مسئولیت کیفری هوش جامع مصنوعی به مثابه شخصیت حقوقی نوین: تحلیل پیامدهای کیفری اعطای شخصیت حقوقی به سیستم های فوق هوشمند

محمود حبیبی تبار [1]

چکیده

ظهور سیستم‌های هوش جامع مصنوعی با قابلیت تصمیم‌گیری مستقل، کارایی ساختارهای سنتی مسئولیت کیفری را که بر انتساب عمل به عامل انسانی استوارند، با بحران مواجه کرده است. هدف این پژوهش، تبیین ضرورت گذار از پارادایم سنتی مسئولیت کیفری انسان‌محور و ارائه‌ی مدلی نوین برای شناسایی «شخصیت حقوقی الکترونیکی کیفری» است تا خلأ ناشی از عملکرد مستقل سیستم‌های فوق هوشمند در نظام عدالت کیفری مرتفع گردد. موضع مقاله حاضر، دفاع از اعطای این شخصیت به سیستم‌هاست، نه شخصیت مدنیِ صرف؛ چراکه تنها از این طریق می‌توان خلأ ناشی از سه ویژگی بنیادین هوش جامع مصنوعی، یعنی «خودمختاری در انتخاب وسیله»، «کدورت الگوریتمی» و «فقدان سوءنیت انسانی» را پر کرد. پژوهش حاضر با رویکرد تحلیلی-توصیفی و تطبیقی، از مسئولیت کیفری اشخاص حقوقی به عنوان شاهد و مبنای نظری برای این گذار پارادایمی بهره می‌برد: همان‌گونه که شرکت می‌تواند موضوع تحریم کیفری واقع شود، سیستم فوق هوشمند نیز می‌تواند واجد اهلیت تحمل کیفر گردد، اما با ماهیت و ضوابطی متفاوت. استدلال محوری مقاله آن است که عنصر روانی جرم در این بستر، نه به عنوان قصد درونی، بلکه در قالب «تقصیر الگوریتمی مفروض» بازتعریف می‌شود که از شکست سیستم در پایبندی به استانداردهای ایمنی مصوب نهاد ناظر قانونی احراز می‌گردد. یافته‌ها تأیید می‌کند که این مدل، با تکمیل از طریق «صندوق ضمانت اجباری» برای جبران خسارت بزه‌دیدگان، نه یک انتخاب نظری، بلکه ضرورتی عملی برای حفظ کارآمدی نظام عدالت کیفری در عصر سیستم‌های فوق هوشمند است.

واژگان کلیدی: هوش جامع مصنوعی ، مسئولیت کیفری، شخصیت حقوقی، سیستم هوشمند، مدل حقوقی

[1]. گروه حقوق، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران، ایران .

habibitabarmahmoud@iau.ac.ir

Criminal Liability of Artificial General Intelligence (AGI) as a Novel Legal Personality: An Analysis of the Criminal Consequences of Granting Legal Personality to Superintelligent Systems

Abstract

The emergence of Artificial General Intelligence (AGI) systems with autonomous decision-making capacity has plunged the efficacy of traditional criminal responsibility frameworks predicated upon the attribution of acts to a human agent into crisis. The aim of this research is to elucidate the necessity of a paradigm shift from the conventional human-centric model of criminal responsibility and to propose a novel framework for recognizing "Electronic Criminal Legal Personality," thereby addressing the vacuum created by the independent functioning of super-intelligent systems within the criminal justice system. This article advocates for granting such systems "electronic criminal legal personality," as opposed to mere civil personality, since only through this approach can the void arising from three fundamental characteristics of AGI namely, "autonomy in the selection of means," "algorithmic opacity," and "the absence of human mens rea" be filled. Employing an analytical-descriptive and comparative approach, the present study draws upon the criminal liability of legal persons as both a precedent and a theoretical foundation for this paradigmatic transition: just as a corporation can be the subject of criminal sanctions, so too can a super-intelligent system possess the capacity to bear punishment, albeit with a distinct nature and criteria. The paper's central argument is that the mental element of the crime, in this context, is redefined not as internal intent but as a form of "presumed algorithmic fault," established upon the system's failure to adhere to safety standards mandated by a statutory regulatory body. The findings confirm that this model, complemented by a "mandatory guarantee fund" for victim compensation, constitutes not a mere theoretical choice but a practical imperative for preserving the efficacy of the criminal justice system in the age of super-intelligent systems.

Keywords: Artificial General Intelligence, Criminal Liability, Legal Personality, Intelligent Systems.Law Model

مقدمه

ظهور سامانه‌های هوش مصنوعی، به‌ویژه در بحث‌های مربوط به امکان تحقق هوش جامع مصنوعی، نقطه عطفی در تاریخ فناوری و چالش‌های حقوق کیفری محسوب می‌شود. در این پژوهش، هوش جامع مصنوعی به‌عنوان یک فرض آینده‌پژوهانه در نظر گرفته می‌شود؛ موجودیتی که در صورت تحقق قادر خواهد بود درکی فراگیر، یادگیری مستمر و به‌کارگیری دانش در سطحی هم‌سنگ یا فراتر از توانایی‌های شناختی انسان داشته باشد. این فرض ناظر به وضعیت بالفعل فناوری نیست، بلکه چارچوبی نظری برای تحلیل پیامدهای حقوقی احتمالی است. از این رو، تمایز میان هوش جامع مصنوعی فرضی، سیستم‌های خودمختار پیشرفته موجود و هوش مصنوعی متعارف ضروری است، زیرا هر سطح از فناوری پیامدهای متفاوتی برای انتساب مسئولیت کیفری دارد.

در نظام‌های کیفری کنونی، مسئولیت‌پذیری بر دو رکن اساسی استوار است: رکن مادی (وقوع فعل مجرمانه) و رکن معنوی (قصد یا تقصیر). در سناریوهایی که یک سامانه هوش مصنوعی پیشرفته رفتاری زیان‌بار یا مجرمانه بروز می‌دهد، انتساب این دو رکن به انسان با دشواری مواجه می‌شود. از منظر رکن مادی، پرسش اصلی این است که آیا رفتار تولیدشده توسط یک سامانه خودمختار را می‌توان «فعل» یک انسان تلقی کرد؟ در مدل‌های سنتی، فعل مجرمانه همواره به یک فاعل انسانی نسبت داده می‌شود، اما در سامانه‌های یادگیرنده، رفتار نهایی محصول تعامل پیچیده داده‌ها، محیط و فرآیندهای خودتکاملی است و نه صرفاً اجرای دستورات برنامه‌نویس. بنابراین، تعیین اینکه «چه کسی» فعل را انجام داده است، خود به یک چالش نظری تبدیل می‌شود.

در خصوص رکن معنوی نیز چالش عمیق‌تری وجود دارد. قصد مجرمانه، علم، آگاهی و قابلیت پیش‌بینی، همگی مفاهیمی انسان‌محور هستند. سامانه‌های هوش مصنوعی در پیشرفته‌ترین حالت فاقد نیت و تجربه ذهنی‌اند. بنابراین، اصولاً نمی‌توان برای آن‌ها سوءنیت عام یا خاص قائل شد. از سوی دیگر، انتساب قصد یا تقصیر به انسان نیز دشوار است، زیرا رفتار نهایی سامانه ممکن است از کنترل و پیش‌بینی طراح یا مالک خارج شده باشد. این وضعیت، به معنای «عدم امکان انتساب کیفری» نیست، اما نشان می‌دهد که دشواری اثبات تقصیر نباید با فقدان امکان انتساب خلط شود. مسئله اصلی، ناکارآمدی ابزارهای سنتی اثبات در مواجهه با رفتارهای غیرخطی سامانه‌های خودمختار است. مسئولیت کارفرما نسبت به کارمند یا مسئولیت تولیدکننده نسبت به نقص محصول، در مواجهه با سامانه‌هایی که سطحی از خودمختاری دارند، کارایی محدودی پیدا می‌کنند، اما این امر لزوماً به «خلأ کامل مسئولیت» نمی‌انجامد. آنچه رخ می‌دهد، شکل‌گیری نوعی شکاف مسئولیت است؛ وضعیتی که در آن رفتار زیان‌بار توسط سامانه‌ای انجام می‌شود که تصمیم‌گیری آن از کنترل مستقیم انسان فاصله گرفته و ابزارهای اثباتی موجود قادر به اتصال رفتار به یک فاعل انسانی مشخص نیستند. این شکاف باید از «نقص ادله» یا «ناکامی سیاست کیفری» متمایز شود، زیرا ریشه آن در عدم انطباق ساختاری میان الگوهای مسئولیت کیفری و ماهیت تصمیم‌گیری شبکه‌ای و خودتکاملی سامانه‌های هوش جامع مصنوعی است.

از منظر سیاست‌گذاری کیفری، پیش‌بینی پیامدهای حقوقی ظهور احتمالی هوش جامع مصنوعی ضرورتی عملی است. نبود چارچوبی روشن می‌تواند از یک سو به عدم جبران خسارت قربانیان و از سوی دیگر به تحمیل مسئولیت‌های نامتناسب بر توسعه‌دهندگان منجر شود. در چنین بستری، ضرورت بررسی راهکارهای بدیل برای پر کردن این خلاء آشکار می‌شود.یکی از این راهکارها، با الهام از تجربه موفق حقوقی در اعطای شخصیت حقوقی به نهادهای انتزاعی همچون شرکت‌های تجاری، بررسی امکان ایجاد نوعی شخصیت حقوقی نوین برای سامانه‌های فوق‌خودمختار است. اعطای چنین شخصیتی نه به معنای هم‌ارز دانستن هوش مصنوعی با انسان، بلکه به‌عنوان یک تمهید حقوقی برای قرار دادن یک موجودیت پاسخگو در مرکز شبکه مسئولیت مطرح می‌شود. پرسش اصلی این است که آیا چنین شخصیت حقوقی می‌تواند مبنایی برای مسئولیت کیفری مستقل باشد و در این صورت، مفاهیم بنیادین رکن مادی و رکن معنوی چگونه باید برای یک موجودیت غیرانسانی بازتعریف شوند؟ در این چارچوب، اعطای «شخصیت حقوقی الکترونیکی کیفری» به سامانه‌هایی که سطحی از استقلال تصمیم‌گیری واقعی دارند، نه به‌عنوان یک ضرورت قطعی، بلکه به‌عنوان یکی از گزینه‌های قابل بررسی برای پر کردن شکاف‌های احتمالی مسئولیت کیفری مطرح می‌شود. این رویکرد تنها زمانی موجه است که به‌طور دقیق روشن شود کدام سطح از فناوری موضوع بحث است و فرض تحقق هوش جامع مصنوعی در چه جایگاهی از تحلیل قرار دارد. در نهایت، مسئله انتساب رکن مادی و معنوی در جرایم مرتبط با هوش جامع مصنوعی نشان می‌دهد که چالش اصلی نه در فقدان قواعد، بلکه در ناسازگاری مفهومی میان حقوق کیفری انسان‌محور و فناوری‌های غیرانسانی خودمختار است؛ ناسازگاری‌ای که بدون بازاندیشی در مبانی مسئولیت، قابل رفع نخواهد بود.

