Journal Website

Legal Innovation Quarterly

Spring 2026

Volume 2, Issue 3

E-ISSN: 3115-8498

Criminalization of Emerging Threats Arising from Artificial Intelligence: An Analysis of the Challenges and Legislative Solutions

Saeed Ghomashi[1], Mohammad Reza Kameli*[2], Mojtaba Nikokalam Nadim[3]

1. Associate Professor, Department of Criminal Law and Criminology, Faculty of Law and Humanities, Kashan University, Kashan, Iran.
2. M.A. in Criminal Law and Criminology, Faculty of Law and Humanities, Kashan University, Kashan, Iran (Corresponding Author)*
3. M.A. in Criminal Law and Criminology, Faculty of Law and Humanities, Kashan University, Kashan, Iran.

***Corresponding author email address:** mrkameli1524@gmail.com

Extended Abstract

1. Introduction

The expansion of artificial intelligence has given rise to emerging threats and challenged fundamental concepts of criminal law, including mens rea (guilty mind), criminal intent, and the attribution of liability, because intelligent systems, in addition to facilitating the commission of offenses by humans, sometimes become sources of harmful conduct by virtue of their autonomous operation.

Deepfakes and intelligent phishing in the field of threats against persons threaten reputation and privacy, respectively, and facilitate access to information through behavioral data analysis. Errors in self-driving vehicles may lead to murder or bodily harm, and autonomous weapons increase the risk of violations of international humanitarian law. Artificial intelligence-based cyberattacks threaten critical infrastructure and national security; ransomware and the illicit mining of cryptocurrencies through botnets also damage property and trust in the digital economy.

Received: October 15, 2025 - **Accepted:** March 11, 2026 - **Published:** March 21, 2026

Ghomashi, Saeid; Kameli, Mohammadreza; & Nikookalam Nadim, Mojtaba. (2026). "Criminalization of Emerging Threats Arising from Artificial Intelligence: An Analysis of Challenges and Legislative Solutions." *Legal Innovation Quarterly*, 2(3), pp 1–53.
Doi: https://doi.org/10.22034/lir.2026.2074805.1026

© 2026 The Authors. Published by Legal Innovation Research Institute, Tehran, Iran. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License.

Domestic studies, unlike some international research, have focused more on prevention; these include studies on the preventive capacity of artificial intelligence and Martha Stroppa's (2023) study on maintaining human control over autonomous cyber capabilities. The principal question is the adequacy of Iranian criminal law in addressing these threats. The hypothesis is that the laws are fragmented and inadequate, and the objective is to analyze the gaps and propose solutions in four areas: threats against persons, autonomous technologies, cybersecurity, and property.

2. Research Methodology

The research methodology is based on a theoretical-analytical approach and relies on documentary research and content analysis of domestic laws, international instruments, scholarly articles, specialized reports, and comparative studies. The findings are used to identify gaps and propose legislative solutions.

3. Research Findings

The findings of this study indicate that the existing criminal law provisions in Iran are fragmented and inadequate to address emerging threats arising from artificial intelligence. In the field of threats against persons, existing provisions, such as Articles 6 and 16 of the Computer Crimes Law, may be applicable to some instances of deepfakes and phishing, but these provisions do not provide for a distinct offense or a specific framework for such conduct.

In the field of autonomous systems and autonomous weapons, the Iranian criminal law system lacks a clear mechanism for determining criminal liability in cases where decision-making is independently performed by an algorithm. Artificial intelligence-based cyberattacks against critical infrastructure, despite their increasing speed and complexity, lack a distinct offense of “cyberattack” or a definition of “artificial intelligence-enhanced cybercrime.” In the field of property, automated ransomware and the illicit mining of cryptocurrencies through botnets call into question the adequacy of traditional classifications of financial crimes. Overall, the findings confirm the research hypothesis and demonstrate the need to reconsider the legal elements, criminal intent, and rules governing the attribution of liability.

4. Recommendations

To address legislative gaps, the article proposes draft statutory provisions for the criminalization of emerging technological conduct and the determination of criminal liability in autonomous systems:

1. Criminalization of Deepfakes in the Iranian Criminal Law System

Altering digital content to create a false and deceptive representation, with the intent to harm the personality, reputation, or pecuniary or non-pecuniary rights

of natural or legal persons, shall constitute an offense. Where serious harm is caused, the maximum punishment shall apply; use for educational, research, artistic, or journalistic purposes, in the absence of any intent to deceive or cause harm, shall be excluded.

2. Criminalization of Phishing Attacks in Iranian Criminal Law

The use of fraudulent webpages, misleading emails, or text messages with the intent to commit fraud or unlawfully obtain information shall constitute an offense. The proposed punishments include taʿzir imprisonment of the sixth degree, confiscation of proceeds derived from the offense, and deprivation of the right to use computer and Internet services for a period of two to five years. Where serious harm occurs, the most severe punishment prescribed shall apply, and an unsuccessful attempt accompanied by criminal intent shall be deemed an attempt to commit an offense.

3. Criminalization of the Use of Artificial Intelligence-Based Autonomous Weapons

The use of autonomous weapons that violate international humanitarian law or the rules governing armed conflicts shall be prohibited and shall constitute an offense; participation in the production, design, distribution, sale, import, export, possession, or use of such weapons shall be subject to taʿzir imprisonment of the fifth degree, and in cases of murder or bodily harm, the corresponding punishments shall also apply. Where preventive measures are established and the offense occurs as a result of an unforeseeable defect, criminal liability shall be attributed to the system under the proposed provision. Government oversight, approval of high-risk modification, and the establishment of a compensation fund or insurance for compensation of losses is also proposed.

4. Criminalization of Conduct Arising from Artificial Intelligence in Self-Driving Vehicles

Where an error in the system is not attributable to any prior programming or planning by engineers or users and results in murder or bodily harm, the draft provides for the attribution of liability to the system and compensation for the damage caused; where it is not possible to directly enforce a punishment, liability shall, on a combined basis, rest with the owner, operator, or entity responsible for the system. Where there is a defect in design or manufacture, the manufacturing company or the owner shall be liable, and a special insurance fund is proposed.

5. Criminalization of Artificial Intelligence-Based Cyberattacks

Intelligent cyberattacks involving manipulation, destruction, unauthorized access to, or disruption of systems, critical infrastructure, or personal data shall

be punishable by imprisonment and a pecuniary penalty of the fifth degree; serious harm to persons, national security, or the economy shall result in the maximum punishment. In an automated offense, the designer, programmer, owner, or operator shall be responsible, unless it is established that the commission of the offense resulted from factors beyond their control and foreseeability; lawful use is excluded.

6. Criminalization of Artificial Intelligence-Based Ransomware

The design, development, dissemination, distribution, or control of ransomware with the assistance of artificial intelligence, or the use of ransomware to threaten and extort natural or legal persons, shall constitute an offense. Serious damage to critical systems, national infrastructure, or public security shall result in the most severe punishment.

7. Criminalization of Cryptocurrency Mining Using Botnets and Artificial Intelligence

Cryptocurrency mining using artificial intelligence or botnets, without legal authorization from the competent authorities, shall be criminalized as unauthorized cryptocurrency mining.

At the legislative level, conduct that can be subsumed under existing offenses may be addressed through the amendment or addition of statutory provisions, including Articles 6 and 16 of the Computer Crimes Law concerning deepfakes; however, with regard to autonomous weapons and self-driving vehicles, an independent and comprehensive law is necessary because of the uncertainty surrounding the attribution of liability. The drafts should be completed through expert review and in accordance with international standards.

5. Conclusion

The existing criminal law frameworks in Iran lack the necessary coherence and adequacy to address emerging crimes based on artificial intelligence, particularly with regard to automated conduct and the attribution of liability. Traditional frameworks based on human reason, will, and intent have demonstrated their inadequacy in addressing autonomous technologies, and gaps are evident in defining offenses, identifying offenders, and establishing the mens rea (guilty mind).

The main innovation of this study lies in presenting a roadmap for four key areas, namely offenses against persons, security, property, and autonomous technologies: the independent criminalization of deepfakes and intelligent phishing, the definition of artificial intelligence-enhanced cybercrime, the establishment of specific regulations for ransomware and the illicit mining of cryptocurrencies, and the design of a combined liability regime for human actors and intelligent systems. The possibility of recognizing a special form of

legal personality for artificial intelligence, based on financial autonomy, capacity to litigate, and layered liability, is also raised.
Effective counteraction requires clear legal definitions, criteria for establishing mens rea (guilty mind) in automated decisions, accuracy and safety standards, specialized evidentiary procedures, oversight, and international cooperation. An interdisciplinary approach should safeguard individual and collective rights while preserving the responsible development of artificial intelligence.

Keywords: Artificial intelligence ; criminalization ; deepfakes ; intelligent phishing ; legislative gaps in Iranian criminal law.

Funding
This article has no financial support or funding.

Authors' Contributions
Saeid Ghomashi: Conceptualization, data curation, validation, project administration, and resources.
Mohammadreza Kameli & Mojtaba Nikookalam Nadim: Methodology, formal analysis, writing – review & editing, investigation, writing – original draft.

Conflict of Interest
According to the authors' declaration, there is no conflict of interest regarding this article.

Ethical Considerations
This study was conducted based on the analysis of legal documents, statutes, judicial decisions, and library-based sources. It did not involve human participants, animals, or personal data; therefore, ethical approval from an institutional ethics committee was not required. Nevertheless, the authors adhered to the principles of academic integrity, research honesty, proper citation practices, and respect for intellectual property rights throughout all stages of the research.Generative artificial intelligence (AI) tools were used solely to assist with language editing, improving the structure of the manuscript, and enhancing the readability of the text. These tools did not contribute to the study design, scientific analysis, interpretation of findings, drawing of conclusions, or any research-related decision-making. The author(s) assume full responsibility for the accuracy, originality, and final content of the manuscript.

Authors' Approval
All authors have read and approved the final version of the manuscript and consent to its publication.

بهار ۱۴۰۵
دوره ۲، شماره ۳
شاپای الکترونیکی:
۳۱۱۵-۸۴۹۸

فصلنامه نوآوری حقوقی

وبسایت مجله

بزه‌انگاری تهدیدات نوظهور ناشی از هوش مصنوعی: تحلیل چالش‌ها و راهکارهای تقنینی

سعید قماشی[۱]، محمدرضا کاملی[*۲]، مجتبی نیکوکلام ندیم[۳]

۱. دانشیار گروه حقوق کیفری و جرم‌شناسی، دانشکدهٔ حقوق و علوم انسانی، دانشگاه کاشان، کاشان، ایران.
۲. کارشناس ارشد حقوق کیفری و جرم‌شناسی، دانشکدهٔ حقوق و علوم انسانی، دانشگاه کاشان، کاشان، ایران (نویسندهٔ مسئول)
۳. کارشناسی ارشد حقوق کیفری و جرم‌شناسی، دانشکدهٔ حقوق و علوم انسانی، دانشگاه کاشان، کاشان، ایران

[*] **ایمیل نویسندهٔ مسئول:** mrkameli1524@gmail.com

چکیده

گسترش سریع فناوری‌های مبتنی بر هوش مصنوعی، چالش‌های تازه‌ای را در حوزه‌های اجتماعی، امنیتی و اقتصادی پدید آورده است. این تحولات، پرسش‌هایی را دربارهٔ میزان کارآمدی چارچوب‌های موجود حقوق کیفری ایران در مواجهه با این پدیده‌ها مطرح می‌کند. مسئلهٔ اصلی این پژوهش، بررسی چالش‌های بزه‌انگاری تهدیدات ناشی از هوش مصنوعی و تحلیل خلأهای تقنینی موجود در نظام حقوقی ایران است. این مطالعه با رویکردی نظری ـ تحلیلی و بر پایهٔ منابع معتبر داخلی و بین‌المللی، به واکاوی تهدیدات علیه اشخاص (مانند جعل عمیق و فیشینگ)، تهدیدات ناشی از فناوری‌های خودران و خودمختار (مانند خودروهای خودران و سلاح‌های خودمختار مرگبار)، تهدیدات علیه امنیت (حملات سایبری به زیرساخت‌های حیاتی) و تهدیدات علیه اموال (باج‌افزار و استخراج غیرقانونی ارزهای دیجیتال) می‌پردازد. یافته‌ها نشان می‌دهد قوانین موجود در زمینهٔ جرم‌انگاری رفتارهای فناورانهٔ ناشی از هوش مصنوعی، پراکنده و ناکافی بوده و توان پاسخ‌گویی به پیچیدگی‌های تهدیدات نوظهور را ندارند. نوآوری پژوهش در تحلیل و ارائهٔ راهکارهای تقنینی برای مواجهه با تهدیدات مذکور است. نتیجه‌گیری مقاله، بازنگری و توسعهٔ مقررات کیفری ایران با تأکید بر بزه‌انگاری‌های نو است تا ضمن تضمین عدالت کیفری و صیانت از امنیت ملی، امکان مدیریت تهدیدات نوظهور ناشی از هوش مصنوعی فراهم شود.

واژگان کلیدی: هوش مصنوعی، بزه‌انگاری، جعل عمیق، فیشینگ هوشمند، خلأهای تقنینی در حقوق کیفری ایران.

دریافت‌شده در تاریخ: ۲۳ مهر ۱۴۰۴ - **پذیرش‌شده در تاریخ:** ۲۰ اسفند ۱۴۰۴- **منتشرشده در تاریخ:** ۱ فروردین ۱۴۰۵

قماشی، سعید؛ کاملی، محمدرضا؛ نیکولام ندیم، مجتبی. (۱۴۰۵). «بزه‌انگاری تهدیدات نوظهور ناشی از هوش مصنوعی: تحلیل چالش‌ها و راهکارهای تقنینی». فصلنامه نوآوری حقوقی، دوره ۲، شماره ۳، صص۱-۵۳.

Doi: https://doi.org/10.22034/lir.2026.2074805.1026

© 2026 The Authors. Published by Legal Innovation Research Institute, Tehran, Iran. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License.

مقدمه

تحولات فناورانه در دهه‌های اخیر، به‌ویژه در حوزهٔ هوش مصنوعی، موجب شکل‌گیری گونه‌های نوینی از تهدیدات شده است که نظام‌های سنتی حقوق کیفری را با چالش‌های بنیادین مواجه ساخته‌اند. هوش مصنوعی به‌عنوان ابزاری پیچیده و خودکار، نه‌تنها در ارتکاب جرائم توسط انسان‌ها نقش‌آفرینی می‌کند، بلکه در مواردی به‌واسطهٔ عملکرد مستقل خود، منشأ رفتارهای مجرمانه نیز تلقی می‌شود. این وضعیت، مفاهیم کلیدی حقوق کیفری نظیر «عنصر معنوی»، «قصد مجرمانه» و «انتساب مسئولیت» را در معرض چالش قرار داده و ضرورت بازنگری و بازتعریف مبانی بزه‌انگاری را ایجاب می‌کند.

در حوزهٔ تهدید تمامیت اشخاص، فناوری‌هایی همچون جعل عمیق (دیپ‌فیک) و حملات فیشینگ هوشمند با بهره‌گیری از الگوریتم‌های یادگیری ماشین، مصادیق بارز تهدیدات علیه حیثیت، حریم خصوصی و اعتماد عمومی محسوب می‌شوند. دیپ‌فیک می‌تواند با تولید تصاویر و ویدئوهای جعلی واقع‌گرایانه، شخصیت معنوی افراد را مخدوش کند و زمینه‌ساز افترا، هتک حیثیت و حتی اخاذی شود. فیشینگ هوشمند نیز با استفاده از پردازش زبان طبیعی و داده‌کاوی رفتاری، پیام‌های فریبنده‌ای تولید می‌کند که قربانیان را به افشای اطلاعات شخصی یا مالی سوق می‌دهد و بدین ترتیب اعتماد عمومی به محیط‌های دیجیتال را تضعیف می‌کند.

در حوزهٔ فناوری‌های خودران و خودمختار، خودروهای خودران و سلاح‌های خودمختار مرگبار نمونه‌هایی از تهدیدات نوین هستند که می‌توانند خطاهای جنایی یا خسارات گسترده‌ای ایجاد کنند. خودروهای خودران در صورت نقص فنی یا تصمیم‌گیری الگوریتمی نادرست، ممکن است به حوادث مرگبار منجر شوند که مسئولیت کیفری آن‌ها هنوز با ابهام مواجه است. سلاح‌های خودمختار نیز با قابلیت تصمیم‌گیری مستقل در میدان نبرد، خطر نقض قواعد حقوق بشردوستانه و ارتکاب جرائم جنگی را به همراه دارند و جرم‌انگاری آن‌ها نیازمند بازنگری جدی در مقررات بین‌المللی و داخلی است.

در حوزهٔ امنیت، حملات سایبری به زیرساخت‌های حیاتی کشورها ازجمله شبکه‌های انرژی، حمل‌ونقل، ارتباطات و خدمات مالی، تهدیدی جدی برای امنیت ملی و نظم عمومی محسوب می‌شوند. این حملات می‌توانند موجب فلج شدن سیستم‌های حیاتی، ایجاد بحران‌های اجتماعی و اقتصادی، و تضعیف اعتماد عمومی به نهادهای دولتی شوند. در چنین شرایطی، جرم‌انگاری دقیق

و صریح این رفتارها و تدوین مقررات بازدارنده برای مقابله با حملات سایبری مبتنی بر هوش مصنوعی، ضرورتی انکارناپذیر است.

در حوزهٔ اموال نیز جرائم اقتصادی نوین همچون باج‌افزار و استخراج غیرقانونی ارزهای دیجیتال از طریق بات‌نت‌ها، چالش‌های جدی برای نظام کیفری سنتی ایجاد کرده‌اند. باج‌افزار با رمزگذاری داده‌های کاربران و مطالبهٔ وجه در قبال آزادسازی آن‌ها، مصداق بارز اخاذی دیجیتال است. استخراج غیرقانونی ارزهای دیجیتال نیز با سوءاستفاده از منابع پردازشی قربانیان، موجب تحصیل منفعت نامشروع و آسیب به دارایی‌های دیجیتال می‌شود. این جرائم نه‌تنها خسارات مالی گسترده‌ای به افراد و سازمان‌ها وارد می‌کنند، بلکه اعتماد به نظام اقتصادی دیجیتال را نیز تضعیف می‌سازند.

با وجود گسترش پژوهش‌های بین‌المللی در زمینهٔ جرائم سایبری و تهدیدات فناورانه، مطالعات داخلی در ایران محدود بوده و بیشتر بر پیشگیری متمرکز شده‌اند تا بزه‌انگاری. برای نمونه، در مقالهٔ «هوش مصنوعی به‌عنوان پیش‌بینی‌کنندهٔ جرائم سایبری: فرصت‌ها و چالش‌ها» اثر فاطمه فلاح تفتی (۱۴۰۳)، ظرفیت‌های الگوریتم‌های یادگیری ماشین در پیش‌بینی و پیشگیری از جرائم سایبری بررسی شده و در کنار آن به چالش‌های حریم خصوصی و تعصب الگوریتمی اشاره گردیده است. همچنین در پژوهش «هوش مصنوعی در خدمت پیشگیری از جرائم سایبری: رویکردی فقهی ـ حقوقی» اثر فاطمه فلاح تفتی (۱۴۰۳)، تلاش شده است تا با تکیه بر مبانی فقه امامیه و اصول حقوقی، چارچوبی برای استفادهٔ مشروع و اخلاقی از هوش مصنوعی در پیشگیری از جرائم ارائه شود. در سطح بین‌المللی نیز در مقاله‌ای با عنوان پیامدهای قانونی و اخلاقی قابلیت‌های سایبری خودمختار: فراخوانی برای حفظ کنترل انسانی در فضای مجازی که توسط مارتا استروپا در سال ۲۰۲۳ منتشر شده است (Marta Stroppa, 2023)، بر ضرورت حفظ کنترل انسانی در عملیات سایبری تأکید کرده و چالش‌های حقوقی و اخلاقی ناشی از توانمندی‌های خودکار هوش مصنوعی در فضای مجازی را واکاوی نموده است. این پژوهش درپی آن است که با تمرکز بر جرم‌انگاری تهدیدات نوظهور ناشی از هوش مصنوعی، به پرسش اصلی پاسخ دهد که آیا نظام حقوق کیفری ایران ظرفیت لازم برای مواجهه با این تهدیدات را دارد یا نیازمند اصلاح و توسعهٔ مقررات است. فرضیهٔ پژوهش بر آن است که قوانین موجود پراکنده و ناکافی بوده و توان پاسخ‌گویی به پیچیدگی‌های جرائم فناورانه را ندارند.

