

ORIGINAL ARTICLE

Balancing Innovation and Human Rights: Legal and Ethical Challenges of Applying Artificial Intelligence in the Criminal Justice System

Fatemeh Ashabi¹, Ali Sadeghinejad²

1. Assistant Professor, Department of Law, Damghan University, Damghan, Iran.

2. Master of Science in Criminology, Faculty of Humanities, Damghan University, Damghan, Iran.

Correspondence:
Fatemeh Ashabi
Email: f.ashabi@du.ac.ir

Received: 06/Oct/2025
Revised: 20/Dec/2025
Accepted: 25/Jan/2026
Published: 11/Sep/2026

How to cite:

Ashabi, F.; Sadeghinejad, A. (2026). Balancing Innovation and Human Rights: Legal and Ethical Challenges of Applying Artificial Intelligence in the Criminal Justice System, Civil Law Knowledge, 15 (1), 43-64.
(DOI: [10.30473/clk.2026.76051.3420](https://doi.org/10.30473/clk.2026.76051.3420))

ABSTRACT

The rapid growth of artificial intelligence (AI) use in legal and judicial fields has posed a dual challenge for legal systems: leveraging this technology while simultaneously safeguarding individual rights. The criminal justice system, as one of the most sensitive areas, requires a balance between technological innovations and adherence to human rights principles. This study, using a descriptive-analytical method, seeks to answer the question of what challenges exist in balancing technology and human rights to achieve justice in the criminal domain. The findings indicate that the use of AI systems, as symbols of innovation and efficiency, necessitates maintaining a balance between reliance on algorithmic data and ensuring transparency, human oversight, and human rights principles. In Iran, documents such as the National AI Strategy (approved 2024), the Seventh Development Plan Law (approved 2023), the Executive Regulation on the Use of Emerging Technologies in Judicial Processes (approved 2025), and the Directive on the Digitalization and Automation of Judicial Processes and Case Files (approved 2024) provide a framework for the responsible use of AI. However, the lack of precise enforcement mechanisms and clear transparency standards increases the risk of violating fundamental rights, such as privacy, the presumption of innocence, and the prohibition of discrimination.

KEYWORDS

Artificial Intelligence, Criminal Justice, Human Rights, Transparency, Oversight.



EXTENDED ABSTRACT

Introduction and Research Innovation

Artificial Intelligence (AI) refers to machine systems capable of automated processes like learning, inference, problem-solving, prediction, and decision-making through data processing (Bagric, Hunter, & Stubbs, 2022). AI's ability to analyze vast datasets offers transformative potential for criminal justice efficiency but raises legal and ethical challenges, including privacy violations, discrimination, and threats to the presumption of innocence. In Iran, AI use in criminal justice is in its early stages, requiring alignment with legal, Sharia, and ethical principles in the Constitution, Criminal Procedure Code (2013), National AI Document (2024), and Seventh Development Plan (2024).

This research innovates by comprehensively analyzing AI applications and challenges in Iran's criminal justice system, emphasizing a balance between innovation and human rights. Unlike prior studies focusing on AI's facilitative role (e.g., Sheikvand et al.), it examines legal and ethical issues using national documents (2025 Regulation, 2024 Directive) and international frameworks like GDPR and Council of Europe guidelines, offering practical regulatory recommendations. Using a descriptive-analytical method with library-based data collection, the study is structured in three sections—AI applications, challenges, and requirements—focusing on transparency, oversight, and human rights.

AI Applications in Criminal Justice

AI's data processing capabilities enable its use in evidence collection, predictive decision-making, and hearing facilitation, requiring compliance with Iran's legal frameworks.

Evidence Collection and Processing: AI supports facial recognition, crime pattern analysis, image identification, and DNA analysis. Global tools like PredPol and HunchLab predict crimes but risk bias without oversight. Forensic AI, such as probabilistic genotyping, aids suspect identification (Gert & Rudin, 2024: 561). European rulings (e.g., *Gaughran v. UK*) warn that prolonged data retention may violate privacy rights (Toumanyants et al., 2023: 147). In Iran, the Criminal Procedure Code (Articles 143, 150, 91, 60) emphasizes privacy, indirectly guiding AI use. The National AI Document promotes judicial AI with Islamic and ethical compliance, and the Seventh Development Plan (2024–2028) supports smart judicial services. The 2019 Dignity Directive regulates data retention for proportionality. However, lacking specific AI regulations risks misuse, privacy breaches, and discrimination.

Predictive Decision-Making: AI assesses crime risk for pretrial detention, sentencing, and parole. Global systems like the UK's Offender Assessment System and Spain's RisCanvi show efficacy but risk false positives/negatives. In the U.S., COMPAS had higher minority error rates (Bagric, Hunter, & Stubbs, 2022: 95). Iran's Criminal Procedure Code (Articles 237, 557) prioritizes human judging. The National AI Document (Clauses 2, 5) and Seventh Plan (Article 114) encourage ethical AI use. The 2019 Prisoner Directive supports predictive tools, but absent regulations risk errors and bias. **Recommendation:** Develop regulations under Article 114 for transparency, with judge training to prevent over-reliance, aligning with Islamic justice (Article 35, Constitution).

Hearings: AI aids judges in legal research and document analysis, acting as an advisory tool. In low-stakes cases, it generates initial documents; in complex cases, it remains assistive (Sourdin, 2018: 1124). The EU (2016) bans autonomous AI decisions. In Iran, the Criminal Procedure Code (Articles 352, 406) emphasizes human judging. The 2022 Data Management Law (Article 5) ensures confidentiality. The National AI Document (Clause 4) and Seventh Plan (Article 114) promote smart systems. The 2019 Dignity Directive prevents full technology reliance. **Recommendation:** Use AI for document analysis with judicial review, regulated via Article 114 for confidentiality and defendant rights.

Legal and Ethical Challenges

AI risks infringing rights like non-discrimination, fair trial, and privacy, addressed in international (e.g., ICCPR) and Iranian documents (2024 Directive, 2025 Regulation, National AI Document, Seventh Plan).

Algorithm Interpretability and Explainability: Interpretability enables understanding decision bases; explainability provides output reasons. "Algorithm aversion" (preferring human decisions) hinders adoption (Bagric, Hunter, & Stubbs, 2022: 758). The U.S. COMPAS case showed opaque algorithms limiting defense rights (Hodges et al., 2020: 13). Iran's 2025 Regulation (Article 3) promotes algorithm training with feedback, and the National AI Document emphasizes transparency. **Challenges:** No source code disclosure or mandatory mechanisms, risking fair trial

violations (Article 34, Constitution). Studies show 40% higher error rates for vulnerable groups. Recommendation: Amend regulations for algorithm disclosure and judge training.

Transparency Legal Frameworks: Transparency requires access to decision factors and oversight for defense rights (Article 35, Constitution). France's 2016 Digital Republic Law ensures information requests (Avramovich & Yovanov, 2023: 161). Iran's 2024 Directive (Article 24) and 2025 Regulation (Article 9) mandate recording proposal changes but lack enforcement, risking defense rights.

Fundamental Rights Protection:

- Privacy: AI surveillance (e.g., facial recognition) threatens privacy but serves crime reduction. Iran's Constitution (Articles 22, 25, 39) and 2025 Regulation (Article 5) protect privacy but lack technical standards (e.g., encryption). GDPR comparison highlights weak "right to explanation."

- Non-Discrimination: AI risks amplifying biases (e.g., COMPAS's 45% minority error rate). Iran's 2025 Regulation (Article 2) and National AI Document aim to prevent bias but lack evaluation protocols, risking discrimination.

- Presumption of Innocence: Predictive tools (e.g., ProKaid) risk "criminalization by association," violating Article 37 (Constitution). The 2024 Directive (Article 22) and 2025 Regulation (Articles 8–9) ensure oversight but lack penalties for violations.

Legal and Ethical Requirements

Fundamental Rights Protection: AI must not violate impartiality, fair trial, privacy, or dignity, per GDPR and AI White Paper (2020).

Data Security and Privacy: Data must be secure. The UK's 2023 Judicial Guidance bans sensitive data in public AI tools. Iran's Data Management Law, National AI Document (Clause 1, Article 5), and Seventh Plan (Articles 65, 113) emphasize security but need implementation.

Non-Discrimination: Algorithms must be unbiased, per EU 2019 guidelines. Iran's 2025 Regulation (Article 2) and Seventh Plan (Article 65) support fairness but need bias evaluation protocols.

Presumption of Innocence: Evidence-less predictions must be avoided. Iran's documents protect dignity but need enforcement.

Transparency and Accountability: Processes must be explainable. France (2016) and Australia (2019) set examples. Iran's 2025 Regulation (Article 3) and National AI Document ensure accountability but need detail.

Human Oversight: Oversight is key for legitimacy (Council of Europe guidelines). Iran's 2025 Regulation (Article 15) and 2024 Directive (Article 34) mandate supervision. Recommendation: Develop real review standards and compensation mechanisms.

In conclusion, AI can enhance Iran's criminal justice efficiency, but specific regulations, judge training, and robust oversight are essential to balance innovation and human rights, integrating international frameworks to mitigate risks. (Approximately 3 pages, ~1800 words, standard font).

«مقاله پژوهشی»

تعادل نوآوری و حقوق بشر: چالش‌های حقوقی و اخلاقی کاربرد هوش مصنوعی در نظام عدالت کیفری

فاطمه اصحابی^۱، علی صادقی‌نژاد^۲

۱. استادیار گروه حقوق، دانشکده علوم انسانی، دانشگاه دامغان، دامغان، ایران.
۲. کارشناس ارشد جزا و جرم‌شناسی، دانشکده علوم انسانی، دانشگاه دامغان، دامغان، ایران.

نویسنده مسئول:

فاطمه اصحابی

ایمانامه: f.ashabi@du.ac.ir

تاریخ دریافت: ۱۴۰۴/۰۷/۱۴

تاریخ بازنگری: ۱۴۰۴/۰۹/۲۹

تاریخ پذیرش: ۱۴۰۴/۱۱/۰۵

تاریخ چاپ: ۱۴۰۵/۰۶/۲۰

استناد به این مقاله:

اصحابی، فاطمه؛ صادقی‌نژاد، علی (۱۴۰۵). تعادل نوآوری و حقوق بشر: چالش‌های حقوقی و اخلاقی کاربرد هوش مصنوعی در نظام عدالت کیفری، دوفصلنامه علمی دانش حقوق مدنی، ۱۵ (۱)، ۴۳-۶۴.

(DOI: 10.30473/clk.2026.76051.3420)

چکیده

رشد شتابان استفاده از هوش مصنوعی (AI) در حوزه‌های حقوقی و قضایی، نظام‌های حقوقی را با چالش بهره‌گیری از این فناوری و همزمان تضمین حقوق افراد مواجه کرده است. نظام عدالت کیفری به‌عنوان یکی از حوزه‌های حساس، نیازمند تعادل بین نوآوری‌های فناورانه و رعایت اصول حقوق بشر است. این پژوهش با روش توصیفی-تحلیلی، در پی پاسخ به این پرسش است که به منظور تعادل بین فناوری و حقوق بشر برای تحقق عدالت در عرصه کیفری چه چالش‌هایی وجود دارد. یافته‌ها نشان می‌دهند که استفاده از سیستم‌های هوش مصنوعی، به‌عنوان نماد نوآوری و کارایی، نیازمند حفظ تعادل بین اتکا به داده‌های الگوریتمی و رعایت شفافیت، نظارت انسانی و اصول حقوق بشر است. در ایران، اسنادی نظیر سند ملی هوش مصنوعی (مصوب ۱۴۰۳)، قانون برنامه هفتم توسعه (مصوب ۱۴۰۳)، آیین‌نامه اجرایی نحوه استفاده از فناوری‌های نوین در فرایندهای قضایی (مصوب ۱۴۰۴)، و دستورالعمل هوشمندسازی و الکترونیکی شدن فرایندها و پرونده‌های قضایی (مصوب ۱۴۰۳) چارچوبی برای استفاده مسئولانه از هوش مصنوعی فراهم می‌کنند. با این حال، فقدان مکانیزم‌های اجرایی دقیق و استانداردهای شفافیت، خطر نقض حقوق بنیادین مانند حریم خصوصی، اصل برائت، و منع تبعیض را افزایش می‌دهد.

واژه‌های کلیدی

شفافیت الگوریتمی، نظارت قضایی انسان‌محور، عدالت پیش‌بینی‌کننده، سوگیری الگوریتمی، حفاظت از حقوق بنیادین.



مقدمه

هوش مصنوعی^۱ به توانایی سامانه‌های ماشینی در انجام فرایندهای خودکار مانند یادگیری، استنتاج، حل مسئله، پیش‌بینی و تصمیم‌گیری از طریق پردازش داده‌ها اطلاق می‌شود (Bagaric, Hunter, & Stobbs, 2022). این فناوری، با قابلیت تحلیل حجم عظیمی از داده‌ها، فرصت‌های بی‌سابقه‌ای برای بهبود کارایی نظام عدالت کیفری فراهم کرده، اما چالش‌های حقوقی و اخلاقی نظیر نقض حریم خصوصی، تبعیض و تهدید اصل برائت را نیز به همراه دارد. در نظام حقوقی ایران، استفاده از هوش مصنوعی در عدالت کیفری در مراحل ابتدایی است و نیازمند تطبیق با اصول حقوقی، شرعی، و اخلاقی مندرج در اسنادی نظیر قانون اساسی، قانون آیین دادرسی کیفری (مصوب ۱۳۹۲) سند ملی هوش مصنوعی (مصوب ۱۴۰۳) و قانون برنامه هفتم توسعه (مصوب ۱۴۰۳) است.

نوآوری این پژوهش در بررسی جامع کاربردها و چالش‌های هوش مصنوعی در عدالت کیفری ایران با تأکید بر تعادل بین نوآوری و حقوق بشر است. برخلاف پژوهش‌های پیشین، مانند «هوش مصنوعی و صدور احکام کیفری» (شیخوند و دیگران) که بر نقش تسهیلگر هوش مصنوعی تمرکز داشتند، این مطالعه چالش‌های حقوقی و اخلاقی را با تکیه بر اسناد ملی (مانند آیین‌نامه ۱۴۰۴ و دستورالعمل ۱۴۰۳) و مقایسه با چارچوب‌های بین‌المللی مانند مقررات حفاظت از داده‌های عمومی^۲ و رهنمودهای شورای اروپا) تحلیل کرده و پیشنهاد عملی برای تنظیم مقررات ارائه می‌دهد. این پژوهش در سه گفتار به بررسی کاربردهای هوش مصنوعی، چالش‌های حقوقی و اخلاقی و الزامات قانونی و نظارتی می‌پردازد. این مطالعه با روش توصیفی-تحلیلی و جمع‌آوری داده‌ها به روش کتابخانه‌ای، فرصت‌های هوش مصنوعی در عدالت کیفری ایران را با تأکید بر شفافیت، نظارت و حقوق بشر بررسی می‌کند.

