Journal of urban law Open Access

Spring (2026) 2(1): 1-14

ORIGINAL ARTICLE

Challenges of Applying Artificial Intelligence in the Field of Citizens' Rights with an Emphasis on Private Law

Mohsen Jafari Behzadkalaei[1*], Mona Rahbar[2]

1. Faculty member of Payam Noor University, Tehran, Iran.
2. PhD student in Artificial Intelligence, University of Tabriz, Tabriz, Iran.

Correspondence:
Mohsen Jafari Behzadkalaei
Email: mohsenjafari@pnu.ac.ir

Received: 10 Nov 2025
Accepted: 04 Jun 2026

How to cite
Jafari Behzadkalaei, M. & Rahbar, M. (2026). Challenges of Applying Artificial Intelligence in the Field of Citizens' Rights with an Emphasis on Private Law. Journal of urban law, 2(1), 1-14.
(DOI:)

ABSTRACT

The technological transformation of artificial intelligence has not only confronted the traditional system of civil liability with a crisis of attribution of fault and causality, but has also challenged broader foundations of citizens' rights, such as the right to explainability, the right to contest automated decisions, and effective access to justice. In the self-adaptive algorithmic environment, citizens interacting with systems such as risk assessment in employment, loan approval, urban services (automated fines, public housing allocation) and machine adjudication face the "black box" phenomenon, which makes reconstructing the logic of decisions and vindicating rights impossible. This research, conducted through a descriptive-analytical method, aims to move from a purely theoretical approach to presenting an operational model based on "civil liability conditional upon operational transparency" in order to preserve technological innovation while securing various dimensions of citizens' rights (including compensation for damages, elimination of algorithmic discrimination, and judicial oversight). Findings show that achieving justice in the self-learning era requires linking technical indicators of transparency and data auditability with the criterion of fault—an approach that replaces mere error liability with traceability-based governance. The proposed model, by requiring documentation of robustness and fairness tests and shifting the burden of proof in case of refusal to provide technical reports, provides the basis for proactive liability. By integrating explainability as a new indicator of ordinary care, this model establishes the legal foundation for building trust in citizen-autonomous system interactions.

KEYWORDS

Citizens' Rights, Artificial Intelligence, Private Law, Civil Liability, Algorithmic Transparency.

Copyright © 2026 The Authors. Published by Payame Noor University.
This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license (https://creativecommons.org/licenses/by-nc/4.0/). Non-commercial uses of the work are permitted, provided the original work is properly cited.

مجله حقوق شهری

سال دوم، شماره یکم، پیاپی دوم، بهار ۱۴۰۵ (۱-۱۴)

«مقاله پژوهشی»

چالش‌های به‌کارگیری هوش مصنوعی در حوزه حقوق شهروندی با تأکید بر حقوق خصوصی

محسن جعفری بهزادکلائی[1*]، مونا رهبر[2]

۱. استادیار گروه حقوق، دانشگاه پیام نور، تهران، ایران.
۲. دانشجوی دکتری هوش مصنوعی، دانشگاه تبریز، تبریز، ایران.

نویسنده مسئول:
محسن جعفری بهزادکلائی
رایانامه: mohsenjafari@pnu.ac.ir

تاریخ دریافت: ۱۴۰۴/۰۸/۱۹
تاریخ پذیرش: ۱۴۰۵/۰۳/۱۴

چکیده

تحول فناورانه هوش مصنوعی، نه تنها نظام سنتی مسئولیت مدنی را با بحران انتساب تقصیر و علیت روبه‌رو کرده، بلکه بنیادهای گسترده‌تری از حقوق شهروندی مانند حق بر توضیح‌پذیری، حق اعتراض به تصمیمات خودکار و دسترسی مؤثر به عدالت را نیز دچار چالش ساخته است. در محیط خودتطبیق الگوریتمی، شهروندان در تعامل با سامانه‌هایی نظیر ارزیابی ریسک در استخدام، اعطای وام، خدمات شهری (جریمه‌های خودکار، تخصیص مسکن عمومی) و دادرسی ماشینی با پدیده «جعبه سیاه» روبه‌رو هستند که بازسازی منطق تصمیم و احقاق حق را ناممکن می‌سازد. هدف این پژوهش که با روش توصیفی-تحلیلی انجام شده، گذار از رویکرد صرفاً نظری به ارائه مدلی عملیاتی بر پایه «مسئولیت مدنی مشروط به شفافیت اجرایی» است تا ضمن حفظ نوآوری فناورانه، ابعاد گوناگون حقوق شهروندی (از جمله جبران خسارت، رفع تبعیض الگوریتمی و نظارت قضایی) تأمین گردد. یافته‌ها نشان می‌دهد تحقق عدالت در عصر خودیادگیری، مستلزم پیوند شاخص‌های فنی شفافیت و قابلیت بازبینی داده با معیار تقصیر است؛ رویکردی که حکمرانی مبتنی بر قابلیت ردیابی را جایگزین مسئولیت محض خطا می‌کند. مدل پیشنهادی با الزام به مستندسازی آزمون‌های استحکام و انصاف و انتقال بار اثبات در صورت امتناع از ارائه گزارش‌های فنی، زمینه اعمال مسئولیت پیش‌دستانه را فراهم می‌آورد. این مدل با ادغام قابلیت توضیح‌پذیری به‌عنوان شاخص جدید مراقبت متعارف، زیربنای حقوقی لازم برای اعتمادسازی در تعاملات شهروند و سیستم‌های خودمختار را می‌سازد.

واژه‌های کلیدی

حقوق شهروندی، هوش مصنوعی، حقوق خصوصی، مسئولیت مدنی، شفافیت الگوریتمی.

استناد به این مقاله:
جعفری بهزادکلائی، محسن و رهبر، مونا. چالش‌های به‌کارگیری هوش مصنوعی در حوزه حقوق شهروندی با تأکید بر حقوق خصوصی. مجله حقوق شهری، ۲(۱)، ۱-۱۴.

حق انتشار این مستند، متعلق به نویسندگان آن است. ۱۴۰۵ ©. ناشر این مقاله، دانشگاه پیام نور است.
این مقاله تحت گواهی زیر منتشرشده و هر نوع استفاده غیرتجاری از آن مشروط بر استناد صحیح به مقاله و با رعایت شرایط مندرج در آدرس زیر مجاز است.
Creative Commons Attribution-NonCommercial 4.0 International license (https://creativecommons.org/licenses/by-nc/4.0/)

مقدمه

تحول بنیادین در مدل‌های محاسباتی هوش مصنوعی، هوشمندسازی علّت را دستخوش دگرگونی کرده است. در گذشته، نظام‌های رایانه‌ای بر منطق دستوری و زنجیره‌ای بنا می‌شدند؛ اما ظهور معماری‌های مبتنی بر یادگیری عمیق و شبکه‌های عصبی، رویکرد محاسباتی را از مدل‌های صلب به الگوهای استنتاجی آماری سوق داده است (Knight, 2017:54). این گذار فنی، چالش‌های جدّی برای حقوق خصوصی، به‌ویژه در حوزه مسئولیت مدنی، برمی‌انگیزد. سیستم‌های پیشرفته‌ای مانند مدل‌های ترانسفورمری یا شبکه‌های مولد تخاصمی[1]، نه بر رابطه علّی خطی، بلکه بر توزیع وزن‌ها و پایگاه‌ها در فضای چندبعدی تکیه دارند. این ساختار فنی، امکان بازسازی مسیر تصمیم‌گیری را به شدت کاهش داده و پدیده‌ای موسوم به «کِدِری الگوریتمی»[2] یا «جعبه سیاه»[3]را پدید آورده است (Burrell, 2016:3). در نتیجه، در دعاوی مدنی مربوط به خسارات ناشی از تصمیمات هوش مصنوعی، اصل سنتی قابلیت انتساب تضعیف می‌شود؛ زیرا انحرافی جزئی در لایه‌های میانی می‌تواند پیامدهای تبعیض‌آمیز یا زیان‌بار ایجاد کند، بدون آنکه بتوان آن را مستقیماً به ورودی خاصی منتسب ساخت. هوش مصنوعی دیگر محصول ثابت نیست، بلکه سیستمی زنده و تطبیق‌پذیر است که به‌طور پویا از داده‌های محیطی یاد می‌گیرد. این ویژگی، محدود کردن مسئولیت به لحظه تولید را ناممکن می‌سازد و مدل‌های سنتی مسئولیت محصول را ناکافی جلوه می‌دهد (Barfield, 2020:112). از سوی دیگر، الگوریتم‌ها قادرند سوگیری‌های تاریخی داده‌ها را بازتولید و در سطح سیستماتیک تشدید کنند؛ امری که می‌تواند منجر به بی‌عدالتی در حوزه‌هایی چون استخدام، وام‌دهی، بیمه و ارزیابی ریسک گردد (O'Neil, 2016:21). از این‌رو، حرکت از رویکرد «پاسخگویی پس از وقوع»[4] به «حکمرانی پیش از وقوع»[5] ضرورتی بنیادین و اجتناب‌ناپذیر در عصر هوش مصنوعی است.

پرسش محوری این پژوهش از همین نقطه آغاز می‌شود که چگونه می‌توان چارچوبی از مسئولیت مدنی طراحی کرد، به‌گونه‌ای که تعادل میان پاسخ‌گویی و آزادی نوآوری حفظ شود. تحقق چنین چارچوبی مستلزم توسعه سازوکارهایی برای قابلیت ردیابی[6]، تفسیرپذیری[7] و ارزیابی شفافیت در تصمیمات الگوریتمی است.

در پاسخ به این پرسش، پژوهش حاضر مدل «مسئولیت مدنی مشروط بر شفافیت اجرایی»[8] را پیش می‌نهد؛ مدلی که در دو سطح متمایز عمل می‌کند: اول، الزام در سطح طراحی، که توسعه‌دهندگان را موظف به مستندسازی اجرای آزمون‌های استحکام و انصاف پیش از استقرار می‌نماید؛ عدم ارائه این مستندات خود مبنای مسئولیت پیش‌دستانه است. دوم، الزام در سطح اجرا، که سلب مسئولیت را منوط به ارائه گزارش تحلیلی جامع از سوی نهاد مسئول (بهره‌بردار/ توسعه‌دهنده) می‌سازد تا انطباق تصمیم با اصول و فقدان سوگیری اثبات شود. در صورت کوتاهی در این شفافیت، بار اثبات به صورت خودکار به سمت نهاد مسئول منحرف می‌گردد. این مکانیزم، مسئولیت مدنی را تابعی از علیت و توضیح‌پذیری می‌داند:

$$Liability\ \alpha\ f(Causality) \times g(Explainability)$$

بر این اساس، قابلیت توضیح‌پذیری (g) از یک عامل فرعی به مؤلفه‌ای حیاتی تبدیل شده است که نادیده گرفتن آن، صرف‌نظر از وجود رابطه‌علّی، اعمال مسئولیت را تسهیل می‌کند و حکمرانی الگوریتمی را متحول می‌سازد.

