ساختارهای حقوقی موجود است. اما نقطه ضعف اصلی آن، ناتوانی در پوشش شکاف مسئولیت در موارد خودمختاری بالا است، جایی که نه سازنده تقصیر محرز دارد و نه کاربر امکان مداخله مؤثر (Zakerinia, 2023).

رویکرد اعطای شخصیت حقوقی الکترونیک (شخص‌انگاری)

رویکرد اعطای «شخصیت حقوقی الکترونیک» یا «شخصیت مجازی» به سیستم‌های هوش مصنوعی پیشرفته، پاسخی جسورانه و آینده‌نگرانه به بن‌بست مسئولیت‌پذیری در عصر هوش مصنوعی است. این ایده که در ادبیات حقوقی جهان تحت عناوینی مانند «شخصیت الکترونیک» مطرح می‌شود، در حقوق ایران نیز توسط برخی محققان مورد بحث قرار گرفته است. این رویکرد در تلاش است تا با استفاده از یک قیاس حقوقی خلاقانه با نهاد شخصیت حقوقی، به‌ویژه شرکت‌های تجاری، چارچوبی برای پاسخگویی مستقیم هوش مصنوعی ایجاد کند (Zakerinia, 2023).

مبنای و قیاس حقوقی؛ از شرکت‌ها تا الگوریتم‌ها

مبنای اصلی این ایده، قیاس سیستم‌های خودمختار هوش مصنوعی با اشخاص حقوقی است. در حقوق مدرن، یک شرکت تجاری، با وجود آنکه مجموعه‌ای از انسان‌ها، دارایی‌ها و قواعد است، به‌عنوان یک شخص حقوقی مستقل شناخته می‌شود. این شخصیت مستقل به شرکت امکان می‌دهد تا دارایی داشته باشد، قرارداد منعقد کند و مسئولیت بپذیرد، در حالی که فاقد «آگاهی» یا «اراده» زیست‌شناختی است. اراده آن از طریق ارگان‌های تصمیم‌گیری متشکل از انسان‌ها اعمال می‌شود. طرفداران این رویکرد استدلال می‌کنند که یک سیستم هوش مصنوعی بسیار پیشرفته نیز می‌تواند به‌عنوان یک واحد عملیاتی خودمختار در نظر گرفته شود. در این قیاس، الگوریتم‌ها، داده‌ها و سخت‌افزار، جایگزین اساسنامه، دارایی و کارکنان شرکت می‌شوند. تصمیمات این شخصیت مجازی نه از طریق هیئت‌مدیره، بلکه از طریق پردازش الگوریتمی داده‌ها اتخاذ می‌شود (Zakerinia, 2023).

سازوکار عملیاتی و مجازات‌های پیشنهادی

هدف از اعطای این شخصیت، ایجاد یک واسطه حقوقی مسئولیت‌پذیر است که بتوان مستقیماً به آن رجوع کرد. سازوکار آن مبتنی بر ایجاد یک «صندوق مالی مسئولیت» برای هر سامانه یا کلاس از سامانه‌های پرریسک است. این صندوق می‌تواند از طریق بیمه اجباری، حق امتیاز بهره‌برداری یا سهمی از درآمد حاصل از عملکرد سامانه تأمین شود و به‌عنوان منبع اصلی جبران خسارات قربانیان عمل کند. تحت این مدل، قربانیان می‌توانند مستقیماً از «شخص هوش مصنوعی» به دلیل عمل زیان‌بارش دعوای جبران خسارت اقامه کنند و دادگاه می‌تواند حکم به پرداخت خسارت از این صندوق را صادر نماید.

مجازات‌های قابل اعمال بر این شخصیت مجازی، ماهیتی کارکردی و مرتبط با ماهیت فناورانه آن خواهند داشت و نه الزاماً انسانی. این مجازات‌ها می‌توانند شامل اصلاح یا به‌روزرسانی

رویکرد مسئولیت مدنی محض یا مبتنی بر ریسک

در مواجهه با «شکاف مسئولیت» ناشی از هوش مصنوعی، برخی حقوقدانان مسئولیت کیفری مبتنی بر تقصیر را رها کرده و به سمت مسئولیت مدنی محض یا مبتنی بر ریسک حرکت کرده‌اند (Katouzian, 2013). این منطق، خسارت‌های ناشی از فناوری‌های پرخطر را صرف‌نظر از تقصیر، قابل جبران توسط ذی‌نفعان می‌داند، مشابه مسئولیت فعالیت‌های پرخطر. اتحادیه اروپا در «قانون هوش مصنوعی» خود، این گرایش را برای سیستم‌های پرخطر نشان داده است (Zakerinia, 2023). مزیت این روش، جبران سریع خسارت قربانی است، اما نقطه ضعف آن فقدان سرزنش اخلاقی و بازدارندگی کیفری و خطر خفه‌سازی نوآوری به دلیل تحمیل بار مالی سنگین بر صنعت است.

