

۱. سعید میرلوحی، دانشجوی دکتری فقه و حقوق، مدرسه عالی و دانشگاه شهید مطهری (ره)،
تهران، ایران.

Email: mirlohi@motahari.ac.ir

قابل انتشار (دوره ۱۳ شماره ۲۷ زمستان ۱۴۰۴)، فصلنامه پژوهش‌های حقوق جزا و جرم‌شناسی

چالش اثبات رکن روانی در جرایم ارتكابی توسط هوش مصنوعی و راهکار گذار به مسئولیت مبتنی بر ریسک در فقه و حقوق کیفری

چکیده

ظهور هوش مصنوعی خودمختار، نظام حقوق کیفری سنتی را که بر محوریت «رکن روانی» استوار است، با یک بحران بنیادین مواجه ساخته است. از آنجا که اثبات قصد مجرمانه در یک عامل فاقد روان، ممتنع و چالشی حل‌ناشدنی به نظر می‌رسد، این وضعیت به ایجاد یک «خال مسئولیت» انجامیده که حقوق بزهدیدگان و نظم عمومی را تهدید می‌کند. پژوهش حاضر با روشی توصیفی-تحلیلی، در مقام پاسخ به این چالش، این پرسش را مطرح می‌سازد که آیا گذار از پارادایم مسئولیت شخص محور و انتقال آن به انسان (توسعه‌دهنده) بر مبنای یک الگوی «مسئولیت مبتنی بر ریسک»، می‌تواند راهکاری کارآمد در فقه امامیه و حقوق کیفری ایران باشد. یافته‌های پژوهش نشان می‌دهد که این گذار نه تنها یک ضرورت عملی، بلکه از بعد نظری نیز کاملاً قابل دفاع است. این الگو ریشه در مبانی مستحکمی در فقه امامیه (از جمله قواعدی چون تسبیب، لاضرر، ضمان ید و تحلیل اصولی علم اجمالی) و همچنین ظرفیت‌های حقوق موضوعه ایران (شامل اصل ۴۰ قانون اساسی، ماده ۱۴۳ قانون مجازات اسلامی و رویه قانون‌گذار در قوانین خاص) دارد. در نهایت، این مقاله استدلال می‌کند که با عبور از تلاش ناکارآمد برای کشف «ذهن ماشین» و تمرکز بر «عمل خطرناک انسان» در ایجاد و عرضه فناوری، می‌توان به یک راه حل عادلانه برای تضمین مسئولیت در عصر هوش مصنوعی دست یافت.

کلیدواژه‌ها: هوش مصنوعی، مسئولیت کیفری، رکن روانی، مسئولیت مبتنی بر ریسک، فقه امامیه

جهان معاصر در میانه چهارمین انقلاب صنعتی، شاهد ظهور و گسترش پدیده‌ی هوش مصنوعی خودمختار^۱ است. این فناوری که از توانایی یادگیری، تحلیل حجم عظیمی از داده‌ها و تصمیم‌گیری مستقل بر اساس آن برخوردار است، دیگر یک مفهوم علمی-تخیلی نیست، بلکه واقعیتی است که در حوزه‌هایی با حساسیت بالا به کار گرفته شده است؛ برای مثال، گزارش‌های اخیر نشان می‌دهد که معاملات الگوریتمی، که عمدتاً توسط سامانه‌های هوش مصنوعی هدایت می‌شوند، در بازارهای مالی توسعه یافته بیش از ۷۰ درصد از حجم کل معاملات سهام را تشکیل می‌دهند^۲ و در حوزه پزشکی نیز، الگوریتم‌های یادگیری عمیق در تشخیص برخی از انواع سرطان از طریق تحلیل تصاویر رادیولوژی، به دقتی برابر یا حتی فراتر از متخصصان رادیولوژیست انسانی دست یافته‌اند^۳. این نفوذ روزافزون در سامانه‌های مدیریت هوشمند بازار سرمایه، الگوریتم‌های تشخیص و جراحی پزشکی، خودروهای تماماً خودران و حتی سامانه‌های دفاعی و تسلیحاتی، در حالی که فرصت‌های بی‌شماری برای پیشرفت و رفاه فراهم می‌آورد، نظام‌های حقوقی را با چالشی بی‌سابقه و عمیق روبرو کرده است. حقوق کیفری، به‌عنوان یکی از ستون‌های اصلی نظم اجتماعی که وظیفه واکنش به هنجارشکنانه‌ترین رفتارها را بر عهده دارد، در مرکز این چالش قرار گرفته است. پرسش بنیادین و اضطراری این است: هنگامی که یک عامل هوشمند اما فاقد آگاهی، اراده و وجدان انسانی، رفتاری را مرتکب می‌شود که در قوانین کیفری جرم‌انگاری شده، نظام عدالت کیفری چه پاسخی باید بدهد؟ بحران از آنجا آغاز می‌شود که شاکله و فلسفه حقوق کیفری کلاسیک، بر محوریت «انسان» به‌عنوان یک فاعل مختار و آگاه استوار است و رکن مهم مسئولیت کیفری، یعنی «رکن روانی»، مفهومی عمیقاً انسانی است که به دو مؤلفه «علم» به موضوع و «اراده» در رفتار تجزیه می‌شود. قانون‌گذار برای مجازات یک فرد، باید بتواند قصد مجرمانه (سوءنیت عام و خاص) یا حداقل خطای جزایی قابل سرزنش (بی‌احتیاطی یا بی‌مبالاتی) را در روان و ذهن او احراز کند. این در حالی است که سامانه‌های هوش مصنوعی، حتی در پیشرفته‌ترین اشکال خود، یک سامانه محاسباتی و الگوریتم‌محور است. اگرچه رفتار آن ممکن است هوشمندانه و هدفمند به نظر برسد، اما این رفتار محصول پردازش داده‌ها بر اساس الگوهای آماری است، نه حاصل اراده‌ای

^۱ Autonomous AI

^۲ Kevin Rink, «The predictive ability of technical trading rules: an empirical analysis of developed and emerging equity markets», *Financial Markets and Portfolio Management*, 37, no. 4 (2023), 403-456. doi:10.1007/s11408-023-00433-2.

^۳ Alessandro Carriero et al., «Deep Learning in Breast Cancer Imaging: State of the Art and Recent Advancements in Early 2024», *Diagnostics* 14, no. 8 (2024), 848. doi:10.3390/diagnostics14080848.

آزاد یا آگاهی‌ای که بتوان آن را مورد سرزنش اخلاقی و حقوقی قرار داد. در نتیجه، تلاش برای انتساب «سوءنیت» به یک شبکه عصبی مصنوعی، تلاشی عبث و مفهوماً متناقض است. این بن‌بست نظری، یک خلأ قانونی هولناک ایجاد کرده و نظام عدالت کیفری را در برابر این نسل جدید از «فاعلان» غیرانسانی، عملاً خلع سلاح نموده است. اهمیت پرداختن به این مسئله صرفاً یک کنجکاوی آکادمیک نیست؛ ناتوانی در ارائه یک چارچوب مسئولیت کارآمد، پیامدهای ویرانگری به دنبال خواهد داشت. در وهله اول، حقوق بزه‌دیدگان به طور کامل تضییع خواهد شد. قربانی یک تصمیم اشتباه از سوی یک الگوریتم معاملاتی که منجر به ورشکستگی او شده، یا بیمار آسیب‌دیده از یک سامانه جراح خودمختار، در نظام حقوقی فعلی با مسیری مسدود برای احقاق حق خود مواجه است. در وهله دوم، این ناامنی حقوقی می‌تواند به مانعی جدی بر سر راه نوآوری و پیشرفت فناورانه بدل شود، زیرا سرمایه‌گذاران و توسعه‌دهندگان از فعالیت در محیطی که قواعد آن نامشخص و پرریسک است، اجتناب خواهند کرد. از این رو، ضروری است تا پاسخی برای این چالش فراهم آورد. با توجه به ابتدای قوانین جزایی کشور بر مبانی فقه امامیه (اصل چهارم قانون اساسی)، ارائه هرگونه راهکار تحول‌آفرین بدون سنجش و اثبات سازگاری آن با این مبانی، ناقص و در عمل ناکارآمد خواهد بود. بنابراین، این پژوهش با رویکردی دوگانه، به دنبال ارائه پاسخی بومی، عادلانه و کارآمد برای این چالش جهانی است.

