امکان سنجی کاربست عدالت کیفری الگوریتمی در مرحله تعیین کیفر

عارف خلیلی پاجی (نویسنده مسئول)
استادیار، گروه حقوق جزا و جرم‌شناسی، واحد ورامین- پیشوا، دانشگاه آزاد اسلامی، ورامین، ایران.
arefkhalilipaji@iau.ir
هدیه داودآبادی
دانشجوی کارشناسی ارشد، گروه حقوق کیفری و جرم‌شناسی، دانشگاه علم و فرهنگ، تهران، ایران.
Hediyeh.davoodabadi@gmail.com
فاطمه ابراهیمی
دانشجوی کارشناسی ارشد، گروه حقوق کیفری و جرم‌شناسی، دانشگاه علم و فرهنگ، تهران، ایران.
fatemehebrahiimii2001@gmail.com

چکیده

استفاده از هوش مصنوعی به‌عنوان ابزاری برای یاری‌رسانی به قضات در فرایند تعیین کیفر، در سال‌های اخیر توجه فزاینده‌ای را در ادبیات نظری به خود جلب کرده است. بااین‌حال، نحوه پیاده‌سازی عملی این فناوری در چارچوب‌های اخلاقی و سیاسی موجود، به‌ویژه در نظام‌های کیفری غیرایده‌آل، کمتر مورد بررسی قرار گرفته است. پژوهش حاضر با هدف پر کردن این خلأ و با اتکا به دیدگاه‌های اخیر جسپر رایبرگ[1]، به امکان‌سنجی کاربست این فناوری در نظام عدالت کیفری می‌پردازد. روش تحقیق مبتنی بر تحلیل دو فرض بنیادین است: «فرض بیش‌کیفرانگاری[2]» که بر وجود مجازات‌های نامتناسب در بسیاری از نظام‌های کیفری تأکید دارد و «فرض حفظ وضع موجود» که پذیرش سامانه‌های الگوریتمی را منوط به عدم اخلال نظام‌مند در نظم کیفری مستقر می‌داند. در پاسخ به چالش‌های ناشی از این دو فرض، «مدل کاربست محدود» به‌عنوان یک چارچوب عملیاتی چهارمرحله‌ای مطرح می‌شود. نوآوری پژوهش حاضر در ارزیابی انتقادی این مدل و تبیین ظرفیت‌ها و محدودیت‌های پذیرش تدریجی الگوریتم‌های تصمیم‌یار در نظام عدالت کیفری نهفته است.

کلیدواژه‌ها: بیش‌کیفرانگاری، تعیین کیفر الگوریتمی، سزاگرایی محدودکننده، عدالت الگوریتمی.

[1] Jesper Ryberg.
[2] Overpunishment Assumption.

Feasibility of Applying Algorithmic Criminal Justice in the Sentencing phase

Abstract

The use of artificial intelligence as a tool to assist judges in the sentencing process has attracted growing attention in contemporary legal scholarship. Nevertheless, the practical implementation of this technology within existing ethical and political frameworks—particularly in non-ideal penal systems—has received comparatively limited examination. This study seeks to address this gap by drawing upon the recent work of Jesper Ryberg and exploring the feasibility of employing algorithmic decision-support systems in criminal justice.

The analysis is based on two foundational assumptions. The first is the overpunishment hypothesis, which maintains that many criminal justice systems impose punishments that exceed what offenders morally deserve. The second is the status quo preservation hypothesis, according to which the institutional acceptance of algorithmic systems depends on their ability to operate without causing systematic disruption to the existing penal order. In response to the challenges posed by these assumptions, the study examines the Restricted Application Model as a four-stage operational framework for the use of algorithmic sentencing support.

The originality of this research lies in its critical evaluation of this model and its assessment of both the opportunities and limitations associated with the gradual integration of decision-support algorithms into criminal justice systems. The findings suggest that the ethical legitimacy and practical feasibility of algorithmic sentencing depend not on replacing judicial discretion, but on carefully constraining the role of artificial intelligence within a human-centered decision-making framework.

Keywords: Overpunishment; Algorithmic Sentencing; Limiting Retributivism; Algorithmic Justice; Artificial Intelligence.

مقدمه

توسعه فناوری‌های دیجیتال و هوش مصنوعی در دهه‌های اخیر، ابعاد مختلف حکمرانی، از جمله نظام‌های قضایی را تحت تأثیر قرار داده و زمینه‌ساز تحولات بنیادین در این حوزه‌ها شده است. شواهد موجود نشان می‌دهد که پیشرفت‌های هوش مصنوعی با سرعتی فزاینده در حال گسترش است و به‌تدریج حوزه‌های بیشتری از زندگی فردی و اجتماعی را دربر می‌گیرد. این تحولات، به‌عنوان بخشی از پیشرفت‌های نوین در عرصه فناوری و فضای سایبری، با سرعت و گستره‌ای کم‌سابقه رخ داده و پیامدهای حقوقی، اجتماعی و نهادی گسترده‌ای را به همراه داشته است. برخلاف بسیاری از پیشرفت‌های فنی و تکنولوژیک گذشته، آثار و پیامدهای این تحولات در مدت زمانی کوتاه در حوزه‌های مختلف نمایان شده است (شفیعی و ابوذری، ۱۳۹۹). نمود این تحولات در حوزه حقوق و قضا بیش از بسیاری از حوزه‌های دیگر قابل مشاهده است؛ زیرا نظام‌های قضایی به‌طور سنتی از جمله نهادهای محافظه‌کار محسوب می‌شوند و اکنون با فناوری‌هایی مواجه‌اند که بسیاری از رویه‌ها و سازوکارهای متعارف تصمیم‌گیری را تحت تأثیر قرار داده‌اند.

ورود الگوریتم‌ها به حوزه عدالت کیفری ــ که در ابتدا عمدتاً در قالب ارزیابی خطر تکرار جرم مطرح بود ــ امروزه به مرحله حساس تعیین میزان مجازات[3] نیز گسترش یافته است. با افزایش استفاده از الگوریتم‌ها و ابزارهای مبتنی بر هوش مصنوعی در مؤسسات حقوقی، میزان آشنایی نظام‌های قضایی با این فناوری نیز افزایش یافته است. امروزه مؤسسات حقوقی از الگوریتم‌ها و هوش مصنوعی برای بررسی اسناد، تنظیم پرونده‌ها و پیش‌بینی نتایج احتمالی دعاوی استفاده می‌کنند (Faggella, 2020). این روند نشان می‌دهد که هوش مصنوعی دیگر صرفاً ابزاری پژوهشی یا فنی نیست، بلکه به‌تدریج در فرایندهای حقوقی و قضایی نیز مورد استفاده قرار می‌گیرد. ازاین‌رو، می‌توان وضعیت کنونی را مرحله‌ای آغازین از گسترش کاربردهای هوش مصنوعی در نظام‌های قضایی دانست. پژوهش‌های اخیر نشان داده‌اند که الگوریتم‌ها در برخی موارد قادرند تصمیمات دادگاه‌ها را با دقت قابل‌توجهی پیش‌بینی کنند (Aletras et al., 2016; Sulea et al., 2017; Katz et al., 2017). همچنین نشان داده شده است که این فناوری در مدیریت پرونده‌های استاندارد و کم‌پیچیدگی ــ که بخش قابل‌توجهی از دعاوی را تشکیل می‌دهند ــ ظرفیت‌های قابل‌توجهی دارد(Mandri, 2019).این تحولات ظرفیت‌های جدیدی را برای بهره‌گیری از ابزارهای هوشمند در فرایند تصمیم‌گیری قضایی فراهم کرده است و زمینه بحث درباره نقش این فناوری در مراحل مختلف دادرسی، از جمله تعیین کیفر، را بیش از پیش برجسته ساخته است.

در خصوص جایگاه هوش مصنوعی در نظام‌های حقوقی، پژوهشگران متعددی به بررسی ابعاد مختلف این موضوع پرداخته و دیدگاه‌های متفاوتی ارائه کرده‌اند (Quattrocolo, 2020). ورود الگوریتم‌ها به حوزه عدالت کیفری، پرسش‌های بنیادینی را مطرح می‌کند: آیا سامانه‌های مبتنی بر هوش مصنوعی قادرند ملاحظات هنجاری و ارزشی نهفته در مفهوم عدالت را بازنمایی کنند؟ آیا این سامانه‌ها توانایی توجه به ویژگی‌ها و شرایط خاص هر پرونده را دارند؟ و مهم‌تر از همه، اگر داده‌های آموزشی مورد استفاده این سامانه‌ها متأثر از سوگیری‌ها و نابرابری‌های ساختاری باشند، تا چه اندازه می‌توان به عادلانه بودن خروجی‌های آن‌ها اعتماد کرد؟

[3] Sentencing.

ضرورت این تحقیق از آنجا ناشی می‌شود که علی‌رغم پیشرفت‌های فنی در افزایش دقت پیش‌بینی‌های الگوریتمی، همچنان شکاف قابل‌توجهی میان نظریه‌های ایده‌آل و واقعیت‌های غیرایده‌آل نظام‌های کیفری وجود دارد. بسیاری از پژوهشگران حوزه اخلاق مجازات بر این باورند که هدف اصلی پژوهش باید هدایت رویه‌های کیفری به سمت وضعیت مطلوب باشد؛ دیدگاهی که از آن با عنوان «فرض تأثیرگذاری» یاد می‌شود. با این حال، واقعیت موجود نشان می‌دهد که بسیاری از نظام‌های کیفری معاصر، به‌ویژه در جوامع غربی، با پدیده «بیش‌کیفرانگاری» مواجه‌اند. از منظر پیامدگرایانی چون مایکل تونی، حبس‌های طولانی‌مدت و سیاست‌های کیفری سخت‌گیرانه، تأثیر محدودی در بازدارندگی و پیشگیری از جرم داشته‌اند و در مواردی حتی به تشدید بزهکاری منجر شده‌اند (Tonry, 2004: 21) . از سوی دیگر، نظریه‌پردازان سزاگرا مانند ریچارد سینگر و جفری مورفی [4] نیز معتقدند که در صورت اجرای صحیح اصول استحقاق اخلاقی، بسیاری از مجازات‌های کنونی نیازمند بازنگری و کاهش خواهند بود. با وجود این، معرفی هرگونه ابزار فناورانه در این زمینه با چالشی مهم تحت عنوان «پوپولیسم کیفری [5]»مواجه است. سیاستمداران و تصمیم‌گیرندگان عمومی، به دلیل نگرانی از هزینه‌های سیاسی ناشی از اتخاذ سیاست‌های کیفری ملایم‌تر، غالباً از پذیرش سازوکارهایی که ممکن است به کاهش مجازات‌ها منجر شود، استقبال نمی‌کنند. این تعارض میان ضرورت اخلاقی اصلاح نظام کیفری و ملاحظات سیاسی ناظر بر حفظ وضعیت موجود، یکی از مهم‌ترین چالش‌هایی است که هر الگوی کاربردی هوش مصنوعی در حوزه عدالت کیفری باید به آن پاسخ دهد.