رویکرد مسئولیت ترکیبی سلسله‌مراتبی (مدل هرمی)

برای پرهیز از افراط و تفریط رویکردهای پیشین، یک راه‌حل میانی تحت عنوان «مسئولیت ترکیبی سلسله‌مراتبی» یا «مدل هرمی» مطرح شده است. این مدل سعی دارد با در نظر گرفتن کل چرخه حیات هوش مصنوعی، از طراحی و تولید تا استقرار و استفاده، مسئولیت را به‌صورت متناسب با سطح کنترل، نفوذ، سود و توانایی پیشگیری هر یک از بازیگران توزیع کند (Keshtkar & Hajilou, 2024).

در قاعده این هرم، توسعه‌دهنده و سازنده قرار دارد که مسئولیت اولیه و محض را در قبال نقایص طراحی، آموزش نادرست یا عدم انجام آزمون‌های کافی بر عهده می‌گیرد. در سطح بالاتر، تأمین‌کننده یا واردکننده مسئول رعایت استانداردهای ایمنی و هشدارهای لازم است. سپس اپراتور یا کاربر حرفه‌ای مسئول استفاده مطابق با دستورالعمل و نظارت مؤثر است. در رأس هرم، برای سامانه‌های با خودمختاری بسیار بالا، می‌توان یک صندوق جبران خسارت اجباری، شبیه به صندوق تأمین خسارت‌های خودرو، در نظر گرفت که به‌صورت خودکار و مستقیم از محل دارایی سامانه یا

اجباری الگوریتم تحت نظارت مرجع ناظر، محدودیت دامنه عملیاتی یا کاهش سطح خودمختاری، جریمه نقدی که از دارایی، صندوق، شخصیت مجازی پرداخت می‌شود، و در نهایت انحلال نهایی یا غیرفعال کردن دائمی سامانه باشد که معادل مجازات «انحلال» برای یک نهاد حقوقی است.

نقدهای وارد بر این رویکرد و پاسخ‌های احتمالی

رویکرد اعطای شخصیت حقوقی به موجودات غیرزنده، به دلیل چالش‌های اخلاقی، فلسفی و حقوقی، با نقدهای جدی مواجه است؛ از جمله نگرانی درباره خلط مرز انسان و ماشین و شکل‌گیری «برده‌داری دیجیتال» که طرفداران آن را به‌عنوان یک ابزار صرفاً حقوقی برای جبران خسارت و نظارت رد می‌کنند و با شرکت‌ها قیاس می‌نمایند. چالش دیگر، ابهام در تعیین آستانه «خودمختاری کافی» برای اعطای این شخصیت است که پاسخ آن، تدوین معیارهای عینی قانونی نظیر توانایی تصمیم‌گیری در محیط‌های پیچیده و ریسک‌زایی برای حقوق دیگران است. قوی‌ترین نقد، خطر فرار مسئولیت انسان‌ها، سازندگان و بهره‌برداران، با انتقال ریسک به «شخصیت هوش مصنوعی» است که برای مقابله با آن، نظام مسئولیت دولایه پیشنهاد می‌شود؛ در این ساختار، لایه اول مسئولیت سریع شخصیت مجازی از طریق صندوق خود را شامل می‌شود و لایه دوم مسئولیت کیفری/مدنی تقصیرمحور انسان‌های پشت صحنه را پوشش می‌دهد. در نهایت، این رویکرد به سمت یک «مدل ترکیبی» گرایش دارد که در آن «شخصیت حقوقی مشروط و ابزاری» برای هوش مصنوعی‌های خاص پذیرفته شده، اما مسئولیت نهایی و سلسله‌مراتبی بر عهده انسان باقی می‌ماند تا شکاف مسئولیت پر شود و عدالت در عصر ماشین‌های خودمختار تحقق یابد (Najafi Abrandabadi & Ebrahimi, 2004).

بیمه اجباری آن، خسارات را پوشش دهد. این مدل انعطاف‌پذیر، هم‌زمان هم جبران خسارت را تضمین می‌کند، هم انگیزه برای رعایت استانداردهای ایمنی در تمام مراحل را ایجاد می‌نماید، و هم با تعریف سطوح مختلف مسئولیت، از تمرکز بار غیرمنصفانه بر یک فرد جلوگیری می‌کند.

بازتعریف عنصر روانی و الگوی ترکیبی انتساب

مباحث پیشین نشان داد که رویکردهای سنتی و جسورانه نسبت به هوش مصنوعی در پر کردن «شکاف مسئولیت» ناکامند؛ محافظه‌کاران پیچیدگی فناوری را درک نمی‌کنند و رادیکال‌ها عاملان انسانی را مبرا می‌سازند. این وضعیت لزوم بازاندیشی در مفاهیم بنیادین حقوق کیفری، به‌ویژه «عنصر روانی» را مطرح می‌کند (Tabrizi et al., 2022). برای حل این معضل، دو گام اساسی پیشنهاد می‌شود: اول، گذار از انسان‌محوری به کارکردمحوری در عنصر روانی، با تمرکز بر «منطق تصمیم‌گیری معیوب» به‌جای حالات ذهنی. دوم، طراحی الگوی عملیاتی سه‌لایه: مسئولیت دارایی‌محور سریع در لایه فاعلی، شخصیت الکترونیک محتاطانه، مسئولیت مبتنی بر تقصیر نوین برای عاملان انسانی در لایه سببیتی، و حمایت نهادهای نظارتی تخصصی (Ghaffari Elahi Kashani, 2018).

