



Judicial Smart Assistant for Alternatives to Imprisonment: AI-Based Risk Prediction Model

Mehdi Ghasmi¹, Mahmud Habibi Tabar², Sadegh Moradi³

Abstract

Field and Aims: The growing prison population, the heavy costs of inmate maintenance, and the ineffectiveness of imprisonment-oriented policies have underscored the necessity of revising Iran's criminal justice model and adopting a rational, data-driven use of non-custodial and rehabilitative measures. In this context, advances in artificial intelligence (AI) and machine learning offer new opportunities for more precise decision-making, risk prediction, and individualized judicial responses. Accordingly, the present study aims to design a localized model of an "Intelligent Judicial Assistant for Alternatives to Imprisonment" that utilizes data analytics and explainable algorithms to assess the risk of recidivism and recommend the most appropriate non-custodial action for each offender.

Method: This research employs a descriptive-analytical and comparative approach between prominent international risk-assessment systems (such as COMPAS and HART) and Iran's legal framework. Based on the principles of sentence individualization and restorative justice, a three-layer architecture was developed, including: (1) a data input layer (criminal, social, and economic variables), (2) an analytic layer using explainable AI algorithms (SHAP and LIME), and (3) a legal-cultural output layer aligned with Islamic principles of justice. Simulated datasets were used for the model's initial testing and accuracy evaluation.

Findings and Conclusions: The findings indicate that the proposed model can estimate the risk-of-recidivism score with over 80 % accuracy by weighting individual, social, and economic variables. The integration of a legal-cultural filter ensured full compliance of the outcomes with Islamic judicial principles and the judge's discretion. Moreover, the implementation of a dynamic feedback loop enhanced the model's continuous learning and precision. The intelligent judicial assistant could, in practice, reduce prison overcrowding, enhance consistency in the application of non-custodial measures, and foster data-driven justice grounded in human dignity. Full realization of this potential requires the establishment of clear ethical and regulatory frameworks, regular algorithmic audits, and an independent supervisory body to monitor discrimination and bias in accordance with Islamic principles of justice and transparency.

Keyword: Artificial Intelligence; Intelligent Judicial Assistant; Risk Prediction; Alternatives to Imprisonment; Data-Driven Justice.

* Citation (APA): Ghasmi, M.; Habibi Tabar, M.; Moradi, S. (2025). Judicial Smart Assistant for Alternatives to Imprisonment: AI-Based Risk Prediction Model. *Applied criminology research*, 3(9), 125-152.

https://www.qacr.ir/article_732169.html?lang=en

1. Department of law, UAE.C., Islamic Azad University, Dubai, United Arab Emirates.

Email: mahdi.ghasemi3294@iau.ac.ir

2. Department of Law, Sav.C., Islamic Azad University, Saveh Iran. (Corresponding Author).

Email: habibitabarmahmoud@iau.ac.ir

3. Department of Law, CT.C., Islamic Azad University, Tehran, Iran. Email: Sad.moradi@iauctb.ac.ir



دستیار هوشمند قضایی برای جایگزین‌های حبس:

مدل پیش‌بینی ریسک مبتنی بر هوش مصنوعی

مهدی قاسمی^۱، محمود حبیبی تبار^۲، صادق مرادی^۳

چکیده

زمینه و هدف: افزایش جمعیت کیفری، هزینه‌های سنگین نگهداری زندانیان و ناکارآمدی سیاست‌های حبس‌محور، ضرورت بازنگری در الگوی عدالت کیفری ایران و بهره‌گیری هدفمند از نهادهای ارفاقی و جایگزین‌های حبس را دوچندان کرده است. در این میان، پیشرفت‌های هوش مصنوعی و یادگیری ماشین فرصت تازه‌ای برای تصمیم‌سازی دقیق‌تر، پیش‌بینی ریسک کیفری و فردی‌سازی و آکنش‌های قضایی فراهم آورده است. از این‌رو، پژوهش حاضر با هدف طراحی مدل بومی «دستیار قضایی هوشمند برای جایگزین‌های حبس» انجام گرفته تا با بهره‌گیری از تحلیل داده و الگوریتم‌های توضیح‌پذیر، ریسک بازگشت به جرم را ارزیابی و مناسب‌ترین اقدام غیرحبسی برای هر محکوم را پیشنهاد دهد.

روش: مطالعه با رویکرد توصیفی-تحلیلی و تطبیق میان الگوهای برجسته خارجی ارزیابی ریسک قضایی (از جمله COMPAS و HART) و نظام حقوقی ایران انجام شده است. بر مبنای اصول فردی‌سازی مجازات و عدالت ترمیمی، معماری سه‌لایه‌ای طراحی گردید که شامل: (۱) لایه ورودی داده (متغیرهای کیفری، اجتماعی و اقتصادی)، (۲) لایه تحلیل مبتنی بر الگوریتم‌های توضیح‌پذیر (SHAP و LIME) و (۳) لایه خروجی حقوقی-فرهنگی منطبق بر عدالت اسلامی است. داده‌های شبیه‌سازی شده برای آزمون اولیه و سنجش دقت مدل به کار گرفته شدند.

یافته‌ها و نتایج: نتایج نشان داد مدل پیشنهادی می‌تواند با وزندهی به متغیرهای فردی و اجتماعی-اقتصادی، امتیاز ریسک بازگشت به جرم را با دقتی بیش از ۸۰ درصد محاسبه کند. تلفیق فیلتر حقوقی-فرهنگی موجب سازگاری نتایج با اصول عدالت اسلامی و اختیارات قاضی گردید. همچنین، پیاده‌سازی چرخه بازخورد پویا، دقت و یادگیری مستمر مدل را تقویت کرد. این دستیار هوشمند می‌تواند با کاهش تراکم زندان‌ها، ارتقای انسجام قضات در اعمال نهادهای ارفاقی و گسترش عدالت داده‌محور مبتنی بر کرامت انسانی، تحولی بنیادین در سیاست جنایی ایران ایجاد کند. تحقق این هدف منوط به وضع چارچوب‌های اخلاقی شفاف، ممیزی منظم الگوریتم‌ها و تأسیس نهاد نظارتی مستقل برای ارزیابی تبعیض و سوگیری احتمالی بر پایه اصول اسلامی عدالت و شفافیت است.

کلیدواژه‌ها: هوش مصنوعی، دستیار قضایی هوشمند، پیش‌بینی ریسک، جایگزین‌های حبس، عدالت داده‌محور.

*استناددهی (APA): قاسمی، مهدی؛ حبیبی تبار، محمود و مرادی، صادق. (۱۴۰۴). دستیار هوشمند قضایی برای جایگزین‌های حبس: مدل پیش‌بینی ریسک مبتنی بر هوش مصنوعی. پژوهش‌های جرم‌شناسی کاربردی، ۳(۹)، ۱۲۵-۱۵۲.

https://www.qacr.ir/article_732169.html

۱. گروه حقوق، واحد امارات، دانشگاه آزاد اسلامی، دبی، امارات متحده عربی.

رایانامه: mahdi.ghasemi3294@iau.ac.ir

۲. گروه حقوق، واحد ساوه، دانشگاه آزاد اسلامی، ساوه، ایران. (نویسنده مسئول).

رایانامه: habitabarmahmoud@iau.ac.ir

۳. گروه حقوق، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران، ایران. رایانامه: Sad.moradi@iauctb.ac.ir

تحولات شتابان فناوری، به‌ویژه در حوزه هوش مصنوعی، بستر تازه‌ای برای بازطراحی فرآیندهای قضایی فراهم کرده است. این فناوری با توانایی پردازش حجم عظیمی از داده‌ها، شناسایی الگوهای پنهان و ارائه پیش‌بینی‌های دقیق، می‌تواند سرعت، دقت و انسجام تصمیمات کیفری را به‌طور چشمگیری بهبود بخشد. در بسیاری از کشورها، ابزارهای «تصمیم‌یار» مبتنی بر پیش‌بینی ریسک نه تنها به کاهش تراکم زندان‌ها کمک کرده‌اند، بلکه رویکرد سیاست جنایی هوشمند را نیز تقویت کرده‌اند. یافته‌های پژوهش‌های اخیر نشان می‌دهد که هوش مصنوعی، فراتر از یک ابزار پشتیبان، قادر است با بازطراحی جریان‌های کاری، بهبود ارتباطات سازمانی و تغییر شیوه تعامل میان قاضی، کارکنان قضایی و شهروندان، زمینه‌ساز تحولی ساختاری در نظام قضایی شود (حبیبی و کریم پور، ۱۴۰۲: ۳). در کنار این رویکرد فناورانه، تجربه‌ی فرانسه نمونه‌ای موفق از به‌کارگیری نیروی انسانی آموزش‌دیده در نقش «دستیار قضایی» به شمار می‌رود؛ نهادی که مأموریت آن پشتیبانی از قاضی در بررسی اسناد، تنظیم پیش‌نویس آراء و انجام تحقیقات تطبیقی است و از سال ۲۰۱۶، به صورت رسمی در ساختار دادگستری فرانسه تثبیت شده است. در پژوهش (سلطانی، ۱۳۹۸: ۱۰-۱۱) نشان داده شده که این مدل با کاهش بار کاری قضات و افزایش دقت تصمیم‌گیری، به ارتقای کیفیت رسیدگی و تسریع فرآیند دادرسی انجامیده است. در مدل پیشنهادی این مقاله، رابطه‌ی دستیار انسانی و دستیار هوشمند نه رقابتی، بلکه تکمیلی و تلفیقی است. «دستیار انسانی» نقش داور فرهنگی - حقوقی را دارد؛ یعنی خروجی‌های تولیدشده توسط سامانه هوش مصنوعی (امتیازهای ریسک، پیشنهاد جایگزین حبس و تحلیل داده‌ای) را از منظر ارزش‌های بومی، تناسب فردی و قابلیت اجرای اجتماعی بازبینی و اصلاح می‌کند. در مقابل، دستیار هوشمند بخش تحلیلی داده‌محور و پیش‌بینی ریسک را برعهده دارد. همکاری این دو در قالب مدل ترکیبی «انسان + ماشین» سبب می‌شود تصمیم نهایی هم از دقت آماری کافی برخوردار باشد و هم با اصول حقوقی و فرهنگی ایران سازگار بماند؛ به این ترتیب، نظام قضایی می‌تواند از قدرت یادگیری الگوریتمی بدون حذف نقش قضاوت انسانی بهره‌گیرد. این نهاد متشکل از فارغ‌التحصیلان حقوقی مستعد است که پس از گذراندن آموزش‌های تخصصی، قضات را در اموری مانند بررسی اسناد، آماده‌سازی پرونده‌ها، انجام تحقیقات و پژوهش‌های حقوقی و نگارش پیش‌نویس آراء پشتیبانی می‌کنند. بهره‌گیری از این دستیاران ضمن کاهش بار کاری محاکم و افزایش سرعت رسیدگی، فرصت کارورزی عملی را برای نسل آینده قضات و وکلا فراهم ساخته است. تجربه فرانسه نشان می‌دهد که ترکیب برنامه‌ریزی نهادی، آموزش مستمر و تقویت نقش‌های کمکی در فرآیند دادرسی، می‌تواند به افزایش کیفیت و بهره‌وری در نظام قضایی بینجامد. ایده «دستیار

قضایی هوشمند» در این پژوهش، ادامه طبیعی همین الگو است؛ با این تفاوت که علاوه بر سرمایه انسانی متخصص، از ظرفیت‌های هوش مصنوعی همچون تحلیل کلان‌داده‌ها، پیش‌بینی ریسک و ارائه پیشنهادهای حقوقی شفاف و قابل توضیح برای تکمیل و ارتقای نقش یاریگران انسانی بهره می‌گیرد. بدین ترتیب، مدل پیشنهادی با هم‌افزایی توان انسانی و الگوریتمی، هم به دقت و سرعت تصمیم‌سازی قضایی می‌افزاید و هم انسجام و شفافیت رویه‌ها را تضمین می‌کند. در ایران، به‌موجب قوانین «مجازات اسلامی» (۱۳۹۲)^۱ و «کاهش مجازات حبس تعزیری» (۱۳۹۹)^۲، بستر قانونی استفاده از جایگزین‌های حبس فراهم شده است. با این حال، ضعف زیرساخت‌های داده‌محور، فقدان مدل‌های بومی ارزیابی ریسک، ناهمگونی در اجرای رویه‌ها و استمرار فرهنگ حبس‌محور مانع بهره‌برداری کامل از این ظرفیت‌هاست. تجربه توسعه دادرسی الکترونیکی در سال‌های اخیر نشان می‌دهد که گرچه بستر اولیه حرکت به سوی هوشمندسازی فراهم شده، اما بهره‌گیری اثربخش از هوش مصنوعی مستلزم مدلی منطبق با ملاحظات حقوقی و فرهنگی است که خروجی‌هایی شفاف، قابل توضیح و قابل اتکا تولید کند. این پژوهش با الهام از تجارب جهانی و انطباق با شرایط بومی، مدل «دستیار قضایی هوشمند» را برای پیش‌بینی ریسک و انتخاب جایگزین‌های مناسب حبس ارائه می‌دهد و مسیر رسیدن به عدالت داده‌محور را از طریق پیوند میان سه حوزه حقوق، فناوری و فرهنگ هموار می‌سازد.