بنابراین، مسئله اصلی آن است که در شرایطی که سیاست‌گذاران بر حفظ رویکردهای سخت‌گیرانه تأکید دارند و سامانه‌های یادگیری ماشین نیز به‌طور پیش‌فرض در معرض بازتولید الگوهای موجود در داده‌های تاریخی قرار دارند، چگونه می‌توان از ظرفیت‌های هوش مصنوعی برای ارتقای عدالت کیفری بهره برد؟ این مقاله با تمرکز بر مدل پیشنهادی جسپر رایبرگ، درصدد تبیین چارچوبی است که در آن الگوریتم، به جای بازتولید الگوهای ناعادلانه موجود، بتواند در قالب ابزاری برای شناسایی و تعدیل برخی نارسایی‌های نظام کیفری مورد استفاده قرار گیرد. این مدل با اتکا به «منطق حداقل» تلاش می‌کند تعارض میان ضرورت سیاسی حفظ نظم موجود و ضرورت اخلاقی کاهش مجازات‌های ناموجه را مدیریت کند(Ryberg, 2025) .

نوآوری این مقاله در ارائه ارزیابی انتقادی مدل کاربست محدود ریبرگ و بررسی ظرفیت‌های آن در پرتو واقعیت‌های حقوقی، سیاسی و نهادی مرتبط با نظام عدالت کیفری ایران است. از این حیث، پژوهش حاضر می‌کوشد فراتر از توصیف صرف ابزارهای الگوریتمی، امکان کاربست عملی این فناوری را در بستر چالش‌ها و محدودیت‌های موجود مورد بررسی قرار دهد.

۱- از بیش‌کیفرانگاری تا حفظ وضع موجود

هرگونه تحلیل پیرامون ورود هوش مصنوعی به فرایند تعیین کیفر، مستلزم شناخت دقیق مفروضاتی است که واقعیت نظام‌های کیفری معاصر را شکل می‌دهند. نظام‌های کیفری کنونی در عمل، افزون بر ملاحظات عدالت‌محور، تحت تأثیر ضرورت‌های سیاسی و فشارهای اجتماعی نیز قرار دارند؛ امری که به شکل‌گیری وضعیتی موسوم به «نظام‌های غیرایده‌آل» انجامیده است. جسپر ریبرگ

[4] Singer, R. G. (1979); Murphy, J. G. (1979).
[5] Penal Populism.

استدلال می‌کند که اخلاق کیفری به‌عنوان یک رشته علمی معمولاً بر اساس «فرض تأثیرگذاری[6]» توجیه می‌شود(Ryberg, 2025) . با این حال، در واقعیت، میان آنچه نظریه‌های اخلاقی تجویز می‌کنند و آنچه در عمل در نظام‌های کیفری رخ می‌دهد، شکاف قابل توجهی وجود دارد.

نظام‌های غیرایده‌آل به حوزه‌های قضایی‌ای اشاره دارند که در آن‌ها رویه‌های کیفری از معیارهای اخلاقی پذیرفته‌شده فاصله گرفته‌اند. این وضعیت غالباً محصول «سیاست‌های ترس» و ملاحظات سیاسی است؛ به‌گونه‌ای که نظام‌های کنترل جرم، به‌ویژه در برخی جوامع غربی، به ساختارهایی تبدیل شده‌اند که اگرچه ممکن است از منظر نظری مورد تأیید نباشند، اما به دلیل نگرانی سیاستمداران از متهم شدن به «نرم‌خویی در برابر جرم[7]»، اصلاح آن‌ها با موانع جدی ساختاری مواجه است(Tonry, 2004) .

۱-۱- فرض بیش‌کیفرانگاری

یکی از ارکان اصلی نقد نظام‌های کیفری معاصر، مسئله «بیش‌کیفرانگاری[8]» است؛ وضعیتی که در آن مجازات‌های اعمال‌شده به‌طور نظام‌مند فراتر از آن چیزی قرار می‌گیرند که بر اساس اصول اخلاقی، اهداف بازدارندگی یا استحقاق فردی قابل توجیه باشد. جسپر ریبرگ استدلال می‌کند که تقریباً تمامی نظریه‌پردازان برجسته اخلاق کیفری، با وجود اختلاف در مبانی نظری، در این نکته اشتراک نظر دارند که شکاف قابل توجهی میان «کیفر عادلانه در سطح نظری» و «کیفر اعمال‌شده در عمل» وجود دارد. در چنین شرایطی، هرگونه سامانه هوش مصنوعی که وارد این نظام شود، ناگزیر بر پایه داده‌هایی آموزش خواهد دید که خود متأثر از الگوهای بیش‌کیفرانگارانه هستند(Ryberg, 2025) .

پیوند این واقعیت با ماهیت داده‌محور هوش مصنوعی، ابعاد مسئله را آشکارتر می‌کند. هوش مصنوعی به‌عنوان سامانه‌ای مبتنی بر تحلیل داده، قادر به شناسایی الگوها و استنتاج آماری از داده‌های تاریخی است؛ با این حال، فاقد درک مستقل از مفاهیم هنجاری نظیر عدالت، انصاف و شفقت بوده و این مفاهیم را صرفاً از خلال الگوهای موجود در داده‌ها بازنمایی می‌کند. در نتیجه، اگر داده‌های آموزشی خود متأثر از رویه‌های ناعادلانه یا نامتناسب باشند، الگوریتم نیز همان الگوها را بازتولید خواهد کرد. برای مثال، ممکن است یک الگوریتم بتواند الگوهای آماری موجود در داده‌ها را شناسایی کند، اما توانایی ارزیابی مستقل ابعاد اخلاقی و هنجاری نهفته در آن الگوها را نداشته باشد. در حوزه تعیین کیفر نیز، الگوریتم ممکن است مجازات‌های سنگینِ موجود در پایگاه داده را به‌عنوان معیار متعارف تصمیم‌گیری شناسایی کرده و در پیشنهادهای خود بازتولید کند (جعفری، ۱۴۰۳).

این نگرانی صرفاً جنبه فنی ندارد، بلکه از منظر نظریه‌های توجیه مجازات نیز قابل بررسی است. در چارچوب دیدگاه پیامدگرایانه، مجازات تنها زمانی موجه تلقی می‌شود که منافع اجتماعی حاصل از آن، از جمله بازدارندگی یا کاهش فرصت ارتکاب جرم، بر رنج تحمیل‌شده و هزینه‌های اقتصادی و اجتماعی آن غلبه کند. با این حال، مطالعات تجربی متعددی نشان داده‌اند که رابطه مستقیمی میان تشدید مجازات‌ها و کاهش نرخ جرم وجود ندارد(Tonry, 2004) . در بسیاری از جوامع غربی، به‌ویژه در بستر پدیده «حبس انبوه[9]»، نظام‌های کیفری به نقطه‌ای از بازده نزولی رسیده‌اند؛ وضعیتی که در آن افزایش مدت یا شدت مجازات نه‌تنها آثار بازدارنده قابل توجهی ایجاد نمی‌کند، بلکه با تحمیل هزینه‌های اقتصادی و اجتماعی گسترده، رفاه عمومی را نیز کاهش می‌دهد. این یافته‌ها مؤید آن است که بسیاری از نظام‌های کیفری از منظر پیامدگرایانه با نوعی بیش‌کیفرانگاری ساختاری مواجه‌اند.

---

[6] Impact Assumption.
[7] Soft on crime.
[8] The Overpunishment Assumption.
[9] Mass incarceration.

نظریه‌پردازان سزاگرا[10] همچون ریچارد سینگر و جفری مورفی نیز بر این باورند که مجازات باید متناسب با «استحقاق اخلاقی[11]»مرتکب باشد (Singer, 1979; Murphy, 1979) .اندرو فون هیرش با طرح اصل «تناسب[12]» استدلال می‌کند که کیفر باید منعکس‌کننده شدت جرم و میزان تقصیر مرتکب باشد (von Hirsch, 1993: 6-15). با این حال، در بسیاری از نظام‌های کیفری معاصر، سقف‌های قانونی مجازات به اندازه‌ای افزایش یافته‌اند که حتی احکام صادرشده در چارچوب قانون نیز ممکن است از منظر استحقاق اخلاقی نامتناسب و بیش از حد شدید تلقی شوند. سینگر در اثر خود با عنوان «استحقاق عادلانه» بیان می‌کند که هرگاه مجازات از حدود استحقاق اخلاقی فراتر رود، دولت در معرض این انتقاد قرار می‌گیرد که رنجی ناموجه بر شهروندان تحمیل کرده است(Singer, 1979).

ادبیات معاصر فلسفه مجازات نیز شواهد و استدلال‌های متعددی در تأیید وجود بیش‌کیفرانگاری در جوامع غربی ارائه کرده است. سارول اسمایلانسکی با طرح مباحثی پیرامون جبرگرایی و مسئولیت اخلاقی، به این نتیجه می‌رسد که نظام‌های کیفری غربی در بسیاری از موارد افرادی را مجازات می‌کنند که مسئولیت اخلاقی آن‌ها محل تردید است یا میزان مجازات تحمیل‌شده بر آنان فراتر از تقصیر واقعی‌شان قرار دارد (Smilansky, 2022: 232-244) .وی با طرح مفهوم «فانیشمنت[13]» نشان می‌دهد که فاصله میان وضعیت کنونی زندان‌ها و آنچه از منظر فلسفی می‌توان عادلانه تلقی کرد، بسیار چشمگیر است. از دیدگاه او، بیش‌کیفرانگاری صرفاً به مدت محکومیت محدود نمی‌شود، بلکه کیفیت و شیوه اجرای مجازات را نیز دربرمی‌گیرد.

۲-۱- واقع‌گرایی سیاسی؛ داده در برابر ضرورت

دومین فرض بنیادین، واقع‌گرایی سیاسی است. بر اساس این دیدگاه، تغییر در نظام‌های کیفری زمانی امکان تحقق می‌یابد که نظم موجود را به‌صورت ناگهانی و بنیادین به چالش نکشد. از این منظر، هر سامانه الگوریتمی که آشکارا در پی کاهش گسترده و فوری مجازات‌ها باشد، با مقاومت جدی نهادهای سیاسی و اجتماعی مواجه خواهد شد و احتمال پذیرش آن کاهش می‌یابد.