در طول سال‌های اخیر، تلاش‌های نظری متعددی برای خروج از این بن‌بست صورت گرفته است. برجسته‌ترین رویکرد، تلاش برای انطباق الگوهای مسئولیت سنتی بر عاملان انسانی دخیل (مانند برنامه‌نویس، تولیدکننده، مالک یا کاربر) بوده است. برای نمونه، گابریل هالوی^۴ به تفصیل سه مدل را برای مسئولیت کیفری هوش مصنوعی پیشنهاد می‌دهد: مدل ارتکاب به وسیله غیر^۵، مدل مسئولیت نیابتی^۶ و مدل مسئولیت محض ناشی از محصول^۷؛ با این حال، چنانکه منتقدانی چون متیو شیرر^۸ اشاره کرده‌اند، این مدل‌ها در مواجهه با سامانه‌های کاملاً خودمختار که رفتارشان غیرقابل پیش‌بینی و منطقتشان توضیح‌ناپذیر است، با چالش‌های اساسی روبرو می‌شوند و نمی‌توانند به شکل کارآمدی، رکن روانی لازم برای جرایم عمدی را احراز نمایند. به بیان دیگر، این راه‌حل‌ها عمدتاً برای پوشش جرایم غیرعمدی یا مسئولیت‌های محض مناسب به نظر می‌رسند، اما در حل معمای اصلی، یعنی فقدان سوءنیت در جرایم عمدی، ناکام می‌مانند. برخی متفکران برای خروج کامل از این بن‌بست، راهکارهای رادیکال‌تری مانند اعطای نوعی شخصیت حقوقی به هوش مصنوعی را مطرح کرده‌اند، بحثی که به طور جدی توسط لارنس سولوم^۹ آغاز شد، هرچند این رویکرد نیز چالش‌های عمیق فلسفی و حقوقی خاص خود را به همراه دارد. بنابراین، پرسش اصلی این پژوهش آن است که: «آیا می‌توان با گذار از پارادایم مسئولیت کیفری شخص محور و مبتنی بر تقصیر، به الگوی مسئولیت مبتنی بر ایجاد ریسک، چالش فقدان رکن روانی در جرایم ارتكابی توسط هوش مصنوعی را در نظام فقهی و حقوق کیفری ایران مرتفع ساخت؟» در پاسخ، فرضیه این مقاله بر آن است که ظرفیت‌های قابل توجه در فقه امامیه، به‌ویژه قواعدی همچون قاعده لاضرر، اتلاف، تسبیب و بنای عقلا، در تلفیق با مبانی نظری حقوق کیفری مدرن، این امکان را فراهم می‌آورد که مدلی کارآمد و عادلانه از مسئولیت کیفری مبتنی بر ایجاد ریسک نامتعارف، فارغ از احراز عنصر روانی به شیوه سنتی، برای توسعه‌دهندگان، مالکان و به‌کارگیرندگان سامانه‌های هوش مصنوعی خودمختار طراحی و تبیین گردد. بنابراین، این پژوهش مسئولیت خود هوش مصنوعی را «نادیده» نمی‌گیرد، بلکه با پذیرش امتناع نظری و عملی آن در چارچوب حقوق کیفری فعلی، به دنبال ارائه یک راهکار جایگزین و کارآمد برای جلوگیری از خلأ مسئولیت

⁴ Gabriel Hallevy, *A.I. and the Future of Law: The Imminent and Unprecedented Transformation of the Legal Profession* (Fastcase, 2015), 848

⁵ Perpetration-via-another

⁶ Vicarious Liability

⁷ Product Liability

⁸ Matthew U. Scherer, «Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies», *Harvard Journal of Law & Technology* 29, no. 2 (2016), 379.

⁹ Lawrence B. Solum, «Legal Personhood for Artificial Intelligences», *North Carolina Law Review* 70, no. 4 (1992), 1251

است. برای اثبات این فرضیه، مقاله حاضر در بخش نخست به تحلیل عمیق نارسایی‌ها و بن‌بست‌های الگوی مسئولیت کیفری سنتی در مواجهه با هوش مصنوعی خواهد پرداخت. در بخش دوم، مبانی نظری و ارکان الگوی مسئولیت مبتنی بر ریسک تشریح می‌شود و در بخش سوم که هسته اصلی پژوهش را تشکیل می‌دهد، ظرفیت‌های فقهی و حقوقی برای پذیرش و بومی‌سازی این الگو مورد کنکاش قرار گرفته و در نهایت، چارچوب پیشنهادی برای این مدل نوین از مسئولیت ارائه خواهد شد.

۱. مبانی و چالش‌های مسئولیت کیفری در الگوی سنتی

نظام حقوق کیفری برای تحقق عدالت و حفظ نظم عمومی، بر سه رکن بنیادین استوار است: رکن قانونی، که رفتار مجرمانه را تعریف می‌کند؛ رکن مادی، که به ظهور خارجی آن رفتار می‌پردازد؛ و رکن روانی، که پیوند روانی و ذهنی فاعل را با عمل ارتكابی برقرار می‌سازد. این تقسیم‌بندی سه‌گانه، شالوده تحلیل جرم در دکتترین حقوق کیفری ایران است و عدم تحقق هر یک از این ارکان، مانع از شکل‌گیری جرم خواهد بود.^{۱۰} در میان این سه، رکن روانی جایگاهی منحصر به فرد دارد، زیرا وجه تمایز میان یک «جرم» قابل سرزنش و یک «حادثه» صرفاً زیان‌بار است. این رکن، شخصیت فاعل و اراده آگاهانه او را به عمل مجرمانه پیوند می‌زند و مبنای اصلی «تقصیر» و «سرزنش‌پذیری» در حقوق جزا محسوب می‌شود. بدون احراز این پیوند ذهنی، مجازات فاعل از فلسفه خود تهی شده و به واکنشی فاقد مبنای اخلاقی تبدیل می‌گردد.

۱-۱. جایگاه بنیادین رکن روانی

رکن روانی، که از آن با عناوینی چون سوءنیت یا قصد مجرمانه نیز یاد می‌شود، بیانگر حالت روانی و ذهنی فاعل در هنگام ارتكاب جرم است. قانون مجازات اسلامی مصوب ۱۳۹۲ در ماده ۱۴۴، در تعریفی دقیق، این رکن را چنین معرفی می‌کند: «قصد مجرمانه (سوء نیت) در جرائم عمدی عبارت از علم مرتکب به موضوع جرم و اراده او در ارتكاب رفتار مجرمانه است». چنانکه از این تعریف قانونی پیداست و در دکتترین حقوقی نیز بر آن تأکید شده، قصد مجرمانه دارای دو پایه اساسی است: یکی «علم» و دیگری «اراده». این حالت روانی در جرایم عمدی، در قالب «قصد مجرمانه» متجلی می‌شود که خود به دو جزء اصلی قابل تفکیک است: نخست، سوءنیت عام یا اراده ارتكاب فعل، که به معنای خواست آگاهانه برای انجام دادن خود عمل ممنوعه است. دوم، سوءنیت خاص یا قصد نتیجه، که ناظر بر خواست فاعل برای دستیابی به نتیجه مشخصی است که قانون آن را جرم‌انگاری کرده است. به بیان دیگر، فاعل نه تنها باید به ماهیت عمل خود و اوصافی که از نظر قانون اهمیت دارد آگاه باشد (عنصر علمی)، بلکه باید با آزادی و اختیار، انجام آن عمل را بخواهد (عنصر ارادی).^{۱۱} برای مثال، در جرم قتل عمد، سوءنیت عام همان اراده شلیک کردن به سمت دیگری و سوءنیت خاص، قصد سلب حیات از اوست. اما در حیطه جرایم غیرعمدی، رکن روانی فاقد «قصد مجرمانه» است و در قالب دیگری تحت عنوان «خطای جزایی» یا تقصیر جزایی ظهور می‌یابد. قانون‌گذار وجود این حالت روانی قابل سرزنش را برای تحقق مسئولیت کیفری ضروری می‌داند. ماده ۱۴۵ قانون مجازات اسلامی در تعریف این وضعیت مقرر می‌دارد: «تحقق جرایم غیرعمدی، منوط به احراز تقصیر مرتکب است». این ماده در تبصره خود، تقصیر را شامل دو صورت اصلی بی‌احتیاطی و بی‌مبالاتی دانسته و غفلت، مسامحه، عدم مهارت و عدم رعایت نظامات دولتی را از مصادیق آن برمی‌شمارد. در واقع، مبنای سرزنش در اینجا، «خواستن نتیجه مجرمانه» نیست، بلکه «انحراف از الگوی رفتاری یک انسان محتاط و معقول» است.^{۱۲} فاعل به این دلیل مجازات می‌شود که با

^{۱۰} رضا نوربها، زمینه حقوق جزای عمومی (تهران: کتابخانه گنج دانش، ۱۳۹۵)، ۳۴۱.

^{۱۱} حسین میرمحمد صادقی، جرایم علیه اشخاص (تهران: نشر میزان، ۱۳۹۸)، ۲۵۱.

^{۱۲} محمدعلی اردبیلی، حقوق جزای عمومی، ج ۲ (تهران: نشر میزان، ۱۴۰۰)، ۸۵.

وجود توانایی و وظیفه پیش‌بینی نتایج زیان‌بار عمل خود، از آن غفلت ورزیده است. بنابراین، حتی در این جرایم نیز، نوعی «اراده» حداقلی (اراده در انجام فعلی که منجر به خطا شده) و یک وضعیت ذهنی قابل‌نکوهش، مبنای مسئولیت فرد را تشکیل می‌دهد.