امکان‌سنجی بازتعریف عنصر روانی؛ از «قصد انسانی» به «عملکرد معیوب برنامه‌ریزی‌شده»

برای آنکه نظام حقوق کیفری ایران بتواند به‌طور مؤثر با چالش هوش مصنوعی روبه‌رو شود، لازم است در تفسیر خود از عنصر روانی تجدید نظر کند. این به معنای کنار گذاشتن کامل مفهوم سنتی نیست، بلکه به معنای توسعه یک درک کارکردگرا و غایت‌نگر است. می‌توان استدلال کرد که در مورد یک سامانه هوش مصنوعی خودمختار، «عنصر روانی» معادل با «عملکرد معیوب برنامه‌ریزی‌شده» یا «انحراف سیستماتیک الگوریتم از اهداف طراحی‌شده ایمن» است (Najafi Abrandabadi &

Ebrahimi, 2004). به عبارت دیگر، اگر ثابت شود که سامانه بر پایه یک الگوی تصمیم‌گیری ذاتاً خطرناک، تبعیض‌آمیز یا خارج از پارامترهای پیش‌بینی‌شده عمل کرده است، این خود می‌تواند به‌عنوان یک «قصد الگوریتمی» در نظر گرفته شود که ناشی از انتخاب‌های اولیه طراحی، مجموعه داده‌های آموزش یا مکانیزم‌های یادگیری است. این «قصد» نه یک حالت ذهنی، بلکه یک ویژگی عینی و قابل شناسایی در عملکرد سامانه است.

برای احراز این عنصر روانی نوین، نیاز به معیارهای عینی داریم: (۱) تخطی از استانداردهای فنی ایمنی مصوب مراجع ذیصلاح؛ (۲) وجود سوگیری سیستماتیک در خروجی‌های سامانه که منجر به آسیب گروه‌های خاص می‌شود؛ (۳) عملکرد غیرقابل توضیح، جعبه سیاه، در موقعیت‌های بحرانی، در حالی که توضیح‌پذیری یک الزام قانونی بوده است (Moazenadegan et al., 2015). پذیرش این تفسیر، مستلزم اصلاح قوانین موجود یا تصویب قانون خاص هوش مصنوعی است که این مفاهیم را تعریف و ضمانت‌های اجرایی آن را مشخص کند.

الگوی پیشنهادی؛ شخصیت حقوقی مجازی مشروط همراه با مسئولیت سلسله‌مراتبی

با ترکیب رویکردهای مختلف و با توجه به بستر حقوقی ایران، یک الگوی ترکیبی به شرح زیر پیشنهاد می‌گردد:

اعطای شخصیت حقوقی مجازی مشروط

به سامانه‌های هوش مصنوعی با درجه خودمختاری «بالا»، طبق تعریف مقررات آینده، یک «شخصیت حقوقی الکترونیک مشروط» اعطا شود. این شخصیت به معنای اهلیت کامل کیفری نیست، بلکه به سامانه این توانایی را می‌دهد که دارایی مستقل، یک صندوق مالی برای بهره‌برداری یا بیمه اجباری، داشته باشد و به‌عنوان طرف مسئول در دعاوی مدنی شناخته شود. مجازات آن محدود به جریمه نقدی از محل همان دارایی، تعلیق مجوز بهره‌برداری یا انحلال و از رده خارج کردن سامانه خواهد بود، شبیه انحلال

شفافیت، استنادپذیری، لاگ‌گذاری اجباری و دسترسی به منطق تصمیم‌گیری برای رفع معضل «جعبه سیاه» ضروری است. در نهایت، این قانون باید چارچوب الزام‌آور توزیع مسئولیت سلسله‌مراتبی، از جمله مسئولیت محض برای سامانه‌های پرخطر و شرایط مسئولیت تضامنی را تعریف نماید (Tavakoli Rad & Miri, 2021).

تشکیل مرکز ملی اخلاق و ایمنی هوش مصنوعی

حقوق نمی‌تواند به‌تنهایی و بدون کمک دانش فنی، بر پیچیدگی‌های هوش مصنوعی نظارت کند. تشکیل یک نهاد فراقوه‌ای و تخصصی متشکل از حقوقدانان، مهندسان هوش مصنوعی، متخصصان اخلاق، فیلسوفان و نمایندگان جامعه مدنی ضروری است. مأموریت‌های کلیدی این مرکز عبارتند از:

الف. رده‌بندی سامانه‌ها بر اساس سطح ریسک. این مرکز با تعیین معیارهای دقیق، مانند دامنه عمل، شدت آسیب، میزان خودمختاری، سامانه‌های هوش مصنوعی را در رده‌های مختلف، نظیر «غیرقابل قبول»، «پرریسک»، «ریسک محدود»، «حداقل ریسک»، طبقه‌بندی می‌کند. این طبقه‌بندی مستقیماً تعیین‌کننده سطح نظارت، الزامات صدور مجوز و نوع نظام مسئولیت حاکم بر هر سامانه خواهد بود.