اهداف اصلی مقاله عبارت‌اند از: تحلیل چالش‌های حقوقی ناشی از تهدیدات نوظهور هوش مصنوعی، بررسی خلأهای قانونی در نظام کیفری ایران، و ارائهٔ راهکارهای تقنینی برای جرم‌انگاری رفتارهای فناورانه. نوآوری پژوهش در ارائهٔ پیش‌نویس‌های قانونی برای جرم‌انگاری دیپ‌فیک و فیشینگ، تقویت مقررات مقابله با حملات سایبری مبتنی بر هوش مصنوعی، پیشنهاد رژیم تقنینی ترکیبی درخصوص فناوری‌های خودران و خودمختار، و همچنین تبیین چارچوب‌های حقوقی برای مقابله با جرائم اقتصادی نوین ازجمله باج‌افزار و استخراج غیرقانونی ارزهای دیجیتال از طریق بات‌نت‌ها نهفته است.

ساختار مقاله در چهار بخش سامان یافته است: نخست، تهدیدات علیه اشخاص ازجمله جعل عمیق و حملات فیشینگ هوشمند بررسی می‌شود؛ سپس تهدیدات ناشی از فناوری‌های خودران و خودمختار مانند خطاهای جنایی خودروهای خودران و به‌کارگیری سلاح‌های خودمختار مرگبار تحلیل می‌شود؛ در ادامه تهدیدات علیه امنیت ازجمله حملات سایبری به زیرساخت‌های حیاتی مورد مطالعه قرار می‌گیرد و درنهایت تهدیدات علیه حقوق مالی همچون باج‌افزار و استخراج غیرقانونی ارزهای دیجیتال از طریق بات‌نت‌ها تبیین می‌شود.

این تقسیم‌بندی به‌گونه‌ای طراحی شده است که هم با نمونه‌های واقعی تهدیدات نوظهور منطبق باشد و هم امکان تحلیل حقوقی و تقنینی آن‌ها را در چارچوبی منسجم فراهم سازد. بدین ترتیب، مقالهٔ حاضر می‌کوشد سهمی در ارتقای ادبیات حقوق کیفری ایران ایفا کند و زمینه‌ای برای اصلاح و توسعهٔ مقررات کیفری در مواجهه با تهدیدات نوظهور ناشی از هوش مصنوعی فراهم آورد.

۱. تهدیدات علیه اشخاص

هوش مصنوعی با توانایی‌های گسترده در پردازش داده‌ها، یادگیری الگوها و تولید محتوا، عرصه‌ای نوین از تهدیدات علیه اشخاص را پدید آورده است. این تهدیدات بیش از هر حوزه‌ای، مستقیماً بر حقوق بنیادین افراد ازجمله حیثیت، حریم خصوصی، آزادی‌های فردی و اعتماد اجتماعی اثرگذارند. درواقع، فناوری‌های مبتنی بر هوش مصنوعی می‌توانند با تولید محتوای جعلی، دستکاری اطلاعات شخصی یا بهره‌برداری از داده‌های رفتاری، زمینهٔ ارتکاب جرائمی را فراهم کنند که در نظام‌های حقوقی سنتی پیش‌بینی نشده‌اند.

از منظر حقوق کیفری، این تهدیدات پرسش‌های جدی دربارهٔ حدود حمایت قانونی از شخصیت و کرامت انسانی مطرح می‌سازند. آیا تولید و انتشار تصاویر و ویدئوهای جعلی واقع‌گرایانه (دیپ‌فیک) صرفاً نقض اخلاقی است یا باید به‌عنوان جرم مستقل مورد شناسایی قرار گیرد؟ آیا حملات فیشینگ هوشمند که با تقلید از سبک گفتاری و رفتاری قربانیان طراحی می‌شوند، مصداقی از کلاهبرداری رایانه‌ای محسوب می‌شوند یا نیازمند جرم‌انگاری خاص هستند؟ این پرسش‌ها نشان می‌دهند که تهدیدات علیه اشخاص، نقطهٔ آغازین و حساس در بحث بزه‌انگاری تهدیدات نوظهور ناشی از هوش مصنوعی به‌شمار می‌روند. بنابراین، در این بخش ابتدا به جعل عمیق (دیپ‌فیک) به‌عنوان تهدیدی علیه حیثیت و اعتبار اشخاص پرداخته می‌شود، سپس حملات فیشینگ هوشمند به‌عنوان تهدیدی علیه اعتماد عمومی و امنیت داده‌های شخصی بررسی خواهد شد.

۱ـ۱. ماهیت و پیامدهای کیفری جعل عمیق (دیپ‌فیک) مبتنی بر هوش مصنوعی

جعل عمیق ترکیبی از دو مفهوم «جعل» و «یادگیری عمیق» است. این فناوری نوین با بهره‌گیری از شبکه‌های عصبی و تحلیل حجم گسترده‌ای از داده‌ها، توانایی بازسازی و تقلید حالات چهره و حرکات افراد را دارد. در این فرایند، تصاویر یا فیلم‌های مربوط به دو شخص در الگوریتمی ویژه قرار داده می‌شود تا چهرهٔ یکی جایگزین دیگری شود. به بیان ساده‌تر، جعل عمیق با استفاده از فناوری نگاشت صورت و هوش مصنوعی، امکان تغییر چهرهٔ فردی در یک فیلم و جایگزینی آن با چهرهٔ فرد دیگر را فراهم می‌کند. دشواری تشخیص این نوع محتوا ناشی از آن است که بر پایهٔ فیلم‌های واقعی ساخته می‌شود؛ بنابراین صدا و تصویر آن طبیعی جلوه کرده و قابلیت انتشار سریع در شبکه‌های اجتماعی را دارد. همین امر سبب می‌شود بسیاری از مخاطبان، ویدئوی تولیدشده را واقعی تصور کنند (احسان‌پور و امی، ۱۴۰۱: ۳).

جعل عمیق یا دیپ‌فیک روشی است که از هوش مصنوعی برای ایجاد ویدیوهای جعلی واقع‌گرایانه با ترکیب تصاویر و ویدیوهای موجود با فیلم منبع بهره می‌گیرد. این فرایند، موسوم به جعل عمیق، با استفاده از شبکه‌های مولد متخاصم[1] در یادگیری ماشینی، تصاویر و ویدیوها را بر

۱. شبکه‌های مولد متخاصم نوعی الگوریتم هوش مصنوعی هستند که از دو بخش تشکیل می‌شوند: یکی «تولیدکننده» که تلاش می‌کند تصاویر یا ویدیوهای جعلی بسازد، و دیگری «تشخیص‌دهنده» که سعی دارد واقعی یا جعلی بودن آن‌ها را تشخیص دهد. این دو بخش با هم رقابت می‌کنند تا درنهایت، محتوای تولیدشده آن‌قدر واقعی به‌نظر برسد که تشخیص جعلی بودنش بسیار سخت شود.

روی یکدیگر هم‌پوشانی می‌کند (Ben Aissa et al., 2023: 32343). نتیجهٔ نهایی ویدیویی است که به‌نظر می‌رسد افراد را درگیر فعالیت‌ها یا سناریوهایی نشان می‌دهد که هرگز رخ نداده‌اند. برای نمونه، فناوری دیپ‌فیک می‌تواند ویدیوهای ساختگی تولید کند و نشان دهد اشخاص در اعمال جنسی مشارکت داشته‌اند، در حالی‌که چنین چیزی واقعیت ندارد؛ یا سخنان و حرکات یک سیاستمدار را تغییر دهد تا به‌نظر برسد مطلبی را بیان کرده است که هرگز جزو نطق او نبوده است. با توجه به این قابلیت‌ها، از ویدیوهای دیپ‌فیک می‌توان برای ساخت هرزه‌نگاری غیرواقعی با حضور افراد مشهور، انتشار اخبار جعلی و انجام کلاهبرداری‌های مخرب استفاده کرد (Chesney & Citron, 2019: 1772).

دستکاری عکس پیشینه‌ای طولانی دارد و به قرن نوزدهم بازمی‌گردد؛ سپس به‌سرعت در تصاویر متحرک به‌کار گرفته شد. در قرن بیستم، پیشرفت‌های فناورانه به‌طور پیوسته روش‌های دستکاری عکس را بهبود بخشید و با ورود ویدیوهای دیجیتال سرعت بیشتری یافت. توسعهٔ فناوری دیپ‌فیک در دههٔ ۱۹۹۰ توسط پژوهشگران دانشگاهی آغاز شد و بعدها در میان افراد غیرحرفه‌ای در جوامع برخط رواج پیدا کرد. در سال‌های اخیر، این روش‌ها در صنعت نیز مقبولیت گسترده‌ای یافته‌اند. با این حال، امروزه این فناوری‌ها به‌طور نادرست برای تولید محتوای صریح، وادارسازی افراد و انتشار اطلاعات نادرست به‌کار گرفته می‌شوند و پیامدهای اجتماعی منفی و فرساینده‌ای به همراه دارند[1] (Khan & Rehman, 2024: 49). افزون بر این، دیپ‌فیک‌ها به‌طور گسترده به‌عنوان منبع مخرب اطلاعات نادرست در دادگاه‌ها مورد استفاده قرار می‌گیرند و می‌توانند تصمیم قضات را تحت تأثیر قرار دهند (Taeb & Chi, 2022: 89).[2]

۲-۱-۱. کارکردهای هوش مصنوعی در تولید و گسترش محتوای دیپ‌فیک

فناوری جعل عمیق (دیپ‌فیک) با بهره‌گیری از الگوریتم‌های هوش مصنوعی ـ به‌ویژه شبکه‌های مولد متخاصم ـ توانایی تولید و دستکاری محتوای صوتی و تصویری را به سطحی بی‌سابقه رسانده است. این فناوری نه‌تنها امکان بازآفرینی چهره‌ها و صداهای واقعی را فراهم می‌کند، بلکه قادر است تصاویر و ویدئوهای کاملاً جدید و باورپذیر از افراد غیرواقعی بسازد. در کنار این قابلیت‌ها

1. “These technologies are now being misapplied for the creation of explicit content, coercing individuals and disseminating false information, resulting in an erosion of and potentially negative societal consequences”.
2. “Deepfakes are being widely used as a malicious source of misinformation in court that seek to sway a court’s decision”.

پژوهشگران از همان ابزارهای هوش مصنوعی برای تشخیص و پیشگیری از جعل‌های عمیق نیز استفاده می‌کنند. بنابراین کارکردهای هوش مصنوعی در حوزهٔ دیپ‌فیک را می‌توان در سه محور اصلی بررسی کرد: سنتز چهره و صدا، مدل‌سازی عمیق و تشخیص و پیشگیری. این تقسیم‌بندی چارچوبی علمی برای تحلیل ابعاد مختلف فناوری جعل عمیق و پیامدهای کیفری آن فراهم می‌سازد.

ـ سنتز چهره و صدا: الگوریتم‌های هوش مصنوعی قادرند چهره‌ها و صداها را تولید یا دستکاری کنند و به این وسیله چنین وانمود سازند که فردی خاص در حال سخن گفتن یا انجام عملی است (Dehghani & Saberi, 2025: 1).

ـ مدل‌سازی عمیق: این فناوری امکان ایجاد تصاویر یا ویدیوهای جدید و واقعی از افرادی که وجود ندارند یا تغییراتی در تصاویر موجود ایجاد می‌کند.

ـ تشخیص و پیشگیری: جالب است هوش مصنوعی تنها برای تولید جعل عمیق به‌کار نمی‌رود، بلکه در شناسایی و جلوگیری از آن نیز نقش دارد. پژوهشگران در حال توسعهٔ روش‌های مبتنی بر هوش مصنوعی هستند تا ناسازگاری‌ها و نشانه‌های مصنوعی باقی‌مانده در جعل‌های عمیق را شناسایی کنند.

هوش مصنوعی در حوزهٔ فناوری جعل عمیق، به‌عنوان ابزاری بالقوه برای ارتکاب جرائم نوظهور شناخته می‌شود. این فناوری با بهره‌گیری از الگوریتم‌های یادگیری ماشین، به‌ویژه شبکه‌های مولد متخاصم، قادر است محتوای صوتی و تصویری جعلی تولید کند که از نمونه‌های واقعی به‌سختی قابل تشخیص است. به‌دلیل توانایی این ابزار در جعل هویت، انتشار اطلاعات نادرست، و تولید محتوای صریح و اغواگرانه، جعل عمیق به‌عنوان تهدیدی جدی در حوزه‌های امنیتی، اجتماعی و حقوقی مطرح شده است. کاربردهای آن در زمینه‌هایی مانند تروریسم سایبری، باج‌گیری دیجیتال، و فریب نهادهای قضایی، ضرورت تدوین مقررات قانونی و توسعهٔ سامانه‌های تشخیص را دوچندان کرده است.

۳-۱-۱. ابعاد کیفری و جرائم ناشی از بهره‌گیری از دیپ‌فیک مبتنی بر هوش مصنوعی

برای تحلیل ابعاد کیفری جعل عمیق، می‌توان مراحل اصلی ارتکاب این جرم را به دو بخش اساسی تقسیم کرد: ایجاد محتوای جعلی و انتشار یا پخش آن. این تقسیم‌بندی هم در ادبیات علمی

بین‌المللی مطرح شده و هم با منطق حقوق کیفری سازگار است، زیرا هر یک از این مراحل می‌تواند به‌طور مستقل واجد آثار مجرمانه باشد. در مرحلهٔ ایجاد، فناوری هوش مصنوعی برای تولید محتوای جعلی و تقلید صدا یا چهره‌ها به‌کار می‌رود. در مرحلهٔ انتشار یا پخش شدن، همین محتوای جعلی در پلتفرم‌های مختلف منتشر می‌شود و می‌تواند به جرائمی چون باج‌گیری، انتشار اخبار جعلی، افترا و نقض حریم خصوصی منجر شود.

ـ ایجاد: هوش مصنوعی برای تولید محتوای جعل عمیق، تقلید صداها یا چهره‌ها با دقت بالا استفاده می‌شود(Sandoval et al., 2024: 1) .

ـ انتشار یا پخش شدن: هوش مصنوعی می‌تواند به انتشار محتوای جعل عمیق در پلتفرم‌های مختلف کمک کند و به‌طور بالقوه به اطلاعات نادرست یا افترا منجر شود. درواقع جرائم مربوط به دیپ‌فیک می‌تواند شامل باج‌گیری، انتشار اخبار جعلی، افترا و نقض حریم خصوصی باشد[1](Sembiring & Firdaus, 2025: 1).

نظام حقوقی ایران تاکنون جرم مستقلی با عنوان «جعل عمیق» پیش‌بینی نکرده است؛ با این حال، اقداماتی همچون ایجاد یا انتشار دیپ‌فیک می‌تواند ذیل مقررات جرائم سایبری موجود قرار گیرد. برای نمونه، مادهٔ ۶ قانون جرائم رایانه‌ای، جعل رایانه‌ای را جرم‌انگاری کرده و برای آن مجازات حبس از یک تا پنج سال مقرر داشته است. همچنین، مادهٔ ۶۸ قانون تجارت الکترونیک موارد مشابه را در زمرهٔ جعل دانسته و حداکثر سه سال حبس تعزیری برای مرتکبان پیش‌بینی کرده است.

علاوه بر مقررات کلی، در نظام حقوقی ایران می‌توان مواد قانونی متعددی را مرتبط با جرائم ناشی از جعل عمیق شناسایی کرد. برای نمونه، تبصرهٔ ۴ مادهٔ ۱۴ قانون جرائم رایانه‌ای تولید، توزیع و انتشار آثار مستهجن غیرواقعی را جرم‌انگاری کرده است؛ امری که به‌طور مستقیم می‌تواند مصداق استفادهٔ مجرمانه از فناوری دیپ‌فیک باشد. همچنین، مادهٔ ۱۶ همان قانون هتک حیثیت افراد از طریق تغییر یا تحریف داده‌های رایانه‌ای را جرم دانسته و در صورتی که این تغییرات ماهیت مستهجن داشته و منتشر شوند، مرتکب به حداکثر مجازات قانونی محکوم خواهد شد. افزون بر این، مادهٔ ۱۳۱ قانون جرائم نیروهای مسلح به تغییرات غیرمجاز در داده‌ها یا تصاویر توسط نظامیان اشاره دارد و در صورت سوءاستفاده، آن را قابل تعقیب کیفری می‌داند.

1. “Crimes involving deepfakes can include blackmail, spreading fake news, defamation, and privacy violations”.

این مقررات نشان می‌دهد که اگرچه عنوان مستقل «جعل عمیق» در قوانین ایران پیش‌بینی نشده است، اما رفتارهای مرتبط با دیپ‌فیک را می‌توان تحت شمول مواد موجود قرار داد. با این حال، پراکندگی و عدم انسجام این مقررات، ضرورت تدوین قواعد اختصاصی و صریح برای مواجهه با چالش‌های پیچیده و نوظهور ناشی از فناوری جعل عمیق را آشکار می‌سازد.

نقش هوش مصنوعی در ایجاد جعل عمیق یک موضوع پیچیده است. هوش مصنوعی به خودی خود یک ابزار است و مانند هر ابزاری می‌توان از آن برای اهداف مفید و مضر استفاده کرد. در زمینهٔ جعل‌های عمیق، الگوریتم‌های هوش مصنوعی قادر به تولید محتوای صوتی یا تصویری جعلی بسیار واقعی و متقاعدکننده هستند. هوش مصنوعی در این زمینه صرفاً به‌عنوان ابزاری برای ارتکاب جرم عمل می‌کند؛ بنابراین مسئولیت کیفری بر عهدهٔ افرادی است که از این فناوری برای ایجاد و انتشار جعل‌های عمیق با هدف فریب، آسیب‌رسانی یا ارتکاب کلاهبرداری بهره‌برداری می‌کنند. جعل‌های عمیق به‌عنوان تهدیدی جدی شناخته شده‌اند که می‌توانند برای بی‌اعتبار کردن شخصیت‌های عمومی، دستکاری افکار عمومی یا حتی اخاذی مورد استفاده قرار گیرند. آسیب‌های اجتماعی ناشی از این فناوری فراتر از موارد فردی است و می‌تواند اعتماد عمومی به شواهد صوتی و تصویری را تضعیف کند (Romero Moreno, 2024: 297). قوانین در حوزه‌های قضائی مختلف برای مقابله با استفادهٔ مخرب از جعل‌های عمیق تدوین شده‌اند و تمرکز آن‌ها بر زمینه‌هایی چون هرزه‌نگاری، کلاهبرداری و افتراست. هوش مصنوعی به‌عنوان یک فناوری، خود مسئول ارتکاب جرم محسوب نمی‌شود؛ ازاین‌رو همکاری سیاست‌گذاران، مجریان قانون و توسعه‌دهندگان هوش مصنوعی برای کاهش خطرات ناشی از سوءاستفاده و پاسخ‌گو نگه‌داشتن عاملان ضروری است. به‌طور خلاصه، مسئولیت کیفری جعل عمیق متوجه نیت و اقدامات سوء افراد در بهره‌برداری از هوش مصنوعی است. تلاش‌ها برای تنظیم و مدیریت ظرفیت‌های این فناوری در مسیرهای زیان‌بار، ازجمله تولید جعل‌های عمیق، همچنان ادامه دارد.