سیاست‌های کیفری در دهه‌های اخیر تا حد زیادی تحت تأثیر گفتمان «سخت‌گیری بر جرم» شکل گرفته‌اند. در بسیاری از نظام‌های سیاسی، اتخاذ سیاست‌های کیفری شدیدتر با کسب حمایت افکار عمومی و افزایش سرمایه سیاسی پیوند خورده است. در چنین فضایی، معرفی سامانه‌های هوش مصنوعی که به کاهش مجازات‌ها یا استفاده گسترده‌تر از جایگزین‌های کیفری منجر شوند، ممکن است به‌عنوان نشانه‌ای از ضعف در مقابله با جرم تلقی شود. در نتیجه، فشارهای سیاسی موجود سبب می‌شوند که ابزارهای فناورانه برای دستیابی به پذیرش نهادی، ناگزیر خود را با سطوح فعلی مجازات و انتظارات موجود تطبیق دهند.

از سوی دیگر، اعتماد قضات و نهادهای قضایی به سامانه‌های الگوریتمی نیز مستلزم آن است که این ابزارها به‌عنوان عاملی برای برهم زدن نظم مستقر تلقی نشوند. پذیرش اجتماعی و نهادی فناوری تا حد زیادی وابسته به آن است که این فناوری در مقام ابزاری

---

[10] Retributivists.
[11] Moral Desert.
[12] Proportionality.
[13] Funishment
به معنای وضعیتی که در آن به دلیل عدم اثبات اراده آزاد مطلق، باید شرایط زندان را به جای شکنجه روانی به امکاناتی رفاهی تبدیل کرد تا صرفاً ناتوان‌سازی صورت گیرد.

برای افزایش ثبات، انسجام و پیش‌بینی‌پذیری تصمیمات قضایی ظاهر شود، نه وسیله‌ای برای ایجاد تغییرات ناگهانی و بنیادین در سیاست جنایی. از همین رو، ریبرگ استدلال می‌کند که پذیرش نهادی الگوریتم‌ها مستلزم آن است که این سامانه‌ها در مراحل اولیه استقرار، بیش از آنکه در پی اصلاح بنیادین رویه‌های کیفری باشند، با الگوهای موجود تصمیم‌گیری سازگار شوند(Ryberg, 2025) .
در اینجا نوعی تنش بنیادین شکل می‌گیرد؛ از یک سو داده‌های تجربی و یافته‌های علمی ممکن است بر ضرورت کاهش برخی مجازات‌ها یا استفاده گسترده‌تر از مجازات‌های جایگزین دلالت داشته باشند و از سوی دیگر، ملاحظات سیاسی و اجتماعی در جهت حفظ ساختارهای موجود عمل کنند. در چنین شرایطی، الگوریتم‌هایی که صرفاً بر مبنای یافته‌های تجربی طراحی شوند، ممکن است از منظر سیاسی با دشواری‌های جدی مواجه شوند. این وضعیت ایجاب می‌کند که مدل‌های کاربردی هوش مصنوعی به‌گونه‌ای طراحی شوند که میان ملاحظات اخلاقی، شواهد تجربی و الزامات سیاسی تعادل برقرار کنند؛ به نحوی که امکان حرکت تدریجی به سوی اصلاح نظام کیفری فراهم شود، بی‌آنکه مقاومت‌های شدید سیاسی و نهادی را برانگیزد.
از این منظر، اهمیت مدل پیشنهادی ریبرگ در آن است که به جای تلاش برای حذف یکباره بیش‌کیفرانگاری، می‌کوشد راهکاری برای ایجاد سازگاری میان الزامات اخلاقی اصلاح نظام کیفری و محدودیت‌های سیاسی و نهادی حاکم بر فرایند تصمیم‌گیری عمومی ارائه دهد. این ویژگی، مدل مذکور را از بسیاری از رویکردهای صرفاً نظری متمایز می‌سازد.

۲ -ارائه مدل کاربردی محدود و امکان مهار بیش‌کیفرانگاری از طریق هوش مصنوعی

جسپر ریبرگ با پذیرش واقعیت‌های نظام‌های کیفری غیرایده‌آل، مدلی را تحت عنوان «مدل کاربردی محدود[۱۴]» پیشنهاد می‌کند. این مدل را می‌توان یکی از نخستین تلاش‌های منسجم برای طراحی یک سازوکار اجرایی در جهت بهره‌گیری از سامانه‌های توصیه‌گر قضایی در شرایط واقعی دانست؛ شرایطی که در آن هدف صرفاً افزایش دقت تصمیم‌گیری نیست، بلکه کنترل و کاهش آثار ناشی از بیش‌کیفرانگاری نیز مدنظر قرار دارد(Ryberg, 2025).
در تبیین ضرورت چنین مدلی، می‌توان به دیدگاه جان مک‌کارتی[۱۵] اشاره کرد که معتقد بود هوش مصنوعی لزوماً نباید همانند انسان بیندیشد، بلکه باید به‌گونه‌ای طراحی شود که بتواند مسائلی را حل کند که انسان نیز قادر به حل آن‌ها است (سازمند، ۱۳۹۷). مدل ریبرگ نیز بر مبنای همین منطق شکل گرفته است. این مدل در پی بازسازی یا شبیه‌سازی شهود اخلاقی قاضی نیست؛ زیرا چنین امری دست‌کم در وضعیت کنونی از توانایی سامانه‌های هوش مصنوعی فراتر است. در عوض، هدف آن بهره‌گیری از ظرفیت‌های محاسباتی الگوریتم‌ها برای کاهش ناهماهنگی در احکام و کنترل انحراف از رویه‌های رایج قضایی است. از این منظر، الگوریتم نه در مقام «قاضی هوشمند»، بلکه به‌عنوان ابزاری کمکی برای حل یک مسئله تصمیم‌گیری در فرایند تعیین کیفر ایفای نقش می‌کند.

۲-۱ -فرایند اجرایی مدل کاربردی محدود در تعیین کیفر

پیش از ورود به حوزه تعیین کیفر، هوش مصنوعی جایگاه خود را در عرصه پیش‌بینی رفتار مجرمانه و ارزیابی خطر تکرار جرم تثبیت کرده است. امروزه از این فناوری برای شناسایی الگوهای رفتاری و ارزیابی احتمال وقوع رفتارهای مجرمانه استفاده می‌شود. در همین راستا پیشینه به‌کارگیری ابزارهای دیجیتال نشان می‌دهد که منطق فازی در سیستم‌های امنیتی نیز کاربردهایی از قبیل

[14] Restricted Application Model.
[15] John McCarthy.

تطبیق چهره، تطبیق صدا و تطبیق مردمک چشم را دارا می‌باشد که برای ورود به سیستم‌ها یا تشخیص مجرم‌ها استفاده می‌شود (ابوذری، ۱۳۹۶: ۲۹).همچنین هوش مصنوعی می‌تواند در اعمال برخی نهادهای ارفاقی نقش مؤثری ایفا کند؛ از جمله در حوزه تعویق صدورحکم که ماده ۴۰ قانون مجازات اسلامی آن را منوط به ارزیابی وضعیت فردی و اجتماعی مرتکب کرده است و در همین راستا مواد ۴۵ و ۴۶ همان قانون، اعمال آن را تا حدی به پیش‌بینی رفتار آینده محکوم وابسته می‌دانند.

علاوه بر این، در زمینه صدور قرارهای تأمین کیفری نیز ظرفیت‌هایی برای بهره‌گیری از ابزارهای ارزیابی ریسک وجود دارد. سامانه‌های ارزیابی ریسک الگوریتمی[۱۶] با تحلیل مجموعه‌ای از داده‌ها، احتمال بروز رفتارهای خاص را برآورد می‌کنند. مدل ریبرگ در واقع گامی فراتر از این کاربردها برداشته و درصدد انتقال ظرفیت‌های محاسباتی هوش مصنوعی به مرحله تعیین کیفر است. با این حال، برای جلوگیری از سلطه الگوریتم بر فرایند تصمیم‌گیری انسانی، یک فرایند چندمرحله‌ای و کنترل‌شده را پیشنهاد می‌کند.

گام نخست، تعیین حکم اولیه توسط قاضی بدون اطلاع از نظر الگوریتم است. در این مرحله، قاضی باید صرفاً بر اساس محتویات پرونده و ارزیابی حقوقی خود، مجازات پیشنهادی را تعیین کند. هدف از این مرحله، جلوگیری از تأثیر «سوگیری لنگراندازی[۱۷]» است. پژوهش‌های روان‌شناختی نشان داده‌اند که مواجهه با یک عدد یا پیشنهاد اولیه می‌تواند قضاوت‌های بعدی افراد را به‌طور معناداری تحت تأثیر قرار دهد. بنابراین، مشاهده توصیه الگوریتم پیش از شکل‌گیری قضاوت مستقل قضایی ممکن است استقلال تصمیم‌گیری قاضی را تضعیف کند. این فرآیندِ مستقل تفکیک با ساختارِ ذاتی استدلال حقوقی انطباق کامل دارد؛ چرا که تصمیم‌گیری قضایی ماهیتاً فرآیندی پیچیده از پیوند میان ارزش، امر واقع، فهم عرفی و تجاربِ شهودی است و اصرار بر عدم مواجهه پیش‌دستانه با ارقام الگوریتمی، تضمین‌کننده استقلال شأن قضایی و صلاحدیدهای منعطف دادرس است (ابوذری، ۱۳۹۶).

گام دوم، ورود هم‌زمان حکم اولیه قاضی و اطلاعات پرونده به سامانه است. پس از ثبت تصمیم اولیه، متغیرهای مرتبط با پرونده همراه با حکم پیشنهادی قاضی وارد سیستم می‌شوند. این مرحله امکان مقایسه میان ارزیابی انسانی و داده‌های استخراج‌شده از رویه‌های پیشین را فراهم می‌سازد.

گام سوم، پردازش داده‌ها و تولید توصیه بر مبنای رویه‌های موجود قضایی است. در این مرحله، الگوریتم با استفاده از پایگاه داده‌ای متشکل از پرونده‌های مشابه، الگوهای غالب در تعیین مجازات را استخراج می‌کند. نتیجه این فرایند، ارائه توصیه‌ای مبتنی بر رویه‌های قضایی پیشین است. چنین رویکردی می‌تواند به تحقق نوعی عدالت توزیعی کمک کند؛ زیرا احتمال برخوردهای متفاوت با پرونده‌های مشابه را کاهش می‌دهد و از تأثیر سلیقه‌های شخصی بر تصمیمات کیفری می‌کاهد(Chiao, 2018: 246).