پیشینه تحقیق

با توجه به تحولات شتابان و بی‌سابقه در حوزه فناوری‌های نوظهور، به‌ویژه ظهور و گسترش روزافزون هوش مصنوعی مولد و سیستم‌های خودمختار، ادبیات پژوهشی حقوقی در سال‌های اخیر دستخوش تغییرات بنیادینی شده است. گذار از مباحث انتزاعی و صرفاً حقوقی به سمت مطالعات کاربردی، میان‌رشته‌ای و دادرسی‌محور، ضرورت واکاوی دقیق پژوهش‌های تجربی را دوچندان می‌کند. در این راستا، بررسی و ارزیابی انتقادی متون علمی معاصر در سطح داخلی و بین‌المللی نشان می‌دهد که ناکارآمدی سازوکارهای سنتی مسئولیت مدنی در مواجهه با فناوری‌های نوین، یک گزاره مورد اجماع علمی است و تلاش‌های

1. Generative Adversarial Networks (GANs)

۲. Opacity یا کدری الگوریتمی، اصطلاحی است که به فقدان شفافیت و قابلیت تبیین در فرآیندهای تصمیم‌گیری هوش مصنوعی (به‌ویژه شبکه‌های عصبی عمیق) اشاره دارد. در این مدل‌ها (معروف به جعبه سیاه)، مسیر رسیدن به نتیجه برای انسان قابل ردیابی و استدلال‌سازی نیست و همین امر، اعمال حق توضیح‌پذیری (Explainability) دشوار می‌سازد.

3. Black-Box

4. Ex-post Liability
5. Ex-ante Governance
6. Interpretability
7. Traceability
8. Civil Liability Conditional upon Operational Transparency

پژوهشی معاصر متمرکز بر طراحی مدل‌های مبتنی بر شفافیت الگوریتمی، حکمرانی داده‌ها و اصلاح قواعد اثبات دعوی در دادگاه‌ها معطوف گشته است .در نظام حقوقی ایران، با وجود خلأ قانونی در خصوص مقررات اختصاصی هوش مصنوعی و عدم تصویب قانون جامع در این حوزه تا سال ۱۴۰۴، تولیدات علمی سال‌های اخیر با رویکردی انتقادی و تحلیلی، کوشیده‌اند تا با تکیه بر قواعد عمومی مسئولیت مدنی، اصول فقهی و رویه قضایی، راهکارهایی برای برون‌رفت از بن‌بست‌های دادرسی و چالش‌های اثبات دعوی ارائه دهند. در پژوهشی جامع با عنوان «بررسی چالش‌های اثبات رابطه سببیت در مسئولیت مدنی ناشی از هوش مصنوعی مولد»، (صادقی و همکاران، ۱۴۰۲: ۴۵) به واکاوی ابعاد حقوقی و فنی عدم قطعیت در خروجی‌های مدل‌های زبانی بزرگ اهتمام ورزیده‌اند. این پژوهش با بهره‌گیری از روش تحلیلی–توصیفی و با مطالعه موردی رویه قضایی در دعاوی فناوری، استدلال می‌نماید که قواعد سنتی اثبات تقصیر در حقوق مدنی ایران، که بر مبنای علم قاضی و امارات متعارف بنا شده‌اند، در مواجهه با پیچیدگی‌های شبکه‌های عصبی عمیق و ماهیت احتمالاتی تصمیم‌گیری ماشین، دچار فروپاشی می‌شوند. نویسندگان با نقد نظریه مسئولیت مبتنی بر تقصیر انسانی، پیشنهاد می‌کنند که با استناد به اصل خطر غیرمتعارف و بهره‌گیری از قواعد مسئولیت ناشی از عمل غیر مستقیم، باید در بار اثبات و نحوه توزیع آن بازنگری اساسی صورت گیرد. با این وجود، این مطالعه اگرچه به درستی تشخیص داده است که ابزارهای اثباتی سنتی و متعارف در دادگاه‌های ایران کارآمد نیستند، اما فاقد طرح دقیق و اجرایی برای استخراج اطلاعات از جعبه سیاه الگوریتم‌ها و تبدیل داده‌های فنی به یک دلیل قانونی معتبر و قابل استناد در دادگاه است. در ادامه، مطالعه‌ای توسط (رحمانیان، ۱۴۰۳) تحت عنوان «الزامات شفافیت الگوریتمی در حقوق ایران: ضرورت یا آرمان» به انجام رسیده است که به طور اختصاصی به چالش‌های حکمرانی داده‌ها و نظارت بر الگوریتم‌های تصمیم‌ساز پرداخته است. محقق با تبیین اصول سیاست‌گذاری کلی در فضای مجازی و اسناد بالادستی، استدلال می‌کند که فقدان ضمانت اجرایی قانونی برای ممیزی الگوریتم و عدم الزام شرکت‌ها به افشای ساختار تصمیم‌گیری، مانع ساختاری و جدی در برابر تحقق عدالت دیجیتال ایجاد می‌کند. مدل پیشنهادی این پژوهش مبتنی بر نهادینه کردن سازوکار گزارش‌دهی دوره‌ای عملکرد است که شرکت‌های فناوری را ملزم به ارائه گزارش‌های فنی و شفاف‌سازی الگوریتم‌های خود می‌نماید. اگرچه این پژوهش از نظر موضوعی و جهت‌گیری کلی با پژوهش حاضر هم‌راستا است، اما تفاوت اساسی آن در این است که ایشان بیشتر بر ابعاد نظری حکمرانی و سیاست‌گذاری عمومی تمرکز داشته و سازوکار حقوقی انتقال بار اثبات در صورت امتناع متهم از ارائه گزارش، که هسته مرکزی و عملیاتی مدل پیشنهادی پژوهش حاضر است، مورد تفحص و طراحی دقیق قرار نگرفته است. همچنین، رساله دکتری (موسوی، ۱۴۰۳) در دانشگاه تهران با موضوع «مسئولیت مدنی توسعه‌دهندگان سیستم‌های خودران»، با روش حقوق تطبیقی و با مقایسه رویه قضایی ایران با کشورهای پیشرو، به بررسی تعارض میان قواعد مسئولیت قراردادی و خارج از قرارداد در حوادث ناشی از رانندگی خودران پرداخته است. این پژوهشگر نشان می‌دهد که در حقوق ایران، اثبات تقصیر توسعه‌دهنده نرم‌افزار یا سیستم هوشمند بدون دسترسی به لاگ‌های سیستمی، داده‌های خام و سوابق ثبت‌شده تصمیمات ماشین، ناممکن است و عملاً زیان‌دیده را در برابر شرکت‌های بزرگ فناوری ناتوان می‌گذارد. وی پیشنهاد می‌کند که قوه قضاییه باید با تجهیز کارشناسان رسمی به ابزارهای پیشرفته تفسیر مدل‌های هوش مصنوعی و ایجاد شعب تخصصی، این خلأ را پر نماید. این رساله بر ضرورت ابزارهای فنی برای کشف رابطه علیت تأکید دارد، اما فاقد یک مدل حقوقی یکپارچه و منسجم است که ابزارهای مذکور را به یک تعهد قانونی تبدیل کرده و ضمانت اجرای قضایی قوی و پیش‌بینانه برای آن پیش‌بینی نماید.

در عرصه بین‌المللی، سال‌های ۲۰۲۳ و ۲۰۲۴ شاهد اوج‌گیری ادبیات علمی پیرامون اجرای قانون هوش مصنوعی اتحادیه اروپا و دستورالعمل مسئولیت مدنی هوش مصنوعی بوده است. پژوهش‌های اخیر در این سطح، به جای پرداختن به مباحث کلی و فلسفی نظریه شخصیت حقوقی ربات‌ها، بر جزئیات فنیِ استانداردهای شفافیت، نحوه اعمال آن در رویه قضایی و تعادل میان حفظ نوآوری و تضمین مسئولیت متمرکز شده‌اند. یکی از مطالعات بنیادین در این حوزه، مقاله (Busch et al., 2024) در مجله معتبر حقوق فناوری اطلاعات با عنوان «هزینه‌های شفافیت: تعادل میان نوآوری و مسئولیت در عصر هوش مصنوعی مولد» است. این محققان با بهره‌گیری از روش‌های پژوهش کمی و کیفی و با بررسی داده‌های اقتصادی شرکت‌های فناوری، اثرات اقتصادی و حقوقی الزامات افشای کد منبع را بر شرکت‌های فناوری در اروپا سنجیده‌اند. ایشان در این مقاله استدلال می‌نمایند که الزامات سخت‌گیرانه قانونی مبنی بر افشای کامل کد منبع برای اثبات تقصیر در دادگاه، می‌تواند به عنوان یک مانع تنظیم‌گری عمل کرده و با ایجاد نگرانی از