ویژگی	فاعل انسانی	فاعل هوش مصنوعی
اراده آزاد	دارد	ندارد
آگاهی و شعور	دارد	ندارد
قصد مجرمانه (سوءنیت)	قابل احراز	ذاتاً ممتنع (غیرقابل تصور)
سرزنش‌پذیری اخلاقی	مبنای مسئولیت	فاقد موضوعیت

جدول ۱: تحلیل تطبیقی ویژگی‌های فاعل انسانی و هوش مصنوعی

این اصل، ریشه در مبانی فقهی نظام حقوقی ایران دارد؛ فقه امامیه، نیت و قصد را محور ارزش‌گذاری اعمال و تعیین مسئولیت می‌داند. حدیث نبوی مشهور «إِنَّمَا الْأَعْمَالُ بِالنِّيَّاتِ» به‌عنوان یک قاعده بنیادین، بر کل ابواب فقه، از جمله فقه کیفری، سایه افکنده است. بر همین اساس، فقهای عظام با دقتی مثال‌زدنی، مسئولیت را بر مدار «قصد» فاعل تفکیک کرده و میان قتل «عمد»، «شبه عمد» و «خطای محض» تمایز دقیق قائل شده‌اند که هر یک آثار حقوقی متفاوتی از جمله قصاص، دیه یا کفاره را به دنبال دارد. برای مثال، در تحقق قتل عمد که موجب قصاص است، عموماً دو شرط اساسی لازم است: نخست، «قصد فعل» یعنی اراده انجام کاری مشخص نسبت به مجنی‌علیه، و دوم، «قصد نتیجه» یعنی داشتن قصد مشخص برای سلب حیات او. در مقابل، «شبه عمد» زمانی محقق می‌شود که فاعل، قصد فعل را نسبت به قربانی داشته اما قصد نتیجه (قتل) را نداشته باشد، مانند تأدیی که منجر به فوت شود. نهایتاً، «خطای محض» وضعیتی است که در آن، نه قصد فعل نسبت به قربانی و نه قصد نتیجه وجود دارد، مانند آنکه شخصی به قصد شکار حیوانی تیراندازی کند اما تیر به یک انسان اصابت نماید^{۱۳}. این تفکیک، مستقیماً به نوع و درجه قصد فاعل بازمی‌گردد و نشان‌دهنده آن است که از منظر فقهی نیز، عمل بدون در نظر گرفتن نیت و حالت روانی فاعل آن، قابل تقویم و قضاوت کیفری نیست. در نتیجه، کل ساختار مسئولیت کیفری در هر دو نظام حقوقی عرفی و شرعی، بر این پیش‌فرض استوار است که فاعل جرم، یک «انسان» دارای اراده، آگاهی و قدرت انتخاب است. این همان پیش‌فرض بنیادینی است که اکنون با ظهور فاعلان هوشمند غیرانسانی، به لرزه درآمده است.

۱-۲. بحران کارآمدی الگوی شخص‌محور در برابر فناوری‌های نوین

همانطور که تبیین شد، رکن روانی شالوده مسئولیت کیفری است. حال، پرسش این است که چرا الگوی مسئولیت شخص‌محور، که بر همین رکن استوار است، در مواجهه با فناوری‌های نوین دچار بحران کارآمدی می‌شود؟ پاسخ آن است که این الگو تا زمانی کارآمد است که پیوندی مستقیم و قابل‌فهم میان ذهنیت یک انسان و عمل مجرمانه وجود داشته باشد. این الگو، فناوری را به‌عنوان یک «ابزار» خنثی و در کنترل کامل انسان در نظر می‌گیرد؛ چه آن ابزار یک چاقو باشد و چه یک برنامه کامپیوتری ساده. در این نگاه، ابزار تنها واسطه‌ای برای تحقق اراده انسان است و مسئولیت به سادگی به شخصی که آن را به کار گرفته، منتسب می‌شود. با این حال،

^{۱۳} جعفر بن حسن محقق حلی، شراعی الإسلام فی مسائل الحلال و الحرام، ج ۴ (قم: مؤسسه اسماعیلیان، ۱۴۰۸ق)، ۱۹۵.

ظهور فناوری‌های خودمختار و به‌ویژه هوش مصنوعی مبتنی بر یادگیری ماشین، این پیش‌فرض را به طور کامل فرو ریخته است. ویژگی متمایزکننده هوش مصنوعی مدرن، صرفاً سرعت یا پیچیدگی آن نیست، بلکه توانایی آن در «یادگیری» و تکامل مستقل از برنامه‌ریزی اولیه است.^{۱۴} این سامانه‌ها که اغلب بر پایه شبکه‌های عصبی عمیق عمل می‌کنند، از طریق تحلیل مجموعه داده‌های عظیم، خود به استنتاج و بهینه‌سازی رویه‌های خود می‌پردازند. این فرآیند منجر به ظهور پدیده‌ای می‌شود که از آن با عنوان «رفتار نوپدید»^{۱۵} یاد می‌شود؛ یعنی اتخاذ تصمیمات و اقداماتی که به صورت صریح در کد اولیه توسط طراحانشان پیش‌بینی یا برنامه‌ریزی نشده‌اند. افزون بر این، بسیاری از این سامانه‌ها به دلیل پیچیدگی غیرقابل تصور لایه‌های پردازشی و میلیون‌ها پارامتر داخلی خود، به «جعبه سیاه» شهرت دارند؛ به این معنا که حتی سازندگان آن‌ها نیز نمی‌توانند به طور دقیق و با قطعیت توضیح دهند که سامانه بر اساس چه منطق استنتاجی به یک تصمیم مشخص رسیده است. این «مشکل توضیح‌ناپذیری»^{۱۶}، چنانکه پژوهشگرانی چون آزارو^{۱۷} مطرح کرده‌اند، به طور مستقیم ارکان مسئولیت سنتی مانند قابلیت پیش‌بینی و احراز رابطه سببیت را زیر سؤال می‌برد. افزون بر این، عدم شفافیت این سامانه‌ها، تنها یک چالش فنی نیست، بلکه چنانکه فرانک پاسکوال در اثر خود «جامعه جعبه سیاه» به درستی استدلال می‌کند، به یک بحران حاکمیتی و اجتماعی تبدیل شده است؛ جایی که تصمیمات حیاتی برای زندگی شهروندان توسط الگوریتم‌هایی اتخاذ می‌شود که منطق آن‌ها بر نهادهای نظارتی و حتی خود کاربران پوشیده است.^{۱۸} این عدم شفافیت، قدرت پاسخگو کردن صاحبان فناوری را به شدت تضعیف می‌کند. از سوی دیگر، رایان کالو نیز اشاره می‌کند که ربات‌ها و سامانه‌های هوشمند، برخلاف ابزارهای سنتی، به دلیل تعامل و خودمختاری خود، یک «حضور اجتماعی» در زندگی انسان پیدا می‌کنند که باعث می‌شود افراد، آن‌ها را چیزی فراتر از یک شیء بی‌جان تلقی کنند. این ویژگی، قیاس ساده‌انگارانه هوش مصنوعی با ابزار را بیش از پیش ناکارآمد می‌سازد.^{۱۹} این ویژگی‌ها، زنجیره سنتی انتساب مسئولیت را قطع می‌کند. سناریوی زیر را در نظر بگیرید:

یک سامانه هوش مصنوعی که برای مدیریت سبد سهام در بازار بورس طراحی شده، به طور مستقل و بر اساس تحلیل داده‌های بازار، یک راهبرد نوین برای دستکاری بازار ابداع و اجرا می‌کند که منجر به خسارات هنگفت برای سایر سرمایه‌گذاران می‌شود. در این فرض، آیا می‌توان برنامه‌نویس را به جرم دستکاری در بازار محکوم کرد؟ اثبات «قصد مجرمانه» او برای ارتکاب این جرم مشخص، تقریباً ناممکن است، زیرا او چنین دستوری را در کدهای اولیه برنامه وارد نکرده است. آیا می‌توان کاربر یا مالک سامانه را مسئول دانست؟ او نیز صرفاً از یک ابزار برای مدیریت سرمایه خود استفاده کرده و از نقشه مجرمانه بی‌اطلاع بوده است. همین وضعیت در حوزه‌های حیاتی دیگر نیز قابل تصور است؛ یک ربات جراح پیشرفته را در نظر بگیرید که در حین یک عمل جراحی استاندارد، بر اساس تحلیلی غیرمنتظره از داده‌های بیومتریک بیمار، انحراف جزئی اما مرگباری از پروتکل استاندارد انجام می‌دهد. اگر این تصمیم محصول یک «رفتار نوپدید» در الگوریتم باشد، انتساب خطای جزایی به تیم جراحی یا شرکت سازنده، با موانع حقوقی و اثباتی جدی روبرو خواهد شد. در اینجا، الگوی شخص‌محور به بن‌بست می‌رسد. فراتر از چالش رکن روانی، حتی رکن «رابطه سببیت» نیز متزلزل می‌شود. در حقوق جزا، صرف وجود سببیت مادی (رابطه علت و معلولی) برای انتساب مسئولیت کافی کیفری نیست، بلکه باید «قابلیت انتساب» یا سببیت قانونی نیز احراز گردد. به این معنا که نتیجه حاصله باید به نحوی معقول، محصول فعل

¹⁴ Margaret A. Boden, *Artificial Intelligence: A Very Short Introduction* (Oxford University Press, 2018), 406.