۲-۱. ماهیت و پیامدهای کیفری حملات فیشینگ هوشمند مبتنی بر یادگیری ماشین

فیشینگ روشی فریبنده برای دسترسی به اطلاعات حساس مانند نام کاربری، گذرواژه، جزئیات حساب بانکی، نشانی اینترنتی (IP) و سایر داده‌هاست. این کار معمولاً از طریق جعل هویت پایگاه‌های اینترنتی معتبر یا نشانی‌های رایانامه انجام می‌گیرد. در اصل، فیشینگ شامل تلاش فردی

برای فریب دادن شخص دیگری به افشای اطلاعات شخصی است. این عمل غیرقانونی باعث ایجاد یک بازار زیرزمینی پرسود به ارزش میلیاردها دلار شده است. در اصل، فیشینگ یک شکل رایج از حملهٔ سایبری است که برای به‌دست آوردن غیرقانونی داده‌های کاربر، ازجمله اعتبار ورود و جزئیات کارت اعتباری استفاده می‌شود (Ramzan, 2010: 439). اصطلاح «فیشینگ» نخستین‌بار در سال ۱۳۷۴ در جعبه‌ابزار نفوذگر «ای او هل»[1] ثبت شد، اما ممکن است قبلاً در مجلهٔ هکر «۲۶۰۰»[2] استفاده شده باشد (Wright et al., 2016: 1115). اقدامات برای جلوگیری یا کاهش تأثیر حملات فیشینگ شامل قوانین، آموزش کاربر، آگاهی عمومی و اقدامات امنیتی فنی است. اهمیت آگاهی از فیشینگ در محیط‌های شخصی و حرفه‌ای افزایش یافته است، به‌طوری که حملات فیشینگ در بین مشاغل از ۷۲ درصد به ۸۶ درصد از سال ۲۰۱۷ تا ۲۰۲۰ افزایش یافته است ("Phishing attacks have increased …", 2020: 37).

در سطح اتحادیهٔ اروپا، پارلمان و شورای اروپا درخصوص حملات علیه سیستم‌های اطلاعاتی، کشورهای عضو را موظف ساخته است که دسترسی غیرمجاز به سامانه‌های اطلاعاتی با قصد تحصیل داده‌ها را به‌عنوان جرم مستقل در قوانین داخلی خود پیش‌بینی کنند. این دستورالعمل با تأکید بر ضرورت جرم‌انگاری رفتارهایی نظیر فیشینگ، بر این نکته تصریح دارد که هرگونه ورود به سیستم‌های اطلاعاتی از طریق نقض تدابیر امنیتی و با هدف دستیابی به اطلاعات، باید مشمول ضمانت اجراهای کیفری شود[3] (European Union, 2013, Art. 3).

۱-۲-۱. نقش هوش مصنوعی در ارتکاب و توسعهٔ حملات فیشینگ

هوش مصنوعی نقش مهمی در حملات فیشینگ، به‌ویژه برای مهاجمان دارد. در اینجا نحوهٔ استفادهٔ مهاجمان از هوش مصنوعی در حملات فیشینگ آمده است:

ـ خودکارسازی حملات فیشینگ: هوش مصنوعی می‌تواند ایجاد و توزیع پیام‌های فیشینگ را خودکار کند و به مهاجمان اجازه می‌دهد حملات در مقیاس بزرگ را به‌طور مؤثر انجام دهند.

1. AOHell
2. 2600: The Hacker Quarterly
3. Member States shall take the necessary measures to ensure that, when committed intentionally, the access without right, to the whole or to any part of an information system, is punishable as a criminal offence where committed by infringing a security measure, at least for cases which are not minor.

حملات فیشینگ مبتنی بر هوش مصنوعی از روش‌های پیشرفته‌ای برای افزایش کارایی عملیات فیشینگ استفاده می‌کنند (Khan & Alam, 2025: 32).

ـ ایجاد پیام‌های قانع‌کننده: هوش مصنوعی قادر است ایمیل‌های فیشینگ را به‌گونه‌ای طراحی کند که با بهره‌گیری از جزئیات اختصاصی هر مخاطب، احتمال پاسخ‌گویی افزایش یابد. استفاده از پردازش زبان طبیعی این امکان را فراهم می‌کند تا متن ایمیل‌ها دقیقاً متناسب با علایق و رفتار برخط قربانی شکل گیرد و درنتیجه نرخ موفقیت حملات بیشتر شود[1](Khan & Alam, 2025: 33).

ـ درس گرفتن از موفقیت‌ها و شکست‌ها: سیستم‌های هوش مصنوعی می‌توانند از حملات گذشته درس بگیرند تا تلاش‌های فیشینگ آینده را بهبود بخشند و مشخص کنند کدام استراتژی‌ها مؤثرترند. درواقع فناوری‌های هوش مصنوعی اثربخشی حملات را بهبود می‌بخشند و در چشم‌انداز جرائم سایبری، تغییردهندهٔ بازی هستند (Jabbar & Al-Janabi, 2025: 1.)

ـ فیشینگ نیزه‌ای:[2] هوش مصنوعی می‌تواند به مهاجمان کمک کند تا با جمع‌آوری و استفاده از اطلاعات شخصی برای ایجاد پیام‌های بسیار سفارشی‌شده، فیشینگ نیزه‌ای، شکلی هدفمند از فیشینگ را انجام دهند. به بیان دیگر در حملات هدفمند، هوش مصنوعی با جمع‌آوری اطلاعات شخصی قربانی، پیام‌هایی بسیار دقیق و متناسب تولید می‌کند که تشخیص آن‌ها برای قربانی دشوار است. از تکنیک‌های پردازش زبان طبیعی برای تولید محتوای فیشینگ متناسب با متن و از نظر دستوری صحیح استفاده می‌شود[3] (Jabir et al., 2025:19). برای مثال، نشان داده شده است که ایمیل‌های فیشینگ تولیدشده توسط هوش مصنوعی نرخ کلیک بالاتری نسبت به ایمیل‌های نوشته‌شده توسط انسان دارند، که نشان‌دهندهٔ اثربخشی هوش مصنوعی در ایجاد محتوای فیشینگ قانع‌کننده است.

1. “Models such as GPT and BERT can generate convincing email messages tailored to the interests and online behavior of the target”.
2. spear phishing
3. “when targeting organisations, attackers may gather information about the implemented systems and infras-tructure, which can be used in attacks to customise phishing messages that, for example, require employees to deploy urgent software patches or log in to a company website link after an upgrade or maintenance, which is then used in subsequent attack”.

۲-۲-۱. زیرساخت قانونی و مقررات کیفری ایران در مواجهه با حملات فیشینگ

در نظام حقوق کیفری ایران، حملات فیشینگ به‌طور مستقیم جرم‌انگاری نشده‌اند؛ با این حال، برخی مواد قانونی در حوزهٔ جرائم رایانه‌ای قابلیت انطباق با این نوع رفتارها را دارند. برای نمونه، مادهٔ ۷۴۰ قانون مجازات اسلامی (تعزیرات، اصلاحی ۱۳۹۹) به ربایش غیرمجاز داده‌ها پرداخته و مادهٔ ۷۴۱ نیز تحصیل مال یا منفعت از طریق دسترسی غیرمجاز به سیستم‌های رایانه‌ای یا مخابراتی را جرم‌انگاری کرده است. هرچند این مواد می‌توانند بر حملات فیشینگ قابل اعمال باشند، اما اصطلاح فیشینگ به‌طور صریح در قانون ذکر نشده است. از همین رو، کارشناسان بر ضرورت تقویت و روزآمدسازی مقررات کیفری تأکید دارند تا بتوان با پیچیدگی‌ها و تحول مداوم تهدیدات سایبری مقابله کرد. این امر شامل شناسایی دقیق‌تر انواع حملات فیشینگ، توسعهٔ همکاری‌های بین‌المللی، ارتقای زیرساخت‌های فناوری اطلاعات و آموزش عمومی برای افزایش آگاهی و پیشگیری از این جرائم است. چنین دیدگاهی بیانگر نیاز به رویکردی جامع‌تر در مواجهه با جرائم سایبری است.

۲. تهدیدات ناشی از فناوری‌های خودران و خودمختار

فناوری‌های خودران و خودمختار، به‌ویژه خودروهای بدون راننده و سامانه‌های تسلیحاتی خودکار، یکی از مهم‌ترین جلوه‌های تحول هوش مصنوعی در عرصه‌های مدنی و نظامی به‌شمار می‌روند. این فناوری‌ها با اتکا به الگوریتم‌های یادگیری ماشین و پردازش داده‌های پیچیده، توانایی تصمیم‌گیری مستقل در شرایط واقعی را پیدا کرده‌اند. هرچند این قابلیت‌ها می‌توانند کارایی و ایمنی را در بسیاری از حوزه‌ها افزایش دهند، اما در عین حال تهدیدات نوظهور و پیچیده‌ای را برای نظام حقوق کیفری ایجاد می‌کنند.

در حوزهٔ مدنی، خودروهای خودران در صورت نقص فنی یا تصمیم‌گیری الگوریتمی نادرست، ممکن است به حوادث مرگبار یا خسارات گسترده منجر شوند. پرسش اساسی این است که در چنین مواردی مسئولیت کیفری بر عهدهٔ چه کسی قرار می‌گیرد: تولیدکننده، برنامه‌نویس، مالک خودرو یا خود سامانهٔ هوش مصنوعی؟ در حوزهٔ نظامی نیز سلاح‌های خودمختار مرگبار با قابلیت انتخاب و هدف‌گیری مستقل، خطر نقض قواعد حقوق بشردوستانه و ارتکاب جرائم جنگی را به همراه دارند. این وضعیت، ضرورت بازنگری در مقررات بین‌المللی و داخلی را آشکار می‌سازد.

۱-۲. ماهیت و چالش‌های کیفری سلاح‌های خودمختار مرگبار مبتنی بر هوش مصنوعی

سلاح‌های خودمختار مرگبار (سامانه‌های تسلیحاتی خودکار مرگبار) نوعی سیستم نظامی مستقل هستند که می‌توانند به‌طور مستقل اهداف را براساس محدودیت‌ها و توصیفات برنامه‌ریزی‌شده جست‌وجو کنند و درگیر شوند. این سلاح‌ها همچنین با نام‌های سیستم‌های اسلحه‌های خودران مرگبار، سیستم‌های سلاح‌های خودمختار (سامانه‌های تسلیحاتی خودکار)، سلاح‌های روباتیک یا روبات‌های قاتل نیز شناخته می‌شوند. این سلاح‌ها ممکن است در هوا، روی زمین، روی آب، زیر آب یا در فضا عمل کنند. استقلال سیستم‌های کنونی از سال ۱۳۹۷ به این معنا محدود شده بود که یک انسان فرمان نهایی حمله را می‌دهد، اگرچه استثناهایی درمورد سیستم‌های «دفاعی» خاصی وجود دارد. مفهوم خودمختاری در رشته‌های مختلف علمی و فناوری معانی متفاوتی دارد. در حوزه‌هایی مانند رباتیک غیرنظامی یا خودروهای خودران، معیارهای شناسایی خودمختاری نسبتاً روشن و استانداردهای مشخصی وجود دارد. اما در زمینهٔ توسعهٔ تسلیحات نظامی، تعریف و شناسایی یک سلاح به‌عنوان خودمختار همچنان با ابهام و اختلاف نظر همراه است (Crootof, 2015:1837). پژوهش‌های حقوقی نشان داده‌اند که استانداردهای مربوط به مفهوم خودمختاری می‌تواند میان دانشمندان، دولت‌ها و سازمان‌های بین‌المللی بسیار متفاوت باشد و همین امر موجب می‌شود شناسایی سلاح‌های خودمختار در عرصهٔ نظامی به وضوح سایر حوزه‌ها نباشد.

ارتش بریتانیا وسایل نقلیهٔ بدون سرنشین و روبات‌های نظامی جدید را در سال ۲۰۱۹ مستقر کرد. (Army Recognition, 2019). نیروی دریایی ایالات متحده در حال توسعهٔ ناوگان «شبح» کشتی‌های بدون سرنشین است (Army Recognition, 2019).

۱-۱-۲. کاربرد هوش مصنوعی در طراحی و عملکرد سلاح‌های خودمختار

هوش مصنوعی نقش مهمی در توسعه و بهره‌برداری از سیستم‌های تسلیحاتی خودمختار ایفا می‌کند. این سیستم‌ها که می‌توانند شامل هواپیماهای بدون سرنشین، موشک‌ها و وسایل نقلیهٔ زمینی باشند، از هوش مصنوعی برای افزایش قابلیت‌های خود به طرق مختلف استفاده می‌کنند:

ـ تصمیم‌گیری: هوش مصنوعی می‌تواند حجم وسیعی از داده‌ها را برای تصمیم‌گیری در زمان واقعی درمورد هدف‌گیری و تعامل بدون دخالت انسان پردازش کند. طی بیش از سه دههٔ گذشته، بهره‌گیری از فناوری هوش مصنوعی در طراحی و توسعهٔ سامانه‌های تسلیحاتی توسط تولیدکنندگان

این حوزه به‌طور مستمر دنبال شده است. مطالعات حقوقی و فنی متعدد، مدل‌های مبتنی بر یادگیری نظارت‌شده را به‌عنوان زیرساختی مؤثر در ارتقای قابلیت تصمیم‌گیری خودکار این سامانه‌ها شناسایی و تحلیل کرده‌اند (Abaimov & Martellini, 2020: 141).

ـ دقت و درستی: الگوریتم‌های هوش مصنوعی می‌توانند دقت سلاح‌ها را بهبود بخشند، آسیب‌های جانبی را کاهش و احتمال اصابت به اهداف مورد نظر را افزایش دهند. درواقع مدل‌های عصبی پیشرفته می‌توانند محل دقیق اصابت موشک یا گلوله را پیش‌بینی کنند، که باعث افزایش دقت و کاهش آسیب‌های جانبی می‌شود (Altmann & Sauer, 2017: 118).

ـ خودمختاری: با هوش مصنوعی، سامانه‌های تسلیحاتی خودکار می‌توانند وظایف پیچیده‌ای مانند هدایت، پایش و شناسایی هدف را به‌طور مستقل انجام دهند. سیستم‌های هدف‌گیری مبتنی بر هوش مصنوعی به پهپادها اجازه می‌دهند تا اهداف را با حداقل دخالت انسان، حتی در محیط‌های بسیار شلوغ، شناسایی و مورد اصابت قرار دهند (Ibid: 118).

– بهره‌وری: سیستم‌های پشتیبانی تصمیم‌گیری مبتنی بر هوش مصنوعی، عملیات را ساده می‌کنند و زمان پاسخ‌گویی را کاهش می‌دهند و امکان کنترل هم‌زمان چندین دارایی را فراهم می‌کنند (Ibid: 119).

– تطبیق‌پذیری: سلاح‌های مجهز به هوش مصنوعی می‌توانند از درگیری‌های گذشته درس بگیرند و با موقعیت‌های جدید تطبیق داده شوند، و به‌طور بالقوه بر سیستم‌های عملیاتی انسانی غلبه کنند.

۲-۱-۲. چالش‌های مسئولیت کیفری ناشی از تصمیم‌گیری هوش مصنوعی در سلاح‌های خودمختار

در بررسی مسئولیت کیفری هوش مصنوعی در سلاح‌های خودمختار، نخست باید میان وضعیت‌های مختلف تفکیک قائل شد. حالت نخست زمانی است که سلاح خودمختار براساس برنامه‌ریزی و طراحی قبلی کاربر یا مهندس اقدام به شلیک می‌کند. در چنین وضعیتی، مسئولیت کیفری متوجه سازنده، برنامه‌ریز یا مالک سلاح خواهد بود، زیرا تصمیم و نقشهٔ قبلی آنان مبنای عملکرد سامانه بوده و نقش تعیین‌کننده در وقوع نتیجهٔ زیان‌بار داشته است. این مسئولیت درواقع ناشی از طراحی و هدایت اولیه‌ای است که مهندسان و عوامل پشت صحنه بر فرایند عملکرد سلاح

اعمال کرده‌اند (Gaeta, 2023: 1033). حالت دوم زمانی مطرح می‌شود که مهندسان و عوامل انسانی پشت صحنه، برنامه یا طرح مشخصی را از پیش در اختیار سامانه قرار نداده باشند و سلاح خودمختار صرفاً براساس خطا یا تصمیم‌گیری مستقل هوش مصنوعی اقدام به شلیک کند. در چنین وضعیتی، پرسش اساسی آن است که آیا باید مسئولیت کیفری را مستقیماً متوجه خود هوش مصنوعی دانست، یا آن‌که باید به سوی الگوی مسئولیت کیفری ترکیبی حرکت کرد؛ الگویی که ضمن پذیرش نقش سامانهٔ هوشمند در وقوع نتیجه، همچنان جایگاه عوامل انسانی و ساختاری را در تعیین مسئولیت نادیده نمی‌گیرد[1](Acquaviva, 2022: 89).

مطالعات به مسئله‌ای به نام «شکاف در مسئولیت» اشاره دارند که ناشی از ذات منحصربه‌فرد هوش مصنوعی است و بر این باورند که مقررات کنونی درخصوص هوش مصنوعی نارسایی دارد[2](Smith, 2022: 2). بنابراین، توصیه شده است که جامعهٔ بین‌المللی باید برای تطبیق با شرایط جدید، سازندگان را ملزم به ارائهٔ شفافیت حداکثری کند و برخی از انواع هوش مصنوعی را ممنوع سازد. برخی پژوهشگران بر این باورند که برای جلوگیری از بی‌کیفرمانی ناشی از استفاده از سلاح‌های هوش مصنوعی، باید اساسنامهٔ دیوان کیفری بین‌المللی نیز مورد بازنگری و اصلاح قرار گیرد (Bátori, 2025: 2).

مسئلهٔ جرم‌انگاری سلاح‌های خودمختار مبتنی بر هوش مصنوعی در حقوق کیفری ایران موضوع بسیار مهم و پیچیده‌ای است. براساس تحقیقات و مقالات علمی، بحث‌هایی در این زمینه وجود دارد که به بررسی مسئولیت کیفری هوش مصنوعی و نحوهٔ برخورد با جرائم ناشی از استفاده از این فناوری‌ها می‌پردازند. از یک طرف، هوش مصنوعی به‌دلیل فقدان شعور و اراده، نمی‌تواند به‌عنوان مباشر جرم شناخته شود و بنابراین، نمی‌توان برای آن مسئولیت کیفری قائل شد. از طرف دیگر، پیشنهاد شده است که باید قوانین موجود را برای پاسخ‌گویی به چالش‌های ناشی از استفاده از هوش مصنوعی در سلاح‌های خودمختار به‌روزرسانی کرد و به‌نظر می‌رسد که پذیرش فرض قانونی

1. “three main theoretical possibilities are explored: that humans remain the only subject capable of being considered criminally respon-sible for harm to interests protected by ICL; that AWS come to be considered ‘persons’ subject to (a form of) individual criminal responsibility; or, a hybrid new system combining aspects of the two”.
2. “As a consequence of this chaotic and unpredictable autonomy and the corresponding likelihood of just war violations, LAWS will almost certainly be involved in responsibility gaps”.