سوءاستفاده رقبا، سرعت نوآوری و سرمایه‌گذاری در حوزه هوش مصنوعی را کاهش دهد. آن‌ها مدلی جایگزین را پیشنهاد می‌کنند که در آن توسعه‌دهندگان ملزم به ارائه مسیر تصمیم‌گیری در سطحی انتزاعی و قابل فهم برای قاضی هستند، بدون آنکه اسرار تجاری و مالکیت معنوی آن‌ها به خطر افتد. این پژوهش از نظر فنی و با تاکید بر حفظ تعادل، با پژوهش حاضر همخوانی دارد، اما تمرکز آن بیشتر بر تحلیل اقتصادی–سیاسی و سیاست‌گذاری است تا ارائه یک سازوکار دادرسی‌پذیر و دقیق برای تغییر بار اثبات در دعاوی خصوصی. در مقاله‌ای دیگر با عنوان «فراتر از جعبه سیاه: کاربرد هوش مصنوعی قابل تفسیر در اثبات مسئولیت مدنی»، (Clemen, Wagner, 2024) به بررسی نقش فناوری‌های نوین هوش مصنوعی قابل تفسیر در فرآیند دادرسی و کشف حقیقت پرداخته‌اند. نویسندگان نشان می‌دهند که ابزارهای جدید «هوش مصنوعی قابل تفسیر»[1] قادرند رابطه میان داده ورودی و تصمیم خروجی را به‌گونه‌ای بازنمایی و ترسیم کنند که برای قاضی و کارشناس دادگاه قابل فهم و استناد باشد. با این وجود، این محققان با رویکردی انتقادی هشدار می‌دهند که خروجی این ابزارهای تفسیری نیز ذاتاً دارای عدم قطعیت و خطای احتمالی هستند و نباید بدون اعتبارسنجی قضایی دقیق، صرفاً مبنای صدور حکم قرار گیرند. آن‌ها پیشنهاد می‌کنند که باید یک استاندارد طلایی برای توضیح‌پذیری تعریف و توسط قانون‌گذار تدوین شود تا از سوءاستفاده‌های احتمالی جلوگیری گردد. این پژوهش دقیقاً همان چالش فنی و حقوقی را هدف گرفته که مدل پیشنهادی پژوهش حاضر با تعریف استانداردهای فنی شفافیت اجرایی و الزام به ارائه گزارش‌های ممیزی، سعی در حل آن دارد. در نهایت، پژوهش (Grundmann, 2024) تحت عنوان «دستورالعمل مسئولیت مدنی هوش مصنوعی اتحادیه اروپا: انقلابی در اثبات تقصیر»، به تحلیل دقیق ابزارهای اثباتی جدید در حقوق اروپا و آثار آن بر رویه قضایی پرداخته است. این نویسنده تبیین می‌کند که دستورالعمل جدید با ایجاد پیش‌فرض‌های علیت، سعی دارد بار اثبات را از دوش شاکی برداشته و فرآیند دادرسی را تسهیل نماید. با این‌حال، این مطالعه بیان می‌کند که این پیش‌فرض‌ها تنها در صورت وجود شواهد کافی و همکاری فعال طرف خوانده در افشای اطلاعات قابل اعمال هستند. این یافته‌ها نشان می‌دهد که هنوز خلأیی جدی در تعیین تکلیف و وضعیت حقوقی زیان‌دیده در صورت عدم همکاری طرف مقابل وجود دارد که مدل پیشنهادی پژوهش حاضر با پیشنهاد انتقال خودکار و قطعی بار اثبات در صورت عدم ارائه گزارش شفافیت، قصد پر کردن این خلأ و تضمین حقوق زیان‌دیدگان را دارد.

۱. مفهوم هوش مصنوعی

هوش مصنوعی در معنای بنیادی خود، به تلاش مهندسی و علمی برای ساخت ماشین‌هایی اشاره دارد که قابلیت‌هایی مشابه یا برتر از هوش انسانی را در زمینه‌هایی نظیر یادگیری، استدلال، ادراک و تصمیم‌گیری از خود نشان دهند (Floridi, 2022:15) . این حوزه با تعریف کلاسیک و تاریخی جان مک‌کارتی در سال ۱۹۵۶ متولد شد که هوش مصنوعی را به عنوان علم و مهندسی ساخت ماشین‌های هوشمند تعریف کرد. در آن دوران اولیه، تمرکز اصلی بر رویکرد نمادین بود؛ بستری که در آن هوش مترادف با دست‌ورزی منطقی نمادها و اجرای قواعد صریح «اگر–آنگاه»[2] تلقی می‌شد؛ رویکردی که به واسطه ساختار قاعده‌محورش، از شفافیت ذاتی برخوردار بود؛ زیرا هر خروجی ماشینی مستقیماً به زنجیره‌ای از دستورات قابل ردیابی فروکاسته می‌شد. این بنیان، یک آرمان برای محاسبات محسوب می‌شد که در آن کنترل کامل و فهم مسیر استدلال همواره در دسترس بود (OECD, 2024:42).

با این‌حال، تحولات محاسباتی در قرن بیست و یکم، به‌ویژه ظهور شبکه‌های عصبی عمیق، تعریف هوش مصنوعی را دچار یک دگرگونی بنیادین کرد. این گذار، هوش را از پیروی از قواعد صریح برنامه‌نویسی شده به استنتاج الگوهای نهفته در داده‌های تجربی انبوه سوق داد. در این معماری نوین که اساس هوش مصنوعی مدرن را شکل می‌دهد، ماشین از طریق فرآیند یادگیری ماشین، وزن‌ها و روابط پیچیده‌ای را در فضای چند بعدی ریاضیاتی می‌آموزد که این دانش استخراج‌شده، دیگر به سادگی قابل تفسیر به زبان منطقی انسانی نیست.

این انتقال از سیستم‌های مبتنی بر قصد صریح به سیستم‌های مبتنی بر استنتاج آماری تطبیقی، ماهیت هوش مصنوعی را از یک ابزار دستوری به یک عامل خودآموخته و پویا تغییر داده است. این عامل، نه تنها قابلیت‌های بی‌سابقه‌ای در پردازش زبان و بینایی پدید آورده، بلکه چالشی مفهومی در مورد چگونگی تعریف هوشمندی و همچنین مسئولیت این سیستم‌های پیچیده و نیمه‌خودمختار ایجاد کرده است؛ چالشی که نیازمند چارچوب‌های

۱. XAI مخفف eXplainable Artificial Intelligence است؛ به مجموعه روش‌ها و ابزارهایی گفته می‌شود که خروجی مدل‌های هوش مصنوعی (به‌ویژه مدل‌های جعبه سیاه) را برای انسان قابل درک و تفسیر می‌سازد.

۲. if–then؛ منظور قواعد شرطی صریح و قابل ردیابی توسط انسان است که در آن، ماشین صرفاً طبق دستورالعمل‌های از پیش تعیین شده عمل می‌کند.

نوین حکمرانی فراتر از الزامات صرفاً فنی مک‌کارتی است (Floridi, 2022:28; OECD, 2024:45).

۲. حقوق شهروندی در نظم داده‌محور

تحول حکمرانی در عصر هوش مصنوعی، مرز میان اقتدار دولتی و حقوق فردی را دگرگون کرده است. هنگامی که تصمیم‌گیری در بستر داده‌های کلان و مدل‌های یادگیری عمیق شکل می‌گیرد، مفهوم حاکمیت دیگر محدود به نهاد سیاسی نیست؛ بلکه به عرصه تعامل میان شهروند و الگوریتم نیز گسترش می‌یابد. در این شرایط، حق مشارکت داده‌ای به مثابه یکی از جدیدترین صور حقوق شهروندی پدیدار می‌شود؛ یعنی حق فرد برای حضور فعال در روند الگوریتمی که بر زندگی او اثر می‌گذارد. این حق نه تنها بر رضایت داده‌ای، بلکه بر شفافیت و امکان بازسازی مسیر تصمیم استوار است؛ به عبارت دیگر، شهروند حق دارد بداند چگونه داده‌هایش به قضاوت، تخصیص یا توصیه بدل می‌شوند. رعایت این حق در سطح حکمرانی، پیش‌شرط اعتماد اجتماعی به سامانه‌های تصمیم‌گیری خودکار است و در غیاب آن، فرآیند هوش مصنوعی از مشروعیت مدنی تهی می‌شود (Wachter et al., 2019:735).

در سطحی عمیق‌تر، حقوق بنیادین بشر در عصر داده‌محور نیازمند بازخوانی است. ظهور الگوریتم‌های بازسازی‌کننده هویت، چالش تازه‌ای پدید آورده است؛ در این فضا باید از حق بر عدم بازسازی داده‌ای سخن گفت؛ حقی که تمامیت هویت انسانی را از برساخت‌های آماری محافظت می‌کند (Barocas et al., 2019:671). رعایت این حق، مستلزم پیروی از اصول ضد تبعیض الگوریتمی و طراحی چارچوب‌هایی برای توازن میان نوآوری و صیانت هویتی است (نجفی، ۱۴۰۳: ۱۰). حاکمیت داده‌محور موظف است تضمین کند که هیچ فردی به واسطه الگوهای آماری ناشی از داده‌های دیگران، در معرض کاهش منزلت یا تصمیم‌های تبعیض‌آمیز قرار نگیرد. طراحی معماری حکمرانی الگوریتمی باید بر پایه اصل شفافیت انتشاری استوار باشد. شفافیت در این معنا، صرفا افشای کد یا داده نیست، بلکه فرآیند پیوستگی از مستندسازی، گزارش عمومی و ارزیابی میان‌دوره‌ای است (Floridi et al., 2022:1131). نهادهای بهره‌بردار هوش مصنوعی باید به ارائه گزارش‌های فنی منظم درباره عملکرد مدل‌ها و میزان انطباق آن‌ها با استانداردهای عدالت آماری مکلف باشند (Raji et al., 2020:10).

تنها از رهگذر چنین معماری‌ای است که اخلاق داده و اعتماد عمومی می‌توانند در ساختار حکمرانی نهادینه شوند. راه برون‌رفت از بن‌بست‌های نظری، نه در انکار قدرت الگوریتم، بلکه در تعبیه سازوکارهای کنترلی فعال در دل معماری حکمرانی نهفته است. حقوق شهروندی در این منظومه، دیگر صرفا مجموعه‌ای از امتیازات منفعل نیست؛ بلکه به مجموعه‌ای از تعهدات فنی قابل ممیزی تبدیل می‌شود که باید پیش از هر عملیاتی در کد، داده و مستندات مدل، تثبیت گردند. این تحول، نهادهای نظارتی مستقل و تخصصی را ناگزیر می‌سازد که به کد و داده دسترسی داشته باشند تا بتوانند مسئولیت مطلق فنی را نه به عنوان یک مجازات پس از وقوع، بلکه به عنوان یک ضمانت اجرایی ساختاری در طول چرخه حیات هوش مصنوعی، احراز و اعمال نمایند (نوذری، ۱۴۰۰: ۱۲). بدین‌ترتیب، حاکمیت دیجیتال با تبدیل تعهد به ساختار، اعتماد شهروندان را به جای تکیه بر حسن نیت، بر استحکام طراحی فنی بنا می‌نهد.