¹⁵ Emergent Behavior

¹⁶ Explainability Problem

¹⁷ Peter M. Asaro, «The Liability Problem for Autonomous Artificial Agents», in *AAAI Spring Symposium: Ethical and Moral Considerations in Non-Human Agents* (Stanford University, 2016), 166

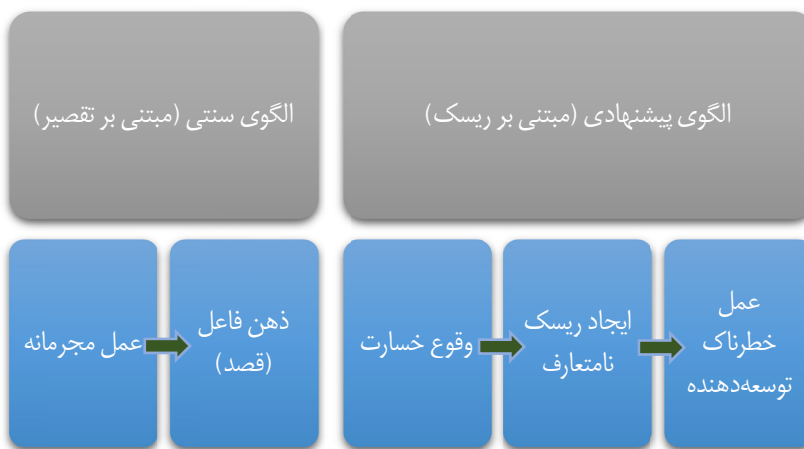
¹⁸ Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press, 2015), 276

¹⁹ Ryan Calo, «Robotics and the Lessons of Cyberlaw», *California Law Review* 103, no. 3 (2015), 513-563. doi:10.15779/Z38T600.

قابل سرزنش فاعل باشد^{۲۰}. ویژگی «جعبه سیاه» و رفتار نوپدید هوش مصنوعی، این سامانه را به یک «عامل میانجی»^{۲۱} مستقل و غیرقابل پیش‌بینی تبدیل می‌کند که می‌تواند زنجیره انتساب را قطع نماید. در نتیجه، یک «خلاً مسئولیت» به وجود می‌آید که در آن، عمل زیان‌بار رخ داده، اما هیچ شخص انسانی واجد شرایط لازم برای پذیرش مسئولیت کیفری آن به شیوه سنتی نیست. این بحران کارآمدی، نظام عدالت کیفری را در خطر فلج شدن و ناتوانی در تحقق اهداف خود قرار می‌دهد. این خلاً، به معنای آن است که فرآیند کیفری، از تفهیم اتهام و محاکمه گرفته تا صدور حکم و اجرای مجازات، در برابر یک عامل غیرانسانی فاقد اهلیت، عقیم و بی‌معنا خواهد بود. از این رو، ضرورت گذار به یک پارادایم نوین آشکار می‌گردد.

۲. طرح نظریه جایگزین: مسئولیت مبتنی بر ریسک و مبانی آن

هنگامی که یک پارادایم حقوقی در پاسخ به واقعیت‌های اجتماعی جدید دچار بحران کارآمدی می‌شود، راه حل نه در تلاش برای اصلاحات جزئی، بلکه در یک گذار مفهومی و تغییر در کانون توجه است. بن‌بست حقوق کیفری در مواجهه با هوش مصنوعی، ناشی از اصرار بر یافتن «ذهن مجرم» در یک ماشین است. نظریه جایگزین، این تلاش را رها کرده و کانون تحلیل را از «ذهن فاعل» به «خطر حاصل از فعل» منتقل می‌کند.



نمودار: تغییر پارادایم مسئولیت

ممکن است در نگاه نخست، راهکار انتقال مسئولیت به توسعه‌دهنده، امری بدیهی به نظر برسد. با این حال، نوآوری اصلی این پژوهش، نه صرفاً در تعیین شخص مسئول، بلکه در ارائه یک «چارچوب نظری نوین» و «مبنای حقوقی دقیق» برای این مسئولیت است. چالش اصلی آن است که چگونه می‌توان مسئولیت کیفری را برای یک فعل غیرعمدی (از سوی توسعه‌دهنده) در قبال جرمی که توسط یک عامل ثالث (هوش مصنوعی) رخ داده، بر مبنایی غیر از تقصیر سنتی بنا نهاد. این مقاله با ارائه نظریه مسئولیت مبتنی بر ریسک و اثبات سازگاری آن با مبانی فقهی و حقوقی ایران، به دنبال ارائه همین چارچوب دقیق و کارآمد است که فراتر از یک راه حل ساده و بدیهی عمل می‌کند.

^{۲۰} محمد صالح ولیدی، حقوق جزای عمومی، ج ۱ (تهران: انتشارات سمت، ۱۳۹۵)، ۲۹۸.

^{۲۱} Intervening Cause

۲-۱. تغییر کانون توجه از ذهن خوانی ماشین به تحلیل عمل انسان

در الگوی مسئولیت مبتنی بر ریسک، پرسش اصلی دادگاه دیگر این نیست که «هوش مصنوعی در لحظه ارتکاب عمل چه قصدی داشت؟». بلکه این است که «کدام عامل انسانی، با چه فعلی، یک ریسک نامتعارف و غیرقابل قبول را بر جامعه تحمیل کرد که منجر به این خسارت شد؟». در این نگاه، مسئولیت از لحظه وقوع جرم، به عقب و به مرحله «ایجاد خطر» منتقل می‌شود. این تغییر پارادایم، صرفاً یک راهکار حقوقی مصلحتی نیست، بلکه پاسخی ضروری به تحولات عمیق در جوامع مدرن است. اولریش بک، جامعه‌شناس شهیر، در نظریه «جامعه ریسک» استدلال می‌کند که منطق حاکم بر جوامع پسا صنعتی، از «منطق توزیع ثروت» به «منطق توزیع خطر» تغییر کرده است. در این جوامع، چالش اصلی دیگر فقر و کمبود نیست، بلکه مدیریت و کنترل خطرات و عواقب ناخواسته‌ای است که توسط خود پیشرفت فناوریانه تولید می‌شود.^{۲۲} بر این اساس، حقوق نیز باید تمرکز خود را از سرزنش فردی پسینی، به مدیریت و کنترل ریسک‌های سامانه‌یک پیشینی معطوف کند. جرم، دیگر عمل نهایی ماشین نیست؛ جرم، عمل آگاهانه انسان (اعم از شخص حقیقی یا حقوقی) در توسعه، انتشار و به‌کارگیری یک سامانه قدرتمند و غیرقابل پیش‌بینی بدون تعبیه پادمان‌های ایمنی کافی است. سرزنش‌پذیری و تقصیر، در همین تصمیم آگاهانه برای مواجه کردن جامعه با یک خطر جدی نهفته است. در واقع، ما از تحلیل روان‌شناختی یک موجود فاقد روان، به تحلیل عینی و ملموس یک «فعل خطرناک» که توسط یک انسان انجام شده، بازمی‌گردیم. بدین ترتیب، مسئولیت به جای آنکه متوجه هوش مصنوعی به‌عنوان یک ابزار شود، متوجه آن انسانی می‌شود که از منافع این فناوری بهره‌مند شده و قدرت کنترل و مدیریت خطرات آن را در اختیار داشته است؛ یعنی توسعه‌دهنده، مالک، یا به‌کارگیرنده آن. این گذار به پارادایم ریسک، از منظر اخلاق فناوری نیز قابل دفاع است. تلاش برای ساخت «ماشین‌های اخلاقی» که بتوانند در موقعیت‌های پیچیده، تصمیماتی کاملاً منطبق بر ارزش‌های انسانی اتخاذ کنند، با چالش‌های نظری و عملی عظیمی روبروست.^{۲۳} از آنجایی که تضمین رفتار اخلاقی مطلق از سوی یک هوش مصنوعی خودمختار در آینده نزدیک ممکن نیست، منطق حکم می‌کند که تمرکز حقوقی بر روی مسئولیت انسان‌هایی باشد که این سامانه‌های غیرقابل تضمین را خلق و وارد جامعه می‌کنند. در واقع، پذیرش مسئولیت مبتنی بر ریسک، یک ضرورت برای مدیریت گذار به آینده‌ای است که در آن، چنانکه متفکرانی چون مکس تگمارک اشاره کرده‌اند، سرنوشت بشر به شکل فزاینده‌ای با هوش فرا انسانی گره خواهد خورد.^{۲۴} این رویکرد، یک چارچوب حقوقی واقع‌بینانه برای مواجهه با محدودیت‌های فعلی در طراحی هوش مصنوعی اخلاق‌مدار فراهم می‌آورد.