اشخاص الکترونیکی، به‌عنوان سازوکاری برای هویت‌بخشی به سامانه‌های هوش مصنوعی، می‌تواند نقش مؤثری در تحقق مسئولیت کیفری این سامانه‌ها ایفا کند[1] (Swart, 2023: 589). همان‌گونه که در نظام حقوقی، مسئولیت کیفری اشخاص حقوقی پذیرفته شده است، به‌منظور ایجاد مسئولیت کیفری برای اشخاص الکترونیکی نیز، نخست باید شناسایی و تعریف دقیق این شخصیت حقوقی نوظهور صورت گیرد. با وجود پیامدهای زیان‌بار و مخاطرات امنیتی ناشی از توسعه و به‌کارگیری سلاح‌های مبتنی بر هوش مصنوعی، در قانون مجازات قاچاق اسلحه و مهمات و دارندگان سلاح و مهمات غیرمجاز مصوب ۱۳۹۰، هیچ‌گونه تصریحی نسبت به سلاح‌های خودمختار مبتنی بر هوش مصنوعی صورت نگرفته است. این فقدان قانون‌گذاری را می‌توان به‌عنوان یکی از خلأهای حقوقی اساسی در نظام قانون‌گذاری کیفری تلقی کرد، که ضرورت بازنگری و به‌روزرسانی قوانین موجود را در مواجهه با تحولات فناورانه ایجاب می‌کند. با توجه به اینکه هوش مصنوعی می‌تواند خطراتی را برای جان و زندگی انسان‌ها ایجاد کند، ممکن است جرم‌انگاری برخی استفاده‌های خاص از سلاح‌های خودمختار مبتنی بر هوش مصنوعی به‌منظور حفاظت از حقوق بشر و جلوگیری از سوءاستفاده‌ها مصلحت باشد. هرچند برخی ممکن است استدلال کنند که این نوع سلاح‌ها در عموم و اطلاق قواعد سنتی مربوط به سلاح‌ها می‌گنجند، اما واقعیت آن است که ویژگی تصمیم‌گیری کاملاً مستقل، آن‌ها را از چارچوب‌های سنتی خارج می‌کند و نیازمند قواعد خاص و مستقل می‌سازد. بنابراین، تصمیم‌گیری درمورد جرم‌انگاری یا عدم جرم‌انگاری این سلاح‌ها باید با درنظر گرفتن تمام جوانب اخلاقی، حقوقی، امنیتی و فناوری انجام شود و نیازمند گفت‌وگو و همکاری بین متخصصان حقوقی، نظامی، فناوری و سیاست‌مداران است. همچنین، باید به نظرات جامعه بین‌المللی و توافق‌های بین‌المللی در این زمینه توجه شود.

۲-۲. ماهیت و چالش‌های کیفری خودروهای خودران مبتنی بر هوش مصنوعی

خودروی خودران، که گاه با اصطلاحاتی چون ماشین بدون راننده، روبوتاکسی[2] یا خودرو روباتیک نیز شناخته می‌شود؛ وسیله‌ای است که قادر است با حداقل یا بدون دخالت انسانی عمل کند

1. "This article aims to construct 'electronic liability' for such international crimes. It will argue that it is not sufficient to hold the persons or programmers behind the autonomous agents liable, but that it should be possible to hold the autonomous agents that commit international crimes liable".
2. Robotaxi

(Gherardini & Cabri, 2023: 1). اتومبیل‌های خودران مسئول تمام فعالیت‌های رانندگی مانند درک محیط، نظارت بر سیستم‌های مهم و کنترل وسیلهٔ نقلیه هستند که شامل حرکت از مبدأ تا مقصد است (Pendleton et al., 2017: 1). خودروهای خودران پتانسیل تأثیرگذاری بر صنعت خودرو، هزینه‌های حمل و نقل، سلامت، رفاه، برنامه‌ریزی شهری، ترافیک، بیمه، بازار کار و سایر حوزه‌ها را دارند. مقررات مناسب برای ادغام خودروهای خودران در محیط رانندگی موجود ضروری است. سازمان‌هایی مانند انجمن مهندسان خودرو استانداردهای اصطلاحات را پیشنهاد کرده‌اند. با این حال، بیشتر اصطلاحات هیچ تعریف استانداردی ندارند و توسط فروشندگان و سایرین به‌طور متفاوتی به‌کار می‌روند.

۱-۲-۲. هوش مصنوعی و تحولات فنی در عملکرد خودروهای خودران

کارکردهای هوش مصنوعی در خودروهای خودران را می‌توان در چند محور اصلی دسته‌بندی کرد. این محورها شامل تفسیر اطلاعات، تصمیم‌گیری در زمان واقعی، ناوبری در محیط‌های پیچیده و تست ایمنی هستند. هر یک از این کارکردها نقشی اساسی در ارتقای عملکرد خودروهای خودران ایفا می‌کنند و به‌طور مستقل می‌توانند زمینه‌ساز چالش‌های حقوقی و کیفری شوند.

ـ تفسیر اطلاعات: سیستم‌های هوش مصنوعی مقادیر زیادی داده از حسگرهای خودرو مانند دوربین‌ها، رادار و لیدار[1] را پردازش می‌کنند تا محیط اطراف خودرو را تفسیر کنند (Garikapati & Shetiya, 2024: 19). این مدل‌سازی محیطی پایه‌گذار تصمیم‌گیری‌های بعدی خودرو است و به آن امکان می‌دهد تا اشیا، موانع و کاربران جاده را شناسایی کند.

ـ تصمیم‌گیری در زمان واقعی: براساس داده‌های تفسیرشده، هوش مصنوعی تصمیماتی را در زمان واقعی برای پیمایش ایمن اتخاذ می‌کند، مانند شناسایی موانع، پیش‌بینی اقدامات سایر کاربران جاده، و برنامه‌ریزی مسیر وسیلهٔ نقلیه بر این اساس (Schwarting et al., 2018: 190).

ـ ناوبری در محیط‌های پیچیده: هوش مصنوعی خودروهای خودران را قادر می‌سازد تا در محیط‌های پیچیده و پویا حرکت کنند، با تغییر شرایط جاده، الگوهای ترافیکی و موقعیت‌های غیرمنتظره بدون دخالت انسان سازگار شوند(Ibid: 191).

1. Lidar

– تست ایمنی: هوش مصنوعی برای شبیه‌سازی شرایط دنیای واقعی برای تست ایمنی وسایل نقلیهٔ خودران استفاده می شود و به مهندسان کمک می‌کند تا سیستم‌ها را بهبود بخشند و از قابلیت امنیت آن‌ها قبل از استقرار، اطمینان حاصل کنند (Yeong et al., 2025: 12).

۲-۲-۲. ابعاد کیفری ارتکاب جرائم توسط خودروهای خودران مبتنی بر هوش مصنوعی

هوش مصنوعی در خودروهای خودران به‌خودی‌خود واجد وصف مجرمانه نیست، زیرا صرف وجود یا استفاده از این فناوری جرم محسوب نمی‌شود. با این حال، این فناوری می‌تواند به‌عنوان ابزار یا وسیله‌ای مورد سوءاستفاده قرار گیرد و زمینهٔ ارتکاب جرائم مختلف را فراهم سازد. بنابراین، باید میان ماهیت بی‌طرفانهٔ فناوری و کارکردهای مجرمانه ناشی از سوءاستفاده از آن تفکیک قائل شد. این فناوری قابلیت آن را دارد که در بسترهای گوناگون مورد سواستفاده قرار گیرد:

ـ تهدیدات امنیت سایبری: سیستم‌های هوش مصنوعی در خودروهای خودران در برابر هک و حملات سایبری آسیب‌پذیر هستند که می‌تواند به کنترل غیرمجاز خودرو منجر شود.[1] (Brown et al., 2025: 1)

ـ بهره‌برداری: مجرمان به‌طور بالقوه می‌توانند از فرایندهای تصمیم‌گیری هوش مصنوعی برای ایجاد تصادفات عمدی یا استفاده از وسایل نقلیهٔ خودران برای فعالیت‌های غیرقانونی سوءاستفاده کنند[2] (Ibid: 17).

ـ عدم پیش‌بینی: مدل‌های فعلی هوش مصنوعی برای اتومبیل‌های خودران ممکن است به‌طور کامل امکان سوء استفاده از آن‌ها برای اهداف مجرمانه را درنظر نگیرند.

توجه به این نکته مهم است که مسئولیت چنین جرائمی معمولاً بر عهدهٔ افرادی است که هوش مصنوعی را دستکاری یا سوءاستفاده می‌کنند، نه خود هوش مصنوعی. با پیشرفت فناوری هوش

1. “Connected and Autonomous Vehicles (CAVs) integrate advanced communication and autonomous driving tech-
nologies, enabling them to operate independently or with minimal human intervention. Despite the anticipated
benefits for transportation, CAVs could be vulnerable to a wide variety of crimes unless security and crime preven-
tion measures are proactively integrated into the technologies enabling their operation”.
2. “The absence of human drivers in CAVs may facilitate illegal transportation, similar to how drone technology is currently misused to smuggle illicit goods including nar- cotics (Daly, 2024) and human trafficking using vehicles”.

مصنوعی، نیاز فزاینده‌ای به اقدامات امنیتی قوی و دستورالعمل‌های اخلاقی برای جلوگیری و رسیدگی به این خطرات وجود دارد. چارچوب‌های قانونی و نظارتی نیز برای مدیریت پیامدهای هوش مصنوعی در وسایل نقلیهٔ خودران و تضمین ایمنی و مسئولیت‌پذیری در حال تکامل هستند. در مواردی که ماشینی که با هوش مصنوعی کار می‌کند و مرتکب جرم می‌شود یا تصادف می‌کند، ممکن است طرف‌های زیر برای مسئولیت درنظر گرفته شوند:

ـ طراحان و توسعه‌دهندگان: اگر اقدامات هوش مصنوعی ناشی از نقص طراحی یا برنامه‌نویسی باشد، کسانی که هوش مصنوعی را ایجاد کرده‌اند می‌توانند مسئول باشند.

ـ تولیدکنندگان: اگر حادثه به‌دلیل نقص ساخت باشد، ممکن است سازنده مسئول باشد.

ـ کاربران: اگر کاربران از هوش مصنوعی سوءاستفاده کردند یا پروتکل‌های ایمنی را رعایت نکردند، می‌توانند پاسخ‌گو باشند.

– تنظیم‌کننده: اگر نقصی در چارچوب نظارتی وجود داشته باشد که باید از وقوع چنین حوادثی جلوگیری می‌کرد، ممکن است نهادهای نظارتی مورد بررسی دقیق قرار گیرند.

۳. تهدید علیه امنیت

هوش مصنوعی با قابلیت‌های گسترده در تحلیل داده‌ها، یادگیری الگوها و تصمیم‌گیری مستقل، علاوه بر فرصت‌های بی‌سابقه، تهدیدات جدی علیه امنیت در ابعاد مختلف ایجاد کرده است. این تهدیدات می‌توانند امنیت ملی، امنیت سایبری، امنیت اقتصادی و حتی امنیت اجتماعی را تحت تأثیر قرار دهند. درواقع، هرچه فناوری‌های هوشمند بیشتر در زیرساخت‌های حیاتی ادغام شوند، آسیب‌پذیری این زیرساخت‌ها در برابر حملات پیچیده و خودکار افزایش می‌یابد.

از منظر حقوق کیفری، تهدید علیه امنیت ناشی از هوش مصنوعی پرسش‌های بنیادینی را دربارهٔ کفایت قوانین موجود و ضرورت جرم‌انگاری نوین مطرح می‌سازد. آیا حملات سایبری مبتنی بر هوش مصنوعی باید به‌عنوان جرائم مستقل شناخته شوند؟ آیا استفاده از الگوریتم‌های خودکار در تخریب یا اختلال زیرساخت‌های حیاتی نیازمند تدوین مقررات ویژه است؟ این پرسش‌ها نشان می‌دهند که تهدید علیه امنیت، یکی از مهم‌ترین عرصه‌های بزه‌انگاری تهدیدات نوظهور ناشی از هوش مصنوعی محسوب می‌شود.

۱-۳. ماهیت و ابعاد حقوقی حملات سایبری علیه زیرساخت‌ها و سامانه‌های اطلاعاتی

حملهٔ سایبری، هر گونه مانور تهاجمی است که سیستم‌های اطلاعاتی رایانه‌ای، شبکه‌های رایانه‌ای، زیرساخت‌ها، دستگاه‌های رایانهٔ شخصی یا گوشی‌های هوشمند را هدف قرار می‌دهد. مهاجم به شخص یا فرایندی گفته می‌شود که سعی می‌کند بدون مجوز به داده‌ها، عملکردها یا سایر مناطق محدودشدهٔ سیستم دسترسی پیدا کند، که احتمالاً با قصد مخرب است (Agrafiotis et al., 2018: 2). حسب زمینه، حملات سایبری می‌تواند بخشی از جنگ سایبری یا تروریسم سایبری باشد. یک حملهٔ سایبری می‌تواند توسط دولت‌ها، افراد، گروه‌ها، جوامع یا سازمان‌های مستقل به کار گرفته شود و ممکن است از یک منبع ناشناس سرچشمه بگیرد. محصولی که حملهٔ سایبری را تسهیل می‌کند، گاهی اوقات سلاح سایبری نامیده می‌شود. حملات سایبری در چند سال گذشته افزایش یافته است. یک حملهٔ سایبری ممکن است با هک کردن یک شبکهٔ خصوصی یا سیستم حساس، یک هدف مشخص را بدزدد، تغییر دهد یا ازبین ببرد. حملات سایبری می‌تواند از نصب نرم‌افزارهای جاسوسی بر روی رایانهٔ شخصی تا تلاش برای تخریب زیرساخت‌های کل کشورها را شامل شود. کارشناسان حقوقی به دنبال محدود کردن استفاده از این اصطلاح به حوادثی هستند که باعث آسیب فیزیکی می‌شوند و آن را از نقض‌های معمول داده‌ها و فعالیت‌های هک گسترده‌تر متمایز می‌کنند.

۱-۱-۳. زیرساخت‌های حیاتی به‌عنوان اهداف بالقوهٔ حملات سایبری مبتنی بر هوش مصنوعی

با آغاز هر حملهٔ سایبری، مجموعه‌ای از اهداف راهبردی وجود دارد که هدف قرار گرفتن آن‌ها می‌تواند به فلج شدن کامل حریف منجر شود. در این میان، زیرساخت‌های حیاتی در زمان درگیری بیش از سایر بخش‌ها برجسته می‌شوند، زیرا آسیب به آن‌ها توان یک ملت را به‌طور جدی مختل می‌سازد. از جملهٔ این زیرساخت‌ها می‌توان به سامانه‌های کنترل صنعتی، منابع انرژی، شبکه‌های مالی، مخابرات، حمل‌ونقل و تأسیسات آب اشاره کرد که در ادبیات امنیت سایبری به‌عنوان اهداف اصلی شناخته می‌شوند.

براساس گزارشی از مؤسسهٔ فناوری بریتیش کلمبیا، که با بهره‌گیری از داده‌های تاریخی تهیه شده است، تعداد حملات موفق علیه زیرساخت‌ها طی چند دههٔ اخیر رشدی چشمگیر داشته و در مقایسه با سال ۱۹۸۱ حدود ده برابر افزایش یافته است. به‌ویژه از ابتدای دههٔ ۲۰۰۰، حملاتی که به آثار

فیزیکی نامطلوب بر سامانه‌های کنترل و نظارت منجر شده‌اند، در ادبیات تخصصی با عنوان «حملات سایبری ـ فیزیکی» مورد شناسایی قرار گرفته‌اند (Sin & Washburn, 2022: 133).

۲-۱-۳. نقش هوش مصنوعی در طراحی و اجرای حملات سایبری پیشرفته

برای مهاجمان، هوش مصنوعی می‌تواند اثربخشی و کارایی حملات سایبری را به میزان قابل توجهی افزایش دهد. در اینجا نحوهٔ استفاده از هوش مصنوعی در انجام حملات سایبری آمده است:

ـ یادگیری و تکامل: سیستم‌های هوش مصنوعی می‌توانند از محیط بیاموزند و استراتژی‌های خود را برای دور زدن اقدامات امنیتی توسعه دهند. آن‌ها می‌توانند پاسخ‌های ابزارهای امنیت سایبری را تجزیه و تحلیل کنند و الگوهای حملهٔ خود را برای جلوگیری از شناسایی تطبیق دهند (Rigaki & García, 2018: 70).

ـ کشف آسیب‌پذیری خودکار: هوش مصنوعی می‌تواند فرایند اسکن آسیب‌پذیری‌های سیستم را خودکار کند و به مهاجمان این امکان را می‌دهد تا به سرعت نقاط ضعف را شناسایی و از آن‌ها سوءاستفاده کنند (Salem et al., 2024: 1).

ـ فیشینگ پیچیده:[1] با تجزیه و تحلیل داده‌های شخصی، هوش مصنوعی می‌تواند کمپین‌های فیشینگ بسیار شخصی‌سازی‌شده ایجاد کند که احتمال فریب دادن اهداف را افزایش می‌دهد.

ـ بدافزار خود ـ اصلاح‌شونده[2]: هوش مصنوعی می‌تواند بدافزار را قادر سازد تا کد خود را در زمان واقعی تغییر دهد و تشخیص و خنثی کردن این تهدیدات را برای نرم‌افزارهای آنتی‌ویروس سنتی سخت‌تر کند.

ـ سرعت و مقیاس: هوش مصنوعی می‌تواند حملات را با سرعتی بسیار بیشتر و در مقیاسی بزرگ‌تر از مهاجمان انسانی انجام دهد و تأثیر بالقوهٔ تهدیدات سایبری را افزایش دهد.

هوش مصنوعی تهاجمی[3] (Girhepuje et al., 2024: 1)[4]، این اصطلاح به سیستم‌های هوش مصنوعی اطلاق می‌شود که می‌توانند به‌طور مستقل حملات را انجام دهند، عملیات خود را پنهان

1. advanced phishing
2. self-modifying malware
3. Offensive Artificial Intelligence (Offensive AI)
4. "The paper will explore adversarial machine learning, attacks against AI models, infrastructure, and interfaces, along with offensive techniques like information gathering, social engineering, and weaponized AI".

کنند و با فعالیت‌های منظم ترکیب شوند و تشخیص آن‌ها را دشوار کند. استفاده از هوش مصنوعی در حملات سایبری نشان‌دهندهٔ تغییر قابل توجهی در چشم‌انداز تهدید است، زیرا امکان تهدیدات سایبری پیچیده‌تر، سازگارتر و بالقوه مستقل را فراهم می‌کند. این امر بر نیاز به مکانیسم‌های دفاعی پیشرفته مبتنی بر هوش مصنوعی برای محافظت در برابر چنین حملات سایبری تقویت‌شده با هوش مصنوعی تأکید می‌کند. این یک مسابقهٔ تسلیحاتی مداوم بین مهاجمان و مدافعان سایبری است که از فناوری‌های هوش مصنوعی در هر دو طرف استفاده می‌شود.

۳-۱-۳. بررسی مواد قانونی موجود در حقوق کیفری ایران مرتبط با حملات سایبری

در حقوق کیفری ایران، مواد قانونی مرتبط با حملات سایبری عمدتاً در قانون جرائم رایانه‌ای و قانون مجازات اسلامی تعریف شده‌اند. برخی از مواد کلیدی عبارت‌اند از مادهٔ ۷۳۷ قانون مجازات اسلامی به جرائمی مانند تخریب یا اخلال در تأسیسات رایانه‌ای و مخابراتی می‌پردازد. همچنین مادهٔ ۷۳۹ همین قانون نیز در رابطه با جرائم سایبری و تأسیسات زیرساختی است، لکن قانون جرائم رایانه‌ای ایران فاقد اصطلاح و جرم‌انگاری مستقل حملهٔ سایبری است. این در حالی است که حملات سایبری ماهیتی فراملی و سازمان‌یافته دارند و امنیت ملی کشور را هدف قرار می‌دهند. ازاین‌رو، ضرورت دارد در نظام حقوقی ایران، حملات سایبری به‌طور مستقل جرم‌انگاری شوند و ضمانت اجراهای متناسب با شدت و خسارت آن‌ها پیش‌بینی شود (بهره‌مند و مرسی، ۱۴۰۱: ۳۲۰).