۳. حریم خصوصی و خودمختاری داده‌ها

حق حریم خصوصی در عصر هوش مصنوعی، دیگر نه یک حق صرفاً سلبی برای سکوت داده‌ها یا عدم افشای اطلاعات، بلکه یک مفهوم فعال، پویا و ساختاری تحت عنوان خودمختاری داده‌ای است که ابعاد نوینی به حقوق شهروندی افزوده است. این رویکرد نوین، حق انسان را در تعیین سرنوشت، نحوه گردش، هدف نهایی استفاده و حتی امکان اعمال وتو بر اطلاعات شخصی خود در بستر پیچیده الگوریتم‌های یادگیری عمیق مستقر می‌سازد. چالش اساسی زمانی پدیدار می‌شود که سیستم‌های هوشمند با استخراج الگوهای آماری بسیار پیچیده از انبوه داده‌های خام، اقدام به بازآفرینی هویت، پیش‌بینی دقیق رفتارهای آتی، استنتاج تمایلات پنهان فردی یا حتی تحلیل روان‌شناختی می‌کنند. این فرآیند که اغلب بدون آگاهی مستقیم فرد، خارج از کنترل او و در پشت پرده جعبه‌های سیاه الگوریتمی رخ می‌دهد، ماهیت حریم خصوصی را دگرگون کرده است(Floridi, 2022:48). در این شرایط، فقدان کنترل کاربر بر نحوه پردازش، استنتاج و به‌کارگیری نهایی داده، مصداق بارز نقض مالکیت اطلاعاتی و محدودسازی جدی آزادی‌های فردی است؛ چرا که قواعد سنتی حقوق مدنی که بر مبنای دعوای مستقیم و ورود فیزیکی بنا شده‌اند، برای پاسخ‌گویی به این نوع آسیب‌های نامحسوس و ناملموس ناکارآمدند(Nissenbaum, 2023:115).

در نظام حقوقی حاکم بر مسئولیت مدنی، رویکردهای کلاسیک که نیازمند اثبات تقصیر مستقیم، ورود مستقیم و وقوع آسیب فیزیکی یا مادی محسوس هستند، در مواجهه با خسارات ناشی از

پردازش داده‌های نامحسوس دچار فروپاشی و کارایی مطلق می‌شوند. آسیب‌های ناشی از نقض حریم خصوصی در عصر دیجیتال، اغلب روانی، معنوی، اجتماعی یا مربوط به از دست رفتن فرصت‌ها و تبعیض الگوریتمی هستند. رابطه سببیت میان عملیات پردازش خودکار ماشین و زیان وارد شده، به دلیل پیچیدگی فنی شبکه‌های عصبی، عدم شفافیت درونی و ماهیت احتمالاتی و غیرقطعی تصمیم‌گیری، در دادگاه‌ها غیرقابل اثبات باقی می‌ماند (Solove, 2021:92).

بنابراین، صرفاً جرم‌انگاری افشای داده یا ورود فیزیکی به حریم شخصی برای پاسداری از کرامت انسانی کفایت نمی‌کند؛ بلکه باید استفاده ابزاری، استنتاجی، استخراجی و بازآفرینی از داده‌ها که منجر به آسیب می‌شود، نیز در قلمرو مسئولیت مدنی قابل پیگیری و جبران باشد (Barocas, Selbst, 2022:155). در پاسخ به این خلأ قانونی و برای پیشگیری از نقض گسترده حقوق شهروندی در فضای دیجیتال، ضروری است سازوکارهای پاسخ‌گویی و مسئولیت‌پذیری به‌طور اساسی بازتعریف شوند. رویکرد پیشنهادی این پژوهش با تغییر ماهیت شفافیت از یک هنجار اخلاقی صرف و توصیه‌ای نرم، به یک الزام حقوقی، بنیان‌های مسئولیت مدنی را دگرگون می‌سازد. در این مدل، شفافیت دیگر یک انتخاب اختیاری یا استراتژی بازاریابی برای شرکت‌های فناوری نیست، بلکه به یک شرط تعلیقی و پیش‌شرط انحصاری برای اعمال مسئولیت تبدیل شده است. بدین معنا که هرگاه بهره‌بردار سیستم، مستندات فنی دقیق، قابل فهم و قابل بازرسی را درباره آزمون‌های استحکام، معیارهای انصاف، اثرات اجتماعی و منطق تصمیم‌گیری در شبکه‌های عصبی ارائه ندهد، عملاً امکان دفاع از عملکرد الگوریتم را در دادگاه از دست می‌دهد. این سازوکار با ایجاد یک وضعیت حقوقی خاص، بار اثبات دشوار را به‌طور معکوس بر دوش بهره‌بردار می‌اندازد. چنانچه نهادهای متولی از افشای اطلاعات فنی امتناع کنند یا شاخص‌های تبیین‌پذیری و تفسیرپذیری را رعایت ننمایند، قرینه قانونی بر تقصیر آن‌ها معتبر شمرده شده و مسئولیت مدنی به صورت پیش‌دستانه و بدون نیاز به اثبات سنتی و دشوار توسط زیان‌دیده، بر عهده آن‌ها گذاشته می‌شود.

نقش این مدل در حفظ حریم خصوصی، تبدیل کردن ابهام فنی و پیچیدگی الگوریتمی به یک امتیاز حقوقی و ابزار قدرت برای صاحبان داده است. در واقع، مدل پیشنهادی با ابزارهای حقوقی قهری، دیوارهای نامرئی جعبه‌های سیاه الگوریتمی را فرو می‌ریزد و اجازه نمی‌دهد پیچیدگی علمی به پناهگاهی برای فرار از مسئولیت تبدیل شود. این مدل به زیان‌دیده این قدرت را می‌دهد که در صورت بروز خدشه‌ای به حریم خصوصی، به جای غرق شدن در پیچیدگی‌های فنی و بن‌بست‌های اثبات رابطه سببیت، بر عدم ارائه شفافیت توسط مسئول فنی سیستم تکیه کند. وقتی قانون بهره‌بردار را مجبور می‌کند تا نحوه به‌کارگیری داده‌ها، وزن‌دهی به متغیرها و منطق تصمیم‌گیری را شفاف سازد، عملاً کنترل اطلاعات و نظارت بر آن به دست صاحب اصلی آن بازمی‌گردد. این امر مانع از سوءاستفاده پنهان، تبعیض‌آمیز یا غیرمنصفانه از داده‌های شخصی می‌شود و تضمین می‌کند که هرگونه استفاده ابزاری یا بازآفرینی از هویت اشخاص، تحت نظارت دقیق قضایی و با رعایت استانداردهای انصاف و عدالت انجام شود. بدین ترتیب، این شیوه با ایجاد امکان نظارت قضایی مؤثر بر عملکرد الگوریتم‌ها، ضمن تضمین استقلال داده‌ای اشخاص و احیای حاکمیت انسان بر ماشین، زمینه را برای ایجاد توازن پایدار میان پیشرفت فناوری و تضمین عدالت مدنی فراهم می‌آورد.

۴. دادرسی ماشینی و مسئله پاسخ‌گویی الگوریتمی

تحول دادرسی در عصر هوش مصنوعی به سوی تفویض بخشی از ارزیابی قضایی به سامانه‌های تصمیم‌یار، مرز میان تشخیص و قضاوت را در هم می‌آمیزد. آنچه پیشتر به مثابه فرایند استدلال انسانی شناخته می‌شد، اکنون در قالب مدل‌های پیش‌بینی و تطبیق داده‌های قضایی بازتعریف می‌گردد. در این بستر، پرسش بنیادین آن است که مفهوم پاسخ‌گویی قضایی در غیاب اراده انسانی چگونه قابل تحقق می‌شود و آیا می‌توان میان خطای محاسباتی و تخلف قضایی تمایز حقوقی برقرار ساخت (Wachter et al., 2019:735).

دادرسی ماشینی بر مبنای الگوریتم‌های استنتاجی عمل می‌کند که نه قصد دارند و نه می‌توانند ادله را بفهمند؛ بلکه صرفاً الگوهای همزمانی میان داده‌ها را تشخیص می‌دهند. در نتیجه، مسئولیت تصمیم در چنین نظامی نه به قاضی انسانی، بلکه به زنجیره‌ای از توسعه‌دهنده، بهره‌بردار و سامانه تصمیم‌یار تعلق می‌گیرد. فقدان شفافیت در منطق آموزش مدل، امکان بازسازی مسیر استدلال قضایی را از میان می‌برد و زمینه بی‌اعتباری ادله تولیدشده را فراهم می‌سازد. به همین دلیل، پاسخ‌گویی الگوریتمی باید در قالب الزام به ثبت و ممیزی منطق تصمیم، در درون ساختار فنی مهندسی نهادینه گردد نه صرفاً به عنوان شرط بیرونی نظارت سنتی. در چنین رویکردی، انتقال بار اثبات از رفتار انسانی به سازوکار داده‌محور، مستلزم اصلاح بنیادی در قواعد دادرسی است. معیار صداقت قضایی باید جای خود را به شاخص‌های فنی نظیر قابلیت ردیابی، توضیح‌پذیری و ممیزی‌پذیری دهد؛ به‌گونه‌ای که امکان دفاع و

اعتراض مبتنی بر منطق فنی تصمیم فراهم شود، نه صرفاً بر اساس فرض درست بودن ابزار. اگر الگوریتم ناقص است، باید بتوان از درون ساختار داده‌ای آن، نه از بیرون، اعتراض حقوقی اقامه کرد (Floridi et al., 2022:1131).

این منطق تازه مستقیماً به حقوق بنیادین شهروندان در فضای قضایی دیجیتال مربوط می‌شود. اصولاً، قابلیت تفسیر خروجی‌های یک مدل پیچیده، مانند شبکه‌های عصبی عمیق، خود به چالشی فنی تبدیل شده است؛ زیرا اغلب توجیهات پس از وقوع جای استدلال پیشین را می‌گیرند. این امر، به‌ویژه در مواردی که مدل بر اساس معیارهای نامرتبط یا سوگیرانه آموزش دیده باشد، مشروعیت تصمیم را زیر سؤال می‌برد. این الزامات فنی، ریشه در تعهدات بین‌المللی در زمینه حقوق بشر دیجیتال دارند (OECD, 2024:42).