۱. مبانی نظری

۱-۱. هوش مصنوعی در عدالت کیفری

هوش مصنوعی^۳ به مجموعه‌ای از روش‌ها، الگوریتم‌ها و سامانه‌های رایانه‌ای گفته می‌شود که با هدف شبیه‌سازی فرایندهای شناختی انسان مانند یادگیری، استدلال، تحلیل و تصمیم‌گیری توسعه یافته‌اند (Norvig & Russell, 2021). این فناوری که ابتدا، در حوزه‌های صنعتی و تجاری کاربرد گسترده پیدا کرد، در دهه‌های اخیر به‌طور فزاینده وارد عرصه عدالت کیفری شده است. از تحلیل داده‌های جرم‌شناختی و پیش‌بینی ریسک ارتکاب مجدد جرم تا بهینه‌سازی تصمیمات قضایی، هوش مصنوعی توانسته در نقش یک ابزار تصمیم‌یار به کار گرفته شود (Angwin و همکاران، 2016). مهم‌ترین مزیت این فناوری، امکان پردازش سریع و دقیق حجم عظیمی از داده‌ها و کشف الگوهای پیچیده‌ای است که ممکن است از دید کارشناسان انسانی پنهان بماند (Hildebrandt, 2018). مطابق یافته‌های مقاله «امکان دادرسی

۱. قانون مجازات اسلامی، مصوب ۱۳۹۲، روزنامه رسمی جمهوری اسلامی ایران، شماره ۱۹۷۷۳ مورخ ۱۳۹۲/۰۳/۰۱.

۲. قانون کاهش مجازات حبس تعزیری، مصوب ۱۳۹۹، روزنامه رسمی جمهوری اسلامی ایران، شماره ۲۱۹۳۵ مورخ ۱۳۹۹/۰۲/۲۳.

بر پایه هوش مصنوعی» (موسوی، ۱۴۰۴)، کاربرت هوش مصنوعی در دادرسی می‌تواند فرایند قضاوت را از طریق افزایش دقت، ایجاد انسجام در آراء و کاهش اطاله دادرسی بهبود دهد؛ به شرط آنکه ملاحظات حقوقی و فرهنگی هر کشور رعایت شود. این مقاله تأکید می‌کند که جایگاه هوش مصنوعی باید محدود به نقش تصمیم‌ساز و یاریگر قاضی باقی بماند، نه جایگزین او؛ چراکه سپردن نقش کامل صدور حکم به یک سامانه هوشمند، خطراتی همچون انسان‌زدایی از فرایند دادرسی، گسست آیینی و تضعیف نقش قاضی به‌عنوان ضامن نهایی عدالت را در پی دارد. علاوه بر این، مطالعات اخیر نشان داده‌اند که چنین سامانه‌هایی قادرند با کاهش تأثیر تعصبات فردی، خطاهای ناشی از خستگی و عوامل غیرحقوقی، ثبات و انسجام در تصمیمات را افزایش دهند. با این حال، همان‌گونه که برخی پژوهشگران داخلی هشدار داده‌اند، عدم شفافیت الگوریتمی و وجود سوگیری در داده‌های آموزشی می‌تواند عدالت را به خطر اندازد؛ از این رو، تفکیک صریح مرز میان «هوش مصنوعی به مثابه مشاور» و «هوش مصنوعی به مثابه قاضی» ضروری است. افزون بر این، خطرات هوش مصنوعی به این موارد محدود نمی‌شود و در صورت نبود ملاحظات حفاظتی، می‌تواند به جعل اسناد و عناوین، تسهیل رفتارهای مجرمانه و سرقت‌های اینترنتی و غیراینترنتی، نقض گسترده حریم خصوصی، تغییر در ماهیت و اعتبار ادله اثبات دعوی و حتی آسیب‌های فیزیکی به اشخاص و جامعه بینجامد (حبیبی و کریم‌پور، ۱۴۰۲: ۳). این تهدیدها به‌روشنی نشان می‌دهد که هرگونه بهره‌گیری مسئولانه از هوش مصنوعی در عدالت کیفری نیازمند چارچوب‌های سخت‌گیرانه حقوقی، ممیزی امنیتی مداوم و کنترل‌های اخلاقی است. تجربه نظام‌های حقوقی پیشرو نشان داده است که موفقیت در استفاده از فناوری‌های هوشمند منوط به توانمندسازی و آموزش مستمر کاربران است. تحقیقات کوهی و همکاران (۱۴۰۲) بیانگر آن است که آموزش دقیق قضات، ضابطان و کارشناسان نه تنها مهارت کار با فناوری‌های نوین را ارتقا می‌دهد، بلکه فرهنگ استفاده آگاهانه و مسئولانه را نیز نهادینه می‌کند. نمونه‌ای شاخص از این رویکرد، نهاد «دستیار قضایی» در فرانسه است که نه تنها برای کاهش حجم کار دادگاه‌ها ایجاد شد، بلکه از همان آغاز بر پایه آموزش تخصصی شکل گرفت. این دستیاران که عمدتاً، از میان فارغ‌التحصیلان حقوق انتخاب می‌شوند، دوره‌های جامع بررسی پرونده، تحلیل حقوقی، پژوهش در قوانین و رویه‌ها و نگارش پیش‌نویس آراء را می‌گذرانند. چنین برنامه‌ای علاوه بر ارتقای کیفیت تصمیمات و کاهش زمان رسیدگی، نیروی انسانی متخصص و آماده برای تصدیق مناصب قضایی و وکالتی در آینده ایجاد می‌کند (سلطانی، ۱۳۹۸). در ایران نیز ادغام آموزش‌های تخصصی هوش مصنوعی در برنامه‌های توسعه حرفه‌ای قضات و کارکنان دستگاه عدالت، نه تنها می‌تواند از بروز خطاهای انسانی و برداشتهای ناصحیح از داده‌های هوشمند

جلوگیری کند، بلکه موجب ارتقای سواد فناوری و افزایش اعتماد عمومی به کاربست این فناوری‌ها در فرایند دادرسی خواهد شد. مطابق یافته‌های پژوهش حبیبی و کریم‌پور (۱۴۰۲، ۲۷۰)، تحقق «عدالت هوشمند» مستلزم سرمایه‌گذاری موازی در آموزش، تدوین آیین‌نامه‌های اخلاقی و ایجاد سازوکارهای نظارتی شفاف است تا قاضی و کارشناس بتوانند رفتار الگوریتم‌ها را تحلیل و کنترل کنند. افزون بر این، هوش مصنوعی در صورتی که به‌درستی در ساختارهای قضایی ادغام شود، قادر است نقشی فراتر از بهبودهای محدود در دقت و سرعت ایفا کرده و با بازطراحی نظام تصمیم‌گیری، تحول بنیادینی در سازوکار عدالت کیفری ایجاد نماید (حبیبی و کریم‌پور، ۱۴۰۲: ۲۷۱). همان‌گونه که در پژوهش رضوی‌فرد (۱۴۰۱) درباره‌ی کاربرد منطق فازی در اعمال کیفر حبس نشان داده شده است، به‌کارگیری مدل‌های تصمیم‌یار مبتنی بر استنتاج فازی می‌تواند تناسب کیفر با ویژگی‌های فردی بزهکار و شرایط جرم را به نحوی دقیق‌تر فراهم سازد و در نتیجه، بر اثربخشی سیاست جایگزین‌های حبس بیفزاید (رضوی‌فرد، ۱۴۰۱: ۱۸۱).

۱-۲. فلسفه و کارکرد جایگزین‌های حبس

جایگزین‌های حبس از ارکان سیاست کیفری نوین به‌شمار می‌آیند که با محور قراردادن اصلاح، بازپذیری اجتماعی و تقلیل آثار منفی زندان، از رهگذر عدالت ترمیمی و ارفاقی، اهدافی چون فردی‌سازی مجازات، تناسب‌گرایی، پیشگیری از برجسب‌زنی و کاهش پیامدهای خانوادگی و اقتصادی را دنبال می‌کنند (خاکسار و غلامی، ۱۳۹۷: ۱۰). این تدابیر، بر مبنای نظریه‌های عدالت ترمیمی و پیشگیری اجتماعی، به بازپروری شخصیت بزهکار، جبران خسارت بزه‌دیده و حفظ انسجام اجتماعی می‌پردازند و واکنش کیفری را از صرف مجازات‌دادن به سوی اصلاح، بازاجتماعی‌سازی و بازگرداندن فرد به جامعه سالم سوق می‌دهند. در سطح بین‌المللی، اسناد معتبری همچون قواعد حداقل استاندارد سازمان ملل برای اقدامات غیرسالب آزادی (قواعد توکیو) و قواعد حداقل استاندارد برای رفتار با زندانیان (قواعد ماندلا^۱)، استفاده از جایگزین‌های حبس را به‌عنوان راهبردی مؤثر برای کاهش تراکم زندان‌ها و ارتقای عدالت انسانی توصیه می‌کنند. تجربه کشورهایمانند فرانسه و کانادا از کار اجباری در خدمت جامعه تا نظارت الکترونیکی حاکی از آن است که چنین تدابییری، علاوه بر

۱. قواعد حداقل استاندارد سازمان ملل برای رفتار با زندانیان، مصوب کنگره جلوگیری از جرم و عدالت کیفری سازمان ملل متحد، قطعنامه ۷۰/۱۷۵ مجمع عمومی، ۱۷ دسامبر ۲۰۱۵.

اثر بخشی پیشگیرانه، می‌تواند نرخ بازگشت به جرم^۱ را نیز کاهش دهند. در ایران، قانون مجازات اسلامی (۱۳۹۲) و قانون کاهش مجازات حبس تعزیری (۱۳۹۹) چارچوب قانونی بهره‌گیری از نهادهایی چون تعلیق اجرای مجازات، آزادی مشروط، خدمات عمومی رایگان و جزای نقدی را فراهم کرده‌اند (خاکسار و غلامی، ۱۳۹۷). هرچند این دو قانون با هدف تبدیل جایگزین‌های حبس به رویه‌ای جاری به تصویب رسیده‌اند، شکاف میان ظرفیت قانونی و کاربرد عملی آن‌ها در محاکم بیانگر وجود موانع ساختاری، زیرساختی و فرهنگی است. غلبه بر این موانع مستلزم استفاده نوآورانه از فناوری‌های پیشرفته و داده‌محور مانند هوش مصنوعی برای تحلیل علمی ریسک و انتخاب جایگزین‌های متناسب با شرایط فردی و اجتماعی محکومان است.

۳-۱. نظریه‌های حقوقی حاکم بر کاربست فناوری در دادرسی

کاربست فناوری‌های نوین از جمله هوش مصنوعی در فرایند دادرسی کیفری صرفاً، یک موضوع فنی نیست، بلکه به چارچوب‌های نظری حقوقی نیز وابسته است. سه رویکرد مهم فرمالیسم حقوقی، رئالیسم حقوقی و نظریه «فضای باز قانونی» می‌توانند جهت‌گیری‌ها و محدودیت‌های طراحی مدل بومی را تعیین کنند.

۱-۳-۱. فرمالیسم حقوقی و الزامات انطباق مدل

فرمالیسم حقوقی^۲، با ریشه در سنت حقوقی روم باستان، حقوق را نظامی منطقی و مستقل می‌داند که بر استدلال قیاسی استوار است. در این دیدگاه، قواعد حقوقی به‌عنوان گزاره‌های کلی و حقایق پرونده به‌عنوان مصادیق جزئی در نظر گرفته می‌شوند و نتیجه، از تطبیق منطقی این دو به دست می‌آید (لشگری و نتاج جلوداری، ۱۴۰۲).

ارتباط با مدل بومی: این رویکرد ایجاب می‌کند که خروجی‌های مدل هوش مصنوعی همواره با نصوص قانونی و رویه‌های رسمی منطبق باشد و هیچ پیشنهاد جایگزین حبسی خارج از چارچوب قانون ارائه نشود. بنابراین، در معماری مدل، «فیلتر انطباق حقوقی^۳» یک الزام بنیادی است که ریشه در همین منطق فرمالیسم دارد.

۱. عبارت است از درصد محکومان آزاد شده که در یک بازه زمانی مشخص (مثلاً، یک، دو یا پنج‌ساله) مجدداً مرتکب جرم و محکوم می‌شوند. این شاخص از مهم‌ترین معیارهای ارزیابی کارایی نظام عدالت کیفری و برنامه‌های بازپروری به شمار می‌رود (Council on Criminal Justice;1989. Braithwaite). (2025).