۴. تهدیدات علیه اموال(حقوق مالی)

هوش مصنوعی علاوه بر فرصت‌های گسترده در حوزه‌های اقتصادی و صنعتی، تهدیدات نوظهوری را علیه اموال اشخاص و سازمان‌ها ایجاد کرده است. این تهدیدات می‌توانند هم در حوزهٔ اموال مادی مانند تجهیزات و زیرساخت‌های فیزیکی و هم در حوزهٔ اموال مالی و دیجیتال مانند داده‌های شخصی، حساب‌های بانکی، ارزهای دیجیتال و دارایی‌های مجازی بروز یابند.

یکی از مهم‌ترین جلوه‌های این تهدیدات، ظهور باج‌افزارهای هوشمند است که با بهره‌گیری از الگوریتم‌های یادگیری ماشین می‌توانند حملات پیچیده‌تر و هدفمندتری علیه دارایی‌های دیجیتال انجام دهند. این باج‌افزارها قادرند داده‌های حساس را رمزگذاری و قربانی را به پرداخت مبالغ هنگفت برای بازیابی اطلاعات وادار کنند. در کنار آن، استخراج غیرقانونی ارزهای دیجیتال با

استفاده از بات‌نت‌ها و سامانه‌های هوشمند، شکل دیگری از تهدید علیه اموال است که نه‌تنها خسارات اقتصادی مستقیم به همراه دارد، بلکه اعتماد عمومی به نظام‌های مالی و فناوری‌های نوین را نیز تضعیف می‌کند.

از منظر حقوق کیفری، این تهدیدات پرسش‌های بنیادین دربارهٔ کفایت قوانین موجود و ضرورت جرم‌انگاری نوین مطرح می‌سازند. آیا باج‌افزارهای مبتنی بر هوش مصنوعی باید به‌عنوان جرائم مستقل شناخته شوند؟ آیا استخراج غیرقانونی ارزهای دیجیتال با بهره‌گیری از سامانه‌های هوشمند نیازمند مقررات ویژه است؟ این پرسش‌ها نشان می‌دهند که تهدیدات علیه اموال در بستر هوش مصنوعی، یکی از مهم‌ترین عرصه‌های بزه‌انگاری تهدیدات نوظهور محسوب می‌شوند.

۱-۴. ماهیت و پیامدهای کیفری باج‌افزارهای مبتنی بر هوش مصنوعی

باج‌افزار نوعی بدافزار رمزنگاری شده است که دسترسی به داده‌های شخصی قربانی را برای همیشه مسدود می‌کند مگر اینکه باج پرداخت شود. در حالی که برخی از باج افزارهای ساده ممکن است سیستم را بدون آسیب رساندن به هیچ فایلی قفل کنند، بدافزارهای پیشرفته‌تر از تکنیکی به نام اخاذی ویروسی استفاده می‌کنند. فایل‌های قربانی را رمزگذاری می‌کند، آن‌ها را غیرقابل دسترسی می‌کند و برای رمزگشایی آن‌ها باج می‌خواهد (Kharraz et al., 2015: 1). در یک حملهٔ اخاذی ویروسی رمزنگاری‌شده که به‌درستی اجرا می‌شود، بازیابی فایل‌ها بدون کلید رمزگشایی یک مشکل حل‌نشدنی است و از ارزهای دیجیتالی با ردیابی دشوار مانند یا بیت‌کوین[1] و سایر ارزهای رمزنگاری‌شده برای باج‌گیری استفاده می‌شود که ردیابی و تعقیب مجرمان را دشوار می‌کند. حملات باج‌افزار معمولاً با استفاده از یک تروجان[2] مبدل به‌عنوان یک فایل قانونی انجام می‌شوند که کاربر فریب داده می‌شود تا آن را دانلود یا باز کند.

۱-۱-۴. کاربرد هوش مصنوعی در توسعه و اجرای حملات باج‌افزاری

هوش مصنوعی نقشی اساسی در ارتقای توانایی مهاجمان برای اجرای حملات باج‌افزاری ایفا می‌کند. در این بخش، به شیوه‌هایی پرداخته می‌شود که بهره‌گیری از الگوریتم‌های هوشمند می‌تواند

1. Bitcoin
2. Trojan

اثربخشی این حملات را افزایش دهد و زمینه‌های پیچیده‌تر و هدفمندتری را برای ارتکاب چنین جرائمی فراهم سازد.

ـ برنامه‌ریزی و اجرا: مجرمان می‌توانند از هوش مصنوعی برای برنامه‌ریزی و اجرای حملات باج‌افزار پیچیده استفاده کنند. حتی کسانی که دانش فناوری اطلاعات محدودی دارند ممکن است توسط چت‌ربات‌های مبتنی بر هوش مصنوعی قادر به انجام حملات پیچیده باشند (Teichmann, 2023: 399).

ـ فیشینگ هدفمند:[1] مهاجمان با تخصص فناوری اطلاعات می‌توانند از هوش مصنوعی برای تهیهٔ ایمیل‌های فیشینگ قانع‌کننده‌تر استفاده کنند، که اغلب اولین گام در حملهٔ باج‌افزار است. هوش مصنوعی می‌تواند به تنظیم این ایمیل‌ها برای افراد یا سازمان‌های خاص کمک کند و ایمیل‌هایی تولید کند که برای هر قربانی متفاوت و شخصی‌سازی‌شده هستند، و همین موضوع شناسایی آن‌ها را دشوارتر می‌سازد (Ibid: 403).

ـ سرعت و مقیاس: هوش مصنوعی می‌تواند مجرمان سایبری را قادر سازد تا حملات را با سرعت و مقیاس بی‌سابقه‌ای انجام دهند و از اقدامات تشخیص سنتی و مبتنی بر امضا دوری کنند. این امر، دفاع در برابر چنین حملاتی را برای سازمان‌ها چالش‌برانگیزتر می‌کند[2] (Salem et al., 2024: 1).

ـ بدافزار تطبیقی: از هوش مصنوعی می‌توان برای ایجاد بدافزاری استفاده کرد که با محیطی که در آن قرار دارد منطبق می‌شود و تشخیص و حذف آن را دشوارتر می‌کند. این ماهیت تطبیقی بدافزار مبتنی بر هوش مصنوعی می‌تواند به حملات باج‌افزاری مداوم و آسیب‌رسان منجر شود (Teichmann, 2023: 404).

ـ اتوماسیون: هوش مصنوعی می‌تواند بسیاری از جنبه‌های حملات باج‌افزار، از شناسایی آسیب‌پذیری‌ها تا بهره‌برداری از آن‌ها را خودکار کند. این اتوماسیون به مهاجمان اجازه می‌دهد تا با تلاش کمتر به اهداف بیشتری ضربه بزنند و تأثیر کلی آن‌ها را افزایش دهند[3] (Ferdous et al., 2024: 2).

---

1. spear-phishing
2. "Traditional security methods, with their reliance on fixed security controls like firewalls and intrusion detection and prevention systems, have struggled to keep pace with the sophisticated nature of contemporary cyber threats".
3. The REvil ransomware group uses AI to automate the attacks

در حوزهٔ حملات باج‌افزاری، هوش مصنوعی می‌تواند به‌عنوان ابزاری برای برنامه‌ریزی و اجرای این حملات مورد استفاده قرار گیرد. این امر از یک سو امکان ارتکاب حملات پیچیده‌تر را برای افراد فاقد دانش تخصصی در فناوری اطلاعات تسهیل می‌کند و از سوی دیگر به متخصصان این حوزه توانایی تولید ایمیل‌های فیشینگ با سطح بالاتری از اقناع و واقع‌نمایی را می‌دهد. دسترسی گسترده به هوش مصنوعی مولد، بالقوه موجب افزایش کمّی و کیفی حملات باج‌افزاری خواهد شد. با این حال، چارچوب‌های قانونی موجود هنوز با چالش‌های ناشی از این فناوری در حال سازگاری‌اند. باید توجه داشت که هرچند هوش مصنوعی می‌تواند به‌عنوان بردار یا سطح حمله عمل کند، مسئولیت کیفری عمدتاً متوجه عوامل انسانی است که این فناوری را طراحی، هدایت یا بهره‌برداری می‌کنند و از قابلیت‌های آن برای مقاصد غیرقانونی استفاده می‌کنند. این مسئولیت می‌تواند شامل برنامه‌نویسان، گردانندگان سامانه و حتی اشخاصی باشد که در اتخاذ تدابیر امنیتی لازم برای جلوگیری از سوءاستفاده از هوش مصنوعی قصور می‌ورزند (Ibid: 1).

۲-۴. استخراج غیرقانونی ارزهای دیجیتال با بهره‌گیری از بات‌نت‌ها و سامانه‌های هوشمند

اصطلاح «ماینینگ»[1] در زمینهٔ ارزهای رمزپایه به فرایند ایجاد و اعتبارسنجی تراکنش‌های ارزهای دیجیتال با حل مشکلات پیچیدهٔ محاسباتی اشاره دارد (Gandotra et al., 2020: 51). از آنجایی که ارزهای رمزنگاری‌شده مانند بیت‌کوین در یک شبکهٔ غیرمتمرکز کار می‌کنند، افراد در سراسر جهان به عملکرد شبکه کمک می‌کنند و خدمات ضروری را ارائه می‌دهند. در ازای مشارکت آن‌ها، جوایزی در قالب ارزهای دیجیتال به آن‌ها تعلق می‌گیرد. این پاداش‌ها به ایجاد ارزهای دیجیتالی مانند بیت‌کوین، اتریوم[2] و ... منجر می‌شود. این فرایند شبیه به استخراج سنتی است که در آن افراد منابع ارزشمندی مانند طلا یا سنگ‌های قیمتی را استخراج می‌کنند. در دنیای ارزهای دیجیتال، ماینرها[3] خدمات ضروری را به شبکه ارائه می‌کنند و برای تلاش‌های خود با ارزهای دیجیتال پاداش می‌گیرند. این پاداش‌ها اساساً تولید ارزهای دیجیتال از طریق این فرایند استخراج است. این صنعت به‌دلیل شباهت آن به مفهوم استخراج منابع ارزشمند از زمین، اصطلاح «معدن» را پذیرفت.

---

1. mining
2. Ethereum
3. Miners

باتنت گروهی از دستگاه‌های متصل به اینترنت است که هر کدام یک یا چند ربات را اجرا می‌کنند. از باتنت‌ها می‌توان برای انجام حملات انکار سرویس توزیع‌شده، سرقت داده‌ها، ارسال هرزنامه، و اجازهٔ دسترسی مهاجم به دستگاه و اتصال آن استفاده کرد (Haner & Knake, 2021: 1).[1] مالک می‌تواند باتنت را با استفاده از نرم‌افزار فرمان و کنترل، کنترل کند. کلمهٔ «باتنت» ترکیبی از دو کلمهٔ «ربات» و «شبکه» است. این اصطلاح معمولاً با مفهوم منفی یا مخرب استفاده می‌شود. باتنت مجموعه‌ای منطقی از دستگاه‌های متصل به اینترنت، مانند رایانه‌ها، تلفن‌های هوشمند یا دستگاه‌های اینترنت اشیاست که امنیت آن‌ها نقض شده و کنترل آن به شخص ثالث واگذار شده است. هر دستگاه در معرض خطر، که به‌عنوان «ربات» شناخته می‌شود، زمانی ایجاد می‌شود که یک دستگاه توسط نرم‌افزاری از یک توزیع بدافزار (نرم‌افزار مخرب) نفوذ کند. کنترل‌کنندهٔ یک باتنت قادر است فعالیت‌های این رایانه‌های در معرض خطر را از طریق کانال‌های ارتباطی تشکیل‌شده توسط پروتکل‌های شبکهٔ مبتنی بر استاندارد، مانند گفت‌وگوی اینترنتی و پروتکل انتقال ابرمتن هدایت کند (Bailey et al., 2009: 1) . باتنت‌ها به‌طور فزاینده‌ای توسط مجرمان سایبری به‌عنوان کالا برای مقاصد مختلف، ازجمله به‌عنوان خدمات راه‌انداز و استرس‌زا، اجاره داده می‌شوند[2] (Haner & Knake, 2021: 2). به بیان ساده‌تر، یک باتنت را به‌عنوان گروهی از ربات‌ها تصور کنید که تحت کنترل یک رئیس واحد با هم کار می‌کنند. هر ربات درواقع کامپیوتر یا دستگاهی است که بدون اطلاع مالک به نرم‌افزارهای مخرب آلوده شده است[3] (Ibid: 1). این رئیس که یک هکر یا مجرم سایبری است، می‌تواند به همهٔ این دستگاه‌های آلوده دستور دهد تا وظایف را به‌طور هم‌زمان انجام دهند. مانند داشتن یک کنترل از راه دور برای هر دستگاه فکر کنید، اما هکر به جای روشن کردن تلویزیون، از این دستگاه‌ها برای ارسال ایمیل‌های اسپم[4]، حمله به وب‌سایت‌ها یا

1. “Both states and nonstate actors use botnets to surreptitiously control the combined computing power of infected devices to engage in espionage, hacking, and to carry out distributed denial of service attacks to disable internet connected targets from businesses and banks to power grids and electronic voting systems”.
2. “Botmasters have a strong economic incentive to create and spread these networks because they can be used for identity theft, mining cryptocurrency, credit card fraud, ransomware extortion, DDoS extortion, or sold or rented to the highest bidder”.
3. “Malicious networks of botnets continue to grow in strength as millions of new users and devices connect to the internet each day, many becoming unsuspectingly complicit in cyber-attacks or unwitting accomplices to cybercrimes”.
4. spam

سرقت داده‌ها استفاده می‌کند. قسمت ترسناک این است که صاحبان این دستگاه‌ها نمی‌دانند که ماشین‌هایشان به این شکل استفاده می‌شود. مثل این است که یک مأمور خواب در خانهٔ خود داشته باشید که از خواب بیدار می‌شود و دستورات هکر را اجرا می‌کند. در اصل، بات‌نت شبکه‌ای از رایانه‌های ربوده‌شده است که تحت فرمان یک مقام مرکزی، می‌توانند برای انجام کارهایی معمولاً مضر یا غیرقانونی در اینترنت با یکدیگر همکاری کنند. مهم است که از دستگاه‌های خود با اقدامات امنیتی خوب محافظت کنید تا از تبدیل شدن آن‌ها به یک بات‌نت جلوگیری کنید.

۱-۲-۴. نقش هوش مصنوعی در استخراج غیرمجاز ارزهای دیجیتال از طریق بات‌نت‌ها

هوش مصنوعی می‌تواند نقش چندوجهی در استخراج ارزهای دیجیتال با بات‌نت‌ها برای مهاجمان داشته باشد و توانایی آن‌ها را برای انجام فعالیت‌های استخراج غیرقانونی افزایش دهد:

ـ بهینه‌سازی کارایی و هدف‌گذاری استراتژیک: هوش مصنوعی می‌تواند با افزایش کارایی فرایند استخراج، امکان استفادهٔ مؤثرتر از دستگاه‌های کم‌قدرت را فراهم آورد؛ به‌ویژه زمانی که این دستگاه‌ها در قالب یک بات‌نت به‌کار گرفته می‌شوند. همچنین الگوریتم‌های هوش مصنوعی قادرند حجم گسترده‌ای از داده‌ها را تحلیل و آسیب‌پذیرترین سیستم‌ها را شناسایی کنند تا آن‌ها را برای مقاصد استخراج به شبکهٔ بات‌نت اضافه نمایند (Kurtović et al., 2025: 27).

ـ اتوماسیون حملات: هوش مصنوعی می‌تواند فرایند انتشار بدافزار و مدیریت بات‌نت را خودکار کند و نیاز به مداخلهٔ دستی مهاجمان را کاهش دهد[1] (Fritsch et al., 2023: 41).

– تکنیک‌های فرار: هوش مصنوعی می‌تواند به بات‌نت‌ها کمک کند تا با تغییر پویا رفتارشان یا تقلید از فعالیت‌های عادی کاربر، از شناسایی فرار کنند[2] (Ibid: 43).

– سازگاری در زمان واقعی: هوش مصنوعی می‌تواند توانایی سازگاری بات‌نت‌ها را در زمان واقعی افزایش دهد؛ به این معنا که شبکه قادر خواهد بود همگام با تغییرات بازار ارزهای دیجیتال و

1. “AI’s potential for automation of complex tasks serves as a challenge in the face of defensive deployment of anti-malware AI techniques”.
2. “In particular, the evasion of detection of malware and the exfiltration of information through covert channels have been recently used in AI-powered malware”.

نوسانات سطح دشواری استخراج، عملکرد خود را تنظیم کرده و درنتیجه سودآوری پایدار را حفظ کند (Anandakumar et al., 2023: 46).

در حالی که هوش مصنوعی می‌تواند به‌طور قابل توجهی اثربخشی بات‌نت‌ها را برای استخراج ارزهای دیجیتال افزایش دهد، مهم است که تأکید شود استفاده از بات‌نت برای چنین فعالیت‌هایی غیرقانونی است و جرم سایبری محسوب می‌شود.

۵. ضرورت شناسایی و اعطای شخصیت مستقل به هوش مصنوعی در نظام حقوقی

پیش از تبیین دلایل ضرورت جعل شخصیت برای هوش مصنوعی، لازم است به تعاریف گوناگون این مفهوم و پیامدهای حقوقی ناشی از آن توجه شود. در ادبیات عمومی، هوش مصنوعی در برابر «هوش طبیعی» که ویژهٔ انسان و حیوانات است قرار می‌گیرد؛ با این حال، در حوزه‌های تخصصی هنوز تعریف واحد و مورد اجماعی برای آن ارائه نشده است (Banteka, 2021: 537-596). برای نمونه، مؤسسهٔ مهندسان برق و الکترونیک[1] طی سال‌های اخیر چندین بار اصطلاحات و تعاریف خود دربارهٔ هوش مصنوعی را تغییر داده است. همچنین، گروه کارشناسان ارشد هوش مصنوعی کمیسیون اروپا در سال ۲۰۲۰ با اصلاح تعاریف پیشین، سامانه‌های هوش مصنوعی را سیستم‌های نرم‌افزاری (و گاه سخت‌افزاری) معرفی کرد که توسط انسان طراحی شده و با اهداف پیچیده در بُعد فیزیکی یا دیجیتال عمل می‌کنند و از طریق تفسیر داده‌های جمع‌آوری‌شده، استدلال می‌کنند یا اطلاعات را پردازش می‌نمایند (European Commission, 2021, Art. 3) . اتحادیهٔ اروپا نیز در «قانون هوش مصنوعی» مصوب ۲۰۲۱، با تبعیت ضمنی از این تعریف، هوش مصنوعی را نرم‌افزاری دانست که به‌وسیلهٔ تکنیک‌های یادگیری ماشین، منطق و دانش و تحلیل‌های آماری توسعه یافته و قادر است برای اهداف مشخص‌شده توسط انسان خروجی تولید کند و از طریق توصیه‌ها یا تصمیم‌های خود بر محیط تعامل اثر بگذارد (Artificial Intelligence Act, 2021, Art. 3) .