۴-۱ حق دسترسی مؤثر به عدالت[1]

اعلامیه حقوق بشر دیجیتال شورای اروپا (مصوب ۲۰۲۴) به عنوان نخستین سند بین‌المللی الزام‌آور در حوزه حقوق دیجیتال، ضمن تکیه بر مبانی تثبیت شده کنوانسیون اروپایی حقوق بشر (به‌ویژه حق دادرسی منصفانه، ماده ۶)[2]، در ماده ۵ خود تصریح می‌کند که حق دسترسی مؤثر به عدالت در زمینه‌های فناوری‌پایه، مشروط به امکان فهم و بررسی مبنای تصمیم‌گیری ماشینی است. این امر مستلزم آن است که منطق پشت تصمیم، حتی اگر پیچیده باشد، به‌گونه‌ای قابل استخراج و ارائه باشد که طرفین دعوی بتوانند بر آن ایراد وارد سازند(CoE, 2024:15). این رویکرد در واقع تفسیر نوین همان ضمانت‌های سنتی دادرسی منصفانه در عصر الگوریتم‌ها است.

۴-۲ حق اعتراض و توضیح[3]

ماده ۸ همان اعلامیه که متأثر از اصل حق استماع[4] (Right to be heard) در نظام حقوق بشر اروپایی است، تصریح می‌کند که حق اعتراض و مطالبه توضیح نسبت به تصمیمات خودکار، جزء لاینفک عدالت داده‌محور است. این حق، دکترین سنتی حق استماع را به حوزه دیجیتال تعمیم می‌دهد و نیازمند سیستمی است که بتواند به طور خودکار، مسیر محاسبه‌ای که منجر به یک نتیجه شده است را بازسازی کند. بدین ترتیب، اعلامیه ۲۰۲۴ نه یک سند نوپا، بلکه تدوین و تحکیم اصول پیشین در بستر فناوری‌های نوین است.

بنابراین، دادرسی ماشینی اگر فاقد سازوکار مستند اعتراض باشد؛ یعنی نتواند توضیح دهد که چرا یک ورودی مشخص منجر به یک خروجی شده است؛ از حیث حقوق بشر دیجیتال ناقص و غیرقابل اعتماد تلقی می‌گردد. نظام حقوقی باید به‌گونه‌ای بازطراحی شود که مسئولیت قضایی در دادرسی ماشینی نه براساس خطای انسانی (مانند سهل‌انگاری قاضی)، بلکه بر مبنای شفافیت و قابلیت ممیزی الگوریتم تعریف گردد. این رویکرد مستلزم توسعه چارچوب‌های قانونی برای اعتبار سنجی الگوریتمی و ایجاد نهادهای نظارتی تخصصی برای ممیزی داده‌ها و مدل‌ها است (Wachter et al., 2019:740). چنین بازآرایی، پاسخ‌گویی را از سطح نهاد انسانی به سطح ساختار فنی ارتقا می‌دهد و پیوند میان عدالت الگوریتمی و حق اعتراض را تحقق می‌بخشد؛ پیوندی که استمرار مشروعیت دادرسی دیجیتال در آن نهفته است. این تغییر رویکرد، دادرسی را از یک فعالیت مبتنی بر باوربه یک فرآیند مبتنی بر شواهد محاسباتی تبدیل می‌کند.

۵. مسئولیت هوش مصنوعی در مقام قاضی

ظهور سامانه‌های هوش مصنوعی در حوزه‌های تصمیم‌گیری حیاتی، به‌ویژه در مقام قضاوت، نقطه تلاقی بحرانی میان پیشرفت فناورانه و اصول بنیادین عدالت است. تفویض بخشی از اختیارات تصمیم‌گیری قضایی به الگوریتم‌ها، صرفاً یک ارتقاء ابزاری نیست، بلکه چالش ساختاری برای نظام مسئولیت مدنی و اصول دادرسی عادلانه ایجاد می‌کند.

۵-۱ بحران مفهومی علیت و تقصیر در تصمیم قضایی الگوریتمی

نظام سنتی مسئولیت مدنی، پاسخ‌گویی را مبتنی بر دو رکن اساسی می‌داند: فعل زیانبار ناشی از تقصیر و رابطه علیت میان آن فعل و خسارت وارده. در حوزه قضاوت مبتنی بر یادگیری عمیق، هر دوی این ارکان دچار فرسایش مفهومی جدی می‌شوند. در مواردی که هوش مصنوعی در مقام قاضی، براساس تعاملات داده‌ای پیچیده و لایه‌های خودتکاملی به رأی می‌رسد، انتساب تقصیر انسانی (به توسعه‌دهنده یا اپراتور) به شکلی که در حقوق سنتی مرسوم است، ناممکن می‌گردد. الگوریتم دیگر یک ابزار تابع نیست؛ بلکه یک

1. Effective Access to justice
2. European Convention on Human Rights (ECHR), Art. 6 (Right to a fair trial)

این ماده در رویه قضایی دیوان اروپایی حقوق بشر، مستلزم استدلال‌آور بودن آرای قضایی و امکان فهم منطق تصمیم توسط طرفین دعوی است.

3. Right to Contest and Explanation
4. Right to be heard

عامل یادگیرنده است. در نتیجه، اگر رأی صادره ناعادلانه باشد، فقدان اراده و قصد زیان‌بار در سویه الگوریتمی و گسست زمانی-فناورانه در سویه انسانی، اثبات عنصر تقصیر را به امری صوری تبدیل می‌کند (عسگری، ۱۴۰۳: ۴۵).

قضاوت، علمی است که نیازمند تفسیر ارزش‌ها، سنجش نیت طرفین و اعمال اصول عدالت و انصاف است. از منظر فلسفه حقوق، این اعمال نیازمند ادراک آگاهانه‌ای می‌باشد که فاقد آن است. صرف به‌کارگیری معیارهای آماری برای بهینه‌سازی خروجی (مانند کمترین نرخ خطا یا بیشترین کارایی)، جایگزین درک انسانی از عدل نمی‌شود. این امر، نظریه شخصیت الکترونیکی را در این حوزه کاملاً مردود می سازد؛ زیرا واگذاری اعمال حاکمیت به موجودیتی فاقد فهم اخلاقی، نقض کرامت انسانی است(Floridi, 2022:52).

۵-۲ تعارض با اصول دادرسی عادلانه و شفافیت به مثابه پیش‌شرط حق دفاع

مهم‌ترین پیامد تفویض قضاوت به هوش مصنوعی، تضعیف بنیادین حقوق شهروندی در فرآیند دادرسی است. حق دفاع یک شهروند در برابر رأی محکومیت، مستلزم آگاهی از منطق رسیدن به آن رأی است. وقتی استنتاج الگوریتمی، به‌ویژه در مدل‌های جعبه سیاه، غیرقابل تبیین باشد، حق دفاع عملاً نادیده گرفته می‌شود. این وضعیت مصداق عینی نقض کرامت انسانی است، چرا که فرد نمی‌تواند بفهمد چرا و براساس چه مبانی‌ای (که شاید آماری و تبعیض‌آمیز باشند) مورد قضاوت قرار گرفته است (محمدی و شفیعی، ۱۴۰۱: ۱۲۰).

براساس مدل مسئولیت مدنی مشروط بر شفافیت اجرایی، فقدان قابلیت تبیین و ردیابی تصمیم باید در این مقام، به مثابه قصور فنی بنیادین تلقی شود. اگر نتوانیم بفهمیم چرا الگوریتم رأی داده است، نمی‌توانیم حقوق دفاعی خود را به درستی اعمال کنیم.

۵-۳ چارچوب پاسخ‌گویی و مسئولیت مدنی مشروط بر قابلیت ممیزی

برای حفظ مشروعیت نظام عدالت در مواجهه با قاضی الگوریتمی، پاسخ‌گویی باید بر محوریت مستندسازی فنی استوار باشد. این امر مستلزم جایگزینی تقصیر سنتی با یک مسئولیت مبتنی بر عملکرد سیستم است. نهاد قضایی یا بهره‌بردار هوش مصنوعی باید ملکف به ارائه گزارشی کامل از مسیر استنتاجی الگوریتم گردد. این گزارش باید شامل داده‌های ورودی کلیدی، پارامترهای فعال‌شده و نحوه تعامل وزن‌های مدل در لحظه صدور رأی شود. از منظر فنی، این گزارش باید امکان بازسازی منطق تصمیم را فراهم آورد، حتی اگر مدل یک شبکه عصبی عمیق باشد (Kearns, et al., 2019:85).

در صورتی که سیستم قضایی نتواند این گزارش شفاف را ارائه دهد، باید فرض تقصیر به سود زیان‌دیده برقرار شود. این اصل، شبیه به رویکردی است که در نظام حقوقی ایران در خصوص عدم ارائه اسناد توسط طرف قوی‌تر (توسعه‌دهنده یا نهاد اجرایی) درنظر گرفته می‌شود. این امر با ماده ۳۰ قانون نظارت بر رفتار قضات هم‌راستا است، زیرا عدم توانایی در ارائه مستندات فنی دقیق، خود نوعی قصور در استفاده صحیح از ابزار تلقی شده و اعتبار فرآیند را خدشه‌دار می‌سازد. هرگونه شواهد مبنی بر اینکه رأی صادره تحت تأثیر سوگیری‌های آماری موجود در داده‌های آموزشی بوده و منجر به نقض اصل عدم تبعیض شده است، باید مستقیماً منجر به مسئولیت مدنی جبرانی برای نهاد استفاده کننده گردد. این رویکرد، جنبه‌ای از عدالت ترمیمی پیشگیرانه را محقق می‌سازد، چرا که نهادها را ملزم به ممیزی مستمر داده‌ها برای حذف تبعیض‌های پنهان می‌سازد (Barocas, Selbst, 2022:160).

تفویض کامل اختیار قضاوت به هوش مصنوعی، در فقدان استانداردهای شفافیت اجرایی غیرقابل قبول است. پاسخ‌گویی در این حوزه تنها با تحمیل یک تکلیف حقوقی جدید مبنی بر اثبات‌پذیری فنی محقق می‌شود که خود به معیار جدید مراقبت متعارف فنی در نظام مسئولیت مدنی بدل می‌گردد. مادامی که نتوانیم مسیر رسیدن به یک حکم را به وضوح و با رعایت اصول دادرسی عادلانه بازسازی کنیم، حکم صادر شده توسط ماشین، فاقد مشروعیت قضایی بوده و مسئولیت جبران خسارت آن باید به‌طور کامل بر عهده نهاد صادرکننده باشد.