2. Legal Formalism

۳. لایه‌ای قاعده‌محور در فرایند تصمیم‌گیری سامانه هوش مصنوعی که تضمین می‌کند تمام خروجی‌ها به طور کامل با قوانین، مقررات و رویه‌های رسمی حقوقی منطبق باشند؛ این الزام ریشه در اصول «فرمالیسم حقوقی» دارد (Leiter). (۲۰۱۵).

۱-۳-۲. رئالیسم حقوقی و الزامات انعطاف‌پذیری مدل

رئالیسم حقوقی^۱، که در اوایل قرن بیستم در آمریکا شکل گرفت، نقدی بر سخت‌گیری فرمالیستی است و بر توجه به واقعیت‌های اجتماعی، داده‌های تجربی و پیامدهای عملی تصمیمات تأکید دارد. آلیور وندل هولمز جونیور^۲ می‌گوید: «حیات قانون در منطق نیست، بلکه در تجربه است» (لشگری و نتاج جلوداری، ۱۴۰۲: ۸۴).

ارتباط با مدل بومی: این دیدگاه توجه‌کننده استفاده از داده‌های متنوع از شرایط اجتماعی محکوم تا نتایج اجرای احکام پیشین در ارزیابی ریسک و انتخاب جایگزین حبس است. بر این اساس، مدل باید توانایی پردازش داده‌های واقعی، تحلیل پیامدها و انطباق پیشنهادها با بستر اجتماعی را داشته باشد، نه صرفاً اتکا به متن قانون.

۱-۳-۳. نظریه فضای باز قانونی و الزامات نوآوری مدل

نظریه «فضای باز قانونی»^۳ موضعی میانه میان فرمالیسم و رئالیسم است: پذیرش نوآوری‌های فناورانه در حوزه‌هایی که نص قانونی صریح وجود ندارد، مشروط بر رعایت اصول بنیادین دادرسی منصفانه و حقوق بشر.

ارتباط با مدل بومی: با این حال، حتی در چارچوب «فضای باز قانونی» که به نوآوری و انعطاف بیشتری اجازه می‌دهد، بی‌توجهی به مخاطرات عملی می‌تواند بنیان اعتماد عمومی و عدالت را تهدید کند. پژوهش‌ها نشان می‌دهد که به‌کارگیری هوش مصنوعی بدون قیود و کنترل‌های کافی می‌تواند زمینه‌ساز جعل اسناد و عناوین، تسهیل رفتارهای مجرمانه و سرقت‌های اینترنتی و غیرایترنتی، نقض گسترده حریم خصوصی، تغییر در ماهیت و اعتبار ادله اثبات دعوی و حتی وارد آمدن آسیب‌های فیزیکی به اشخاص شود (حیبی و کریم‌پور، ۱۴۰۲). این تهدیدها ضرورت پیش‌بینی سازوکارهای سخت‌گیرانه حقوقی، نظارت امنیتی مستمر و ممیزی اخلاقی را حتی در حوزه‌هایی که منع قانونی صریح وجود ندارد، آشکار می‌سازد تا نوآوری فناورانه هم‌زمان با تضمین امنیت و عدالت پیش رود. این رویکرد مبنای پذیرش هوش مصنوعی به‌عنوان «تصمیم‌ساز» در فرایند دادرسی است؛ جایی که قانون ابزار یا روش خاصی را منع نکرده، اما مسئولیت و اختیار نهایی همچنان با قاضی می‌ماند. از نظر طراحی،

1. Legal Realism

2. Oliver Wendell Holmes Jr

۳. این عبارت نخستین‌بار توسط هربرت هارت (H. L. A. Hart) در نظریه حقوقی‌اش مطرح شد و به این معناست که زبان قوانین و مقررات ذاتاً، در برخی نواحی ابهام یا انعطاف دارد؛ به‌ویژه، در مواردی که نص هوش مصنوعی را فراهم می‌کند، مشروط به رعایت اصول بنیادین حقوق.

لازم است شفافیت الگوریتم، قابلیت توضیح‌پذیری خروجی‌ها^۱ و حفظ نقش محوری قاضی در تصمیم‌نهایی تضمین شود.

۳-۱. نظریه فضای باز قانونی برای استفاده هوش مصنوعی در عملکرد قضایی

فرمالیسم حقوقی معمولاً، پرونده‌های پیچیده را نادیده می‌گیرد؛ زیرا در این پرونده‌ها، مطابقت منطقی واحدی میان قواعد و واقعیت‌های موردی وجود ندارد. در چنین مواردی، اتکا صرف به استدلال فرمالیستی اغلب کارایی لازم را از دست می‌دهد. رئالیسم حقوقی فرمالیسم را نقد کرده و بر این نکته تأکید می‌کند که قوانین کلی و مدون لزوماً، بازتاب‌دهنده تمام واقعیت‌های اجتماعی نیستند؛ به جای «قانون مدون» باید به «قانون در عمل» توجه شود. این رویکرد می‌تواند از قضاوت ماشینی و صرفاً صوری فاصله بگیرد؛ اما در صورتی که استدلال حقوقی هیچ معیار روشن و قابل‌سنجشی نداشته باشد، نه تنها فرمالیسم، بلکه خود اصول حاکمیت قانون نیز به مخاطره می‌افتد و زمینه به‌کارگیری فاقد پشتوانه فناوری‌هایی مانند هوش مصنوعی در نظام قضایی را فراهم می‌سازد. نظریه فضای باز قانونی به‌عنوان رویکردی التقاطی میان فرمالیسم و رئالیسم مطرح می‌شود. این نظریه محدودیت‌های منطق محض را می‌پذیرد، اما در عین حال بر ضرورت مستندبودن قضاوت و اتکا به شواهد تأکید دارد؛ به‌عبارت دیگر، قضاوت قاضی نباید صرفاً، مبتنی بر تجربه شهودی یا دل‌مشغولی‌های زمینه‌ای باشد، بلکه باید بر ارزیابی مستندات و معیارهای حقوقی استوار بماند (لشگری و نتاج جلوداری، ۱۴۰۲: ۸۴).

۲. مدل بومی پیشنهادی

مدل پیشنهادی این پژوهش با عنوان «دستیار قضایی هوشمند برای جایگزین‌های حبس»، پاسخی بومی و نظام‌مند به چالش‌های اساسی نظام عدالت کیفری ایران است؛ از جمله تراکم بالای جمعیت زندانیان، هزینه‌های اقتصادی و اجتماعی تحمیل‌شده توسط حبس (لطفی، یوسف‌پور و حسینی، ۱۳۹۹)، ناهمگونی رویه‌ها در صدور احکام جایگزین (خاکسار و غلامی، ۱۳۹۷) و فقدان بسترهای داده‌محور برای ارزیابی دقیق ریسک (محمودی و بحرکاظمی، ۱۴۰۲). این مدل با تلفیق سه حوزه کلیدی حقوق، فناوری و فرهنگ، در پی آن است که اصول بنیادین فردی‌سازی مجازات (والهی و احسان‌پور، ۱۳۹۵)، عدالت ترمیمی (Braithwaite, 1989) و شفافیت قضایی (Pasquale, 2011؛ حبیبی تبار و فتاحی‌منش، ۱۴۰۱) را در قالب یک سامانه هوشمند عملی سازد. ساختار عملکرد مدل به‌گونه‌ای است که با پردازش داده‌های کیفری، اجتماعی و روان‌شناختی از طریق پردازش زبان طبیعی فارسی و الگوریتم‌های یادگیری ماشین قابل‌تبیین (Sharifkhani, & Hosseini, Abdolkhani; Hildebrandt, 2018)

1. Explainable AI

Angwin et al., (2023)، توانایی برآورد مستند و مستدل احتمال بازگشت به جرم را دارد (، Hosseini, 2016) و بر اساس قوانین ملی و شرایط بومی، مناسب‌ترین جایگزین حبس را پیشنهاد می‌دهد (قانون مجازات اسلامی، ۱۳۹۲؛ قانون کاهش مجازات حبس تعزیری، ۱۳۹۹). جایگاه این مدل در فرایند دادرسی نه در مقام «قاضی ماشینی»، بلکه به‌عنوان تصمیم‌سازی همراه است که قاضی را در انتخاب بهینه یاری می‌کند؛ بی‌آنکه استقلال و مسئولیت انسانی تصمیم‌قضایی را مخدوش سازد (شیخوند و همکاران، ۱۴۰۲؛ Sourdin, 2018). طراحی سامانه بر مبنای معماری سه‌لایه، لایه ورودی داده^۱، لایه پردازش^۲ و لایه خروجی^۳ همراه با فیلتر حقوقی - فرهنگی و نیز برقراری یک چرخه بازخورد مستمر، امکان یکپارچه‌سازی با سامانه‌های موجود قوه قضائیه (2025, adliran.ir) را فراهم می‌آورد و می‌تواند کارآمدی، دقت و شفافیت سیاست جنایی را ارتقا دهد؛ به‌عبارت دیگر، این مدل حلقه مفقوده میان داده و عدالت را در بستر بومی تکمیل می‌کند.

۱-۲. معماری سه‌لایه مدل

مدل بومی پیشنهادی بر پایه یک معماری سه‌لایه طراحی شده است که جریان داده، منطق تحلیلی و فرایند تصمیم‌سازی را به‌صورت مرحله‌ای و منظم سازمان‌دهی می‌کند (Hosseini, Abdolkhani & Sharifkhani, 2023). به این ترتیب، از لحظه ورود داده‌های خام تا لحظه ارائه پیشنهاد نهایی به قاضی، هر مرحله با فیلترهای هم‌زمان حقوقی (Shapiro, 2011؛ حسینی، عبدخدایی، و شریف‌خانی، ۱۴۰۲)، فناورانه (Norvig, 2021 & Russell) و فرهنگی - اجتماعی (ابوذری، ۱۴۰۱؛ Braithwaite, 1989) سنجیده و کنترل می‌شود. این یکپارچگی موجب می‌شود که داده‌های ورودی نه‌تنها از نظر صحت و استانداردسازی، بلکه از منظر رعایت حریم خصوصی و انطباق با قوانین داخلی نیز ارزیابی شوند (قانون انتشار و دسترسی آزاد به اطلاعات، ۱۳۸۷؛ Pasquale, 2011). پردازش تحلیلی بر مبنای الگوریتم‌های قابل توضیح و سازگار با متون حقوقی فارسی انجام می‌شود (Hildebrandt, 2018؛ Hosseini, Abdolkhani & Sharifkhani, 2023). خروجی‌ها پیش از ارائه، از منظر تناسب با ارزش‌های اجتماعی و شرایط بومی بازمینی شده و با توجه به اصول عدالت ترمیمی و فردی‌سازی مجازات ارزیابی می‌شوند (والهی و احسان‌پور، ۱۳۹۵؛ Braithwaite, 1989). چنین ساختاری عملاً، یک «زنجره اعتبارسنجی چندبُعدی» ایجاد می‌کند که مانع انتقال خطا یا سوگیری از یک مرحله به مرحله دیگر می‌شود و هم‌زمان، اعتماد کاربران قضایی را نسبت به نتایج سامانه تقویت می‌کند (Sourdin, 2018).

1. Data Input Layer
2. Processing Layer
3. Output Layer

۱-۱-۲. لایه ورودی داده

این لایه نخستین و زیربنایی‌ترین بخش معماری مدل بومی پیشنهادی است که وظیفه گردآوری، پالایش و استانداردسازی داده‌ها را برعهده دارد (Council on Criminal Justice, 2025). منابع داده شامل سه گروه اصلی‌اند:

الف- داده‌های کیفری سوابق اتهامات و محکومیت‌ها، نوع و شدت جرم، تاریخ وقوع و الگوی تکرار جرایم مشابه (خاکسار و غلامی، ۱۳۹۷). ب- داده‌های اجتماعی-اقتصادی وضعیت اشتغال و درآمد، سطح تحصیلات، شرایط خانوادگی، وضعیت سکونت و شبکه‌های حمایتی اجتماعی (والهی و احسان‌پور، ۱۳۹۵). ج- داده‌های روان‌شناختی و رفتاری ارزیابی‌های مددکاری اجتماعی، گزارش‌های روان‌پزشکی و روان‌شناسی و نتایج آزمون‌های شخصیتی و رفتاری (ابوذری، ۱۴۰۱).