کاربردهای فناوری هوش مصنوعی در تحقق عدالت کیفری در نظام های حقوقی

با قابلیت پردازش و تحلیل حجم گسترده‌ای از داده‌ها، پتانسیل بالایی برای تحول در نظام عدالت کیفری دارد. این فناوری

می‌تواند در حوزه‌های مختلفی از جمله جمع‌آوری و پردازش شواهد، تصمیم‌گیری‌های پیش‌بینی‌کننده و تسهیل جلسات رسیدگی به کار گرفته شود. با این حال، استفاده از هوش مصنوعی در نظام حقوقی ایران نیازمند تطبیق با اصول حقوقی، شرعی، اخلاقی و همچنین رعایت چارچوب‌های قانونی و آیین‌نامه‌های مربوطه است. در این نوشتار، کاربردهای هوش مصنوعی در نظام عدالت کیفری بررسی شده و ارتباط آن با قوانین داخلی، آیین‌نامه‌های رئیس قوه قضائیه، سند ملی هوش مصنوعی و مواد مرتبط در برنامه هفتم توسعه مورد تحلیل قرار می‌گیرد.

جمع‌آوری و پردازش شواهد

هوش مصنوعی در جمع‌آوری و پردازش شواهد کیفری کاربردهای متعددی دارد، از جمله تشخیص چهره، تحلیل رفتار و الگوهای جنایی، شناسایی تصاویر و اشیاء و تجزیه و تحلیل DNA به عنوان مثال، الگوریتم‌هایی مانند «پردپل»^۴ در لس‌آنجلس از داده‌های تاریخی جرم برای پیش‌بینی وقوع جرایم استفاده می‌کنند. همچنین، الگوریتم «هانکلاب»^۵ با تحلیل سوابق عمومی جرایم و درخواست‌های پلیس، به شناسایی الگوهای جرم کمک می‌کند. این ابزارها، اگرچه در کاهش جرم مؤثرند، اما خطر سوگیری و تبعیض را به همراه دارند؛ به‌ویژه اگر بدون نظارت مناسب استفاده شوند. هوش مصنوعی پزشکی قانونی در ارتباط با ترکیبات پیچیده دی‌ان‌ای به دادگاه‌ها معرفی شده است. الگوریتم‌های طراحی‌شده در این حوزه قادر هستند مخلوط‌های دی‌ان‌ای را که شامل چندین مشارکت‌کننده و در برخی موارد افراد ناشناس است، تحلیل کنند. برای تفسیر نتایج آزمایشی طراحی شده است تا مشخص شود آیا مظنون ممکن است در نمونه‌ای از جرم مشارکت داشته باشد یا خیر. جامعه علمی این رویکردها، که به "ژنوتایپینگ احتمالی" معروف است، را امیدوارکننده ارزیابی کرده است (Garrett & Rudin, 2024: 561). با این حال، آرای دادگاه اروپایی حقوق بشر، مانند پرونده‌های *گراغان علیه بریتانیا* و *مارپر علیه بریتانیا*، نشان می‌دهند که نگهداری طولانی‌مدت داده‌های هویتی (مانند DNA و اثر انگشت) می‌تواند نقض ماده ۸ کنوانسیون اروپایی حقوق بشر (حق حریم خصوصی) باشد، مگر اینکه با اصل تناسب

4. PredPol
5. HunchLab
6. Probabilistic genotyping

1. Artificial Intelligence
2. General Data Protection Regulation (GDPR)
Deoxyribonucleic Acid (مخفف دی‌ان‌ای)

و محدودیت‌های قانونی همراه باشد (Tumanyants, Krytska, & Verkhoglyad-Gerasymenko, 2023: 147). در نظام حقوقی ایران، استفاده از هوش مصنوعی در جمع‌آوری و پردازش شواهد هنوز در مراحل ابتدایی است. قانون آیین دادرسی کیفری (مصوب ۱۳۹۲) در موادی مانند ۱۴۳ (ورود به اماکن و تحقیقات)، ۱۵۰ (استفاده از وسایل شنود)، ۹۱ (محرمانگی تحقیقات) و ۶۰ (بازجویی از متهم) بر حفظ حریم خصوصی و رعایت حقوق شهروندی تأکید دارد. این مواد، اگرچه به صورت صریح به هوش مصنوعی اشاره ندارند، اما چارچوبی برای استفاده مسئولانه از فناوری‌های نوین ارائه می‌دهند. به عنوان مثال، هرگونه استفاده از هوش مصنوعی برای تحلیل داده‌های شخصی باید با رعایت اصل محرمانگی و محدودیت‌های قانونی باشد.

سند ملی هوش مصنوعی (مصوب شورای عالی انقلاب فرهنگی) بر توسعه فناوری‌های هوش مصنوعی با تأکید بر رعایت ارزش‌های اسلامی و اخلاقی تأکید دارد. این سند، استفاده از هوش مصنوعی در بخش‌های عمومی از جمله قوه قضاییه را تشویق می‌کند، اما بر لزوم تدوین چارچوب‌های قانونی برای جلوگیری از نقض حقوق افراد تأکید دارد. همچنین، برنامه هفتم توسعه (۱۴۰۳-۱۴۰۷) در مواد مرتبط با توسعه فناوری‌های نوین؛ به‌ویژه در بندهایی که به هوشمندسازی خدمات عمومی اشاره دارند، زمینه‌ساز استفاده از هوش مصنوعی در دستگاه قضایی است. با این حال، فقدان آیین‌نامه‌های مشخص برای کاربرد هوش مصنوعی در جمع‌آوری شواهد، چالش‌هایی مانند سوءاستفاده، نقض حریم خصوصی و تبعیض را به همراه دارد. دستورالعمل‌های رئیس قوه قضاییه، مانند دستورالعمل حفظ کرامت و ارزش‌های انسانی (مصوب ۱۳۹۸)، بر رعایت حقوق متهمان و حفظ حریم خصوصی تأکید دارند. این دستورالعمل می‌تواند به عنوان مبنایی برای تنظیم استفاده از هوش مصنوعی در پردازش شواهد مورد استفاده قرار گیرد؛ برای مثال، هرگونه نگهداری داده‌های هویتی باید با رعایت اصل تناسب و محدود به زمان و اهداف مشخص باشد تا از نقض حقوق افراد جلوگیری شود.

تصمیم‌گیری‌های پیش‌بینی‌کننده

در نظام‌های عدالت کیفری معاصر، استفاده از ابزارهای مبتنی بر

هوش مصنوعی برای ارزیابی ریسک و پیش‌بینی رفتار مجرمانه به منظور تسهیل تصمیم‌گیری‌های قضایی به طور فزاینده‌ای رواج یافته است. صدها سامانه ارزیابی ریسک در سطح جهانی مورد بهره‌برداری قرار گرفته‌اند که مهم‌ترین کاربرد آنها کمک به اتخاذ تصمیمات کلیدی دادگاه، از جمله تعیین آزادی یا بازداشت پیش از محاکمه، انتخاب نوع و میزان مجازات و تعیین شرایط مجازات‌های مشروط است. این ابزارها که در چارچوب مفهوم «عدالت پیش‌بینی‌کننده» طبقه‌بندی می‌شوند، معمولاً بر پایه تحلیل داده‌های گسترده‌ای مانند سوابق کیفری، رفتارهای پیشین، شرایط اجتماعی-اقتصادی و دیگر عوامل مرتبط بنا شده‌اند.

در برخی کشورهای اروپایی، سامانه‌هایی نظیر «سیستم ارزیابی مجرم» و «مقیاس محکومیت مجدد گروه مجرم» توسعه یافته و به کار گرفته شده‌اند. این سامانه‌ها با هدف ارزیابی احتمال ارتکاب مجدد جرم طراحی شده و در گزارش‌های پیش از صدور حکم و پس از محکومیت به کار می‌روند. به عنوان مثال، سیستم ارزیابی مجرم در تعیین مجازات و تصمیم‌گیری درباره آزادی مشروط، به ویژه برای مجازات‌های کمتر از دو سال، کاربرد دارد. همچنین، «مقیاس محکومیت مجدد گروه مجرم» پس از صدور حکم جهت ارزیابی ریسک تکرار جرم استفاده می‌شود (مطابق تجربه انگلستان و ولز) (Kaplan, Smith, & Brown, 2023: 147). در اسپانیا، سامانه «رایس کانوی» برای ارزیابی ریسک پس از صدور حکم به کار گرفته می‌شود. در اوکراین نیز، برنامه «کاساندر»^۲ با هدف خودکارسازی گزارش‌های پیش از محاکمه توسط افسران آزمایشی توسعه یافته است که می‌تواند در تصمیم‌گیری درباره بازداشت در مرحله محاکمه مؤثر باشد و قابلیت تحلیل تصمیمات قضایی جهت شناسایی موارد ناعادلانه را داراست. علاوه بر این، ابزارهای ارزیابی ریسک آسیب جهت شناسایی مظنونان و پیش‌بینی احتمال ارتکاب مجدد جرم در تصمیم‌گیری‌های جایگزین کیفری، نظیر تعیین اتهام رسمی یا استفاده از روش‌های توافقی خارج از دادگاه، مورد استفاده قرار می‌گیرند (Ibid). در نظام عدالت کیفری ایالات متحده، چهار دسته اصلی سامانه‌های هوش مصنوعی کاربرد دارند: تحلیل امنیت عمومی برای توصیه درباره بازداشت یا آزادی پیش از محاکمه؛ پروفایل مدیریت مجرمان اصلاحی برای تحلیل ریسک ارتکاب مجدد؛ تحلیل ریسک پیش از محاکمه برای ارزیابی خطر

همچنین، دستورالعمل ساماندهی زندانیان و کاهش جمعیت کیفری (مصوب ۱۳۹۸، رئیس قوه قضاییه) در بندهای ۶ و ۸، زمینه را برای استفاده از ابزارهای پیش‌بینی‌کننده در ارزیابی شرایط آزادی مشروط یا بازداشت موقت فراهم می‌کند. این دستورالعمل بر کاهش جمعیت کیفری با رعایت عدالت و حقوق متهمان تأکید دارد، که می‌تواند با ابزارهای هوش مصنوعی هماهنگ شود. با این حال، فقدان آیین‌نامه‌های مشخص برای استفاده از هوش مصنوعی در تصمیم‌گیری‌های پیش‌بینی‌کننده، خطراتی مانند خطاهای الگوریتمی، تبعیض و نقض حقوق متهمان را به همراه دارد. دستورالعمل هوشمندسازی و الکترونیکی شدن فرایندهای قضایی (مصوب ۱۴۰۰) در بند ۳ بر ضرورت تدوین چارچوب‌های قانونی برای استفاده از فناوری‌های نوین تأکید دارد و می‌تواند به‌عنوان مبنایی برای تنظیم این ابزارها عمل کند.

برای بهره‌گیری مسئولانه از ابزارهای پیش‌بینی‌کننده در ایران، بر اساس ماده ۱۱۴ برنامه هفتم توسعه و ماده ۳ دستورالعمل هوشمندسازی، آیین‌نامه‌ای برای تنظیم استفاده از الگوریتم‌های پیش‌بینی‌کننده تدوین شود که شفافیت، نظارت قضایی و رعایت اصل تناسب را تضمین کند. قضات باید در تحلیل خروجی‌های الگوریتمی آموزش ببینند تا از سوگیری و وابستگی بیش از حد به فناوری جلوگیری شود و تصمیم‌گیری‌های مبتنی بر هوش مصنوعی باید با اصول عدالت اسلامی و حفظ کرامت انسانی (مطابق اصل ۳۵ قانون اساسی و بند ۲ سند ملی هوش مصنوعی) هماهنگ باشد.

جلسات رسیدگی

سامانه‌های مبتنی بر هوش مصنوعی می‌توانند به قاضی در جست‌وجوی منابع حقوقی مرتبط کمک کنند و نتایج بازایی‌شده را به صورت اولویت‌بندی شده ارائه دهند. همچنین، این فناوری با ایجاد ساختارهایی مبتنی بر پیوند میان متون و اطلاعات، قابلیت ورود به مرحله تحلیل و نتیجه‌گیری را نیز دارد (Sourdin, 2018: 1124). خروجی این سامانه‌ها می‌تواند نقش مشاوره مؤثر را برای قاضی ایفا کرده و فرایند تصمیم‌گیری را تسهیل کند. در برخی نظام‌های حقوقی که ابتدا نسبت به این فناوری محتاط بودند، پس از مشاهده عملکرد قابل اعتماد آن در تولید مستندات و تحلیل‌های با کیفیت، اجازه مشارکت بیشتر هوش مصنوعی حتی در پرونده‌های با

متهمان و ارزیابی ریسک پس از محکومیت که به عنوان پیشرفته‌ترین سامانه نسل چهارم، علاوه بر عوامل ایستا مانند سابقه کیفری، متغیرهای پویا نظیر وضعیت شغلی، تحصیلات و روابط خانوادگی را نیز لحاظ می‌کند و در تصمیمات مربوط به آزادی مشروط کاربرد دارد (Bagaric, Hunter, & Stobbs, 2022: 95). شایان ذکر است در فرایندهای پیش‌بینی ریسک، دو نوع خطای اصلی ممکن است رخ دهد: خطای مثبت کاذب، هنگامی که فرد به اشتباه پرخطر تشخیص داده شده و محدودیت‌های غیرضروری بر وی اعمال می‌شود و خطای منفی کاذب، زمانی که فرد کم‌خطر محسوب شده، اما پس از آزادی مرتکب جرم می‌شود. اکثریت این سامانه‌ها از نوع «شفاف» هستند که در آنها نحوه امتیازدهی عوامل به روشنی قابل بررسی است. این ویژگی اهمیت فراوانی دارد؛ زیرا به‌رغم نقش اطلاع‌رسانی این ابزارها، اختیار نهایی در تصمیم‌گیری همچنان در دست مراجع قضایی باقی می‌ماند (Ibid).

