Islamic Maaref University

Scientific Journal

PAZHUHESH NAME-E AKHLAQ

Vol. 18, Sprign 2025, No. 67

Legal Challenges Arising from Artificial Intelligence and Ethical Technology Based Solutions

Mahdieh Latifzadeh [1]

1. Assistant Professor, Department of Private Law, Research Group of Islamic Jurisprudence and Law, Research Institute for Islamic Studies in the Humanities, Ferdowsi University of Mashhad, Mashhad, Iran.
latifzadeh@um.ac.ir

Abstract Info

Article Type:
Research Article

Received:
2024/09/25

Accepted:
2025/03/11

Abstract

The rapid advancement of artificial intelligence technology and its deep penetration into various layers of social life have confronted legal systems worldwide with unprecedented challenges. Automated decision-making, large-scale processing of personal data, and the autonomous functioning of intelligent systems have challenged traditional legal concepts such as liability, consent, and compensation for damages. In this context, ethical principles and norms can play a key role in preventing and managing these challenges. The central issue of this research concerns emerging legal challenges resulting from the development and application of AI systems in different domains of social life, raising the question of how ethical principles and considerations can function as preventive or moderating factors in reducing AI-related legal challenges. Employing a descriptive–analytical method and drawing on library resources, legal documents, and judicial practices, the study examines three major legal challenges: violations of informational privacy and data protection, conflicts between individual rights and collective interests, and cybersecurity threats in intelligent infrastructures. The findings demonstrate that adopting an integrated approach combining regulatory frameworks, technical mechanisms, and ethical strategies represents the most effective method for addressing these legal challenges. The study concludes that artificial intelligence technology must consistently align with human values and reinforce the foundations of human rights.

Keywords Technology ethics; artificial intelligence regulation; personal data; EU Artificial Intelligence Act; privacy; cybersecurity.

Cite this article: Latifzadeh, Mahdieh (2025). Legal Challenges Arising from Artificial Intelligence and Ethical Technology Based Solutions. *Pazhuhesh Name-E Akhlaq*. 18 (1). 169-188. DOI: 10.22034/18.67.169

DOI: https://doi.org/10.22034/18.67.169

Publisher: Islamic Maaref University, Qom, Iran.

مجلة علمية
النشرة الاخلاقية
السنة ١٨ / ربيع عام ١٤٤٦ / العدد ٦٧
جامعة المعارف الإسلامية

التحدّيات القانونيّة الناشئة عن الذكاء الاصطناعي وحلولها في ضوء أخلاقيّات التكنولوجيا

مهديّه لطيف‌زاده [١]

١. أستاذة مساعدة، قسم القانون الخاص، المجموعة البحثيّة للفقه والقانون الإسلامي، معهد دراسات العلوم الإنسانيّة الإسلاميّة، جامعة فردوسي مشهد، مشهد، إيران.
latifzadeh@um.ac.ir

معلومات المادة

نوع المقال: بحث

تاريخ الاستلام: ١٤٤٦/٠٣/٢١

تاريخ القبول: ١٤٤٦/٠٩/١٠

ملخّص البحث

أدّى التطوّر المتسارع لتقنيّات الذكاء الاصطناعي وتغلغلها العميق في مختلف مناحي الحياة الاجتماعيّة إلى مواجهة الأنظمة القانونيّة العالميّة لتحدّيات غير مسبوقة. فقد وضعت عمليّات اتّخاذ القرار الآلي، والمعالجة الواسعة للبيانات الشخصيّة، والأداء المستقلّ للأنظمة الذكيّة، مفاهيمَ قانونيّة تقليديّة ـ كالمسؤوليّة، والرضا، والتعويض ـ موضع تساؤل. وفي هذا السياق، يمكن للمبادئ والمعايير الأخلاقيّة أن تؤدّي دورًا محوريًا في الوقاية من هذه التحدّيات وإدارتها. وتتمحور إشكاليّة البحث حول التحدّيات القانونيّة المستجدّة الناشئة عن تطوير وتطبيق أنظمة الذكاء الاصطناعي في مجالات الحياة الاجتماعيّة المختلفة، مع التساؤل عن كيفيّة إسهام المبادئ الأخلاقيّة في الحدّ من هذه التحدّيات أو تعديل آثارها. اعتمد البحث المنهج الوصفي ـ التحليلي، مستندًا إلى المصادر المكتبيّة، والوثائق القانونيّة، والسوابق القضائيّة، ودرس ثلاثة تحدّيات قانونيّة رئيسة: انتهاك الخصوصيّة المعلوماتيّة وحماية البيانات، التعارض بين الحقوق الفرديّة والمصالح العامّة، والتهديدات السيبرانيّة في البيئات الذكيّة. وتشير النتائج إلى أنّ تبنّي مقاربة تكامليّة تجمع بين الأطر التنظيميّة، والآليّات التقنيّة، والاستراتيجيّات الأخلاقيّة يُعَدّ أنجع السبل لمواجهة هذه التحدّيات. وتخلص الدراسة إلى أنّ تقنيّات الذكاء الاصطناعي ينبغي أن تكون منسجمة دومًا مع القيم الإنسانيّة، وأن تُسهم في تعزيز أسس الحقوق الإنسانيّة.

الألفاظ المفتاحية

أخلاقيّات التكنولوجيا، تنظيم الذكاء الاصطناعي، البيانات الشخصيّة، قانون الذكاء الاصطناعي للاتحاد الأوروبي، الخصوصيّة، الأمن السيبراني.

الاقتباس:

لطيف‌زاده، مهديّه (١٤٤٦). التحدّيات القانونيّة الناشئة عن الذكاء الاصطناعي وحلولها في ضوء أخلاقيّات التكنولوجيا. *مجلة علمية النشرة الاخلاقية*. ١٨ (١). ١٨٨ ـ ١٦٩. DOI: 10.22034/18.67.169

رمز DOI: https://doi.org/10.22034/18.67.169

الناشر: جامعة المعارف الإسلامية، قم، ايران.

نشریه علمی
پژوهش‌نامه اخلاق
سال ۱۸، بهار ۱۴۰۴، شماره ۶۷

چالش‌های حقوقی برخاسته از هوش مصنوعی و راهکارهای برون‌رفت از آن بر اساس اخلاق فناوری

مهدیه لطیف‌زاده [۱]

۱. استادیار، گروه حقوق خصوصی، گروه پژوهشی فقه و حقوق اسلامی، پژوهشکده مطالعات اسلامی در علوم انسانی دانشگاه فردوسی مشهد، مشهد، ایران.
latifzadeh@um.ac.ir

اطلاعات مقاله

نوع مقاله: پژوهشی
(۱۸۸ ـ ۱۶۹)

تاریخ دریافت:
۱۴۰۳/۰۷/۰۴

تاریخ پذیرش:
۱۴۰۳/۱۲/۲۱

چکیده

پیشرفت شتابان فناوری هوش مصنوعی و نفوذ عمیق آن در لایه‌های مختلف زندگی اجتماعی، نظام‌های حقوقی جهان را با چالش‌های بی‌سابقه‌ای مواجه ساخته است. تصمیم‌گیری‌های خودکار، پردازش انبوه داده‌های شخصی و عملکرد مستقل سیستم‌های هوشمند، مرزهای متعارف مفاهیم حقوقی همچون مسئولیت، رضایت و جبران خسارت را به چالش کشیده است. در این میان، اصول و هنجارهای اخلاقی می‌توانند نقشی کلیدی در پیشگیری و مدیریت این چالش‌ها ایفا نمایند. مسئله اصلی پژوهش، چالش‌های حقوقی نوظهوری است که در اثر توسعه و کاربرد سیستم‌های هوش مصنوعی در حوزه‌های مختلف حیات اجتماعی پدیدار گشته و این پرسش را مطرح می‌سازد که چگونه اصول و ملاحظات اخلاقی می‌توانند به‌عنوان عامل پیشگیرانه یا تعدیل‌کننده، چالش‌های حقوقی هوش مصنوعی را کاهش دهند. روش پژوهش توصیفی ـ تحلیلی است و با بهره‌گیری از منابع کتابخانه‌ای، اسناد حقوقی و رویه قضایی، سه چالش عمده حقوقی شامل نقض حریم خصوصی اطلاعاتی و حفاظت از داده‌ها، تعارض میان حقوق فردی و منافع جمعی، و تهدیدات امنیت سایبری در بسترهای هوشمند را مورد بررسی قرار داده است. یافته‌ها نشان می‌دهد که اتخاذ رویکردی تلفیقی متشکل از چهارچوب‌های تنظیم‌گری، سازوکارهای فنی و راهبردهای اخلاقی، می‌تواند کارآمدترین شیوه برای مواجهه با چالش‌های حقوقی یادشده باشد. نتیجه پژوهش آن است که فناوری هوش مصنوعی باید همواره هم‌سو با ارزش‌های بشری باشد و بنیان حقوق انسانی را مستحکم سازد.