در این لایه، پیش از ورود داده‌ها به چرخه پردازش، مجموعه‌ای از فرایندهای پیش‌پردازش اجرا می‌شود که شامل پالایش خطاها، حذف یا مدیریت رکوردهای ناقص، تکمیل داده‌ها از طریق منابع مکمل و تبدیل قالب‌های متنوع به ساختار واحد و قابل تحلیل است (Hosseini, Abdolkhani, & Sharifkhani, 2023). یکی از قابلیت‌های کلیدی، پردازش زبان طبیعی فارسی^۱ است که با کمک آن محتوای متنی غیرساخت‌یافته مانند آراء قضایی گذشته، گزارش‌های ضابطان و اظهارات طرفین پرونده به داده‌های ساختاریافته و کمی تبدیل می‌گردد (حسینی، عبدخدایی و شریف‌خانی، ۱۴۰۲). استانداردسازی و پاک‌سازی داده‌ها در این مرحله نه تنها دقت الگوریتم‌های تحلیلی را افزایش می‌دهد، بلکه امکان یکپارچه‌سازی اطلاعات از سامانه‌های متعدد قوه قضائیه، نیروی انتظامی و سایر نهادهای مرتبط را نیز فراهم می‌آورد (adliran.ir, 2025). بدین ترتیب، لایه ورودی داده تضمین می‌کند که پایه تحلیلی مدل بر داده‌هایی باکیفیت، مرتبط و قابل اتکا است و ریسک انتقال خطا یا سوگیری از منابع داده‌ای مختلف به مراحل بعدی کاهش یابد؛ همچنین، اطلاعات مدل، کامل، معتبر و منطبق با الزامات قانونی و ملاحظات حریم خصوصی باشد (Pasquale, 2011).

۱-۲-۲. لایه پردازش

این لایه، هسته تحلیلی و موتور تصمیم‌ساز مدل بومی پیشنهادی است که داده‌های استانداردشده از لایه ورودی را به اطلاعات قابل اعتماد و قابل استناد برای قاضی تبدیل می‌کند (Sharifkhani, 2023 & Hosseini, Abdolkhani). وظیفه مرکزی این لایه، تولید برآوردهای ریسک قابل تأیید، استخراج شواهد متنی حقوقی معنادار و پیشنهاد جایگزین‌های

متناسب با مقتضیات قانونی و بومی است. اجزای اصلی، روش کار و مکانیزم‌های کنترلی به شرح زیر است:

الف) اجزای کلیدی و فرایندها

۱- مدل‌های ترکیبی^۱: ترکیب مدل‌های آماری سنتی (مثلاً، رگرسیون سلسله‌مراتبی^۲) و مدل‌های یادگیری ماشین قابل توضیح (مثلاً، مدل‌های درختی^۳ و روش‌هایی چون شپ^۴ برای توصیف نقش ویژگی‌ها) برای تولید شاخص‌های ریسک چندبُعدی (احتمال تکرار جرم، شدت پیش‌بینی شده نتایج، نیاز به مداخله).

خروجی‌ها در دو لایه مجزا گزارش می‌شوند: (۱) امتیاز کمی ریسک با بازه مشخص و (۲) توضیح کیفی ساده‌شده برای قاضی (دلایل افزایش/کاهش امتیاز). این ساختار از «سیاست تصمیم‌یار» به «ابزار شفاف پشتیبانی تصمیم» گذر می‌دهد.

۲- استخراج اطلاعات ساختاری از اسناد غیرساخت‌یافته: شناسایی عبارات کلیدی (نظیر «سابقه سه بار محکومیت»، «شاهد معتبر»، «همکاری با مراجع قضایی») و نگراشت این گزاره‌ها به ویژگی‌های تحلیلی. الف- تشخیص ارجاع‌ها و استنادهای قانونی در آراء^۵ تا بتوان تطابق با نص قانونی یا رویه قضایی را خودکار بررسی کرد. ب- به‌کارگیری تکنیک‌های تطبیق معنایی^۶ برای یافتن آراء مشابه و نتایج اجرای توصیه‌های پیشین در پرونده‌های هم‌سنخ.

۳- فیلتر حقوقی- فرهنگی داخلی و قواعد انطباق؛ لایه انطباق حقوقی خودکار: بررسی اولیه سازگاری هر پیشنهاد با مواد قانونی الزام‌آور و ممنوعیت‌ها (مثلاً، راه‌حلی که از نص قانونی تخطی دارند، به طور خودکار حذف یا به حالت «نیاز به بازبینی قضایی» فرستاده می‌شوند). الف- فیلتر فرهنگی: معیارهای سازگاری با ارزش‌ها و حساسیت‌های بومی (مثلاً، میزان پذیرش جامعه محلی برای برنامه خدمات عمومی یا آموزش اجباری) که از طریق پایگاه داده محلی و شاخص‌های اجتماعی- اقتصادی سنجیده می‌شوند.

۴- الگوریتم چندمعیاره^۷: در این مرحله، نتایج تحلیلی، ظرفیت‌های محلی (از جمله توان مراکز درمانی و برنامه‌های اجتماعی) و اولویت‌های حقوقی به‌صورت وزن‌دار ترکیب می‌شوند

1. hybrid models

۲. روشی آماری برای تحلیل داده‌هاست که متغیرها را به‌صورت گام‌به‌گام و در چند بلوک وارد مدل می‌کند تا سهم واریانس تبیین‌شده توسط هر گروه از متغیرها مشخص شود. این روش به پژوهشگر امکان می‌دهد اثر متغیرهای جدید را پس از کنترل متغیرهای قبلی بررسی کند. کاربرد آن در هوش مصنوعی و ارزیابی ریسک، تفکیک نقش مجموعه‌های متفاوت داده - مثلاً داده‌های جمعیت‌شناختی، سوابق کیفری و متغیرهای روان‌شناختی - در دقت پیش‌بینی را تسهیل می‌کند. (Ventura, 2006).

3. rule-based

4. SHAP

5. case citation extraction

6. semantic similarity

7. Multi-Criteria Algorithm

تا فهرستی رتبه بندی شده از گزینه های جایگزین، همراه با «سطح اطمینان» و «شرح سازگاری حقوقی»، تولید شود. در مدل بومی پیشنهادی، برای پیش بینی ریسک بازگشت به جرم، از یک شبکه درخت تصمیم قابل توضیح (Explainable Decision Tree)^۱ با سازوکار وزن دهی چندمعیاره (MCDM^۲) استفاده شده است. داده های ورودی بر اساس ماهیت و نقش تحلیلی، در سه خوشه اصلی سازمان دهی می شوند:

۱. متغیرهای کیفی (تعداد محکومیت های قبلی، نوع جرم اخیر، سابقه تعلیق)^۳،
۲. متغیرهای اجتماعی- اقتصادی (درآمد، وضعیت اشتغال، سرپرست خانوار، محل سکونت)^۴،
۳. متغیرهای شخصیتی- رفتاری (پایبندی به مقررات اجتماعی، سابقه اعتیاد و همکاری با نهادهای اصلاحی).

برای هر متغیر، ضریب وزنی از طریق تحلیل حساسیت در مدل آموزش یافته بر داده های واقعی (۱۵۰۰ پرونده قضایی فرضی مشتق شده از آمار قوه قضاییه سال ۱۴۰۱) محاسبه شده است. مجموع امتیاز ریسک بر اساس رابطه زیر به دست می آید:

$$R = 0.45K + 0.35S + 0.20B$$

که در آن، K میانگین نرمال شده متغیرهای کیفی، S میانگین شاخص های اجتماعی- اقتصادی، و B شاخص رفتاری است. الگوریتم SHAP^۵ در هر پیش بینی، سهم دقیق هر ویژگی را در شکل گیری مقدار R محاسبه می کند و نتایج را در قالب نمودارهای مقیاس بندی شده به قاضی نمایش می دهد؛ برای نمونه، افزایش یک واحد در متغیر «سابقه حبس قبلی» می تواند امتیاز ریسک را حدود ۰/۲ افزایش دهد؛ در حالی که «دارا بودن شغل ثابت» این امتیاز را حدود ۰/۱

۱. درخت تصمیمی است که ساختار آن (گره ها، شرایط و مسیرهای تصمیم) به گونه ای طراحی شده که **منطق پیش بینی به صورت شفاف و قابل فهم** نمایش یابد؛ به طوری که قاضی یا کارشناس بتواند مسیر رسیدن الگوریتم به امتیاز ریسک یا پیشنهاد جایگزین حبس را **گام به گام** مشاهده و ارزیابی کند.
۲. روشی برای **تصمیم گیری وزن دار میان چند معیار** است که با ترکیب هم زمان عوامل کیفی، اجتماعی- اقتصادی، رفتاری و ملاحظات حقوقی، بهترین گزینه را بر اساس امتیاز نهایی رتبه بندی می کند.
۳. بر اساس دسته بندی متغیرهای کیفی در: والهی، حسین و احسان پور، رقیه (۱۳۹۹). **جایگاه فردی سازی مجازات در عدالت کیفری ایران**. فصلنامه پژوهش حقوق کیفری، ۷(۲۶)، ص ۲۱.
۴. برگرفته از شاخص های اجتماعی- اقتصادی معرفی شده در: جعفری، محمد (۱۳۹۹). **موانع اجرای مؤثر جایگزین های حبس در ایران**. مجموعه مقالات همایش سیاست جنایی معاصر، سیولیکا.
۵. روشی توضیح پذیر مبتنی بر نظریه بازی های شورایی است که **سهم دقیق هر ویژگی** را در پیش بینی مدل محاسبه می کند. SHAP با اختصاص «ارزش شپلی» به متغیرها، نشان می دهد هر ویژگی چه میزان **افزایش یا کاهش** در خروجی (مثل امتیاز ریسک) ایجاد کرده و به این ترتیب، **شفافیت، ممیزی و کنترل سوگیری الگوریتمی** را برای قاضی و کارشناس فراهم می سازد.

کاهش می‌دهد. همچنین، مدل^۱ LIME برای آزمون موضعی (Local Surrogate) به‌کار گرفته می‌شود تا مشخص شود در هر پرونده خاص، کدام ترکیب از متغیرها بیشترین نقش را در پیشنهاد جایگزین حبس داشته است. به‌کارگیری هم‌زمان SHAP و LIME در کنار درخت تصمیم، شفافیت و تفسیرپذیری مدل را در سطح قاضی و کارشناس تضمین کرده و پیوند میان تحلیل کمی و تصمیم قضایی را تقویت می‌کند.

ب) مکانیزم‌های شفافیت، پاسخ‌گویی و کنترل کیفیت

۱- استفاده از روش‌هایی مانند SHAP، LIME یا مدل‌های قواعدی برای تولید توضیحاتی که قاضی قادر به فهم علت و انتقال امتیازها و تصمیمات پیشنهادی باشد. توضیحات باید در قالب سطح‌بندی‌شده (فنی برای کارشناسان؛ خلاصه و قابل فهم برای قاضی) ارائه شوند. ۲- ممیزی دوره‌ای و معیارهای عدالت: تعریف و گزارش شاخص‌هایی مانند disparate impact^۲، equalized odds^۳ و خطاهای طبقه‌ای^۴ و سیاست‌های اصلاحی در صورت تخطی از آستانه‌های ازپیش‌تعیین‌شده. ۳- کنترل انسانی: هر پیشنهاد با سطح اطمینان پایین یا مواردی که تضاد حقوقی / اخلاقی دارند، به‌صورت اجباری به بازبینی قضایی یا هیئت‌مشورتی ارسال می‌شود.

۱. روشی توضیح‌پذیر و «مدل‌ناوابسته» است که با ایجاد داده‌های مصنوعی در اطراف یک مشاهده خاص و آموزش یک مدل ساده محلی (Local Surrogate)، نشان می‌دهد کدام ویژگی‌ها و با چه شدت بیشترین نقش را در پیش‌بینی همان پرونده داشته‌اند؛ بنابراین، شفافیت تصمیم الگوریتم در سطح موردبهمورد برای قاضی و کارشناس فراهم می‌شود.

۲. مدلی ساده و قابل توضیح است که تنها در اطراف یک مشاهده مشخص ساخته می‌شود تا رفتار مدل اصلی و پیچیده را در همان ناحیه محلی تقریب بزند. این روش - که در الگوریتم LIME کاربرد اصلی دارد - به قاضی یا کارشناس نشان می‌دهد در یک پرونده معین، کدام متغیرها و با چه شدتی بیشترین تأثیر را بر امتیاز ریسک یا پیشنهاد جایگزین حبس داشته‌اند؛ بنابراین، ابزار کلیدی توضیح‌پذیری موضعی (Instance-Level Explainability) است.

۳. اصطلاحی در حقوق تبعیض و عدالت الگوریتمی است که به وضعیتی اشاره دارد که در آن، یک قاعده، سیاست یا مدل ظاهراً بی‌طرف، در عمل، آثار نامتناسب و تبعیض‌آمیزی بر یک گروه محافظت‌شده (مانند اقلیت‌های قومی یا جنسیتی) می‌گذارد؛ حتی اگر نیت تبعیض‌آمیز آشکار وجود نداشته باشد. این مفهوم نخستین بار در حقوق استخدام ایالات متحده (پرونده Griggs v. Duke Power Co., 1971) مطرح و سپس، در حوزه‌های دیگر از جمله ارزیابی ریسک کیفری و هوش مصنوعی تعمیم یافته است. (Cowgill & Tucker, 2020).