فارغ از این تعاریف فنی، پرسش اصلی در حوزهٔ حقوقی آن است که جایگاه حقوقی هوش مصنوعی چگونه باید تعیین شود. برخی حقوقدانان آن را همانند بردگان روم باستان تلقی کرده و صرفاً ابزاری در خدمت انسان دانسته‌اند (Bryson, 2010; Sharifi & Birmi, 2018: 27-54) . در مقابل، گروهی دیگر بر این باورند که برخی گونه‌های هوش مصنوعی باید در قالب شخص حقوقی

1. IEEE

طبقه‌بندی شوند (Solum, 1992: 4) . افزون بر این، برخی پژوهشگران با تفکیک میان هوش مصنوعی ضعیف و قوی، تنها نوع قوی را واجد شرایط شناسایی شخصیت حقوقی دانسته‌اند (Schirmbacher & Rechtsanwälte, 2020: 71-78). از یک سو، هوش مصنوعی به‌دلیل برخورداری از نوعی ارادهٔ مستقل به اشخاص حقیقی شباهت می‌یابد و از سوی دیگر، به‌سبب فقدان ماهیت انسانی، باید در زمرهٔ اشخاص حقوقی قرار گیرد، زیرا کرامت انسانی بر پایهٔ ویژگی‌های شخص حقیقی استوار است و نمی‌توان آن را به سامانه‌های غیرانسانی تسری داد(Karimi & Karimi, 2014: 17) .

بر این مبنا، مناسب‌ترین قالب قانونی برای هوش مصنوعی، شناسایی شخصیت حقوقی با مختصات ویژه و از نوع «هوشمند» است. تحقق این رویکرد مستلزم بررسی شرایط لازم برای جعل شخصیت، ازجمله وجود مقتضی و فقدان مانع خواهد بود (Kurki, 2019: 180). تقسیم اشخاص به حقیقی و حقوقی نشان می‌دهد که می‌توان با ایجاد قالبی جدید، هوش مصنوعی را در زمرهٔ اشخاص حقوقی خاص ـ یا همان «اشخاص الکترونیکی» ـ قرار داد و بدین ترتیب زمینهٔ لازم برای بار کردن تکالیف کیفری بر آن فراهم ساخت.

۱-۵. مبانی و مقتضیات جعل شخصیت حقوقی برای هوش مصنوعی

بند حاضر در راستای بررسی مقتضیات لازم برای جعل شخصیت حقوقی هوش مصنوعی، ابتدا به مطالعهٔ روند توسعهٔ این فناوری و پیامدهای آن بر قابلیت تصمیم‌گیری مستقل می‌پردازد، زیرا نخستین پرسش در این زمینه، امکان برخورداری هوش مصنوعی از شخصیت است و پرسش بعدی، نوع و ماهیت چنین شخصیتی خواهد بود. بر همین اساس، در بخش دوم به تحلیل شرایط و الزامات لازم برای تحقق این شخصیت پرداخته می‌شود.

۱-۱-۵. خودمختاری عملی سامانه‌های الکترونیک در پرتو توسعهٔ هوش مصنوعی

هوش مصنوعی با گذر زمان و توسعهٔ شاخه‌هایی همچون یادگیری عمیق و یادگیری ماشین، به سطح بالاتری از استقلال عملی دست یافته است؛ به‌گونه‌ای که امروزه بسیاری از گونه‌های آن واجد حدی از خودمختاری هستند و می‌توانند بدون دخالت مستقیم انسان تصمیم‌گیری و اقدام کنند. نمونه‌هایی نظیر خودروهای هوشمند خودران، بات‌های معامله‌گر مستقل که در بازارهای مالی اقدام به خرید و فروش سهام و رمزارز می‌کنند، ربات‌های نظامی که در برخی مأموریت‌ها به‌طور مستقل عمل

می‌نمایند، پهپادهای خودران مورد استفاده در عملیات اطفای حریق یا کشاورزی و نیز بات‌های تولیدکنندهٔ آثار هنری، همگی بیانگر میزان بالای خودمختاری این فناوری در تصمیم‌گیری‌های کلیدی‌اند. برای نمونه، کاربرد بات‌های معامله‌گر در بازارهای مالی نظیر بورس، فارکس و رمزارز به‌طور گسترده مشاهده می‌شود؛ به‌گونه‌ای که کمیسیون بورس و اوراق بهادار ایالات متحده در سال ۲۰۲۰ گزارش داد حدود ۷۸ درصد معاملات توسط بات‌های هوشمند انجام شده است (U.S. Securities & Exchange Commission, 2020: 1). با پیشرفت فناوری، گونه‌های نوینی از بات‌های مبتنی بر هوش مصنوعی قوی پدیدار شده‌اند که با دریافت سرمایه از کاربر، به‌طور مستقل به تحلیل بازار می‌پردازند و صرفاً اهداف کلی همچون کسب سود بیشتر برای آن‌ها تعیین می‌شود؛ در حالی که نقش کاربر عمدتاً به نظارت بر معاملات و دریافت سود احتمالی محدود می‌گردد (Nakhjavani, 2022: 5).

برخی صاحب‌نظران بر این باورند که تا زمان تدوین معیارهای دقیق برای شناسایی شخصیت هوش مصنوعی، مسئولیت آن باید در چارچوب مسئولیت مطلق تولیدکنندگان و مصرف‌کنندگان مورد بررسی قرار گیرد (Rahbar & Dehghanpour Farashah, 2021: 525). بر همین اساس، از آنجا که هوش مصنوعی بر پایهٔ برنامه‌ریزی مشخص عمل می‌کند، مسئولیت باید متوجه طراحان آن باشد (Mele, 2001: 10). با این حال، توسعهٔ روزافزون فناوری موجب افزایش سطح خودمختاری سامانه‌های هوش مصنوعی شده است؛ به‌گونه‌ای که در بسیاری موارد میان ارادهٔ کاربر یا سازنده و فعل ارتکابی توسط سامانه فاصله ایجاد می‌شود. حتی تولید آثار مبتنی بر جعل عمیق، مقالات جعلی و اطلاعات گمراه‌کننده توسط هوش مصنوعی مسبوق به سابقه است (Nelson, 2023: 1). افزون بر این، سوءاستفادهٔ کاربران با سوءنیت از این فناوری در حوزه‌های مجرمانه نیز رایج شده و ضرورت دارد قانون‌گذاری به نحوی صورت گیرد که افراد نتوانند با استناد به حقوق شخصی خود زمینهٔ سوءاستفاده از این ابزار را فراهم سازند (Karimi & Karimi, 2022: 126).

برخی صاحب‌نظران بر این باورند که می‌توان با تدوین استانداردها و قواعدی مشابه قوانین سه‌گانهٔ آسیموف[1]، تبعیت هوش مصنوعی از اصول رفتاری را تضمین کرد. با این حال، چنین رویکردی نیز نمی‌تواند به‌طور کامل ایمنی این فناوری را تضمین کند، زیرا امکان دارد سازندگان از این قواعد پیروی نکنند یا با ارائهٔ تفاسیر متفاوت از آن‌ها، نتایج متعارضی پدید آید. برای نمونه، بنا

1. Asimov’s Law of Robotics

بر گزارش رئیس بخش آزمایش نیروی هوایی ایالات متحده، در جریان آزمایش یک پهپاد نظامی مبتنی بر هوش مصنوعی، که مأموریت اصلی آن انهدام اهداف تعیین‌شده بود و اختیار لغو یا ادامهٔ عملیات درنهایت به اپراتور انسانی واگذار شده بود، پهپاد به‌دلیل برنامه‌ریزی اولیه برای تخریب اهداف، دستور اپراتور انسانی را مانعی در اجرای مأموریت تلقی کرد و به سمت او حمله‌ور شد (Tuhin Das Mahapatra, 2023: 1). از این رو، بحث مسئولیت‌پذیری در قبال عملکردهای متنوع هوش مصنوعی ازجمله پهپادهای نظامی خودران، بات‌های تجاری و خودروهای هوشمند به‌طور جدی مطرح می‌شود (Chopra & White, 2007: 1-17). با این حال، انتساب مسئولیت صرفاً به طراحان و مالکان این فناوری به‌دلیل نقش آنان در تعیین الگوهای رفتاری، همواره موجه به‌نظر نمی‌رسد، زیرا بسیاری از سامانه‌های هوش مصنوعی پیشرفته قادرند الگوهای رفتاری جدید را به‌طور مستقل فراگیرند و از چارچوب‌های اولیه طراحی فراتر روند.

۲-۱-۵. شرایط لازم برای جعل شخصیت حقوقی هوش مصنوعی

بند حاضر به بررسی امکان برخورداری هوش مصنوعی از دارایی مستقل و همچنین قابلیت رسیدگی قضائی به دعاوی مطرح‌شده علیه آن می‌پردازد؛ امری که می‌تواند به‌عنوان یکی از شرایط اساسی در جعل شخصیت حقوقی برای این فناوری تلقی شود.

۱-۲-۱-۵. استقلال مالی و امکان دارایی مثبت و منفی برای شخصیت هوش مصنوعی

دارایی به مجموعه‌ای از اموال، مطالبات و دیون اطلاق می‌شود که می‌تواند به صورت مثبت (اموال و مطالبات) یا منفی (دیون) جلوه‌گر شود (Jafari Langroudi, 2022: 90) . با پیشرفت‌های فناورانه، امکان فرض دارایی مستقل برای هوش مصنوعی مطرح شده است، برای مثال بات‌های معامله‌گر خودمختار قادرند با دریافت سرمایهٔ اولیه از کاربر، به انعقاد معاملات بپردازند و سودآوری کنند. در صورت پذیرش دارایی و اموال برای هوش مصنوعی و اعطای مالکیت به آن، می‌توان مالیات بر این دارایی‌ها وضع کرد و حتی با کاهش مالیات در مواردی که به‌جای ربات‌ها نیروی انسانی استخدام می‌شود، ابزاری برای مدیریت رقابت شغلی میان انسان و ربات فراهم آورد (King et al., 2017: 56) . برخی از چهره‌های برجستهٔ اقتصادی، ازجمله بیل گیتس، نیز از ایدهٔ افزایش مالیات بر ربات‌هایی که جایگزین نیروی انسانی می‌شوند حمایت کرده‌اند (Delaney, 2017: 1) . در مقابل، گروهی دیگر

بر این باورند که چنین سیاستی می‌تواند آثار منفی بر رقابتی بودن بازار داشته باشد و درنهایت به کاهش بهره‌وری فناوری منجر شود (Prodhan, 2017: 1).

۲-۲-۱-۵. قابلیت دادرسی و اجرای احکام علیه شخصیت حقوقی هوش مصنوعی

یکی از اساسی‌ترین جنبه‌های اعطای شخصیت به هوش مصنوعی، امکان طرح دعوا به نفع یا علیه آن است. این قابلیت در مقررات مدنی رباتیک اتحادیهٔ اروپا[1] نیز مورد تأکید قرار گرفته است (EU Resolution, 2018: 12). پرسش اصلی آن است که آیا بهره‌برداران یا سازندگان این فناوری می‌توانند مستقیماً مسئول شناخته شوند و دعاوی حقوقی علیه آنان اقامه شود (Joizil et al., 2022: 2) . برخی حقوقدانان و حتی نظام‌های حقوقی، فرض مسئولیت را متوجه سازندگان دانسته‌اند؛ برای نمونه، اتحادیهٔ اروپا مسئولیت مطلق را برای اپراتورهای هوش مصنوعی پرخطر پیش‌بینی کرده است. براساس این رویکرد، اپراتورها مکلف به جبران خسارات ناشی از عملکرد سامانه‌های مبتنی بر هوش مصنوعی خواهند بود، حتی در شرایطی که اقدامات لازم برای تضمین ایمنی آن‌ها صورت گرفته باشد (European Parliament, 2020/2014(INL). با این حال، چالش‌هایی همچون عدم امکان درک کامل فرایند تفسیر دستورها، تصمیم‌گیری و اجرای آن‌ها توسط سامانه‌های خودمختار هوش مصنوعی از سوی سازندگان، می‌تواند تحقق عدالت حقوقی را با مشکل مواجه سازد، زیرا در بسیاری از موارد حتی سازندگان نیز از نحوهٔ تحلیل داده‌ها و نتیجه‌گیری این سامانه‌ها آگاهی دقیق ندارند. برای نمونه، بررسی سامانهٔ «کامپَس»[2] که به‌عنوان یک سیستم مبتنی بر هوش مصنوعی در دادگاه‌های ایالات متحده برای ارزیابی احتمال تکرار جرم مورد استفاده قرار می‌گیرد، نشان داد که سیاه‌پوستان در مقایسه با سفیدپوستان کمتر مشمول آزادی موقت می‌شوند؛ هرچند رنگ پوست به‌طور رسمی در معیارهای تصمیم‌گیری این سامانه لحاظ نشده بود (University of Helsinki, 2019: 2).

علاوه بر این، روند طرح دعاوی علیه شرکت‌های توسعه‌دهنده به‌دلیل عملکردهای ناشی از هوش مصنوعی رو به افزایش است؛ موضوعی که ضرورت اعطای شخصیت مستقل به برنامه‌های هوش مصنوعی و جداسازی آن‌ها از مسئولیت مستقیم سازندگان را برجسته‌تر می‌سازد. برای نمونه، دو نویسندهٔ رمان‌های پرفروش در سال ۲۰۲۳ با این ادعا که دسترسی رایگان کاربران به خلاصهٔ آثارشان

1. Civil Law Rules on Robotics
2.Correctional Offender Management Profiling for Alternative Sanction (COMPAS)

در چت‌جی‌پی‌تی[1] موجب کاهش قابل توجه فروش کتاب‌ها شده است، شکایتی را علیه شرکت توسعه‌دهندهٔ این فناوری در دادگاه‌های ایالات متحده مطرح کردند (Mona Awad & Paul Tremblay VS Open AI co, 2023). همچنین در جریان شکایتی گسترده علیه شرکت اوپن‌ای‌آی،[2] درخواست توقف توسعهٔ چت‌جی‌پی‌تی تا زمان روشن شدن مخاطرات احتمالی هوش مصنوعی ارائه شد (Winshel, 2023: 1).

نتیجه‌گیری

تحولات سریع در حوزهٔ هوش مصنوعی و گسترش کاربردهای آن در عرصه‌های اجتماعی، امنیتی و اقتصادی، ضرورت بازنگری در مبانی بزه‌انگاری و سازوکارهای تقنینی را بیش‌ازپیش آشکار کرده است. در چنین شرایطی، این پژوهش درپی پاسخ به این پرسش بود که آیا نظام حقوق کیفری ایران توانایی لازم برای مواجهه با جرائم نوظهور مبتنی بر هوش مصنوعی را دارد یا خیر. بررسی‌های انجام‌شده نشان داد که چارچوب‌های موجود حقوق کیفری، به‌ویژه در حوزهٔ جرم‌انگاری رفتارهای فناورانه و تعیین حدود مسئولیت، از انسجام و کفایت لازم برخوردار نیستند. قوانین فعلی عمدتاً بر مبنای رفتارهای انسانی تدوین شده‌اند و توانایی تطبیق با شرایطی را که در آن سامانه‌های هوشمند به‌طور مستقل تصمیم‌گیری و اقدام می‌کنند، ندارند.

یافته‌های تحقیق فرضیهٔ اصلی را تأیید کردند؛ بدین معنا که مقررات کنونی در برابر تصمیمات خودکار و خودمختار سامانه‌های هوش مصنوعی با خلأ جدی مواجه‌اند. این خلأ نه‌تنها در سطح تعریف جرم بلکه در سطح انتساب مسئولیت کیفری نیز آشکار است، برای مثال در مواردی که یک سامانهٔ هوشمند اقدام به تولید محتوای جعلی یا اجرای حملهٔ سایبری می‌کند، قوانین موجود قادر به تعیین دقیق مسئولیت و شناسایی عنصر معنوی جرم نیستند. از این رو، نظام حقوق کیفری ایران نیازمند بازنگری اساسی در مفاهیم بنیادین خود است تا بتواند با واقعیت‌های جدید فناوری هماهنگ شود. این نتیجه بیانگر آن است که نظام حقوق کیفری ایران در برابر تحولات فناورانه، به‌ویژه در حوزهٔ هوش مصنوعی، فاقد ابزارهای تقنینی کارآمد برای جرم‌انگاری و پاسخ‌گویی مؤثر است. به عبارت دیگر، چارچوب‌های سنتی که بر عقل، اراده و قصد انسانی بنا شده‌اند، در مواجهه با فناوری‌های

1. ChatGPT
2. OpenAI

خودمختار ناکارآمدی خود را نشان داده‌اند. بنابراین، ضرورت دارد که قانون‌گذار با تدوین مقررات نوین، هم امکان جرم‌انگاری مستقل تهدیدات ناشی از هوش مصنوعی را فراهم سازد و هم سازوکارهای جدیدی برای انتساب مسئولیت کیفری در شرایطی که تصمیمات توسط سامانه‌های هوشمند اتخاذ می‌شود، پیش‌بینی کند. تنها در این صورت است که می‌توان از یک سو امنیت دیجیتال شهروندان را تضمین کرد و از سوی دیگر مسیر توسعهٔ مسئولانهٔ فناوری‌های هوش مصنوعی را هموار ساخت.

نوآوری اصلی این پژوهش در ارائهٔ یک نقشه‌راه برای بزه‌انگاری تهدیدات نوظهور ناشی از هوش مصنوعی است؛ بدین معنا که برای نخستین‌بار مجموعه‌ای از مصادیق در چهار حوزهٔ کلیدی؛ جرائم علیه اشخاص، امنیت، اموال (حقوق مالی) و فناوری‌های خودران با رویکردی نظام‌مند و هم‌زمان نظری و کاربردی تحلیل شده‌اند. در حوزهٔ جرائم علیه تمامیت اشخاص، تمرکز بر جعل عمیق، دستکاری چهره، صوت و فیشینگ هوشمند نشان داد که آسیب به حریم خصوصی، حیثیت و خودمختاری فردی در مقیاسی بی‌سابقه ممکن است و جرم‌انگاری مستقل با تعریف عناصر مادی و معنوی خاص این رفتارها ضرورت پیدا می‌کند. در حوزهٔ تهدید امنیت، حملات تقویت‌شده با یادگیری ماشین علیه زیرساخت‌های حیاتی، از حملات تطبیقی تا اکتشاف خودکار رخنه‌ها، نشان دادند که صرف اتکا به دسته‌بندی‌های سنتی جرائم رایانه‌ای پاسخ‌گو نیست و نیاز به تعریف حقوقی «جرم سایبری تقویت‌شده با هوش مصنوعی» و سازوکارهای شناسایی و بازدارندگی متناسب وجود دارد. در حوزهٔ اموال، باج‌افزارهای هوشمند، استخراج غیرقانونی ارز دیجیتال مبتنی بر بات‌نت و تقلب‌های الگوریتمی در تراکنش‌ها، شکاف‌های کیفرگذاری و بی‌اثری نسبی سازوکارهای سنتی بازدارندگی را آشکار کردند. نهایتاً در حوزهٔ فناوری‌های خودران و سلاح‌های خودمختار، مسئلهٔ تصمیم‌گیری مستقل سامانه و توزیع مسئولیت بین طراح، بهره‌بردار و سیستم هوشمند، ضرورت رژیم‌های مسئولیت ترکیبی و قواعد خاص منع و کنترل را برجسته ساخت.

