
عدالت الگوریتمی و هوش مصنوعی در نظام عدالت کیفری

باقر بهبودی یزدی

کارشناسی ارشد جزا و جرم‌شناسی

چکیده

این مقاله با تمرکز بر عدالت الگوریتمی و حکمرانی هوش مصنوعی در نظام عدالت کیفری، به بررسی چالش‌ها و فرصت‌های ناشی از به کارگیری سامانه‌های هوش مصنوعی در تصمیم‌های کیفری می‌پردازد. هدف پژوهش، تحلیل آسیب‌پذیری‌های حقوقی و فناورانه‌ای نظیر تبعیض الگوریتمی، نبود شفافیت و قابلیت توضیح، تهدید حریم خصوصی و خلأ مسئولیت‌پذیری است. روش تحقیق، تحلیل اسنادی و تطبیقی با مطالعه‌ی گزارش‌ها و اسناد ملی و بین‌المللی است. یافته‌ها نشان می‌دهد که تحقق عدالت الگوریتمی مستلزم چارچوب حکمرانی جامع شامل داده‌های باکیفیت، نظارت انسانی معنادار، حسابرسی منظم و حق اعتراض است. نتیجه‌گیری مقاله آن است که بدون تنظیم‌گری هدفمند، تبیین مسئولیت‌پذیری نهادها و انطباق با مبانی حقوقی و فقهی، اتکای کیفری به سامانه‌های هوشمند می‌تواند به نقض حقوق بنیادین و تضعیف عدالت منجر شود.

واژگان کلیدی: عدالت الگوریتمی، حکمرانی هوش مصنوعی، نظام عدالت کیفری، تبعیض الگوریتمی، حسابرسی

مقدمه

تحول دیجیتال در دهه های اخیر، نظام های عدالت کیفری را نیز تحت تأثیر قرار داده و استفاده از سامانه های مبتنی بر هوش مصنوعی در فرایندهای کشف جرم، ارزیابی ریسک، پیش بینی تکرار جرم و حتی تعیین مجازات رو به گسترش است. این فناوری ها با هدف افزایش سرعت، دقت و کارآمدی تصمیم گیری های قضایی به کار گرفته می شوند، اما هم زمان پرسش های بنیادینی درباره عدالت، بی طرفی و رعایت حقوق اساسی متهمان ایجاد کرده اند. مفهوم «عدالت الگوریتمی» در همین بستر مطرح می شود و ناظر بر این است که تصمیمات مبتنی بر داده و الگوریتم تا چه اندازه می توانند منصفانه، شفاف، پاسخگو و عاری از تبعیض باشند. مسئله اصلی این پژوهش آن است که بهره گیری از هوش مصنوعی در نظام عدالت کیفری، در کنار ظرفیت های فنی، چه مخاطرات حقوقی و اخلاقی ای به همراه دارد و چه سازوکارهایی برای تضمین انطباق آن با اصول بنیادین دادرسی عادلانه ضروری است.

روش تحقیق

این پژوهش با رویکردی توصیفی - تحلیلی و با استفاده از روش اسنادی و تطبیقی انجام شده است. بدین معنا که ابتدا ادبیات نظری مرتبط با عدالت الگوریتمی، حکمرانی هوش مصنوعی و استانداردهای حقوق بشری در حوزه عدالت کیفری بررسی شده و سپس اسناد بین المللی، گزارش های سیاست گذاری و چارچوب های حقوقی برخی کشورها مورد مطالعه تطبیقی قرار گرفته اند. هدف از این رویکرد، شناسایی چالش های مشترک و استخراج مؤلفه های لازم برای طراحی یک چارچوب حکمرانی بومی و کارآمد در زمینه استفاده از هوش مصنوعی در نظام عدالت کیفری بوده است.