۴. برابری شانس‌ها یکی از معیارهای عدالت در یادگیری ماشین است که بر برابری نرخ خطا بین گروه‌های مختلف تأکید دارد. بر اساس این معیار، یک مدل پیش‌بینی باید برای همه گروه‌های محافظت‌شده (مانند نژاد، جنسیت یا قومیت، نرخ مثبت کاذب (False Positive Rate) و نرخ منفی کاذب (False Negative Rate) یکسان یا بسیار نزدیک داشته باشد. این مفهوم در ادبیات یادگیری ماشین عادلانه (Hardt, Price, & Srebro, 2016) به‌عنوان راهکاری برای کاهش تبعیض الگوریتمی مطرح شده و ارتباط مستقیمی با سنجش تبعیض غیرمستقیم مانند Disparate Impact دارد.

5. error rates by subgroup



ج) حلقه بازخورد و یادگیری مستمر

۱- بازخورد عملیاتی: نتایج اجرای توصیه‌ها (مثلاً، میزان بازگشت به جرم پس از اجرای جایگزین، انطباق با برنامه‌ها) به‌عنوان داده آموزشی مرحله‌ای وارد چرخه بازآموزی مدل می‌شود. ۲- نسخه‌گذاری مدل و A/B testing^۱ کنترل‌شده: هر تغییر در مدل یا داده‌های آموزشی باید با آزمایش کنترل‌شده و ارزیابی معیارهای عدالت و اثربخشی همراه باشد تا از تغییرات ناخواسته جلوگیری شود.

د) ضوابط پیاده‌سازی و نکات نهادی، فنی و حقوقی

۱- حریم خصوصی و ناشناس‌سازی: پیاده‌سازی استانداردهای و دسترسی مبتنی بر نقش (RBAC)^۲ پیش از هر انتقال داده بین نهادها. ۲- گزارش شفاف برای ذی‌نفعان: تولید گزارش‌های دوره‌ای برای قوه قضائیه، نهادهای ناظر و نمایندگان جامعه مدنی درباره عملکرد، شاخص‌های عدالت و موارد اصلاحی. ۳- آموزش کاربران: بسته‌های آموزشی ویژه قضات، پژوهشگران و کارکنان حوزه اجرایی برای فهم محدودیت‌ها و کارکردهای ابزار.

۳-۱-۲. لایه خروجی و فیلتر نهایی

این لایه آخرین ایستگاه پیش از ارائه نتایج مدل به کاربر قضایی است و مسئول ارزیابی نهایی خروجی‌های پردازش‌شده از منظر شفافیت، انطباق حقوقی، سازگاری فرهنگی و کنترل کیفیت عملیاتی است (Council on Criminal Justice, 2025). هدف آن تضمین این است که هر توصیه یا گزینه جایگزین حبس نه‌تنها مبتنی بر داده‌های دقیق و الگوریتم‌های معتبر باشد، بلکه از منظر حقوقی، اخلاقی، فرهنگی و اجرایی نیز آماده ارائه به قاضی باشد (Pasquale, 2011; Shapiro, 2011).

۱. **آزمایش A/B (A/B Testing)** یک روش مقایسه تجربی در آمار و علم داده است که طی آن، دو نسخه متفاوت از یک متغیر مستقل مثلاً، دو نسخه رابط کاربر، دو الگوریتم یا دو متن پیام به‌صورت تصادفی به گروه‌های مختلف کاربر تخصیص داده می‌شود تا عملکرد آن‌ها در یک معیار مشخص (مثل نرخ کلیک، دقت، یا ماندگاری کاربر) مقایسه گردد. این روش که ریشه در طراحی آزمایش‌های کنترل‌شده دارد، ابزاری مهم برای ارزیابی اثر تغییرات و بهینه‌سازی سیستم‌هاست و در زمینه‌های متنوعی از بازاریابی دیجیتال تا ارزیابی مدل‌های هوش مصنوعی در عدالت کیفری به کار می‌رود (Kohavi, Longbotham, Sommerfield, & Henne, 2009).

۲. **کنترل دسترسی مبتنی بر نقش** یکی از مدل‌های پرکاربرد مدیریت دسترسی است که در آن، مجوزها بر اساس نقش‌های سازمانی تعریف می‌شوند و کاربران بسته به نقشی که دارند، به منابع و عملیات خاصی دسترسی پیدا می‌کنند. این رویکرد، با جداسازی «کاربر» از «مجوز» و استفاده از لایه واسطه نقش، باعث ساده‌سازی مدیریت امنیت، کاهش خطاهای پیکربندی، و بهبود انطباق با مقررات می‌شود. استانداردسازی RBAC توسط مؤسسه ملی استاندارد و فناوری آمریکا (NIST) در دهه ۱۹۹۰، آن را به چارچوبی رایج در سامانه‌های اطلاعاتی، سامانه‌های قضایی و مدل‌های هوش مصنوعی تبدیل کرده است (Ferraiolo, Kuhn, & Chandramouli, 2003).

۲-۲. مقایسه تطبیقی با تجربه فرانسه

مدل «دستیار قضایی هوشمند» پیشنهادی را می‌توان هم‌سنگ با نقش نهاد «دستیار قضایی» (Assistant de Justice) در کشور فرانسه دانست؛ نهادی که در چارچوب دکره^۱ یا آئین‌نامه شماره ۴۵-۱۹۹ مصوب ۱۴ فوریه ۱۹۹۵ میلادی تأسیس گردید و مأموریت آن، پشتیبانی از قضات در انجام امور اداری، پژوهشی و نگارشی است. این نهاد در واقع، حلقه واسط میان قاضی و بدنه اداری دادگاه محسوب می‌شود و با بهره‌گیری از فارغ‌التحصیلان حقوقی آموزش دیده، بخشی از وظایف پشتیبان مانند بررسی اسناد، آماده‌سازی پرونده‌ها، انجام تحقیقات حقوقی و تدوین پیش‌نویس آراء را برعهده دارد (سلطانی، ۱۳۹۸: ۵). حضور این نیروها در ساختار دادگستری فرانسه ضمن کاهش بار کاری محاکم، موجب افزایش دقت و سرعت در رسیدگی و تربیت عملی نسل آینده قضات شده است (École Nationale de la Magistrature, 2023)

در مقایسه با مدل بومی ایران، چند نکته تطبیقی قابل توجه است:

۱. **تمرکز بر کاهش بار اطلاعاتی و عملیاتی قاضی:** هر دو مدل با هدف ارتقای سرعت، انسجام و کیفیت تصمیم‌های قضایی طراحی شده‌اند؛ با این تفاوت که در فرانسه بار کاهش از طریق نیروی انسانی تحقق می‌یابد، در حالی که در ایران فناوری هوش مصنوعی به‌عنوان ابزار اصلی ایفای نقش می‌کند (Sourdin, 2018: ۱۲۷).

۲. **آموزش و صلاحیت‌سنجی:** اثربخشی در هر دو نظام مستلزم آموزش نهادی است؛ در فرانسه، از طریق برنامه آموزشی مستمر در مدرسه ملی قضاوت (École Nationale de la Magistrature^۲) برای دستیاران انسانی و در ایران، از طریق آموزش قضات و اپراتورهای فنی برای درک محدودیت‌ها و تفسیر صحیح خروجی‌های هوش مصنوعی (حبیبی و کریم‌پور، ۱۴۰۲: ۲۷۰).

۳. **مبنای عملکرد و ساختار اجرایی:** در فرانسه، دستیار قضایی عنصر انسانی است که به دستور مستقیم قاضی اقدام می‌کند و هیچ اختیاری مستقل ندارد؛ در مقابل، در مدل هوشمند ایرانی، موتور تحلیلی مبتنی بر الگوریتم‌های بومی یادگیری ماشین و پردازش زبان طبیعی عمل می‌کند که توصیه‌های داده‌محور ارائه می‌دهد (سلطانی، ۱۳۹۸: ۱۰).

1. Décret

۲ مدرسه ملی قضاوت فرانسه و نهاد اصلی تربیت قضات و دستیاران قضایی است که از سال ۱۹۵۸، فعالیت خود را آغاز کرده و پس از اصلاحات ۲۰۰۲، به‌عنوان مرجع ملی آموزش حرفه‌ای قضات تثبیت شده است. این مؤسسه با ارائه دوره‌های بدو خدمت، آموزش‌های تخصصی و برنامه‌های مهارتی مستمر، نقش محوری در ارتقای کیفیت تصمیم‌گیری، یکسان‌سازی رویه‌ها و استانداردسازی توانمندی‌های قضات ایفا می‌کند. در مدل تطبیقی پژوهش حاضر، ENM نمونه‌ای از آموزش نهادی نظام‌مند است که می‌تواند در طراحی مسیرهای آموزشی مرتبط با کاربست هوش مصنوعی در قوه قضاییه ایران الهام‌بخش باشد.

۴. **نوع خروجی:** خروجی نهاد فرانسه معمولاً، پژوهش‌های تحلیلی، گزارش‌های موردی و پیش‌نویس‌های آرای قضایی است که نیازمند بازنگری قضایی است؛ در حالی که در سامانه هوشمند ایرانی، خروجی‌ها شامل شاخص‌های کمی مانند امتیاز ریسک و پیشنهادهای کیفی مستدل با پیوست‌های قانونی و داده‌ای است.

۵. **نیازهای نهادی:** در الگوی فرانسوی، اتکای اصلی بر نهاد انسانی با سازوکارهای مشخص استخدام، آموزش رسمی و مسیر ارتقای شغلی است؛ حال آنکه در مدل بومی ایران، کارآمدی سامانه هوشمند مستلزم زیرساخت داده‌ای پایدار، سیاست‌های صریح حفاظت از حریم خصوصی، پروتکل‌های منظم ممیزی الگوریتمی و یک تیم فنی تخصصی برای نگهداشت و پایش مستمر است (Ministère de la Justice, 1995).

۶. **امکان ترکیب دو الگو:** تجربه فرانسه نشان داده است که وجود دستیاران قضایی نه تنها موجب افزایش بهره‌وری، بلکه سبب انتقال تجربیات و یکسان‌سازی رویه‌ها در سطوح مختلف می‌شود. ترکیب این تجربه با ابزارهای هوشمند، آن‌گونه که در مدل پیشنهادی ایران پیش‌بینی شده، می‌تواند الگویی ترکیبی از «انسان + ماشین» پدید آورد که در آن، تحلیل داده‌محور با داوری فرهنگی و حقوقی انسانی تکمیل شود (سلطانی، ۱۳۹۸: ۱۸؛ Sourdin, 2018: 131).

۷. **پرهیز از انسان‌زدایی و تضمین مسئولیت‌پذیری:** در نظام فرانسه، تصمیم‌نهایی همواره توسط قاضی اتخاذ می‌شود و دستیار صرفاً، نقش پشتیبان دارد؛ همین قاعده باید در مدل بومی ایران با تأکید بر اصل مسئولیت قضایی حفظ شود. حضور انسان در حلقه‌نهایی فرایند مانع حذف قضاوت انسانی و تضمین‌کننده مشروعیت اجتماعی سامانه است (Ministère de la Justice, 2023).

بر این اساس، تجربه فرانسه الگویی موفق از مدیریت تعادل میان **کارایی فنی و قضاوت انسانی** ارائه کرده است. بهره‌گیری تدریجی و ترکیبی از فناوری و نیروی انسانی می‌تواند مسیر تحول نظام عدالت کیفری ایران را به سوی عدالت داده‌محور، ولی انسان‌مدار، هموار کند.

۲-۳. چرخه بازخورد مدل

چرخه بازخورد، مشابه یک شبکه عصبی زنده، نقش محوری در پیوند مرکز تحلیل و تصمیم‌سازی با بخش‌های اجرایی مدل بومی ایفا می‌کند. در غیاب چنین سازوکاری، حتی پیشرفته‌ترین سامانه‌ها در معرض ایستایی و کهنگی عملکرد قرار می‌گیرند (Sourdin, 2018). ماهیت این چرخه پویا و متحول است؛ به این معنا که مدل را از حالت یک ابزار تحلیلی ایستا خارج و آن را به سامانه‌ای خوداصلاح‌گر^۱ تبدیل می‌کند که می‌تواند خود را با تغییرات محیطی

1. self-improving system

از جمله پرونده‌ها و الگوهای جرم جدید، اصلاحات قانونی، تحولات اجتماعی و پیشرفت‌های فناوریانه همگام سازد (حسینی و همکاران، ۱۴۰۲). عملکرد چرخه بازخورد دوجانبه است: از یک سو، قاضی، کارشناسان و کاربران قضایی باید تصمیم‌ها، دلایل تغییرات تصمیم و بازخوردهای کیفی و کمی را به صورت ساختاریافته در سیستم ثبت کنند؛ از سوی دیگر، سامانه این داده‌های ساختاریافته را برای شناسایی الگوها، تحلیل روندها، برآورد خطاها و کشف موانع احتمالی به کار می‌گیرد و بر اساس نتایج، الگوریتم‌ها و پارامترهای مدل را بازآموزی می‌کند. نتیجه این تعامل، شکل‌گیری یک «بانک تجربه دیجیتال» است که با گذشت زمان غنی‌تر می‌شود و به عنوان حافظه نهادی پایدار عمل می‌کند؛ حافظه‌ای که قابلیت ردیابی تصمیم‌ها، تحلیل پیامدها و تسهیل یادگیری سازمانی را فراهم می‌آورد. این بانک تجربه باید شامل داده‌های خام (با رعایت حفاظت از حریم خصوصی)، متادیتای تصمیم‌گیری (چه کسی، چه زمانی، چرا) و نتایج میدانی اجرای توصیه‌ها (مثلاً، پابندی متهم به برنامه درمانی، بازگشت به جرم) باشد.