مطالعات میدانی در برخی ایالت‌های آمریکا نشان داده است که بهره‌گیری از الگوریتم‌های ارزیابی ریسک نسبت به قضاوت سنتی قضاات منجر به بهبود ثبات و انصاف در تصمیمات وثیقه شده است. برای نمونه، در ویرجینیا و نیوجرسی، استفاده از این سامانه‌ها باعث افزایش تعداد متهمان آزاد شده بدون افزایش جرائم پیش از محاکمه و کاهش چشمگیر جمعیت زندانیان پیش از محاکمه گردیده است. با این حال، تجربه کنتاکی نشان داده که استفاده اجباری از الگوریتم‌ها می‌تواند منجر به تبعیض‌های نژادی گردد (Ibid). در ایران، استفاده از ابزارهای پیش‌بینی‌کننده هوش مصنوعی در عدالت کیفری در مراحل ابتدایی قرار دارد. قانون آیین دادرسی کیفری (مصوب ۱۳۹۲) در مواد ۲۳۷ (بازداشت موقت) و ۵۵۷ (قرار تعویق، تعلیق و آزادی مشروط) بر نقش قاضی انسانی در تصمیم‌گیری تأکید دارد و هیچ اشاره‌ای به استفاده از هوش مصنوعی ندارد. با این حال، سند ملی هوش مصنوعی (۱۴۰۳) در بندهای ۲ و ۵، توسعه فناوری‌های هوش مصنوعی در بخش‌های عمومی، از جمله قوه قضاییه، را با رعایت اصول شرعی و اخلاقی تشویق می‌کند. برنامه هفتم توسعه (۱۴۰۳-۱۴۰۷) در ماده ۱۱۴ بر هوشمندسازی خدمات قضایی و الکترونیکی شدن فرایندها تأکید دارد و استفاده از فناوری‌های نوین را برای افزایش کارایی و شفافیت قضایی ترویج می‌کند.

اهمیت کمتر صادر شده است. این روند می‌تواند به تدریج منجر به افزایش نقش هوش مصنوعی در محاکم عالی گردد، مشروط بر آنکه این سامانه‌ها آزمون‌های سخت‌گیرانه‌ای را پشت سر گذاشته و توانایی تحلیل موضوعات پیچیده را کسب کنند (Ming, 2022: 36). با این وجود، در دعاوی مهم و پیچیده، نقش هوش مصنوعی باید محدود به کمک‌کننده و تسهیل‌گر باشد و تحت نظارت مستقیم قضات قرار گیرد تا قاضی بتواند خروجی سامانه را با دیدگاهی انتقادی ارزیابی کند. این رویکرد در مقابل نگرانی‌هایی نظیر سوگیری یا اتکای مطلق به فرایندهای خودکار، به عنوان تدبیری حفاظتی برای دستگاه قضا مطرح می‌شود (Danielle, 2008: 1272). هوش مصنوعی می‌تواند با تحلیل مستندات، جست‌وجوی منابع حقوقی و ارائه تحلیل‌های اولویت‌بندی‌شده، نقش مشاوره‌ی مؤثر برای قضات ایفا کند. در پرونده‌های کم‌اهمیت، این فناوری برای تولید مستندات اولیه استفاده می‌شود، اما در دعاوی پیچیده، نقش آن به ابزار کمکی محدود است. مصوبه پارلمان اروپا (۲۰۱۶) توسعه‌دهندگان را ملزم می‌کند که سامانه‌های هوش مصنوعی به‌صورت کاملاً خودکار تصمیم‌گیری نکنند، بلکه تحت نظارت انسانی عمل کنند.

در ایران، استفاده از هوش مصنوعی در جلسات رسیدگی در مراحل ابتدایی است. قانون آیین دادرسی کیفری (مصوب ۱۳۹۲) در مواد ۳۵۲ (علنی بودن جلسات) و ۴۰۶ (حضور بودن رسیدگی) بر نقش قاضی انسانی و اصول دادرسی عادلانه تأکید دارد. قانون مدیریت داده‌ها و اطلاعات ملی (مصوب ۱۴۰۱) در ماده ۵ بر حفظ محرمانگی اطلاعات پرونده‌ها تأکید می‌کند که در استفاده از هوش مصنوعی در تحلیل مستندات اهمیت دارد. سند ملی هوش مصنوعی در بند ۴، استفاده از این فناوری به‌عنوان ابزار کمکی در بخش‌های قضایی را تشویق می‌کند، مشروط بر نظارت قضایی و رعایت اصول اخلاقی. برنامه هفتم توسعه در ماده ۱۱۴ (بند ب) بر توسعه سامانه‌های هوشمند برای مدیریت پرونده‌ها و تحلیل مستندات تأکید دارد. دستورالعمل حفظ کرامت و ارزش‌های انسانی (مصوب ۱۳۹۸) در ماده ۴ بر رعایت حقوق متهمان و جلوگیری از وابستگی کامل به فناوری تأکید می‌کند. همچنین، دستورالعمل هوشمندسازی و الکترونیکی شدن فرایندهای قضایی بر توسعه ابزارهای هوش مصنوعی برای تسریع رسیدگی و حفظ محرمانگی داده‌ها تأکید دارد. برای استفاده از هوش مصنوعی در جلسات رسیدگی در ایران، نقش

هوش مصنوعی باید به‌عنوان ابزار کمکی برای تحلیل مستندات و جست‌وجوی منابع حقوقی استفاده شود. تمام خروجی‌های هوش مصنوعی باید توسط قاضی انسانی بررسی شود و آیین‌نامه‌ای بر اساس ماده ۱۱۴ برنامه هفتم توسعه تدوین شود تا استفاده از هوش مصنوعی در محاکم را با تأکید بر محرمانگی (ماده ۵ قانون مدیریت داده‌ها و اطلاعات ملی) و حقوق متهمان تنظیم کند. هوش مصنوعی می‌تواند با هوشمندسازی فرایندهای قضایی، کارایی و انصاف در نظام عدالت کیفری ایران را بهبود بخشد. سند ملی هوش مصنوعی و برنامه هفتم توسعه زمینه‌ساز توسعه این فناوری هستند، با این حال، تدوین آیین‌نامه‌های مشخص، آموزش قضات و نظارت قضایی دقیق برای جلوگیری از خطاها و رعایت اصول شرعی و حقوقی ضروری است.

چالش‌های حقوقی و اخلاقی فناوری هوش مصنوعی در تحقق عدالت کیفری: نگاهی به چارچوب‌های ملی و بین‌المللی

بررسی استفاده از هوش مصنوعی در دادرسی کیفری نشان می‌دهد که این فناوری می‌تواند برخی از حقوق بنیادین انسان را که در اسناد بین‌المللی مانند کنوانسیون اروپایی حقوق بشر و میثاق بین‌المللی حقوق مدنی و سیاسی ذکر شده است، تحت تأثیر قرار دهد. برخی از این تأثیرات عبارت‌اند از: نقض اصل منع تبعیض، نقض حق دادرسی عادلانه (از جمله اصل برائت و رسیدگی توسط دادگاهی مستقل و بی‌طرف) و نقض حق حریم خصوصی. این چالش‌ها که در ادامه مورد بررسی قرار خواهند گرفت، نیازمند تحلیل علمی دقیق هستند تا تعادل میان مزایای فناوری و حفاظت از حقوق بشر برقرار شود. در حقوق ایران، این موضوع با تلفیق دستورالعمل هوشمندسازی و الکترونیکی شدن فرایندها و پرونده‌های قضایی (مصوب ۱۴۰۳)، آیین‌نامه اجرایی نحوه استفاده از فناوری‌های نوین در فرایندهای قضایی (مصوب ۱۴۰۴)، سند ملی هوش مصنوعی جمهوری اسلامی ایران (مصوب ۱۴۰۳) و مواد مرتبط قانون برنامه پنج‌ساله هفتم پیشرفت جمهوری اسلامی ایران (مانند بند الف ماده ۱۱۳) قابل مدیریت است. این اسناد، بر اساس سند تحول و تعالی قوه قضاییه و سیاست‌های کلی نظام، بر استفاده از هوش مصنوعی به‌عنوان ابزار تصمیم‌یار، اما با حفظ اختیار قاضی، نظارت انسانی و رعایت اصول اخلاقی مانند شفافیت و امنیت داده‌ها (ماده ۵ آیین‌نامه ۱۴۰۴، ماده ۱۸ دستورالعمل ۱۴۰۳، بند ۱ ماده ۵ سند

که افراد، پس از مشاهده یک اشتباه از سوی سیستم هوش مصنوعی، سریع‌تر اعتماد خود را نسبت به آن از دست می‌دهند تا در برابر اشتباهات انسانی. از نظر منطقی توجیه‌پذیر نیست، اما به عنوان یک واقعیت روان‌شناختی باید مورد توجه قرار گیرد. از این‌رو، هرگونه پیشنهاد مبنی بر ادغام هوش مصنوعی در نظام عدالت کیفری باید این پدیده را در نظر گرفته و راهکارهایی برای کاهش این نگرانی‌ها ارائه دهد (Hunter, Bagaric, & Stobbs, 2020: 758). در پرونده‌ای در آمریکا با موضوع درخواست تجدیدنظر در استفاده از ابزار ارزیابی ریسک کومپاس^۱ بود. متهم به جرایم کوچک ناشی از یک تیراندازی از خودرو اعتراف کرده بود. در آماده‌سازی برای صدور حکم، دولت تخمینی از خطر ارتکاب جرم مجدد را بر اساس تجزیه و تحلیل کومپاس از مصاحبه با متهم و اطلاعاتی از تاریخچه جنائی او ارائه داد. متهم به حبس محکوم شد و استفاده از این ابزار را بر اساس نقض حقوق اساسی خود به چالش کشید و ادعا کرد که حقوق او برای صدور حکم بر اساس اطلاعات دقیق و منحصر به فرد نقض شده است. به ویژه، متهم ادعا کرد که چون الگوریتم به صورت یک جعبه سیاه است، او قادر به چالش کشیدن نحوه رسیدن آن به نتیجه‌گیری‌هایش نبود. دادگاه بدوی دفاع متهم را رد کرد و دادگاه عالی ویسکانسین^۲ این حکم را تأیید کرد. اگرچه سوالاتی درباره این ابزار مطرح شد، دادگاه عالی ویسکانسین با احتیاط استفاده از آن را تأیید کرد و اشاره کرد که ارزیابی تنها یکی از عوامل در تصمیم‌گیری قضایی است که قاضی می‌تواند آن را نادیده بگیرد. علاوه بر این، دادگاه تأکید کرد که "پیامدهای حقوق اساسی ما را وادار می‌کند که به دادگاه‌های حوزه قضایی هشدار دهیم که چون نمرات ارزیابی ریسک کومپاس بر اساس داده‌های گروهی است، این ارزیابی‌ها قادر به شناسایی گروه‌های مجرم با ریسک بالا هستند، نه یک فرد خاص با ریسک بالا (Hedges, Gottehrer & Francis, 2020: 13). در نتیجه در این پرونده، متهم ممکن است نداند چه اطلاعاتی در هنگام ارزیابی ریسک توسط الگوریتم در نظر گرفته می‌شود. دوم، متهم تقریباً قطعاً نمی‌داند که الگوریتم چگونه داده‌های مختلف را وزن‌دهی می‌کند تا دفاعیات خود را ارائه کند. در حالی که یک متهم ممکن است درخواست کشف الگوریتم را مطرح کند.

ملی هوش مصنوعی). تأکید دارند. با وجود این، بدون شفافیت الگوریتم‌ها، این فناوری می‌تواند سوگیری‌های اجتماعی را تقویت کند، همان‌طور که مطالعات نشان‌دهنده نرخ خطای بالاتر در گروه‌های اقلیت است.

تفسیرپذیری و توضیح‌پذیری الگوریتم‌ها (با نگاه چالشی به قوانین، مقررات و دستورالعمل‌ها)

تفسیرپذیری به توانایی کاربران در درک مبنای تصمیم‌گیری سیستم اشاره دارد، در حالی که توضیح‌پذیری به قابلیت ارائه دلایل روشن برای خروجی‌های سیستم، حتی پس از تولید نتایج، مربوط می‌شود. یکی از دلایلی که پژوهشگران را به سمت مطالعه فناوری‌های هوش مصنوعی در دادرسی کیفری سوق می‌دهد، غیرقابل پیش‌بینی بودن تصمیمات قضایی؛ به‌ویژه در مورد دادرسی با حضور هیئت منصفه است. طبق قانون آیین دادرسی کیفری فدراسیون روسیه، اعضای هیئت منصفه ملزم به توضیح رأی خود نیستند. این امر امکان ارزیابی میزان ارتباط و استحکام تصمیم آنها را از طرفین دعوی سلب کرده و باعث می‌شود که درک روشنی از مبانی قضاوت هیئت منصفه وجود نداشته باشد. به همین دلیل، تصمیمات هیئت منصفه اغلب خودسرانه تلقی می‌شود (Spiridonov, 2023: 481)، اما مانع مهم در پذیرش هوش مصنوعی در فرایندهای قضایی، تمایل ذاتی انسان به ترجیح تصمیم‌گیری انسانی بر تصمیم‌گیری مبتنی بر رایانه است. افراد معمولاً اشتباهات و خطاهای انسانی را آسان‌تر می‌پذیرند، اما در برابر خطاهای ناشی از سیستم‌های رایانه‌ای حساسیت بیشتری نشان می‌دهند. این پدیده که تحت عنوان گریز الگوریتمی شناخته می‌شود، مانعی مهم در پذیرش هوش مصنوعی در فرایندهای قضایی به شمار می‌رود. مطالعات نشان داده‌اند که الگوریتم‌های مبتنی بر داده‌های تجربی، پیش‌بینی‌های دقیق‌تری نسبت به تحلیل‌های انسانی ارائه می‌دهند. با این حال، هنگامی که افراد باید بین پیش‌بینی‌های انسانی و پیش‌بینی‌های آماری الگوریتمی یکی را انتخاب کنند، غالباً تحلیل انسانی را ترجیح می‌دهند. این پدیده، که ناشی از گریز الگوریتمی است، هزینه‌های قابل توجهی را به همراه دارد و درک علل آن حائز اهمیت است. حتی اگر این الگوریتم‌ها عملکرد بهتری نسبت به پیش‌بینی‌های انسانی داشته باشند. این امر به این دلیل است

در ایران، ماده ۳ آیین‌نامه اجرایی نحوه استفاده از فناوری‌های نوین در فرایندهای قضایی (مصوب ۱۴۰۴) بر ارتقای مستمر دقت و کارایی پیشنهاد سامانه از طریق آموزش مکرر با استفاده از بازخوردهای دریافتی از کاربران تأکید دارد، که این امر مستقیماً به توضیح‌پذیری الگوریتم‌ها کمک می‌کند؛ زیرا بازخوردها می‌توانند نقاط کور جعبه سیاه را شناسایی کنند. همچنین شاخص‌های ارزیابی سند ملی هوش مصنوعی جمهوری اسلامی ایران (مصوب ۱۴۰۳) بر قابلیت تفسیر و توضیح‌پذیری مدل‌ها تمرکز دارد که این شاخص‌ها می‌توانند با اندازه‌گیری میزان شفافیت الگوریتم‌ها (مانند استفاده از مدل‌های جعبه شیشه‌ای)، چالش‌های روان‌شناختی مانند گریز الگوریتمی را کاهش دهند. این رویکردها نشان می‌دهد که ترکیب بازخورد انسانی با شاخص‌های ارزیابی می‌تواند دقت پیش‌بینی‌ها را تا ۲۰-۳۰ درصد افزایش دهد، اما بدون اجرای عملی، خطر عدم پذیرش فناوری توسط قضات باقی می‌ماند (Hunter, Bagaric, & Stobbs, 2020: 758).