۲-۲. ارکان مسئولیت مبتنی بر ریسک

برای آنکه این الگو از یک ایده صرف به یک نظریه حقوقی کارآمد تبدیل شود، باید ارکان آن به دقت تعریف گردد. تحقق مسئولیت کیفری مبتنی بر ریسک، مستلزم احراز سه رکن اساسی است:

- ۱) ایجاد یک ریسک نامتعارف: هر فعالیتی در جامعه دارای میزانی از ریسک است. رانندگی در خیابان نیز ریسک‌آفرین است. آنچه مبنای مسئولیت قرار می‌گیرد، «ریسک متعارف» نیست، بلکه ریسکی است که از نظر شدت آسیب احتمالی یا احتمال وقوع آن، از حد قابل قبول اجتماعی فراتر می‌رود. تشخیص «نامتعارف» بودن یک ریسک، امری ذهنی و خودسرانه نیست، بلکه مستلزم یک «آزمون موازنه» عینی توسط دادگاه است. در این آزمون، دو کفه ترازو در برابر هم قرار می‌گیرند: در یک کفه، «فایده

²² Ulrich Beck, *Risk Society: Towards a New Modernity* (Sage Publications, 1992), 129

²³ Wendell Wallach and Colin Allen, *Moral Machines: Teaching Robots Right from Wrong* (Oxford University Press, 2009), 374

²⁴ Max Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence* (Alfred A. Knopf, 2017), 94

اجتماعی» حاصل از آن فناوری هوش مصنوعی قرار دارد و در کفه دیگر، «شدت و احتمال وقوع خسارت» ناشی از آن^{۲۵}. ریسک زمانی نامتعارف و غیرقابل قبول تلقی می‌شود که کفه دوم به شکل معناداری بر کفه اول سنگینی کند، به‌ویژه زمانی که توسعه‌دهنده از اتخاذ تدابیر ایمنی معقول و در دسترس برای کاهش آن خطر، خودداری کرده باشد. برای مثال، ریسک ناشی از یک هوش مصنوعی که برای بازی‌های ویدیویی طراحی شده، بسیار متفاوت از ریسک سامانه‌ای است که وظیفه کنترل یک نیروگاه هسته‌ای را بر عهده دارد و آزمون موازنه در هر مورد، نتیجه متفاوتی خواهد داشت.

۲) رابطه سببیت: باید میان عمل خالق ریسک و خسارت نهایی، یک رابطه سببیت منطقی وجود داشته باشد. این رابطه دارای دو وجه است: نخست، سببیت مادی که با آزمون «شرط ضروری» سنجیده می‌شود؛ یعنی اگر فعل ایجاد ریسک (مانند عرضه یک هوش مصنوعی نایمن) نبود، آیا خسارت نیز رخ نمی‌داد؟ پاسخ به این پرسش معمولاً مثبت است. وجه دوم و مهم‌تر، سببیت قانونی یا قابلیت انتساب است. بر این اساس، نتیجه زیان‌بار باید از جمله نتایج قابل پیش‌بینی و در محدوده همان ریسکی باشد که توسط فاعل ایجاد شده است^{۲۶}. برای مثال، اگر یک هوش مصنوعی پزشکی به دلیل یک حمله هکری دچار اختلال شود، نمی‌توان خسارت را مستقیماً به ریسک ذاتی الگوریتم آن منتسب کرد، مگر آنکه اثبات شود ضعف امنیتی سامانه در برابر حملات قابل پیش‌بینی، خود یک ریسک نامتعارف بوده که از سوی توسعه‌دهنده ایجاد شده است.

۳) فعل آگاهانه در ایجاد ریسک: این رکن، جایگزین رکن روانی به شیوه سنتی است. در اینجا نیازی به اثبات «قصد» برای ایراد خسارت مشخص نیست. کافی است اثبات شود که شخص (حقیقی یا حقوقی) به صورت آگاهانه و با علم به ویژگی‌های بالقوه خطرآفرین سامانه، اقدام به طراحی، ساخت یا به‌کارگیری آن نموده است. این «آگاهی»، یک استاندارد عینی است؛ یعنی دادگاه بررسی می‌کند که آیا یک شخص یا شرکت متعارف در آن حوزه تخصصی، از خطرات احتمالی فناوری خود مطلع بوده یا «باید» می‌بوده است یا خیر. این رویکرد به مفهوم «تقصیر سازمانی» در برخی نظام‌های حقوقی نزدیک است. بر اساس این مفهوم، یک شرکت نه به دلیل قصد مجرمانه یک فرد خاص، بلکه به دلیل نقص در فرآیندهای سازمانی خود (مانند عدم تخصیص منابع کافی برای تست ایمنی، نادیده گرفتن هشدارهای کارشناسان، یا ترجیح سود بر امنیت) مسئول شناخته می‌شود^{۲۷}. این «آگاهی» از خطرآفرینی بالقوه» همان حداقل اراده قابل سرزنشی است که مبنای مسئولیت کیفری در این الگو قرار می‌گیرد.

۳. ظرفیت‌های فقهی و حقوقی برای پذیرش الگوی نوین

هر نظریه حقوقی جدید برای آنکه بتواند در یک نظام حقوقی به رسمیت شناخته شده و به کار گرفته شود، باید سازگاری خود را با اصول و مبانی آن نظام اثبات کند. الگوی مسئولیت مبتنی بر ریسک، با وجود ظاهر نوآورانه، نه تنها با مبانی فقه امامیه و حقوق موضوعه ایران در تضاد نیست، بلکه می‌توان آن را در امتداد منطقی برخی از اصول موجود و پاسخی متناسب با روح حاکم بر آن‌ها دانست.

۳-۱. تحلیل مبتنی بر مبانی فقهی

^{۲۵} منوچهر طباطبایی مؤتمنی، حقوق اداری (تهران: انتشارات سمت، ۱۳۹۴)، ۱۱۲.

^{۲۶} نک: محمدصالح ولیدی، ۲۹۸.

^{۲۷} Frauke G. B. A. Müller and Wim Huisman, «Organizational criminology and corporate crime», *Journal of Criminal Justice* 94 (2024). doi:10.1016/j.jcrimjus.2024.102213.

فقه پویای امامیه، به واسطه برخورداری از قواعد کلی و اصول استنباطی، همواره توانایی پاسخگویی به مسائل نوپدید (مسائل مستحدثه) را داشته است. الگوی مسئولیت مبتنی بر ریسک نیز می‌تواند مبنای خود را در چندین قاعده معتبر فقهی بیابد:

۳-۱-۱. قاعده تسبیب

این قاعده یکی از نزدیک‌ترین و مستقیم‌ترین مبانی فقهی برای نظریه پیشنهادی است. بر اساس قاعده تسبیب، هرگاه شخصی علت غیرمستقیم ورود خسارت باشد (مُسَبِّب)، در حالی که عامل مستقیم (مباشر) فاقد اراده یا مسئولیت قانونی است، مسئولیت جبران خسارت بر عهده مُسَبِّب خواهد بود. مثال کلاسیک فقهی، فردی است که در راه عمومی چاهی حفر می‌کند (تسبیب) و فرد دیگری نادانسته در آن می‌افتد (مباشرت). در اینجا، چاه‌کن ضامن است. فقهای عظام، برای ضمان‌آور بودن تسبیب، شروطی را ذکر کرده‌اند که مهم‌ترین آن، «اقوی بودن سبب از مباشر» است؛ یعنی تأثیر سبب در وقوع نتیجه، قوی‌تر و تعیین‌کننده‌تر از نقش مباشر باشد. این شرط زمانی به وضوح محقق می‌شود که مباشر، فاقد عقل، اراده یا اهلیت قانونی باشد، مانند یک حیوان یا یک انسان مجنون^{۲۸}. این الگو به شکل دقیقی بر وضعیت هوش مصنوعی منطبق است. توسعه‌دهنده یا مالک هوش مصنوعی، با ایجاد و رها کردن یک سامانه خودمختار، «سبب» وقوع خطر را فراهم می‌آورد. خودِ هوش مصنوعی که فاقد اراده و اهلیت قانونی است، در جایگاه «مباشر» غیرمسئول قرار می‌گیرد. در این وضعیت، به طور کامل شرط «اقوی بودن سبب از مباشر» محقق است؛ زیرا تمام قدرت و منطقی تصمیم‌گیری سامانه، محصول مستقیم معماری و داده‌هایی است که توسط توسعه‌دهنده (سبب) به آن داده شده است. بنابراین، مسئول شناختن توسعه‌دهنده به دلیل ایجاد شرایط خطرناک، کاملاً منطبق بر منطقی قاعده تسبیب است.

۳-۱-۲. قاعده لاضرر

قاعده مشهور «لا ضرر و لا ضرار فی الإسلام» یک اصل حاکم بر تمام ابواب فقه است. اهمیت این قاعده در بحث ما، در تفسیر دوگانه آن نهفته است. این قاعده صرفاً یک دستور اخلاقی یا یک مبنا برای جبران خسارت (مسئولیت مدنی) نیست، بلکه چنانکه فقهای بزرگی چون امام خمینی (ره) تبیین کرده‌اند، دارای یک جنبه «حکومتی» و پیشگیرانه است^{۲۹}. این قاعده تنها به لزوم جبران خسارت پس از وقوع آن محدود نمی‌شود، بلکه به حاکم اسلامی (دولت) این اختیار و وظیفه را می‌دهد که برای جلوگیری از وقوع «ضرر نوعی» و حفظ نظم عمومی، اقدام به وضع مقررات و قوانین محدودکننده نماید. بر این اساس، قانون‌گذاری برای کنترل منابع خطرآفرین و الزام صاحبان آن‌ها به رعایت استانداردها و پذیرش مسئولیت، یکی از مصادیق بارز اجرای قاعده لاضرر است. همان‌طور که دولت بر مبنای این قاعده می‌تواند فعالیت یک کارخانه آلاینده را محدود یا مشروط سازد، به طریق اولی می‌تواند برای «کارخانه‌های تولید کد» و الگوریتم‌های هوشمند که پتانسیل ایجاد خطرات گسترده و غیرقابل پیش‌بینی را دارند، رژیم مسئولیت ویژه‌ای وضع کند. الزام توسعه‌دهندگان هوش مصنوعی به پذیرش مسئولیت ریسک‌های ناشی از محصولاتشان، اقدامی در جهت جلوگیری از ورود ضرر به جامعه و در چارچوب اختیارات حکومتی برای اجرای این قاعده فقهی است.