ب. تدوین و تصویب استانداردهای فنی. این مرکز مسئول تبدیل اصول کلی قانون به دستورالعمل‌های عملیاتی و استانداردهای فنی برای صنعت خواهد بود.

ج. بررسی حوادث و صدور نظریات کارشناسی. در صورت وقوع حادثه، این مرکز به‌عنوان مرجع کارشناسی بی‌طرف، به بررسی فنی ریشه‌های حادثه می‌پردازد. نظریه کارشناسی این مرکز باید در دادگاه‌ها از اعتبار ویژه‌ای برخوردار باشد و بتواند در احراز رابطه سببی و تقصیر کمک شایانی به قضات نماید.

شخص حقوقی. این مکانیسم، شکاف مسئولیت را بدون ورود به مباحث پیچیده مجازات‌های سنتی می‌پوشاند (Tabrizi et al., 2022).

مسئولیت سلسله‌مراتبی و توأم انسان‌ها

مدل «مسئولیت سلسله‌مراتبی و توأم» چارچوبی عملی برای انتساب مسئولیت در قبال «عملکرد معیوب برنامه‌ریزی‌شده» هوش مصنوعی ارائه می‌دهد. سطح اول مسئولیت کیفری توسعه‌دهنده بر مبنای «تقصیر شدید» یا سوءنیت در طراحی است (Fahimi, 2001). سطح دوم مسئولیت کیفری کاربر حرفه‌ای مشروط به «تخطی از تکلیف نظارتی» و قصور در مداخله معنادار است (Farahmand, 2024). در نهایت، سطح سوم، مسئولیت مدنی محض را برای جبران قطعی خسارت قربانی از طریق صندوق یا بیمه اجباری توسعه‌دهنده ایجاد می‌کند (Mousavi, 2011). این سطوح سه‌گانه مانع‌الجمع نیستند و شبکه ایمنی حقوقی جامعی برای عدالت توزیعی و بازدارندگی ایجاد می‌کنند.

الزامات نهادی و تقنینی برای اجرای الگو

اجرای موفقیت‌آمیز الگوی ترکیبی بازتعریف عنصر روانی و مسئولیت سلسله‌مراتبی که در فصل پیشین ارائه شد، مستلزم فراتر رفتن از مباحث نظری و استقرار یک چارچوب نهادی، قانونی و اجرایی قدرتمند است. بدون این بسترها، بهترین ایده‌های حقوقی نیز در عمل با شکست مواجه خواهند شد. چهار رکن اصلی این چارچوب به شرح زیر است.

تصویب قانون جامع حکمرانی و مسئولیت هوش مصنوعی

نخستین الزام، خروج از پراکندگی تقنینی موجود در قوانین عمومی ایران و تصویب یک قانون جامع برای هوش مصنوعی است. این قانون باید تعاریف حقوقی پایه شامل «شخصیت الکترونیک مشروط» و دارایی مجازی آن را مشخص کند. همچنین، قانون موظف به تعیین استانداردهای ایمنی طراحی و به‌روزرسانی آن‌ها است که معیار «تقصیر شدید» توسعه‌دهنده خواهد بود. الزامات

در کنار قاضی، برای تحلیل ادله فنی، بررسی گزارش‌های مرکز ملی اخلاق و ایمنی و تفسیر عملکرد سامانه.

ج. آیین دادرسی متناسب. طراحی سازوکارهایی برای حفظ و بررسی ادله دیجیتال، رعایت حریم اسرار تجاری در حین شفافیت لازم برای دادرسی، و استفاده از ابزارهای تحلیل داده در فرآیند رسیدگی (Jalali Farahani & Bagheri Asl, 2008).

بنابراین، این چهار بستر نه به صورت مجزا، بلکه به شکل یک اکوسیستم به هم پیوسته عمل می‌کنند. قانون جامع، چارچوب کلی و الزامات را تعیین می‌کند. مرکز ملی، این الزامات کلی را به استانداردهای فنی تبدیل و بر اجرای آن نظارت تخصصی می‌کند. نظام بیمه، پشتوانه مالی و تضمین عملی جبران خسارت را فراهم می‌سازد. و در نهایت، دادگاه تخصصی، حل و فصل عادلانه و آگاهانه اختلافات در پرتو تمامی این سازوکارها را بر عهده می‌گیرد. تنها با استقرار هم‌زمان این ارکان است که الگوی پیشنهادی این مقاله می‌تواند از قوه به فعل درآمده و پاسخگوی چالش‌های بی‌سابقه هوش مصنوعی در عرصه حقوق کیفری باشد.