گام چهارم، اعمال «منطق حداقل[۱۸]» و حذف توصیه‌های تشدیدکننده مجازات است. این مرحله مهم‌ترین وجه تمایز مدل ریبرگ محسوب می‌شود. بر مبنای فرض بیش‌کیفرانگاری، استفاده از هوش مصنوعی در تعیین کیفر تنها زمانی قابل توجیه است که خروجی‌های الگوریتمی در جهت کاهش یا تعدیل مجازات به کار گرفته شوند و نه تشدید آن. در غیر این صورت، خطر بازتولید سوگیری‌های ساختاری موجود در داده‌های تاریخی افزایش خواهد یافت.

بر این اساس، اگر مجازات پیشنهادی الگوریتم شدیدتر از حکم اولیه قاضی باشد، این پیشنهاد کنار گذاشته می‌شود و تصمیم قاضی ملاک عمل قرار می‌گیرد. اما در صورتی که الگوریتم مجازاتی سبک‌تر پیشنهاد کند، نتیجه به قاضی ارائه می‌شود تا امکان بازنگری

[16] Risk Assessment
[17] Anchoring Bias..
[18] The Minimum Logic.

در تصمیم اولیه فراهم شود. این سازوکار یک‌سویه تضمین می‌کند که مداخله الگوریتم صرفاً در جهت کاهش آثار بیش‌کیفرانگاری صورت گیرد (Ryberg, 2025). از این منظر، می‌توان میان مدل ریبرگ و نهادهای تخفیف در حقوق کیفری ایران نیز نوعی هم‌پوشانی مشاهده کرد؛ زیرا جهات تخفیف مقرر در ماده ۳۸ قانون مجازات اسلامی، از جمله ندامت، فقدان سابقه کیفری مؤثر و شرایط خاص مرتکب، قابلیت آن را دارند که به‌عنوان متغیرهای ورودی در طراحی چنین سامانه‌هایی مورد استفاده قرار گیرند.

۲-۲ - مسئله ورودی و محدودیت‌های داده‌ای در سامانه‌های تعیین کیفر

مدل پیشنهادی ریبرگ برخلاف سامانه‌های پیشنهاددهنده باز یا الگوریتم‌های الزام‌آور، بر یک ساختار تصمیم‌گیری نامتقارن استوار است. در این مدل، رابطه میان قضاوت انسانی و تحلیل آماری ماشین به‌گونه‌ای طراحی می‌شود که نقش نهایی همچنان در اختیار تصمیم‌گیرنده انسانی باقی بماند. در همین راستا، گسترش فناوری‌های دیجیتال در حوزه عدالت کیفری، توجه پژوهشگران را به مفهومی تحت عنوان «عدالت الگوریتمی»، «عدالت خودکار» یا «عدالت دیجیتال» جلب کرده است (عباچی، ۱۴۰۰).

فرمول‌بندی این مدل بر انتخاب کمترین مقدار[19] میان دو پیشنهاد استوار است. سادگی این ساختار نباید موجب نادیده گرفتن اهمیت آن شود؛ زیرا همین قاعده تضمین می‌کند که الگوریتم هرگز نتواند مجازاتی شدیدتر از آنچه قاضی به‌عنوان سقف قانونی و اخلاقی پذیرفته است، پیشنهاد کند. برای مثال، در نظام حقوقی ایران، حدود قانونی مجازات‌های تعزیری در ماده ۱۸ قانون مجازات اسلامی مشخص شده‌اند و هرگونه تصمیم‌گیری باید در همین چارچوب صورت گیرد.

این رویکرد همچنین با نظریه «سزاگرایی محدودکننده[20]» نوروال موریس هماهنگی دارد. بر اساس این نظریه، استحقاق اخلاقی نقش تعیین‌کننده سقف مجازات را بر عهده دارد و سایر ملاحظات تنها می‌توانند در جهت کاهش آن عمل کنند (Morris, 1974: 59). از این منظر، یکی از نگرانی‌های رایج درباره هوش مصنوعی، یعنی امکان خروج آن از کنترل انسانی[21]، در مدل ریبرگ تا حد زیادی مهار می‌شود. در این مدل، سقف مجازات همواره توسط یک تصمیم‌گیرنده انسانی تعیین می‌شود و الگوریتم نه‌تنها مجاز به عبور از آن نیست، بلکه مأموریت آن صرفاً شناسایی فرصت‌های احتمالی برای تعدیل مجازات است. در نتیجه، حتی در صورت بروز خطاهای داده‌ای نیز احتمال تبدیل شدن سامانه به ابزاری برای تشدید بی‌عدالتی کاهش می‌یابد. به بیان دیگر، الگوریتم در این چارچوب، اختیار تخفیفی اعطاشده به قاضی را به شکلی نظام‌مند و ساختارمند پشتیبانی می‌کند.

از سوی دیگر، یکی از مهم‌ترین چالش‌های طراحی سامانه‌های هوش مصنوعی قضایی، نحوه تبدیل واقعیت‌های پیچیده پرونده‌های کیفری به داده‌های قابل پردازش است؛ مسئله‌ای که در ادبیات این حوزه از آن با عنوان «مسئله ورودی[22]» یاد می‌شود. ماتیس شوارتز و جولیان رابرتز تصریح کرده‌اند که مانع اصلی در طراحی الگوریتم‌های تعیین کیفر، نه در مرحله پردازش داده‌ها، بلکه در مرحله تعریف و ورود داده‌های اولیه قرار دارد(Schwarze & Roberts, 2022: 211) . ریبرگ نیز این چالش را به دو تفسیر اصلی تقسیم می‌کند(Ryberg, 2025) .

[19] Minimum Value.
[20] Limiting Retributivism.
[21] Human Oversight.
[22] Input problem.

نخست، تفسیر کمی[۲۳] که بر مسئله حجم اطلاعات تمرکز دارد. بر اساس این دیدگاه، توصیف دقیق یک رفتار مجرمانه مستلزم ثبت تعداد بسیار زیادی از متغیرها است. برای نمونه، عنوان کلی «ضرب و جرح» به‌تنهایی نمی‌تواند شدت واقعی رفتار، شرایط وقوع جرم، وضعیت بزه‌دیده، رابطه میان طرفین و سایر عوامل مؤثر را منعکس کند. در نتیجه، کاهش تعداد متغیرها ممکن است به ساده‌سازی مفرط واقعیت منجر شود و افزایش آن‌ها نیز هزینه‌ها و احتمال خطا را افزایش دهد.

دوم، تفسیر کیفی[۲۴] که به دشواری ترجمه مفاهیم انتزاعی به داده‌های دیجیتال اشاره دارد. بسیاری از مفاهیم حقوقی و قضایی ماهیتی بافتارمحور دارند و به‌سادگی قابل تبدیل به متغیرهای عددی نیستند. برای مثال، مفهوم «ندامت» را در نظر بگیریم. قاضی ممکن است این مفهوم را از طریق نحوه بیان، رفتار، واکنش‌ها یا تعامل متهم با بزه‌دیده درک کند. تبدیل چنین مؤلفه‌هایی به متغیرهای کمی یا دودویی، همواره با خطر از دست رفتن بخشی از معنای واقعی آن‌ها همراه است. ازاین‌رو، توصیه الگوریتمی را باید نوعی تقریب از واقعیت تلقی کرد، نه بازنمایی کامل آن. با وجود این چالش در تبدیل مفاهیم، باید توجه داشت که مبنای ریاضیاتی این سیستم‌ها پتانسیل خاص خود را دارد؛ چرا که این تئوری دارای یک چارچوب ریاضی و ابزار ریاضی گونه جدید است که اجازه می‌دهد مدل‌های کیفی، قابل تحلیل و بررسی با کامپیوترهای مرسوم باشند (ابوذری، ۱۳۹۶: ۲۸). در تایید این چالش، پژوهش‌های داخلی نیز نشان می‌دهند که اساساً واقعیت‌های نظام قضایی و سازه‌های علوم اجتماعی و حقوقی، ویژگی چندبعدی و تشکیکی دارند؛ به طوری که حاکمیت منطق کلاسیک و ارسطویی (که بر پایه ارزش‌های صوری و دودویی صفر و یک استوار است) در تحلیل این پدیده‌ها کارآمد نبوده و اصرار بر قالب‌های قطبی و جزمی، سیستم را به سمت تقلیل‌گرایی مفرط سوق می‌دهد (ابوذری، ۱۳۹۶).

برای درک بهتر پیچیدگی مرحله ورودی، مطالعه آنتونی دوب و نورمن پارک اهمیت ویژه‌ای دارد. این پژوهشگران در تلاش برای مدل‌سازی متغیر «سابقه کیفری»، شش مؤلفه مختلف را شناسایی کردند: تعداد محکومیت‌های قبلی، فاصله زمانی از آخرین محکومیت، وجود سابقه خشونت، سن فرد در نخستین محکومیت، نسبت شدت جرم فعلی به جرایم پیشین و نوع مجازات‌های قبلی اعمال‌شده است.نتایج پژوهش آنان نشان داد که حتی با در نظر گرفتن سه یا چهار سطح برای هر یک از این مؤلفه‌ها، بیش از ۷۰۰ ترکیب متفاوت تنها برای متغیر سابقه کیفری ایجاد می‌شود(Doob & Park, 1987: 61-62) . اوری شیلد نیز اشاره می‌کند که با ترکیب تمامی متغیرهای مؤثر در یک پرونده کیفری، تعداد حالت‌های ممکن به ده‌ها هزار وضعیت مختلف خواهد رسید؛ وضعیتی که فراهم کردن داده‌های آموزشی کافی برای تمامی آن‌ها را بسیار دشوار می‌سازد. همین مسئله، ادعای دقت بالای الگوریتم‌ها در پرونده‌های پیچیده را با چالش‌های جدی مواجه می‌کند(Schild, 1998: 151-202) .

۳- چالش‌های اخلاقی و نهادی عدالت کیفری الگوریتمی در مرحله تعیین کیفر

مدل پیشنهادی ریبرگ برای به‌کارگیری الگوریتم‌های تصمیم‌یار در تعیین مجازات، در بستر نظام‌های کیفری «غیرایده‌آل» مطرح می‌شود؛ نظام‌هایی که تصمیم‌گیری کیفری در آن‌ها نه‌تنها تابع ملاحظات اخلاقی، بلکه متأثر از محدودیت‌های نهادی، ساختاری و انسانی است. در چنین بستری، انتقادات وارد بر این مدل صرفاً ناظر به دقت یا کارآمدی فنی الگوریتم نیست، بلکه پرسش‌هایی

[23] Amount Interpretation.
[24] Quality Interpretation.

بنیادین درباره امکان‌پذیری، مطلوبیت و پیامدهای عملی مداخله الگوریتمی در فرایند تعیین مجازات را مطرح می‌کند. پاسخ‌های ریبرگ نیز در همین چارچوب شکل می‌گیرند و بر این پیش‌فرض استوارند که ارزیابی اخلاقی استفاده از هوش مصنوعی در تعیین مجازات باید هم‌زمان محدودیت‌های واقعی نظام عدالت کیفری و خطرات ناشی از جایگزینی یا تضعیف قضاوت انسانی را مورد توجه قرار دهد.