Garrett & Rudin, 2024: 561). با این حال، نگاه چالشی به این قوانین و مقررات نشان‌دهنده نقاط ضعفی است که می‌تواند کارایی آنها را زیر سؤال ببرد. نخست، آیین‌نامه ۱۴۰۴، هرچند بر آموزش مکرر و بازخورد تأکید دارد، اما فاقد مکانیسم‌های الزامی برای افشای کد منبع الگوریتم‌ها یا دسترسی متهمان به جزئیات وزن‌دهی داده‌ها - مشابه چالش‌های پرونده کمپاس در آمریکا است. این کمبود می‌تواند منجر به نقض اصل دادرسی عادلانه (اصل ۳۴ قانون اساسی ایران) شود؛ زیرا متهمان قادر به چالش کشیدن خروجی‌های سامانه نیستند و این امر خطر خودسرانگی تصمیمات را افزایش می‌دهد. بر اساس مطالعات اخیر، نرخ خطای الگوریتم‌ها در گروه‌های آسیب‌پذیر تا ۴۰ درصد بالاتر است، که این با تأکید آیین‌نامه بر "حفظ مسئولیت شخصی قاضی" ناسازگار است؛ زیرا قاضی بدون دسترسی به جزئیات نمی‌تواند نظارت مؤثری اعمال کند. دوم، سند ملی هوش مصنوعی ۱۴۰۳، با وجود شاخص‌های ارزیابی، در اجرای عملی چالش‌برانگیز است؛ تا سال ۱۴۰۴، ستاد هوش مصنوعی تشکیل شده، اما دوگانگی سیاست‌گذاری (مانند تصویب سند جداگانه در هیئت وزیران) می‌تواند منجر به ناهماهنگی شود، و تأکید بر "زیست‌بوم نوآورانه" بدون مقررات سخت‌گیرانه برای توضیح‌پذیری، خطر

گریز الگوریتمی را تشدید می‌کند - پدیده‌ای که در ایران، با توجه به فرهنگ قضایی سنتی، می‌تواند پذیرش فناوری را به تأخیر بیندازد. همچنین، مقایسه با استانداردهای بین‌المللی مانند رهنمودهای اتحادیه اروپا (۲۰۱۹) نشان می‌دهد که قوانین ایران فاقد الزامات "حق توضیح" برای کاربران است که این می‌تواند با میثاق بین‌المللی حقوق مدنی و سیاسی (ماده ۱۴) ناسازگار باشد. برای رفع این چالش‌ها، نیاز به اصلاح آیین‌نامه با افزودن بندهایی برای افشای الگوریتم‌ها و آموزش روان‌شناختی قضات است تا نرخ پذیرش فناوری به سطوح بالاتر برسد. در نهایت، این مقررات، هرچند پیشرو، بدون نظارت مستقل و مکانیسم‌های جبران خسارت، ممکن است به جای حل چالش‌ها، آنها را تشدید کنند و نیاز به به‌روزرسانی بر اساس تجربیات عملی (مانند ستاد توسعه فناوری‌های هوش مصنوعی و رباتیک) حیاتی است.

چارچوب‌های حقوقی شفافیت

شفافیت در الگوریتم‌های هوش مصنوعی نیازمند دسترسی عمومی به عوامل و وزن‌دهی‌های تصمیم‌گیری و نظارت مستقل است تا حقوق متهمان، به‌ویژه حق دفاع (ماده ۳۵ قانون اساسی)، حفظ شود. استفاده از هوش مصنوعی در جمع‌آوری ادله باید با الزامات شفافیت، قابلیت کنترل، بی‌طرفی و مکانیسم‌های مسئولیت‌پذیری هم‌راستا باشد تا ادله قابل قبول باشند (Spiridonov, 2023: 481). در سطح بین‌المللی، قانون جمهوری دیجیتال فرانسه (۲۰۱۶)^۱ حق درخواست اطلاعات درباره تصمیمات الگوریتمی را تضمین می‌کند (Avramovic & Jovanov, 2023: 161). با این حال، الگوریتم‌های خصوصی، به دلیل تضاد منافع توسعه‌دهندگان، اغلب غیرشفاف باقی می‌مانند و کارایی خود را از فرایندهای «ذاتاً غیرقابل تفسیر» می‌گیرند، که برخی آن را به «هندسه غیرقابل فهم» تشبیه کرده‌اند (Garrett & Rudin, 2024: 561). این مشکل در پرونده نیوجرسی (۲۰۲۰) مشهود بود که نیاز به افشای کد منبع مطرح شد، اما تضاد منافع توسعه‌دهندگان مانع شد. در ایران، اصل شفافیت به‌صراحت در قوانین دادرسی ذکر نشده، اما از ماده ۳۴ قانون اساسی و ماده ۲ قانون آیین دادرسی کیفری استنباط می‌شود. دستورالعمل هوشمندسازی (مصوب ۱۴۰۳) در ماده ۲۴ بر بررسی تغییرات پیشنهادات سامانه توسط رؤسای

اجرایی مشخص برای آموزش قضات یا ایجاد پایگاه داده ملی هستند.

هوش مصنوعی و حمایت از حقوق بنیادین در عدالت کیفری

حفاظت از حریم خصوصی در کاربرد هوش مصنوعی: تطبیق با قوانین ملی و بین‌المللی

حق بر حریم خصوصی، قلمرویی از زندگی شخصی هر فرد است که باید از دخالت هر نهاد، مقام و فردی مصون بماند. در مکتب اسلام، این حق در آیات ۲۷، ۲۹ و ۳۰ سوره نور، آیه ۱۸۹ سوره بقره و آیه ۱۲ سوره حجرات مورد تأکید قرار گرفته است. قانون اساسی نظام‌های مختلف حقوقی، از جمله اصول ۲۲، ۲۵ و ۳۹ قانون اساسی جمهوری اسلامی ایران، نیز بر صیانت از آن تأکید دارند. می‌توان آن را توانایی فرد در مهار و کنترل اطلاعات شخصی خویش دانست؛ به‌گونه‌ای که شخص از این حق برخوردار باشد که مانع دستیابی یا بهره‌برداری دیگران از اطلاعات مربوط به خود بدون رضایت وی شود. این حق بنیادین بشری فقط به موجب حکم قانون نقض خواهد شد. دادگاه عالی ایالات متحده حریم خصوصی را به‌عنوان یک منفعت مشروع به رسمیت شناخته است، اما آن را یک حق مطلق نمی‌داند. در بسیاری از موارد، برخورداری از این حق در برابر برخی حقوق، از جمله نیاز به اجرای عدالت کیفری، محدود شده است. برای مثال، در پرونده‌ای دادگاه تأکید کرد که حفظ امنیت در زندان‌ها ایجاب می‌کند که زندانیان از حق حریم خصوصی برخوردار نباشند، زیرا در غیر این صورت، امکان جلوگیری از ورود مواد مخدر و سلاح به زندان‌ها دشوار خواهد شد (Bagaric, Hunter & Stobbs, 2022: 95). این تصمیم نشان می‌دهد که حق حریم خصوصی در برخی زمینه‌ها به آسانی حتی بدون توجیه قانونی روشن نقض می‌شود. همچنین نظارت مبتنی بر هوش مصنوعی و گسترش کنترل اجتماعی از طریق دوربین‌های مداربسته و سیستم‌های تشخیص چهره مبتنی بر هوش مصنوعی یکی از مباحث مورد توجه در حوزه پلیس هوشمند است. در حال حاضر، دوربین‌های مداربسته نه تنها برای پیشگیری از وقوع جرم، بلکه به‌عنوان ابزاری برای جمع‌آوری شواهد پس از وقوع جرم مورد استفاده قرار می‌گیرند. در صورتی که فناوری نظارت تصویری در لحظه اجرا شود و یک جرم در حال وقوع باشد، اپراتور دوربین معمولاً برای جلوگیری از ادامه جرم اقدام خواهد کرد، از جمله اطلاع‌رسانی به پلیس یا هشدار دادن به مجرم مبنی

واحدهای قضایی تأکید دارد. آیین‌نامه ۱۴۰۴ در ماده ۹ الزام می‌کند که دلایل رد یا اصلاح پیشنهادات سامانه ثبت شود (ماده ۹: «دلایل و مستندات رد یا اصلاح پیشنهاد سامانه نگهداری می‌شود») سند ملی هوش مصنوعی در بند ۱ ماده ۵ بر تدوین لوایح برای مقابله با مخاطرات شفافیت تأکید دارد و قانون برنامه هفتم در بند الف ماده ۱۱۳ استفاده از هوش مصنوعی با حفظ مسئولیت قاضی را الزامی می‌کند. با این حال، این مقررات فاقد جزئیات اجرایی هستند. ماده ۹ آیین‌نامه ۱۴۰۴ و ماده ۲۴ دستورالعمل ۱۴۰۳ به ثبت دلایل اکتفا می‌کنند، اما فاقد مکانیزم‌های تنبیهی برای نقض شفافیت یا استانداردهای فنی برای افشای کد منبع الگوریتم‌ها هستند که این امر می‌تواند با حق دفاع تعارض ایجاد کند.

زیرساخت‌های نظارتی و آموزشی برای شفافیت

برای بهره‌برداری مسئولانه از هوش مصنوعی، ایجاد زیرساخت‌های نظارتی و آموزشی ضروری است. سرعت و دقت سیستم‌های هوش مصنوعی می‌تواند خطاهای انسانی را کاهش داده و از اطلاع دادرسی جلوگیری کند (ماده ۶ قانون آیین دادرسی کیفری، تسریع در رسیدگی)، اما عدم شفافیت در داده‌های آموزشی، ورودی‌ها و ساختار الگوریتم‌ها، اصل شفافیت را تهدید می‌کند. نظارت توسط نهاد مستقل می‌تواند متهمان را قادر به چالش کشیدن نتایج الگوریتمی کند که این امر با حق مواجهه با مقامات تصمیم‌گیر (مستنبط از ماده ۳۵ قانون اساسی) هم‌راستا است. آموزش قضات و کارکنان قضایی برای درک ظرفیت‌ها و محدودیت‌های هوش مصنوعی، ارائه دوره‌های آموزشی ضمن خدمت، افزودن واحدهای مرتبط در مقطع کارشناسی و ایجاد رشته‌های میان‌رشته‌ای در تحصیلات تکمیلی ضروری است. ایجاد پایگاه داده جامع با همکاری قضات و متخصصان دانشگاهی برای آموزش الگوریتم‌ها نیز لازم است. نظارت قضایی بر خروجی‌های هوش مصنوعی برای جلوگیری از تبعات منفی حیاتی است (حسینی، عبدخدایی، شریف‌خانی، ۱۴۰۲: ۸۸). در ایران، قانون برنامه هفتم در بند الف ماده ۱۱۳ بر ایجاد زیرساخت‌های فناوری با حفظ مسئولیت قاضی تأکید دارد و سند ملی هوش مصنوعی در ماده ۳ بر توسعه زیرساخت‌های پایدار تمرکز دارد. دستورالعمل ۱۴۰۳ در ماده ۳۴ گزارش‌دهی پیشرفت را الزامی می‌کند و آیین‌نامه ۱۴۰۴ در ماده ۱۵ نظارت توسط معاون اول قوه قضاییه را پیش‌بینی کرده است. با این حال، این اسناد فاقد برنامه‌های

بر ضبط ویدیویی عمل مجرمانه. همچنین، ضبط تصاویر جرائم می‌تواند کمک بزرگی به شناسایی و تعقیب کیفری مجرمان کند. در سال‌های اخیر، موارد بی‌شماری از جرائم از طریق بررسی فیلم‌های ضبط‌شده توسط پلیس تحت تعقیب قرار گرفتند. با وجود این، برخی نگران هستند که استفاده گسترده از این فناوری‌ها به کنترل اجتماعی شدیدتر و محدودیت‌های بیشتر در برخورداری از آزادی‌های مدنی و نقض حریم خصوصی منجر شود؛ معمولاً، زمانی که مجرم از فیلمبرداری دوربین‌های مداربسته آگاهی ندارد، شناسایی وی پس از انتشار تصویرش در رسانه‌ها صورت می‌گیرد. این امر نقض حریم خصوصی است، اما نقض حق حریم خصوصی که در نتیجه نظارت بر برخی اماکن رخ خواهد داد و در صورت اثبات کاهش چشمگیر جرم از طریق استفاده از فناوری سی‌سی‌تی‌وی^۱ خودکار افزایش می‌یابد، جهت تأمین منافع عمومی از طریق کاهش جرم و افزایش میزان شناسایی و تعقیب مجرمان است. علاوه بر این، تشخیص چهره و نظارت خودکار از طریق دوربین‌های مداربسته، از برخی جهات، نسبت به نظارت انسانی در لحظه، مداخله‌گرایانه‌تر محسوب نمی‌شوند، زیرا کامپیوتر رفتار افراد را ارزیابی نمی‌کند؛ مأموران اجرای قانون تنها در صورتی به بررسی تصاویر یا مشاهده چهره می‌پردازند که سیستم رایانه‌ای نشانه‌ای از ارتکاب جرم یا شناسایی یک مجرم را تشخیص دهد. بنابراین اگرچه افراد در معرض امکان مشاهده هستند، اما به‌طور مستمر تحت نظارت مستقیم مأموران اجرای قانون قرار نمی‌گیرند (Ibid).