نتیجه‌گیری و پیشنهادات

فناوری هوش مصنوعی خودمختار، مفاهیم پایه‌ای حقوق کیفری از جمله عنصر روانی و انتساب جرم را به چالش کشیده است. نظام حقوقی ایران، که ذاتاً شخص‌محور و مبتنی بر اراده و قصد انسانی است، در وضعیت فعلی فاقد ابزارهای لازم برای پاسخگویی مؤثر به رفتارهای زیان‌بار ناشی از این سامانه‌هاست. رویکردهای موجود، از انکار مسئولیت مستقل تا اعطای شخصیت حقوقی، هر کدام نقاط قوت و ضعف خود را دارند. بررسی این رویکردها نشان می‌دهد که هیچ‌یک به تنهایی قادر به حل کامل معضل «شکاف مسئولیت» نیستند.

هوش مصنوعی یک «چالش وجودی» برای حقوق کیفری محسوب می‌شود. عبور از این چالش، نه با انجماد بر مفاهیم قرن نوزدهمی، بلکه با شجاعت نظری برای بازخوانی این مفاهیم در

د. نظارت بر رعایت اصول اخلاقی. نظارت بر مسائلی مانند جلوگیری از سوگیری الگوریتمی، حفظ حریم خصوصی داده‌ها و شفافیت نیز از وظایف این مرکز است (Farahmand, 2024).

نظام بیمه اجباری برای سامانه‌های پرخطر

برای تضمین جبران سریع و قطعی خسارت به قربانیان و ایجاد اطمینان در جامعه، ایجاد یک نظام بیمه اجباری برای سامانه‌های طبقه‌بندی‌شده در رده «پرریسک»، مانند خودروهای خودران، ربات‌های جراح، سیستم‌های کنترل حیاتی، ضروری است.

الف. الزام قانونی. قانون باید توسعه‌دهنده یا بهره‌بردار اصلی را موظف به تهیه بیمه‌نامه مسئولیت مدنی با پوشش کافی نماید. این بیمه، منبع مالی اولیه برای جبران خسارات وارده توسط سامانه خواهد بود و مکمل «صندوق مسئولیت» شخصیت الکترونیک مشروط عمل می‌کند.

ب. کارکردهای دوگانه. این نظام نه تنها از حقوق قربانیان حمایت می‌کند، بلکه با تعرفه‌گذاری خطر محور، انگیزه‌ای اقتصادی برای صنعت ایجاد می‌نماید تا با رعایت استانداردهای ایمنی بیشتر، هزینه بیمه خود را کاهش دهند. شرکت‌های بیمه نیز خود به عنوان ناظران ثانویه، بر رعایت پروتکل‌های ایمنی نظارت خواهند کرد (Shirzad, 2009).

تأسیس دادگاه‌های تخصصی جرایم سایبری و فناوری

رسیدگی به پرونده‌های پیچیده ناشی از هوش مصنوعی در دادگاه‌های سنتی، به دلیل نیاز به درک هم‌زمان عمیق از حقوق و فناوری، با دشواری روبه‌روست. تشکیل شعب تخصصی در دادگاه‌ها با ترکیب زیر ضروری است:

الف. قضات آموزش‌دیده. قضاتی که دوره‌های تخصصی در زمینه حقوق فناوری، جرایم سایبری و مبانی هوش مصنوعی را گذرانده‌اند.

ب. هیئت‌های مشاوره فنی دائم. حضور دائم یا موردی مشاوران فنی، مانند مهندسان هوش مصنوعی، کارشناسان امنیت سایبری،

۳. تدوین قانون مسئولیت سلسله‌مراتبی. قانونی تصویب شود که مسئولیت کیفری برای انسان‌ها و مدنی برای سامانه و انسان‌ها را به‌صورت توأم و بر اساس سطح کنترل و تقصیر تعیین کند. در این قانون، مسئولیت اولیه جبران خسارت بر عهده صندوق سامانه یا بیمه اجباری، و مسئولیت کیفری در موارد تقصیر شدید بر عهده توسعه‌دهنده یا کاربر باشد.

۴. اصلاح قوانین آیین دادرسی کیفری. با توجه به معضل «جعبه سیاه»، مقرراتی برای حق دسترسی به منطق الگوریتم، حق تفسیر، در فرآیند دادرسی، با حفظ اسرار تجاری ضروری، پیش‌بینی شود. همچنین، ادله الکترونیکی تولیدشده توسط هوش مصنوعی باید با ضوابط دقیق صحت و استنادپذیری، مانند استفاده از امضای دیجیتال و زنجیره بلوکی، مورد پذیرش قرار گیرند.

۵. تقویت همکاری‌های بین‌المللی. با توجه به فرامرزی بودن فضای سایبر، ایران باید در زمینه تعیین صلاحیت، استرداد مجرمان و همکاری‌های قضایی در پرونده‌های مرتبط با هوش مصنوعی، با سایر کشورها و نهادهای بین‌المللی همکاری فعالانه داشته باشد.