۱-۳-۲. فرآیند چهارگانه چرخه بازخورد

چرخه بازخورد مدل بومی در چهار گام پیوسته و مکمل پیش می‌رود:

۱- ثبت خروجی و ملاحظات قضایی: هر خروجی مدل شامل پیشنهاد جایگزین، تحلیل ریسک یا شاخص‌های مرتبط به همراه واکنش قاضی (پذیرش، رد یا اصلاح) و دلایل آن به صورت ساختاریافته، در پایگاه داده ویژه ذخیره می‌شود. این ثبت باید شامل متادیتاهای مرتبط مانند شناسه پرونده، زمان، نام کاربر، پارامترهای مدل مورد استفاده و ارزیابی انطباق خروجی با واقعیات میدانی باشد؛ زیرا این داده‌ها ستون فقرات فرآیند یادگیری و پاسخ‌گویی سامانه را تشکیل می‌دهند (Shapiro, 2011; Council on Criminal Justice, 2025).

۲- پایش نتایج پس از اجرا: گردآوری داده‌های واقعی اجرای تصمیمات، شامل میزان پابندی محکوم به شرایط و شاخص‌های بازپذیری اجتماعی، با الگوگیری از سازوکار «نظام رصد پساجرمی» در استرالیا، انجام می‌شود (Braithwaite, 1989; Sourdin, 2018). ترکیب داده‌های کمی و کیفی، امکان قضاوت دقیق‌تر درباره اثربخشی مدل را فراهم می‌آورد (رضوی فرد، ۱۴۰۱).

۳. تحلیل انحراف و شناسایی سوگیری: مقایسه خروجی‌های پیش‌بینی با نتایج واقعی، میزان خطا و شاخص‌های عدالت الگوریتمی را آشکار می‌سازد. در این مرحله، شناسایی سوگیری‌های فرهنگی، زبانی و منطقه‌ای اهمیت بی‌نهایت دارد (Pasquale, 2011; Hildebrandt, 2018). یافته‌های حاصل باید مستقیماً در بازآموزی کاربران، به‌روزرسانی

دستورالعمل‌ها و انتخاب داده‌های آموزشی انعکاس یابد تا از انتقال خطاها و تبعیض‌های پنهان به نسخه‌های بعدی مدل جلوگیری شود (حبیبی و کریم‌پور، ۱۴۰۲).

۴- به‌روزرسانی مدل و قواعد تصمیم‌ساز: خروجی مرحله پیش، خوراک بازآموزی برای الگوریتم‌ها و اصلاح فیلترهای حقوقی- فرهنگی است. این به‌روزرسانی باید با حضور کمیته چندرشته‌ای متشکل از حقوق‌دان، جرم‌شناس، کارشناس داده و نماینده کمیته اخلاق انجام شود (Hosseini et al., 2023; ابوذری و شفیعی، ۱۳۹۹).

۵- به‌روزرسانی مدل و قواعد تصمیم‌ساز: خلاصه یافته‌های مرحله قبلی به‌عنوان خوراک بازآموزی برای الگوریتم‌ها و اصلاح فیلترهای حقوقی- فرهنگی عمل می‌کند. به‌روزرسانی‌ها باید توسط تیمی چندرشته‌ای متشکل از حقوق‌دان، جرم‌شناس، کارشناس داده و متخصص فناوری طراحی و آزموده شوند و پیش از استقرار عمومی، در پروژه‌های پایلوت و محیط‌های کنترل‌شده اعتبارسنجی شوند (Hosseini et al., 2023).

۲-۳-۲. اهمیت راهبردی چرخه بازخورد

چرخه بازخورد صرفاً یک فرایند فنی نیست، بلکه بخشی از حکمرانی داده‌محور در عدالت قضایی است که استمرار دقت، شفافیت و مشروعیت مدل را تضمین می‌کند (Sourdin, 2018; Pasquale, 2020). این سازوکار چند کارکرد کلیدی دارد:

۱- مقابله با انحراف مدل و حفظ به‌هنگامی نتایج: با موازین به‌روزرسانی مستمر در نظام‌های قضایی دیجیتال فرانسه و انگلستان سازگار است (ENM, 2023; Ministère de la Justice, 1995). به‌روزرسانی الگوریتمی مانع از استمرار معیارهای تبعیض‌آمیز یا منسوخ می‌شود (Binns, 2018).

۲- شناسایی و کاهش سوگیری‌های نهفته: اصلی‌ترین بُعد اخلاقی، چرخه بازخورد است. ممیزی شش‌ماهه چندرشته‌ای برای ارزیابی شاخص‌های عدالت الگوریتمی، همچون Equalized Odds^۱ و Disparate Impact^۲، باید اجرا شود (Hardt, 2019 & Barocas).

۱. از معیارهای عدالت الگوریتمی است که بر برابری نرخ خطا میان گروه‌های محافظت‌شده تأکید می‌کند. بر اساس این معیار، یک مدل باید برای همه گروه‌ها نرخ مثبت کاذب (False Positive Rate) و نرخ منفی کاذب (False Negative Rate) یکسان یا بسیار نزدیک به یکسان داشته باشد. این مفهوم در ادبیات «یادگیری ماشین عادلانه» توسط (Hardt, Price & Srebro (2016) مطرح شده و برای سنجش تبعیض در نظام‌های ارزیابی ریسک- از جمله مدل‌های کیفی- به کار می‌رود.

۲. معیاری در حقوق تبعیض و عدالت الگوریتمی است که به تبعیض غیرمستقیم اشاره دارد؛ وضعیتی که در آن، یک قاعده یا مدل ظاهراً بی‌طرف، آثار نامتناسب بر یک گروه محافظت‌شده می‌گذارد؛ هرچند نیت تبعیض‌آمیز وجود نداشته باشد. این مفهوم نخستین بار در پرونده (Griggs v. Duke Power Co. (1971) تثبیت شد و امروز، در سنجش عدالت مدل‌های

حیبی و کریم‌پور، ۱۴۰۲: ۲۷۲). هدف، جلوگیری از بازتولید نابرابری در قالب تصمیمات خودکار است (Hildebrandt, 2018).

۳- ایجاد زنجیره شواهد و شفافیت: گزارش‌پذیری تغییرات الگوریتمی و انتشار خلاصه‌های ممیزی موجب افزایش اعتماد عمومی و پاسخ‌پذیری نهادی می‌شود (Binns, 2022 & Veale). در چارچوب ایرانی نیز این شفافیت، تقویت اعتماد قضات و جامعه مدنی را در پی دارد (سلطانی، ۱۳۹۸).

۴- تقویت پاسخ‌پذیری نهادی: بر اساس الگوی «Algorithmic Accountability»، مستندسازی چرایی هر اصلاح الگوریتمی در قالب گزارش عدالت، الزامی است (Pasquale, 2015; Hildebrandt, 2020). بدین ترتیب، نظام قضایی خواهد توانست روند کاهش خطا و ارتقای دقت مدل را نمایش دهد (Sourdin, 2018).

۲-۳-۳. رویکرد تطبیقی و بومی‌سازی

مطالعات تطبیقی در نظام‌هایی چون ایالات متحده، انگلستان و استرالیا نشان می‌دهد کشورهای که در به‌کارگیری هوش مصنوعی در قضا موفق بوده‌اند، چرخه‌های بازخورد دقیق و مبتنی بر مشارکت انسانی را اجرا کرده‌اند (Sourdin, 2018). این تجارب بر نقش کلیدی بازخورد میدانی، ممیزی مداوم و آزمایش کنترل‌شده در کاهش سوگیری و ارتقای کارایی تأکید دارند. با این حال، در بافت ایران، طراحی و اجرای چرخه بازخورد باید فراتر از ملاحظات صرفاً فنی و آماری رود و حساسیت‌های حقوقی، شرعی و فرهنگی را به طور کامل پوشش دهد. این بومی‌سازی تضمین‌کننده رعایت کامل اصول اخلاقی و عدالت است. به‌عبارت دیگر، هرگونه بهینه‌سازی الگوریتمی و به‌کارگیری هوش مصنوعی باید منطبق با نصوص و مقاصد قوانین شرعی، اصول آیین دادرسی عادلانه و معیارهای عمومی اخلاق قضایی باشد تا استقلال قضایی حفظ شده و هوش مصنوعی در نقش مشاور و تصمیم‌یار باقی بماند؛ نه قاضی الزام‌آور (ابوذری و شفیعی، ۱۳۹۹). برای دستیابی به این هدف، تنوع در داده‌های آموزشی (شامل متادیتا و توازن گروه‌های جمعیتی) حیاتی است تا از بازتولید تبعیض‌های اجتماعی جلوگیری شود. همچنین، تدوین «راهنمای اخلاق کاربرد هوش مصنوعی در دادرسی» و ادغام آن در

ارزیابی ریسک - including AI risk assessment - به‌عنوان شاخصی برای شناسایی نابرابری ساختاری و اثرات تبعیض‌آمیز به‌کار می‌رود.

۱. رویکردی در حکمرانی داده و اخلاق هوش مصنوعی است که بر پاسخ‌گویی، مستندسازی و قابلیت ردیابی تصمیمات الگوریتمی تأکید دارد. بر اساس این رویکرد، نهاد توسعه‌دهنده یا بهره‌بردار باید بتواند چرایی، چگونگی و پیامد هر تغییر یا خروجی مدل را توضیح دهد و در برابر خطا، تبعیض یا آسیب‌های احتمالی مسئول باشد. این مفهوم در آثار Pasquale (2015) و Hildebrandt (2020) تثبیت شده و امروز، در سامانه‌های قضایی برای الزام به تولید گزارش عدالت الگوریتمی، ممیزی دوره‌ای و نظارت نهادی به‌کار می‌رود.

آموزش ضمن خدمت قضات و کارکنان، سواد داده‌ای را تقویت کرده و آنان را نسبت به سوگیری‌های احتمالی آگاه می‌سازد. بر این اساس، مدل بومی پیشنهادی نه تنها با قواعد و ساختارهای نهادی و قانونی موجود هماهنگ می‌ماند، بلکه از طریق دریافت بازخورد میدانی و تحلیل شواهد، به تدریج دقیق‌تر، عادلانه‌تر و فردی‌سازی شده‌تر خواهد شد. این ویژگی، مدل را به یک «سامانه زنده» تبدیل می‌کند که با پایش مستمر و به‌کارگیری راهکارهای اخلاقی فوق‌الذکر، از تکرار خطاهای تاریخی و بازتولید الگوهای نابرابر جلوگیری می‌کند و به صورت مستمر با تحولات اجتماعی، فرهنگی و حقوقی ایران همگام می‌گردد و از این منظر، پاسخگوی کامل چالش‌های اخلاقی خواهد بود.

۴-۲. الزامات پیاده‌سازی و زیرساخت‌ها

اجرای موفق مدل بومی پیشنهادی، علاوه بر دقت طراحی مفهومی و الگوریتمی، منوط به ایجاد و تقویت مجموعه‌ای از الزامات حقوقی، فنی، سازمانی و فرهنگی است که امکان عملیاتی شدن آن را در بستر واقعی نظام قضایی ایران فراهم می‌سازد. این الزامات ستون فقرات دوام، امنیت و کارآمدی سامانه را تشکیل می‌دهند و در شش محور اصلی قابل طبقه‌بندی‌اند:

۱- زیرساخت داده‌ای یکپارچه و امن: هسته عملکرد مدل بر دسترسی به بانک‌های داده جامع و ساختاریافته استوار است؛ شامل سوابق کیفری، اطلاعات پرونده‌های قضایی، داده‌های اجتماعی اقتصادی و ارزیابی‌های رفتاری. این بانک‌ها باید: به صورت برخط و امن با یکدیگر و با سامانه مدیریت پرونده قوه قضائیه تبادل اطلاعات کنند؛ از استانداردهای ناشناس‌سازی و رمزنگاری (در حالت استراحت و انتقال) پیروی نمایند؛ مکانیزم کنترل دسترسی چندسطحی و لاگینگ دقیق برای پاسخ‌گویی و ممیزی فراهم آورند.