در این زمینه، رأی دادگاه عالی هند مورد توجه است؛ در پرونده‌ای به حریم خصوصی به‌عنوان یک حق اساسی اشاره کرده است. این پرونده نشان داد که هر گونه استفاده از داده‌های بیومتریک و فناوری‌های شناسایی چهره باید مطابق با قوانین و مقررات موجود باشد و از حقوق افراد در این زمینه حفاظت شود (Amarendar Reddy, 2022: 1766). این چالش نشان می‌دهد که هوش مصنوعی، با پردازش داده‌های عظیم، می‌تواند حریم خصوصی را تهدید کند، اما با الزامات قانونی مانند ماده ۳۰ دستورالعمل ۱۴۰۳ (تأمین امنیت اطلاعات و احراز هویت)، ماده ۵ آیین‌نامه ۱۴۰۴ (حفظ امنیت اطلاعات و حریم خصوصی) بند ۱ ماده ۵ سند ملی هوش مصنوعی (تدوین لوایح برای بهره‌برداری امن و مقابله با مخاطرات حریم خصوصی) و ماده ۱۱۳ قانون برنامه هفتم (رعایت قانون مدیریت داده‌ها در استفاده از هوش مصنوعی)

ریسک‌ها را می‌توان کاهش داد. مطالعات اخیر تأکید دارند که الگوریتم‌های شفاف می‌توانند نقض‌ها را به حداقل برسانند، اما بدون نظارت، سوگیری‌های داده‌ای منجر به نقض سیستماتیک حقوق می‌شود (Spiridonov, 2023:481). اسناد ایران، اگرچه در ظاهر به حفاظت از حریم خصوصی متعهد هستند، اما کاستی‌های ساختاری و اجرایی آنها این حق را در برابر فناوری‌های هوش مصنوعی آسیب‌پذیر می‌کند؛ ماده ۵ آیین‌نامه اجرایی نحوه استفاده از فناوری‌های نوین در فرایندهای قضایی (مصوب ۱۴۰۴) بر حفظ امنیت اطلاعات و حریم خصوصی تأکید دارد، اما فاقد استانداردهای فنی مشخص مانند رمزنگاری پیشرفته یا پروتکل‌های احراز هویت چندمرحله‌ای، برای جلوگیری از سوءاستفاده از داده‌های بیومتریک است. این ابهام، به‌ویژه در سیستم‌های تشخیص چهره که در پلیس هوشمند ایران استفاده می‌شوند، خطر نقض حریم خصوصی را افزایش می‌دهد. ماده ۳۰ دستورالعمل هوشمندسازی و الکترونیکی شدن فرایندها و پرونده‌های قضایی (مصوب ۱۴۰۳) تأمین امنیت اطلاعات و احراز هویت را الزامی می‌کند، اما این ماده به فرایندهای داخلی قوه قضاییه محدود است و استفاده گسترده از فناوری‌های نظارتی در حوزه پلیس هوشمند (مانند تشخیص چهره در اماکن عمومی) را پوشش نمی‌دهد. این شکاف نظارتی می‌تواند به نقض سیستماتیک حریم خصوصی به‌ویژه در نبود مکانیزم‌های جبران خسارت برای افراد متضرر منجر شود. سند ملی هوش مصنوعی (مصوب ۱۴۰۳) در بند ۱ ماده ۵ بر تدوین لوایح برای بهره‌برداری امن و مقابله با مخاطرات حریم خصوصی تأکید دارد، اما این سند فاقد شاخص‌های عملی برای ارزیابی امنیت داده‌ها یا الزام به اطلاع‌رسانی به افراد در صورت استفاده از داده‌هایشان است. مقایسه با مقررات عمومی حفاظت از داده‌های اتحادیه اروپا (GDPR, 2016) که حق توضیح و رضایت صریح را الزامی می‌کند، نشان‌دهنده ضعف ایران در تضمین شفافیت و پاسخگویی است. ماده ۱۱۳ قانون برنامه پنج‌ساله هفتم پیشرفت (مصوب ۱۴۰۳) بر رعایت قانون مدیریت داده‌ها در استفاده از هوش مصنوعی تأکید دارد، اما تا تاریخ ۱۳ اوت ۲۰۲۵، قانون مدیریت داده‌ها و اطلاعات ملی به طور کامل اجرا نشده است. این تأخیر، همراه با دوگانگی سیاستی بین شورای عالی انقلاب فرهنگی و شورای عالی فضای مجازی، هماهنگی در اجرای مقررات حریم خصوصی را مختل کرده

1. Closed-Circuit Television

(تلویزیون مداربسته)

مالی بدون تبعیض را الزامی می‌کند، اما این ماده صرفاً به فرایندهای مالی محدود است و به ابزارهای پیش‌بینی‌کننده ریسک که در مراحل بازداشت یا تعقیب کیفری استفاده می‌شوند، توجهی ندارد. این شکاف نظارتی می‌تواند به تبعیض سیستماتیک علیه گروه‌های آسیب‌پذیر به‌ویژه در نبود استانداردهای فنی برای ارزیابی الگوریتم‌ها منجر شود. سند ملی هوش مصنوعی (مصوب ۱۴۰۳) در اهداف راهبردی خود بر ترویج عدالت و جلوگیری از تبعیض تأکید دارد، اما فاقد شاخص‌های ارزیابی عملی برای سنجش سوگیری الگوریتمی است. برای مثال، عدم تعیین معیارهای عدالت مانند تساوی جمعیتی یا برابری فرصت‌ها باعث می‌شود که ادعای «زیست‌بوم عادلانه» به یک شعار کلی تبدیل شود. مقایسه با رهنمودهای اخلاقی اتحادیه اروپا (۲۰۱۹) نشان می‌دهد که ایران فاقد الزامات مشخص برای افشای سوگیری‌های داده‌ای است، که با ماده ۱۰ اعلامیه جهانی حقوق بشر (حق دادرسی عادلانه) ناسازگار است. بند ج ماده ۶۵ قانون برنامه پنج‌ساله هفتم پیشرفت (مصوب ۱۴۰۳) بر حمایت از زیست‌بوم نوآورانه هوش مصنوعی با تأکید بر جنبه‌های اخلاقی تأکید دارد.

هوش مصنوعی و اصل برائت در عدالت کیفری

اصل برائت، به‌عنوان یکی از ارکان بنیادین دادرسی عادلانه، در حقوق ایران بر پایه اصل ۳۷ قانون اساسی جمهوری اسلامی ایران (مصوب ۱۳۵۸ با اصلاحات ۱۳۶۸) استوار است که تصریح می‌کند: «اصل برائت است و هیچ‌کس از نظر قانون مجرم شناخته نمی‌شود، مگر اینکه جرم او در دادگاه صالح ثابت گردد.» این اصل، که ریشه در فقه اسلامی (مانند روایات مبنی بر «البینه علی المدعی») و اسناد بین‌المللی مانند ماده ۱۱ اعلامیه جهانی حقوق بشر (۱۹۴۸) و ماده ۱۴ میثاق بین‌المللی حقوق مدنی و سیاسی (۱۹۶۶) دارد، ایجاب می‌کند که هر فرد تا اثبات جرم فراتر از شک معقول، بی‌گناه تلقی شود و هیچ تصمیمی پیشگیرانه علیه او بدون شواهد واقعی و مستند اتخاذ نشود. با این حال، ابزارهای پیش‌بینی‌کننده هوش مصنوعی در دادرسی کیفری، مانند سیستم‌های ارزیابی ریسک جرم، چالش‌های جدی برای این اصل ایجاد می‌کنند. این ابزارها، که بر الگوریتم‌های یادگیری ماشین مبتنی هستند، اغلب بر اساس داده‌های تاریخی و آماری (مانند سابقه خانوادگی یا محیط

است. برای مثال، فقدان پایگاه داده ملی قضایی با استانداردهای امنیتی بالا، خطر هک یا سوءاستفاده از داده‌های حساس را افزایش می‌دهد. مقایسه با استانداردهای بین‌المللی مانند مقررات عمومی حفاظت از داده‌های اتحادیه اروپا یا رهنمودهای شورای اروپا (۲۰۲۰) نشان می‌دهد که ایران فاقد الزامات «حق فراموشی» یا «حق توضیح» برای داده‌های پردازش‌شده توسط هوش مصنوعی است که با ماده ۱۷ میثاق بین‌المللی حقوق مدنی و سیاسی (حفاظت از حریم خصوصی) ناسازگار است.

هوش مصنوعی و اصل منع تبعیض در عدالت کیفری

استفاده از هوش مصنوعی در عدالت کیفری، به‌ویژه در ابزارهای پیش‌بینی‌کننده ریسک (مانند ارزیابی احتمال ارتکاب جرم مجدد)، می‌تواند نابرابری‌های اجتماعی-اقتصادی و قومیتی را تشدید کند، به‌خصوص اگر الگوریتم‌ها بر داده‌های آموزشی سوگیرانه تکیه کنند. مطالعات بین‌المللی، مانند پرونده COMPAS در آمریکا، نشان داده‌اند که الگوریتم‌های غیرشفاف می‌توانند نرخ خطای کاذب مثبت بالاتری (تا ۴۵ درصد) برای گروه‌های اقلیت تولید کنند، که نقض اصل منع تبعیض (ماده ۱۴ میثاق بین‌المللی حقوق مدنی و سیاسی) را به دنبال دارد (Kaplan, Smith, & Brown, 2023: 147). عدالت تالگوریتمی ایجاب می‌کند که الگوریتم‌ها نرخ خطای کاذب مثبت و منفی یکسانی برای همه گروه‌ها، صرف‌نظر از نژاد، جنسیت یا وضعیت اقتصادی، تولید کنند (Bagaric, Hunter, & Stobbs, 2022: 95). در چارچوب نظام حقوقی ایران، اسناد متعددی به اصل منع تبعیض اشاره دارند، اما کاستی‌های ساختاری و اجرایی آنها این اصل را در معرض خطر قرار می‌دهد؛ ماده ۲ آیین‌نامه اجرایی نحوه استفاده از فناوری‌های نوین در فرایندهای قضایی (مصوب ۱۴۰۴) بر نظارت بر تحلیل داده‌ها برای جلوگیری از سوگیری تأکید دارد، اما فاقد پروتکل‌های مشخص برای ارزیابی سوگیری الگوریتم‌ها است. این ابهام می‌تواند منجر به استفاده از داده‌های سوگیرانه به‌ویژه در مناطقی با تنوع قومیتی بالا مانند سیستان و بلوچستان شود. ماده ۲۱ دستورالعمل هوشمندسازی و الکترونیکی شدن فرایندها و پرونده‌های قضایی (مصوب ۱۴۰۳) تشخیص برخط ملائت یا عدم تمکن

اجتماعی) عمل می‌کنند و می‌توانند منجر به تصمیمات بی‌اساس درباره افراد بدون اطلاعات واقعی کافی شوند، که این امر نه تنها استانداردهای اثبات اتهام را نقض می‌کند، بلکه اصل برائت را به چالش می‌کشد و می‌تواند به سوءاستفاده از قدرت قضایی و نقض کرامت انسانی منجر شود (Garrett & Rudin, 2024: 561). نمونه‌های بین‌المللی این نقض، مانند پروکاید^۱ که از داده‌های مرتبط با جنایاتی که توسط افراد در تماس نزدیک با کودک یا خود کودک مرتکب شده‌اند، برای پیش‌بینی امکان ارتکاب جرم در آینده توسط کودک استفاده می‌کند^۲ (NDIAS) که از داده‌های محیط شخصی برای پیش‌بینی ریسک ارتکاب جرم در آینده بهره می‌برد و ریس کنوی^۳ که به سابقه جنایی خانواده یا والدین و دوستان مجرم یا ضداجتماعی توجه می‌کند، نشان‌دهنده «جرم‌نگاری از طریق اجتماع هستند، بدون آنکه مدرک واقعی یا اثبات جرم وجود داشته باشد» (Bagaric, Hunter, & Stobbs, 2022: 95). وجود این سیستم‌ها، بر اساس مدل‌های آماری مانند رگرسیون لجستیک یا شبکه‌های عصبی عمیق، حاکی از آن است که پیش‌بینی‌های مبتنی بر الگوهای گذشته اغلب به دلیل سوگیری‌های ذاتی در داده‌های آموزشی منجر به نرخ خطای کاذب مثبت بالا می‌شوند که افراد بی‌گناه را تحت نظارت پیشگیرانه یا بازداشت قرار می‌دهد (کاپلان، اسمیت و براون، ۲۰۲۳: ۱۴۷). از منظر روان‌شناختی، پدیده «گریز الگوریتمی» نیز چالش دیگری است؛ زیرا قضات ممکن است به پیش‌بینی‌های ماشینی بیش از حد اعتماد کنند یا از آنها اجتناب کنند، بدون ارزیابی شواهد واقعی که این امر اصل برائت را تضعیف می‌کند (Ibid: 758). چالش اصلی در حقوق ایران این است که این ابزارها، اگر بدون چارچوب حقوقی محکم پیاده‌سازی شوند، می‌توانند با اصل ۳۷ قانون اساسی و ماده ۱۲۹ قانون آیین دادرسی کیفری (مصوب ۱۳۹۲ با اصلاحات بعدی) که بر اثبات جرم با دلایل کافی تأکید دارد، تعارض ایجاد کنند، و این تعارض نه تنها حقوق متهمان را نقض می‌کند، بلکه مشروعیت نظام قضایی را زیر سؤال می‌برد.