واژگان کلیدی
اخلاق فناوری، تنظیم‌گری هوش مصنوعی، داده شخصی، قانون هوش مصنوعی اتحادیه اروپا، حریم خصوصی، امنیت سایبری.

استناد:
لطیف‌زاده، مهدیه (۱۴۰۴). چالش‌های حقوقی برخاسته از هوش مصنوعی و راهکارهای برون‌رفت از آن بر اساس اخلاق فناوری. *پژوهشنامه اخلاق*. ۱۸ (۱). ۱۸۸ ـ ۱۶۹. DOI: 10.22034/18.67.169

کد DOI:
https://doi.org/10.22034/18.67.169

ناشر:
دانشگاه معارف اسلامی، قم، ایران.

طرح مسئله

در جامعه معاصر، فناوری نقشی محوری در زندگی روزمره ایفا می‌کند و اخلاق هوش مصنوعی اهمیتی فزاینده یافته است. اخلاق هوش مصنوعی مجموعه‌ای از اصول و رهنمودهایی است که طراحی، توسعه و کاربرد مسئولانه سیستم‌های هوشمند را هدایت می‌کند. با پیشرفت سریع فناوری هوش مصنوعی، این حوزه اخلاقی به‌عنوان چهارچوبی راهبردی برای تضمین عملکرد مسئولانه افراد، سازمان‌ها و جوامع در تعامل با سیستم‌های هوشمند عمل می‌کند. به دیگر سخن گسترش روزافزون کاربردهای هوش مصنوعی در حوزه‌های مختلف زندگی و ظهور قابلیت‌های پیشرفته در این سیستم‌ها، ضرورت توجه به اخلاق هوش مصنوعی را برجسته‌تر ساخته است. این ضرورت از آنجا ناشی می‌شود که توسعه سریع سیستم‌های هوشمند، همواره با خطر سوءاستفاده و آسیب‌رسانی همراه است و تحقق اخلاق هوش مصنوعی می‌تواند به‌عنوان سپری محافظتی در برابر این مخاطرات عمل کند. در این راستا بررسی‌های انجام‌شده نشان می‌دهد که این حوزه اخلاقی به دلیل گستردگی موضوعات مرتبط با فناوری هوش مصنوعی، دربرگیرنده ابعاد متعددی است (Marghalani, 2019). این ابعاد متعدد شامل جنبه‌های فنی (مانند شفافیت الگوریتمی)، حقوقی (مانند مسئولیت‌پذیری در برابر تصمیمات خودکار)، اجتماعی (مانند جلوگیری از تبعیض الگوریتمی) و فرهنگی (مانند احترام به ارزش‌های متنوع انسانی در طراحی سیستم‌های هوشمند) می‌شود. در این میان، تعامل میان نظام حقوقی و سیستم‌های هوش مصنوعی، یکی از مهم‌ترین حوزه‌های چالش‌برانگیز است. منظور از مواجهه بسترهای فناورانه هوش مصنوعی با عرصه حقوق، نقاط تلاقی‌ای است که این فناوری با الزامات حقوقی پیدا می‌کند. برای مثال، زمانی که یک سیستم هوش مصنوعی برای تصمیم‌گیری اعتباری در بانک‌ها استفاده می‌شود، باید مطابق با قوانین منع تبعیض و حفاظت از داده عمل نماید. رعایت این الزامات حقوقی (مانند شفافیت در فرایند تصمیم‌گیری یا امکان توضیح‌پذیری تصمیمات) به تحقق اخلاق هوش مصنوعی کمک می‌کند، زیرا حقوق افراد را محترم می‌شمارد و از تبعیض ناعادلانه جلوگیری می‌نماید.

با توجه به آنچه بیان شد، پژوهش حاضر با رویکردی توصیفی ـ تحلیلی به بررسی نقش اخلاق هوش مصنوعی در پیشگیری از چالش‌های حقوقی در حوزه این فناوری می‌پردازد. این چالش‌ها شامل مسائل مربوط حریم خصوصی اطلاعاتی و حفاظت از داده، تعارض میان حقوق فردی و منافع جمعی و تهدیدات امنیت سایبری در بسترهای هوشمند است. شناسایی دقیق این چالش‌ها و ارائه راهکارهای مناسب می‌تواند به تحقق اخلاق هوش مصنوعی کمک کند و متقابلاً، پایبندی به اصول اخلاقی در طراحی و توسعه هوش مصنوعی می‌تواند به کاهش چالش‌های حقوقی در تعامل با این فناوری بینجامد

(Floridi et al., 2019: 10). مسائل اصلی این پژوهش از جمله این است که چالش‌های حقوقی عمده در حوزه هوش مصنوعی کدامند. اخلاق هوش مصنوعی چگونه می‌تواند در پیشگیری از چالش‌های حقوقی این فناوری نقش‌آفرینی کند. همچنین چه راهکارهایی برای مواجهه مؤثر با چالش‌های حقوقی سیستم‌های هوش مصنوعی وجود دارد؟ در پاسخ به این پرسش‌ها، فرضیه‌های پژوهش بر این اساس استوار است که پایبندی به اصول اخلاقی در طراحی و توسعه هوش مصنوعی می‌تواند به کاهش چالش‌های حقوقی در تعامل با این فناوری بینجامد و همچنین رعایت الزامات حقوقی مانند شفافیت در فرایند تصمیم‌گیری و امکان توضیح‌پذیری تصمیمات، به تحقق اخلاق هوش مصنوعی کمک می‌کند. همچنین هدفِ پژوهش حاضر، بررسی و ارائه راهکارهای مؤثر برای مواجهه با چالش‌های حقوقی سیستم‌های هوش مصنوعی با تأکید بر نقش تکمیلی اخلاق هوش مصنوعی است. بدین منظور، ابتدا به تبیین چالش‌های حقوقی عمده در حوزه هوش مصنوعی پرداخته خواهد شد و سپس راهکارهایی مؤثر برای کاهش این چالش‌ها با محوریت اصول اخلاقی ارائه خواهد شد. جنبه نوآوری این پژوهش در بررسی هم‌زمان و تلفیقی حوزه‌های اخلاق و حقوق در مواجهه با چالش‌های هوش مصنوعی است که این رویکرد بین‌رشته‌ای می‌تواند به‌عنوان چهارچوبی راهنما برای سیاست‌گذاران، توسعه‌دهندگان و کاربران هوش مصنوعی عمل کند.

بحث

الف) چالش‌های حقوقی فناوری‌های هوش مصنوعی

چالش‌هایی که با گسترش سیستم‌های هوش مصنوعی در زمینه حقوق ایجاد شده‌اند، متعدد می‌باشند، لیکن برخی از آنها دامنه‌ای وسیع‌تر و اهمیت بیشتری دارند که در ادامه به تبیین این چالش‌ها از منظر حقوقی پرداخته می‌شود.

۱. مسائل مربوط به حفظ حریم خصوصی اطلاعاتی و حفاظت از داده

حریم خصوصی اطلاعاتی و حمایت از داده شخصی، حق بنیادینی است که در اسناد متعدد بین‌المللی مورد شناسایی قرار گرفته است. ماده ۱۲ اعلامیه جهانی حقوق بشر[1] و ماده ۱۷ میثاق بین‌المللی حقوق مدنی و سیاسی،[2] صراحتاً بر حق افراد در محافظت در برابر مداخله خودسرانه در حریم خصوصی تأکید نموده‌اند (UN General Assembly, 1948 & UN General Assembly, 1966). در

1. Universal Declaration of Human Rights.
2. International Covenant on Civil and Political Rights.

نظام‌های حقوقی داخلی نیز، این حق در قالب قوانین متعدد حفاظت از داده‌ها مانند مقررات عمومی حفاظت از داده اتحادیه اروپا[1] و قانون حریم خصوصی مصرف‌کننده کالیفرنیا[2] تجلی یافته است (European Parliament & Council, 2016 & California Legislature, 2018).