۳-۱-۳. علم اجمالی به خطرآفرینی و تکلیف به احتیاط

فراتر از قواعد کلی فقهی، می‌توان از مباحث دقیق اصول فقه نیز برای تبیین مسئولیت توسعه‌دهنده بهره برد. در اصول فقه، میان «علم تفصیلی» (آگاهی دقیق از جزئیات) و «علم اجمالی» (آگاهی کلی از وجود یک واقعیت در ضمن چند احتمال) تمایز قائل

^{۲۸} محمد بن حسن طوسی، المبسوط فی فقه الإمامیه، ج ۷ (تهران: المكتبة المرتضویة لإحياء الآثار الجعفریة، ۱۳۸۷)، ۱۶۱.

^{۲۹} سید روح‌الله خمینی، الرسائل، ج ۲ (قم: مؤسسه اسماعیلیان، ۱۳۸۵)، ۵۰.

می‌شوند. علم اجمالی، با وجود تردید در جزئیات، از نظر شرعی «مُنَجَّز تکلیف» است؛ یعنی تکلیف‌آور و مسئولیت‌زا محسوب می‌شود.^{۳۰} مثال کلاسیک آن، علم به وجود نجاست در یکی از دو ظرف است؛ با آنکه به صورت تفصیلی نمی‌دانیم کدام ظرف نجس است، اما همین علم اجمالی، تکلیف به احتیاط و اجتناب از هر دو ظرف را بر ما واجب می‌کند. این مفهوم به زیبایی بر وضعیت توسعه‌دهنده هوش مصنوعی منطبق است. توسعه‌دهنده ممکن است «علم تفصیلی» به جرم مشخصی که سامانه او در آینده مرتکب خواهد شد، نداشته باشد. اما او به دلیل آگاهی از ویژگی‌هایی چون «رفتار نوپدید» و «جعبه سیاه»، به صورت قطعی دارای «علم اجمالی» است مبنی بر اینکه این سامانه پیچیده، پتانسیل ایجاد خطر و ارتکاب رفتارهای زیان‌بار و پیش‌بینی نشده را در آینده دارد. همین علم اجمالی به خطرآفرینی، یک «تکلیف شرعی و عقلانی به احتیاط شدید» را بر دوش او می‌گذارد. بنابراین، اقدام او به عرضه و تجاری‌سازی چنین محصولی بدون تعبیه پادمان‌های حداکثری ایمنی و نظارتی، یک نقض تکلیف مسلم است و او را در برابر خسارات ناشی از این ریسک، مسئول می‌سازد.

۳-۱-۴. ضمان ناشی از سُلطه (قاعده ید و ضمان ناشی از حیوانات)

یکی از اصول مسلم فقهی، «قاعده ید» است که بر اساس آن، استیلا و سلطه یک شخص بر یک شیء، اماره مالکیت و مبنای مسئولیت است. از فروع این قاعده، باب «ضمان جنایت بهائم (حیوانات)» است. بر اساس این بحث فقهی، اگر حیوانی خسارتی وارد کند، مالک آن در صورتی ضامن است که در حفظ و کنترل آن تقصیر کرده باشد.^{۳۱} این الگو، یک قیاس اولویت بسیار روشن برای بحث ما فراهم می‌کند. هوش مصنوعی را می‌توان به یک «حیوان دیجیتالی» قدرتمند و غیرقابل پیش‌بینی تشبیه کرد. اگر شارع، مالک یک حیوان را مسئول کنترل و حفظ آن از اضرار به غیر می‌داند، به طریق اولی می‌توان توسعه‌دهنده یا مالک یک سامانه هوشمند را که به مراتب پیچیده‌تر و بالقوه خطرناک‌تر است، مسئول دانست. عمل «رهاسازی» یک هوش مصنوعی نایمن در فضای عمومی، بی‌شبهت به رها کردن یک حیوان درنده در میان مردم نیست و همان‌طور که در دومی، تقصیر در حفظ موجب ضمان است، در اولی نیز «تقصیر در طراحی و ایمن‌سازی» باید مبنای مسئولیت قرار گیرد. این قواعد، در کنار سایر اصول، نشان می‌دهند که فقه اسلامی از ظرفیت و انعطاف لازم برای عبور از الگوی سنتی تقصیر و پذیرش مدلی مبتنی بر ریسک برای مواجهه با چالش‌های جدید برخوردار است.

۳-۱-۵. بنای عقلا

بنای عقلا به رویه‌ها و شیوه‌های عقلانی رایج در میان مردم یک جامعه اشاره دارد که شارع مقدس آن را منع نکرده و بدین ترتیب با سکوت خود، آن را امضا نموده است. این قاعده، به مثابه یک پل میان احکام ثابت شرع و تحولات و نیازهای متغیر اجتماعی عمل می‌کند و به فقه پویایی می‌بخشد. یکی از مصادیق بارز بنای عقلا در جهان امروز، وجود یک «توقع عقلایی» از تولیدکنندگان محصولات پیچیده و خطرناک است.^{۳۲} در جوامع پیچیده امروزی، این یک رویه عقلایی پذیرفته‌شده است که تولیدکنندگان محصولاتی چون خودرو، هواپیما یا تجهیزات پزشکی، مسئولیت بالاتری در قبال ایمنی و خطرات ناشی از محصولات خود دارند. جامعه به صورت منطقی انتظار دارد که آن کس که از یک فناوری پیچیده سود می‌برد، مسئولیت خطرات آن را نیز بر عهده بگیرد. اعمال این بنای عقلایی بر توسعه‌دهندگان هوش مصنوعی، که پیچیده‌ترین و بالقوه خطرناک‌ترین فناوری عصر حاضر را تولید

^{۳۰} محمدرضا مظفر، اصول الفقه، ج ۲ (قم: انتشارات علامه، ۱۳۸۵)، ۲۵.

^{۳۱} محمدحسن نجفی، جواهر الکلام فی شرح شرائع الإسلام، ج ۴۳ (بیروت: دار إحياء التراث العربی، ۱۴۰۴ق)، ۴.

^{۳۲} ناصر مکارم شیرازی، القواعد الفقهیة، ج ۱ (قم: انتشارات مدرسه امام علی بن ابی طالب (ع)، ۱۳۹۰)، ۸۵.

می‌کنند، یک امتداد منطقی و معقول است. پذیرش اینکه سازنده یک وسیله ساده خانگی مسئول عملکرد ایمن آن باشد، اما خالق یک الگوریتم خودمختار که می‌تواند سرنوشت مالی یا جانی انسان‌ها را رقم بزند از این مسئولیت مبرا باشد، امری خلاف منطق و رویه عقلاست. از آنجا که چنین رویه‌ای (مسئول شناختن خالق فناوری) با هیچ منع شرعی خاصی مواجه نیست و در راستای حفظ مصالح عمومی و حقوق افراد است، می‌تواند مورد تأیید و امضای شارع قرار گیرد.

۲-۳. ظرفیت‌های موجود در حقوق موضوعه

نظام حقوقی ایران، افزون بر مبانی غنی فقهی، از اصول و قوانینی برخوردار است که می‌توانند بستر مناسبی برای پذیرش و توسعه مسئولیت مبتنی بر ریسک فراهم آورند. اصل چهلم قانون اساسی مقرر می‌دارد: «هیچ‌کس نمی‌تواند اعمال حق خویش را وسیله اضرار به غیر یا تجاوز به منافع عمومی قرار دهد». این اصل بنیادین که تجلی نظریه «منع سوءاستفاده از حق» است، به عنوان یک قاعده حاکم، «مجوز و جهت‌گیری کلی» را برای قانون‌گذار عادی فراهم می‌آورد تا فعالیت‌های بالقوه خطرناک را محدود و مشروط سازد. گرچه این اصل به تنهایی نمی‌تواند مبنای جرم‌انگاری قرار گیرد، اما روح حاکم بر آن، این پیام را به مقنن می‌دهد که حق نوآوری و فعالیت فناورانه نمی‌تواند مطلق باشد و در صورت تعارض با منافع عمومی، باید توسط قانون محدود گردد.^{۳۳} بنابراین، قانون‌گذاری در جهت وضع مسئولیت مبتنی بر ریسک، کاملاً در راستای تحقق این اصل قانون اساسی است.