از این منظر، اهمیت نظریه ریبرگ در آن است که به جای نادیده گرفتن انتقادات وارد بر کاربرد هوش مصنوعی در تعیین مجازات، آن‌ها را در پرتو واقعیت‌های نهادی نظام کیفری بازخوانی می‌کند و می‌کوشد راه‌حل‌هایی سازگار با شرایط واقعی ارائه دهد. بنابراین، ارزیابی این مدل صرفاً به سنجش توان فنی الگوریتم‌ها محدود نمی‌شود، بلکه باید نسبت آن با اصول بنیادین عدالت کیفری، پاسخگویی نهادی و حفظ جایگاه قضاوت انسانی نیز مورد بررسی قرار گیرد(Ryberg, 2025) .

۱-۳ -چالش پذیرش و مطلوبیت اخلاقی

منتقدان بر این باورند که حتی اگر یک الگوریتم تصمیم‌یار بتواند از منظر اخلاقی یا تحلیلی به کاهش افراط‌های تنبیهی کمک کند، لزوماً در عرصه عملی و نهادی مورد پذیرش قرار نخواهد گرفت؛ زیرا پذیرش سیاستی و سازمانی ابزارهای جدید، تابع ملاحظاتی فراتر از درستی اخلاقی آن‌ها است. هسته اصلی این انتقاد بر این فرض استوار است که ابزارهایی که به‌صورت آشکار در جهت کاهش سطح مجازات‌ها حرکت می‌کنند، از سوی سیاست‌گذاران به‌عنوان تهدیدی برای اقتدار نظام عدالت کیفری و نظم کیفردهی تلقی می‌شوند(Ryberg, 2025) .در چنین شرایطی، مسئله اصلی نه کارآمدی فنی الگوریتم، بلکه هزینه سیاسی ناشی از انتساب «نرم‌خویی در برابر جرم» به تصمیم‌گیران است. ریبرگ با پذیرش این واقعیت توضیح می‌دهد که ابزارهایی که آشکارا درصدد برهم زدن نظم کیفری موجود هستند، معمولاً امکان عبور از فرایندهای رسمی تصمیم‌گیری را پیدا نمی‌کنند. از این رو، مانع اصلی پذیرش را باید در منطق سیاست جنایی و سازوکارهای حفظ مشروعیت جست‌وجو کرد، نه در ضعف محاسباتی الگوریتم‌ها (Ryberg, 2025).افزون بر این، پژوهش‌های مربوط به نابرابری در تعیین مجازات نشان می‌دهد که نظام‌های کیفری موجود، حتی در چارچوب قواعد رسمی، تفاوت‌های گسترده و پایداری در احکام تولید می‌کنند. در نتیجه، مداخلاتی که به جای مدیریت این تفاوت‌ها، کاهش کلی مجازات‌ها را هدف قرار می‌دهند، با احتمال بیشتری با مقاومت و طرد نهادی مواجه خواهند شد.

بر همین اساس، ریبرگ از یک راهبرد واقع‌گرایانه دفاع می‌کند. به اعتقاد وی، اگر قرار است ابزارهای الگوریتمی در نظام عدالت کیفری پذیرفته شوند، باید در ظاهر نقش حافظ ثبات را ایفا کنند و اصلاحات مورد نظر خود را به‌صورت تدریجی و غیرتقابلی پیش ببرند (Ryberg, 2025). این تحلیل با چارچوب‌های اخلاقی استقرار هوش مصنوعی نیز هم‌خوانی دارد؛ زیرا پژوهش‌ها نشان داده‌اند که کاربرد فاقد ضابطه فناوری می‌تواند به تضعیف حاکمیت قانون و کاهش اعتماد عمومی به نظام قضایی منجر شود (Hunter et al., 2020: 749-800). در ادبیات داخلی نیز تأکید شده است که اتکای صرف به سامانه‌های هوشمند ممکن است مشروعیت و بی‌طرفی فرایندهای قضایی را مخدوش ساخته و پذیرش نهادی آن‌ها را با دشواری مواجه کند (حاجی‌زاده، ۱۴۰۰).

نقد مطلوبیت، پرسش متفاوتی را مطرح می‌کند. حتی اگر استفاده از الگوریتم‌های تصمیم‌یار از منظر پذیرش نهادی امکان‌پذیر باشد، آیا به لحاظ اخلاقی نیز مطلوب و موجه است؟ منتقدان استدلال می‌کنند که یکی از مهم‌ترین اهداف ورود ابزارهای الگوریتمی به فرایند تعیین مجازات باید کاهش تفاوت‌های سلیقه‌ای و تقویت اصل «رفتار یکسان با موارد مشابه» باشد؛ اصلی که از ارکان

بنیادین عدالت رویه‌ای و اعتماد عمومی به نظام عدالت کیفری محسوب می‌شود(Ryberg, 2025) . بر این اساس، مشروعیت الگوریتم زمانی قابل دفاع خواهد بود که به افزایش هماهنگی و انسجام تصمیمات قضایی کمک کند، نه آنکه صرفاً در برخی موارد به سمت ارفاق حرکت نماید.

منتقدان همچنین نگران‌اند که مدل‌های اصلاح‌گرایانه، به‌ویژه هنگامی که جهت‌گیری آن‌ها کاهش مجازات باشد، خود منشأ شکل‌گیری نابرابری‌های جدید شوند. ممکن است دو متهم با شرایط مشابه، صرفاً به دلیل تفاوت در نحوه مداخله الگوریتم یا واکنش قاضی به توصیه ماشینی، مجازات‌های متفاوتی دریافت کنند. چنین وضعیتی، حتی اگر به سود یکی از محکومان باشد، می‌تواند احساس بی‌عدالتی و بی‌ثباتی رویه‌ای را تقویت کند. از این رو، این پرسش مطرح می‌شود که آیا کاهش رنج گروهی از محکومان می‌تواند توجیه‌کننده تضعیف اصل برابری صوری باشد؟

ریبرگ این انتقاد را می‌پذیرد، اما پاسخ خود را بر تمایز میان «نظام‌های ایده‌آل» و «نظام‌های غیرایده‌آل» بنا می‌کند. به اعتقاد او، در نظام‌هایی که با بیش‌کیفرانگاری ساختاری مواجه‌اند، همسان‌سازی ریاضی و ثبات صوری ممکن است به معنای توزیع برابر بی‌عدالتی باشد. در چنین شرایطی، پایبندی مطلق به اصل برابری نه‌تنها لزوماً به تحقق عدالت منجر نمی‌شود، بلکه می‌تواند رنج‌های ناموجه موجود را تثبیت کند(Ryberg, 2025) . از این منظر، کاهش رنج ناشی از افراط‌های تنبیهی ممکن است در برخی موارد از ارزش اخلاقی بیشتری نسبت به حفظ کامل یکنواختی صوری برخوردار باشد.

این تحلیل با ادبیات داخلی نیز هم‌سو است. در مطالعات تطبیقی مربوط به تعیین مجازات تأکید شده است که بهره‌گیری از هوش مصنوعی باید به گونه‌ای صورت گیرد که استقلال قضایی، کرامت انسانی و اصل شخصی بودن مسئولیت کیفری را مخدوش نکند و الگوریتم به معیار نهایی داوری تبدیل نشود (صالحی و همکاران، ۱۴۰۳). همچنین در ادبیات تصمیم‌گیری الگوریتمی تصریح شده است که مطلوبیت اخلاقی صرفاً به خروجی عددی وابسته نیست، بلکه به سازگاری فرایند تصمیم‌سازی با ارزش‌های بنیادین عدالت، کرامت انسانی و پاسخگویی نهادی نیز بستگی دارد (فاضلی و فاضلی، ۱۴۰۲).

در نتیجه، پاسخ ریبرگ را نمی‌توان دفاعی از ناهماهنگی در تعیین مجازات دانست. هدف او ایجاد تعادل میان دو ارزش مهم، یعنی برابری رویه‌ای و کاهش رنج ناموجه است. در این چارچوب، ثبات رویه ارزشمند باقی می‌ماند، اما در نظام‌های غیرایده‌آل و مبتلا به بیش‌کیفرانگاری، اخلاق ممکن است کاهش محدود و کنترل‌شده مجازات‌های نامتناسب را بر حفظ کامل یکنواختی صوری ترجیح دهد. بدین ترتیب، ارزش مدل ریبرگ در آن است که می‌کوشد بدون برهم زدن بنیادهای نظام عدالت کیفری، راهی تدریجی برای اصلاح پیامدهای ناشی از بیش‌کیفرانگاری ساختاری ارائه کند. ضرورتِ ضابطه‌مند کردن معیارهای کیفردهی از این رو است که نگاه سنتی قطب‌بندی شده به پرونده‌ها، همواره بستر ایجاد تشتت آراء و اعمال سلایق گوناگون در تعیین مجازات‌ها را فراهم می‌آورد. از این حیث، تعبیه یک سیستم هوشمندِ پشتیبان تصمیم‌یار که بر مبنای درجه‌بندی منظم عمل کند، می‌تواند بدون خدشه به استقلال قضات، آن‌ها را به سمت نوعی وحدت رویه ساختاری و ساختارمند شدن معیارهای کیفردهی هدایت نماید (ابوذری، ۱۳۹۶).

۲-۳- نقد کارایی [۲۵] و شفافیت الگوریتمی [۲۶]

نقد کارایی ناظر بر این پرسش است که حتی اگر یک مدل الگوریتمی از حیث پذیرش نهادی و مطلوبیت اخلاقی قابل دفاع باشد، آیا در عمل نیز می‌تواند به‌گونه‌ای مؤثر و قابل اعتماد عمل کند یا خیر. در سطح اجرایی، مهم‌ترین چالش به نحوه تعامل قاضی با

[25] Workability Objection.
[26] Accuracy vs Transparency.