حریم خصوصی اطلاعاتی به‌عنوان حق کنترل افراد بر اطلاعات شخصی خود تعریف می‌شود که مستلزم رعایت اصول حقوقی مهمی همچون اصل رضایت آگاهانه، اصل محدودیت هدف، اصل حداقل‌سازی داده و اصل پاسخ‌گویی می‌باشد. این اصول در نظام‌های حقوقی مختلف به‌عنوان مبنای قانون‌گذاری در زمینه حفاظت از داده‌ها مورد استناد قرار گرفته‌اند. در عصر هوش مصنوعی، همان‌طور که فناوری هوش مصنوعی بیشتر با حوزه‌های حقوقی ادغام می‌شود، چالش‌های قانونی متعددی در خصوص حفاظت از داده‌های اشخاص موضوع داده بروز می‌نماید. یکی از مهم‌ترین این چالش‌ها، تعارض میان نیاز سیستم‌های هوش مصنوعی به حجم گسترده داده‌ها و اصل حداقل‌سازی داده در قوانین حفاظت از داده‌ها می‌باشد. به‌عنوان مثال، ماده ۵ مقررات عمومی حفاظت از داده صراحتاً بر لزوم کافی، مرتبط و محدود بودن داده‌های جمع‌آوری‌شده به اهداف مشخص تأکید دارد، حال آنکه کارآمدی سیستم‌های هوش مصنوعی اغلب مستلزم دسترسی به داده‌های گسترده است. چالش حقوقی دیگر در این زمینه، مسئله انتقال فرامرزی داده‌ها می‌باشد. در نظام‌های حقوقی متعدد، انتقال داده‌های شخصی به خارج از حوزه قضایی تابع مقررات خاصی است. به‌عنوان نمونه، فصل پنجم مقررات پیش‌گفته، شرایط و محدودیت‌های انتقال داده‌های شخصی به کشورهای ثالث را مشخص نموده است. با توجه به ماهیت جهانی سیستم‌های هوش مصنوعی و ضرورت استفاده از داده‌های متنوع، رعایت این مفاد قانونی، چالش‌های قانونی متعددی را ایجاد می‌نماید. همچنین در زمینه فناوری‌های هوش مصنوعی، مسئله شفافیت و قابلیت توضیح تصمیمات از منظر حقوقی اهمیت ویژه‌ای دارد. ماده ۲۲ مقررات مذکور، حق افراد را در عدم قرارگیری تحت تصمیم‌گیری خودکار با آثار حقوقی یا پیامدهای مشابه به رسمیت شناخته است. این ماده همچنین بر لزوم اتخاذ تدابیر مناسب برای حفاظت از حقوق و آزادی‌های افراد تأکید دارد. با این حال، ماهیت پیچیده و «جعبه سیاه» برخی از الگوریتم‌های هوش مصنوعی، تحقق این الزام قانونی را با چالش مواجه می‌سازد (European Parliament & Council, 2016).

افزون بر این، چالش حقوقی دیگر، مسئله سوگیری و تبعیض در سیستم‌های هوش مصنوعی از منظر حقوقی است. این موضوع با اصل منع تبعیض که در اسناد متعدد حقوق بشری و قوانین داخلی مورد

1. General Data Protection Regulation (GDPR).
2. California Consumer Privacy Act (CCPA).

تأکید قرار گرفته، در ارتباط است. به‌عنوان مثال، ماده ۱۴ کنوانسیون اروپایی حقوق بشر[1] و ماده ۲۱ منشور حقوق بنیادین اتحادیه اروپا[2] (European Parliament, Council & Commission, 2000 & Council of Europe, 1950)، صراحتاً هرگونه تبعیض را منع نموده‌اند. سیستم‌های هوش مصنوعی که بر اساس داده‌های دارای سوگیری آموزش دیده‌اند، می‌توانند منجر به تصمیماتی شوند که از منظر حقوقی، نقض اصل منع تبعیض محسوب می‌گردند (See. Turillazzi et al., 2023: 90). در واقع سیستم‌های هوش مصنوعی که از داده‌های شخصی برای تصمیم‌گیری استفاده می‌کنند، باید علاوه بر رعایت قوانین حفاظت از داده‌ها، مطابق با قوانین منع تبعیض نیز عمل نمایند. به‌عنوان مثال، استفاده از داده‌های مربوط به نژاد، جنسیت یا مذهب در یک سیستم استخدامی مبتنی بر هوش مصنوعی، می‌تواند هم نقض حریم خصوصی (در صورت عدم رعایت اصول حفاظت از داده) و هم نقض اصل منع تبعیض محسوب گردد. در پرونده‌های متعددی در حوزه قضایی آمریکا، استفاده از الگوریتم‌های تصمیم‌گیر با سوگیری نژادی به‌عنوان نقض قانون فرصت‌های برابر استخدامی[3] مورد چالش قرار گرفته است (U. S. Congress, 1964).

۲. مسائل مربوط به تعارض بین حقوق فردی و منافع جمعی

تعارض میان حقوق فردی و منافع جمعی در حوزه فناوری هوش مصنوعی، از منظر حقوقی چالش‌های متعددی را ایجاد می‌نماید. این تعارض در نظام‌های حقوقی مختلف به رسمیت شناخته شده و قوانین متعددی برای ایجاد توازن میان این دو دسته از حقوق و منافع وضع گردیده است. حقوق فردی در زمینه داده‌های شخصی در اسناد متعدد بین‌المللی و قوانین داخلی مورد حمایت قرار گرفته است. ماده ۸ منشور حقوق بنیادین اتحادیه اروپا صراحتاً بر حق حفاظت از داده‌های شخصی تأکید نموده و بیان می‌دارد: «هر فرد حق دارد از داده‌های شخصی مربوط به خود محافظت کند» (European Parliament, Council & Commission, 2000). همچنین ماده نخست مقررات عمومی حفاظت از داده اتحادیه اروپا نیز حفاظت از حقوق و آزادی‌های بنیادین افراد، به‌ویژه حق آنها بر حفاظت از داده‌های شخصی را به‌عنوان یکی از اهداف اصلی این مقررات معرفی می‌نماید. در مقابل، منافع جمعی نیز در قوانین مختلف مورد توجه قرار گرفته است. به‌عنوان مثال، ماده ۲۳ مقررات پیش‌گفته به کشورهای عضو اجازه می‌دهد تا در مواردی مانند امنیت ملی، دفاع، امنیت عمومی یا اهداف مهم منافع عمومی، محدودیت‌هایی را بر حقوق فردی در زمینه داده‌های شخصی اعمال نمایند (European Parliament & Council, 2016).

---

1. European Convention on Human Rights.
2. Charter of Fundamental Rights of the European Union.
3. Equal Employment Opportunity Act.

همچنین، در نظام حقوقی ایالات متحده، قانون آزادی اطلاعات[1] و قانون حفظ حریم خصوصی،[2] تلاش نموده‌اند تا توازنی میان حق دسترسی عموم به اطلاعات دولتی و حفاظت از حریم خصوصی افراد ایجاد نمایند (U. S. Congress, 1966 & U. S. Congress, 1974).

تعارض حقوقی مذکور در زمینه هوش مصنوعی به شکل خاصی نمود می‌یابد. به‌عنوان مثال، استفاده از داده‌های شخصی برای آموزش سیستم‌های هوش مصنوعی که می‌تواند منافع جمعی مانند پیشرفت علمی یا بهبود خدمات عمومی را به همراه داشته باشد، ممکن است با حقوق فردی مانند حق کنترل بر داده‌های شخصی در تعارض قرار گیرد. در این زمینه، ماده ۸۹ مقررات عمومی حفاظت از داده، شرایط خاصی را برای پردازش داده‌های شخصی با اهداف تحقیقاتی، آماری یا آرشیوی در راستای منافع عمومی مشخص نموده است (See. Wagner & Benecke, 2017: 354).

در این راستا دادگاه‌های مختلف نیز در آرای خود به این تعارض پرداخته‌اند. به‌عنوان نمونه، دیوان دادگستری اتحادیه اروپا[3] در پرونده Google Spain SL v AEPD and Mario Costeja González (۲۰۱۴) بر لزوم ایجاد توازن میان حق فرد بر محو داده‌های شخصی (حق فراموش شدن) و حق عموم بر دسترسی به اطلاعات تأکید نموده است. همچنین، در پرونده La Quadrature du Net and Others v Premier ministre (۲۰۲۰)، این دیوان مقرر داشت که نگهداری فراگیر و تفکیک‌نشده داده‌های ترافیکی و مکانی، حتی با هدف مبارزه با جرایم سنگین و حفظ امنیت ملی با حقوق بنیادین مندرج در منشور حقوق بنیادین اتحادیه اروپا سازگار نیست. توضیح بیشتر اینکه پرونده Google Spain SL v AEPD and Mario Costeja González (۲۰۱۴) یک پرونده مهم است که «حق فراموش شدن» را در قوانین اتحادیه اروپا بنیان نهاد. ماریو کاستخا گونزالس متوجه شد که جست‌وجوی نام او در گوگل، پیوندهایی به مقالات روزنامه‌ای از سال ۱۹۹۸ درباره بدهی‌های تأمین اجتماعی او را نمایش می‌دهد. با وجود اینکه مسائل مربوط به بدهی سال‌ها قبل حل شده بود، اطلاعات دیجیتال همچنان قابل دسترسی بود. گونزالس شکایتی علیه روزنامه و گوگل نزد آژانس حفاظت از داده‌های اسپانیا (AEPD) مطرح کرد و خواستار حذف این اطلاعات شد. دیوان دادگستری اتحادیه اروپا حکم داد که موتورهای جست‌وجو «کنترل‌کننده» داده‌های شخصی به موجب دستورالعمل حفاظت از داده هستند. دادگاه اعلام کرد که افراد حق دارند از موتورهای جست‌وجو بخواهند پیوندهای حاوی اطلاعات شخصی را حذف کنند، زمانی که چنین اطلاعاتی نادرست، ناکافی، نامربوط یا بیش از حد برای اهداف پردازش

1. Freedom of Information Act (FOIA).
2. Privacy Act.
3. Court of Justice of the European Union (CJEU).

داده باشد. این حق باید با حق عموم برای دسترسی به اطلاعات متوازن شود، به‌ویژه اگر فرد نقشی در زندگی عمومی داشته باشد (Court of Justice of the European Union, 2014).