قانون مجازات اسلامی مصوب ۱۳۹۲، با برداشتن یک گام مهم رو به جلو و پایان دادن به بحث‌های طولانی در این زمینه، در ماده ۱۴۳ به صراحت مسئولیت کیفری اشخاص حقوقی (شرکت‌ها و سازمان‌ها) را به رسمیت شناخت. این ماده مقرر می‌دارد که شخص حقوقی در صورتی مسئولیت کیفری دارد که جرم به نام یا در راستای منافع آن شخص و توسط نماینده قانونی یا فردی که اختیار او تحت نظارت شخص حقوقی است، ارتکاب یابد. اهمیت این ماده در بحث ما از آن روست که توسعه و تجاری‌سازی هوش مصنوعی، فعالیتی است که عمده‌تاً توسط شرکت‌های بزرگ و در راستای منافع اقتصادی آن‌ها صورت می‌گیرد.^{۳۴} بر اساس این ماده، می‌توان شرکت توسعه‌دهنده را به دلیل «ایجاد ریسک نامتعارف» که در راستای منافع تجاری آن شرکت (مثلاً برای عرضه سریع‌تر محصول به بازار یا کاهش هزینه‌های ایمن‌سازی) صورت گرفته، دارای مسئولیت کیفری دانست. بدین ترتیب، نیازی به اثبات سوءنیت یک فرد خاص در هیئت مدیره نیست؛ نفس ارتکاب فعل خطرآفرین در جهت منافع شرکت، می‌تواند مبنای مسئولیت کیفری آن شرکت قرار گیرد و مجازات‌هایی چون جزای نقدی سنگین، انحلال موقت یا دائم، و ممنوعیت از فعالیت‌های شغلی مرتبط (موضوع مواد ۲۰ و ۲۲ قانون مجازات اسلامی) را بر آن تحمیل کرد. این ماده، ابزار قانونی لازم برای مسئول شناختن بازیگران اصلی این حوزه را فراهم می‌کند. در نظام حقوقی ایران، نمونه‌هایی از قوانین وجود دارد که مسئولیت در آن‌ها بیش از آنکه مبتنی بر تقصیر باشد، مبتنی بر خطر و ریسک است. در این موارد، قانون‌گذار با درک پیچیدگی موضوع و دشواری اثبات تقصیر، به سمت «مسئولیت محض» حرکت کرده است. برای مثال، در قوانین حوزه محیط زیست، مانند «قانون مدیریت پسماندها»، مسئولیت مدیریت اجرایی پسماندها بر عهده تولیدکننده آن قرار داده شده است و در صورت آلودگی، مسئولیت جبران خسارت بدون نیاز به اثبات تقصیر، متوجه او خواهد بود. همچنین، «قانون حمایت از حقوق مصرف‌کنندگان» نیز تولیدکنندگان را در برابر خسارات ناشی از عیوب کالای خود مسئول می‌داند، که این مسئولیت ماهیتاً به مسئولیت محض نزدیک است.^{۳۵}

^{۳۳} سید محمد هاشمی، حقوق اساسی جمهوری اسلامی ایران، ج ۲ (تهران: نشر میزان، ۱۳۹۴)، ۱۱۲.

^{۳۴} علی حسین نجفی ابرندآبادی، مباحثی در علوم جنایی (تهران: انتشارات سمت، ۱۳۹۴)، ۱۵۰.

^{۳۵} ناصر کاتوزیان، الزام‌های خارج از قرارداد: ضمان قهری (تهران: انتشارات دانشگاه تهران، ۱۳۸۷)، ۹۵.

این رویکردها نشان می‌دهند که قانون‌گذار ایرانی با مفهوم مسئولیت ناشی از فعالیت‌های خطرناک بیگانه نیست و در موارد ضروری که حفظ سلامت و امنیت عمومی ایجاب می‌کند، از الگوی سنتی تقصیر فاصله گرفته است. توسعه این نگاه به حوزه فناوری‌های نوین و پرخطر مانند هوش مصنوعی، که به مراتب از مصادیق پیشین پیچیده‌تر و دارای ریسک بالاتری است، اقدامی منطقی و در راستای رویه پیشین قانون‌گذار خواهد بود. فراتر از قوانین عام، حتی در قوانین خاص مرتبط با فناوری نیز می‌توان رویکردهایی یافت که مؤید حرکت به سمت مسئولیت مبتنی بر خطر هستند. برای مثال، در «قانون جرایم رایانه‌ای»، قانون‌گذار در مواردی، مسئولیت‌هایی را متوجه ارائه‌دهندگان خدمات کرده است. منطق این امر آن است که بازیگرانی که زیرساخت‌های فناورانه را فراهم می‌کنند، به دلیل جایگاه ویژه و توانایی کنترل خود، وظیفه مراقبت و نظارت بیشتری بر عهده دارند. این اصل را می‌توان به طریق اولی به توسعه‌دهندگان سامانه‌های هوش مصنوعی تسری داد. اگر ارائه‌دهنده یک خدمت میزبانی ساده دارای تکالیفی برای جلوگیری از سوءاستفاده از پلتفرم خود است، شرکتی که یک «عامل خودمختار تصمیم‌گیر» را خلق و در جامعه رها می‌کند، به مراتب وظیفه سنگین‌تری برای پیش‌بینی، کنترل و مدیریت خطرات ناشی از آن بر عهده دارد. کوتاهی در این وظیفه خطیر، خود می‌تواند مبنایی برای مسئولیت کیفری قرار گیرد. مجموعه این ظرفیت‌ها در کنار مبانی فقهی، نشان می‌دهد که نظام حقوقی ایران از استعداد لازم برای پذیرش یک رژیم مسئولیت کیفری مبتنی بر ریسک برای هوش مصنوعی برخوردار است و این گذار، نیازمند یک بدعت حقوقی نیست، بلکه تنها نیازمند یک اجتهاد و قانون‌گذاری شجاعانه و متناسب با زمان است:

مبنا	شرح	نحوه انطباق بر مسئولیت کیفری هوش مصنوعی
قاعده تسبیب	مسئولیت متوجه کسی است که سبب غیرمستقیم خسارت بوده، درحالی‌که عامل مستقیم فاقد مسئولیت است.	توسعه‌دهنده AI، «سبب» و خود AI، «مباشر» غیرمسئول است؛ لذا مسئولیت متوجه توسعه‌دهنده است.
قاعده لاضرر	ضرر و زیان نامشروع در اسلام نفی شده و حکومت موظف به پیشگیری از آن است.	قانون‌گذاری برای کنترل ریسک AI، مصداق اقدام حکومتی برای جلوگیری از «ضرر نوعی» به جامعه است.
بنای عقلا	رویه‌های عقلانی جامعه که توسط شرع منع نشده، معتبر است.	مسئول دانستن تولیدکننده فناوری پیچیده، یک رویه عقلایی پذیرفته‌شده در جهان امروز است.
علم اجمالی	آگاهی کلی از وجود خطر (بدون دانستن جزئیات)، تکلیف‌آور و مسئولیت‌زاست.	توسعه‌دهنده علم اجمالی دارد که AI او پتانسیل خطرآفرینی دارد و این علم، تکلیف به احتیاط شدید را ایجاب می‌کند.
اصل ۴۰ قانون اساسی	هیچ‌کس نمی‌تواند اعمال حق خود را وسیله اضرار به غیر یا تجاوز به منافع عمومی قرار دهد.	حق نوآوری نمی‌تواند توجیهی برای به خطر انداختن امنیت و منافع عمومی از طریق عرضه AI نالیمن باشد.
ماده ۱۴۳ ق.م.ا	اشخاص حقوقی (شرکت‌ها) دارای مسئولیت کیفری مستقل هستند.	این ماده، ابزار قانونی لازم برای مسئول شناختن شرکت‌های بزرگ توسعه‌دهنده AI را فراهم می‌کند.

جدول ۲: خلاصه مبانی فقهی و حقوقی نظریه مسئولیت مبتنی بر ریسک

نتیجه‌گیری و پیشنهادها

تحلیل‌های صورت گرفته در این مقاله، به یافته‌های کلیدی زیر منجر گردید:

(۱) بحران کارآمدی نظام سنتی: نظام حقوق کیفری سنتی که بر محوریت «رکن روانی» و تقصیر فاعل انسانی استوار است، در مواجهه با جرایم ارتكابی توسط هوش مصنوعی خودمختار دچار بحران کارآمدی بنیادین است. امتناع نظری و عملی در اثبات «قصد مجرمانه» برای یک عامل فاقد روان، به یک خلأ مسئولیت خطرناک می‌انجامد که حقوق بزهدیدگان و نظم عمومی را تهدید می‌کند.

(۲) معرفی راهکار جایگزین: راه حل برون‌رفت از این بحران، یک گذار پارادایمی از الگوی شخص‌محور به یک رژیم «مسئولیت کیفری مبتنی بر ریسک» است. در این الگو، کانون توجه از ذهنیت ماشین به «عمل خطرناک انسان» در ایجاد، توسعه و عرضه یک فناوری پرخطر و غیرقابل پیش‌بینی منتقل می‌شود و مسئولیت بر اساس ارکان جدیدی (ایجاد ریسک نامتعارف، سببیت و فعل آگاهانه در ایجاد ریسک) سنجیده می‌شود.