توصیه‌های الگوریتمی بازمی‌گردد. پژوهش‌های مرتبط با تصمیم‌گیری قضایی نشان می‌دهد که استفاده از سامانه‌های هوشمند می‌تواند تصمیم‌گیران انسانی را در معرض پدیده‌ای موسوم به «سوگیری اتوماسیون[۲۷]» قرار دهد؛ وضعیتی که در آن اعتماد بیش از اندازه به خروجی ماشین، به‌تدریج استقلال قضاوت انسانی را تضعیف می‌کند. در چنین شرایطی، قاضی ممکن است به جای ارزیابی انتقادی شواهد و اوضاع‌واحوال پرونده، به توصیه‌های الگوریتمی به عنوان یک مرجع معتبر اتکا کند. این خطر در فرایند تعیین مجازات اهمیت بیشتری می‌یابد؛ زیرا تصمیم کیفری محصول ارزیابی هم‌زمان عوامل متعدد قانونی، شخصیتی و اجتماعی است و راستی‌آزمایی مستقل خروجی الگوریتم در بسیاری از موارد برای قاضی دشوار و زمان‌بر خواهد بود. ازاین‌رو، حتی الگوریتم‌های دقیق نیز ممکن است به نادیده گرفتن برخی ویژگی‌های خاص پرونده یا تبعیت ناخودآگاه از پیشنهاد ماشینی منجر شوند(Kazim & Tomlinson, 2023).در مقابل، خطر دیگری نیز وجود دارد که می‌توان آن را «مقاومت قضایی» نامید. برخی قضات ممکن است به منظور حفظ استقلال حرفه‌ای، جلوگیری از تضعیف جایگاه قضاوت انسانی یا پرهیز از تلقی «قضاوت ماشینی»، توصیه‌های الگوریتمی را به طور کامل نادیده بگیرند. در چنین وضعیتی، سامانه تصمیم‌یار عملاً به ابزاری تشریفاتی تبدیل می‌شود که تأثیر واقعی بر کیفیت تصمیم‌گیری ندارد. در ادبیات داخلی نیز بر این نکته تأکید شده است که هوش مصنوعی باید در خدمت قاضی قرار گیرد و تصمیم نهایی همواره با لحاظ ویژگی‌های خاص هر پرونده اتخاذ شود؛ زیرا مشابهت ظاهری پرونده‌ها لزوماً به معنای یکسان بودن تمام ابعاد آن‌ها نیست و ذهن تحلیل‌گر قاضی نباید تحت سلطه سامانه قرار گیرد (فاضلی و فاضلی، ۱۴۰۲).

در نتیجه، چالش اصلی نه پذیرش کامل توصیه‌های الگوریتمی و نه طرد کامل آن‌ها، بلکه یافتن نقطه تعادل میان این دو وضعیت است. در همین راستا، بخش مهمی از ادبیات جدید هوش مصنوعی حقوقی از محدودسازی آگاهانه نقش الگوریتم دفاع می‌کند. بر اساس این رویکرد، سامانه‌های هوشمند نباید جایگزین قاضی شوند، بلکه باید به‌عنوان «تلنگر اصلاحی» عمل کنند؛ یعنی قاضی را به بازاندیشی در تصمیم اولیه و ارائه استدلال‌های دقیق‌تر وادار سازند(Alimardani & Istiqomah, 2025) . چنین برداشتی با مدل ریبرگ نیز همخوانی دارد؛ زیرا در این مدل، الگوریتم اختیار تصمیم‌گیری مستقل ندارد و صرفاً در مواردی که امکان کاهش مجازات وجود دارد، نقش هشداردهنده و اصلاحی ایفا می‌کند.
این رویکرد با تأکید بر «وظیفه ارائه دلیل» تکمیل می‌شود. به بیان دیگر، استفاده از سامانه‌های الگوریتمی زمانی می‌تواند مشروعیت خود را حفظ کند که فرایند عملکرد آن‌ها شفاف، قابل توضیح و همراه با نظارت انسانی مؤثر باشد (Hendrickx, 2025). همچنین در چارچوب‌های اخلاقی حاکم بر استقرار هوش مصنوعی توصیه شده است که خروجی سامانه‌ها ماهیتی غیرالزام‌آور داشته باشد و قاضی بتواند در صورت وجود دلایل کافی، از توصیه الگوریتم عدول کند. چنین سازوکاری از یک سو خطر سوگیری اتوماسیون را کاهش می‌دهد و از سوی دیگر مانع شکل‌گیری مقاومت نهادی در برابر فناوری می‌شود (Hunter et al., 2020) .افزون بر این، پژوهش‌ها نشان می‌دهد که کارایی واقعی سامانه‌های تصمیم‌یار صرفاً به دقت فنی آن‌ها وابسته نیست، بلکه به کیفیت داده‌ها، امکان نقد و بازبینی آن‌ها و نیز وجود سازوکارهای پاسخگویی بستگی دارد(Davies & Douglas, 2020).

در کنار نقد کارایی، مسئله شفافیت الگوریتمی نیز از مهم‌ترین چالش‌های اخلاقی کاربرد هوش مصنوعی در تعیین مجازات محسوب می‌شود. یکی از مباحث کلاسیک در این حوزه آن است که افزایش دقت پیش‌بینی معمولاً مستلزم استفاده از مدل‌های پیچیده‌تر

[27] Automation Bais.

است و همین پیچیدگی، قابلیت فهم و ارزیابی مستقل تصمیمات را کاهش می‌دهد. به عبارت دیگر، هرچه مدل‌های یادگیری ماشین پیشرفته‌تر می‌شوند، توضیح نحوه رسیدن آن‌ها به یک نتیجه مشخص دشوارتر خواهد شد(Kazim & Tomlinson, 2023).
این مسئله در حوزه تعیین مجازات اهمیت مضاعف پیدا می‌کند؛ زیرا الگوریتم‌ها معمولاً ده‌ها عامل مختلف را به طور هم‌زمان پردازش و وزن‌دهی می‌کنند و همین امر تشخیص مبنای واقعی تصمیم را برای قاضی، متهم و سایر ذی‌نفعان دشوار می‌سازد (Alimardani & Istiqomah, 2025) .در نتیجه، تعارض میان «دقت» و «شفافیت» صرفاً یک چالش فنی نیست، بلکه مستقیماً با حق دفاع، امکان اعتراض و نظارت عمومی بر اعمال قدرت کیفری ارتباط پیدا می‌کند. از این منظر، هرگونه افزایش دقت که به بهای کاهش قابلیت فهم و ارزیابی تصمیم حاصل شود، می‌تواند هزینه‌های سنگینی برای عدالت رویه‌ای به همراه داشته باشد.

مطالعات جدید نشان می‌دهد که گسترش استفاده از سامانه‌های الگوریتمی در فرایند دادرسی، وظیفه سنتی قاضی در ارائه دلایل رأی را با چالش‌های تازه‌ای مواجه می‌کند. ازاین‌رو، برخی پژوهشگران پیشنهاد کرده‌اند که در تصمیمات مبتنی بر هوش مصنوعی، قاضی علاوه بر تبیین دلایل حقوقی رأی، باید نقش سامانه، حدود تأثیر آن و منطق فنی به‌کاررفته در تصمیم‌سازی را نیز توضیح دهد (Hendrickx, 2025).اگر شفافیت را از عناصر بنیادین دادرسی منصفانه بدانیم، مدل‌های «دقیق اما غیرقابل توضیح» تنها زمانی قابل پذیرش خواهند بود که با سازوکارهای جبرانی نظیر اعلام استفاده از الگوریتم، امکان اعتراض مؤثر و فرایندهای بازبینی مستقل همراه شوند.
در پاسخ به این چالش، برخی نویسندگان پیشنهاد کرده‌اند که به جای طراحی یک سامانه واحد و پیچیده که تمامی عوامل مؤثر در تعیین مجازات را هم‌زمان تحلیل کند، از مجموعه‌ای از ابزارهای تخصصی‌تر استفاده شود که هر یک تنها یک مؤلفه خاص را ارزیابی می‌کنند. در چنین مدلی، امکان راستی‌آزمایی، نقد و ارزیابی نتایج افزایش می‌یابد و رابطه میان داده‌های ورودی و خروجی سامانه روشن‌تر می‌شود (Alimardani & Istiqomah, 2025) .مزیت این رویکرد آن است که به جای انتخاب میان دوگانه «دقت بالا و شفافیت پایین» یا «شفافیت بالا و دقت پایین»، نوعی تعادل میان این دو ارزش برقرار می‌کند.
در مجموع، مهم‌ترین نقطه قوت مدل ریبرگ در مواجهه با این انتقادات آن است که الگوریتم را به جای یک تصمیم‌گیر مستقل، به ابزاری کمکی و محدود تبدیل می‌کند. با این حال، حتی در چنین مدلی نیز کارایی و مشروعیت سامانه منوط به حفظ نقش فعال قاضی، شفافیت نسبی فرایند تصمیم‌سازی، امکان نظارت و اعتراض مؤثر و کنترل مستمر کیفیت داده‌های مورد استفاده خواهد بود. در غیر این صورت، هوش مصنوعی نه به ابزاری برای اصلاح نظام عدالت کیفری، بلکه به سازوکاری برای بازتولید خطاها و ابهامات موجود تبدیل خواهد شد.

۳-۳ -تناسب،فردی‌سازی مجازات و خطر تبعیض سیستمی

حتی اگر مدل‌های محدود و واقع‌گرایانه استفاده از هوش مصنوعی در تعیین مجازات از حیث پذیرش نهادی و کارایی عملی قابل دفاع باشند، همچنان مجموعه‌ای از چالش‌های اخلاقی باقی می‌ماند که نمی‌توان آن‌ها را نادیده گرفت. بخش مهمی از این چالش‌ها به ماهیت داده‌محور الگوریتم‌ها و نحوه بازنمایی عدالت در فرایندهای محاسباتی مربوط می‌شود. از یک سو، تنش میان دقت محاسباتی و شفافیت الگوریتمی این پرسش را مطرح می‌کند که آیا خروجی‌های دقیق اما غیرقابل توضیح می‌توانند با الزامات مشروعیت قضایی، پاسخگویی و دادرسی منصفانه سازگار باشند (Hendrickx, 2025). از سوی دیگر، استفاده از الگوهای آماری

برای تعیین مجازات این نگرانی را ایجاد می‌کند که عدالت کیفری به تدریج از ارزیابی فردی پرونده‌ها فاصله گرفته و به بازتولید الگوهای غالب موجود در داده‌های تاریخی محدود شود (Ryberg, 2025). افزون بر این، مطالعات مربوط به تبعیض الگوریتمی نشان می‌دهد که سامانه‌های هوش مصنوعی حتی بدون استفاده مستقیم از متغیرهای حساس نیز می‌توانند تبعیض‌های ساختاری نهفته در داده‌های آموزشی را بازتاب داده یا تشدید کنند (Davies & Douglas, 2020) .ازاین‌رو، ارزیابی اخلاقی هوش مصنوعی در تعیین مجازات مستلزم توجه هم‌زمان به مسئله تناسب، شفافیت، پاسخگویی و عدالت توزیعی است.