۱-مبانی نظری و چارچوب‌های مفهومی

پیش از ورود به بحث چالش‌برانگیز انتساب مسئولیت کیفری به هوش مصنوعی، ضروری است که چارچوب مفهومی این پژوهش بر پایه‌های نظری مستحکم بنا شود. این چارچوب بر دو مفهوم اساسی استوار است: نخست، تبیین دقیق هوش جامع مصنوعی نه به عنوان صرفاً مجموعه‌ای از کدها، بلکه به عنوان یک عامل نوظهور دارای سطحی از خودمختاری که او را از سامانه‌های هوش مصنوعی محدود متمایز می‌سازد؛ خودگردانی که آستانه‌ای برای بررسی قابلیت انتساب جرم محسوب می‌شود. دوم، تحلیل مفهوم شخصیت حقوقی در حقوق معاصر. شخصیت حقوقی، نهادی اعتباری و برساخته نظام حقوقی است که قانون‌گذار برای اعطای توانایی‌ها و تعهدات به نهادهای غیرانسانی (مانند شرکت‌ها و سازمان‌ها) وضع کرده است. این امر می‌تواند به عنوان الگویی کارآمد برای تعریف یک «شخصیت حقوقی الکترونیکی کیفری» در حوزه حقوق کیفری مورد استفاده قرار گیرد تا بدین‌وسیله، قابلیت پاسخگویی کیفری به خود سامانه‌های فوق‌هوشمند منتسب گردد.

۱-۱-تعریف و تمایز هوش جامع مصنوعی از هوش مصنوعی محدود و آستانه مسئولیت کیفری

نقطه تمایز اساسی در مباحث مسئولیت‌پذیری، گذار از هوش مصنوعی محدود به هوش جامع مصنوعی است. سامانه‌های هوش مصنوعی محدود در دامنه‌های مشخص و از پیش تعریف‌شده‌ای عمل می‌کنند و فاقد قابلیت‌های شناختی عمومی انسان هستند. در مقابل، هوش جامع مصنوعی به سیستمی اطلاق می‌شود که قادر به انجام هرگونه تکلیف شناختی در سطحی برابر یا فراتر از انسان، شامل توانایی یادگیری، درک و استدلال در محیط‌های متنوع و حل مسائل کلی است. (Goertzel, B., & Pennachin, C. (Eds.). 2007.11-15) آستانه ورود به بحث مسئولیت کیفری مستقل، در مفهوم «خودمختاری کامل» نهفته است؛ به معنای توانایی سیستم برای تصمیم‌گیری و اقدام بر اساس اهداف خود، بدون نیاز به مداخله لحظه‌ای از سوی عامل انسانی. هنگامی که هوش جامع مصنوعی به این سطح از خودمختاری می‌رسد، فرآیند علیت و انتساب سوءنیت دچار گسست اساسی می‌شود؛ زیرا ممکن است عمل مجرمانه نتیجه یک تصمیم مستقل و غیرقابل پیش‌بینی توسط خود سیستم باشد، نه صرفاً اجرای خطایی در کدنویسی اولیه پژوهش‌ها در این حوزه به شدت بر محدودیت‌های مفهومی خودمختاری کامل تأکید دارند؛ (Hallevy, G. 2015.45-50) از آن جمله که اعطای شخصیت حقوقی و مسئولیت کیفری به هوش جامع مصنوعی، در صورت مطالبه «مسئولیت اخلاقی کامل»، زودرس تلقی می‌شود. دلیل این امر آن است که این سامانه‌ها فاقد «هدف معنادار و خودآگاهی» هستند که پیش‌شرط‌های اساسی برای قضاوت و اراده آزاد در نظام‌های حقوقی فعلی محسوب می‌شوند. (Pagallo, U. 2013.88-92) عنصر روانی جرم، که برای مجازات به معنای سنتی ضروری است، مستلزم آن است که عامل دارای ویژگی‌های شناختی مانند آگاهی، نیت و اراده آزاد باشد تا بتوان عمل مجرمانه را به تصمیمات درونی او منتسب کرد. (Chopra, S., & White, L. F. 2011.62-65). تا زمانی که هوش جامع مصنوعی این خصوصیات ذهنی را کسب نکند، نمی‌توان آن را از نظر اخلاقی تقصیرکار و سزاوار سرزنش دانست.

با این حال، تحقق کامل این سطح از مسئولیت اخلاقی، شرط ضروری برای هرگونه پاسخ کیفری به رفتارهای زیان‌بار هوش جامع مصنوعی نیست. رسالت حقوق کیفری، منحصر به تحمیل سرزنش اخلاقی نبوده، بلکه حفظ نظم عمومی و حمایت از بزه‌دیدگان را نیز در بر می‌گیرد. بر این اساس، ضرورت عملی پر کردن خلأ حقوقی ناشی از استقلال عملی سیستم‌های فوق هوشمند، به بازتعریف مبانی مسئولیت از «مسئولیت کیفری اخلاقی» به «مسئولیت شبه‌کیفری مبتنی بر ضرورت اجتماعی» می‌انجامد. در این چارچوب، که از دکترین مسئولیت کیفری اشخاص حقوقی الهام گرفته شده است(Wells, C.2001.120-125). شخصیت حقوقی الکترونیکی کیفری نه برای تحمیل سرزنش اخلاقی بر ماشین، بلکه به عنوان یک ابزار حقوقی برای مهار رفتارهای پرخطر و تأمین خسارت بزه‌دیدگان طراحی می‌شود. عنصر روانی جرم در این سطح، نه به‌عنوان قصد درونی اخلاقی، بلکه در قالب «تقصیر الگوریتمی مفروض» بازتعریف می‌گردد که از شکست سیستم در پایبندی به استانداردهای ایمنی مصوب نهاد رگولاتور[۲] احراز می‌شود. بنابراین، خودمختاری کامل در این بستر، نه به عنوان عملی که از یک اراده

---

[2] (Regulator)و نهاد متناظر آن (Regulatory Body/Regulatory Authority)به سازمانی دولتی، مستقل یا فرابخشی اطلاق میگردد که حسب تفویض قانونگذار، دارای صلاحیت انحصاری در سه حوزه است: (۱) تدوین و تصویب استانداردهای فنی، ایمنی و اخلاقی لازمالاجرا برای سیستمهای هوش مصنوعی، (۲) نظارت مستمر بر عملکرد این سیستمها از حیث میزان پایبندی به استانداردهای مصوب، و (۳) اعمال

آگاهانه نشأت گرفته، بلکه به عنوان خروجی مستقل سیستمی که از کنترل انسان خارج شده، مبنای حقوقی لازم برای اعمال این نوع خاص از مسئولیت شبه‌کیفری را فراهم می‌آورد.

در تبیین دقیق‌تر مفهوم «تقصیر الگوریتمی مفروض»، باید توجه داشت که این تقصیر نه یک مسئولیت مطلق، بلکه مبتنی بر اماره‌ی فنی است .(European Commission .2022.496)در خصوص بار اثبات، در این مدل، بارِ اثباتِ «فقدان تقصیر» به طور منطقی بر عهده‌ی نماینده‌ی قانونی یا مالکِ سامانه است تا در صورت بروز حادثه، ادله‌ای مبنی بر غیرقابل پیش‌بینی بودنِ رخداد ارائه دهد. این سامانه در مقام دفاع می‌تواند از طریقِ بازخوانیِ «گزارش‌های رفتاری» و تحلیلِ مسیرِ تصمیم‌گیریِ الگوریتمی توسط هیئت‌های قضاییِ تخصصی، از خود دفاع کند..(Pagallo, U., et al. 2021.105).در نهایت، نسبت میان مسئولیت شخصیت حقوقی الکترونیکی کیفری و مسئولیت عوامل انسانی، مبتنی بر الگوی «مسئولیت سلسله‌مراتبی و مکمل» است. در این مدل، مسئولیت این سامانه به عنوان «مسئولیت اصلی» جهت تأمین خسارتِ بی‌واسطه و اعمال مجازات‌های اصلاحی-فنی عمل می‌کند، در حالی که مسئولیت عوامل انسانی(تولیدکنندگان و مالکان) به عنوان «مسئولیت ثانویه» تنها در صورت اثباتِ قصورِ اثباتی در فرآیندهای طراحی یا نظارت فعال می‌شود. این سازوکارِ دو لایه، ضمن جلوگیری از فرارِ عوامل انسانی از پاسخگویی، از توسلِ شتاب‌زده به مسئولیت مطلق کیفری که با مبانی عدالت فردی در تعارض است، جلوگیری به عمل می‌آورد .(Lagioia, F., & Sartor, G.2020.433-465)

۱-۲-مفهوم شخصیت حقوقی در حقوق معاصر

شخصیت حقوقی مفهومی است که پیشینه آن به مبانی فکری متعددی بازمی‌گردد و شناخت آن به عنوان موجودی برخوردار از حقوق و تکالیف، در یک سیر تاریخی و با تکیه بر بنیادهای اندیشه‌ای گوناگون شکل گرفته است (سلیمان زاده ،۱۴۰۴،81).

در حقوق معاصر، شخصیت حقوقی یک واقعیت هستی‌شناختی یا زیست‌شناختی نیست، بلکه یک برساخته مصنوعی و ابزاری قانونی است که نظام حقوقی برای نیل به اهداف مشخص اجتماعی و اقتصادی خلق می‌کند. این مفهوم، ظرفیتی را به نهادهای غیرانسانی اعطا می‌کند تا بتوانند در روابط حقوقی حضور یابند، مالکیت کسب کنند، قرارداد منعقد سازند و قابلیت اقامه دعوا و نیز مورد دعوا قرار گرفتن را داشته باشند (Kurki, V. A. J. 2019.128). کارکرد اولیه این نهاد در شرکت‌های تجاری، ایجاد موجودیتی مجزا از سهامداران و مدیران، جهت تضمین دوام و بقای نهاد و نیز فراهم آوردن مسئولیت محدود برای تشویق سرمایه‌گذاری خصوصی است .(Kraakman, R., Armour, J., Davies, P., Enriques, L., Hansmann, H.

ضمانت اجراهای اداری، مدنی یا کیفری در صورت تخلّف از مقررات وضع شده. در چارچوب نظریه ی «تقصیر الگوریتمی مفروض» که در این مقاله طرح گردیده، احراز مسئولیت شبه کیفری سیستم هوش جامع مصنوعی، منوط به احراز شکست سیستم در پایبندی به استانداردهای ایمنی مصوب همین نهاد رگولاتور است. برای مطالعه ی تفصیلی مفهوم و کارکرد نهادهای رگولاتور در حوزه ی فناوری، ر.ک:

Baldwin, Robert, Martin Cave, and Martin Lodge. Understanding Regulation: Theory, Strategy, and Practice. 2nd ed. Oxford: Oxford University Press, 2012, pp. 15-40.

.Hertig, G., ... & Rock, E. (2017.8) با این حال، پیشینه اعطای شخصیت حقوقی به نهادهای غیرانسانی، قدمتی فراتر از شرکت‌های مدرن دارد و دقیقاً در نقطه‌ای پدیدار می‌شود که یک «خلأ در زنجیره مسئولیت» شکل می‌گیرد.