همچنین پرونده La Quadrature du Net and Others v Premier ministre (۲۰۲۰) به چالش قوانین نظارتی فرانسه، بلژیک و انگلستان پرداخت که نگهداری موسع داده‌های ارتباطی شهروندان را الزامی می‌کرد. دیوان دادگستری اتحادیه اروپا در حکمی، تعارض میان امنیت ملی و حریم خصوصی را بررسی کرد و اعلام نمود نگهداری فراگیر داده‌های ترافیکی و مکانی، مداخله‌ای جدی در حقوق بنیادین است که می‌تواند جزئیات دقیقی از زندگی خصوصی افراد را آشکار سازد. دادگاه ضمن شناسایی اهمیت امنیت ملی، رویکردی متعادل اتخاذ کرد و اقدامات نظارتی را تنها در شرایط استثنایی، با محدودیت زمانی، هدفمندی مشخص، نظارت قضایی و تضمین‌های کافی مجاز دانست. این پرونده اصل تناسب را در مرکز تصمیم‌گیری‌های مربوط به نظارت دیجیتال قرار داد و الگویی جهانی برای ایجاد توازن میان منافع فردی و جمعی ارائه کرد و تأکید نمود که حتی اهداف مشروع امنیتی نمی‌توانند توجیهی برای نقض سیستماتیک حقوق بنیادین شهروندان باشند (Court of Justice of the European Union, 2020).

فارغ از آنچه بیان شد، نقض حریم خصوصی و عدم رعایت تعادل میان حقوق فردی و منافع جمعی، علاوه بر پیامدهای حقوقی مانند جریمه‌های مالی سنگین (به‌عنوان نمونه جریمه ۵۰ میلیون یورویی گوگل بر اساس الزامات حقوقی مقررات اروپایی حفاظت از داده) می‌تواند منجر به کاهش اعتماد عمومی به فناوری‌های نوین گردد. این امر به نوبه خود، توسعه و پذیرش این فناوری‌ها را با چالش مواجه می‌سازد (See. Jarvis, 2019: 43).

۳. مسائل مربوط به تهدیدهای امنیت سایبری در بسترهای هوشمند

تهدیدهای امنیت سایبری در عصر هوش مصنوعی، چالش‌های حقوقی ویژه‌ای را ایجاد می‌نمایند که متمایز از تهدیدات سنتی امنیت سایبری می‌باشند. این چالش‌ها از منظر حقوقی، مستلزم بررسی دقیق و تدوین چهارچوب‌های قانونی متناسب با پیچیدگی‌های فناوری هوش مصنوعی هستند (Taddeo et al., 2019). یکی از چالش‌های حقوقی اختصاصی در زمینه امنیت سایبری سیستم‌های هوش مصنوعی، مسئله حملات خصمانه[1] می‌باشد. در این نوع حملات، مهاجمان با دستکاری هوشمندانه داده‌های ورودی، سیستم‌های هوش مصنوعی را به اتخاذ تصمیمات نادرست هدایت می‌کنند. از منظر حقوقی، این مسئله چالش‌های جدیدی را در زمینه تعیین مسئولیت مدنی و کیفری ایجاد می‌نماید. به‌عنوان مثال، در صورت بروز خسارت ناشی از تصمیم نادرست یک سیستم هوش مصنوعی که مورد حمله خصمانه قرار گرفته،

1. Adversarial Attacks.

تعیین مسئول جبران خسارت (طراح سیستم، به‌کارگیرنده یا مهاجم) از منظر حقوقی پیچیده می‌باشد (Čerka et al., 2015). در این راستا، ماده ۲۲ دستورالعمل اتحادیه اروپا در زمینه مسئولیت محصول و اصلاحیه‌های پیشنهادی آن[1] برای پوشش محصولات مبتنی بر هوش مصنوعی، تلاشی برای پاسخ به این چالش حقوقی است (European Commission, 1985).

چالش حقوقی دیگر، مسئله استفاده از هوش مصنوعی برای انجام حملات سایبری پیشرفته می‌باشد. سیستم‌های هوش مصنوعی می‌توانند برای شناسایی آسیب‌پذیری‌های امنیتی، ایجاد بدافزارهای پیچیده یا حتی اجرای حملات فیشینگ هدفمند استفاده شوند (Brundage et al., 2018). این مسئله نیازمند بازنگری در قوانین موجود مبارزه با جرایم سایبری است. به‌عنوان نمونه، کنوانسیون بوداپست (کنوانسیون جرایم سایبری شورای اروپا)[2] که در سال ۲۰۰۱ تصویب شده، فاقد مقررات خاص برای مقابله با حملات سایبری مبتنی بر هوش مصنوعی می‌باشد (Council of Europe, 2001). همچنین مسئله حفاظت از داده‌های آموزشی هوش مصنوعی نیز چالش حقوقی دیگری است. سیستم‌های هوش مصنوعی برای آموزش به حجم وسیعی از داده‌ها نیاز دارند و حمله به این داده‌ها می‌تواند عملکرد سیستم را مختل نماید (Papernot et al., 2016). این مسئله با مباحث مالکیت داده و حفاظت از دارایی‌های فکری مرتبط است. ماده ۴ دستورالعمل اتحادیه اروپا در زمینه حفاظت از اسرار تجاری[3] به‌طور غیرمستقیم به این موضوع می‌پردازد، اما هنوز چهارچوب حقوقی جامعی برای حفاظت از داده‌های آموزشی هوش مصنوعی وجود ندارد (European Parliament & Council, 2016).

چالش حقوقی دیگر در زمینه امنیت سایبری هوش مصنوعی، مسئله مسئولیت‌پذیری در برابر نقض امنیتی سیستم‌های خودمختار می‌باشد. در سیستم‌های سنتی، مسئولیت نقض امنیتی معمولاً متوجه به‌کارگیرنده یا طراح سیستم است، اما در مورد سیستم‌های هوش مصنوعی خودمختار، تعیین مسئول از منظر حقوقی پیچیده‌تر است (Scherer, 2015). در این راستا، قطعنامه پارلمان اروپا در زمینه قواعد حقوق مدنی در رباتیک[4] پیشنهاد ایجاد وضعیت حقوقی خاص برای ربات‌ها و سیستم‌های هوش مصنوعی را مطرح نموده است (European Parliament, 2017). همچنین، استفاده از سیستم‌های هوش مصنوعی برای دفاع سایبری نیز چالش‌های حقوقی خاصی را ایجاد می‌نماید. این سیستم‌ها می‌توانند به صورت خودکار به حملات سایبری پاسخ دهند، اما این پاسخ‌ها ممکن است پیامدهای حقوقی غیرمنتظره‌ای داشته باشند. به‌عنوان مثال، پاسخ متقابل به یک حمله سایبری ممکن است منجر

1. European Union Product Liability Directive (85 / 374 / EEC).
2. Budapest Convention (Council of Europe Convention on Cybercrime).
3. European Union Trade Secrets Directive (EU / 2016 / 943).
4. European Parliament Resolution on Civil Law Rules on Robotics (2015 / 2103(INL)).

به نقض قوانین کشور ثالث شود. در این زمینه، راهنمای تالین ۲/۰ در مورد قانون بین‌المللی قابل اعمال در عملیات سایبری (۲۰۱۷)[1] چهارچوبی برای ارزیابی قانونی بودن عملیات سایبری ارائه می‌دهد، اما به‌طور خاص به استفاده از هوش مصنوعی در این عملیات نمی‌پردازد (Schmitt, 2017). در نهایت، مسئله حفظ حریم خصوصی در سیستم‌های امنیتی مبتنی بر هوش مصنوعی نیز چالش حقوقی مهمی است. این سیستم‌ها برای تشخیص تهدیدات، ممکن است حجم وسیعی از داده‌های شخصی را پردازش کنند که با اصول حفاظت از داده‌ها در تعارض قرار می‌گیرد. ماده ۶ (۱) مقررات عمومی حفاظت از داده اتحادیه اروپا پردازش داده‌ها را بر اساس «منافع مشروع» مجاز می‌داند (European Parliament & Council, 2016)؛ اما تعیین مرز دقیق بین امنیت سایبری به‌عنوان یک منفعت مشروع و نقض حریم خصوصی، همچنان چالش حقوقی مهمی است (Edwards & Veale, 2018).

ب) راهکارهای رفع و تقلیل چالش‌های حقوقی هوش مصنوعی

راهکارهای مؤثر برای مواجهه با چالش‌های حقوقی هوش مصنوعی را می‌توان در سه محور اصلی بررسی نمود. این موارد چهارچوب‌های تنظیم‌گری، تدابیر فنی و امنیتی و راهبردهای اخلاقیِ مکمل است. این محورها که به صورت هم‌افزا عمل کرده و یکدیگر را تقویت می‌نمایند در ادامه مورد بررسی قرار خواهند گرفت.