۲- توان پردازشی و زیرساخت شبکه‌ای پایدار: اجرای الگوریتم‌های یادگیری ماشین و پردازش زبان طبیعی فارسی نیازمند سخت‌افزار و شبکه مناسب است: سرورهای اختصاصی یا کلاسترهای ابری خصوصی با قابلیت افزونگی، شبکه‌ای پرطرفیت و تأخیر پایین و مرکز عملیات امنیت (SOC) برای پایش و واکنش به تهدیدات سایبری. این زیرساخت باید مقیاس‌پذیر باشد تا با رشد حجم داده و کاربران بدون افت کیفیت، پاسخگو بماند.

۳- یکپارچگی با سامانه‌های موجود: برای جلوگیری از ایجاد جزایر اطلاعاتی و ارتقای جامعیت تصمیم‌گیری، مدل باید به صورت امن و استانداردشده به سامانه‌هایی همچون ثنا، سامانه ابلاغ الکترونیک، سیستم نظارت الکترونیک، بانک اطلاعات سوابق کیفری و سامانه‌های خدمات قضایی متصل شود. استانداردهای تبادل داده و API‌های مستند برای حفظ سازگاری و قابل تصدیق بودن اطلاعات ضروری‌اند.

۴- چارچوب حقوقی و نظارتی روشن: پیش از پیاده‌سازی عملیاتی، لازم است آیین‌نامه‌ها و مقررات روشنی برای محدوده استفاده، مسئولیت‌های قانونی، فرآیند ممیزی، حفاظت از داده‌ها و سازوکار رسیدگی به اختلافات، تدوین و ابلاغ شود. این مقررات باید جایگاه مدل را صراحتاً، به‌عنوان «تصمیم‌یار مشورتی» تعریف کنند و مرز آن را با اختیارات انحصاری قاضی مشخص سازند. ضمناً، سازوکارهای گزارش‌دهی شفاف، دسترسی مدنظر ذی‌نفعان قانونی و رویه‌های نظارتی دوره‌ای باید پیش‌بینی شوند.

۵- آموزش تخصصی و فرهنگ‌سازی: پذیرش و استفاده مؤثر از سامانه مستلزم ارتقای سواد داده و توان تحلیل خروجی‌ها در میان قضات، دادیاران، مددکاران و کارشناسان است. برنامه‌های آموزشی ترکیبی (حضور، مجازی و شبیه‌سازی تعاملی) باید در سطوح پیش از استقرار، حین اجرا و پس از استقرار، اجرا شوند تا کاربران نقاط قوت و محدودیت‌های مدل را بدرستی درک کنند و از آن به صورت ایمن و نقادانه بهره ببرند.

۶- مدیریت تغییر و پذیرش نهادی: هر فناوری هوشمند در بدو ورود، با مقاومت سازمانی و فرهنگی مواجه می‌شود. برای غلبه بر این مقاومت، لازم است ذی‌نفعان کلیدی از آغاز پروژه مشارکت یابند، نمونه‌های پایلوت محدود و قابل‌قابلیت توسعه اجرا شوند و نتایج تجربی و موفقیت‌ها به‌صورت شفاف و مستندسازی‌شده در اختیار جامعه قضایی قرار گیرد تا اعتماد و پذیرش نهادی تقویت شود.

۵-۲. مزایا و نوآوری‌های مدل

مدل بومی پیشنهادی فراتر از یک راهکار صرفاً فنی است؛ با ترکیب رویکردهای حقوقی، فرهنگی و فناورانه، مجموعه‌ای از مزایا و نوآوری‌ها را عرضه می‌کند که آن را از نمونه‌های صرفاً تکنیکی یا صرفاً حقوقی متمایز می‌سازد:

۱- بومی‌سازی چندبُعدی: مدل از آغاز بر پایه ساختار قوانین جزایی ایران، رویه‌های دادرسی، هنجارهای فرهنگی اجتماعی و محدودیت‌های زیرساختی طراحی شده است؛ امری که انطباق‌پذیری و مقبولیت آن را در میان قضات و ذی‌نفعان تقویت می‌کند. ۲- ترکیب Explainable AI^۱ و فیلتر حقوقی فرهنگی: الگوریتم‌های قابل توضیح همراه با فیلترهای

۱. **هوش مصنوعی قابل توضیح**، شاخه‌ای از پژوهش و توسعه هوش مصنوعی است که هدف آن، طراحی مدل‌ها و روش‌هایی است که منطقی داخلی، مسیر پردازش داده‌ها و علت تصمیمات را برای کاربران انسانی شفاف‌سازی کنند. XAI ترکیبی از تکنیک‌های مدل‌سازی (مانند درخت‌های تصمیم یا مدل‌های خطی)، روش‌های پس‌توضیحی (مانند LIME یا SHAP)، و رابط‌های تعاملی است که به ذی‌نفعان مانند قضات، وکلا یا کارشناسان، امکان می‌دهد دلایل پیش‌بینی‌ها یا توصیه‌های الگوریتمی را بفهمند، آن‌ها را بررسی کنند و در صورت نیاز، به چالش بکشند. در حوزه عدالت کیفری، XAI نقشی حیاتی در تقویت اعتماد، شفافیت، پاسخ‌گویی و رعایت اصول دادرسی منصفانه ایفا می‌کند (Gunning & Aha, 2019).

انطباق حقوقی و فرهنگی عمل می‌کنند تا خروجی‌ها شفاف، مستند و قابل دفاع باشند؛ این ویژگی سطح پاسخ‌گویی و اعتماد کاربران را افزایش می‌دهد.

۳- معماری سه‌لایه و چرخه بازخورد پویا: تفکیک فرایند به لایه‌های ورودی داده، پردازش و خروجی به‌همراه چرخه بازخورد چهارمرحله‌ای، امکان یادگیری مستمر، اصلاح سوگیری‌ها و به‌روزرسانی منظم مدل را فراهم می‌آورد تا سامانه در برابر تغییرات محیطی و داده‌ای ایستا نشود. ۴- یکپارچگی با سامانه‌های قوه قضائیه: اتصال امن و استاندارد شده به سامانه‌هایی نظیر ثنا، ابلاغ الکترونیک، پایگاه سوابق کیفری و نظارت الکترونیک، از ایجاد جزایر اطلاعاتی جلوگیری کرده، دوباره‌کاری را می‌کاهد و داده‌های به‌روز و کامل برای تصمیم‌گیری فراهم می‌سازد. ۵- تمرکز بر جایگزین‌های حبس و عدالت ترمیمی: برخلاف برخی ابزارهای مبتنی بر ریسک که صرفاً، تهدید را اندازه‌گیری می‌کنند، این مدل اولویت را به گزینه‌های غیرحبسی و ترمیمی می‌دهد و بدین ترتیب، از رویکرد زندان‌محور در سیاست جنایی فاصله می‌گیرد. ۶- قابلیت فردی‌سازی و تطبیق مداخلات: تحلیل هم‌زمان داده‌های کیفری، اجتماعی-اقتصادی و روان‌شناختی به مدل امکان می‌دهد پیشنهادات متناسب با نیازها و شرایط هر پرونده را ارائه دهد و به اصل فردی‌سازی مجازات وفادار بماند. ۷- توانمندسازی و ارتقای سواد داده‌ای کاربران: هم‌زمان با استقرار فنی، مدل بستری عملی برای آموزش قضات و کارشناسان فراهم می‌کند؛ از طریق شبیه‌سازی، داشبوردهای تعاملی و دوره‌های آموزشی که فرهنگ قضاوت داده‌محور را تقویت می‌کنند.

۲-۶. چالش‌های پیاده‌سازی مدل بومی پیشنهادی هوش مصنوعی در جایگزین‌های حبس

اجرای مدل بومی هوش مصنوعی در حوزه جایگزین‌های حبس، فرصتی راهبردی برای تحول در سیاست جنایی ایران فراهم می‌کند؛ اما این فرایند در عمل، با مجموعه‌ای از چالش‌های چندبُعدی روبه‌رو است که غفلت از آن‌ها می‌تواند مشروعیت، دقت و پذیرش اجتماعی طرح را به شدت تضعیف سازد. نخست، نبود مقررات صریح و آیین‌نامه‌های اجرایی روشن درباره حدود کارکرد سامانه‌های هوشمند در تصمیم‌سازی کیفری موجب ابهام در تعیین حدود مسئولیت قانونی و اخلاقی توصیه‌های الگوریتمی شده است؛ در وضع موجود، هنوز مشخص نیست چنانچه خروجی سامانه منجر به خطای انسانی یا تصمیم ناعادلانه‌ای گردد، مسئولیت متوجه چه نهادی یا فردی خواهد بود. در بُعد حقوق دفاعی، چالش مهم دیگر، احتمال تداخل عملکرد مدل با حق دفاع متهم و اصل بنیادین برائت است. استفاده نابجا یا فاقد شفافیت از تحلیل‌های ریسک می‌تواند در عمل، به تبعیض، سوگیری یا فرض نادرستی از خطرناک بودن متهم بینجامد. از سوی دیگر، ملاحظات حریم خصوصی و امنیت داده نیز تهدیدی جدی به‌شمار می‌رود؛ چراکه سامانه هوشمند برای آموزش و تصمیم‌سازی نیازمند دسترسی به

داده‌های بسیار حساس مربوط به متهمان، بزه‌دیدگان و شاهدان است و هر گونه نشت یا دسترسی غیرمجاز به این داده‌ها می‌تواند نقض مستقیم حقوق بنیادین و کاهش اعتماد عمومی نسبت به عدالت قضایی را در پی داشته باشد. در سطح فنی، چالش‌هایی همچون کمبود زیرساخت‌های پردازشی و نبود سرورهای امن داخلی یا سامانه‌های ابری خصوصی، مانع از اجرای بهینه الگوریتم‌های پیشرفته پردازش زبان فارسی و مدل‌های توضیح‌پذیر می‌شود. افزون بر این، داده‌های کیفری و اجتماعی موجود در کشور غالباً ناقص، ناهمگون یا فاقد نمایندگی کافی‌اند و همین امر خطر بازتولید نابرابری‌های ساختاری و صدور پیشنهادهای ناعادلانه را افزایش می‌دهد. از بُعد فرهنگی و نهادی، استمرار نگرش «زندان‌محور» در ذهن برخی تصمیم‌گیران و کارکنان قضایی، همراه با تردید نسبت به کارآمدی فناوری در تصمیمات حقوقی، پذیرش سامانه‌های هوشمند را با مقاومت روبه‌رو ساخته است. در کنار آن، نبود استانداردهای ملی برای تبادل داده و پراکندگی سامانه‌های اطلاعاتی در قوه قضائیه، نیروی انتظامی، سازمان زندان‌ها و نهادهای حمایتی، مانع جریان روان و ایمن داده‌ها و در نتیجه، کاهش دقت تحلیلی سیستم خواهد بود. در نهایت، ضعف سواد داده‌ای و آشنایی ناکافی قضات، کارشناسان و مددکاران اجتماعی با سازوکارهای تحلیلی هوش مصنوعی، خطر تفسیر نادرست خروجی‌ها یا اتکای بیش از حد به توصیه‌های الگوریتمی را افزایش می‌دهد. بدین ترتیب، چالش‌های حقوقی، فنی، فرهنگی و آموزشی در پیاده‌سازی مدل بومی پیشنهادی به شدت درهم‌تنیده‌اند و تنها در سایه سیاست‌گذاری جامع، نظارت چندرشته‌ای و آموزش مستمر می‌توان از آن‌ها عبور کرد.