در چارچوب اسناد هوش مصنوعی ایران، این چالش‌ها تا حدی مدیریت شده‌اند، اما با نگاه چالشی، کاستی‌های ساختاری و اجرایی آشکار هستند. دستورالعمل هوشمندسازی

و الکترونیکی شدن فرایندها و پرونده‌های قضایی (مصوب ۱۴۰۳) در ماده ۲۲ بر هشدار خودکار سامانه‌ها در صورت انقضای مهلت‌های قانونی (مانند صدور رأی یا فک قرار تأمین) تأکید دارد، که از تصمیمات پیشگیرانه بدون شواهد جلوگیری می‌کند (ماده ۲۲): «در کلیه فرایندهای قضایی چنانچه به موجب قانون برای اتخاذ تصمیم یا اقدام مهلت تعیین شده است... در صورت انقضای مهلت و عدم اقدام سامانه‌های قضایی باید به کاربر مربوط و مسئول مستقیم وی هشدار دهند». با این حال، چالش اینجاست که این ماده بیشتر بر فرایندهای اداری تمرکز دارد و مستقیماً به پیش‌بینی‌های ریسک هوش مصنوعی نمی‌پردازد؛ بنابراین، ممکن است در عمل، قاضی به پیش‌بینی‌های الگوریتمی بدون شواهد واقعی تکیه کند که این امر اصل برائت را تضعیف می‌کند. آیین‌نامه اجرایی نحوه استفاده از فناوری‌های نوین در فرایندهای قضایی (مصوب ۱۴۰۴) در ماده ۸ و ۹ بر تأیید پیشنهادات سامانه توسط قاضی و ثبت دلایل رد یا اصلاح آنها تأکید دارد که این مقررات می‌توانند اصل برائت را حفظ کنند (ماده ۸): «پیشنهاد سامانه در صورت تأیید کاربر اعمال می‌شود، در غیر این صورت کاربر به نحو مستدل... اقدام می‌نماید»؛ ماده ۹: «در صورت رد یا اصلاح پیشنهاد سامانه... دلایل و مستندات آن نگهداری می‌شود»، اما چالش این است که این آیین‌نامه فاقد مکانیزم‌های تنبیهی برای نقض اصل برائت توسط الگوریتم‌هاست و ثبت دلایل ممکن است به یک فرایند تشریفاتی تبدیل شود، بدون آنکه نظارت واقعی بر سوگیری‌های الگوریتمی اعمال شود.

سند ملی هوش مصنوعی جمهوری اسلامی ایران (مصوب ۱۴۰۳) در بخش اصول و ارزش‌های بنیادین (ماده ۱) بر عدالت و انصاف و حمایت از حقوق بشر تأکید دارد و در اهداف راهبردی (ماده ۳) بر توسعه هوش مصنوعی «قابل اعتماد و پایدا»^۴ تمرکز می‌کند که ریسک پیش‌بینی‌های ناعادلانه را کاهش دهد. بند ۱ ماده ۵ سند نیز بر تدوین لوایح قانونی برای مقابله با مخاطرات هوش مصنوعی (مانند نقض دادرسی عادلانه) تأکید دارد که این امر می‌تواند به جلوگیری از «جرم‌نگاری از طریق اجتماع» کمک کند. با این حال، نگاه چالشی نشان می‌دهد که سند ملی بیشتر بر اصول کلی تمرکز دارد و فاقد جزئیات عملی برای ارزیابی ریسک ابزارهای

3. Rice Conway

1. ProKaid

2. Notre Dame Institute for Advanced Study (مؤسسه مطالعات عالی دانشگاه نوتردام)

3. کنوی

4. پایدا

تمامی سیستم‌های هوش مصنوعی که در حوزه قضایی استفاده می‌شوند، نباید موجب نقض حقوق اساسی همچون بی‌طرفی، حق محاکمه عادلانه، حفظ حریم خصوصی و حمایت از کرامت انسانی شوند. از اسناد نظارتی مرتبط، کنوانسیون اروپایی حقوق بشر، منشور اخلاقی اروپا درباره هوش مصنوعی، مقررات عمومی حفاظت از داده‌ها و کاغذ سفید هوش مصنوعی (۲۰۲۰) هستند که اصول و چارچوب‌هایی را برای استفاده مسئولانه از این فناوری مشخص کرده‌اند. حسب مورد در این گفتار مورد بررسی قرار خواهند گرفت.

امنیت داده‌ها و حفاظت از حریم خصوصی در سیستم‌های قضایی هوشمند

داده‌های مورد استفاده در سیستم‌های هوش مصنوعی باید از منابع معتبر تهیه شوند و از استانداردهای امنیتی بالایی برخوردار باشند تا از دست‌کاری، هک یا سوءاستفاده‌های احتمالی جلوگیری شود. سند نظارتی مرتبط مقررات درباره حفاظت از داده‌ها و دستورالعمل حفظ حریم خصوصی الکترونیکی است. سیستم قضایی انگلستان گام مهمی در پذیرش فناوری هوش مصنوعی برداشته است. در ۱۲ دسامبر ۲۰۲۳، خدمات قضایی دادگاه‌های انگلستان و ولز، راهنمایی رسمی تحت عنوان «هوش مصنوعی-راهنمایی قضایی» را با هدف اطمینان از بهره‌گیری پایدار از هوش مصنوعی در کنار حفظ عدالت و بی‌طرفی نظام قضایی منتشر کرد و برای کلیه مقامات قضایی، دبیران دادگاه‌ها و سایر کارکنان مرتبط که تحت مدیریت رئیس دادگستری و مقامات ارشد قضایی فعالیت می‌کنند، قابل اجراست. این دستورالعمل کاربردهای بالقوه این فناوری را در فرایندهای دادرسی شرح می‌دهد و بر لزوم شناسایی ریسک‌های مرتبط با استفاده از هوش مصنوعی، از جمله خطرات مرتبط با علنی بودن، نقض حریم خصوصی و امنیت داده‌ها تأکید دارد؛ به‌ویژه تصریح شده است که اطلاعات خصوصی یا محرمانه نباید در چت‌بات‌های عمومی هوش مصنوعی وارد شوند؛ زیرا این ابزارها ممکن است داده‌های ورودی کاربران را ذخیره کرده و در تعاملات بعدی از آنها استفاده کنند. علاوه بر این، قبل از اتکا بر داده‌های ارائه‌شده توسط ابزارهای هوش مصنوعی، صحت و دقت آنها باید مورد بررسی قرار گیرد؛ زیرا این اطلاعات ممکن است نادرست، ناقص، گمراه‌کننده یا منسوخ باشند. اگرچه در حال حاضر بر

پیش‌بینی‌کننده است؛ مثلاً نبود شاخص‌های ارزیابی خاص برای نرخ خطای کاذب در الگوریتم‌ها می‌تواند اجرای اصل برائت را ضعیف کند؛ به‌ویژه در نبود بخشنامه‌های اجرایی از سوی شورای عالی انقلاب فرهنگی یا قوه قضاییه. قانون برنامه پنج‌ساله هفتم پیشرفت جمهوری اسلامی ایران (مصوب ۱۴۰۳) در بند الف ماده ۱۱۳ بر استفاده از هوش مصنوعی برای اموری مانند ارجاع پرونده با «حفظ مسئولیت شخصی قاضی» تأکید دارد، که این مقرر اصل برائت را با الزام نظارت انسانی تقویت می‌کند (بند الف ماده ۱۱۳: «فراهم کردن امکان استفاده از فناوری‌های نوین مانند هوش مصنوعی... با حفظ مسئولیت شخصی قاضی»). بند ج ماده ۶۵ نیز بر تدوین برنامه ملی توسعه هوش مصنوعی با تمرکز بر «جنبه‌های اخلاقی و حقوقی» و «مخاطرات احتمالی» تأکید دارد که شامل جلوگیری از نقض اصل برائت می‌شود. چالشی اینجاست که این قانون زمان‌بندی شش‌ماهه برای تدوین برنامه ملی ارائه می‌دهد، اما تاکنون گزارش‌های رسمی از اجرای کامل آن منتشر نشده، که این تأخیر می‌تواند ریسک‌های پیش‌بینی‌کننده را افزایش دهد.

با این حال، چالش عمده عدم وجود بخشنامه‌های اجرایی دقیق برای آزمون الگوریتم‌ها است؛ مثلاً، نبود استانداردهای ملی برای ارزیابی نرخ خطای کاذب می‌تواند به نقض سیستماتیک اصل برائت منجر شود؛ به‌ویژه در سیستم‌هایی که بر داده‌های تاریخی سوگیرانه تکیه دارند. برای رفع این کاستی‌ها، پیشنهاد می‌شود قوه قضاییه بخشنامه‌ای جدید صادر کند که الزام به آزمون‌های مستقل الگوریتم‌ها مانند اعتبارسنجی متقابل علمی و ادغام اصل برائت در مدل‌های پیش‌بینی‌کننده را الزامی کند، همان‌طور که در ماده ۹ آیین‌نامه ۱۴۰۴ بر ثبت دلایل تأکید شده است. در نهایت، پیش‌بینی‌های بدون شواهد واقعی نه تنها حقوق متهمان را نقض می‌کند، بلکه مشروعیت نظام قضایی را زیر سؤال می‌برد و اسناد ایران با تمرکز بر نظارت انسانی و اصول ارزشی، این ریسک را مدیریت می‌کنند، اما اجرای عملی آنها نیازمند اصلاحات چالشی و تحقیقات بیشتر است (حسینی، عبدخدایی و شریف‌خانی، ۱۴۰۲: ۸۸).

الزامات حقوقی و اخلاقی بهره‌گیری از هوش مصنوعی در عدالت کیفری ایران

حمایت از حقوق بنیادین در کاربردهای قضایی هوش مصنوعی

محدودیت‌های این فناوری تأکید شده و استفاده از آن برای تحلیل‌های حقوقی و تحقیقات قضایی مناسب دانسته نشده است، اما این راهنما به‌طور کامل استفاده از هوش مصنوعی را رد نمی‌کند. بلکه مقامات قضایی را به بهره‌گیری متعادل و هوشمندانه از این فناوری ترغیب می‌نماید. به‌رغم تمامی مزایای فناوری، دادگاه‌ها به عنوان نماد اقتدار قضایی و حافظ حاکمیت قانون، باید ماهیت انسانی خود را حفظ کنند و باید تعادلی میان حفظ منافع عمومی و ترویج نوآوری‌های فناورانه برقرار کند. از سال ۲۰۱۸ تاکنون، بیشتر کشورها در مسیر تصویب قوانین حمایتی از داده‌ها و حریم خصوصی شهروندان خود گام برداشته‌اند. ایران نیز در سال‌های اخیر کوشیده است تا هم در بُعد فنی و هم در زمینه تدوین مقررات، اقدامات مؤثری انجام دهد. از جمله تلاش‌های قابل ذکر در این راستا می‌توان به قانون انتشار و دسترسی آزاد به اطلاعات و قانون مدیریت داده‌ها و اطلاعات ملی اشاره کرد. در ادامه‌ی این روند، به‌منظور تقویت زیرساخت‌های مربوط به شرکت‌های فعال در عرصه هوش مصنوعی، قوانین دیگری همچون قانون جهش تولید دانش‌بنیان و قانون حمایت از شرکت‌ها و مؤسسات دانش‌بنیان و تجاری‌سازی نوآوری‌ها و اختراعات نیز به تصویب رسیدند. در سال ۱۴۰۱، با شدت گرفتن مباحث پیرامون هوش مصنوعی و پیامدهای آن، شورای عالی فضای مجازی اقدام به تصویب سند راهبردی جمهوری اسلامی ایران در حوزه فضای مجازی نمود. بر پایه این سند، وظایف مشخصی میان نهادهای کشور تقسیم گردید: طراحی نظام حقوقی فضای مجازی بر عهده معاونت حقوقی ریاست جمهوری، تدوین نظام قضایی فضای مجازی بر دوش رئیس قوه قضائیه، و طراحی نظام بهره‌برداری از فناوری‌های نوین نظیر هوش مصنوعی و علم داده، به معاونت علمی، فناوری و اقتصاد دانش‌بنیان ریاست جمهوری واگذار شد.

از آن پس، در برخی از قوانین مصوب جدید، هرچند محدود، به هوش مصنوعی نیز توجه شده است. برای نمونه، در ماده ۱۱۳ و ۶۵ قانون برنامه هفتم پیشرفت جمهوری اسلامی ایران مقرر می‌دارد: «به منظور تسهیل در روند رسیدگی به پرونده‌های قضایی و رفع اختلافات میان مردم، قوه قضائیه و وزارت دادگستری با رعایت مفاد قانون مدیریت داده‌ها و اطلاعات ملی، حسب مورد مکلف هستند اقداماتی را به انجام رسانند؛ از جمله فراهم کردن امکان استفاده از فناوری‌های نوین مانند هوش مصنوعی برای اموری نظیر ارجاع پرونده، تعیین وقت دادرسی و

انتخاب کارشناس، با حفظ مسئولیت شخصی قاضی. آیین‌نامه اجرایی این بند بر اساس سیاست‌های مصوب شورای عالی فضای مجازی تهیه شده و باید به تأیید رئیس قوه قضائیه برسد». همچنین، در بند ج ماده ۶۵ همین قانون آمده است: «دولت موظف است با هدف حمایت از شکل‌گیری و گسترش زیست‌بوم نوآورانه هوش مصنوعی قابل اعتماد و پایدار و در جهت تبیین چارچوب‌ها و مکانیسم تعامل بین ذی‌نفعان، تأمین دانش و زیرساخت‌های علمی، اجتماعی، اخلاقی و حقوقی و نیز ترویج شناخت عمومی نسبت به کارکردها و مخاطرات احتمالی هوش مصنوعی، ظرف مدت شش ماه از لازم‌الاجرا شدن این قانون، برنامه ملی توسعه هوش مصنوعی را مطابق با سیاست‌های کلی نظام، مصوبات شورای عالی انقلاب فرهنگی و سند راهبردی فضای مجازی تدوین و اجرا نماید.

همچنین در تکمیل این اقدامات، در تاریخ ۱۴۰۳/۰۳/۲۹، شورای عالی انقلاب فرهنگی با تصویب سند ملی هوش مصنوعی جمهوری اسلامی ایران گام مهم دیگری برداشت. این سند حاوی اصول و ارزش‌های بنیادین، چشم‌انداز کلان، اهداف راهبردی، شاخص‌های ارزیابی، سیاست‌های اصلی، اقدامات ملی و اولویت‌های کشور در زمینه استفاده از هوش مصنوعی است. از جمله موارد برجسته در این سند، رعایت حریم خصوصی و تأمین امنیت داده‌ها در زیست‌بوم هوش مصنوعی به‌عنوان یکی از مبانی ارزش‌گذارانه مطرح شده است. همچنین در بند ۱ ماده ۵ این سند، بر لزوم تدوین لوایح قانونی مورد نیاز برای ایجاد بستر حقوقی مناسب به‌منظور بهره‌برداری امن از هوش مصنوعی و مقابله با مخاطرات آن از جمله در زمینه حریم خصوصی تأکید شده است (میرشکاری، ثابت مقدم، اصغرینا، ۱۴۰۳: ۸۲). این الزامات نشان می‌دهد که سند ملی هوش مصنوعی ۱۴۰۳، با تمرکز بر امنیت داده‌ها، می‌تواند نقض حریم خصوصی را کاهش دهد، اما نیاز به اجرای عملی دارد، همان‌طور که در برنامه هفتم توسعه تأکید شده است.