۱. توسعه چهارچوب‌های تنظیم‌گری هوش مصنوعی

چهارچوب‌های تنظیم‌گری نقش بنیادینی در مدیریت چالش‌های حقوقی هوش مصنوعی ایفا می‌نمایند. این چهارچوب‌ها با ایجاد مرزهای مشخص برای توسعه و استفاده از سیستم‌های هوش مصنوعی، حقوق اشخاص را تضمین می‌کنند. این چهارچوب‌ها می‌توانند در دو سطح مورد توجه قرار گیرند.

یک. تدوین قوانین و مقررات خاصِ هوش مصنوعی

نظام‌های حقوقی پیشرو در سراسر جهان، در حال اجرا، تدوین و تصویب قوانین اختصاصی برای تنظیم فناوری هوش مصنوعی هستند. به‌عنوان مثال قانون هوش مصنوعی اتحادیه اروپا[2] که در آوریل ۲۰۲۴ به تصویب رسید، اولین چهارچوب جامع قانونی در جهان برای تنظیم سیستم‌های هوش مصنوعی محسوب می‌شود. این قانون، رویکردی مبتنی بر خطر را اتخاذ نموده و سیستم‌های هوش مصنوعی را بر اساس سطح خطر طبقه‌بندی می‌کند. در حوزه سیاست‌گذاری، این قانون مصادیق مشخصی را برای

1. Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations (2017).
2. Artificial Intelligence Act (AI Act).

تنظیم هوش مصنوعی ارائه می‌دهد. به‌عنوان نمونه، سیستم‌های هوش مصنوعی پرخطر (مانند سیستم‌های استخدام، ارزیابی اعتباری یا تشخیص چهره) ملزم به رعایت الزامات سخت‌گیرانه‌ای همچون ارزیابی انطباق قبل از ورود به بازار، شفافیت الگوریتمی، نظارت انسانی و مستندسازی دقیق هستند. به‌طور مشخص، ماده ۱۰ این مقررات، توسعه‌دهندگان سیستم‌های هوش مصنوعی پرخطر را ملزم می‌سازد تا از داده‌های آموزشی با کیفیت بالا و عاری از سوگیری استفاده نمایند تا از تبعیض الگوریتمی جلوگیری شود (European Commission, 2024). در ایالات متحده نیز، دستور اجرایی رئیس جمهور در اکتبر ۲۰۲۳ با عنوان «توسعه و استفاده ایمن، امن و قابل اعتماد از هوش مصنوعی» الزامات مشخصی را برای توسعه‌دهندگان سیستم‌های هوش مصنوعی مقرر نموده است. به‌عنوان مثال، این دستور اجرایی شرکت‌های توسعه‌دهنده مدل‌های بزرگ هوش مصنوعی را ملزم می‌سازد تا نتایج آزمون‌های امنیتی خود را به دولت فدرال گزارش دهند (White House, 2023).

دو. انطباق بسترهای حقوقی موجود با هوش مصنوعی

علاوه بر تدوین قوانین جدید، انطباق و به‌روزرسانی بسترهای حقوقی موجود ـ منظور از بسترهای حقوقی مجموعه قوانین، مقررات، رویه‌های قضایی و دکترین‌های حقوقی است که چهارچوب حقوقی برای تنظیم فناوری هوش مصنوعی را تشکیل می‌دهند ـ نیز از اهمیت بسزایی برخوردار است. قوانین حفاظت از داده، مانند مقررات عمومی حفاظت از داده اتحادیه اروپا، چهارچوب‌های مهمی برای حفاظت از حقوق افراد در عصر هوش مصنوعی ارائه می‌دهند. این مقررات، اصول مشخصی را برای پردازش داده‌های شخصی مقرر نموده است که در زمینه هوش مصنوعی نیز کاربرد دارند. به‌عنوان نمونه، ماده ۲۲ این مقررات، حق افراد را در عدم قرارگیری تحت تصمیم‌گیری خودکار با آثار حقوقی یا پیامدهای مشابه به رسمیت شناخته است. این اصل در مورد سیستم‌های هوش مصنوعی تصمیم‌گیر مانند سیستم‌های ارزیابی اعتباری، به معنای الزام به نظارت انسانی و امکان درخواست بازبینی تصمیمات است.

در این مسیر همچنین می‌توان از رویه قضایی موجود نیز بهره برد. به‌عنوان نمونه پرونده CNIL علیه Clearview AI در سال ۲۰۲۱ که مقام حفاظت از داده فرانسه شرکت Clearview را به دلیل جمع‌آوری و پردازش تصاویر چهره افراد بدون رضایت آنها برای آموزش سیستم تشخیص چهره مبتنی بر هوش مصنوعی، به پرداخت جریمه ۲۰ میلیون یورویی محکوم نمود. توضیح بیشتر اینکه پرونده CNIL علیه Clearview AI در سال ۲۰۲۱ یکی از مهم‌ترین پرونده‌های حقوقی در زمینه حفاظت از داده‌های شخصی و کاربرد هوش مصنوعی در سال‌های اخیر است که اصول اساسی حفاظت از داده‌ها را در عصر هوش مصنوعی تبیین نموده است. در دسامبر ۲۰۲۱، کمیسیون ملی انفورماتیک و آزادی‌های

فرانسه (CNIL)،[1] به‌عنوان نهاد ناظر حفاظت از داده‌های این کشور، شرکت آمریکایی Clearview AI را به پرداخت جریمه‌ای به مبلغ ۲۰ میلیون یورو محکوم کرد. این تصمیم پس از تحقیقات گسترده‌ای اتخاذ شد که نشان می‌داد Clearview AI بدون رضایت افراد، میلیون‌ها تصویر چهره را از اینترنت و شبکه‌های اجتماعی جمع‌آوری کرده و از آنها برای آموزش سیستم تشخیص چهره مبتنی بر هوش مصنوعی خود استفاده نموده است. CNIL تشخیص داد که Clearview AI داده‌های زیست‌سنجی افراد را بدون مبنای حقوقی پردازش ـ مقرر در ماده ۶ مقررات عمومی حفاظت از داده اتحادیه اروپا ـ پردازش کرده است و هیچ‌گونه رضایت صریح از افراد برای جمع‌آوری و پردازش داده‌های زیست‌سنجی آنها کسب نکرده بود، در حالی که طبق ماده ۹ مقررات عمومی حفاظت از داده، پردازش داده‌های زیست‌سنجی مستلزم رضایت صریح است. همچنین این شرکت به حقوق اشخاص موضوع داده از جمله حق دسترسی به داده و حق حذف داده احترام نگذاشته و اطلاعات کافی در مورد نحوه جمع‌آوری، پردازش و استفاده از داده‌های شخصی به افراد ارائه نکرده بود (CNIL, 2021).

این پرونده بر اهمیت اصل نظارت انسانی بر سیستم‌های تصمیم‌گیر هوش مصنوعی تأکید می‌کند. طبق ماده ۲۲ مقررات مذکور، افراد حق دارند که موضوع تصمیم‌گیری‌های صرفاً خودکار که اثرات حقوقی یا مشابه قابل توجهی بر آنها دارد، قرار نگیرند. این اصل مستلزم نظارت انسانی بر تصمیمات مهم اتخاذشده توسط سیستم‌های هوش مصنوعی است. پرونده Clearview AI نشان می‌دهد که سیستم‌های هوش مصنوعی باید قادر به ارائه توضیح در مورد تصمیمات خود باشند و امکان بازبینی انسانی این تصمیمات وجود داشته باشد؛ موضوعی که به‌ویژه در سیستم‌های ارزیابی اعتباری، استخدام یا تصمیم‌گیری‌های قضایی اهمیت دارد. این پرونده همچنین نشان می‌دهد که علی‌رغم اهمیت نوآوری در حوزه هوش مصنوعی، این نوآوری نمی‌تواند به بهای نقض حقوق بنیادین افراد، به‌ویژه حق حریم خصوصی و حفاظت از داده‌های شخصی، صورت گیرد. علاوه بر این، پرونده Clearview AI نمونه بارزی از اِعمال فراسرزمینی مقررات عمومی حفاظت از داده است. این امر یعنی علی‌رغم اینکه Clearview AI یک شرکت آمریکایی است، به دلیل پردازش داده‌های شهروندان اروپایی، مشمول این مقررات شناخته شده است. افزون بر این، پرونده حاضر تأثیر قابل توجهی بر نحوه توسعه و استفاده از سیستم‌های هوش مصنوعی داشته است. توسعه‌دهندگان سیستم‌های هوش مصنوعی باید از همان مراحل اولیه طراحی، ملاحظات حفاظت از داده‌ها را در نظر بگیرند و جمع‌آوری داده از منابع عمومی مانند اینترنت، بدون رضایت افراد، برای آموزش سیستم‌های هوش مصنوعی می‌تواند غیرقانونی باشد. همچنین

1. The Commission Nationale de l'informatique et des libertés (CNIL).

سیستم‌های هوش مصنوعی باید به‌گونه‌ای طراحی شوند که قابل توضیح باشند و تصمیمات آنها قابل بررسی و بازبینی باشد. در نهایت، پرونده CNIL علیه Clearview AI نقطه عطفی در تنظیم مقررات فناوری‌های هوش مصنوعی محسوب می‌شود که نشان می‌دهد استفاده از فناوری‌های هوش مصنوعی باید در چهارچوب احترام به حقوق بنیادین افراد صورت گیرد و اصل نظارت انسانی و امکان بازبینی تصمیمات سیستم‌های هوش مصنوعی، به‌عنوان یک اصل کلیدی در این زمینه، مورد تأکید قرار گرفته است.