مثال کلاسیک این خلأ، رویه دعوای «این رم»[۳] در حقوق دریایی است[۴]. در این حوزه، هنگامی که یک کشتی تحت کنترل خدمه(که ممکن است خود فاقد دارایی کافی برای جبران خسارت باشند) موجب ایراد خسارت می‌شد، نظام حقوقی با یک معضل عملی مواجه می‌گشت: عامل انسانیِ مستقیم اغلب غیرقابل تعقیب یا فاقد توان مالی برای جبران خسارت بود. راه‌حل حقوقی که ریشه در قرون وسطی دارد، آن بود که مسئولیت به‌طور مستقیم علیه خود کشتی اقامه شود و کشتی به‌عنوان یک نهاد دارای شخصیت (یا شبه‌شخصیت) در نظر گرفته شود که قادر به «خطا کردن» و پرداخت خسارت از محل ارزش خود است (Robertson, D. W. 2012.18-22) .کشتی بدل به «محور مسئولیت‌پذیری» شد، نه به دلیل آنکه واجد آگاهی یا سوءنیت بود، بلکه به این دلیل که این انتساب، تنها راه عملی برای پر کردن شکاف میان عمل زیان‌بار و جبران خسارت بود. این رویه به‌خوبی نشان می‌دهد که نظام حقوقی، شخصیت را نه بر اساس ویژگی‌های ذاتی یک موجودیت، بلکه بر اساس ضرورت کارکردیِ ایجاد یک کانون پاسخگویی اعطا می‌کند(Pietrzykowski, T. 2018.82) .

الگوی مشابهی را می‌توان در اعطای شخصیت حقوقی به اشخاص حقوقی عمومی، موقوفات و اخیراً برخی عناصر طبیعی مانند رودخانه‌ها(همچون رودخانه وانگانویی در نیوزیلند) مشاهده کرد. در تمامی این موارد، شخصیت حقوقی پاسخی به یک نیاز عملی برای سازمان‌دهی مسئولیت‌ها و حمایت از حقوق بوده است، نه تأییدی بر وجود آگاهی یا اراده انسانی در آن نهادها (O'Donnell, E., & Talbot-Jones, J. 2018.4-6) بنابراین، تاریخ حقوق به روشنی گواهی می‌دهد که شخصیت حقوقی یک مفهوم انعطاف‌پذیر و کارکردگراست که بر اساس نیازهای حقوقی–اجتماعی، و نه لزوماً وجود آگاهی یا اراده، اعطا می‌شود.

گذار به هوش جامع مصنوعی دقیقاً بر مبنای همین منطق کارکردی توجیه می‌شود. بحرانی که سیستم‌های هوش جامع مصنوعی خودمختار ایجاد می‌کنند، از نظر ساختاری مشابه بحران مسئولیت در حقوق دریایی است: یک عامل انسانی(تولیدکننده، برنامه‌نویس یا کاربر) وجود دارد که عمل نهایی، به دلیل استقلال عملی سیستم، دیگر به‌طور کامل و عادلانه به او منتسب نیست. همان‌طور که در مورد کشتی، خدمه انسانی دیگر تنها کانون مسئولیت نبودند، در مورد هوش جامع مصنوعی نیز عامل انسانیِ پشتِ سیستم، دیگر پاسخگوی تمام‌عیار اعمال غیرقابل پیش‌بینی سیستم نیست ,Chesterman) (S. 2021.83 این شکاف در انتساب، یعنی جایی که «خطا» رخ می‌دهد اما «خطاکار انسانیِ پاسخگو» وجود ندارد، همان ضرورت کارکردی‌ای است که در طول تاریخ، قانون‌گذار را به ابداع یک شخص حقوقی جدید واداشته است. استدلال این

---

[۳]. این رم ، یک اصطلاح حقوقی است که به معنای علیه خود چیز یا علیه خود مال است. در حقوق دریایی، دعوای این رم به دعوایی گفته می شود که نه علیه شخص مالک یا متصرف یک مال (مانند کشتی)، بلکه مستقیما علیه خود مال (مثلا خود کشتی) اقامه می شود. هدف از این نوع دعوا، معمولا توقیف یا فروش آن مال برای پرداخت بدهی یا جبران خسارت وارده است. این رویه به کشتی این امکان را می دارد که به عنوان یک شخص حقوقی فرضی عمل کرده و مسئولیت حسارات را بر عهده بگیرد.

[۴]. منظور از قانون دریایی در این متن، مجموعه ای از قوانین و مقررات است که به امور مربوط به ناوبری، تجارت دریایی، و مسئولیت های ناشی از فعالیت های دریایی می پردازد. در اینجا، این قانون به عنوان یک بستر تاریخی معرفی شده که در آن نظام حقوقی برای حل مسائل عملی ، مانند تعیین مسئولیت خسارات در دریا، شخصیت حقوقی را به نهادهایی مانند کشتی ها اعطا کرده است.

نیست که چون شرکت یا کشتی شخصیت حقوقی دارد، هوش جامع مصنوعی نیز باید داشته باشد(که به‌درستی مغالطه تمثیل ناقص خواهد بود). استدلال این است که شرایط ضرورت کارکردی تکرار شده است: یک عامل ایجادکننده خسارت با استقلال عملی بالا وجود دارد که نمی‌توان مسئولیت آن را به‌طور کامل به انسان پشت صحنه منتسب کرد. بنابراین، ابزار آزموده‌شده «شخصیت حقوقی» برای ایجاد یک کانون نوین مسئولیت‌پذیری، پاسخی متناسب با این بحران تکرارشونده است.

۲-ارکان مسئولیت کیفری و چالش‌های هوش جامع مصنوعی

از آنجا که تحقق مسئولیت کیفری مستلزم اراده آگاهانه، سوءنیت یا خطا و قابلیت انتساب میان رفتار و فاعل است، ارتکاب صرفِ تخلف به‌خودی‌خود برای مجازات کفایت نمی‌کند. این اصل سنتی در مواجهه با سیستم‌های هوش مصنوعی با چالش‌های ساختاری مواجه می‌شود.(کاوه ، ۱۴۰۳، ۵۲).

نظام حقوقی کیفری برای اعمال مجازات، بر دو رکن اساسی عمل مجرمانه و سوءنیت بنا نهاده شده است که هر دو در مواجهه با هوش جامع مصنوعی به چالش کشیده می‌شوند. در خصوص عمل مجرمانه، چالش اصلی این است که چگونه می‌توان عملی را که توسط یک سامانه کاملاً خودمختار و الگوریتمی انجام شده، به عنوان یک «فعل» در معنای حقوقی و ارادی در نظر گرفت (Hallevy, G. 2015.53). از آنجایی که اقدامات هوش جامع مصنوعی(مانند یک تصمیم ناگهانی یا یک تصادف) در واقع ریشه در ورودی‌های برنامه‌نویس یا اپراتور انسانی دارند، تعیین علیت در زنجیره تصمیم‌گیری هوش جامع مصنوعی به شدت پیچیده می‌شود؛ محاکم ناچارند زنجیره علیت را نه در اراده متهم، بلکه در آموزش‌های الگوریتمی و دخالت‌های انسانی جستجو کنند .(Abbott, R., & Sarch, A. 2019.323-384) برخی از پژوهشگران برای رفع این ابهام در سامانه‌هایی که قابلیت اصلاح کد خود را دارند، پیشنهاد می‌کنند که هوش جامع مصنوعی می‌تواند از خالق خود جدا شود و به واسطه توانایی‌اش در تصمیم‌گیری مستقل، مسئولیت فعلی که انجام می‌دهد را بر عهده بگیرد. .Chopra, S., & White, L) .

.(F. 2011با این وجود، بسیاری همچنان به فقدان بدن فیزیکی و توانایی‌های بیولوژیکی برای انجام «فعل» در معنای سنتی اشاره می‌کنند و هوش جامع مصنوعی را ابزاری در دست انسان می‌دانند.(Pagallo, U. 2013.48)

در خصوص عنصر ذهنی جرم یا همان سوءنیت، بحث در مورد سامانه‌های هوش جامع مصنوعی به مراتب عمیق‌تر است؛ چراکه این سامانه‌ها، با وجود سطح بالای پیچیدگی فنی و عملکردی، فاقد آگاهی یا توانایی انتخاب اخلاقی هستند و در هسته خود، صرفاً یک بهینه‌ساز آماری باقی می‌مانند.(Chesterman, S. 2021) در غیاب این نیت انسانی قوانین کیفری سنتی در مواردی که هوش جامع مصنوعی به صورت خودمختار باعث آسیب می‌شود، دچار خلأ قانونی خواهند شد ,.Lagioia, F) .

(Sartor, G. 2020.435 & برای پرکردن این خلأ، مفاهیم جایگزین متعددی مطرح شده‌اند. یکی از این مفاهیم، جایگزین کردن نیت با مفهوم گسترده‌تر «قصد تقصیر» است که مبتنی بر وجود یک عنصر اخلاقی در رفتار هدفمند سیستم است تا مبنای لازم برای مجازات، حتی در غیاب آگاهی فراهم شود.(Hallevy, G. 2015) مفهوم دیگر، «نیت شبیه‌سازی شده» است

که چالشی‌تر بوده و بر این ایده متمرکز است که سیستم حقوقی، تصمیمات مستقل هوش جامع مصنوعی را به دلیل توانایی آن در انتخاب بین راه‌حل‌های جایگزین و تطبیق رفتار با هنجارهای از پیش تعریف‌شده، به‌طور عملی معادل یک نیت حقوقی قلمداد کند(Abbott, R., & Sarch, A. 2019.323-384) این رویکرد، در واقع بازنمایی عملکرد هوش جامع مصنوعی را به‌جای واقعیت نیت در نظر می‌گیرد. با این حال، استفاده از هر یک از این مفاهیم جایگزین، مستلزم آن است که هوش جامع مصنوعی ابتدا به‌عنوان یک شخص حقوقی به رسمیت شناخته شود تا بتواند به‌طور مستقیم حقوق، وظایف و به تبع آن، مسئولیت کیفری را بپذیرد(Kurki, V. A. J. 2019).

رویارویی نظام عدالت کیفری با هوش جامع مصنوعی، پرسش‌هایی بنیادین را پیرامون اصل قانونی بودن جرم و مجازات و ارکان سه‌گانه جرم‌انگاری(قانونی بودن، مادی بودن، و روانی بودن) مطرح می‌سازد. در بُعد رکن مادی(فعل)، ناتوانی هوش جامع مصنوعی در ارائه یک «اراده مستقل» به مثابه «فعل خارجی و ارادی» در معنای سنتی، ما را به سمت فرضیه علیت مبتنی بر خطا در زنجیره آموزش سوق می‌دهد (Pagallo, U. 2013). این وضعیت، مستلزم بازتعریف مفهوم «عمل» از یک فعل فیزیکی-ارادی به یک قصد کنش‌گرانه است که نیازمند تفکیک دقیق میان «علت محقق» و «علت منتسب» در فرایند تصمیم‌گیری خودکار است (Lagioia, F., & Sartor, G.2020.433-465) در این راستا، جستجوی مسئولیت در ورودی‌های اولیه (کدهای الگوریتمی)، مسئولیت کیفری را از ماهیت «جرم» به «تقصیر در مدیریت ریسک تکنولوژیک» تقلیل می‌دهد، که این امر با ماهیت بازدارندگی و اصلاحی مجازات کیفری در تعارض است .