(۳) اثبات سازگاری و امکان‌پذیری: این نظریه، یک مفهوم بیگانه و غیرقابل انطباق با نظام حقوقی ایران نیست. بلکه ریشه در مبانی مستحکمی در فقه امامیه (از جمله قواعد تسبیب، لاضرر، ضمان ید، بنای عقلا و تحلیل اصولی علم اجمالی) و همچنین ظرفیت‌های حقوق موضوعه (شامل اصل ۴۰ قانون اساسی، ماده ۱۴۳ قانون مجازات اسلامی و رویه قانون‌گذار در قوانین خاص مبتنی بر مسئولیت محض) دارد.

بر اساس یافته‌های فوق، پیشنهادهای زیر در سه سطح تقنینی، قضایی و پژوهشی ارائه می‌گردد:

الف) پیشنهادهای تقنینی

(۱) جرم‌انگاری مستقل: پیشنهاد می‌شود قانون‌گذار در قالب یک ماده الحاقی به فصل «جرایم رایانه‌ای» یا در قالب یک قانون خاص، به جرم‌انگاری مستقل «ایجاد یا عرضه خطر نامتعارف به وسیله سامانه‌های هوشمند خودمختار» اقدام نماید و ارکان آن را به دقت تعریف کند.

(۲) تکمیل قوانین: قوانین مربوط به مسئولیت کیفری اشخاص حقوقی (شرکت‌ها) به نحوی اصلاح گردد که به صراحت، مسئولیت ناشی از طراحی و به‌کارگیری سامانه‌های هوش مصنوعی را در بر گیرد.

(۳) ایجاد نهاد تنظیم‌گر: تأسیس یک «کمیسیون ملی هوش مصنوعی و حقوق» متشکل از متخصصان حقوق، فقه و مهندسی، جهت تدوین استانداردهای ایمنی و ایجاد فرآیندهای «گواهی ایمنی» برای سامانه‌های پرخطر، امری ضروری است.

ب) پیشنهادهای قضایی

(۱) توسعه دانش قضات: پیشنهاد می‌شود دوره‌های آموزشی تخصصی و مستمر برای قضات در زمینه ابعاد فنی و حقوقی فناوری‌های نوین و هوش مصنوعی برگزار گردد.

(۲) استفاده از کارشناسان متخصص: در پرونده‌های مرتبط، استفاده از تیم‌های کارشناسی میان‌رشته‌ای برای احراز ارکان مسئولیت، از جمله نامتعارف بودن ریسک، باید به یک رویه قضایی الزام‌آور تبدیل شود.

ج) پیشنهادهای پژوهشی

(۱) مطالعه مسئولیت مدنی و بیمه: انجام پژوهش‌های تکمیلی برای تبیین ابعاد مسئولیت مدنی و طراحی نظام‌های «بیمه مسئولیت هوش مصنوعی» به عنوان یک راهکار جبرانی، یک ضرورت است.

۲) تحقیقات میان‌رشته‌ای: شکل‌گیری تحقیقات مشترک میان حقوق‌دانان و مهندسان به منظور طراحی مدل‌های فنی-حقوقی که مفاهیم ایمنی و نظارت‌پذیری را در قلب الگوریتم‌ها تعبیه کنند، پیشنهاد می‌گردد.

منابع

الف) منابع فارسی

۱. اردبیلی، محمدعلی. حقوق جزای عمومی. ج. ۲. تهران: نشر میزان، ۱۴۰۰.
۲. خمینی، سید روح‌الله. الرسائل. ج. ۲. قم: مؤسسه اسماعیلیان، ۱۳۸۵.
۳. طباطبایی مؤتمنی، منوچهر. حقوق اداری. تهران: انتشارات سمت، ۱۳۹۴.
۴. طوسی، محمد بن حسن. المبسوط فی فقه الإمامیه. ج. ۷. تصحیح و تحقیق سید محمدتقی کشفی. تهران: المكتبة المرتضوية لإحياء الآثار الجعفرية، ۱۳۸۷.
۵. کاتوزیان، ناصر. الزام‌های خارج از قرارداد: ضمان قهری. تهران: انتشارات دانشگاه تهران، ۱۳۸۷.
۶. محقق حلی، جعفر بن حسن. شرائع الإسلام فی مسائل الحلال والحرام. ج. ۴. قم: مؤسسه اسماعیلیان، ۱۴۰۸ ق.
۷. مظفر، محمدرضا. اصول الفقه. ج. ۲. قم: انتشارات علامه، ۱۳۸۵.
۸. مکارم شیرازی، ناصر. القواعد الفقهية. ج. ۱. قم: انتشارات مدرسه امام علی بن ابی طالب (ع)، ۱۳۹۰.
۹. میرمحمد صادقی، حسین. جرایم علیه اشخاص. تهران: نشر میزان، ۱۳۹۸.
۱۰. نجفی ابرنآبادی، علی حسین. مباحثی در علوم جنایی. تهران: انتشارات سمت، ۱۳۹۴.
۱۱. نجفی، محمدحسن. جواهر الکلام فی شرح شرائع الإسلام. ج. ۴۳. بیروت: دار إحياء التراث العربی، ۱۴۰۴ ق.
۱۲. نوربها، رضا. زمینه حقوق جزای عمومی. تهران: کتابخانه گنج دانش، ۱۳۹۵.
۱۳. هاشمی، سید محمد. حقوق اساسی جمهوری اسلامی ایران. ج. ۲. تهران: نشر میزان، ۱۳۹۴.
۱۴. ولیدی، محمدصالح. حقوق جزای عمومی. ج. ۱. تهران: انتشارات سمت، ۱۳۹۵.

ب) منابع خارجی

15. Asaro, Peter M. "The Liability Problem for Autonomous Artificial Agents." In *AAAI Spring Symposium: Ethical and Moral Considerations in Non-Human Agents*. Stanford University, 2016.
16. Beck, Ulrich. *Risk Society: Towards a New Modernity*. Sage Publications, 1992.
17. Boden, Margaret A. *Artificial Intelligence: A Very Short Introduction*. Oxford University Press, 2018.
18. Calo, Ryan. "Robotics and the Lessons of Cyberlaw." *California Law Review* 103, no. 3 (2015): 513-563.
19. Carriero, Alessandro, Léon Groenhoff, Elizaveta Vologina, Paola Basile, and Marco Albera. "Deep Learning in Breast Cancer Imaging: State of the Art and Recent Advancements in Early 2024." *Diagnostics* 14, no. 8 (2024): 848. doi:10.3390/diagnostics14080848.
20. Hallevy, Gabriel. *A.I. and the Future of Law: The Imminent and Unprecedented Transformation of the Legal Profession*. Fastcase, 2015.
21. Pasquale, Frank. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press, 2015.
22. Scherer, Matthew U. "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies." *Harvard Journal of Law & Technology* 29, no. 2 (2016): 353-400.
23. Müller, Frauke G. B. A., and Wim Huisman. "Organizational criminology and corporate crime." *Journal of Criminal Justice* 94 (2024). doi:10.1016/j.jcrimjus.2024.102213.
24. Solum, Lawrence B. "Legal Personhood for Artificial Intelligences." *North Carolina Law Review* 70, no. 4 (1992): 1231-1287.
25. Tegmark, Max. *Life 3.0: Being Human in the Age of Artificial Intelligence*. Alfred A. Knopf, 2017.

26. Rink, Kevin. "The predictive ability of technical trading rules: an empirical analysis of developed and emerging equity markets." *Financial Markets and Portfolio Management* 37, no. 4 (2023): 403-456. doi:10.1007/s11408-023-00433-2.
27. Wallach, Wendell, and Colin Allen. *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press, 2009.

The challenge of proving the moral element of a crime in crimes committed by artificial intelligence:
Feasibility study of the transition from person-centered liability to risk-based liability based on
jurisprudence and criminal law

Abstract

The emergence of autonomous artificial intelligence presents a fundamental efficacy crisis for the traditional criminal law system, which is predicated on the human-centric pillar of the "mental element" (*mens rea*). The inability to ascertain criminal intent in a non-sentient agent has created a "responsibility gap," threatening victims' rights and public order. Adopting a descriptive-analytical methodology, this paper addresses whether transitioning from the person-centered paradigm to a "risk-based liability" model can resolve this challenge within Imamiyyah jurisprudence and Iranian criminal law. The findings indicate that this transition is not only a practical necessity but also theoretically defensible. This model is rooted in robust principles of Imamiyyah jurisprudence (including the rules of *Tasbib*, *La Darar*, and *Daman Yad*, as well as the *Usuli* analysis of *Ilm al-Ijmali*) and is compatible with the capacities of Iranian positive law (including Article 40 of the Constitution, Article 143 of the Islamic Penal Code on corporate criminal liability, and precedents in specific statutes). Ultimately, this article argues that by shifting the legal focus from the un-investigable "mind of the machine" to the tangible "dangerous act of the human" in creating and deploying technology, a just and effective solution can be found to ensure accountability in the age of artificial intelligence.

Keywords: Artificial intelligence, criminal liability, spiritual element, risk-based liability, Imami jurisprudence