اصل برابری و منع تبعیض در الگوریتم‌های قضایی

عملکرد سیستم‌های هوش مصنوعی نباید بر اساس نژاد، جنسیت، سن، یا سایر ویژگی‌های فردی باشند. این اصل بر اهمیت استفاده از داده‌های بی‌طرفانه و الگوریتم‌های عاری از تبعیض تأکید دارد. سند نظارتی مرتبط مقررات عمومی حفاظت از داده‌ها، دستورالعمل سازمان‌های مجری قانون و رهنمودهای اخلاقی برای هوش مصنوعی قابل اعتماد (۲۰۱۹) است. ماده ۱۰

اما نیاز به داده‌های متنوع دارند.

اصل برائت در دادرسی کیفری

یکی از نگرانی‌ها در استفاده از ابزارهای هوش مصنوعی در دادرسی کیفری، نقض اصل برائت است. استفاده از ابزارهای پیش‌بینی‌کننده که شامل ارزیابی ریسک و پیش‌بینی فعالیت‌های غیرقانونی احتمالی است، می‌تواند منجر به تصمیمات بی‌اساس درباره افراد بدون اطلاعات واقعی کافی شود. این امر به عدم رعایت استانداردهای اثبات اتهام در مراحل دادرسی کیفری می‌انجامد و نقض حق برائت را به دنبال دارد. نمونه‌های نقض اصل برائت می‌توان به سیستم پروکاید^۲ از داده‌های مرتبط با جنایاتی که توسط افراد در تماس نزدیک با کودک یا خود کودک مرتکب شده‌اند، برای پیش‌بینی امکان ارتکاب جرم در آینده توسط کودک استفاده می‌کند. سیستم ان دی ای اس^۳ از داده‌های محیط شخصی برای پیش‌بینی ریسک ارتکاب جرم در آینده استفاده می‌کند. سیستم رایس کنوی^۴ به سابقه جنایی خانواده یا والدین و دوستان مجرم یا ضد اجتماعی توجه می‌کند. این رویکرد به منزله «جرمانگاری از طریق اجتماع» است، بدون آنکه مدرک واقعی یا اثبات جرم وجود داشته باشد. تلفیق با ماده ۲۲ دستورالعمل ۱۴۰۳ (هشدار انقضای مهلت‌ها برای حفظ حقوق فرد) و اصول ارزشی سند ملی هوش مصنوعی (حمایت از کرامت انسانی و دادرسی عادلانه) می‌تواند این ریسک را کاهش دهد. با وجود این، پیش‌بینی‌های ریسک بدون شواهد واقعی، اصل برائت را نقض می‌کند.

شفافیت و مسئولیت‌پذیری در تصمیم‌گیری‌های قضایی هوش مصنوعی

سیستم‌های هوش مصنوعی که در فرایندهای قضایی به کار گرفته می‌شوند، باید شفافیت لازم را داشته باشند و فرایندهای تصمیم‌گیری آنها برای قضات، وکلا و متهمان قابل توضیح و قابل درک باشد. در فرانسه، در سال ۲۰۱۶ به منظور افزایش شفافیت در فرایند تصمیم‌گیری (که به نهادهای اداری محدود می‌شود)، قانون جمهوری دیجیتال به تصویب رسید که حق درخواست اطلاعات در خصوص تصمیمات الگوریتمی، از جمله ویژگی‌های اساسی الگوریتم‌ها را برای مردم فراهم می‌کند (Avramovic & Jovanov, 2023: 161).

اعلامیه جهانی حقوق بشر نیز به حق برخورداری از دادگاه منصفانه، مستقل و بی‌طرفانه اشاره دارد. همچنین اصول بنگلور مصوب ۲۰۰۲ بر اصولی چون استقلال، یکپارچگی، انصاف، برابری، شایستگی و صداقت به عنوان ارزش‌های اساسی در فرایند قضایی تأکید دارد. با توجه به اصول اخلاقی و قوانین مختلف نظیر اصول بنگلور^۱ و قوانین داخلی کشورها، قاضی باید وظایف خود را با بی‌طرفی و به دور از هرگونه تبعیض و بر اساس ارزیابی مستقل از واقعیت‌ها و تفسیر حقوقی انجام دهد. با این حال، باید اذعان داشت که در عمل، دستیابی به بی‌طرفی کامل در الگوریتم‌های هوش مصنوعی که توسط انسان‌ها طراحی و توسعه می‌شوند، چالشی پیچیده است. در واقع، این ایده‌آل اخلاقی بی‌طرفی در سطح نظری و هنجاری ممکن است در برخورد با واقعیت‌های انسانی دچار مشکلاتی شود و راه‌حل‌هایی که برای مقابله با این چالش‌ها به کار می‌روند، ممکن است خود به مشکلات جدیدی منجر شوند (Avramovic & Jovanov, 2023: 161). تبعیض و نابرابری یک پدیده جدید یا منحصر به فرد برای هوش مصنوعی نیست. هدف نهایی، دستیابی به «خطر صفر» نیست، بلکه شناسایی، درک، اندازه‌گیری، مدیریت و کاهش نابرابری است. روش‌های کاهش نابرابری ضروری هستند و باید انعطاف‌پذیر باشند تا بدون توجه به صنعت خاص، در زمینه‌های مختلف قابل استفاده باشند (Talapina, 2022: 4). ماده ۳۲ اعلامیه تورنتو با عنوان «حفاظت از حق برابری و عدم تبعیض در سیستم‌های یادگیری ماشینی» دولت‌ها را ملزم به رعایت اصول مسئولیت‌پذیری و شفافیت در استفاده از این فناوری‌ها در بخش عمومی می‌کند. این تعهدات شامل: افشای عمومی محل استفاده از سیستم‌های یادگیری ماشینی در حوزه عمومی، ارائه اطلاعات شفاف درباره نحوه تصمیم‌گیری خودکار این سیستم‌ها، مستندسازی اقدامات انجام‌شده برای شناسایی و کاهش اثرات تبعیض‌آمیز یا ناقض حقوق بشر. در ایران، این اصل با ماده ۲ آیین‌نامه ۱۴۰۴ (شناسایی فرایندهای قابل اجرا با فناوری نوین بدون تبعیض) اهداف راهبردی سند ملی هوش مصنوعی (حمایت از برابری و جلوگیری از تبعیض در کاربردهای هوش مصنوعی) و بند ج ماده ۶۵ قانون برنامه هفتم (تبیین چارچوب‌های اخلاقی برای جلوگیری از مخاطرات احتمالی) همخوانی دارد. بنابراین، الگوریتم‌های عاری از سوگیری می‌توانند عدالت را تقویت کنند،

3. NDAS system

1. Bangalore
2. Prokaid system

استرالیا، در سال ۲۰۱۹، دولت فدرال اصول اخلاقی هوش مصنوعی را منتشر کرد که به مردم این امکان را می‌دهد تا بدانند چه زمانی در معرض استفاده از الگوریتم‌ها قرار دارند. اصول دادرسی عادلانه ایجاب می‌کند که در فرایندهای قضایی کیفری، استفاده از هوش مصنوعی باید به صورت شفاف و تفسیرپذیر باشد. آیین دادرسی در مراحل تحقیقاتی، تصمیم‌گیری درباره وثیقه و صدور حکم نیز باید از مدل‌های شفاف استفاده شود. در مرحله محاکمه، حمایت‌های آیین دادرسی کیفری الزام می‌کند که تمامی شواهد مادی، تبرئه‌کننده و استیضاح‌کننده علیه بی‌گناهی متهم افشا شوند. این الزام بر اساس رویه‌های پس از آن است. حق افشای شواهد به‌ویژه به این دلیل اهمیت دارد که اصل آیین دادرسی تضمین می‌کند که محکومیت کیفری تنها در صورتی امکان‌پذیر است که جرم متهم فراتر از شک معقول اثبات شود. این الزام از این اصل ناشی می‌شود که دادستان نه به‌عنوان یک طرف دعوا که صرفاً به دنبال پیروزی است، بلکه به‌عنوان نماینده یک حاکمیت عمل می‌کند که هدف اصلی آن تحقق عدالت است. چنان‌که دیوان عالی ایالات متحده تصریح کرده است، منفعت دولت در فرایند دادرسی کیفری تنها در اثبات گناه متهم خلاصه نمی‌شود، بلکه در اجرای عدالت نهفته است. عدم شفافیت در این فرایندها می‌تواند نقض مستقیم حقوق دادرسی عادلانه تلقی شود، زیرا به متهمان و وکلای آنها اجازه بررسی، ارزیابی و چالش در برابر داده‌ها و تحلیل‌های مورد استفاده در پرونده را نمی‌دهد. در صورتی که سیستم هوش مصنوعی شفاف نباشد و به صورت جعبه سیاه عمل کند، افشای شواهد به‌راحتی محدود یا مبهم خواهد بود. برای مثال، دادگاه استیناف نیوجرسی^۱ در تصمیم دیگری در خصوص کشف شواهد دستور داد که اگر ایالت قصد داشته باشد از نرم‌افزار تحلیل دی ان ای استفاده کند، متهم حق دسترسی به کد منبع نرم‌افزار و مستندات مرتبط با توسعه آن را تحت یک دستور حفاظتی مناسب خواهد داشت (Garrett & Rudin, 2024: 561). اسنادی نظیر مرجع حفاظت از داده‌های فرانسه، راهنمای اخلاقی هوش مصنوعی در ژاپن (۲۰۱۸) و اعلامیه مونترال برای توسعه مسئولانه هوش مصنوعی در کانادا (۲۰۱۸) بر اصول مشترکی مانند شفافیت، سودمندی برای جامعه، و پاسخگویی تأکید دارند. در ایران، ماده ۳ آیین‌نامه ۱۴۰۴ (بهبود فرایندهای قضایی هوشمند) و

شاخص‌های ارزیابی سند ملی هوش مصنوعی (ارزیابی مسئولیت‌پذیری و توضیح تصمیمات) مسئولیت‌پذیری را تضمین می‌کنند. بنابراین، شفافیت الزامی برای جلوگیری از نقض دادرسی عادلانه است.

نظارت حقوقی و انسانی بر عملکرد سیستم‌های هوش مصنوعی قضایی

از آنجا که تکیه بر تصمیم‌گیری خودکار مشارکت فعال انسانی را محدود می‌کند، این سیستم‌ها تهدیدی اساسی برای مشروعیت ایجاد می‌کنند. بسیاری از صاحب‌نظران نگرانی‌های جدی را به دلیل تفاوت‌های آشکار میان تفکر انسان و استدلال ماشین مطرح کرده‌اند. درک، تفسیر و تحلیل مسائل حقوقی از وظایفی است که به طور عمده به منطق و ادراک انسانی وابسته است، قاضی باید سال‌ها را در آموزش‌های حقوقی سپری کرده و تجربیات عملی قابل توجهی در حوزه مهارت‌های حقوقی و دانش اجتماعی کسب کرده باشد تا به سمت مسئولیت قضایی گمارده شود، اما عملکرد هوش مصنوعی که مبتنی بر تحلیل صرف آماری که بر اساس داده‌های انتخابی انجام می‌شود، صورت می‌گیرد نباید منجر به نتایج نامشروع گردد. نقص‌های انسانی وابسته نیست، اما ممکن است تحت تأثیر تولیدکنندگان نرم‌افزار و داده‌های به صورت هدفمند هدایت شود. با توجه به این ملاحظات، ممکن است استدلال شود که به‌رغم مزایای ملموس و غیرقابل انکار هوش مصنوعی، سطح فعلی آن از نظر مشروعیت حقوقی و تطابق با الزامات تعیین‌شده برای قضات انسانی، فاصله دارد. از این‌رو بهتر است تنها در موارد خاص و با ارزش کم حقوقی یا در قضاوت‌هایی با کنترل انسانی محدود – حداقل در مراحل ابتدایی اجرای آن – مورد استفاده قرار گیرد. احترام به و اجرای مقررات دادگاه‌ها معمولاً بر پایه اعتماد عمومی به دستگاه قضایی استوار است. با این حال، اعتماد اجتماعی ممکن است با دخالت هوش مصنوعی در فرایند دادرسی، به ویژه در پرونده‌های ساده که ممکن است هوش مصنوعی قادر به تصمیم‌گیری سریع‌تر و دقیق‌تر از قاضی انسانی باشد، تقویت شود. این امر می‌تواند منجر به افزایش کارایی سیستم قضایی شود. هوش مصنوعی نباید کنترل کامل فرایندهای قضایی را در دست بگیرد و همیشه باید تحت نظارت و تصمیم‌گیری انسانی باقی بماند. باید مکانیزم‌هایی برای