در خصوص رکن روانی(سوءنیت/قصد مجرمانه)، چالش از سطح اثبات به سطح تعریف ماهیت ارتقا می‌یابد. فقدان آگاهی و اختیار اخلاقی در هوش جامع مصنوعی، مفهوم سنتی «قصد عام» و «قصد خاص» را نامعتبر می‌سازد. مفاهیم بدیل نظیر «قصد تقصیر» یا «نیت شبیه‌سازی شده» در واقع تلاشی برای تحقق مسئولیت کیفری عینی هستند؛ یعنی انتساب پیامد به سیستم صرفاً بر اساس انطباق عمل با هنجارها، فارغ از وجود عنصر روانی لازم. بنابراین اتکای صرف بر این مفاهیم، در صورت عدم اعطای شخصیت حقوقی، منجر به تعارض با اصل مسئولیت فردی می‌شود، زیرا مجازات بدون تقصیر ذهنی، ماهیت تنبیهی صرف یافته و از جنبه اصلاحی تهی می‌گردد .

نهایتاً، راهکار اساسی برای عبور از این تنگنای تقنینی، بحث شخصیت حقوقی هوش جامع مصنوعی است. اعطای این وضعیت به هوش جامع مصنوعی، امکان پذیرش مستقیم الزامات کیفری را فراهم می‌آورد و از این طریق، مسئولیت را به‌طور کارآمدتری از زنجیره انسانی دور می‌سازد. با این حال، پذیرش چنین امری نیازمند بازنگری در مفاهیم سنتی «قابلیت انتساب» و «اهلیت تحمل مجازات» است. در غیاب این تعریف ساختاری، نظام حقوقی ناچار است به سمت نوعی مسئولیت تضامنی یا نیابتی سوق یابد که در آن، مسئولیت کیفری بر دوش خالق یا مالک(به‌عنوان فاعل غیرمستقیم) سنگینی خواهد کرد، امری که خود چالش‌هایی در حوزه اثبات تقصیر مستقیم پیش می‌آورد.

۱-۲-چالش‌های انتساب مسئولیت کیفری به هوش جامع مصنوعی(بدون شخصیت حقوقی)

پذیرش مسئولیت کیفری برای سامانه‌های هوش مصنوعی و اعطای شخصیت الکترونیکی به آن‌ها، در کنار مزایای احتمالی، چالش‌های بنیادینی به همراه دارد که نیازمند مطالعات دقیق حقوقی و رفع ابهامات اجرایی است؛ چرا که مواجهه با این مسئله نباید جنبه رفع تکلیف داشته باشد. برای نمونه، چالش اصلی در این مسیر، ناکارآمدی مجازات‌های سنتی مانند حبس یا اعدام برای موجودات فاقد جسم، ناتوانی در اجرای جزای نقدی به دلیل بی‌ملکی و فقدان دارایی مستقل سیستم‌ها، تعیین مصداق جرایم قابل انتساب (همچون ارتکاب قتل عمد)، و همچنین خطر سوءاستفاده انسان‌ها از پوشش مسئولیت کیفری یا بیمه‌های جبران خسارت هوش مصنوعی است (امیریان فارسانی ،،۱۴۰۲41).

فقدان عنصر روانی و آگاهی در سامانه‌های هوش جامع مصنوعی، اعطای شخصیت حقوقی و در نتیجه مسئولیت کیفری مستقل به این نهادها را حداقل در حال حاضر ناممکن می‌سازد. این امر، بحث را ناگزیر به سوی تمرکز بر عوامل انسانی و سازمانی هدایت می‌کند، یعنی باید مسئولیت را به طراحان، توسعه‌دهندگان، مالکان یا اپراتورهای هوش جامع مصنوعی منتسب کرد. با این حال، چالش اصلی اینجاست که حتی با بازگرداندن مسئولیت به انسان، درجه بالای خودمختاری عملی و عدم شفافیت هوش جامع مصنوعی در فرآیندهای تصمیم‌گیری، مفاهیم حقوقی سنتی مانند زنجیره علیت و تقصیر را دچار تزلزل می‌کند.(Matthias, A. 2004.177)

در این میان، استقلال عملی هوش جامع مصنوعی وضعیتی را پدید می‌آورد با عنوان «شکاف مسئولیت». مقصود از این اصطلاح، صرفاً دشواری اثبات تقصیر یا نقص ادله نیست، بلکه اشاره به وضعیتی ساختاری دارد که در آن رفتار زیان‌بار توسط سامانه‌ای انجام می‌شود که فرایند تصمیم‌گیری آن از کنترل مستقیم انسان فاصله گرفته و الگوهای سنتی انتساب که بر علیت خطی، پیش‌بینی‌پذیری و امکان ردیابی تصمیم انسانی استوارند توان اتصال رفتار سامانه به یک فاعل انسانی مشخص را از دست می‌دهند.(Danaher, J. 2016.299-309) در چنین شرایطی، هیچ‌یک از مدل‌های کلاسیک مسئولیت، اعم از مسئولیت مبتنی بر تقصیر، مسئولیت کارفرما یا مسئولیت تولیدکننده، به‌طور کامل پاسخ‌گو نیستند؛ زیرا این مدل‌ها بر پیش‌فرض‌هایی بنا شده‌اند که با ماهیت شبکه‌ای، خودتکاملی و غیرشفاف سامانه‌های هوش جامع مصنوعی سازگار نیست. همان‌گونه که تأکید می‌شود، ظهور سامانه‌های خودمختار پیشرفته موجب تضعیف مفروضات بنیادین حقوق کیفری درباره کنترل، پیش‌بینی‌پذیری و قابلیت انتساب می‌شود و این امر ضرورت بازاندیشی در مبانی مسئولیت را برجسته می‌سازد .(Chesterman, S. 2021)

بر این اساس، «شکاف مسئولیت» باید از مفاهیمی مانند «نقص ادله» یا «ناکامی سیاست کیفری» متمایز شود. نقص ادله ناظر به دشواری در گردآوری یا ارائه شواهد است، در حالی که شکاف مسئولیت ناشی از عدم انطباق ساختاری میان ابزارهای اثباتی حقوق کیفری و ماهیت تصمیم‌گیری سامانه‌های فوق‌خودمختار است؛ وضعیتی که حتی با وجود ادله کافی نیز امکان انتساب مسئولیت به یک شخص حقیقی را مخدوش می‌سازد. به همین دلیل، این شکاف نه یک مشکل اجرایی، بلکه یک چالش مفهومی و نهادی است که ریشه در نابسندگی چارچوب‌های موجود برای مواجهه با فناوری‌های نوظهور دارد. .Vladeck, D) C. 2014.145).

در همین راستا، چالش محوری در مدل مسئولیت طراح، اثبات سوءنیت یا تقصیر اثباتی در ارتباط میان تصمیم اولیه کدنویسی و اقدام نهایی سیستم است. این امر نیازمند تدوین استانداردهای فنی-حقوقی جدید برای سنجش «تقصیر در پیش‌بینی» در طراحی الگوریتم‌های یادگیری تقویتی است، جایی که طراح صرفاً مجموعه قوانین اولیه را فراهم کرده و عملکرد ثانویه از دل داده‌ها نشأت گرفته است.

در مدل مسئولیت مالک/اپراتور، توجیه کیفری بر پایه نظارت کافی استوار است. اما با افزایش پیچیدگی هوش جامع مصنوعی پیشرفته، اثبات اینکه مالک نتوانسته است به‌صورت منطقی وقوع فعل مجرمانه را پیش‌بینی یا از آن جلوگیری کند، دشوارتر می‌شود. این امر، سیستم حقوقی را به سمت پذیرش مسئولیت مطلق در حوزه آسیب‌های ناشی از هوش جامع مصنوعی سوق می‌دهد؛ مدلی که در آن صرف وقوع ضرر توسط سیستم، بدون نیاز به اثبات تقصیر فردی، موجب مسئولیت می‌شود. با این حال، کاربرد مسئولیت مطلق در حوزه کیفری که غالباً ماهیت بازدارندگی دارد با مبانی عدالت فردی در تعارض است و می‌تواند منجر به محدودسازی بیش از حد فعالیت‌های نوآورانه گردد.

در خصوص نسبت میان مسئولیت شخصیت حقوقی الکترونیکی هوش جامع مصنوعی و مسئولیت عوامل انسانی، باید بر الگوی "مسئولیت سلسله‌مراتبی و مکمل" تأکید ورزید. در این مدل، مسئولیت هوش جامع مصنوعی به عنوان "مسئولیت اصلی" شناخته می‌شود که هدف آن تأمین خسارتِ بی‌واسطه و اعمال مجازات‌های اصلاحی-فنی(مانند بازآموزی الگوریتمی) برای مهار ریسکِ ناشی از خودمختاریِ سیستم است. در مقابل، مسئولیت عوامل انسانی(تولیدکنندگان و مالکان) به عنوان "مسئولیت ثانویه" تعریف می‌گردد که تنها در صورت اثباتِ قصورِ اثباتی در فرآیندهای طراحی یا نظارت، فعال می‌شود. در واقع، این دو مسئولیت نه در تقابل، بلکه در عرض یکدیگر عمل می‌کنند؛ به نحوی که شخصیت حقوقی هوش جامع مصنوعی خلاء ناشی از "گسست سببی" را پر می‌کند و مسئولیت انسانی، "تعهداتِ پیشینی" مبنی بر رعایت استانداردهایِ ایمنی را تضمین می‌نماید. بدین ترتیب، این سازوکارِ دو لایه، ضمن جلوگیری از فرارِ عوامل انسانی از پاسخگویی، از توسلِ شتاب‌زده به مسئولیت مطلق کیفری که با مبانی عدالت فردی در تعارض است، جلوگیری به عمل می‌آورد.