یکی از مهم‌ترین پیامدهای کاربرد الگوریتم‌های پیش‌بینی‌محور در تعیین مجازات، گرایش به استانداردسازی بیش از حد تصمیمات کیفری است. در این رویکرد، «تناسب مجازات» نه بر پایه ارزیابی کیفی و فردی هر پرونده، بلکه بر اساس الگوهای آماری استخراج‌شده از پرونده‌های پیشین تعیین می‌شود. بدین ترتیب، تناسب از یک مفهوم تفسیری و زمینه‌محور به مفهومی کمّی و نسبتاً ثابت تقلیل می‌یابد(Alimardani & Istiqomah, 2025) .

منتقدان معتقدند چنین رویکردی خطر جایگزینی «تناسب مطلق» با «تناسب نسبی» را در پی دارد. منظور از تناسب مطلق، تناسبی است که بر اساس میانگین‌ها، الگوهای غالب و داده‌های آماری شکل می‌گیرد؛ در حالی که تناسب نسبی بر ویژگی‌های خاص هر پرونده، شخصیت مرتکب، شرایط وقوع جرم و سایر عوامل زمینه‌ای استوار است (شیخوند و همکاران، ۱۴۰۲). مسئله اساسی آن است که الگوریتم‌ها تمایل دارند عدالت را در سطح کلان و جمعی بهینه‌سازی کنند، حال آنکه منطق سنتی حقوق کیفری بر فردی‌سازی مجازات و توجه به ویژگی‌های منحصربه‌فرد هر پرونده تأکید دارد.

در همین راستا، در تحلیل چالش‌های کاربرد هوش مصنوعی در مرحله تعیین مجازات تصریح شده است که استفاده از ابزارهای هوشمند ارزیابی خطر و پیشنهاد مجازات، در صورت بی‌توجهی به اصل فردی‌سازی کیفر، می‌تواند به تضعیف یکی از بنیادی‌ترین اصول حقوق کیفری بینجامد و معیارهای کلی و آماری را جایگزین ملاحظات شخصی و وضعیتی مرتکب سازد (صالحی و همکاران، ۱۴۰۳). این هشدار به‌خوبی نشان می‌دهد که خطر اصلی نه استفاده از فناوری، بلکه تبدیل شدن داده‌های آماری به معیار نهایی عدالت کیفری است. در چنین وضعیتی، «تناسب آماری» جایگزین «تناسب حقوقی» می‌شود؛ حال آنکه تناسب حقوقی ماهیتی تفسیری، انسانی و وابسته به شرایط خاص هر پرونده دارد.

پاسخ ریبرگ به این انتقاد نیز از منطق کلی مدل او تبعیت می‌کند. وی معتقد است که الگوریتم نباید مرجع نهایی تعیین تناسب باشد، بلکه صرفاً باید به‌عنوان یک نقطه مرجع یا ابزار کمکی در اختیار قاضی قرار گیرد (Ryberg, 2025). در این چارچوب، قاضی همچنان اختیار دارد با ارائه استدلال حقوقی، از پیشنهاد الگوریتم فاصله بگیرد یا آن را تعدیل کند. بنابراین، مدل ریبرگ درصدد جایگزینی قضاوت انسانی نیست، بلکه می‌کوشد نوعی «تناسب نسبی هدایت‌شده» ایجاد کند؛ تناسبی که از مزایای تحلیل آماری بهره می‌برد، اما تصمیم نهایی را به تشخیص انسانی واگذار می‌کند.

در کنار مسئله تناسب، خطر تبعیض سیستمی نیز یکی از جدی‌ترین چالش‌های اخلاقی استفاده از هوش مصنوعی در تعیین مجازات محسوب می‌شود. سامانه‌های یادگیری ماشین بر مبنای داده‌های تاریخی آموزش می‌بینند و این داده‌ها غالباً بازتاب‌دهنده ساختارهای اجتماعی، اقتصادی و حقوقی موجود هستند. در نتیجه، اگر این داده‌ها متضمن سوگیری یا تبعیض باشند، الگوریتم نیز همان الگوها را فرا خواهد گرفت و در تصمیمات آینده بازتولید خواهد کرد.

اهمیت این مسئله در حوزه عدالت کیفری دوچندان است؛ زیرا متغیرهایی همچون وضعیت اقتصادی، محل سکونت، سطح تحصیلات یا پیشینه اجتماعی می‌توانند حتی بدون ورود مستقیم به مدل، از طریق متغیرهای جانشین بر تصمیم نهایی اثر بگذارند (شیخوند و همکاران، ۱۴۰۲). در چنین شرایطی، ادعای بی‌طرفی مطلق الگوریتم‌ها با تردید جدی مواجه می‌شود؛ زیرا آنچه به‌ظاهر یک تصمیم فنی و خنثی به نظر می‌رسد، ممکن است در عمل بازتاب‌دهنده تبعیض‌های انباشته‌شده در داده‌های تاریخی باشد.

در ادبیات داخلی نیز بر این مسئله تأکید شده است که اتکای بی‌قیدوشرط به داده‌های تاریخی در تصمیم‌گیری‌های قضایی می‌تواند به بازتولید تبعیض‌های ساختاری موجود در جامعه منجر شود؛ زیرا داده‌ها خود محصول روابط قدرت، سیاست‌های کیفری و عملکرد نهادهایی هستند که همواره از بی‌طرفی کامل برخوردار نبوده‌اند (فاضلی و فاضلی، ۱۴۰۲). از این منظر، تبعیض الگوریتمی صرفاً یک نقص فنی نیست، بلکه بازتابی از نابرابری‌های اجتماعی و حقوقی موجود در بستر تولید داده‌ها است.

در همین نقطه است که اهمیت مدل محدود ریبرگ آشکار می‌شود. ریبرگ برخلاف دیدگاه‌هایی که به هوش مصنوعی نقش تعیین‌کننده می‌دهند، معتقد است که استفاده از این فناوری تنها در صورتی قابل دفاع است که نقش آن به‌طور نهادی محدود شده و تحت نظارت مستمر قرار گیرد (Ryberg, 2025) .به بیان دیگر، الگوریتم باید در معرض ارزیابی مداوم، امکان اعتراض، بازبینی داده‌ها و کنترل قضایی قرار داشته باشد تا آثار تبعیض‌آمیز احتمالی آن شناسایی و اصلاح شود.

در نتیجه، مهم‌ترین چالش اخلاقی استفاده از هوش مصنوعی در تعیین مجازات نه صرفاً دقت فنی یا توان پیش‌بینی آن، بلکه نحوه مواجهه با خطر استانداردسازی افراطی و بازتولید تبعیض‌های ساختاری است. ارزش مدل ریبرگ در آن است که می‌کوشد میان بهره‌گیری از ظرفیت‌های فناوری و حفظ اصول بنیادین عدالت کیفری تعادل برقرار کند. در این چارچوب، محدودسازی نقش الگوریتم نه یک عقب‌نشینی فناورانه، بلکه شرط لازم برای استفاده مسئولانه و اخلاقی از هوش مصنوعی در نظام عدالت کیفری است؛ زیرا بدون چنین محدودیتی، فناوری ممکن است به جای اصلاح نابرابری‌ها، به ابزاری برای تثبیت و بازتولید آن‌ها تبدیل شود.

نتیجه‌گیری

بررسی امکان‌سنجی کاربست عدالت کیفری الگوریتمی در مرحله تعیین کیفر نشان داد که مسئله اصلی نه قابلیت فنی هوش مصنوعی، بلکه نحوه تعریف جایگاه آن در ساختار تصمیم‌گیری کیفری است. یافته‌های پژوهش حاضر حاکی از آن است که استفاده از الگوریتم‌ها در نظام‌های کیفری غیرایده‌آل، چنانچه بر مبنای الگوهای متعارف پیش‌بینی ریسک یا تشدید کنترل کیفری طراحی شود، به احتمال زیاد موجب بازتولید سوگیری‌ها، نابرابری‌ها و الگوهای بیش‌کیفرانگارانه موجود خواهد شد؛ زیرا داده‌های آموزشی این سامانه‌ها خود محصول نظام‌هایی هستند که از منظر اخلاقی و کیفرشناختی با چالش‌های جدی مواجه‌اند. در چنین شرایطی، هوش مصنوعی نه ابزاری برای اصلاح عدالت کیفری، بلکه عاملی برای تثبیت و بازتولید کاستی‌های آن خواهد بود.

تحلیل دیدگاه جسپر ریبرگ نشان داد که امکان دفاع اخلاقی از کاربرد هوش مصنوعی در تعیین مجازات، تنها زمانی فراهم می‌شود که نقش الگوریتم از یک ابزار تصمیم‌گیر یا پیش‌بینی‌کننده به یک سازوکار محدود اصلاحی تغییر یابد. «مدل کاربردی محدود» با اتکا بر فرض بیش‌کیفرانگاری و بهره‌گیری از «منطق حداقل»، الگوریتم را در موقعیتی قرار می‌دهد که صرفاً امکان تعدیل و کاهش مجازات را داشته باشد و هرگونه توصیه در جهت تشدید کیفر از فرایند تصمیم‌گیری حذف شود. اهمیت این رویکرد در آن است که برخلاف بسیاری از مدل‌های رایج هوش مصنوعی، هدف آن افزایش قدرت کیفری دولت نیست، بلکه مهار بخشی از پیامدهای نامطلوب نظام کیفری موجود است. ازاین‌رو، استقرار چنین الگویی در نظام‌های حقوقی مختلف، از جمله نظام حقوقی ایران، مستلزم آن است که هوش مصنوعی نه به عنوان مرجع تصمیم‌گیری مستقل، بلکه در جایگاه یک ابزار مشورتی و غیرالزام‌آور مورد استفاده قرار گیرد تا استقلال قضایی و مسئولیت شخصی قاضی همچنان حفظ شود.

پژوهش حاضر همچنین نشان داد که حتی این مدل محدود نیز با چالش‌های مهمی مواجه است. مسئله ورودی، دشواری تبدیل مفاهیم کیفی و زمینه‌مند حقوقی به داده‌های قابل پردازش، خطر سوگیری اتوماسیون، کاهش شفافیت تصمیم‌گیری، امکان تضعیف اصل فردی‌سازی مجازات و بازتولید تبعیض‌های ساختاری از جمله مهم‌ترین موانع پیش روی عدالت کیفری الگوریتمی محسوب می‌شوند. به همین دلیل، کاربست هوش مصنوعی در تعیین کیفر را نمی‌توان صرفاً یک مسئله فنی دانست؛ بلکه این موضوع در درجه نخست یک مسئله حقوقی، اخلاقی و نهادی است که نیازمند چارچوب‌های نظارتی و تضمین‌های هنجاری مؤثر است. در این راستا، هرگونه بهره‌گیری از سامانه‌های تصمیم‌یار باید همراه با امکان نظارت مستمر بر داده‌های آموزشی، ارزیابی دوره‌ای عملکرد الگوریتم‌ها و شناسایی و اصلاح الگوهای تبعیض‌آمیز باشد؛ زیرا بدون چنین سازوکارهایی، خطر تبدیل فناوری به ابزاری برای تثبیت نابرابری‌های موجود همچنان باقی خواهد ماند.