مشارکت نویسندگان

در نگارش این مقاله تمامی نویسندگان نقش یکسانی ایفا کردند.

تعارض منافع

در انجام مطالعه حاضر، هیچ‌گونه تضاد منافی وجود ندارد.

EXTENDED ABSTRACT

The rapid diffusion of artificial intelligence into transportation, medicine, finance, public administration, commercial contracting, and judicial decision-making has generated a profound challenge for criminal law, particularly in relation to the concepts of the mental element and attribution of criminal responsibility. Classical criminal law is built upon the assumption that crime is committed by a subject capable of intention, knowledge, negligence, moral blameworthiness, and voluntary action. This assumption becomes

پرتو واقعیت‌های جدید ممکن است. پیشنهاد‌های این مقاله، اگرچه نیازمند بحث‌های گسترده‌تر و اصلاحات جزئی هستند، اما مسیری را نشان می‌دهند که می‌تواند تعادل مناسبی بین حمایت از قربانیان، حفظ انصاف برای توسعه‌دهندگان و کاربران، و عدم خفه‌سازی نوآوری فناوریانه برقرار کند. زمان برای اقدام اکنون است، قبل از آنکه پیچیدگی فناوری از توانایی قانون برای تنظیم آن پیشی بگیرد و «بی‌کیفرمانی فناوریانه» به هنجار جدیدی در عصر هوش مصنوعی تبدیل شود.

با توجه به یافته‌های این پژوهش، پیشنهاد‌های زیر برای نظام حقوقی ایران ارائه می‌گردد:

۱. بازتعریف قانونی عنصر روانی. در قانون آینده هوش مصنوعی، مفهوم «عملکرد معیوب برنامه‌ریزی‌شده» یا «انحراف سیستماتیک الگوریتم» به‌عنوان معادل عنصر روانی برای سامانه‌های خودمختار شناسایی شود. احراز آن منوط به اثبات نقض استانداردهای ایمنی، وجود سوگیری ذاتی یا عملکرد غیرقابل توضیح در شرایط بحرانی باشد.

۲. پذیرش شخصیت حقوقی الکترونیک مشروط. برای سامانه‌های با درجه خودمختاری بالا، طبق تعریف مرجع ملی، شخصیت حقوقی محدود جهت دارایی مستقل، صندوق جبران خسارت، و مسئولیت مدنی مستقیم به رسمیت شناخته شود. این شخصیت مجوزی برای مجازات کیفری به معنای سنتی نیست.

unstable when harmful conduct results from autonomous or semi-autonomous artificial intelligence systems that operate through algorithmic learning, probabilistic inference, and independent decision-making rather than direct human command. In Iranian criminal law, as in many civil-law systems, criminal liability remains fundamentally person-centered and presupposes a human or legally recognized person capable of bearing responsibility. The emergence of autonomous artificial intelligence therefore creates a structural tension between traditional

doctrines of criminal liability and the practical realities of algorithmic agency. Existing legal frameworks concerning cybercrime, electronic evidence, data governance, and digital technologies provide partial guidance, but they do not fully resolve whether an intelligent system can itself be treated as a criminal actor or whether responsibility must remain exclusively attached to designers, manufacturers, programmers, operators, or users (Alipour, 2011; Sieber & Nouri, 2011; Tabrizi et al., 2022; Tavakoli Rad & Miri, 2021). The central issue addressed in this study is the feasibility of redefining the mental element and the attribution of criminal responsibility in criminal law related to artificial intelligence. The mental element traditionally refers to the internal relationship between the offender and the prohibited act or result, including intent, knowledge, recklessness, negligence, and fault. These categories are meaningful when applied to human beings because they presuppose consciousness, rational choice, moral understanding, and the capacity to be blamed. Artificial intelligence systems, however complex, do not possess consciousness, emotional awareness, moral regret, or free will in the philosophical and legal sense. Their outputs are produced through programmed architectures, training data, optimization functions, and computational processes. Accordingly, attributing intent or negligence to artificial intelligence in the classical sense appears conceptually problematic. Nevertheless, the increasing autonomy of such systems, especially in high-risk contexts, makes it insufficient to treat them merely as passive tools. When a self-driving vehicle, medical robot, financial algorithm, or autonomous weapon produces harmful consequences that were not directly intended or predicted by any human actor, criminal law faces a responsibility gap. This gap is particularly serious where neither the developer nor the

user can easily be shown to have acted with traditional fault, while the harmful outcome is directly linked to the autonomous operation of the system (Hekmatnia et al., 2019; Keshtkar & Hajilou, 2024; Masoumi Janaqerd & Zarei, 2024; Zakerinia, 2023).