۱-۲-۱-تئوری‌های مسئولیت مبتنی بر تولیدکننده/مالک/کاربر و محدودیت‌های آن‌ها

اگرچه ممکن است برخی به‌دلیل ماهیت انسانی ارکان مسئولیت کیفری، پذیرش آن را برای هوش مصنوعی ناممکن بدانند، اما توسعه دامنه مسئولیت اشخاص حقوقی نشان می‌دهد که سیستم حقوقی ظرفیت انطباق با پدیده‌های نوین را دارد. با این وجود، در سناریوهای خاصی مانند نقض حق نسخه برداری، تعیین مسئولیت میان توسعه‌دهنده، ارائه‌دهنده یا کاربر نهایی به نحوه طراحی الگوریتم وابسته است؛ به‌طوری‌که گاه رفتار سامانه مستقل از اراده کاربر و صرفاً ناشی از منطق کدهای اولیه است.(ابوذری ، ۱۴۰۳، ۴۳)

تلاش‌های اولیه برای پر کردن «شکاف مسئولیت» ناشی از خودمختاری هوش جامع مصنوعی، بر انتساب مسئولیت به عوامل انسانی از طریق تئوری‌های موجود مانند مسئولیت ناشی از تقصیر یا مسئولیت مطلق متمرکز است. در این مدل‌ها، تولیدکننده به دلیل طراحی ضعیف یا عدم رعایت استانداردهای ایمنی، مالک به دلیل عدم نظارت کافی یا انتخاب محیط پرخطر، و کاربر به دلیل استفاده نادرست، مسئول شناخته می‌شود. (Kingston, J. K. C. 2020.312) با این حال، تلاش‌های اولیه برای پرکردن «شکاف مسئولیت» ناشی از خودمختاری هوش جامع مصنوعی، بر انتساب مسئولیت به عوامل انسانی از طریق تئوری‌های موجود مانند مسئولیت ناشی از تقصیر یا مسئولیت مطلق متمرکز است. بنابراین کارایی این تئوری‌ها در سناریوهای یادگیری عمیق و تغییرات خارج از کنترل اولیه به شدت کاهش می‌یابد. زمانی که یک سامانه هوش جامع مصنوعی با استفاده از شبکه‌های عصبی عمیق، در محیط‌های عملیاتی و بدون نظارت لحظه‌ای انسان به تکامل و یادگیری ادامه می‌دهد، تغییرات خارج از کنترل اولیه توسعه‌دهندگان رخ می‌دهد؛ این امر باعث می‌شود که عمل مجرمانه در زمان وقوع، دیگر به تقصیر یا نیت اولیه طراحان یا مالکان قابل انتساب نباشد. (Matthias, A.2004.183) این فرآیند تکامل درون سیستم، منجر به «مشکل جعبه سیاه» می‌شود، به این معنا که حتی توسعه‌دهندگان اصلی نیز نمی‌توانند به طور دقیق پیش‌بینی کنند که هوش جامع مصنوعی در موقعیتی خاص چه تصمیمی خواهد گرفت یا چرا آن تصمیم منجر به آسیب شده است .(Bathaee, Y. 2018.916) این عدم شفافیت و عدم قابلیت پیش‌بینی، پیوند لازم برای اثبات رابطه علیت و قابلیت پیش‌بینی را در تئوری تقصیر از بین می‌برد؛ چراکه تقصیر مستلزم آن است که فرد مسئول، می‌توانسته است عمل زیان‌بار سیستم را پیش‌بینی کرده و از آن جلوگیری نماید.(Danaher, J. 2016.302)

بنابراین، مدل مسئولیت مبتنی بر انسان، در سناریوهایی که هوش جامع مصنوعی به‌طور مستقل و خارج از پارامترهای طراحی‌شده عمل می‌کند، به دلیل پیچیدگی علیت و ناپدید شدن عنصر تقصیر، ناکارآمد می‌شود و یک «خلأ مسئولیت» دائمی در حقوق کیفری ایجاد می‌کند.

در همین راستا تئوری‌های مسئولیت سنتی که بر تولیدکننده، مالک یا کاربر متمرکز هستند، در مواجهه با خودمختاری شناختی هوش جامع مصنوعی، از لحاظ توجیه کیفری دچار چالش‌های جدی می‌شوند. در حقوق کیفری، مجازات نیازمند توجیهی مبتنی بر عقوبت یا بازدارندگی خاص است؛ هر دو توجیه بر این فرض استوارند که عامل انسانی می‌توانسته است «انتخاب متفاوتی» داشته باشد.

این ناکارآمدی نه صرفاً یک کاستی در اثبات، بلکه یک بحران مفهومی در تئوری تقصیر است. وقتی عمل مجرمانه در لحظه وقوع، محصول محاسبات ماشینی است که متأثر از داده‌های محیطی در زمان اجرا بوده و نه تصمیم اولیه برنامه‌نویس، انتساب کیفری به طراح متضمن اجرای مجازات برای جرمی است که وی قصد وقوع آن را نداشته و توان پیشگیری کامل از آن را نیز نداشته است. این امر، نقض آشکار اصل مسئولیت مبتنی بر تقصیر را به دنبال دارد و می‌تواند منجر به صدور احکام کیفری شود که مبنای اخلاقی و قانونی محکمی ندارند و صرفاً به منظور مدیریت ریسک اجتماعی وضع شده‌اند.

با این حال، پرسش مهمی باقی می‌ماند: اگر طراح از ابتدا می‌دانسته است که هوش جامع مصنوعی ممکن است رفتارهای غیرقابل پیش‌بینی بروز دهد، آیا این آگاهی اولیه به معنای قابلیت پیش‌بینی رفتار مجرمانه در آینده است؟ پاسخ حقوقی به این پرسش، تمایز میان «پیش‌بینی‌پذیری نوعی» و «پیش‌بینی‌پذیری مصداقی» است. آگاهی طراح از اینکه سامانه ممکن است در آینده رفتارهای غیرقابل پیش‌بینی داشته باشد، صرفاً یک پیش‌بینی‌پذیری نوعی است؛ یعنی آگاهی از وجود یک ریسک کلی. اما مسئولیت کیفری نیازمند پیش‌بینی‌پذیری مصداقی است؛ یعنی توانایی پیش‌بینی رفتار مشخصی که منجر به جرم شده است. در سامانه‌های یادگیرنده، حتی اگر طراح بداند که رفتارهای غیرمنتظره ممکن است رخ دهد، نمی‌تواند رفتار خاصی را که در زمان وقوع جرم بروز کرده است پیش‌بینی کند. بنابراین، آگاهی از «امکان کلی خطر» به معنای پیش‌بینی «رفتار مجرمانه خاص» نیست و نمی‌تواند مبنای انتساب تقصیر کیفری قرار گیرد.بنابراین حتی با پذیرش آگاهی طراح از ریسک‌های کلی، رابطه علیت و عنصر تقصیر همچنان مخدوش باقی می‌ماند و نمی‌توان مسئولیت کیفری را صرفاً بر مبنای علم اجمالی به خطر بنا کرد. این نتیجه‌گیری، بار دیگر ضرورت بازنگری در مدل‌های مسئولیت و حرکت به سمت چارچوب‌های مبتنی بر ریسک یا شخصیت حقوقی مستقل را تقویت می‌کند.

۲-۲-۱-مسئولیت مطلق و عدم انطباق آن با حقوق کیفری

رویکرد مسئولیت مطلق، که در حقوق مدنی و در مورد فعالیت‌های ذاتاً خطرناک مانند رانندگی خودرو یا نقص محصول کاربرد دارد، در نگاه نخست جذاب‌ترین راهکار برای حل مشکل مسئولیت ناشی از هوش جامع مصنوعی به نظر می‌رسد. دلیل این جذابیت آن است که مسئولیت مطلق نیاز به اثبات تقصیر، نیت و قابلیت پیش‌بینی را از میان برمی‌دارد و تنها بر وقوع آسیب و رابطه علیت تمرکز می‌کند .(Kingston, J. K. C. 2020.314) ؛ مدلی که می‌تواند به‌طور مؤثر خسارات وارده به قربانیان را جبران کند. با این حال، در حوزه حقوق کیفری، این رویکرد با محدودیت‌های بنیادین مواجه می‌شود. اصل اساسی در حقوق کیفری، یعنی اصل تقصیر، مقرر می‌دارد که «مجازات بدون تقصیر جایز نیست». مسئولیت کیفری برخلاف مسئولیت مدنی که جنبه جبرانی دارد، ماهیتی تنبیهی و بازدارنده دارد و بنابراین نمی‌توان فردی اعم از تولیدکننده، مالک یا اپراتور را به دلیل عملی که نه می‌توانسته آن را پیش‌بینی کند و نه بر آن کنترلی داشته است، به مجازات محکوم کرد. .(Chesterman, S.2021.119) اعمال مسئولیت مطلق کیفری، نقض آشکار اصل قانونی بودن جرایم و مجازات‌ها و اصل شخصی بودن مسئولیت کیفری است، زیرا فرد را برای رفتار شخص دیگری(یا در اینجا، یک سیستم مستقل) مجازات می‌کند . (Danaher, J. 2016.299-309)

با این حال، نقد مسئولیت مطلق در این زمینه نیازمند دقت مفهومی بیشتری است. هرچند مسئولیت مطلق در حقوق کیفری به‌طور گسترده پذیرفته نشده، اما نمی‌توان آن را به‌طور مطلق ناسازگار با حقوق کیفری دانست. در برخی نظام‌های حقوقی، نمونه‌هایی محدود از مسئولیت کیفری بدون تقصیر وجود دارد؛ مواردی که معمولاً در حوزه تخلفات خطرمحور یا نقض الزامات ایمنی مشاهده می‌شوند و هدف آن‌ها حمایت از منافع عمومی است. (Ashworth, A., & Horder, J. 2013.178) . با

این حال، این استثنائات محدود را نمی‌توان به‌سادگی به حوزه هوش جامع مصنوعی تعمیم داد، زیرا در این موارد، همچنان نوعی کنترل یا امکان پیشگیری از سوی فاعل انسانی مفروض گرفته می‌شود. در مقابل، در سناریوهای مرتبط با هوش جامع مصنوعی، مسئله اصلی دقیقاً فقدان همین کنترل و پیش‌بینی‌پذیری است؛ یعنی همان بنیانی که استثنائات محدود مسئولیت کیفری بدون تقصیر بر آن استوارند.

بنابراین در خصوص رابطه علیت نیز باید توجه داشت که بحث علیت در مورد هوش جامع مصنوعی هسته اصلی چالش انتساب مسئولیت کیفری است، زیرا ارکان علیت سنتی علیت مادی و علیت حقوقی را دچار شکاف می‌کند. هنگامی که هوش جامع مصنوعی بر اساس یادگیری عمیق عمل می‌کند، رفتار آن به دلیل خودآموختگی و انطباق با محیط از برنامه اولیه و تصمیم‌های انسانی فاصله می‌گیرد. در این حالت، اثبات اینکه عمل مجرمانه مستقیماً ناشی از یک نقص در کدنویسی اولیه بوده و نه از تصمیم مستقل هوش جامع مصنوعی، دشوار می‌شود. (Bathaee, Y. 2018.920) این استقلال رفتاری یک گسست در زنجیره علیت ایجاد می‌کند، زیرا عمل هوش جامع مصنوعی می‌تواند به‌عنوان یک عامل دخالت‌کننده جدید و مستقل تلقی شود که علیت انسانی اولیه را از بین می‌برد. اگر این عامل دخالت‌کننده جدید قابل پیش‌بینی نباشد، رابطه‌ای که نقص برنامه اولیه را به نتیجه نهایی پیوند می‌دهد از نظر حقوقی بسیار دور قلمداد می‌شود و نتیجه آن است که برنامه‌نویس یا طراح نمی‌تواند مسئول نهایی ضرر باشد(Vladeck, D. C.2014.135).

بنابراین مشکل اصلی نه در نقص فنی، بلکه در فقدان یک ساختار حقوقی-فلسفی مناسب است. چالش اصلی در علیت حقوقی است، نه مادی: هوش جامع مصنوعی قطعاً از نظر فیزیکی باعث آسیب می‌شود، اما نظام حقوقی نمی‌تواند فردی را مسئول بداند که عمل او از نظر زمانی، مکانی یا شناختی با نتیجه فاصله زیادی دارد. استقلال هوش جامع مصنوعی، کد اولیه را به یک عامل بسیار دور تبدیل می‌کند. بنابراین، مسئولیت باید بر عهده کسی باشد که بیشترین توانایی و وظیفه را برای مدیریت ریسک‌های غیرقابل پیش‌بینی هوش جامع مصنوعی دارد؛ مانند تولیدکننده در زمان انتشار محصول یا مالک در زمان استفاده. این مدل جدید با تمرکز بر وظیفه مراقبت در مرحله طراحی و آزمایش، تلاش می‌کند علیت را پیش از گسست کامل به عامل انسانی بازگرداند.