از سوی دیگر، یافته‌ها نشان می‌دهد که موفقیت عملی این الگو وابسته به نحوه استقرار تدریجی آن در نظام عدالت کیفری است. به نظر می‌رسد مناسب‌ترین بستر برای آزمون چنین سامانه‌هایی، حوزه جرایم تعزیری و مواردی است که قانونگذار دامنه بیشتری از اختیار را برای قاضی پیش‌بینی کرده است؛ زیرا در این حوزه‌ها امکان ارزیابی آثار الگوریتم بدون ایجاد مخاطرات جدی برای حقوق و آزادی‌های بنیادین افراد فراهم خواهد بود. افزون بر این، ارتقای سواد الگوریتمی قضات و سایر کنشگران نظام عدالت کیفری شرطی اساسی برای موفقیت این فرایند محسوب می‌شود؛ چراکه قاضی فاقد شناخت کافی از منطق و محدودیت‌های سامانه‌های هوشمند یا به‌صورت غیرانتقادی به توصیه‌های آن‌ها اعتماد خواهد کرد یا به‌طور کامل آن‌ها را نادیده خواهد گرفت و در هر دو حالت، کارکرد اصلاحی مدل از میان خواهد رفت.

بر این اساس، فرضیه اصلی پژوهش مبنی بر امکان استفاده محدود، کنترل‌شده و اخلاقاً موجه از هوش مصنوعی در مرحله تعیین کیفر قابل تأیید است؛ اما این امکان مشروط به رعایت مجموعه‌ای از محدودیت‌های بنیادین است. هوش مصنوعی نباید جایگزین قاضی شود، نباید اختیار تشدید مجازات داشته باشد، نباید به منبع مستقل مشروعیت کیفری تبدیل گردد و نباید بدون نظارت انسانی معنادار مورد استفاده قرار گیرد. در واقع، مشروعیت کاربرد آن تنها زمانی قابل دفاع است که الگوریتم در جایگاه «تصمیم‌یار اصلاحی» باقی بماند و تصمیم نهایی همچنان محصول قضاوت انسانی باشد.

در نهایت، عدالت کیفری الگوریتمی را نباید تلاشی برای واگذاری قضاوت به ماشین دانست، بلکه باید آن را کوششی محدود برای استفاده از ظرفیت‌های فناوری در جهت کاهش بی‌عدالتی‌های موجود تلقی کرد. ارزش واقعی مدل ریبرگ نیز نه در «هوشمندسازی» فرآیند تعیین کیفر، بلکه در تلاش برای تبدیل فناوری از ابزاری در خدمت گسترش قدرت کیفری به ابزاری در خدمت محدودسازی و تعدیل آن نهفته است. از این منظر، آینده عدالت کیفری الگوریتمی نه در حذف قاضی، بلکه در ایجاد نوعی تعامل کنترل‌شده میان قضاوت انسانی و تحلیل الگوریتمی رقم خواهد خورد؛ تعاملی که اگر با الزامات شفافیت، پاسخگویی، کرامت انسانی و نظارت نهادی همراه شود، می‌تواند زمینه کاهش بخشی از بی‌عدالتی‌های ساختاری نظام‌های کیفری معاصر را فراهم سازد.

فهرست منابع

الف) منابع فارسی

۱-ابوذری،مهرنوش(۱۳۹۶).کاربرد منطق فازی در حقوق کیفری ایران،تهران:نشر میزان،چاپ اول.

۲-عباچی، مریم. (۱۴۰۰). «پیشگفتار ریاضیات و علوم جنایی». در: دایره‌المعارف ریاضیات و علوم جنایی. تهران: انتشارات میزان، صص. ۲۵-۴۹.

۳-جعفری، رضا. (۱۴۰۳). «تأملی در باب هوش مصنوعی و ارتقای قضاوت اخلاقی: واکاوی چالش‌ها و فرصت‌ها». مطالعات دینی رسانه.

۴-حاجی‌زاده، نادر. (۱۴۰۰). «تأثیر هوش مصنوعی و فناوری‌های نوین دیجیتال بر فرآیند کشف و تعقیب جرایم در نظام عدالت کیفری ایران». پژوهش‌های نوین در علوم انسانی و حقوق.

۵-سازمند، بهاره. (۱۳۹۷). «هوش مصنوعی در جهان (۳): جمهوری خلق چین». مرکز پژوهش‌های مجلس شورای اسلامی، صص. ۶۴۱-۶۹۳.

۶-سارتور، گیواندی؛ برنتینگ، کارل. (۱۳۹۹). کاربردهای قضایی هوش مصنوعی. ترجمه: مهرنوش ابوذری و محمدسعید شفیعی. تهران: انتشارات میزان.

۷-شیخوند، محمدصادق؛ آشوری، محمد؛ مینایی، بهروز؛ کردعلیوند، روح‌الدین؛ و مهدوی ثابت، محمدعلی. (۱۴۰۲). «هوش مصنوعی و صدور احکام کیفری؛ تصمیم‌سازی یا تصمیم‌گیری؟». پژوهش‌های حقوق تطبیقی، ۲۷(۴)، ۱۳۸-۱۶۷.

۸-فاضلی، معصومه؛ و فاضلی، مژده. (۱۴۰۲). «تصمیم‌گیری الگوریتمی در نظام‌های قضایی: چالش‌های حقوقی، رویکردهای تطبیقی و راهکارهای پیشنهادی». فصلنامه حقوق و فناوری.

۹- صالحی، محمدخلیل،حاجی‌ده‌آبادی، محمدعلی؛ هاشمی‌دمنه، فاطمه‌سادات. (۱۴۰۳). «چالش‌های کاربست هوش مصنوعی در فرایند تعیین مجازات؛ مطالعه تطبیقی نظام حقوقی ایران و کامن‌لا». آموزه‌های حقوق کیفری، ۲۱(۲۸)، ۳۲۹-۳۵۸.

ب)منابع انگلیسی

1-Aletras, N., Tsarapatsanis, D., Preoțiuc-Pietro, D., & Lampos, V. (2016). *Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective. PeerJ Computer Science, 2, e93*. https://doi.org/10.7717/peerj-cs.93

2-Alimardani, A., & Istiqomah, M. (2025). Beyond black boxes and biases: Advancing artificial intelligence in sentencing. *Current Issues in Criminal Justice*. https://doi.org/10.1080/10345329.2025.2527994

3-Chiao, V. (2018). Predicting proportionality: *The case for algorithmic sentencing*. *Criminal Justice Ethics, 37*(3), 238–258.
https://utoronto.scholaris.ca/bitstreams/1179503d-907c-43bb-879c-a2a010f204c3/download

4-Davies, M., & Douglas, H. (2020). *Learning to discriminate: The "perfect proxy" problem in data-driven sentencing. In The algorithmic society: Technology, power, and knowledge. Cambridge University Press.*

5-Doob, A. N., & Park, N. W. (1987). *Computerized sentencing information for judges: An aid to the sentencing process. Criminal Law Quarterly, 30, 54–72.*

6-Faggella, D. (2020). *AI in law and legal practice: A comprehensive view of 35 current applications*. https://emerj.com/ai-in-law-legal-practice-current-applications/

7-Hendrickx, V. (2025). *Rethinking the judicial duty to state reasons in the age of automation. Cambridge Forum on AI: Law and Governance, 1, e26.* https://doi.org/10.1017/cfl.2025.11

8-Hirsch, A. von. (1993). *Censure and sanctions*. Clarendon Press.

9-Hunter, D., Bagaric, M., & Stobbs, N. (2020). *A framework for the efficient and ethical use of artificial intelligence in the criminal justice system. Florida State University Law Review, 47*(4), 749–800.

10-Kazim, T., & Tomlinson, J. (2023). *Automation bias and the principles of judicial review. Judicial Review, 28(1), 9–16.* https://doi.org/10.1080/10854681.2023.2189405

11-Katz, D. M., Bommarito, M. J., & Blackman, J. (2017). *A general approach for predicting the behavior of the Supreme Court of the United States. PLOS ONE, 12(4), e0174698*. https://doi.org/10.1371/journal.pone.0174698

12-Mandri, J. (2019). *Kohtunikud saavad robotabilised: Riik otsib võimalusi kohtusüsteemis tehisintellektirakendamiseks.*

13- Morris, N. (1974). *The Future of Imprisonment*. Chicago: University of Chicago Press. https://repository.law.umich.edu/cgi/viewcontent.cgi?article=4386&context=mlr

14-Murphy, J. G. (1979). *Retribution, justice, and therapy: Essays in the philosophy of law*. D. Reidel. https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=1106&context=faculty_articles

15-Quattrocolo, S. (2020). *Artificial intelligence, computational modelling and criminal proceedings*. Springer. https://www.academia.edu/92478046/Artificial_Intelligence_Computational_Modelling_and_Criminal_Proceedings

16-Ryberg, J. (2025). *Artificial intelligence and criminal justice: How to use algorithmic sentencing support in real life (and ethically non-ideal) penal systems? AI and Ethics*. https://doi.org/10.1007/s43681-024-00655-8

17-Schilder, U. J. (1998). *Criminal sentencing and intelligent decision support. Artificial Intelligence and Law, 6, 151–202.*

18-Schwarze, M., & Roberts, J. V. (2022). *Reconciling artificial and human intelligence: Supplementing not supplanting the sentencing judge. In J. Ryberg & J. V. Roberts* (Eds.), *Sentencing and artificial intelligence* (p. 211). Oxford University Press.

19-Singer, R. G. (1979). *Just deserts: Sentencing based on equality and desert*. Ballinger Publishing. https://books.google.com/books/about/Just_Deserts.html?id=MFIqAQAAMAAJ

20-Smilansky, S. (2022). *Overpunishment and the punishment of the innocent. Analytic Philosophy, 63*(4), 232–244. https://openurl.ebsco.com/contentitem/doi%3A10.1111%252Fphib.12235?sid=ebsco:ocu:record&id=ebsco:doi:10.1111%2Fphib.12235&bquery=IS%202153-9596%20AND%20VI%2063%20AND%20IP%204%20AND%20DT%202022&page=1&link_origin=&searchDescription=Analytic%20Philosophy,%202022,%20Vol%2063,%20Issue%204

21-Tonry, M. (2004). *Thinking about crime: Sense and sensibility in American penal culture*. Oxford University Press. https://api.pageplace.de/preview/DT0400.9780198032335_A24389130/preview-9780198032335_A24389130.pdf