۱. طرح مسئله

گسترش فناوری های یادگیری ماشین، پردازش زبان طبیعی و کلان داده موجب شده تصمیمات مبتنی بر داده به تدریج جایگزین بخشی از قضاوت های انسانی در نظام عدالت کیفری شود. امروزه الگوریتم های پیش بینی گر در فرایندهایی مانند پیش بینی نقاط پرخطر وقوع جرم، ارزیابی احتمال تکرار جرم، مدیریت هوشمند پرونده های قضایی و تحلیل خودکار اسناد و آرای دادگاه ها به کار گرفته می شوند. این تحول اگرچه کارآمدی و سرعت تصمیم گیری را افزایش داده است، اما هم زمان پرسشی بنیادین را مطرح می کند: آیا اتکای روزافزون نظام عدالت کیفری به الگوریتم ها، عدالت را تقویت می کند یا خطر بازتولید تبعیض و بی عدالتی را در خود پنهان دارد؟

۲. چارچوب نظری: عدالت، فناوری و حاکمیت قانون

در نظریه عدالت کیفری، دو مؤلفه اساسی یعنی عدالت رویه ای و عدالت ماهوی نقش محوری دارند. عدالت رویه ای به بی طرفی فرآیند، حق دفاع، شفافیت و امکان اعتراض اشاره دارد، در حالی که عدالت ماهوی بر برابری نتایج و پرهیز از تبعیض تأکید می کند. در صورتی که الگوریتم ها بر داده های تاریخی سوگیرانه آموزش ببینند، نابرابری های ساختاری گذشته بازتولید می شوند. بنابراین فناوری ذاتاً عادلانه یا ظالمانه نیست، بلکه کیفیت داده ها و شیوه نظارت بر آن تعیین کننده عدالت است.

از منظر اصول حقوق عمومی، هر تصمیم باید قابل توضیح باشد، مرجع تصمیم گیرنده باید پاسخگو باشد و سازوکاری برای اعتراض مؤثر در نظر گرفته شود. در همین راستا، قانون هوش مصنوعی اتحادیه اروپا با طبقه بندی «سیستم های پریسک» در حوزه عدالت کیفری، نظارت انسانی را یک الزام غیرقابل چشم پوشی قرار داده است.

۳. مبانی اخلاقی و فقهی در حقوق ایران

در نظام حقوق اسلامی و اخلاقی ایران، اصولی چون کرامت ذاتی انسان، اصل برائت، بی طرفی قاضی و حرمت تجسس جایگاه بنیادی دارند. تصمیم گیری کاملاً خودکار و حذف مداخله انسانی در فرآیند دادرسی با این مبانی سازگار نیست. بنابراین، به کارگیری هوش مصنوعی در فرایندهای کیفری تنها زمانی مشروع و قابل دفاع است که در چارچوب نظارت انسانی مؤثر و با رعایت معیارهای اخلاقی و شرعی انجام گیرد.

۴. کاربردهای عملی هوش مصنوعی در عدالت کیفری

فناوری هوش مصنوعی در چند حوزه کلیدی نظام عدالت کیفری به کار گرفته می شود. نخست، پلیس پیش بین با تحلیل داده های مکانی و زمانی جرایم، نقاط مستعد وقوع جرم را شناسایی کرده و امکان تخصیص بهینه منابع انتظامی را فراهم می سازد، اما در عین حال می تواند سبب تمرکز بیش از حد بر مناطق خاص و تشدید انگ اجتماعی شود. دوم، الگوریتم های ارزیابی ریسک نظیر COMPAS در ایالات متحده که برای سنجش احتمال تکرار جرم پیش از صدور حکم یا آزادی مشروط استفاده می شوند، به دلیل عدم شفافیت معیارها و تفاوت آماری نتایج میان گروه های نژادی، مورد نقدهای جدی قرار گرفته اند. سوم، تحلیل خودکار اسناد و آرای قضایی با بهره گیری از پردازش زبان طبیعی، به دسته بندی پرونده ها، استخراج الگوهای رأی و پیش بینی نتایج احتمالی دعاوی کمک می کند. این فرایند ضمن کاهش خطای انسانی و ارتقای سرعت رسیدگی، خطر وابستگی بیش از حد به خروجی نرم افزار را نیز به همراه دارد.

۵. یافته های تحلیلی پژوهش

تحلیل تطبیقی داده های پژوهش نشان می دهد که کارآمدی فناوری هوش مصنوعی به کیفیت و بی طرفی داده ها وابسته است. اصل مشهور «Garbage In, Garbage Out» بدین معنا است که داده های ناقص یا سوگیرانه، به خروجی های ناعادلانه منجر خواهند شد. افزون بر آن، الگوریتم ها باید نقش ابزار تصمیم یار را داشته باشند نه جایگزین قاضی انسان؛ زیرا الگوی مطلوب حکمرانی در هوش مصنوعی همان Human-in-the-Loop است که در آن انسان تصمیم نهایی را اتخاذ می کند.