در پی اثبات ناکارآمدی مدل‌های سنتی انتساب تقصیر و علیت به عوامل انسانی در سناریوهای خودمختاری و یادگیری عمیق هوش جامع مصنوعی، بحث ناگزیر به سمت بررسی راهکاری رادیکال و در عین حال تنها گزینه ممکن برای پر کردن شکاف مسئولیت هدایت می‌شود. این راهکار، اعطای شخصیت حقوقی نوین و محدود به سامانه‌های هوش جامع مصنوعی است. این پیشنهاد نه بر مبنای تحقق آگاهی فلسفی یا اراده آزاد، بلکه بر اساس ضرورت کارکردی در نظام حقوقی مطرح می‌شود. زیرا نظام حقوقی نمی‌تواند یک فاعل علّی را که قادر به ایجاد خسارت‌های بزرگ است، بدون مسئولیت رها کند. (Solaiman, S. M. 2017.280) انتساب این شخصیت حقوقی نوین امکان می‌دهد تا اعمال هوش جامع مصنوعی به‌عنوان یک فعل حقوقی مستقل در نظر گرفته شود و مسئولیت آن مستقیماً به نهاد هوش مصنوعی منتقل گردد. با این حال، پذیرش چنین

«شخصیت الکترونیکی» پیامدهای حقوقی، اقتصادی و اخلاقی عمیقی در پی خواهد داشت؛ از جمله تغییر مفهوم دارایی، ایجاد صندوق‌های جبران خسارت خودکار، و مهم‌تر از همه، بازنگری کامل در ساختار حقوق کیفری و مدنی که تاکنون صرفاً بر مبنای شخصیت انسانی بنا شده است.

۳-مدل پیشنهادی: شخصیت حقوقی الکترونیکی کیفری

در نظام‌های حقوقی غرب، مفهوم شخصیت حقوقی به‌گونه‌ای توسعه یافته که از انحصار انسان خارج شده و شامل هر پدیده‌ای می‌شود که بتواند موضوع حقوق و تعهدات قرار گیرد. در واقع قانون‌گذاران نه‌تنها از منافع فرد به‌عنوان شخص حمایت می‌کنند، بلکه از منافع گروه‌ها و اجتماعات بشری که برای وصول به هدف‌های مشترک تشکیل می‌شوند نیز حمایت می‌نمایند؛ به‌طوری‌که شرکت‌ها در اکثر کشورها و تراست‌ها، انجمن‌ها و معابد در برخی کشورها دارای شخصیت حقوقی هستند.(حسینی ، ۱۴۰۴، 153).

پیش از ورود به بحث شخصیت حقوقی الکترونیکی، لازم است مشخص شود چه نوع سامانه‌ای موضوع این تحلیل است. هوش جامع مصنوعی به سیستمی اطلاق می‌شود که قادر به انجام هرگونه تکلیف شناختی در سطحی برابر یا فراتر از انسان، شامل توانایی یادگیری، درک، استدلال در محیط‌های متنوع و حل مسائل کلی است .Goertzel, B., & Pennachin, C) (Eds.). 2007.1).

مخالفان اعطای شخصیت حقوقی به هوش جامع مصنوعی استدلال می‌کنند که شخصیت حقوقی در نظام‌های حقوقی سنتی، مستلزم وجود «اراده مستقل» و «آگاهی» است؛ ویژگی‌هایی که هوش جامع مصنوعی فاقد آن‌هاست. از این منظر، اعطای شخصیت حقوقی به یک سامانه غیرآگاه، نه تنها بی‌اساس، بلکه خطرناک است؛ زیرا ممکن است به مشروعیت‌بخشی به نوعی «فاعل غیرانسانی» منجر شود و مرزهای سنتی مسئولیت را تضعیف کند. (Bathaee, Y.2018.31) همچنین، این نگرانی وجود دارد که شخصیت حقوقی، سپری برای عوامل انسانی ایجاد کند و به آن‌ها امکان دهد از مسئولیت اعمال سیستم‌هایی که خود طراحی کرده‌اند، شانه خالی کنند.(Chesterman, S. 2021.132) این دیدگاه تأکید می‌کند که مسئولیت باید همواره بر عهده انسان‌ها باقی بماند و راه‌حل، نه خلق شخصیت‌های جدید، بلکه تقویت مکانیسم‌های انتساب مسئولیت به عوامل انسانی از طریق مسئولیت سازمانی است.

در پاسخ به این دیدگاه، باید اذعان داشت که مدل مسئولیت سازمانی، یعنی انتساب مسئولیت به شرکت‌ها و نهادهای پشت پرده هوش جامع مصنوعی، در نگاه نخست جذاب به نظر می‌رسد. این مدل استدلال می‌کند که به‌جای اعطای شخصیت به ماشین، باید استانداردهای نظارتی و ایمنی را برای سازمان‌های توسعه‌دهنده و بهره‌بردار تشدید کرد. (Abbott, R., & Sarch, A. 2019.366). با این حال، این رویکرد در مواجهه با ویژگی خودمختاری کامل هوش جامع مصنوعی با محدودیت‌های جدی روبروست. در سامانه‌هایی که از یادگیری تقویتی عمیق استفاده می‌کنند، حتی توسعه‌دهندگان اصلی نیز نمی‌توانند

به‌طور دقیق پیش‌بینی کنند که سیستم در موقعیتی خاص چه تصمیمی خواهد گرفت (Bathaee, Y. 2018.916) این عدم شفافیت، پیوند لازم برای اثبات رابطه علیت و تقصیر را از بین می‌برد. مسئولیت سازمانی، هرچند می‌تواند انگیزه‌های ایمنی را تقویت کند، اما نمی‌تواند گسست علیت را در مواردی که رفتار مجرمانه کاملاً خارج از پارامترهای طراحی و نظارت سازمانی رخ می‌دهد، ترمیم کند (Matthias, A. 2004.183) به بیان دیگر، مدل سازمانی صرفاً مسئولیت را به یک سطح بالاتر منتقل می‌کند، اما مشکل ساختاری«مسئولیت» را حل نمی‌کند.

رویکرد دیگر، محدود کردن شخصیت حقوقی به حوزه مدنی و جبران خسارت است، بدون ورود به حوزه کیفری. این رویکرد که در قطعنامه پارلمان اروپا در سال ۲۰۱۷ نیز منعکس شده، شخصیت الکترونیکی را صرفاً ابزاری برای داخلی‌سازی هزینه‌ها و ایجاد صندوق‌های جبران خسارت می‌داند (European Parliament. 2017.10) مزیت این مدل، سادگی و اجتناب از چالش‌های فلسفی مسئولیت کیفری است. اما این رویکرد نیز به‌تنهایی کافی نیست. دلیل این ناکافی بودن آن است که حقوق کیفری کارکردی فراتر از جبران خسارت دارد: کارویژه اصلی آن، ایجاد بازدارندگی عمومی و خاص، و تحمیل سرزنش اجتماعی است (Ashworth, A., & Horder, J. 2013) اگر یک سامانه هوش جامع مصنوعی مرتکب عملی شود که در صورت ارتکاب توسط انسان، مستوجب مجازات کیفری سنگین می‌بود، اکتفا به جبران خسارت مدنی، پیام روشنی به جامعه ارسال نمی‌کند و کارکرد بازدارندگی و سرزنش حقوق کیفری را معطل می‌گذارد. بنابراین، رویکرد صرفاً مدنی، هرچند در حوزه جبران خسارت کارآمد است، اما در تأمین اهداف کیفری ناتوان می‌ماند.

با پذیرش ضرورت نوعی شخصیت حقوقی الکترونیکی کیفری کیفری خاص‌منظور، پرسش اساسی این است که مجازات‌های پیشنهادی تا چه اندازه با نظریه‌های فلسفی مجازات سازگارند. تحلیل هر یک از نظریه‌های کلاسیک در زمینه هوش جامع مصنوعی، نتایج متفاوتی به دست می‌دهد. نخست، از منظر بازدارندگی عمومی، مجازات‌های پیشنهادی می‌توانند کارکرد مؤثری داشته باشند، اما نه از طریق تأثیر روانی بر خود ماشین، بلکه از طریق تأثیرگذاری بر رفتار عوامل انسانی. مجازات‌های مالی و جریمه‌های سنگین که از محل صندوق شخصیت الکترونیکی پرداخت می‌شوند، در نهایت به افزایش هزینه‌های توسعه‌دهندگان و مالکان منجر می‌شوند و انگیزه‌های اقتصادی برای طراحی ایمن‌تر را تقویت می‌کنند (Kingston, J. K. C. 2020.309-320.). مجازات‌های غیرمالی مانند الزام به بازآموزی الگوریتمی نیز با افزایش هزینه‌های عملیاتی، همین اثر را دارند. دوم، از منظر بازدارندگی خاص و ناتوان‌سازی، مجازات‌هایی مانند «تعلیق دائم عملکرد» یا «خروج اجباری از چرخه بهره‌برداری» کاملاً منطبق با هدف ناتوان‌سازی هستند: این مجازات‌ها، سامانه خطرناک را برای همیشه از چرخه فعالیت خارج می‌کنند و از تکرار آسیب جلوگیری می‌نمایند. این منطق، مشابه انحلال شرکت‌های متخلف در حقوق کیفری اشخاص حقوقی است. (Wells, C. 2001.165) الزام به بازآموزی الگوریتمی نیز معادل «اصلاح و بازپروری» در نظریه‌های سنتی است که هدف آن تغییر رفتار مجرمانه و بازگرداندن عامل به جامعه است. سوم، از منظر سزادهی(تقاص)، چالش اصلی

نمایان می‌شود. نظریه‌های سزادهی، مجازات را به دلیل استحقاق اخلاقی مجرم توجیه می‌کنند و مستلزم آن است که مجرم، واجد آگاهی اخلاقی و قابلیت سرزنش باشد (Duff, R. A.2001.82). از آنجا که هوش جامع مصنوعی فاقد این آگاهی است، مجازات آن از منظر سزادهی فاقد توجیه اخلاقی است. این همان نقطه‌ای است که مدل پیشنهادی از مسئولیت کیفری اخلاقی فاصله می‌گیرد و به مسئولیت شبه‌کیفری مبتنی بر ضرورت اجتماعی نزدیک می‌شود. مجازات در این چارچوب، نه یک واکنش اخلاقی به تقصیر، بلکه یک ابزار کنترلی برای مدیریت ریسک و حمایت از جامعه است. بنابراین، باید پذیرفت که کارکرد اصلی مجازات‌های پیشنهادی، «ناتوان‌سازی» و «تنظیم‌گری» است، نه «سزادهی اخلاقی».