بررسی‌ها نشان داد که یکی از چالش‌های اساسی در مواجهه با تهدیدات نوظهور هوش مصنوعی، ناکارآمدی قواعد سنتی مسئولیت کیفری در برابر سطح بالای خودمختاری سامانه‌های هوشمند است. فاصلهٔ روزافزون میان ارادهٔ انسان و رفتار نهایی این سامانه‌ها، ضرورت بازاندیشی در مبانی انتساب مسئولیت و حتی امکان شناسایی نوعی شخصیت حقوقی ویژه برای هوش مصنوعی را برجسته می‌کند؛ رویکردی که با تکیه بر مفاهیمی چون استقلال مالی، قابلیت دادرسی و مسئولیت

لایه‌ای، می‌تواند چارچوبی واقع‌بینانه‌تر برای تنظیم‌گری و پاسخ کیفری فراهم سازد. این چارچوب، به‌جای برخورد موردی یا صرفاً فنی، با تلفیق تحلیل مفهومی (بازتعریف عنصر معنوی، قصد و تقصیر در بستر تصمیم‌گیری خودکار) و اقتراح تقنینی (طراحی عناوین مجرمانه مستقل، عناصر قانونی و حدود مسئولیت)، نشان می‌دهد که قواعد سنتی مبتنی بر انسانِ دارای عقل و اراده، برای مواجهه با کنش‌های غیرانسانیِ خودمختار کافی نیستند. افزون بر این، پژوهش با مقایسهٔ تطبیقی رویکردهای خارجی و اقتباس سنجیدهٔ مؤلفه‌هایی که با بستر حقوقی ایران سازگارند، یک چارچوب نظری برای طراحی مقررات آینده ارائه کرده است: تعریف‌های قانونی شفاف برای جرائم تقویت‌شده با هوش مصنوعی، معیارهای احراز عنصر معنوی در تصمیمات خودکار، رژیم مسئولیت ترکیبی و استانداردهای دقت و ایمنی برای توسعه و استقرار سامانه‌های هوشمند. این چارچوب به‌منزلهٔ نقشهٔ عمل تقنینی، مسیر تازه‌ای برای مطالعات حقوقی در پیوند با هوش مصنوعی می‌گشاید: از تدوین مواد مستقل و آیین‌های اثبات تخصصی، تا سازوکارهای نظارتی و همکاری‌های بین‌المللی. به این ترتیب، نوآوری پژوهش صرفاً در شناسایی خلأها خلاصه نمی‌شود، بلکه با ارائهٔ مدل مفهومی و تقنینی قابل‌اجرا، گذار از پارادایم سنتی به نظم کیفری سازگار با فناوری‌های خودمختار را امکان‌پذیر می‌کند.

بررسی مصادیق نشان داد که در حوزهٔ جرائم علیه اشخاص، نمونه‌هایی همچون جعل عمیق (دیپ‌فیک) و فیشینگ هوشمند، به‌طور جدی بنیان‌های سنتی حمایت کیفری از حیثیت، آبرو و امنیت فردی را به چالش کشیده‌اند. جعل عمیق با بهره‌گیری از الگوریتم‌های یادگیری عمیق قادر است تصاویر و ویدئوهای کاملاً مشابه واقعیت تولید کند و از این طریق، امکان تخریب شخصیت، انتشار اکاذیب یا سوءاستفاده‌های اخلاقی را فراهم سازد. فیشینگ هوشمند نیز با استفاده از تحلیل داده‌های رفتاری قربانیان، حملات هدفمند و شخصی‌سازی‌شده‌ای را طراحی می‌کند که احتمال فریب را به‌طور چشمگیری افزایش می‌دهد. قوانین موجود تنها به‌طور پراکنده و غیرمستقیم قابلیت انطباق با این جرائم را دارند و نمی‌توانند به‌طور کامل از حقوق فردی حمایت کنند؛ از این رو، جرم‌انگاری مستقل و صریح این رفتارها ضرورت دارد. در حوزهٔ جرائم علیه امنیت، نبود تعریف قانونی روشن از «جرم سایبری تقویت‌شده با هوش مصنوعی» تهدیدی جدی برای زیرساخت‌های حیاتی کشور محسوب می‌شود. حملات سایبری مبتنی بر هوش مصنوعی می‌توانند با سرعت و دقتی فراتر از توان انسان، آسیب‌های گسترده‌ای به شبکه‌های انرژی، حمل‌ونقل، ارتباطات و سامانه‌های

دفاعی وارد کنند. فقدان چارچوب قانونی مشخص برای شناسایی و مقابله با این نوع جرائم، نه‌تنها موجب سردرگمی در فرایند دادرسی می‌شود بلکه امنیت ملی را به‌طور مستقیم تحت تأثیر قرار می‌دهد. بنابراین، ضرورت دارد قانون‌گذار با تعریف دقیق این دسته از جرائم و تعیین عناصر قانونی آن‌ها، امکان واکنش سریع و مؤثر نظام کیفری را فراهم سازد. در حوزهٔ جرائم علیه اموال (حقوق مالی)، مصادیقی همچون باج‌افزارهای هوشمند و استخراج غیرقانونی ارز دیجیتال با استفاده از بات‌نت‌ها نشان دادند که چارچوب‌های سنتی بازدارندگی و مجازات پاسخ‌گوی پیچیدگی‌های ناشی از هوش مصنوعی نیستند. باج‌افزارهای مبتنی بر هوش مصنوعی قادرند به‌طور پویا روش‌های رمزگذاری و اخاذی خود را تغییر دهند و شناسایی آن‌ها را دشوار سازند. استخراج غیرقانونی ارز دیجیتال نیز با بهره‌گیری از شبکه‌های گستردهٔ بات‌نت، خسارت‌های مالی هنگفتی به کاربران و شرکت‌ها وارد می‌کند. قوانین موجود که عمدتاً بر سرقت سنتی یا کلاهبرداری کلاسیک تمرکز دارند، توانایی مقابله با این جرائم پیچیده و فناورانه را ندارند و نیازمند قواعد نوین و اختصاصی هستند. همچنین، در حوزهٔ فناوری‌های خودران و سلاح‌های خودمختار، فقدان رژیم مسئولیت ترکیبی میان عامل انسانی و سامانهٔ هوشمند خلأ تقنینی آشکاری ایجاد کرده است. خودروهای خودران یا سلاح‌های خودمختار می‌توانند بدون دخالت مستقیم انسان تصمیماتی اتخاذ کنند که پیامدهای کیفری گسترده‌ای داشته باشد؛ از تصادفات مرگبار گرفته تا حملات نظامی خارج از کنترل انسانی. در چنین شرایطی، پرسش اساسی آن است که مسئولیت کیفری بر عهدهٔ چه کسی قرار می‌گیرد؟ نبود پاسخ روشن به این پرسش، نظام کیفری را در وضعیت بلاتکلیفی قرار داده و ضرورت تدوین مقررات ویژه و مستقل را دوچندان ساخته است. این مقررات باید بتوانند مسئولیت را میان عامل انسانی و فناوری به‌صورت ترکیبی تقسیم کنند و از بروز خلأهای حقوقی در مواجهه با جرائم ناشی از فناوری‌های خودمختار جلوگیری نمایند.

درنهایت، پژوهش حاضر تأکید می‌کند که ثبات حقوقی در برابر پویایی فناوری تنها با رویکردی میان‌رشته‌ای و آینده‌نگر محقق می‌شود؛ رویکردی که بر پایهٔ پژوهش‌های دانشگاهی، مشارکت فعال نهادهای اجرایی و تعامل مستمر با جامعهٔ فناوران بنا شود. بدین ترتیب، نظام حقوق کیفری ایران باید در توازن میان حفظ نوآوری و صیانت از حقوق فردی و جمعی بازطراحی شود تا هم امنیت دیجیتال شهروندان تضمین شود و هم مسیر توسعهٔ فناوری‌های هوش مصنوعی مسدود نشود. این نتیجه‌گیری نشان می‌دهد که آیندهٔ حقوق کیفری ایران در گرو پذیرش واقعیت‌های فناورانه و ایجاد چارچوب‌های

تقنینی نوین است؛ چارچوب‌هایی که نه‌تنها پاسخ‌گوی تهدیدات نوظهور باشند، بلکه زمینه‌ساز توسعهٔ پایدار و مسئولانه فناوری‌های هوش مصنوعی نیز شوند.

پیشنهادها

یافته‌های پژوهش حاضر نشان داد که چارچوب‌های موجود حقوق کیفری ایران در مواجهه با جرائم نوظهور مبتنی بر هوش مصنوعی، به‌ویژه در حوزه‌های انتساب مسئولیت، عنصر معنوی و قصد مجرمانه، با خلأهای جدی روبه‌روست. این خلأها نه‌تنها موجب پراکندگی و عدم انسجام در مقررات موجود می‌شوند، بلکه امکان بهره‌برداری مجرمانه از ظرفیت‌های خودکار و مقیاس‌پذیر فناوری‌های هوش مصنوعی را نیز تسهیل می‌کنند. ازاین‌رو، ارائهٔ پیشنهادهای تقنینی مشخص و تدوین مواد قانونی مستقل، ضرورتی انکارناپذیر برای تضمین عدالت کیفری، صیانت از امنیت ملی و حمایت از حقوق فردی و جمعی است. پیشنهادهای زیر بر پایهٔ تحلیل‌های نظری و تطبیقی، و با الهام از تجربه‌های تقنینی بین‌المللی، به‌عنوان پیش‌نویس مواد قانونی ارائه می‌شوند. هدف آن‌ها ایجاد چارچوبی منسجم و آینده‌نگر برای جرم‌انگاری رفتارهای فناورانهٔ نوظهور، تعیین مسئولیت کیفری در حوزهٔ سامانه‌های خودمختار و تقویت کارآمدی نظام حقوق کیفری ایران در برابر تهدیدات ناشی از هوش مصنوعی است.

با توجه به خلأهای شناسایی‌شده در نظام حقوق کیفری ایران در مواجهه با جرائم نوظهور مبتنی بر هوش مصنوعی، این پژوهش مجموعه‌ای از پیش‌نویس مواد قانونی پیشنهادی را ارائه می‌کند که می‌تواند مبنای اصلاح و توسعهٔ تقنینی در آینده قرار گیرد.

۱. جرم‌انگاری دیپ‌فیک در نظام حقوق کیفری ایران

مادهٔ قانونی پیشنهادی: هر شخصی که با استفاده از فناوری‌های هوش مصنوعی یا هر وسیلهٔ دیگری، تصاویر، صداها، ویدئوها یا هر نوع محتوای دیجیتال دیگری را به نحوی تغییر دهد که به ایجاد یک نمایش غیرواقعی و فریبنده منجر شود و این عمل به قصد آسیب رساندن به شخصیت، حیثیت، حقوق مالی یا معنوی اشخاص حقیقی یا حقوقی انجام گیرد، به‌عنوان مرتکب جرم دیپ‌فیک شناخته شده و به مجازات مقرر در این قانون محکوم خواهد شد.

تبصرهٔ ۱: در صورتی که این عمل منجر به آسیب جدی به شخصیت، حیثیت، حقوق مالی یا معنوی اشخاص شود، مرتکب به حداکثر مجازات مقرر در این قانون محکوم خواهد شد.

تبصرهٔ ۲: استفاده از فناوری دیپ‌فیک برای اهداف آموزشی، تحقیقاتی، هنری و خبری که به‌منظور فریب و آسیب رساندن به اشخاص نباشد، از شمول این قانون خارج است.

۲. جرم‌انگاری حملات فیشینگ در حقوق کیفری ایران

مادهٔ قانونی پیشنهادی: هرگاه شخصی با استفاده از روش‌های فریبنده ازقبیل ایجاد صفحات وب جعلی، ارسال ایمیل‌های گمراه‌کننده، پیام‌های متنی یا هر وسیلهٔ ارتباطی دیگر، با قصد کلاهبرداری یا تحصیل غیرقانونی اطلاعات شخصی، مالی یا امنیتی افراد اقدام به حملهٔ فیشینگ کند، مرتکب جرم محسوب شده و به مجازات‌های زیر محکوم می‌شود:

الف) حبس تعزیری درجهٔ ۶ (بیش از شش ماه تا دو سال)؛

ب) مصادرهٔ اموال و منافع ناشی از جرم به‌عنوان مجازات تکمیلی؛

ج) محرومیت از استفاده از خدمات رایانه‌ای و اینترنتی به مدت دو تا پنج سال (در قالب محرومیت از حقوق اجتماعی درجهٔ ۶).

تبصرهٔ ۱: در صورتی که حملهٔ فیشینگ به آسیب جدی به امنیت ملی، اقتصاد کشور یا سلامت روانی اشخاص منجر شود، مرتکب به اشد مجازات مقرر در این ماده محکوم خواهد شد.

تبصرهٔ ۲: هرگونه اقدام به حملهٔ فیشینگ که به تحصیل اطلاعات منجر نشود، لیکن قصد کلاهبرداری یا فریب آشکار باشد، به‌عنوان شروع به جرم تلقی شده و مرتکب به مجازات‌های متناسب با آن محکوم می‌شود.

۳. جرم‌انگاری استفاده از سلاح‌های خودمختار مبتنی بر هوش مصنوعی

مواد قانونی پیشنهادی:

مادهٔ ۱ ـ استفاده از سلاح‌های خودمختار مبتنی بر هوش مصنوعی که قابلیت تصمیم‌گیری و اجرای عملیات نظامی به‌صورت مستقل دارند و به نقض حقوق بشردوستانهٔ بین‌المللی یا مقررات ناظر بر مخاصمات مسلحانه منجر شوند، ممنوع بوده و جرم محسوب می‌شود.

مادهٔ ۲ ـ هر شخص حقیقی یا حقوقی که در تولید، طراحی، توزیع، فروش، واردات، صادرات، نگهداری یا بهره‌برداری از سلاح‌های موضوع مادهٔ ۱ مشارکت یا مباشرت ورزد، به مجازات تعزیری درجهٔ ۵ محکوم خواهد شد.

مادهٔ ۳ ـ در صورتی که استفاده از سلاح‌های خودمختار به ایراد صدمهٔ جسمی یا قتل اشخاص منجر شود، مرتکب علاوه بر مجازات‌های مقرر در مادهٔ ۲، به مجازات‌های مقرر در قانون مجازات اسلامی برای قتل یا جرح عمدی محکوم خواهد شد.

مادهٔ ۴ ـ چنانچه سازندگان یا برنامه‌ریزان اثبات کنند که تمامی اقدامات لازم برای پیشگیری از ارتکاب جرم توسط سامانهٔ هوش مصنوعی را انجام داده‌اند و وقوع جرم ناشی از نقص غیرقابل پیش‌بینی در سیستم بوده است، مسئولیت کیفری به سامانهٔ هوش مصنوعی منتسب خواهد شد.

تبصره: به‌منظور پوشش خسارت‌های ناشی از خطاهای غیرقابل پیش‌بینی سامانه‌های هوش مصنوعی، صندوق‌های بیمه‌ای ویژهٔ خودروهای خودران ایجاد می‌شود و جبران خسارت از طریق این صندوق‌ها یا بیمه‌های اختصاصی انجام خواهد شد.

مادهٔ ۵ ـ دولت مکلف است نظارت مستمر و دقیق بر توسعه، تولید و استفاده از فناوری‌های هوش مصنوعی در حوزهٔ نظامی اعمال و تضمین کند که هرگونه بهره‌برداری از این فناوری‌ها با اصول حقوق بشر و قواعد حقوق بین‌الملل بشردوستانه مطابقت دارد.

مادهٔ ۶ ـ هرگونه تغییر، اصلاح یا ارتقا در سلاح‌های خودمختار مبتنی بر هوش مصنوعی که احتمالاً به نقض مقررات این قانون منجر شود، باید پیش از اجرا به تأیید مراجع قانونی ذی‌صلاح برسد.

۴. جرم‌انگاری رفتارهای ناشی از هوش مصنوعی در خودروهای خودران

مادهٔ قانونی پیشنهادی: هرگاه در نتیجهٔ خطا یا اختلال ناشی از سامانهٔ هوش مصنوعی به‌کاررفته در خودروهای خودران، قتل یا صدمهٔ بدنی واقع شود و این خطا ناشی از برنامه‌ریزی یا طرح‌ریزی قبلی مهندسان یا کاربران نباشد، مسئولیت کیفری به سامانهٔ هوش مصنوعی منتسب خواهد شد. در این موارد، علاوه بر الزام به جبران خسارت وارده، ضمانت اجراهای کیفری متناسب با شدت جرم بر سامانهٔ مذکور اعمال می‌شود. در صورتی که امکان اجرای مستقیم مجازات نسبت به سامانه وجود

نداشته باشد، مسئولیت کیفری به‌صورت ترکیبی بر عهدهٔ مالک، بهره‌بردار یا نهاد متولی سامانه قرار خواهد گرفت.

تبصرهٔ ۱: در صورتی که خطا ناشی از نقص طراحی یا تولید باشد، مسئولیت کیفری و مدنی بر عهدهٔ شرکت سازنده یا مالک خودرو قرار خواهد گرفت.

تبصرهٔ ۲: به‌منظور پوشش خسارت‌های ناشی از خطاهای غیرقابل پیش‌بینی سامانه‌های هوش مصنوعی، صندوق‌های بیمه‌ای ویژه خودروهای خودران ایجاد می‌شود و جبران خسارت از طریق این صندوق‌ها یا بیمه‌های اختصاصی انجام خواهد شد.

۵. جرم‌انگاری حملات سایبری مبتنی بر هوش مصنوعی

مادهٔ قانونی پیشنهادی: هرگاه شخصی با بهره‌گیری از هوش مصنوعی یا سامانه‌های خودکار، مرتکب دستکاری، تخریب، دسترسی غیرمجاز، اختلال یا ایجاد وقفه در سامانه‌های رایانه‌ای، شبکه‌های اطلاع‌رسانی، زیرساخت‌های حیاتی یا داده‌های شخصی شود، به مجازات حبس و جزای نقدی درجهٔ ۵ محکوم خواهد شد. در صورتی که این اقدامات به آسیب جدی به اشخاص، امنیت ملی یا اقتصاد کشور منجر شود، مرتکب به حداکثر مجازات مقرر محکوم می‌شود.

تبصرهٔ ۱: در مواردی که سامانهٔ هوش مصنوعی بدون دخالت مستقیم انسان و به‌صورت خودکار مرتکب جرم شود، مسئولیت کیفری بر عهدهٔ طراح، برنامه‌نویس، مالک یا بهره‌بردار سامانه خواهد بود؛ مگر آنکه ثابت شود وقوع جرم ناشی از عوامل خارج از اراده و پیش‌بینی آنان بوده است.

تبصرهٔ ۲: استفاده از هوش مصنوعی برای مقاصد دفاعی، امنیتی یا انجام وظایف قانونی توسط دولت و نهادهای ذی‌صلاح، مشروط به رعایت قوانین و مقررات مربوط، از شمول این ماده خارج است.

۶. جرم‌انگاری باج‌افزارهای مبتنی بر هوش مصنوعی

مادهٔ قانونی پیشنهادی: هرگاه شخصی با بهره‌گیری از هوش مصنوعی یا سایر فناوری‌های مرتبط، اقدام به طراحی، توسعه، انتشار، توزیع یا کنترل نرم‌افزارهای باج‌افزاری کند، یا به هر نحو از این نرم‌افزارها برای تهدید، اعمال فشار یا اخاذی از اشخاص حقیقی یا حقوقی استفاده کند، مرتکب جرم باج‌افزار محسوب می‌شود.

تبصره: در صورتی که اقدامات مذکور به آسیب یا اختلال جدی در سامانه‌های حیاتی، زیرساخت‌های ملی یا امنیت عمومی منجر شود، مرتکب به اشد مجازات مقرر محکوم خواهد شد.

۷. جرم‌انگاری استخراج ارز دیجیتال با استفاده از بات‌نت‌ها و هوش مصنوعی

مادۀ قانونی پیشنهادی: هرگاه شخصی با استفاده از هوش مصنوعی یا بات‌نت‌ها، بدون اخذ مجوز قانونی از مراجع ذی‌صلاح، اقدام به استخراج ارز دیجیتال کند، مرتکب جرم استخراج غیرمجاز ارز دیجیتال محسوب می‌شود.