نتایج تطبیقی همچنین نشان می دهد کشورهای در تحقق عدالت الگوریتمی موفق تر بوده اند که نظارت دوره ای بر سوگیری الگوریتم ها، بازبینی داده های آموزشی و سازوکار شکایت شهروندان را نهادینه کرده اند. در ایران، پیاده سازی بدون انطباق با مبانی فقهی، ساختار آیین دادرسی کیفری و فرهنگ حقوقی، می تواند پیامدهای منفی به دنبال داشته باشد؛ از این رو بومی سازی الزام آور است و نباید مدل های غربی را به صورت مستقیم به کار بست.

بنابراین، یافته‌های پژوهش نشان می‌دهد که هرچند سامانه‌های هوش مصنوعی می‌توانند در تحلیل حجم گسترده‌ای از داده‌های کیفی و شناسایی الگوهای پنهان مؤثر باشند، اما در صورت اتکا به داده‌های ناقص، سوگیرانه یا غیرنماینده، خطر بازتولید یا حتی تشدید تبعیض‌های ساختاری وجود دارد. الگوریتم‌های ارزیابی ریسک ممکن است به‌طور غیرمستقیم متغیرهایی را در نظر بگیرند که با وضعیت اقتصادی، نژادی یا اجتماعی افراد همبستگی دارد و در نتیجه تصمیماتی اتخاذ شود که با اصل برابری در برابر قانون ناسازگار است. افزون بر این، پیچیدگی فنی بسیاری از مدل‌های یادگیری ماشین موجب کاهش شفافیت و دشواری در تبیین دلایل یک تصمیم می‌شود و این امر می‌تواند حق دفاع و امکان اعتراض مؤثر را با چالش مواجه سازد. پژوهش همچنین نشان می‌دهد که نبود سازوکارهای مشخص مسئولیت‌پذیری نهادی، خلأهای قانونی در تعیین مسئولیت خطاهای الگوریتمی و فقدان نظام‌های حسابرسی مستقل از مهم‌ترین موانع تحقق عدالت الگوریتمی به شمار می‌روند.

۶. چالش‌های بنیادین عدالت الگوریتمی

چالش‌های اصلی تحقق عدالت الگوریتمی در نظام کیفری را می‌توان در چند محور خلاصه نمود: جعبه‌سیاه بودن الگوریتم‌ها و ناتوانی در تبیین دلایل تصمیم، سوگیری داده‌های تاریخی، تهدید حریم خصوصی و حقوق داده‌ای افراد، ابهام در تعیین مسئولیت حقوقی خطاهای الگوریتمی و کاهش نقش قضاوت انسانی در روند دادرسی. این عوامل در مجموع ظرفیت نقض اصول بنیادین دادرسی عادلانه را افزایش می‌دهند.

۷. الگوی پیشنهادی حکمرانی مسئولانه

برای تحقق عدالت الگوریتمی در نظام کیفری ایران، مدلی پنج‌بعدی قابل پیشنهاد است. بعد نخست، قانون‌گذاری شفاف است که باید در آن تعریف سیستم‌های پریسک و حدود مسئولیت کیفری و مدنی نهادهای به‌کارگیرنده مشخص گردد. بعد دوم، نظارت انسانی الزامی است که تصمیم‌گیری کاملاً خودکار را در امور کیفری منع می‌کند. بعد سوم بر حسابرسی فنی مستقل تمرکز دارد تا کیفیت داده‌ها و میزان سوگیری الگوریتم‌ها به‌صورت منظم ارزیابی و نتایج به‌طور عمومی گزارش شوند. بعد چهارم، تضمین حق دسترسی و اعتراض است تا متهم از استفاده از الگوریتم در فرایند تصمیم مطلع گردد و امکان بررسی مستقل خروجی‌ها فراهم شود. نهایتاً بعد پنجم، آموزش تخصصی قضات است که هدف آن ارتقای سواد داده‌ای، آشنایی با منطق الگوریتمی و افزایش توان تحلیل فناورانه دستگاه قضایی است. در مجموع این مدل پنج‌بعدی می‌تواند زمینه تحقق حکمرانی مسئولانه و صیانت از عدالت را در عصر هوش مصنوعی فراهم آورد.