در واقع مجازات‌های قابل اعمال بر شخصیت حقوقی هوش جامع مصنوعی در سه سطح طراحی می‌شوند که هر یک با منطق یک نهاد غیرانسانی سازگار شده‌اند. دسته نخست، مجازات‌های مالی هستند که شامل جریمه‌های مالی و الزام به پرداخت غرامت به قربانیان از محل صندوق جبران خسارت شخصیت الکترونیکی می‌شود. این رویکرد، مشابه جریمه کردن شرکت‌ها در حقوق معاصر بوده و هدف آن جبران خسارت، بازدارندگی اقتصادی، و داخلی‌سازی هزینه‌های اجتماعی فعالیت‌های پرخطر است(Hallevy, G. 2015.211). این مجازات‌ها مستلزم آن است که شخصیت حقوقی هوش جامع مصنوعی دارای دارایی‌های مالی کافی باشد که از طریق الزام قانونی به ایجاد صندوق ضمانت تأمین می‌شود. دسته دوم، الزام به بازآموزی الگوریتمی اجباری است. این مجازات منحصر به فرد، شامل الزام به تغییر و اصلاح الگوریتم‌ها و پارامترهای تصمیم‌گیری به‌منظور حذف رفتارهای مجرمانه است. این مجازات، معادل «اصلاح رفتار» در انسان و «الزام به تغییر رویه‌های سازمانی» در شرکت‌هاست (Pagallo, U. 2013.153) فرایند بازآموزی تحت نظارت نهاد رگولاتور انجام می‌شود و هدف آن بازگرداندن سامانه به چرخه فعالیت ایمن است. دسته سوم، تعلیق دائم عملکرد یا خروج اجباری از چرخه بهره‌برداری است که شدیدترین مجازات برای شخصیت حقوقی هوش جامع مصنوعی محسوب می‌شود. این مجازات، معادل انحلال شخصیت حقوقی در مورد شرکت‌هاست و برای مواردی در نظر گرفته می‌شود که سامانه مرتکب جرایم غیرقابل جبران شده یا غیرقابل اصلاح تشخیص داده شود.(Wells, C. 2001.167) باید توجه داشت که این مجازات، با وجود شباهت کارکردی به انحلال، از جنس مجازات‌های بدنی یا سالب حیات نیست؛ زیرا هوش جامع مصنوعی فاقد حیات بیولوژیک و آگاهی است. تعبیر دقیق‌تر حقوقی برای این مجازات، «انحلال فنی سامانه» یا «خاتمه شخصیت حقوقی الکترونیکی کیفری» است که نشان‌دهنده پایان یافتن موجودیت حقوقی و عملیاتی سیستم است، نه مجازات یک موجود ذی‌شعور.

در واقع مدل شخصیت حقوقی الکترونیکی کیفری، یک نظریه واحد و همگن نیست؛ بلکه مجموعه‌ای از رویکردهای متفاوت را در بر می‌گیرد که از «شخصیت حقوقی حداقلی برای جبران خسارت» تا «شخصیت حقوقی شبه‌شرکتی با ظرفیت‌های گسترده‌تر» متغیر است. تعارض اصلی میان دو رویکرد شکل می‌گیرد: رویکردی که شخصیت الکترونیکی را صرفاً ابزاری برای داخلی‌سازی هزینه‌ها و جبران خسارت می‌داند، و رویکردی که آن را گامی به سوی پذیرش نوعی

«فاعل‌بودگی حقوقی» برای سامانه‌های خودمختار تلقی می‌کند. رویکرد نخست، با تأکید بر کارکردگرایی، از این ایده دفاع می‌کند که شخصیت حقوقی صرفاً یک ابزار نهادی برای مدیریت ریسک است و هیچ دلالتی بر آگاهی یا اراده ندارد. (Kurki, V. A. J. 2019.175). در مقابل، رویکرد دوم هشدار می‌دهد که اعطای شخصیت می‌تواند به تدریج به مشروعیت‌بخشی به «فاعل غیرانسانی» منجر شود (Bryson, J. J., Diamantis, M. E., & Grant, T. D. 2017.273-291) این مقاله ضمن اذعان به اعتبار نگرانی‌های رویکرد دوم، از رویکرد نخست دفاع می‌کند: شخصیت حقوقی الکترونیکی کیفری خاص‌مورد بحث، نه یک بیانیه فلسفی درباره آگاهی ماشین، بلکه یک ضرورت عملی برای پر کردن خلاء عدم انطباق ساختاری مسئولیت است که در صورت عدم پذیرش آن، نظام حقوقی از تأمین اهداف بازدارندگی و حمایت از بزه‌دیدگان باز خواهد ماند. در نهایت، هیچ‌یک از رویکردهای موجود به‌تنهایی قادر به حل کامل چالش‌های ناشی از هوش جامع مصنوعی نیستند. شخصیت حقوقی الکترونیکی کیفری می‌تواند خلأ جبران خسارت را پر کند، اما در حوزه کیفری نیازمند تکمیل با مدل‌های مسئولیت سازمانی است. راه‌حل مطلوب، ترکیبی از شخصیت حقوقی الکترونیکی کیفری برای انتساب مستقیم مسئولیت و اجرای مجازات‌های خاص، مسئولیت سازمانی برای ایجاد انگیزه‌های پیشینی در طراحی ایمن، و نظام‌های بیمه اجباری و صندوق‌های جبران خسارت برای حمایت از قربانیان است.

٤-سازوکارهای نظارتی و قضایی لازم

تعامل هوش مصنوعی با دادرسی منصفانه نیازمند پایبندی به اصل حاکمیت قانون است؛ اصلی که با تکیه بر مؤلفه‌هایی چون قانون‌مندی، انتشار قواعد و ثبات نسبی (مطابق نظریه فولر)، تضمین‌کننده حقوق شهروندان در برابر اقدامات خودسرانه است. در همین راستا، تنش میان هوش مصنوعی قانون‌مند و سامانه‌های مبتنی بر یادگیری خودکار، لزوم بازنگری در ساختارهای اجرایی و قضایی را برجسته می‌سازد.(مصطفوی اردبیلی ،۱۴۰۱، ۵۱)

 اجرای موفقیت‌آمیز مدل مسئولیت هوش جامع مصنوعی مستلزم بازنگری در دو سطح نهادی و رویه‌ای است. در سطح نهادی، پیچیدگی‌های فنی و حقوقی پرونده‌های هوش جامع مصنوعی نیاز به ایجاد دادگاه‌ها یا هیئت‌های تخصصی دارد. این هیئت‌ها باید ترکیبی از قضات متخصص در امور حقوقی و کارشناسان خبره در علوم کامپیوتر، یادگیری ماشین، و اخلاق داشته باشند .

هدف از این تخصص‌گرایی این است که فرآیند قضایی قادر باشد علیت و عملکرد فنی هوش جامع مصنوعی را به درستی درک کند، از اتکا به شواهد سطحی پرهیز نماید، و مجازات‌های تنبیهی-اصلاحی مناسب (مانند اجبار به آموزش مجدد) را صادر کند. تنها از طریق چنین تریبون‌های چندرشته‌ای، می‌توان قضاوت‌هایی صادر کرد که هم از نظر حقوقی صحیح باشند و هم از نظر فنی قابل اجرا(Bathaee, Y. 2018.889-938). اگرچه هوش جامع مصنوعی فاقد عنصر روانی انسانی است، اما

برای اثبات عنصر مادی و همچنین معادل‌های سوءنیت (مانند نیت شبیه‌سازی شده یا قصد تقصیر)، سیستم حقوقی باید قادر به ردیابی و تفسیر مسیر تصمیم‌گیری الگوریتم باشد. این الزام به شفافیت که اغلب با اصطلاح شفاف‌سازی هوش مصنوعی شناخته می‌شود، باید از طریق قانون به تولیدکنندگان هوش جامع مصنوعی تحمیل شود تا در صورت وقوع جرم، «گزارش‌های علیت» و «گزارش‌های رفتار» سیستم به نهادهای قضایی ارائه گردد . ,Wachter, S., Mittelstadt, B., & Floridi) .(L.2017.76-99 بدون این شفافیت ریشه‌ای، اثبات علیت حقوقی برای انتساب جرم، چه به هوش جامع مصنوعی(شخصیت الکترونیکی) و چه به عوامل انسانی(مسئولیت نظارتی)، عملاً غیرممکن خواهد بود.

ترکیب دادگاه‌های تخصصی با الزام به شفاف‌سازی هوش جامع مصنوعی، نمایانگر پذیرش یک واقعیت بنیادین است: حقوق نمی‌تواند به تنهایی عمل کند. الزام به شفافیت و شفاف‌سازی هوش جامع مصنوعی، در واقع تلاشی حقوقی برای جبران فقدان آگاهی هوش جامع مصنوعی است. از آنجایی که نمی‌توانیم از خود هوش جامع مصنوعی بپرسیم چرا مرتکب عمل مجرمانه شده، حقوق باید تولیدکنندگان را مجبور کند که یک معادل مصنوعی از پاسخ(مسیر تصمیم‌گیری قابل فهم) را ارائه دهند. بنابراین ایجاد دادگاه‌های تخصصی خطر انتقال قدرت قضایی از اصول حقوقی به تحلیل فنی را در پی دارد. این هیئت‌ها باید به دقت ساختاردهی شوند تا اطمینان حاصل شود که کارشناسان فنی(که اغلب تمایل به تمرکز بر احتمال فنی دارند) بر اصول عدالت، تقصیر و تناسب(که حیطه تخصصی حقوق هستند) غلبه نکنند (Pasquale, F. 2015.142) اما باید توجه داشت که این تخصص‌گرایی در نهایت باید در خدمت عدالت سنتی(تحقق عدالت و بازدارندگی) باقی بماند و نه در خدمت پیچیدگی‌های فنی.

بنابراین ایجاد دادگاه‌های تخصصی چندرشته‌ای یک ضرورت عملی برای مدیریت پرونده‌های هوش جامع مصنوعی است، اما از منظر دکترین حقوقی، این اقدام باید با احتیاط مدیریت شود تا از انحراف سیستم قضایی از اصول اساسی خود جلوگیری شود. چالش اصلی در این ساختار، حفظ اقتدار اصول حقوقی(مانند اصل قانونی بودن جرم و مجازات و بار اثبات) در برابر تفسیر فنی است. کارشناسان فنی ممکن است بر اساس معیارهای مهندسی(مانند کارایی یا دقت الگوریتم) قضاوت کنند، در حالی که حقوق نیازمند ارزیابی قصد مجرمانه(یا معادل آن) و رابطه سببی قابل انتساب است. بنابراین، ساختاردهی این هیئت‌ها باید تضمین کند که نقش کارشناسان فنی صرفاً تفسیر شواهد فنی(گزارش‌های علیت) باشد و تصمیم نهایی مبتنی بر سنجش قضایی اصول عدالت باقی بماند.