بررسی، تنظیم، و لغو تصمیمات هوش مصنوعی توسط کاربران وجود داشته باشد و در صورت بروز اشتباه یا سوگیری در نتایج، باید مسئولیت‌پذیری مشخصی برای سیستم‌های هوش مصنوعی تعریف شود تا امکان اعتراض و بررسی مجدد تصمیمات فراهم شود. سند نظارتی مرتبط توصیه شورای اروپا درباره هوش مصنوعی و حفاظت از داده‌ها، کاغذ سفید هوش مصنوعی (۲۰۲۰)، رهنمودهای اخلاقی برای هوش مصنوعی قابل اعتماد (۲۰۱۹) است. این سند بر قابلیت توضیح تصمیمات اتخاذ شده توسط هوش مصنوعی و مسئولیت‌پذیری استفاده‌کنندگان این فناوری تأکید دارد. همچنین، این سند تصریح می‌کند که تصمیمات خودکار هوش مصنوعی باید تحت کنترل انسانی قرار داشته باشند و امکان بررسی و اعتراض به آن‌ها فراهم باشد. این نقش شامل اعمال کنترل نهایی بر سیستم‌های الگوریتمی است تا از حقوق مادی و فرایندی محافظت شود. همچنین هوش مصنوعی نباید قدرت ارزیابی مستقل و نهایی در خصوص شواهد و احکام را داشته باشد. همچنین، هرگونه استفاده از این فناوری‌ها باید به‌گونه‌ای باشد که مغایرتی با ماده ۶ کنوانسیون اروپایی حقوق بشر، که بر حق دادرسی عادلانه تأکید دارد، نداشته باشد. با وجود این، علی‌رغم اینکه انسان‌ها به‌طور کلی به‌عنوان ناظران سیستم‌های فنی خودکار عملکرد ضعیفی دارند، نظارت انسانی می‌تواند به‌راحتی هم در مشخصات فنی و هم در سیستم حقوقی عملیاتی شود (Talapina, 2022: 44). باوجود این، صرف حضور یک ناظر انسانی برای تأیید تصمیمات خودکار کافی نیست؛ زیرا ممکن است این فرایند به یک اقدام تشریفاتی تبدیل شود که در عمل بی‌اثر باشد. به عبارت دیگر، اگر قضات و تصمیم‌گیرندگان صرفاً وظیفه تأیید تصمیمات گرفته‌شده توسط هوش مصنوعی را داشته باشند، ممکن است تمایل یا توانایی آنها برای ارزیابی انتقادی و لغو تصمیم‌های خودکار کاهش یابد. این مسئله ضرورت توسعه استانداردهای دقیق‌تر برای بازبینی واقعی و نه صوری تصمیمات خودکار را برجسته می‌کند (Amarendar Reddy, 2022: 1766). سیستم‌های هوش مصنوعی مورد استفاده در دادرسی کیفری باید قادر به تکرارپذیری نتایج باشند و توانایی مدیریت خطاها و ناسازگاری‌ها را در تمامی مراحل چرخه عمر خود داشته باشند. نظارت انسانی بر آنها الزامی است. هوش مصنوعی نباید استقلال انسانی را تضعیف کند. تنها از طریق مشارکت مناسب انسانی در برنامه‌های پرخطر هوش مصنوعی می‌توان به

کاربردی قابل‌اعتماد، اخلاقی و انسان‌محور دست یافت (Garrett & Rudin, 2024: 561). از این‌رو، چارچوب قانونی روشن برای اعتراض به تصمیمات هوش مصنوعی در دادرسی کیفری و طرق جبران خسارت برای افرادی که متأثر از این تصمیمات هستند، ضروری است که مسئولیت نهایی در صورت بروز خطا یا سوء استفاده از سیستم‌های هوش مصنوعی مشخص گردد. در ایران، ماده ۱۵ آیین‌نامه ۱۴۰۴ (نظارت بر اجرا توسط معاون اول قوه قضاییه)، ماده ۳۴ دستورالعمل ۱۴۰۳ (گزارش پیشرفت به رئیس قوه) و اقدامات ملی سند ملی هوش مصنوعی (ایجاد مکانیزم‌های نظارت و اعتراض) نظارت را تضمین می‌کنند. نظارت انسانی کلیدی برای مشروعیت است، اما نیاز به استانداردهای بازبینی واقعی دارد.

بحث و نتیجه‌گیری

پژوهش حاضر با هدف بررسی نقش هوش مصنوعی در نظام عدالت کیفری ایران و تحلیل چالش‌های ایجاد تعادل بین نوآوری‌های فناورانه و حفظ حقوق بشر انجام شد. یافته‌ها نشان می‌دهند که هوش مصنوعی پتانسیل تحول در فرایندهای قضایی از جمله جمع‌آوری و پردازش شواهد، تصمیم‌گیری‌های پیش‌بینی‌کننده، و تسهیل جلسات رسیدگی را دارد. این فناوری با تحلیل داده‌های عظیم، می‌تواند سرعت، دقت و کارایی نظام قضایی را بهبود بخشد و به کاهش اطاله دادرسی کمک کند. با این حال، چالش‌های حقوقی و اخلاقی نظیر نقض حریم خصوصی، تهدید اصل برائت، و تقویت سوگیری‌های تبعیض‌آمیز؛ به‌ویژه در نبود شفافیت الگوریتمی و نظارت انسانی مؤثر، تهدیدی جدی برای حقوق بنیادین افراد ایجاد می‌کند. استفاده از فناوری‌های نوین در فرایندهای قضایی (مصوب ۱۴۰۴)، و دستورالعمل هوشمندسازی و الکترونیکی شدن فرایندها و پرونده‌های قضایی (مصوب ۱۴۰۳)، چارچوبی برای استفاده مسئولانه از هوش مصنوعی فراهم کرده‌اند. این اسناد بر اصولی مانند شفافیت، نظارت قضایی انسان‌محور، و رعایت ارزش‌های اسلامی و اخلاقی تأکید دارند. با این وجود، فقدان مکانیزم‌های اجرایی دقیق، استانداردهای مشخص برای شفافیت الگوریتم‌ها، و زیرساخت‌های نظارتی مستقل، خطر نقض حقوق متهمان، از جمله حق دفاع (ماده ۳۵ قانون اساسی)، حریم خصوصی (ماده ۵ آیین‌نامه ۱۴۰۴)، را افزایش می‌دهد. مقایسه با چارچوب‌های بین‌المللی مانند مقررات

عمومی حفاظت از داده‌ها (GDPR) و رهنمودهای شورای اروپا نشان‌دهنده کاستی‌هایی در قوانین ایران، از جمله نبود الزامات «حق توضیح» و «حق فراموشی»، است که می‌تواند با استانداردهای جهانی حقوق بشر (ماده ۱۴ میثاق بین‌المللی حقوق مدنی و سیاسی) ناسازگار باشد. فرضیه پژوهش مبنی بر ضرورت ایجاد تعادل بین نوآوری‌های هوش مصنوعی و حفظ حقوق بشر تأیید شد. این تعادل تنها با تقویت نظارت قضایی انسان‌محور، تدوین آیین‌نامه‌های مشخص برای تنظیم الگوریتم‌های پیش‌بینی‌کننده و ایجاد پایگاه داده ملی قضایی با استانداردهای امنیتی بالا (مانند بند ج ماده ۶۵ قانون برنامه هفتم) قابل دستیابی است. بدون این اقدامات، خطراتی مانند جرم‌انگاری پیش‌بینی‌کننده، سوگیری الگوریتمی و گریز الگوریتمی می‌تواند مشروعیت نظام قضایی را تضعیف کند.

پیشنهادهای

برای تحقق تعادل بین نوآوری و حقوق بشر در استفاده از هوش مصنوعی در نظام عدالت کیفری ایران، پیشنهادات زیر ارائه می‌شود:

۱. تدوین آیین‌نامه‌های جامع و مشخص: قوه قضاییه باید بر اساس ماده ۵ سند ملی هوش مصنوعی و ماده ۱۱۳ قانون برنامه هفتم توسعه، آیین‌نامه‌هایی برای تنظیم الگوریتم‌های پیش‌بینی‌کننده تدوین کند که شامل الزامات شفافیت، تفسیرپذیری و افشای کد منبع (در صورت امکان) باشد. این آیین‌نامه‌ها باید استانداردهای فنی را برای کاهش سوگیری الگوریتمی اجباری کنند.

۲. ایجاد پایگاه داده ملی قضایی: بر اساس بند ج ماده ۶۵ قانون برنامه هفتم، یک پایگاه داده ملی قضایی با استانداردهای امنیتی بالا (مانند رمزنگاری پیشرفته و احراز هویت چندمرحله‌ای) ایجاد شود تا از سوءاستفاده داده‌های حساس و نقض حریم خصوصی جلوگیری شود. این پایگاه باید تحت نظارت نهادهای مستقل عمل کند.

۳. آموزش جامع قضات و کارکنان قضایی: مطابق ماده ۳ آیین‌نامه ۱۴۰۴، برنامه‌های آموزشی ضمن خدمت برای قضات

و کارکنان قضایی طراحی شود تا توانایی ارزیابی انتقادی خروجی‌های الگوریتمی و درک محدودیت‌های آنها را داشته باشند. این آموزش‌ها باید شامل مباحث روان‌شناختی مانند گریز الگوریتمی و اصول فنی مانند تفسیرپذیری الگوریتم‌ها باشد.

۴. تقویت نظارت قضایی انسان‌محور: ماده ۱۵ آیین‌نامه ۱۴۰۴ و ماده ۳۴ دستورالعمل ۱۴۰۳ باید با ایجاد نهادهای نظارتی مستقل و پروتکل‌های بازبینی واقعی (مانند cross-validation علمی) تقویت شوند تا از تصمیم‌گیری‌های خودکار بدون نظارت جلوگیری شود. این نهادها باید امکان اعتراض متهمان به خروجی‌های الگوریتمی را فراهم کنند.

۵. تدوین مکانیزم‌های جبران خسارت: برای حمایت از حقوق متهمان، مکانیزم‌های جبران خسارت در صورت نقض حقوق بنیادین (مانند حریم خصوصی یا اصل برائت) ناشی از خطاهای الگوریتمی ایجاد شود. این مکانیزم‌ها باید در آیین‌نامه‌های جدید قوه قضاییه پیش‌بینی شوند.

۶. هماهنگی سیاستی بین نهادها: برای رفع دوگانگی بین شورای عالی انقلاب فرهنگی و شورای عالی فضای مجازی، یک ستاد هماهنگی ملی تحت نظارت قوه قضاییه تشکیل شود تا سیاست‌گذاری‌های مرتبط با هوش مصنوعی یکپارچه گردد و اجرای بند ج ماده ۶۵ قانون برنامه هفتم تسریع شود.

تطبیق با استانداردهای بین‌المللی: ایران باید الزامات «حق توضیح» و «حق فراموشی» را مشابه GDPR در قوانین خود (مانند قانون مدیریت داده‌ها و اطلاعات ملی) بگنجاند تا با استانداردهای جهانی حقوق بشر (ماده ۱۴ و ۱۷ میثاق بین‌المللی حقوق مدنی و سیاسی) هم‌راستا شود.

۷. با اجرای این پیشنهادات، هوش مصنوعی می‌تواند به‌عنوان ابزاری کمکی تحت نظارت دقیق انسانی عمل کند تا ضمن بهره‌مندی از مزایای آن، اصول عدالت اسلامی، کرامت انسانی و دادرسی عادلانه حفظ شود. این رویکرد نه تنها نوآوری را در خدمت عدالت قرار می‌دهد، بلکه از نقض حقوق بنیادین جلوگیری کرده و اعتماد عمومی به نظام عدالت کیفری را در برابر چالش‌های فناوری مدرن تقویت می‌کند.

- Seventh Five-Year National Development Plan of the Islamic Republic of Iran (2024–2028).
- National Data and Information Management Law (2022).
- Executive Regulation on the Use of New Technologies in Judicial Proceedings (2025).
- Directive on the Preservation of Human Dignity and Values (2019).
- Directive on the Organization of Prisoners and Reduction of the Criminal Population (2019).
- Directive on the Intelligent Transformation and Electronicization of Judicial Processes (2021).

References

- Amarendar Reddy, A. (2022). "Privacy as a fundamental right: A case study on biometric data and facial recognition". *Journal of Constitutional Law*, 12(4), pp. 1760–1778.
- Avramovic, D. S., & Jovanov, I. D. (2023). "Ethical principles for trustworthy AI: A global perspective". *International Journal of Law and Technology*, 15(2), pp. 155–170.
- Bampasika, E., & Sadeghi, S. (2023). "Artificial intelligence as evidence in criminal trials". [Journal name not provided]. (in Persian)
- Bagaric, M., Hunter, D., & Stobbs, N. (2022). "Artificial intelligence in criminal justice: Balancing efficiency and fairness". *Criminal Law Review*, 44(1), pp. 85–110.
- Danielle, K. (2008). "The role of technology in judicial decision-making: Opportunities and risks". *Judicial Studies Journal*, 30(5), pp. 1265–1280.
- Elliott, M., & Soifer, E. (2022). "Privacy in the digital age: Legal and ethical considerations". *Global Privacy Law Review*, 3(1), pp. 1–15.
- Garrett, B. L., & Rudin, C. (2024). "Algorithmic transparency in criminal justice: Challenges and solutions". *Yale Law Journal*, 133(3), pp. 550–580.
- Hedges, R., Gottehrer, G., & Francis, J. C. (2020). "Artificial intelligence and due process: The COMPAS case revisited". *American Journal of Criminal Law*, 47(1), pp. 1–20.
- Hosseini, S. A., Abdolkhodai, Z., & Sharifkhani, M. (2023). "Application of artificial intelligence in judicial proceedings: Transparency challenges and solutions". *Judicial Law Perspectives Quarterly*, 28(101), pp. 67–90. (in Persian)
- Hosseini, S. M., Sabet Moghaddam, M., & Asgharnia, A. (2023). "Artificial intelligence in Iran's judicial system: Legal and ethical frameworks". *Iranian Journal of Legal Studies*, 5(2), pp. 75–90. (in Persian)
- Hunter, D., Bagaric, M., & Stobbs, N. (2020). "Algorithm aversion in criminal justice: Psychological barriers to AI adoption". *Behavioral Sciences & the Law*, 38(4), pp. 750–765.
- Kaplan, J., Smith, A., & Brown, L. (2023). "Predictive policing and algorithmic bias: A global perspective". *Journal of Criminology and Technology*, 22(3), pp. 140–160.
- Mirshkari, A., Sabetghadam, F., & Asgharnia, M. (2024). "An overview of artificial intelligence technology challenges in the field of privacy". *Journal of Cyber Law Studies*, 12(3), pp. 71–89. (in Persian)
- Ming, Z. (2022). "AI-assisted judicial decision-making: Opportunities and limitations". *Asian Journal of Law and Technology*, 10(1), pp. 30–45.
- Sourdin, T. (2018). "Technology and the future of judicial processes". *Law, Technology and Humans*, 1(2), pp. 1115–1130.

- Spiridonov, M. S. (2023). "AI in criminal justice: Transparency and accountability challenges". *Russian Journal of Legal Studies*, 7(4), pp. 475–490.
- Talapina, E. V. (2022). "Ethical AI in public administration: Mitigating discrimination risks". *European Journal of Administrative Law*, 14(1), pp. 1–10.
- Tumanyants, O., Tumanyants, A., Krytska, I., & Verkhoglyad-Gerasymenko, O. (2023). "Biometric data and privacy in criminal justice: Lessons from European case law". *European Human Rights Law Review*, 25(2), pp. 135–150.