بحث و نتیجه‌گیری

بر اساس یافته‌ها، تحقق عدالت الگوریتمی مستلزم طراحی یک چارچوب حکمرانی جامع است که هم‌زمان ابعاد فنی، حقوقی و اخلاقی را در بر گیرد. چنین چارچوبی باید بر کیفیت و تنوع داده‌های آموزشی تأکید کند، امکان نظارت انسانی معنادار بر تصمیمات الگوریتمی را فراهم آورد و نظام‌های حسابرسی دوره‌ای و ارزیابی اثرات حقوق بشری را پیش‌بینی نماید. همچنین تضمین حق اطلاع از سامانه‌های هوشمند، حق دسترسی به توضیح قابل فهم درباره منطق تصمیم و حق اعتراض

مؤثر به نتایج، از الزامات اساسی سازگاری این فناوری با اصول دادرسی عادلانه است. در سطح تقنینی نیز تدوین مقررات شفاف درباره حدود استفاده از هوش مصنوعی، تعیین مسئولیت نهادهای به کارگیرنده و پیش‌بینی ضمانت‌های اجرایی مؤثر، نقش تعیین‌کننده‌ای در جلوگیری از نقض حقوق بنیادین خواهد داشت.

در مجموع، استفاده از هوش مصنوعی در نظام عدالت کیفری ظرفیتی قابل توجه برای ارتقای کارآمدی و تحلیل داده‌های پیچیده فراهم می‌کند، اما بدون تنظیم‌گری هدفمند و استقرار سازوکارهای پاسخگویی و شفافیت، می‌تواند به تضعیف اصول بنیادین عدالت منجر شود. عدالت الگوریتمی نه صرفاً یک مسئله فنی، بلکه موضوعی عمیقاً حقوقی و هنجاری است که نیازمند تلفیق دانش فناوری با مبانی حقوق بشر و اصول دادرسی منصفانه است. از این رو، حرکت به سوی بهره‌گیری مسئولانه از هوش مصنوعی در عدالت کیفری مستلزم ایجاد توازن میان نوآوری فناورانه و صیانت از کرامت انسانی و حقوق بنیادین افراد خواهد بود.

منابع

۱. پژوهشکده هوش مصنوعی و حقوق اسلامی. (۱۴۰۱). گزارش تخصصی هوش مصنوعی و عدالت کیفری.
۲. گزارش مرکز پژوهش‌های مجلس شورای اسلامی. (۱۴۰۰). راهکارهای حقوقی و اخلاقی بکارگیری الگوریتم‌های پیش‌بینی‌گر.
۳. مطهری، م. (۱۳۹۸). حقوق کیفری و فناوری‌های نوین. پژوهشنامه حقوق اسلامی.
۴. نظری، ح. (۱۳۹۹). بررسی جایگاه الگوریتم‌های یادگیری ماشین در عدالت کیفری ایران. فصلنامه حقوق فناوری اطلاعات.
۵. قوانین مصوب جمهوری اسلامی ایران در خصوص فناوری اطلاعات، داده‌کاوی و حقوق شهروندی (جدیدترین ویرایش تا سال ۱۴۰۲).
۶. Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (۲۰۱۶). *Machine Bias: There's Software Used Across the Country to Predict Future Criminals—and It's Biased Against Blacks*. ProPublica.
۷. Buolamwini, J., & Gebru, T. (۲۰۱۸). *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*. Proceedings of Machine Learning Research.
۸. CEPEJ. (۲۰۱۸). *Guidelines on Artificial Intelligence and Criminal Justice*. Council of Europe.
۹. European Union. (۲۰۲۱). *European Artificial Intelligence Act (EU AI Act)*. Brussels: EU Publications.
۱۰. Ferguson, A. (۲۰۱۷). *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. NYU Press.
۱۱. Russell, S., & Norvig, P. (۲۰۲۰). *Artificial Intelligence: A Modern Approach (4th ed.)*. Pearson.
۱۲. Wachter, S., Mittelstadt, B., & Floridi, L. (۲۰۱۷). *Why a Right to Explanation of Automated Decision-Making Exists in the GDPR*. International Data Privacy Law, ۷(۳), ۲۴۳-۲۶۰.
۱۳. Zavrnsnik, A. (۲۰۲۰). *Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings*. European Journal of Criminology.