سه. تأثیر چهارچوب‌های حقوقی بر اخلاق‌مداری هوش مصنوعی

چهارچوب‌های حقوقی نقش مهمی در اخلاق‌مدار نمودن فضای هوش مصنوعی ایفا می‌نمایند. این تأثیر از طریق سازوکارهای مختلفی محقق می‌شود که در ادامه خواهد آمد. یکی از این موارد الزام به شفافیت الگوریتمی است. در واقع چهارچوب‌های حقوقی با الزام توسعه‌دهندگان به شفاف‌سازیِ نحوه عملکرد سیستم‌های هوش مصنوعی، امکان نظارت عمومی و پاسخ‌گویی را افزایش می‌دهند. به‌عنوان مثال، ماده ۱۳ مقررات هوش مصنوعی اتحادیه اروپا، توسعه‌دهندگان سیستم‌های پرخطر را ملزم به ارائه مستندات فنی دقیق می‌نماید. این شفافیت منجر به اعتمادسازی و تقویت رفتار اخلاقی می‌شود (European Commission, 2024). جلوگیری از تبعیض الگوریتمی مورد بعدی است. این امر یعنی چهارچوب‌های حقوقی با منع استفاده از داده‌های دارای سوگیری و الزام به آزمون سیستم‌ها برای شناسایی تبعیض، اصل اخلاقی عدالت و انصاف را تقویت می‌کنند. و در نهایت نیز مسئولیت‌پذیری قابل بیان است.

در این مورد، چهارچوب‌های حقوقی با تعیین مسئولیت حقوقی برای خسارات ناشی از سیستم‌های هوش مصنوعی، توسعه‌دهندگان را به رعایت استانداردهای اخلاقی ترغیب می‌کنند. در این راستا، اصلاحیه پیشنهادی دستورالعمل مسئولیت محصول اتحادیه اروپا، محصولات مبتنی بر هوش مصنوعی را تحت پوشش نظام مسئولیت محصول قرار می‌دهد (European Commission, 2022).

۲. بهره‌مندی از تدابیر فنی و امنیتی نسبت به هوش مصنوعی

تدابیر فنی و امنیتی، مجموعه اقدامات و راهکارهای عملی هستند که در مراحل طراحی، توسعه و استفاده از سیستم‌های هوش مصنوعی به کار گرفته می‌شوند تا امنیت، قابل اعتماد بودن و سازگاری این سیستم‌ها با اصول حقوقی و اخلاقی تضمین گردد. این تدابیر در مقایسه با چهارچوب‌های حقوقی که جنبه الزام‌آور و تنبیهی دارند، بیشتر ماهیت پیشگیرانه و فناورانه داشته و در سطح طراحی و معماری سیستم‌ها اعمال می‌شوند. مهم‌ترین این تدابیر در ادامه مورد بررسی قرار خواهند گرفت.

یک. حریم خصوصی از طریق طراحی و پیش‌فرض

یکی از مهم‌ترین تدابیر فنی، پیاده‌سازی اصول «حریم خصوصی از طریق طراحی»[1] و «حریم خصوصی به صورت پیش‌فرض»[2] در سیستم‌های هوش مصنوعی است. این رویکرد که در ماده ۲۵ مقررات عمومی حفاظت از داده نیز مورد تأکید قرار گرفته است، مستلزم رعایت الزامات حفاظت از داده از همان مراحل اولیه طراحی سیستم است. به‌عنوان مثال، فناوری یادگیری فدرال[3] به‌عنوان یک راهکار فنی، امکان آموزش مدل‌های هوش مصنوعی را بدون نیاز به جمع‌آوری متمرکز داده‌های شخصی فراهم می‌آورد. در این روش، مدل به دستگاه‌های کاربران فرستاده می‌شود و آموزش روی داده‌های محلی انجام می‌شود، سپس تنها پارامترهای مدل (و نه داده‌های اصلی) به سرور مرکزی ارسال می‌گردد. گوگل از این فناوری در قابلیت پیشنهاد کلمات کیبورد استفاده می‌کند تا حریم خصوصی کاربران حفظ شود (Bonawitz et al., 2019).

دو. شفافیت و توضیح‌پذیری الگوریتمی

شفافیت و توضیح‌پذیری الگوریتمی از دیگر تدابیر فنی مهم برای مواجهه با چالش‌های حقوقی هوش مصنوعی است. فناوری‌های هوش مصنوعی قابل توضیح[4] امکان درک و تفسیر تصمیمات سیستم‌های هوش مصنوعی را فراهم می‌آورند. به‌عنوان نمونه، شرکت IBM با توسعه کتابخانه AIX360، ابزارهایی را برای توضیح‌پذیری مدل‌های یادگیری ماشین ارائه داده است. این ابزارها به کاربران امکان می‌دهند تا دلایل تصمیم‌گیری مدل‌ها را درک کنند که برای رعایت الزامات قانونی مانند «حق دریافت توضیح» در ماده ۲۲ مقررات عمومی حفاظت از داده، ضروری است (Arya et al., 2019).

سه. تحقق امنیت سایبری پیشرفته

تأمین امنیت سیستم‌های هوش مصنوعی در برابر حملات خصمانه، یکی دیگر از تدابیر فنی مهم است. فناوری‌های یادگیری متخاصم[5] به توسعه‌دهندگان امکان می‌دهد تا مدل‌های مقاوم در برابر حملات خصمانه ایجاد کنند. به‌عنوان مثال، محققان با توسعه روش‌های دفاعی خاص، رویکردی را برای افزایش مقاومت شبکه‌های عصبی عمیق در برابر حملات نمونه‌های متخاصم ارائه داده‌اند. این روش‌ها به‌طور قابل توجهی موفق به کاهش اثربخشی حملات انتقال و افزایش استحکام مدل‌های هوش مصنوعی شده‌اند (Papernot et al., 2016).

---

1. Data protection by design.
2. Data protection by default.
3. Federated Learning.
4. Explainable AI / XAI.
5. Adversarial Learning.

۳. راهبردهای اخلاقیِ مکمل برای هوش مصنوعی انسان‌محور

راهبردهای اخلاقی، مکمل مهمی برای چهارچوب‌های قانونی و تدابیر فنی محسوب می‌شوند. این راهبردها با تمرکز بر ارزش‌های انسانی و اصول اخلاقی، زمینه را برای توسعه و استفاده مسئولانه از هوش مصنوعی فراهم می‌آورند. اهم این موارد به شرح ذیل است:

یک. طراحی هوش مصنوعی اخلاق‌مدار

طراحی هوش مصنوعی اخلاق‌مدار به معنای لحاظ کردن ملاحظات اخلاقی در تمام مراحل چرخه حیات سیستم‌های هوش مصنوعی است. در این رویکرد، مسئولیت اخلاقی متوجه طراحان، توسعه‌دهندگان و کاربران سیستم‌های هوش مصنوعی ـ نه خود فناوری ـ است. به عبارت دیگر، هوش مصنوعی به خودی خود دارای اخلاق نیست، بلکه این انسان‌ها هستند که باید اصول اخلاقی را در طراحی و استفاده از این سیستم‌ها رعایت نمایند. با این حال، پیاده‌سازی این اصول در عمل با چالش‌های متعددی روبه‌رو است. این چالش‌ها عمدتاً از ماهیت پیچیده تصمیم‌گیری‌های اخلاقی و محدودیت‌های ذاتی سیستم‌های هوش مصنوعی در درک مفاهیم انتزاعی نشئت می‌گیرند. پیچیدگی در طراحی هوش مصنوعی اخلاق‌مدار به چند عامل مشخص بازمی‌گردد. نخست، تعارض میان اهداف مختلف مانند دقت و عدالت؛ به‌عنوان مثال، در سیستم‌های تشخیص چهره، افزایش دقت ممکن است منجر به عملکرد نابرابر برای گروه‌های مختلف نژادی شود. دوم، عدم قطعیت در پیش‌بینی پیامدهای بلندمدت؛ مانند سیستم‌های توصیه‌گر که می‌توانند با ایجاد حباب‌های فیلتر به قطبی شدن جامعه کمک کنند. سوم، چالش تفسیر مفاهیم انتزاعی اخلاقی در قالب کد و الگوریتم؛ مثلاً تعریف دقیق «عدالت» یا «منفعت عمومی» برای یک سیستم هوش مصنوعی دشوار است (Dignum et al., 2018).