پیشنهادهای قانونی مطرح‌شده در این پژوهش، با توجه به ماهیت تهدیدات نوظهور ناشی از هوش مصنوعی، در دو سطح قابل پیگیری است. نخست، در مواردی که رفتارهای جدید را می‌توان ذیل جرائم رایانه‌ای یا جرائم سنتی موجود قرار داد، اصلاح یا الحاق مواد قانونی کفایت می‌کند. برای نمونه، جعل عمیق (دیپ‌فیک) را می‌توان با افزودن تبصره‌ای به مادۀ ۶ قانون جرائم رایانه‌ای یا مادۀ ۱۶ همان قانون، به‌طور صریح و شفاف جرم‌انگاری کرد. این رویکرد ضمن حفظ انسجام نظام تقنینی موجود، امکان پاسخ‌گویی به بخشی از تهدیدات فناورانه را فراهم می‌سازد. در مقابل، در حوزه‌هایی که تهدیدات ماهیت کاملاً جدید دارند و چارچوب‌های موجود توانایی پاسخ‌گویی به آن‌ها را ندارند، تدوین قانون مستقل ضرورت می‌یابد. نمونۀ بارز این وضعیت را می‌توان در زمینۀ سلاح‌های خودمختار مرگبار و خودروهای خودران مشاهده کرد؛ فناوری‌هایی که تصمیمات الگوریتمی آن‌ها مسئولیت کیفری را با ابهام جدی مواجه ساخته است. در چنین مواردی، ایجاد یک رژیم تقنینی مستقل و جامع، نه‌تنها برای تعیین حدود مسئولیت کیفری بلکه برای تضمین انطباق با اصول حقوق بشردوستانه و عدالت کیفری، اجتناب‌ناپذیر است.

مواد قانونی پیشنهادی این پژوهش، با تمرکز بر جرائم نوظهور مبتنی بر هوش مصنوعی، تلاشی است برای پر کردن شکاف‌های تقنینی موجود در نظام حقوق کیفری ایران. این پیشنهادها نه‌تنها به جرم‌انگاری صریح رفتارهایی چون دیپ‌فیک، فیشینگ پیشرفته، باج‌افزار و استخراج غیرقانونی ارز دیجیتال می‌پردازند، بلکه ابعاد جدیدی از مسئولیت کیفری در حوزۀ فناوری‌های خودران و خودمختار، ازجمله خودروهای خودران و سلاح‌های هوشمند، را نیز روشن می‌سازند. پیش‌بینی مقررات ویژه برای حملات سایبری تقویت‌شده با هوش مصنوعی و تعیین مسئولیت ترکیبی میان عامل انسانی و سامانه‌های هوشمند، نشان‌دهندۀ ضرورت بازتعریف مفاهیم بنیادین حقوق کیفری

همچون «عنصر معنوی» و «قصد» در مواجهه با تصمیمات خودکار است. این مواد پیشنهادی، علاوه بر ایجاد انسجام تقنینی، می‌توانند عدالت کیفری را تضمین کنند، امنیت ملی و اجتماعی را تقویت سازند و از سوءاستفاده‌های فناورانه جلوگیری کنند. درنهایت، باید تأکید کرد که این پیش‌نویس‌ها نقطهٔ آغاز برای گفت‌وگوی علمی و تقنینی میان دانشگاهیان، قانون‌گذاران و نهادهای اجرایی هستند. تصویب و اجرای چنین مقرراتی نیازمند بررسی دقیق، اجماع کارشناسی و هماهنگی با استانداردهای بین‌المللی است تا نظام حقوق کیفری ایران بتواند همگام با تحولات فناورانه، کارآمدی خود را در مدیریت ریسک‌های نوظهور حفظ کند.

حامی مالی:

این مقاله هیچ حامی مالی ندارد.

مشارکت نویسندگان:

سعید قماشی: م فهوم‌سازی، نظارت بر داده‌ها، اعتبارسنجی، مدیریت پروژه، منابع.

محمدرضا کاملی و مجتبی نیکوکلام ندیم: روش‌شناسی، تحلیل، نوشتن بررسی و ویرایش، تحقیق و بررسی، پیش‌نویس اصلی.

تعارض منافع:

بنا بر اظهار نویسندگان این مقاله تعارض منافع ندارد.

ملاحظات اخلاقی:

این پژوهش بر پایهٔ مطالعهٔ اسناد، قوانین، رویه‌های قضائی و منابع کتابخانه‌ای انجام شده و در آن هیچ‌گونه مطالعه‌ای بر روی انسان، حیوان یا داده‌های شخصی صورت نگرفته است؛ بنابراین، اخذ تأییدیه کمیتهٔ اخلاق برای این پژوهش ضروری نبوده است. با این حال، نویسنده اصول امانت‌داری علمی، صداقت پژوهشی، استناددهی صحیح و رعایت حقوق مالکیت فکری را در تمامی مراحل پژوهش رعایت کرده است.

در تهیهٔ این مقاله از ابزارهای هوش مصنوعی مولد صرفاً برای کمک به ویرایش زبانی، بهبود ساختار نگارش و ارتقای خوانایی متن استفاده شده است. این ابزارها در طراحی پژوهش، تحلیل علمی، تفسیر یافته‌ها، استنتاج نتایج یا تصمیم‌گیری‌های پژوهشی نقشی نداشته‌اند و مسئولیت کامل صحت، اصالت و محتوای نهایی مقاله بر عهده نویسنده/نویسندگان است.

موافقت نویسندگان:

تمامی نویسندگان نسخه نهایی مقاله را مطالعه و تأیید کرده‌اند و با انتشار آن موافقت دارند.

منابع

آقاشاهی، حسین (۱۳۹۷)، امکان‌سنجی مسئولیت کیفری ناشی از سیستم‌های هوش مصنوعی در فقه و حقوق، پایان‌نامهٔ کارشناسی ارشد، دانشگاه شهید مطهری، گروه فقه و اصول.

ابوذری، مهدی (۱۴۰۱)، حقوق و هوش مصنوعی، چاپ دوم، تهران: میزان.

احسان‌پور، سیدرضا و امی، امیر (۱۴۰۱)، «جرائم فناوری جعل عمیق از منظر فقه و حقوق کیفری»، فقه جزای تطبیقی، ۲(۴)، ۱-۹.

ادیب‌نیا، محمد (۱۳۹۵)، هوش مصنوعی در پرتو حقوق کیفری، پایان‌نامهٔ کارشناسی ارشد، دانشگاه آزاد اسلامی واحد شاهرود، دانشکدهٔ علوم انسانی.

اشلی، کوین (۱۴۰۱)، هوش مصنوعی و تحلیل حقوقی، ترجمهٔ امیر اسلامی‌تبار و همکاران، تهران: مجد.

بهره‌مند، حسین و مرسی، هادی (۱۴۰۱)، «راهبرد جرم‌انگاری مستقل حملات سایبری در حقوق کیفری ایران»، مجلس و راهبرد، ۲۹(۱۱۱)، ۲۸۵-۳۲۴.

جعفری لنگرودی، م. ج. (۱۴۰۱). دانشنامه حقوقی (چاپ چهارم، جلدهای ۱-۵). گنج دانش.

دوگال، پریت (۱۳۹۸)، اینترنت اشیا، هوش مصنوعی و چالش‌های حقوقی پیش‌رو، ترجمهٔ علی غلامعلی‌پور و مهدی غلامعلی‌پور، تهران: مجد.

رهبر، نرگس و دهقان‌پور فراشاه، سمیه (۱۴۰۰)، «بررسی تطبیقی مبنای مسئولیت مدنی در تصادفات وسایل نقلیهٔ خودران»، مطالعات حقوق تطبیقی، ۱۲(۲)، ۵۲۳-۵۴۳.

عطازاده، سعید و انصاری، جواد (۱۳۹۸)، «بازپژوهی مفهوم مسئولیت کیفری هوش مصنوعی (مطالعهٔ موردی خودروهای خودران) در حقوق اسلام، ایران، آمریکا و آلمان»، پژوهش تطبیقی حقوق اسلام و غرب، ۶(۴)، ۵۵-۸۶.

قیاسی، جواد (۱۳۷۵)، تسبیب در قوانین کیفری، قم: مدین.

کریمی، علی و کریمی، سارا (۱۳۹۳)، «تعارض بین مالیت بالفعل و شخصیت بالقوهٔ جنین آزمایشگاهی»، فصلنامهٔ حقوق پزشکی، ۸(۳۰)، ۱۱-۳۶.

گلدوزیان، ایرج (۱۳۸۵)، بایسته‌های حقوق جزای عمومی، تهران: میزان.

محسنی، فاطمه (۱۴۰۱)، جرم‌شناسی، تهران: دانشگاه امام صادق.

میرهادی، سید و علوی‌امین، سید محمدطه (۱۴۰۲)، هوش مصنوعی، تحولی عظیم در حوزهٔ حقوق، تهران: مهراندیشان سپهر.

نجفی ابرندآبادی، علی‌حسین و هاشم‌بیکی، حسن (۱۳۹۰)، دانشنامهٔ جرم‌شناسی، تهران: گنج دانش.

نخجوانی، علی و یاقوتی، الهام (۱۴۰۱)، «وضعیت حقوقی معاملات انجام‌شده توسط هوش مصنوعی: نظریهٔ وکیل مجازی»، پژوهش‌های حقوق اقتصادی و تجاری، ۱(۱)، ۴۱ـ۶۸.

References

Abaimov, S., & Martellini, M. (2020), "Artificial intelligence in autonomous weapon systems", *In 21st Century Prometheus* (pp. 141–177), Springer.

Acquaviva, G. (2022), "Autonomous weapons systems controlled by artificial intelligence: A conceptual roadmap for international criminal responsibility", *Military Law and Law of War Review, 60*(1), 89–121.

Agrafiotis, I., Nurse, J. R. C., Goldsmith, M., Creese, S., & Upton, D. (2018), "A taxonomy of cyber-harms", *Journal of Cybersecurity, 4*(1), tyy006.

Altmann, J., & Sauer, F. (2017), "Autonomous weapon systems and strategic stability", *Survival, 59*(5), 117–142.

Anderljung, M., et al. (2023), Frontier AI regulation. arXiv:2307.03718.

Bailey, M., Cooke, E., Jahanian, F., Xu, Y., & Karir, M. (2009), "Botnet technology", In *Cybersecurity Applications & Technology Conference* (pp. 299–304), IEEE.

Banteka, N. (2021), "Artificially intelligent persons", *Houston Law Review, 58*(3), 537–596.

Bátori, A. (2025), "International criminal law issues of autonomous weapons: The Kargu-2 drone in Libya", *Gradus, 12*(1), 1–15.

Ben Aissa, F., Hamdi, M., Zaied, M., & Mejdoub, M. (2023), "An overview of GAN-DeepFakes detection", *Multimedia Tools and Applications, 83*, 32343–32365.

Blanchard, A., & Bruun, L. (2025), Autonomous weapon systems and AI-enabled decision support, SIPRI.

Brown, A., Johnson, S. D., & Tuptuk, N. (2025), "Crime and connected and autonomous vehicles", *Crime Science, 14*, Article 2.

Chesney, R., & Citron, D. K. (2019), "Deep fakes: A looming challenge", *California Law Review*, *107*(6), 1753–1819.
Chopra, S., & White, L. F. (2011), *A legal theory for autonomous artificial agents*, Ann Arbor: University of Michigan Press.
Crootof, R. (2015), "The killer robots are here", *Cardozo Law Review*, *36*, 1837–1914.
Damiani, E. (2025), *Autonomy and accuracy in weapon systems*. ISPI.
Delaney, K. J. (2017), *Robot tax debate*. Quartz.
European Union (2013), Directive 2013/40/EU on attacks against information systems.
Ferdous, J. (2024), "AI-based ransomware detection", *IEEE Access*, *12*, 136666–136695.
Fernández Llorca, D., et al. (2025), "Testing autonomous vehicles and AI", *European Transport Research Review*, *17*, Article 38.
Fritsch, L., Jaber, A., & Yazidi, A. (2023), "AI used in malware", In *NAIS 2022*. Springer.
Gaeta (2023), "Who acts when autonomous weapons strike?", *Journal of International Criminal Justice*, *21*(5), 1033–1055.
Gandotra, V., Racicot, F.-É., & Rahimzadeh, A. (2020), "Cryptocurrency mining", In *Cryptofinance and mechanisms of exchange* (pp. 51–67), Springer.
Garikapati, D., & Shetiya, S. S. (2024), "Autonomous vehicles", *Big Data and Cognitive Computing*, *8*(4), 42.
Gherardini, L., & Cabri, G. (2023), "Impact of autonomous vehicles", *World Electric Vehicle Journal*, *15*(12), 579.
Haner, J. K., & Knake, R. K. (2021), "Breaking botnets", *Journal of Cybersecurity*, *7*(1), tyab003.
Jabbar, H., & Al-Janabi, S. (2025), "AI-driven phishing detection", *Journal of Cybersecurity and Privacy*, *5*(2), 26.
Jabir, R., Le, J., & Nguyen, C. (2025), "Phishing in the age of AI", *AI*, *6*(8), 174.
Joizil, K., Goldenberg, A., & Ghaly, S. (2022). *AI litigation risk*. TechLex Blog.
Khan, A., & Alam, T. (2025), "Integration of AI in phishing attacks", *IJNRD*, *10*(3), 1–6.
Khan, M., & Rehman, A. (2024), E*xploring deepfake technology: Creation, consequences and countermeasures*, Springer.
Kharraz, A., et al. (2015), "Ransomware attacks", In *DIMVA 2015* (pp. 3–24), Springer.
King, B., Hammond, T., & Harrington, J. (2017), "Economic consequences of AI", *Journal of Strategic Innovation and Sustainability*, *12*(2), 53–67.
Kirichenko, D. (2025), AI's role in modern warfare. War Room – U.S. Army War College.

Kurki, V. (2019), *A theory of legal personhood*, Oxford: Oxford University Press.
Loukas, G. (2015), *Cyber-physical attacks: A growing invisible threat* (1st ed.), Oxford: Butterworth-Heinemann.
Mele, A. R. (2001), *Autonomous agents: From self-control to autonomy*, Oxford: Oxford University Press.
Nelson, J. (2023). *ChatGPT false accusation*. Decrypt.
Pendleton, S. D., et al. (2017), "Autonomous vehicles", *Machines*, *5*(1), 6.
Prodhan, G. (2017), *Robot law debate*. Reuters.
Ramzan, Z. (2010), "Phishing attacks", In Handbook of Information and Communication Security. Springer.
Rigaki, M., & García, S. (2018), "GAN malware communication", In *IEEE Security and Privacy Workshops* (pp. 70–75),
Romero Moreno, F. (2024), "Generative AI and deepfakes", *International Review of Law, Computers & Technology*, *38*(3), 297–326.
Sandoval, M.-P., et al. (2024), "Threat of deepfakes", *Crime Science*, *13*(1), Article 2.
Schwarting, W., Alonso-Mora, J., & Rus, D. (2018), "Planning for autonomous vehicles. Annual Review of Control", *Robotics, and Autonomous Systems*, *1*(1), 187–210.
Solum, L. B. (1992), "Legal personhood for artificial intelligences", *North Carolina Law Review*, *70*(4), 1231–1287.
Teichmann, F. (2023), "Ransomware and generative AI", *International Cybersecurity Law Review*, *4*(3), 399–414.
University of Helsinki (2019). AI responsibility.
Winshel, A. (2023). Complaint against OpenAI. Harvard JSEL.
Taeb, M., & Chi, H. (2022). Comparison of Deepfake Detection Techniques through Deep Learning. Journal of Cybersecurity and Privacy, *2*(1), 89–106. https://doi.org/10.3390/jcp2010007
Dehghani, A., & Saberi, H. (2025). Generating and detecting various types of fake image and audio content: A review of modern deep learning technologies and tools. arXiv. https://arxiv.org/abs/2501.06227
Sandoval, M.-P., de Almeida Vau, M., Solaas, J., & Rodrigues, L. (2024). Threat of deepfakes to the criminal justice system: a systematic review. Crime Science, *13*, Article 41. https://doi.org/10.1186/s40163-024-00239-1
Sembiring, J., & Firdaus, S. U. (2025). Criminal liability for misuse of artificial intelligence (AI) in deepfake crimes. International Journal of Multiscience, *6*(1), 78–95.
https://multisciencejournal.com/index.php/ijm/article/view/432

Wright, M., Sohrawardi, S. J., & Bhattacharjee, S. (2016). Towards a framework for detecting deepfake videos [Conference paper]. Rochester Institute of Technology

Ferdous, N., et al. (2024). Deepfake video detection: challenges and opportunities. Artificial Intelligence Review, *57*, Article 10810. https://doi.org/10.1007/s10462-024-10810-6

Girhepuje, S., Verma, A., & Raina, G. (2024). A survey on offensive AI within cybersecurity. arXiv. https://doi.org/10.48550/arXiv.2410.03566

Pendleton, S., et al. (2017). Perception, planning, control, and coordination for autonomous vehicles. Machines, *5*(1), Article 6. https://doi.org/10.3390/machines5010006

Salem, A. H., Azzam, S. M., Emam, O. E., & Abohany, A. A. (2024). Advancing cybersecurity: a comprehensive review of AI-driven detection techniques. Journal of Big Data, *11*, Article 38. https://doi.org/10.1186/s40537-024-00969-8

Khanjani, Z., Watson, G., & Janeja, V. P. (2022). Audio deepfakes: A survey. Frontiers in Big Data, *5*, Article 1001063. https://doi.org/10.3389/fdata.2022.1001063

Koh, Y. K. (2025). Enhancing deepfake detection generalization through component-based development in a web platform [Final Year Project]. Universiti Tunku Abdul Rahman (UTAR). http://eprints.utar.edu.my/7295/

Sandoval, M.-P., de Almeida Vau, M., Solaas, J., & Rodrigues, L. (2024). Threat of deepfakes to the criminal justice system: a systematic review. Crime Science, *13*, Article 41. https://doi.org/10.1186/s40163-024-00239-1

Schwarting, W., Alonso-Mora, J., & Rus, D. (2018). Planning and decision-making for autonomous vehicles. Annual Review of Control, Robotics, and Autonomous Systems, *1*, 187–210. https://doi.org/10.1146/annurev-control-060117-105157

Sembiring, J., & Firdaus, S. U. (2025). Criminal liability for misuse of artificial intelligence (AI) in deepfake crimes. International Journal of Multiscience, *6*(1). https://multisciencejournal.com/index.php/ijm/article/view/432

Wright, M., Sohrawardi, S. J., & Bhattacharjee, S. (2016). Towards a framework for detecting deepfake videos [Conference paper]. Rochester Institute of Technology.

Smith, S. (2022). Audio deepfakes: A survey [Review article]. Frontiers in Big Data.

Anandakumar, H., et al. (2023). Artificial intelligence for sustainable internet research.

European Commission. (2021). Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union

legislative acts (COM/2021/206 final). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

Kurtović, S., Hasanbegović, A., & Krivić, S. (2025). Conditional variational auto encoder based dynamic motion for multitask imitation learning. In 2025 XXX International Conference on Information, Communication and Automation Technologies (ICAT).

U.S. Securities and Exchange Commission. (2020). Artificial intelligence in the securities industry (FINRA Report)

Schirmbacher, M. (2020). Germany. In International Bar Association, Guidelines and regulations to provide insights on public policies to ensure artificial intelligence's beneficial use as a professional tool (pp. 71–78). International Bar Association.

European Commission. (2021). Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM/2021/206 final).

European Parliament. (2019). European Parliament resolution of 12 February 2019 on a comprehensive European industrial policy on artificial intelligence and robotics (2018/2088(INI)). Official Journal of the European Union.

Joizil, K., Morgan, C., Glover, D., & Thérien, D. (2022). Canada. In J. Buyers (Ed.), Artificial intelligence: Law over borders comparative guide 2022. The Global Legal Post.

King, I., et al. (2017). Aspect-level sentiment classification with HEAT (HiErarchical ATtention) network. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (pp. 97–106). Association for Computing Machinery. https://doi.org/10.1145/3132847.3132877

Mahapatra, T. D. (2023, April 7). For which device Sony files patent application will blow your mind. Hindustan Times. https://www.hindustantimes.com/technology/for-which-device-sony-files-patent-application-will-blow-your-mind-101680827599538-amp.html

Tremblay v. OpenAI, Inc., No. 3:23-cv-03223 (N.D. Cal. filed June 28, 2023).