۶. چالش‌های متقاضیان کار با هوش مصنوعی

تحول دیجیتال در فرایندهای استخدام، ظهور سامانه‌های هوش مصنوعی را به یکی از عوامل اصلی تصمیم‌گیری شغلی بدل ساخته است. این امر گرچه موجب تسریع و کارایی بیشتر نظام گزینش شده، اما از دید حقوقی، چالش‌های عمیقی را در زمینه شفافیت، تبعیض، مالکیت داده و مسئولیت کارفرما ایجاد کرده است. در این بخش تلاش می‌شود با تکیه بر نظریه مسئولیت مدنی مشروط بر شفافیت اجرایی، ابعاد حقوقی قضاوت الگوریتمی در فرایند استخدام واکاوی گردد.

در نظام‌های مبتنی بر هوش مصنوعی، تصمیم‌گیری استخدامی غالباً بر پایه الگوهای آماری و مدل‌های یادگیری عمیق انجام می‌شود. این فرایند اگر فاقد توضیح‌پذیری باشد، عملاً حق دفاع و

کرامت انسانی را نقض می‌کند؛ زیرا کارجو قادر نیست منطق تصمیم را درک کند. براساس اصل شفافیت مندرج در ماده ۵ قانون دسترسی آزاد به اطلاعات و ماده ۸ منشور حقوق شهروندی، هر تصمیمی که به حقوق کاری افراد اثر بگذارد باید قابل تبیین باشد. در فقدان این قابلیت مسئولیت مدنی کارفرما یا بهره‌بردار سامانه بر مبنای قصور فنی محرز است. از دید نظری، هر سامانه استخدامی یابد از مدل‌هایی استفاده کندکه مسیر تصمیم را مستند کنند؛ بدین معنا که داده‌های ورودی، الگوریتم فعال‌شده و معیارهای اولویت‌دهی قابل ثبت و بازبینی باشند. عدم رعایت این الزام، مصداق تقصیر نسبت به معیار مراقبت متعارف فنی در ماده ۳ قانون جرایم رایانه‌ای تلقی می‌شود (عسگری، ۱۴۰۳: ۳).

۶-۱ تبعیض آماری و عدالت الگوریتمی

اصل عدم تبعیض در استخدام، که در ماده ۳۸ قانون کار و ماده ۲ اعلامیه جهانی حقوق بشر تضمین شده، در بستر یادگیری ماشینی، چالش‌برانگیز است؛ زیرا داده‌های آموزشی ممکن است حامل سوگیری‌های اجتماعی باشند. به‌کارگیری سیستمی که الگوهای ناعادلانه موجود را بازتولید کند، نوعی قصور فنی محسوب می‌شود و مسئولیت مدنی کارفرما را نسبت به اثرات تصمیم غیرمنصفانه، فعال می‌سازد.

معیارهایی چون برابری جمعیتی[۱] و شانس برابر[۲] ابزار سنجش انصاف در مدل هستند؛ فقدان آن‌ها نقض عدالت استخدامی است. در صورت اثبات تبعیض آماری، جبران خسارت باید بر مبنای عدالت ترمیمی انجام گیرد؛ یعنی الزام فنی کارفرما به بازآموزی مدل و حذف سوگیری‌های موجود، مطابق ماده ۱ قانون مسئولیت مدنی (پورقلی، ۱۴۰۲: ۱۲۳). برای مثال، اگر مدل نرخ پذیرش یک گروه اقلیت (A) را به‌طور سیستماتیک پایین تر از نرخ هدف (P) نگه دارد:

$$P(Y = 1|A) < P(Y = 1|Non - A)$$

این وضعیت نشان‌دهنده نقض اصل تساوی آماری است و کارفرما مسئولیت جبران آن را برعهده دارد.

۶-۲ مالکیت داده و حق خودمختاری کارجو

حریم خصوصی در فرآیند استخدام با هوش مصنوعی، صرفاً حفاظت از داده‌ها نیست بلکه حق تعیین سرنوشت اطلاعات شخصی است. طبق ماده ۵۸ قانون تجارت الکترونیکی و ماده ۱۰۳ قانون کار، جمع‌آوری و پردازش خودکار داده‌های فردی بدون رضایت، مصداق نقض مالکیت اطلاعاتی است.

ابزارهای فنی نظیر یادگیری فدرال[۳] و حریم خصوصی[۴] تفاضلی می‌توانند سطح کنترل شخصی را حفظ کنند، اما تحقق کامل این حق نیازمند چارچوب حقوقی الزام‌آور است. هرگونه افشای غیرمجاز یا استفاده خودکار از داده حتی بدون آسیب فیزیکی، براساس ماده ۱ قانون مسئولیت مدنی قابل مطالبه خسارت است. این موضوع شامل داده‌های استنباطی که از تحلیل داده‌های اصلی به دست می‌آید نیز می‌شود.

۶-۳ مسئولیت کارفرما در برابر خطاهای الگوریتمی

در نظام سنتی، مسئولیت مدنی بر تقصیر انسانی استوار بود؛ حال آنکه در قضاوت الگوریتمی، عنصر تقصیر به شکل کلاسیک قابل احراز نیست (Selbst, 2024:120). برای حفظ مشروعیت تصمیم استخدامی، باید مسئولیت بر محور قابلیت ممیزی فنی استوار شود. مطابق اصل انتقال بار اثبات، چنانچه کارفرما نتواند منطق تصمیم الگوریتمی را مستند کند، فرض تقصیر علیه او برقرار می شود.

این قاعده با ماده ۳ قانون نظارت بر رفتار مستخدمان و اصول عدالت رویه‌ای مندرج در منشور حقوق شهروندی هماهنگ است. هر رأی یا انتخاب مبتنی بر مدل غیر شفاف که منجر به تحمیل زیان شغلی یا نقض برابری شود، به منزله قصور در مراقبت متعارف فنی تلقی و موجب مسئولیت جبرانی خواهد بود. این مسئولیت می‌تواند از نوع مسئولیت مطلق در قبال خسارات ناشی از فرآیندهای پرخطر نیز تلقی می‌شود، زیرا به‌کارگیری هوش مصنوعی در تصمیم‌گیری‌های حیاتی شغلی، سطح جدیدی از ریسک را به بازار کار وارد کرده است. قضاوت الگوریتمی در استخدام اگر فاقد شفافیت، عدالت و کنترل داده باشد، از دید حقوقی مشروعیت خود را ازدست می‌دهد. مسئولیت

---

۱. Demographic parity: معیاری که بر اساس آن، نرخ تصمیم‌گیری مطلوب (مانند پذیرش) در گروه‌های مختلف جمعیتی (مانند جنسیت، نژاد) باید برابر باشد؛ به عبارت دیگر، خروجی مدل نباید به متغیرهای محافظت‌شده وابسته باشد.

۲. Equalized Odds: معیاری که بر اساس آن، نرخ خطای مدل (هم خطای مثبت کاذب و هم خطای منفی کاذب) در گروه‌های مختلف جمعیتی باید یکسان باشد؛ به این ترتیب، مدل نباید برای یک گروه خاص، خطای سیستماتیک بیشتری داشته باشد.

۳. Federated Learning (یادگیری فدرال یا مشارکتی): رویکردی در یادگیری ماشین که در آن مدل بدون انتقال داده‌های خام کاربران به سرور مرکزی، تنها به‌روزرسانی‌های محلی وزن‌ها را مبادله می‌کند؛ بدین ترتیب حریم خصوصی داده‌های فردی حفظ می‌شود.

۴. Differential Privacy (حریم خصوصی تفاضلی): چارچوبی ریاضی برای تضمین حریم خصوصی که با افزودن نوفه (نویز) کنترل‌شده به خروجی پرس‌وجوهای آماری، امکان شناسایی یک فرد خاص را از پایگاه داده تقریباً غیرممکن می‌سازد.

کارفرما یا بهره‌بردار سیستم باید مبتنی بر قابلیت تبیین، ممیزی و اصلاح سوگیری‌ها باشد؛ زیرا این سازوکارها مصداق مراقبت فنی متعارف در حقوق نوین‌اند (Wachter et al,. 2019: 50). اجرای دقیق این اصول، تنها راه حفظ کرامت انسانی در تعامل انسان و ماشین در بازار کار آینده است.

۷. دیپ فیک و زوال حق بر اصالت دیجیتال

فناوری دیپ‌فیک[1] نه تنها یک چالش رسانه‌ای، بلکه یک چالش ساختاری در بنیادهای مسئولیت مدنی است. توانایی شبکه‌های مولد در خلق محتوای بصری–صوتی که از منظر حسی و شواهد فنی قابل تفکیک از واقعیت نیست، مستقیماً اصل تمامیت محتوایی[2] را هدف قرارمی‌دهد. این اصل، شالوده حفظ اعتبار و کرامت افراد در فضای دیجیتال محسوب می‌شود. در نظام سنتی، تقصیر به اراده و نیت انسان گره خورده است اما دیپ‌فیک، فعل زیان‌بار را از دایرهٔ قصد خارج کرده و آن را به یک خروجی آماری– الگوریتمی تبدیل می‌کند. در نتیجه، پیوند علیت حقوقی از عامل انسانی به نقص فنی در زنجیره اعتبار سنجی و تولید منتقل می‌گردد. تحلیل حقوقی دیگر نباید بر نیت تولید کننده متمرکز شود، بلکه باید بر کیفیت طراحی و تضمین اصالت فنی مدل معطوف گردد (حسینی، ۱۴۰۰: ۱۰۲).

برای دستیابی به پاسخگویی مؤثر، لازم است مسئولیت در تولید محتوای مصنوعی براساس تفکیک کارکردی بازیگران زنجیره صورت پذیرد؛ این رویکرد، محل وقوع نقص فنی را به عنوان معیار اصلی انتساب، می‌پذیرد.

۷-۱ مسئولیت توسعه‌دهنده و قصور در طراحی پیشگیرانه

توسعه‌دهنده‌ای که مدل‌های هوش مصنوعی را آموزش می‌دهد، مسئولیتی بنیادین برعهده دارد که قصور او نه در نتیجه نهایی، بلکه در عدم رعایت تعهد به مراقبت پیشگیرانه تعریف می‌شود. این تعهد مستلزم پیاده‌سازی ساختارهای فنی دقیقی است که در رأس آن‌ها واترمارکینگ غیرقابل حذف قرار دارد؛ امری که به معنای تعبیه یک امضای دیجیتالی نشان‌دهنده ماهیت مصنوعی محتوا است تا از طریق فرآیندهای بازتولید نیز قابل حذف نباشد. در کنار آن، ماژول‌های حفاظتی ضد جعل باید به‌گونه‌ای طراحی شوند که احتمال تولید محتوای تقلبیِ مضر را از مرحله آموزش کاهش دهند.