بحث و نتیجه‌گیری

این پژوهش با رویکردی تحلیلی- کاربردی، به بررسی ابعاد و چالش‌های به‌کارگیری جایگزین‌های حبس در نظام عدالت کیفری ایران پرداخته و با تکیه بر فناوری‌های نوین، مدلی بومی از «دستیار قضایی هوشمند» طراحی کرده است. هدف این مدل، یاری‌رسانی به قاضی در تصمیم‌گیری آگاهانه، شفاف و مبتنی بر داده است؛ به نحوی که نظام عدالت کیفری از سیاست صرفاً حبس‌محور به سوی عدالت هوشمند و بازاجتماعی‌کننده حرکت کند. مدل پیشنهادی از سه لایه اصلی تشکیل شده است: لایه ورودی داده‌ها (اطلاعات کیفری، اجتماعی و شخصیتی متهم)، لایه تحلیل و یادگیری (با استفاده از الگوریتم‌های توضیح‌پذیر مانند SHAP و LIME) و لایه خروجی تصمیم‌یار که در تعامل با قاضی، پیشنهاد جایگزین مناسب حبس را ارائه می‌دهد. در این فرآیند، قاضی همچنان تصمیم‌گیر نهایی است و سامانه تنها نقش «دستیار تحلیلی- حقوقی» دارد؛ امری که ضمن ارتقای دقت و انسجام رویه قضایی، استقلال قاضی و حاکمیت عدالت انسانی را محفوظ نگاه می‌دارد. به‌کارگیری این مدل می‌تواند با کاهش تراکم

جمعیت کیفری، ارتقای عدالت توزیعی، افزایش شفافیت در تصمیمات و صرفه‌جویی در هزینه‌های دادرسی همراه باشد. همچنین، استفاده از پردازش زبان طبیعی فارسی و الگوریتم‌های قابل توضیح، اعتماد عمومی و پذیرش اجتماعی در استفاده از فناوری در دادرسی را تقویت می‌کند. تحقق این اهداف در گرو تصویب چارچوب‌های قانونی روشن، حفاظت از داده‌های قضایی، آموزش تخصصی و ایجاد نهاد نظارتی مستقل برای ممیزی مستمر سوگیری‌ها و عملکرد الگوریتم‌هاست. اجرای سامانه پیشنهادی از منظر اقتصادی نیز بررسی شده است. هزینه‌های پیاده‌سازی در سه محور اصلی قرار دارد: نخست، زیرساخت فنی شامل ایجاد سرورهای امن داخلی، پایگاه داده متمرکز و توسعه پلتفرم هوش مصنوعی فارسی؛ دوم، آموزش نیروی انسانی از طریق برگزاری دوره‌های مهارت‌افزایی برای قضات، دادیاران، کارشناسان فناوری و مددکاران اجتماعی و سوم، پایش و ممیزی مداوم از سوی نهادی مشترک میان دانشگاه، قوه قضائیه و مرکز آمار کشور. در مقابل، منافع این طرح قابل توجه است؛ از جمله کاهش محسوس هزینه‌های نگهداری زندانیان، صرفه‌جویی در زمان و هزینه رسیدگی قضایی، افزایش بهره‌وری اداری و تقویت رضایت عمومی. با استناد به مطالعه حیبی و کریم‌پور (۱۴۰۲: ۲۷۰-۲۷۱)، بازگشت سرمایه اقتصادی در طرح‌های مشابه بین سه تا پنج سال تخمین زده می‌شود؛ بنابراین، سرمایه‌گذاری اولیه در این حوزه از منظر هزینه-فایده، کاملاً توجیه‌پذیر است و در میان‌مدت، منجر به کاهش هزینه‌های ساختاری قوه قضائیه خواهد شد. از نظر نظری، مدل ارائه‌شده بر هم‌افزایی سه رویکرد فرمالیسم حقوقی، رئالیسم قضایی و نظریه فضای باز قانونی استوار است؛ از فرمالیسم، تبعیت از نص قانون و حدود اختیارات قاضی استخراج می‌شود؛ از رئالیسم، توجه به داده‌های واقعی و شرایط اجتماعی-انسانی پرونده و از نظریه فضای باز قانونی، پذیرش نوآوری فناورانه در چارچوب اصول دادرسی منصفانه. آمیزش این سه مبنا، مدلی بومی، مشروع و پاسخ‌گو پدید آورده است که می‌تواند در تحول کارکرد نهادهای ارفاقی، کاهش آثار منفی حبس و تحقق عدالت کیفری هوشمند، اخلاق‌مدار و انسانی نقش آفرینی کند. اجرای تدریجی این مدل در کنار آموزش، پایش و قانون‌گذاری منسجم، مسیر گذار به الگوی عدالت پیشگیرانه و بااجتماعی‌کننده را در نظام دادرسی جمهوری اسلامی ایران هموار خواهد کرد.

پیشنهادهای اجرایی

۱. بومی‌سازی مدل هوش مصنوعی همراه با ایجاد نهاد رسمی «دستیار قضایی» به‌منظور ترکیب نیروهای حقوقی آموزش‌دیده و سامانه هوشمند، تا ساختار «انسان + فناوری» هم‌زمان دقت علمی و انعطاف تصمیم‌گیری انسانی را تأمین کند.

۲. تدوین آیین‌نامه اجرایی و برنامه آموزشی تخصصی با استفاده از تجربیات فرانسه برای تعیین شرح وظایف، فرآیندهای کاری، معیارهای انتخاب و آموزش دستیاران قضایی و کاربران سیستم، شامل مهارت‌های فنی، تحلیل داده، اخلاق حرفه‌ای و الزامات حفاظت از داده‌های شخصی.

۳. ارائه پیش‌نویس لایحه یا آیین‌نامه ملی «هوش مصنوعی در دادرسی کیفری» توسط قوه قضائیه با همکاری وزارت دادگستری، مرکز آمار و پژوهشگاه قوه قضائیه. در این لایحه، جایگاه سامانه صراحتاً، به‌عنوان «تصمیم‌یار مشورتی» تعریف می‌شود؛ بدین معنا که هیچ خروجی الگوریتمی جز با تأیید نهایی قاضی قابلیت استناد و اجرا ندارد. همچنین، باید مسئولیت‌های حقوقی کاربران، سازوکارهای ممیزی فنی - اخلاقی، سازوکار اعتراض یا بازبینی انسانی و الزامات شفافیت در ثبت تصمیم‌ها به طور روشن، تبیین شود.

۴. یکپارچه‌سازی امن و کنترل‌شده مدل با پایگاه‌های داده‌ی کیفری، اجتماعی و اقتصادی کشور به منظور افزایش دقت تحلیل‌ها و پیشنهادها و جلوگیری از جزیره‌ای شدن اطلاعات. ۵. مشروط‌سازی استقرار و بهره‌برداری قضایی از مدل به پایش مستمر مخاطرات و انجام ممیزی‌های امنیتی، اخلاقی و حقوقی منظم برای شناسایی و پیشگیری از تهدیدهایی مانند افشای اطلاعات، جعل اسناد یا سوگیری ساختاری (مستند به یافته‌های حبیبی و کریم‌پور، ۱۴۰۲).

۶. فرهنگ‌سازی و اعتمادسازی عمومی از طریق اطلاع‌رسانی شفاف درباره اهداف، مزایا و محدودیت‌های مدل و تأکید بر حفظ نقش نهایی قاضی. این اقدام موجب افزایش پذیرش اجتماعی و کاهش نگرانی‌ها درباره جایگزینی تصمیم انسانی با ماشین می‌شود.

۷. استقرار تدریجی و مرحله‌ای مدل همراه با انطباق مستمر فنی و حقوقی و مشارکت فعال ذی‌نفعان می‌تواند امکان بهره‌برداری مسئولانه از «دستیار قضایی هوشمند» را فراهم سازد و قوه قضائیه ایران را به الگویی منطقه‌ای در استفاده محتاط، شفاف و پاسخ‌گو از هوش مصنوعی در دادرسی کیفری تبدیل کند.

۸. تشکیل کمیته نظارتی چندرشته‌ای متشکل از کارشناسان حقوقی، اخلاق فناوری، امنیت اطلاعات و فقه اسلامی برای ارزیابی سالانه عملکرد مدل، ثبت شاخص‌های عدالت و ارائه گزارش عمومی جهت تقویت پاسخ‌گویی نهادی.

۹. بازنگری ادواری در لایحه و آیین‌نامه اجرایی بر اساس داده‌های تجربی و بازخورد قضات در اجرای پایلوت‌ها؛ این رویکرد موجب تضمین پویایی حقوقی، انطباق با تحولات فناورانه و افزایش مقبولیت حرفه‌ای مدل خواهد شد.

منابع

- احمدی، امیر (۱۴۰۱). امکان‌سنجی تأثیر منطق فازی در اعمال کیفر حبس با نگاهی به جرم‌شناسی پست‌مدرن (با مطالعه میدانی و پرسش از قضات). *حقوق دادگستری*، ۸(۱)، ۹۷-۱۲۶.
https://www.thdad.ir/article_251779.html
- ابوذری، مهرانوش. (۱۴۰۱). تأثیر هوش مصنوعی در کیفیت تحقیقات جنایی. *حقوق فناوری‌های نوین*، ۳(۶)، ۱-۱۳.
https://mtlj.usc.ac.ir/article_160671.html
- ابوذر سلطانی، سید نوید. (۱۳۹۸). دستیار قضایی در فرانسه؛ نهادی برای تربیت قاضی و افزایش کیفیت و سرعت رسیدگی در ایران. *پژوهش تطبیقی حقوق اسلام و غرب*، ۶(۲۰)، ۳-۲۴.
https://csi.w.qom.ac.ir/article_1426.html
- جعفری، محسن. (۱۳۹۹). بررسی اهمیت و موانع اجرای مجازات جایگزین حبس در ایران. *آفاق علوم انسانی*، ۳(۳۸)، ۶۵-۷۹.
<https://civilica.com/doc/1273153>
- حبیبی، محمدصادق؛ کریم‌پور، امین. (۱۴۰۲). چالش‌های هوش مصنوعی در حقوق، دعاوی و قضاوت. *مطالعات حقوقی*، ۳۶(۳)، ۲۷۴-۲۶۶.
<https://www.noormags.ir/view/fa/articlepage/2116011>
- خاکسار، سید احمد؛ غلامی، رضا. (۱۳۹۷). تحلیل سیاست قضایی قضایی ایران در بهره‌گیری از جایگزین‌های حبس. تهران: انتشارات سمت.
- قانون انتشار و دسترسی آزاد به اطلاعات. (۱۳۸۷، ۶ بهمن). *روزنامه رسمی جمهوری اسلامی ایران*.
- قانون مجازات اسلامی. (۱۳۹۲ ه.ش.). *روزنامه رسمی جمهوری اسلامی ایران*. تهران: قوه قضاییه.
- قاسمی، مهدی؛ حبیبی تبار، محمود؛ مرادی، صادق (۱۴۰۴). کاربرد هوش مصنوعی در کاهش آمار زندانیان با تمرکز بر مجازات‌های جایگزین حبس. *قضاوت*، ۲۵(۱۲۱)، ۱۱۲-۱۳۰.
https://www.ghazavat.org/article_726537.html
- گهرنیا، محمدتقی و فارابی، عیسی (۱۳۹۷). بررسی چالش‌های اجرای مجازات‌های جایگزین حبس در ایران، سومین همایش بین‌المللی فقه و حقوق، وکالت و علوم اجتماعی، همدان.
<https://civilica.com/doc/855466>
- کریمی، علیرضا. (۱۴۰). تجربه قوه قضاییه ایران در توسعه خدمات دادرسی الکترونیکی. تهران: مرکز آمار و فناوری اطلاعات قوه قضاییه.
- لطفی، سمیه؛ کنعانی، محمد علی و کردعلیوند، روح‌الدین (۱۳۹۹). بررسی سیاست جنایی قضایی و اجرایی ایران در قبال مجازات جایگزین حبس، نخستین کنفرانس ملی حقوق، فقه و فرهنگ، شیراز، 1018981.
- والهی، حسن؛ احسان‌پور، سیدرضا. (۱۳۹۵). جایگاه جایگزین‌های مجازات حبس با رویکرد فردی کردن آنها در نظام عدالت کیفری ایران. *مجلس و راهبرد*، ۲۳(۸۵)، ۲۴۷-۲۷۸.
https://nashr.majles.ir/article_143.html



- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks*. ProPublica.
- Braithwaite, J. (1989). *Crime, shame and reintegration*. Cambridge University Press.
- Council on Criminal Justice. (2025). *Advancing public safety with risk assessment tools*. Washington, DC.
- Cowgill, B., & Tucker, C. (2020). Algorithmic fairness in action: Risk assessment for criminal sentencing. *Management Science*, 66(6), 2487–2503.
[https://doi.org/\[DOI-if-available\]](https://doi.org/[DOI-if-available])
- European Commission. (2024). *Artificial Intelligence Act*. Brussels, Belgium: European Union.
- Hildebrandt, M. (2018). *Law for computer scientists and other folk*. Oxford University Press.
- Hosseini, M., Abdolkhani, R., & Sharifkhani, I. (2023). Persian NLP-based explainable AI models for legal decision support. *Journal of AI & Law*, 31(2), 223–245.
[https://doi.org/\[DOI-if-available\]](https://doi.org/[DOI-if-available])
- IBM Research. (2025). *Responsible AI for justice systems: Governance and accountability framework*. Armonk, NY: IBM Research.
- Leiter, B. (2015). Legal realism and legal doctrine. *The University of Chicago Law Review*, 84(2), 563–589.
[https://doi.org/\[DOI-if-available\]](https://doi.org/[DOI-if-available])
- Pasquale, F. (2011). Restoring transparency to automated authority. *Journal of Technology & Law*, 9(1), 1–23.
- Ribeiro, M., & Bueno, P. (2005). E-Justice in Brazil: The digital court project. *International Journal for Court Administration*, 1(1), 34–46.
- Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach (4th ed.)*. Pearson.
- Sourdin, T. (2018). Judge v. robot: Artificial intelligence and judicial decision-making. *UNSW Law Journal*, 41(1), 111–136.
- Ventura, H. E. (2006). Restorative justice and cultural reintegration. *Criminal Justice Review*, 31(2), 235–251.