با این حال برای طراحی هوش مصنوعی اخلاق‌مدار، می‌توان از رویکردهای متعددی استفاده نمود. این موارد از جمله تشکیل گروه‌های میان‌رشته‌ای است. در این گروه همکاری متخصصان فنی با فیلسوفان، حقوق‌دانان، جامعه‌شناسان و روان‌شناسان لازم است. به‌عنوان نمونه، شرکت DeepMind با تأسیس واحد اخلاق و جامعه[1] در سال ۲۰۱۷، متخصصان حوزه‌های مختلف را گرد هم آورده تا پیامدهای اخلاقی فناوری‌های خود را بررسی نمایند. همچنین ارزیابی تأثیرات اخلاقی قابل اشاره است. این امر به معنای انجام ارزیابی‌های جامع از پیامدهای اخلاقی سیستم‌های هوش مصنوعی قبل از استقرار است. به‌عنوان مثال گوگل در سال ۲۰۲۰ چهارچوبی برای ارزیابی اخلاقی محصولات هوش مصنوعی خود معرفی کرد که شامل بررسی عدالت، شفافیت و پاسخ‌گویی است (Raji et al., 2020). در نهایت آموزش اخلاق به

1. Ethics & Society.

متخصصان فنی نیز می‌تواند مؤثر باشد. این امر شامل گنجاندن دروس اخلاق کاربردی در برنامه‌های آموزشی علوم کامپیوتر و هوش مصنوعی است. دانشگاه استنفورد با ارائه دوره «اخلاق، سیاست عمومی و هوش مصنوعی» به دانشجویان مهندسی، این رویکرد را پیاده کرده است (Stanford University, 2023).

دو. ترویج فرهنگ استفاده اخلاقی از هوش مصنوعی

ترویج فرهنگ استفاده اخلاقی از هوش مصنوعی، مستلزم اقدامات چندبعدی در سطوح مختلف فردی، سازمانی و اجتماعی است. منظور از رویکرد چندبعدی، توجه هم‌زمان به جنبه‌های آموزشی، فرهنگی، حقوقی و فنی استفاده از هوش مصنوعی است. به‌عنوان مثال، در بعد آموزشی، آگاهی‌بخشی عمومی درباره مزایا و خطرات هوش مصنوعی؛ در بعد فرهنگی، ترویج ارزش‌های مسئولیت‌پذیری و احترام به حقوق دیگران؛ در بعد حقوقی، آشنایی با قوانین مرتبط؛ و در بعد فنی، آموزش نحوه استفاده ایمن از فناوری‌ها لازم است. همچنین استانداردهای رفتاری در استفاده از هوش مصنوعی، مجموعه‌ای از اصول و قواعد مشخصی است که رفتار مسئولانه کاربران را هدایت می‌کند. به‌عنوان نمونه اصل شفافیت مستلزم آن است که کاربران باید هنگام تعامل با دیگران؛ استفاده از سیستم‌های هوش مصنوعی (مانند چت‌بات‌ها) را آشکار سازند. همچنین به موجب اصل بعدی یعنی اصل صحت‌سنجی، کاربران باید اطلاعات دریافتی از سیستم‌های هوش مصنوعی را قبل از استفاده یا انتشار، از منابع معتبر راستی‌آزمایی کنند. یکی دیگر از این اصول نیز اصل احترام به مالکیت فکری است که به موجب آن کاربران باید حقوق مالکیت فکری دیگران را در استفاده از سیستم‌های هوش مصنوعی رعایت کنند. پرونده‌های حقوقی متعددی مانند شکایت نویسندگان از OpenAI در سال ۲۰۲۳ به دلیل استفاده بدون اجازه از آثارشان برای آموزش ChatGPT، اهمیت این اصل را نشان می‌دهد.

نتیجه

پژوهش حاضر با هدف بررسی نقش اخلاق هوش مصنوعی در پیشگیری از چالش‌های حقوقی در این حوزه انجام شده است. یافته‌های این پژوهش نشان می‌دهد که فناوری هوش مصنوعی، ضمن ارائه فرصت‌های بی‌سابقه برای پیشرفت بشر، چالش‌های حقوقی متعددی را نیز ایجاد نموده که نیازمند رویکردی چندوجهی برای مواجهه با آنها می‌باشد. در بخش نخست این پژوهش، چالش‌های حقوقی عمده در حوزه هوش مصنوعی مورد بررسی قرار گرفت. مسائل مربوط به حفظ حریم خصوصی اطلاعاتی و حفاظت از داده یکی از مهم‌ترین این چالش‌ها محسوب می‌شود. ماهیت سیستم‌های هوش مصنوعی که مستلزم دسترسی به حجم وسیعی از داده‌ها برای آموزش و عملکرد مؤثر می‌باشد، تعارضی با اصول

حقوقی حاکم بر پردازش داده دارد. چالش حقوقی دیگر، مسئله تعارض بین حقوق فردی و منافع جمعی در استفاده از هوش مصنوعی است. این تعارض در نظام‌های حقوقی مختلف مورد تصریح است و چهارچوب‌های حقوقی متعددی در تلاش برای ایجاد توازن میان این دو دسته از حقوق و منافع وضع گردیده است. همچنین تهدیدهای امنیت سایبری در سیستم‌های هوش مصنوعی، چالش حقوقی دیگری است که در این پژوهش مورد بررسی قرار گرفت. حملات خصمانه به سیستم‌های هوش مصنوعی، استفاده از هوش مصنوعی برای انجام حملات سایبری پیشرفته و مسئله حفاظت از داده‌های آموزشی هوش مصنوعی، چالش‌های حقوقی نوینی هستند که قوانین و مقررات موجود برای مواجهه با آنها کافی نمی‌باشند.

با توجه به چالش‌های مذکور در بخش دوم پژوهش، راهکارهای رفع و تقلیل چالش‌های حقوقی هوش مصنوعی در سه محور اصلی مورد بررسی قرار گرفت. این موارد تحقق چهارچوب‌های تنظیم‌گری، بهره‌مندی از تدابیر فنی و امنیتی، و توجه به راهبردهای اخلاقیِ مکمل است. در زمینه چهارچوب‌های تنظیم‌گری، تدوین قوانین اختصاصی هوش مصنوعی مانند قانون هوش مصنوعی اتحادیه اروپا، گامی مهم در مواجهه با چالش‌های حقوقی این فناوری محسوب می‌شود. همچنین، انطباق قوانین و مقررات موجود مانند مقررات عمومی حفاظت از داده با هوش مصنوعی، راهکار دیگری است که در این پژوهش مورد توجه قرار گرفت. از سویی دیگر یافته‌های پژوهش حاکی از آن است که چهارچوب‌های حقوقی نقش مهمی در اخلاق‌مدار نمودن فضای هوش مصنوعی ایفا می‌نمایند. این تأثیر از طریق سازوکارهایی مانند الزام به شفافیت الگوریتمی، جلوگیری از تبعیض الگوریتمی و ایجاد مسئولیت‌پذیری محقق می‌شود. فارغ از حوزه تنظیم‌گری، بهره‌مندی از تدابیر فنی و امنیتی، محور دوم راهکارهای مورد بررسی در این پژوهش بود. این تدابیر شامل پیاده‌سازی اصول «حریم خصوصی از طریق طراحی» و «حریم خصوصی به صورت پیش‌فرض» در سیستم‌های هوش مصنوعی، استفاده از فناوری‌های هوش مصنوعی قابل توضیح برای افزایش شفافیت و توضیح‌پذیری الگوریتمی و به‌کارگیری فناوری‌های امنیت سایبری پیشرفته مانند یادگیری متخاصم برای مقابله با حملات خصمانه می‌باشد. در نهایت راهبردهای اخلاقیِ مکمل، محور سوم راهکارهای مورد بررسی بود. در این راستا طراحی هوش مصنوعی اخلاق‌مدار مستلزم توجه به ملاحظات اخلاقی در تمام مراحل چرخه حیات سیستم‌های هوش مصنوعی است. علاوه بر این ترویج فرهنگ استفاده اخلاقی از هوش مصنوعی نیز مستلزم اقدامات چندبعدی در سطوح مختلف فردی، سازمانی و اجتماعی است. تدوین استانداردهای رفتاری مانند اصل شفافیت، اصل صحت‌سنجی و اصل احترام به مالکیت فکری از جمله اقدامات مؤثر در این زمینه می‌باشند.