The study adopts a descriptive–analytical approach and examines the theoretical foundations of criminal liability, the nature of artificial intelligence autonomy, and possible models for legal attribution. The analysis shows that the classical Iranian criminal law framework is not easily compatible with direct criminal liability of artificial intelligence. The reasons are both theoretical and practical. Theoretically, artificial intelligence lacks legal consciousness, moral agency, culpability, and capacity for blame. Practically, traditional punishments such as imprisonment, rehabilitation, deterrence, and moral condemnation cannot be meaningfully applied to a non-human digital entity. Even where disabling, deleting, restricting, or updating a system may prevent future harm, such measures are functionally regulatory rather than punitive in the traditional criminal-law sense. This limitation also affects the purposes of punishment, including retribution, reform, rehabilitation, and deterrence, which are designed around human psychology and social responsibility. Although criminal law has already expanded the notion of legal personality to include corporations and institutions, the analogy between legal persons and artificial intelligence is incomplete, because corporate criminal liability still ultimately depends on human organs, representatives, or decision-makers. Artificial intelligence, by contrast, may generate operational outcomes through computational processes that are not reducible to a single human intention or decision (Basami, 2020; Katouzian, 1995; Najafi Abrandabadi & Ebrahimi, 2004; Rahimi Nejad, 2010; Rosenhan et al., 2010; Sadeghi Neshat, 1997).

A major finding of the study is that the responsibility gap is intensified by the black-box problem. Many advanced artificial intelligence models, especially those based on machine learning and deep learning, produce decisions through complex layers of parameters that are difficult to interpret even for their designers. This creates evidentiary and procedural difficulties in criminal adjudication. First, proving causation becomes difficult when the precise computational pathway between input data and harmful output cannot be clearly reconstructed. Second, proving human fault becomes challenging when no specific defect in design, training, deployment, or supervision can be identified with certainty. Third, the right of defense may be undermined where a judicial or administrative decision relies on an algorithmic assessment whose reasoning cannot be meaningfully challenged. These concerns demonstrate that artificial intelligence is not merely a technological problem but also a procedural and evidentiary problem for criminal justice. The reliability, admissibility, integrity, and traceability of electronic evidence become crucial in such cases. Therefore, any future legal framework must include duties of transparency, explainability, logging, auditability, and preservation of digital evidence, especially for high-risk systems whose outputs may affect life, liberty, property, privacy, or public security (Farahmand, 2024; Fathi, 2017; Ghaffari Elahi Kashani, 2018; Jafari & Rahbarpour, 2017; Moazenzadegan et al., 2015; Salehi, 2019; Tabrizi et al., 2022).

The article argues that a purely conservative approach, which treats artificial intelligence only as a tool, is insufficient in cases of high autonomy, while a radical approach, which grants full criminal personality to artificial intelligence, is also theoretically and normatively problematic. A more defensible solution is a hybrid model based on functional

reinterpretation of the mental element and hierarchical attribution of responsibility. Under this model, the mental element should not be transferred mechanically from human psychology to machines. Instead, for autonomous artificial intelligence systems, the functional equivalent of the mental element may be understood as programmed defective performance, systematic algorithmic deviation, or dangerous decision-making logic embedded in the design, training data, learning mechanism, or operational architecture of the system. This does not mean that artificial intelligence has consciousness or guilt; rather, it means that the law may identify objective indicators of culpability-like malfunction, including violation of technical safety standards, foreseeable algorithmic bias, failure of explainability in critical contexts, inadequate testing, defective data governance, or deployment of a system beyond its validated operational domain. At the same time, responsibility should be distributed hierarchically among developers, manufacturers, suppliers, professional operators, users, insurers, and regulatory bodies according to their level of control, benefit, expertise, and capacity to prevent harm. For very high-risk systems, limited conditional electronic legal personality may be recognized only for civil liability, financial responsibility, insurance, compensation funds, suspension of operation, or deactivation, without equating such systems with human criminal persons (Calerick et al., 2010; Fahimi, 2001; Farajiha & Alamdari, 2017; Katouzian, 2013; Khodaghali, 2004; Mousavi, 2011; Shirzad, 2009).

In conclusion, the study maintains that autonomous artificial intelligence requires a careful reconstruction, not abandonment, of core criminal-law concepts. The traditional categories of intent, negligence, fault, attribution, and punishment cannot be applied to artificial intelligence without conceptual

adjustment; however, the growing autonomy of intelligent systems makes it equally unacceptable to leave serious harms without effective legal response. The most balanced path is a hybrid framework that preserves human accountability while creating functional mechanisms for risk allocation, victim compensation, regulatory oversight, and technical accountability. Such a framework should include a comprehensive artificial intelligence responsibility law, mandatory safety and explainability standards, compulsory insurance for high-risk systems, specialized regulatory institutions, expert technical review, and specialized courts or judicial branches for cyber and technology-related offenses. This approach can reduce impunity, protect victims, encourage responsible innovation, and prevent the excessive criminalization of technological development. The future of criminal law in the age of artificial intelligence therefore depends on its ability to move beyond rigid human-centered assumptions while still preserving the normative foundations of justice, fairness, accountability, and social protection.