شکست در تعبیه این ساختارهای ایمنی، مصداق بارز قصور فنی در طراحی است و مسئولیت مدنی را مستقیماً بر عهده توسعه‌دهنده قرار می‌دهد؛ این همان تقصیر الگوریتمی است که نیازمند اثبات از سوی زیان‌دیده نیست. بنابراین، عدم توجه به این الزامات فنی در فرآیند طراحی و آموزش مدل، نه تنها یک نقص فنی محسوب می‌شود، بلکه پایه‌ای برای ایجاد مسئولیت مدنی مستقیم برای توسعه‌دهنده به شمار می‌رود و نشان می‌دهد که حفاظت پیشگیرانه در برابر سوءاستفاده‌های احتمالی، بخشی جدایی‌ناپذیر از فرآیند توسعه هوش مصنوعی است.

۷-۲ مسئولیت بهره‌بردار و پلتفرم ناشی از نقض تعهد ممیزی محتوایی

بهره‌بردار یا پلتفرم‌های انتشار، مسئولیت ثانویه‌ای دارند که از تعهد به ممیزی محتوایی نشأت می‌گیرد. پس از استقرار، وظیفه آن‌ها رصد محیط و به‌کارگیری ابزارهای تشخیصی به‌روز است. کوتاهی در ادغام و به‌کارگیری رابط‌های برنامه‌نویسی کاربردی[3] پیشرفته تشخیص دیپ‌فیک، نقض این تعهد محسوب می‌شود. در این سطح، مسئولیت بیشتر از جنس تخلف از وظیفه نظارتی در قبال انتشار محتوای گمراه‌کننده است و شباهت زیادی به مسئولیت ناشر در قبال نشر محتوای توهین‌آمیز دارد (فلاحی، ۱۴۰۱: ۴۲).

۷-۳ جبران خسارت از جبران ماهوی تا تضمین فنی اصالت

جبران خسارت ناشی از نقض حق بر اصالت دیجیتال، نیازمند رویکردی چندوجهی است که فراتر از پرداخت‌های مالی رایج باشد و بُعد فناورانه را دربرگیرد. در سطح جبران فنی، زیان‌دیده حق دارد ابطال الگوریتمی درخواست کند؛ امری که شامل الزام به ارائه سند ابطال رسمی یا توسعه یک الگوریتم اصلاحی توسط عامل فنی است تا تأثیر مخرب محتوای جعلی را خنثی سازد. همچنین در جبران ماهوی، در مواردی که زیان اعتباری غیرقابل بازگشت باشد (مانند آسیب به اعتبار حرفه‌ای)، بار اثبات خسارت باید به نفع زیان‌دیده تغییر یابد؛ به این معنا که اگر عامل فنی نتواند اثبات کند که از کلیه الزامات طراحی پیشگیرانه پیروی کرده، فرض تقصیر بر او برقرار می‌شود. این دو سطح جبران تضمین می‌کنند که کرامت انسانی در فضای دیجیتال، نه یک مفهوم صرفاً اخلاقی، بلکه یک حق قابل احراز و اجرا باقی بماند. پاسخ‌گویی در عصر دیپ‌فیک،

---

1. Deep fake
2. Content Integrity Axiom
۳. API (Application Programming Interface) یا «رابط برنامه‌نویسی کاربردی»، مجموعه‌ای از قواعد فنی است که امکان ارتباط، ادغام و تبادل داده میان دو سامانه نرم‌افزاری را فراهم می‌کند. در اینجا، منظور ابزاری است برای ادغام خدمات خودکار تشخیص دیپ‌فیک توسط پلتفرم‌ها.

مستلزم گذار کامل از چارچوب‌های سنتی به مسئولیت سازمانی–فناورانه است؛ مدلی که پاسخ‌گویی را در محل وقوع نقص فنی بنیادین، خواه در مرحله آموزش (توسعه‌دهنده) و خواه در مرحله اِعمال (بهره‌بردار)، تثبیت می‌کند. تنها از طریق پذیرش تفکیک کارکردی و الزامی کردن ابزارهای فنی مانند فراداده‌های ریشه‌دار[1] و مهر شفافیت دیجیتال[2] است که می‌توان مشروعیت دادرسی را در برابر محتوای مصنوعی حفظ کرد (Price, 2019: 45). این رویکرد، نه تنها عدالت ترمیمی را محقق می‌سازد، بلکه زمینه‌ساز حکمرانی هوش مصنوعی‌ای است که در آن فناوری در برابر اصول بنیادین حقوق شهروندی پاسخ‌گو باشد.

۸. مروری بر تجربیات جهانی

مسئولیت مدنی در هر قانونی، بازتاب درک جامعه از منشأ زیان و مسئولیت است. در نظام حقوقی ایران، این عدالت بر محوریت تقصیر انسانی بنا شده که وابسته به اراده و قصد است. سامانه‌های خودیادیگر هوش مصنوعی، این پیوند را در عمل شکسته است. در دنیایی که تصمیم‌ها براساس محاسبات آماری و بدون دخالت مستقیم انسان گرفته می‌شوند، اتکا به مفاهیم سنتی، دیگر پاسخ‌گو نیست. از این‌رو، این بخش بر الزام بازتعریف مبنای پاسخ‌گویی تأکید دارد: گذار از تقصیر به مسئولیت مطلق فنی؛ تنها راهکار برای حفظ عدالت در عصر الگوریتم.

نظام‌های پیشرو در مواجهه با این معضل، دو راهکار متضاد را به نمایش گذاشته‌اند که هیچ‌کدام برای ساختار حقوقی ایران کامل نیستند.

۸-۱ مدل آمریکا و تقصیر پنهان در پوشش محصول

رویکرد ایالات متحده تلاش می‌کند هوش مصنوعی را زیر چتر قوانین مسئولیت محصول جای دهد. این راهکار از ابتدا با مشکل روبه‌رو است زیرا محصول خودیادگیر ثابت نیست بلکه یک فرایند زنده است که دائماً در حال تغییر است. تفکیک لحظه عیب از تکامل الگوریتم عملاً غیرممکن است. این مدل مسئولیت را به یک نقطه ایستا در زنجیره تولید محدود می‌کند و از پاسخ‌گویی در برابر خطرات آینده فرار می‌کند (Matwyshyn, 2019: 210).

تعریف محصول معیوب در سیستمی که یادگیری آن در زمان استفاده رخ می‌دهد مبهم است. آیا عیب ناشی از داده‌های آموزشی است یا ناشی از به‌روزرسانی‌های حین بهره‌برداری. سیستم‌های سنتی محصول مانند یک خودرو در زمان فروش ثابت بودند اما در مورد هوش مصنوعی این ثبات وجود ندارد (Wachter et al., 2017: 105).

۸-۲ مدل چین و کنترل حکومتی در برابر آزادی فردی

در سوی دیگر، مدل چین با تمرکز بر نظارت متمرکز و الزام به ردیابی عملیاتی، نوعی پاسخ‌گویی ایجاد می‌کند. این مدل بر پایه ثبت دقیق فعالیت‌ها و توانایی دولت برای مداخله و اصلاح یا تعلیق سیستم‌ها بنا شده است. اما این مدل، یک پاسخ‌گویی مبتنی بر اقتدار حکومتی است، نه عدالت مدنی. این دیدگاه، اساس حقوق خودمختاری فردی را تضعیف کرده و با مبنای حقوق خصوصی ایران همخوانی ندارد، زیرا تمرکز بر انقیاد تکنولوژیک است نه جبران زیان شهروند(Roberts, 2019: 120).

۸-۳ راهکار حقوقی ایران و مسئولیت مطلق با اهرم شفافیت

راه‌حل مناسب، استفاده از نقاط قوت مدل‌های جهانی و استوار ساختن آن بر مبنای مطلق است؛ یعنی حذف نیاز به اثبات تقصیر و جایگزینی آن با یک الزام فنی قابل راستی‌آزمایی. چون استفاده از هوش مصنوعی در حوزه‌های اثرگذار بر زندگی افراد مانند استخدام، اعطای وام، تشخیص پزشکی یا خدمات عمومی، یک فعالیت ذاتاً پرخطر است، باید اصل مسئولیت مطلق اعمال شود (Rahban et al., 2023: 45). این اصل که در موارد خطرناک مانند رانندگی با وسایل نقلیه پرسرعت یا حمل مواد خطرناک در حقوق ایران سابقه دارد، در مورد سامانه‌های خودکار باید تعمیم یابد. در این حالت، صرف اثبات زیان ناشی از خروجی الگوریتم، مسئولیت جبران خسارت را فعال می‌کند و نیازی به اثبات تقصیر در طراحی، آموزش یا استقرار نیست. این امر، تحقق عدالت جبرانی را برای زیان‌دیده تضمین می‌کند.

اعمال مسئولیت مطلق بدون مکانیسم دفاعی، می‌تواند منجر به بار بیش از حد بر نوآوری شود. برای جلوگیری از مسئولیت‌دهی نامتوازن، باید یک اهرم دفاعی تعریف شود که فرض تقصیر قابل ابطال باشد. مسئولیت مطلق کارفرما یا توسعه‌دهنده تنها در صورتی قابل رفع است که او بتواند با ارائه مدارک مستدل بررسی فنی استاندارد سیستم را اثبات کند(Raji et al., 2020: 112). این ممیزی باید شامل گزارش‌های دقیق سنجش سوگیری و قابلیت تبیین‌پذیری باشد و نشان دهد که توسعه‌دهنده داده‌های آموزشی را

1. Metadata Provenance

2. Digital Transparency Seal

بر اساس استانداردهای اخلاقی و قانونی مرتبط با حوزه عملکرد مانند عدم تبعیض نژادی یا جنسیتی غربال کرده است و از مدل‌هایی استفاده کرده است که در صورت بروز خطا، امکان تحلیل ریشه‌ای فراهم باشد. فقدان این مستندات یا عدم ارائه گزارش ممیزی فنی معتبر، به طور خودکار منجر به برقراری فرض تقصیر می‌شود و بار اثبات را به‌طور کامل به دوش طرف توسعه‌دهنده منتقل می‌کند. این سازوکار، مراقبت فنی متعارف را از سطح توصیه‌ای به سطح الزام حقوقی ارتقا می‌دهد و اگر توسعه‌دهنده نتواند استاندارد مراقبت فنی خود را مستند کند، مقصر شناخته می‌شود.