با توجه به آنچه بیان شد، این پژوهش مبین آن است که مواجهه مؤثر با چالش‌های حقوقی هوش

مصنوعی، مستلزم رویکردی جامع و چندوجهی است که در آن، چهارچوب‌های حقوقی، تدابیر فنی و راهبردهای اخلاقی به صورت هم‌افزا عمل نمایند. قوانین و مقررات، چهارچوب الزام‌آوری را برای تضمین رعایت حداقل‌های حقوقی فراهم می‌آورند، تدابیر فنی امکان پیاده‌سازی عملی این الزامات را میسر می‌سازند و راهبردهای اخلاقی، با تمرکز بر ارزش‌های انسانی و اصول اخلاقی، زمینه را برای توسعه و استفاده مسئولانه از هوش مصنوعی فراهم می‌آورند. البته با توجه به ماهیت پویا و تحول‌پذیر فناوری هوش مصنوعی، ضروری است که چهارچوب‌های حقوقی و اخلاقی نیز به‌طور مستمر مورد بازنگری قرار گیرند. در نهایت باید توجه داشت که هدف نهایی از تمامی این تلاش‌ها، تضمین این امر است که فناوری هوش مصنوعی در خدمت انسان و ارزش‌های انسانی باقی بماند و به ابزاری برای تقویت کرامت، آزادی و رفاه بشر تبدیل گردد.

منابع و مآخذ

- لطیف‌زاده، مهدیه و سید محمدمهدی قبولی درافشان (۱۴۰۱). چگونگی انتقال بین‌المللی داده‌های شخصی (مطالعه تطبیقی در حقوق اتحادیه اروپا و نظام حقوقی ایران). ***پژوهشنامه حقوق تطبیقی***. ۶ (۱). ۲۳۰ ـ ۲۰۷. شناسه برنمود دیجیتالی 10.22080/lps.2022.23212.1309

- Arya, V., Bellamy, R. K., Chen, P. Y., Dhurandhar, A., Hind, M., Hoffman, S. C.,... & Zhang, Y. (2019). One explanation does not fit all: A toolkit and taxonomy of ai explainability techniques. ***arXiv preprint arXiv:1909.03012***.

- Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V.,... & Roselander, J. (2019). Towards federated learning at scale: System design. ***Proceedings of machine learning and systems***. ***1***. 374 - 388.

- Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B.,... & Amodei, D. (2018). The malicious use of artificial intelligence: Forecasting. prevention. and mitigation. ***arXiv preprint arXiv:1802.07228***.

- California Legislature. (2018). California Consumer Privacy Act of 2018 [CCPA]. Cal. Civ. Code § 1798.100 - 1798.199.

- Cath, C. (2018). Governing artificial intelligence: ethical, legal and technical opportunities and challenges. ***Philosophical Transactions of the Royal Society A: Mathematical. Physical and Engineering Sciences***. 376. https://DOI.org/10.1098/rsta.2018.0080.

- Čerka, P., Grigienė, J., & Sirbikytė, G. (2015). Liability for damages caused by artificial intelligence. Computer law & security review. 31(3). 376 - 389.

- CNIL (2021). Facial recognition: the CNIL orders CLEARVIEW AI to stop reusing photographs available on the Internet. Commission Nationale de l'Informatique et des Libertés.

- Council of Europe. (1950). European Convention for the Protection of Human Rights and Fundamental Freedoms. as amended by Protocols Nos. 11 and 14. ETS 5.

- Council of Europe. (2001). ***Convention on Cybercrime***. European Treaty Series. No. 185. Budapest.

- Court of Justice of the European Union. (2020). La Quadrature du Net and Others v Premier ministre and Others (Joined Cases C-511/18, C-512/18 and C-520/18). ECLI: EU:C:2020:791.
- Dignum, V., Baldoni, M., Baroglio, C., Caon, M., Chatila, R., Dennis, L.,... & De Wildt, T. (2018, December). Ethics by design: Necessity or curse?. In ***Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*** (pp. 60 - 66).
- Edwards, L., & Veale, M. (2018). Enslaving the algorithm: From a "right to an explanation" to a "right to better decisions"?. ***IEEE Security & Privacy***. ***16***(3). 46 - 54.
- Eija, S. (2018). Applying General Data Protection Regulation In Small Organizations Simplified Framework and Templates for Managing a Privacy. ***School of Business and Culture.*** Available at: https://www.theseus.fi/bitstream/handle/10024/158605/Eija_Syrjanen.pdf
- European Commission (2024). ***Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act***)
- European Commission. (1985). ***Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products***. Official Journal L 210. 07/08/1985.
- European Commission. (2022). ***Proposal for a Directive of the European Parliament and of the Council on liability for defective products. COM(2022) 495 final***. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52022PC0495
- European Parliament & Council. (2016). ***Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure***. Official Journal L 157. 15/06/2016.
- European Parliament & Council. (2016). ***Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)***. Official Journal L 119. 04/05/2016.
- European Parliament, Council & Commission. (2000). Charter of Fundamental Rights of the European Union. Official Journal of the European Union. C 364/1.
- European Parliament. (2017). ***European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))***.
- Floridi, L., Cath, C., & Taddeo, M. (2019). Digital ethics: its nature and scope. ***The 2018 yearbook of the digital ethics lab***. 9 - 17. DOI:10.1007/978-3-030-17152-0_2
- Jarvis, L. (2019). The Age of Big Data Analytics: A cross-national comparison of the implementation of Article 23 of the GDPR in the United Kingdom, France, Germany and Italy (pp. 1 - 50). Available at: https://discovery.ucl.ac.uk/id/eprint/10083534/
- John, A., U., A., & Panachakel, J. (2023). Ethical Challenges of Using Artificial Intelligence in Judiciary. ***2023 IEEE International Conference on Metrology for extended Reality. Artificial Intelligence and Neural Engineering (MetroXRAINE)***. 723 - 728. https://DOI.org/10.1109/MetroXRAINE58569.2023.10405688.

- Kubben, P., Dumontier, M., & Dekker, A. (Eds.). (2019). Fundamentals of Clinical Data Science. ***Springer International Publishing.*** DOI: https://doi.org/10.1007/978-3-319-99713-1
- Madiega, T., Car, P., Niestadt, M., & van de Pol, L. (2022). Metaverse, Opportunities, risks and policy implications (pp. 1 - 12).
  Available at: https://www.europarl.europa.eu/thinktank/en/document/EPRS_IDA(2022)733557
- Marghalani, A. (2019). ***Digital ethics and privacy: A study about digital ethics issues, implications, and how to solve them***. DOI:10.13140/RG.2.2.23675.16169
- Ourt of Justice of the European Union. (2014). Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González (Case C-131/12). ECLI: EU:C:2014:317.
- Papernot, N., McDaniel, P., Sinha, A., & Wellman, M. (2016). Towards the science of security and privacy in machine learning. ***arXiv preprint arXiv:1611.03814***.
- Papernot, N., McDaniel, P., Wu, X., Jha, S., & Swami, A. (2016, May). Distillation as a defense to adversarial perturbations against deep neural networks. In ***2016 IEEE symposium on security and privacy (SP)*** (pp. 582 - 597). IEEE.
- Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B.,... & Barnes, P. (2020, January). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In ***Proceedings of the 2020 conference on fairness, accountability, and transparency*** (pp. 33 - 44).
- Roossien, C. C., de Jong, M., Bonvanie, A. M., & Maeckelberghe, E. L. M. (2021). Ethics in design and implementation of Technologies for Workplace Health Promotion: a call for discussion. ***Frontiers in Digital Health, 3***, 644539. DOI: https://doi.org/10.3389/fdgth.2021.644539
- Scherer, M. U. (2015). Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. ***Harv. JL & Tech***. 29. 353.
- Schmitt, M. N. (Ed.). (2017). ***Tallinn Manual 2.0 on the international law applicable to cyber operations***. Cambridge University Press.
- Stanford University (2023). CS 182: Ethics, Public Policy, and Technological Change. Stanford Computer Science.
- Taddeo, M., McCutcheon, T., & Floridi, L. (2019). Trusting artificial intelligence in cybersecurity is a double-edged sword. ***Nature Machine Intelligence***. 1(12), 557 - 560.
- Turillazzi, A., Taddeo, M., Floridi, L., & Casolari, F. (2023). The digital services act: an analysis of its ethical, legal, and social implications. ***Law, Innovation and Technology***. 15(1), 83 - 106. https://DOI.org/10.1080/17579961.2023.2184136
- U.S. Congress. (1964). Title VII of the Civil Rights Act of 1964. 42 U.S.C. § 2000e et seq.
- U.S. Congress. (1966). Freedom of Information Act. 5 U.S.C. § 552.
- U.S. Congress. (1974). Privacy Act. 5 U.S.C. § 552a.
- UN General Assembly. (1948). Universal Declaration of Human Rights.
- UN General Assembly. (1966). International Covenant on Civil and Political Rights. United Nations, Treaty Series. vol. 999. p. 171.
- Wagner, J., & Benecke, A. (2017). National Legislation within the Framework of the GDPR. ***European Data Protection Law Review. 2***(3). 353 - 361. DOI: https://doi.org/10.21552/edpl/2017/3/8

- Wang, Y., Su, Z., Zhang, N., Liu, D., Xing, R., Luan, T. H., & Shen, X. (2022). A Survey on Metaverse: Fundamentals, Security, and Privacy. In ***Cornell University*** (pp. 1 - 31). DOI: https://DOI.org/10.48550/arXiv.2203.13202
- White House (2023). ***Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence***. The White House.