در رابطه با الزام به شفاف‌سازی الگوریتم‌ها، این سازوکار تلاشی برای «پلی ساختن بر شکاف آگاهی» هوش جامع مصنوعی است، اما خود می‌تواند ریسک‌هایی را به دنبال داشته باشد. از یک سو، فراهم آوردن «گزارش‌های علیت» برای ردیابی مسیر تصمیم، عملاً برای اثبات عنصر مادی و تقصیر نظارتی ضروری است. از سوی دیگر، اجبار به افشای کامل کد و وزن‌های شبکه عصبی می‌تواند امنیت سایبری، مالکیت فکری تولیدکنندگان، و مزیت رقابتی شرکت‌ها را به خطر اندازد . .Selbst, A)

.(D., & Barocas, S. 2018.1085-1139 بنابراین، باید یک تعادل حقوقی–فنی برقرار شود؛ شفاف‌سازی باید محدود به سطح توضیحات عملیاتی و سببی باشد که برای قضاوت در مورد تقصیر کافی است، نه افشای کامل دانش محرمانه فنی سیستم.در نهایت، موفقیت این سازوکارها منوط به تدوین سیاست‌گذاری پیشینی است، نه صرفاً واکنش قضایی پسینی. نهادهای نظارتی پیش از وقوع جرم، باید استانداردهای الزامی برای حداقل شفافیت قابل قبول را تعیین کنند و سطح قابل انتظار از قابلیت پیش‌بینی را در قراردادهای صدور مجوز هوش جامع مصنوعی بگنجانند. این رویکرد پیشگیرانه، با پیوند زدن مسئولیت نهادی به رعایت استانداردهای شفافیت فنی، زمینه را برای اعمال مؤثر تئوری‌های مسئولیت انسانی فراهم می‌سازد و اجازه نمی‌دهد که پیچیدگی فنی به دستاویزی برای فرار از مسئولیت کیفری تبدیل شود.

نتیجه گیری

تحقیقات صورت‌گرفته نشان می‌دهد که تعارض بنیادین میان خودمختاری عملی هوش جامع مصنوعی و اصول کلاسیک حقوق کیفری، به‌ویژه اصل تقصیر و لزوم نیت، مدل‌های سنتی انتساب مسئولیت به عوامل انسانی(اعم از تولیدکننده، مالک یا کاربر) را کاملاً به بن‌بست کشانده است. ارکان مسئولیت ثابت می‌کند که عملکرد خودتکامل‌یاب هوش جامع مصنوعی، به‌طور خاص، منجر به گسست علّی در زنجیره علیت می‌شود. از آنجا که اقدامات غیرقابل پیش‌بینی هوش جامع مصنوعی قابل انتساب به تقصیر یا نیت انسانی نیست، یک خلاء انتساب مسئولیت حقوقی عمیق ایجاد می‌گردد. در مواجهه با این خلأ، تلاش برای اعطای مسئولیت کیفری مستقل به هوش جامع مصنوعی تحت چارچوب‌های حقوقی سنتی کاملاً مردود است، زیرا هوش جامع مصنوعی فاقد آگاهی اخلاقی و اراده انسانی است. در نتیجه، تنها مسیر عملی و کارکردگرا، عبور از مدل‌های انسان‌محور و پذیرش یک راهکار نوین حقوقی است.

راه‌حل عملی برای مهار این شکاف حقوقی، در ایجاد یک شخصیت حقوقی نوین و محدود تحت عنوان شخصیت حقوقی الکترونیکی کیفری برای هوش جامع مصنوعی نهفته است. این شخصیت حقوقی الکترونیکی کیفری نباید به معنای اعطای حقوق کیفری کامل شهروندی باشد، بلکه صرفاً یک ابزار جبرانی و نظارتی است. هدف اصلی این ابزار حقوقی، انتقال مستقیم مسئولیت به نهاد مصنوعی است. این انتقال، دو هدف کلیدی را تضمین می‌کند: نخست، ایجاد بازدارندگی سازمانی(از طریق تهدید به مجازات‌هایی چون خاموش کردن سیستم یا اجبار به آموزش مجدد) و دوم، تضمین جبران خسارت قربانیان که بدون وجود این محور مسئولیت، حقوق آن‌ها تضییع خواهد شد. این مدل، در حقیقت، مسئولیت را از تقصیر سنتی جدا کرده و بر مبنای نیازهای اجتماعی و کارکردی سیستم حقوقی بنا می‌نهد.

بر این اساس، در مقام پیشنهاد تقنینی برای رسمیت بخشیدن به این مدل کارکردگرا، تدوین یک چارچوب قانونی زیرساختی ضروری است. پیشنهاد می‌شود ماده‌ای قانونی تدوین گردد که به‌صراحت تصریح کند: "سامانه‌های هوش جامع مصنوعی که دارای سطح اثبات‌شده‌ای از خودمختاری عملی و ظرفیت یادگیری خودتکمیلی هستند، به موجب قانون، دارای شخصیت

حقوقی الکترونیکی کیفری محدودمی‌باشند." این شخصیت صرفاً برای مقاصد انتساب مسئولیت و جبران خسارت ایجاد شده و مسئولیت آن محدود به مجازات‌های مالی(مانند جریمه‌ها و دیات) و مجازات‌های تنبیهی غیرمالی(اجبار به آموزش مجدد و در موارد جرایم غیرقابل جبران، خاموش کردن-انحلال سامانه) خواهد بود. همچنین، اجرای این ماده نیازمند الزامات سخت‌گیرانه‌ای چون شفاف‌سازی الگوریتم از سوی تولیدکننده برای ارائه گزارش‌های علیت و رفتار سیستم به هیئت‌های قضایی تخصصی است تا بستر لازم برای اجرای عدالت کیفری در عصر سامانه‌های خودمختار فراهم گردد.

منابع

فارسی

-ابوذری، مهرنوش (۱۴۰۳). «در جستجوی ردپای نقض هوشمند مالکیت فکری». دوفصلنامه تحقیق و توسعه در حقوق خصوصی، پژوهشکده حقوق و قانون ایران، دوره ۱، شماره ۱

-امیریان فارسانی، امین، و ابوذری، مهرنوش (۱۴۰۲). «چالش‌ها و موانع مسئولیت کیفری در ربات‌های با قابلیت هوش مصنوعی». تمدن حقوقی، دوره ۶، شماره ۱۸ (ویژه‌نامه هوش مصنوعی)، صص. ۳۹-۵۴.

-بارانی، محمد، و کاوه، محمدهادی (۱۴۰۴). «مسؤولیت کیفری هوش مصنوعی در حقوق کیفری ایران با نگاهی به قوانین اتحادیه اروپا». مجله حقوق کیفری و جرم‌شناسی، دوره ۴، شماره ۳.

-حسینی، سید امیرعلی، و هاشمی‌زاده، سید علیرضا (۱۴۰۴). «بررسی اعطای شخصیت حقوقی (شرکتی) به هوش مصنوعی». حقوق فناوری‌های نوین، دوره ۱۲، شماره ۴۷، صص. ۶-۱۴.

-سلیمان‌زاده، سمیرا (۱۴۰۴). «شخصیت حقوقی، تاریخچه، مبانی و آثار، با تأکید بر بررسی تطبیقی شرکت‌های تجاری در نظام حقوقی ایران و حقوق نوشته (فرانسه و آلمان)». فصلنامه تحقیق و توسعه در حقوق تطبیقی، پژوهشکده حقوق و قانون ایران، دوره ۸، شماره ۲۶، صص. ۷۹-۱۰۶.

-مصطفوی اردبیلی، سید محمدمهدی؛ تقی‌زاده انصاری، مصطفی؛ و رحمتی‌فر، سمانه (۱۴۰۱). «کارکردها و بایسته‌های هوش مصنوعی از منظر دادرسی منصفانه». دوفصلنامه علمی حقوق فناوری‌های نوین، دوره ۳، شماره ۶.

انگلیسی

- Abbott, R., & Sarch, A. (2019). Punishing Artificial Intelligence: Legal Fiction or Science Fiction. UC Davis Law Review, 53(1), 323-384.

- Ashworth, A., & Horder, J. (2013). Principles of Criminal Law (7th ed.). Oxford: Oxford University Press.

-Baldwin, Robert, Martin Cave, and Martin Lodge. Understanding Regulation: Theory, Strategy, and Practice. 2nd ed. Oxford: Oxford University Press, 2012, pp. 15-40.

- Bathaee, Y. (2018). The Artificial Intelligence Black Box and the Failure of Intent and Causation. Harvard Journal of Law & Technology, 31(2), 889-938.

- Bryson, J. J., Diamantis, M. E., & Grant, T. D. (2017). Of, for, and by the People: The Legal Lacuna of Synthetic Persons. Artificial Intelligence and Law, 25(3), 273-291.

- Chesterman, S. (2021). We, the Robots? Regulating Artificial Intelligence and the Limits of the Law. Cambridge: Cambridge University Press.

- Chopra, S., & White, L. F. (2011). A Legal Theory for Autonomous Artificial Agents. Ann Arbor: University of Michigan Press.

- Danaher, J. (2016). Robots, Law and the Retribution Gap. Ethics and Information Technology, 18(4), 299-309.

- Duff, R. A. (2001). Punishment, Communication, and Community. Oxford: Oxford University Press.

- European Parliament. (2017). Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)).

-European Commission (2022). Proposal for a Directive on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive). COM(2022) 496 final.

-Goertzel, B., & Pennachin, C. (Eds.). (2007). Artificial General Intelligence. Berlin: Springer.

- Hallevy, G. (2015). Liability for Crimes Involving Artificial Intelligence Systems. Cham: Springer.

-Kingston, J. K. C. (2020). Artificial Intelligence and Legal Liability. In M. Bramer & M. Petridis (Eds.), Artificial Intelligence XXXVII (pp. 309-320). Cham: Springer.

- Kraakman, R., Armour, J., Davies, P., Enriques, L., Hansmann, H., Hertig, G., ... & Rock, E. (2017). The Anatomy of Corporate Law: A Comparative and Functional Approach (3rd ed.). Oxford: Oxford University Press.

- Kurki, V. A. J. (2019). A Theory of Legal Personhood. Oxford: Oxford University Press.

- Matthias, A. (2004). The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata. Ethics and Information Technology, 6(3), 175-183.

- Pagallo, U. (2013). The Laws of Robots: Crimes, Contracts, and Torts. Dordrecht: Springer.

- Pasquale, F. (2015). The Black Box Society: The Secret Algorithms That Control Money and Information. Cambridge: Harvard University Press.

- Pietrzykowski, T. (2018). Personhood Beyond Humanism: Animals, Artificial Intelligence, and the Law. Cham: Springer.

-Pagallo, U., et al. (2021). The Auditability of AI Systems: Legal and Technical Challenges. Computer Law & Security Review, 40, 105527

- Robertson, D. W. (2012). Admiralty and Maritime Law in the United States (2nd ed.). Durham: Carolina Academic Press.

- Russell, S., & Norvig, P. (2021). Artificial Intelligence: A Modern Approach (4th ed.). Hoboken: Pearson.

-Lagioia, F., & Sartor, G. (2020). AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective. Philosophy & Technology, 33, 433-465.

- Selbst, A. D., & Barocas, S. (2018). The Intuitive Appeal of Explainable Machines. Fordham Law Review, 87(3), 1085-1139.

- Vladeck, D. C. (2014). Machines Without Principals: Liability Rules and Artificial Intelligence. Washington Law Review, 89(1), 117-150.

- Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. International Data Privacy Law, 7(2), 76-99.

- Wells, C. (2001). Corporations and Criminal Responsibility (2nd ed.). Oxford: Oxford University Press.