مدل پیشنهادی، یک پل ارتباطی میان عدالت مدنی و واقعیت‌های تکنولوژیک فراهم می‌آورد. مدلی که بر مسئولیت مطلق فنی استوار بوده و با مکانیسم فرض تقصیر مشروط به شفافیت اجرایی تقویت شود، یک مبنای نوین ارائه می‌دهد. این مبنا نه تنها جبران خسارت را تضمین می‌کند، بلکه با الزام به مستندسازی علمی منطق الگوریتم‌ها، کرامت انسانی و حق فردی شهروندان را در برابر ماشین‌ها صیانت می‌بخشد؛ زیرا فرد زیان‌دیده مجبور نیست پیچیدگی‌های یادگیری عمیق را در دادگاه اثبات کند، بلکه تنها باید زیان خود را ثابت نماید و توسعه‌دهنده موظف به اثبات بی‌تقصیری فنی خود است.

نتیجه

تحول هوش مصنوعی در ساختار تصمیم‌گیری‌های انسانی، بنیان‌های کلاسیک مسئولیت مدنی را با چالش مواجه ساخته است؛ چرا که با کمرنگ شدن نقش اراده مستقیم انسانی در فرآیندهای تصمیم‌گیری خودکار و جایگزینی آن با تعامل پیچیده داده‌ها و الگوریتم‌ها، مفاهیم سنتی همچون تقصیر، علیت و قصد انسانی در رویکرد یادگیری ماشین به سهولت قابل احراز نیستند. در این شرایط، تداوم پاسخ‌گویی مستلزم تغییر کانون مسئولیت از رفتار انسانی به عملکرد فنی و سیستمی است؛ به این معنا که معیارهایی نظیر «تبیین‌پذیری» (قابلیت بازسازی منطق تصمیم برای انسان)، «شفافیت اجرایی» (دسترسی به لاگ‌ها و مستندات آزمون‌های مدل) و «قابلیت ممیزی» (امکان بازرسی بیرونی) باید به‌عنوان ارکان کمکی در کنار عناصر سنتی مسئولیت مدنی مورد توجه قرار گیرند. هدف از این تغییر کانون، فراهم آوردن امکان مطالبه عدالت از درون ساختار فناوری است، نه صرفاً در پیامدهای آن.

در حوزه حقوق شهروندی، این تحول به رسمیت شناختن حق تبیین تصمیم الگوریتمی را ایجاب می‌کند؛ حقی که بر اساس ماده ۸ آیین نامه هوش مصنوعی اتحادیه اروپا و اسناد مشابه، به فرد اجازه می‌دهد منطق تصمیم خودکاری را که بر سرنوشت او اثر می‌گذارد (مانند رد درخواست وام توسط سامانه اعتباری، جریمه خودکار ترافیک شهری، یا رد صلاحیت شغلی در فرآیند استخدام) درک و نسبت به آن اعتراض کند. در حقوق خصوصی، قراردادها و مسئولیت‌ها باید بر مبنای مراقبت فنی متعارف تنظیم شوند؛ مراقبتی که حداقل استانداردهای آن شامل مستندسازی داده‌های آموزشی، آزمون‌های انصاف (مانند برابری جمعیتی و شانس برابر) و قابلیت بازبینی تصمیمات کلیدی است. فقدان این مستندات، می‌تواند به‌عنوان قرینه یا اماره قصور فنی درنظر گرفته شود که بار اثبات را به نفع زیان‌دیده به سمت بهره‌بردار یا توسعه‌دهنده منتقل می‌کند (اصل انتقال نسبی بار اثبات)، بدون آنکه مسئولیت مطلق و بی‌قید و شرطی بر آن‌ها تحمیل گردد.

نوآوری اصلی این پژوهش طرح «مدل مسئولیت مدنی مشروط بر شفافیت اجرایی» است. در این مدل، شفافیت فنی و قابلیت تبیین الگوریتم از یک توصیه اخلاقی به یک شرط قانونی برای برخورداری از فرض عدم تقصیر تبدیل می‌شود. برخلاف نظریه‌هایی که کوشیده‌اند برای ماشین اراده‌ای تقصیرآمیز متصور شوند، مدل پیشنهادی با پذیرش ماهیت غیرانسانی هوش مصنوعی، مسئولیت را بر عملکرد سامانه و تعهدات فنی عوامل انسانی (توسعه‌دهنده و بهره‌بردار) متمرکز می‌سازد. بدین ترتیب، ناتوانی در ارائه مستندات فنی آزمون‌های استحکام و انصاف، خود می‌تواند به‌عنوان قرینه‌ای برای اعمال مسئولیت (مسئولیت پیش‌دستانه به معنای جابجایی بار اثبات) تلقی شود، نه اینکه به خودی خود مسئولیت‌آور باشد.

برای عملیاتی‌سازی این مدل در نظام حقوقی ایران، پیشنهادهای زیر قابل طرح است: اول، در اصلاحات آتی قانون مسئولیت مدنی، ماده‌ای پیش‌بینی شود که در خسارات ناشی از سیستم‌های هوش مصنوعی پرخطر (مانند سامانه‌های استخدام، اعتبارسنجی، جریمه‌گذاری خودکار شهری و تشخیص پزشکی)، در صورت امتناع بهره‌بردار از ارائه گزارش شفافیت فنی، قرینه قانونی تقصیر به نفع زیان‌دیده برقرار گردد. دوم، نهادهای نظارتی (مانند مرکز ملی فضای مجازی یا سازمان تنظیم مقررات ارتباطات) موظف شوند برای سیستم‌های پرخطر، «گواهی انطباق با اصول انصاف و شفافیت» صادر کنند که اعتبار آن حداکثر دوساله بوده و ارائه آن در دادگاه به نفع توسعه‌دهنده تلقی شود. سوم، دادگاه‌ها با استعانت از کارشناسان رسمی متخصص در حوزه تفسیرپذیری مدل‌های هوش مصنوعی، امکان ارزیابی مستندات فنی ارائه شده را فراهم آورند. این اقدامات، در کنار حفظ کرامت انسانی و حق اعتراض، امکان برقراری توازن میان نوآوری فناورانه و عدالت مدنی را افزایش می‌دهد.

منابع

فارسی

پورقلی، فرهاد. (۱۴۰۲). حریم خصوصی در برابر کارایی: بررسی سازوکارهای حفظ حریم خصوصی داده‌ها در مدل‌های یادگیری ماشین. *مجله امنیت اطلاعات و حریم خصوصی* ، ۲(۲)، ۱۱۵–۱۳۹.

حسینی، مریم. (۱۴۰۰). مسئولیت مدنی پیشگیرانه در قبال تعصبات الگوریتم. *فصلنامه حقوق عمومی* ، ۱۲(۴)، ۹۰–۱۱۴.

رفیعی، حسن. (۱۴۰۱). بررسی تطبیقی مسئولیت توسعه‌دهنده و بهره‌بردار در خسارات ناشی از نرم‌افزارهای قضایی. *مجله حقوق فناوری اطلاعات و ارتباطات* ، ۳(۲)، ۴۰–۶۲.

صادقی، کیومرث. (۱۳۹۹). الزامات دادرسی عادله در عصر دیجیتال و اصول سازمان ملل متحد. *مجله حقوق بشر معاصر* ، ۱۱(۳)، ۶۰–۷۹.

عسکری، وحید. (۱۴۰۳). نقد مفهوم تقصیر سنتی در خسارات ناشی از سیستم‌های خودران هوشمند. *پژوهش حقوق مدنی*، ۸(۳)، ۴۵–۶۸.

فلاحی، نسرین. (۱۴۰۱). کرامت انسانی در مواجهه با قضاوت ماشینی: مطالعه مبانی فلسفی حقوقی. *مجله مطالعات فلسفی حقوق*، ۱۵، ۳۰–۵۵.

محمدی، سارا؛ و شفیعی، بهنام. (۱۴۰۱). تأثیر معماری‌های یادگیری عمیق بر قابلیت تبیین‌پذیری در فرآیندهای قضایی. *فصلنامه پژوهش‌های محاسباتی نوین* ، ۱۴، ۶۵–۸۸.

میرزایی، امیر؛ و کاویانی، رضا. (۱۴۰۲). ارزیابی روش‌های ممیزی فنی برای مدل‌های جعبه سیاه در نهادهای دولتی. *مجله مدیریت و ارزیابی فناوری*، ۷(۲)، ۷۸–۱۰۱.

نوذری، پریوش. (۱۴۰۰). مسئولیت جبرانی بر اساس اصل شفافیت اجرایی. *فصلنامه حقوق کاربردی*، ۹(۴)، ۱۲–۳۴.

نجفی، فاطمه. (۱۴۰۳). نقش داده‌های آموزشی در تداوم سوگیری‌های ساختاری در احکام الگوریتمی. *فصلنامه علوم داده و جامعه*، ۶(۱)، ۱۰–۳۲.

انگلیسی

Barocas, S. and Selbst, A. D. (2016). Big Data's Disparate Impact. *California Law Review* , 104(3), 671-735.

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* . St. Martin's Press.

Floridi, L. and Mökander, J. (2022). From algorithmic accountability to governance. *AI & Society* , 37(3), 1131-1138.

Goel, S. and Liu, J. (2023). The Legal Implications of Generative AI: A Call for a New Liability Framework. *Stanford Technology Law Review*, 26(2), 201-235.

Goodall, S. (2022). Ethical Accountability in AI Systems: Beyond the Technical Fix. *Journal of Business Ethics*, 175(3), 567-580.

Goodman, B. and Flaxman, S. (2023). Does Data or Code Drive Decision Making? A Testable Hypothesis for Explainable AI. In *Proceedings of the International Conference on Machine Learning (ICML)*.

Husser, J. (2021). *Algorithmic Decision-Making and the Law: Challenges for the European Legal Order*. Routledge.

Kearns, M. et al. (2019). Toward a Rigorous Science of Interpretable Machine Learning. *ACM Transactions on Machine Learning Research (TMLR)*, 4, Article 112.

Mittelstadt, B. D. (2019). Principles alone cannot guarantee compliance: Seven challenges in the UK’s approach to AI. *AI & Society* , 34(2), 417-425.

OECD. (2024). *The 2024 OECD AI Principles Update*. OECD Publishing.