References

- Alipour, H. (2011). *Information Technology Criminal Law*. Khorsandi.
- Basami, A. (2020). Analysis of Ideal Concurrence of Crimes in Iranian Criminal Law: With Reference to Article 131 of the 2013 Code. *Criminal Law Research*, 49, 41-65.
- Calerick, A. M., Janczewski, L., & Ebrahim Nejad, S. (2010). *An Introduction to Cyber Warfare and Cyber Terrorism*. Boostan Hamid.
- Code of Criminal Procedure, (2013).
- Fahimi, M. (2001). Computer Crimes and Methods of Combating and Preventing Them. *Legal Perspectives Quarterly, Faculty of Judicial Sciences and Administrative Services*(23/24), 21-36.
- Farahmand, F. (2024). Personal Data and Information Protection Bill with a Comparative Study of the General Data Protection Regulation. *Law of New Technologies*, 5(9), 17-25.
- Farajiha, M., & Alamdari, A. (2017). A Comparative Study of the Criteria for Criminalization in Cyberspace in Iran and Germany. *Comparative Law Studies*, 8(2), 637-653.
- Fathi, Y. (2017). Scope and Domain of Privacy in Cyberspace. *Law and Politics Biannual*, 45, 91-118.
- Ghaffari Elahi Kashani, N. (2018). Territorial Jurisdiction of Iranian Courts over Crimes Committed in Cyberspace. *Legal Research*, 185-218.
- Hekmatnia, M., Mohammadi, M., & Vatheghi, M. (2019). Civil Liability Arising from the Production of Autonomous Artificial Intelligence Robots. *Islamic Law*, 15(16), 231-255.
- Jafari, A., & Rahbarpour, M. R. (2017). Civil Liability for Violation of Data Privacy in Imami Jurisprudence and Positive Law. *Private Law Research Quarterly*, 5(18), 44-52.
- Jalali Farahani, A., & Bagheri Asl, R. (2008). Social Prevention of Cyber Crimes as the Main Solution for Institutionalizing Cyber Ethics. *Rahavard Noor Information Science and Librarianship Quarterly*, 24(41), 121-156.
- Katouzian, N. (1995). *Non-Contractual Obligations: Civil Liability*. University of Tehran Press.
- Katouzian, N. (2013). *Civil Liability*. University of Tehran Press.
- Keshtkar, Z., & Hajilou, A. (2024). *The Impact of the Use of Artificial Intelligence in International and Domestic Arbitration*. Masharegh al-Anvar Publications.
- Khodagholi, Z. (2004). *Computer Crimes*. Ariyan Publications.
- Masoumi Janaqerd, I., & Zarei, G. (2024). Consequences and Effects of the Use of Artificial Intelligence in Commercial Contracts from Tourism, Legal, and Business Perspectives. *Journal of Geography and Human Relations*, 7(3), 533-557.
- Moazenadegan, H. A., Soleiman Dehkordi, E., & Youshi, M. (2015). Preserving the Integrity and Admissibility of Electronic Evidence Using Biometrics and Cryptography. *Criminal Law Research*, 4(12), 69-97.
- Mousavi, S. R. (2011). Situational Prevention of Cyber Crimes through Technical Measures and Their Limitations. Regional Conference on the Challenges of Computer Crimes in the Present Age.
- Najafi Abrandabadi, A. H., & Ebrahimi, S. (2004). *Topics in Criminal Sciences: Notes from Criminology Courses at the Doctoral and Master's Levels*. Shahid Beheshti University.
- Rahimi Nejad, E. (2010). *Criminology*. Forouzesh.
- Rosenhan, D. L., Seligman, M. E. P., & Seyed Mohammadi, Y. (2010). *Abnormal Psychology*. Savalan.
- Sadeghi Neshat, A. (1997). *Protection of the Rights of Computer Software Creators*.
- Salehi, J. (2019). Access to the Data Center Information of a Foreign Government in Conflict with the Principles of Territorial Criminal Jurisdiction and Sovereignty in International

- Law. *International Legal Scientific Quarterly*, 26(60), 185-210.
- Shirzad, K. (2009). *Computer Crimes from the Perspective of Iranian Criminal Law and International Law*. Behineh Faragir.
- Sieber, U., & Nouri, M. A. (2011). *Computer Crimes*. Ganj Danesh.
- Tabrizi, S., Elahi Mansh, M. ج., Alipour, H., Tahmasbi, J., & Fazli, M. (2022). The Principle of Legality in the Seizure of Data and Systems in Criminal Procedure: Effects and Guarantees. *Legal Excellence*, 13(2), 125-134.
- Tavakoli Rad, R., & Miri, H. S. (2021). Governance Transformation: From Electronic Government to Digital Governance. *Exalted Governance Quarterly*, 2(7), 115-162.
- Zakerinia, H. (2023). Nature and Basis of Civil Liability Arising from Artificial Intelligence in Iranian Law and European Union Countries. *Scientific Journal of Private Law*, 20(1), 135-152.