

## Criminal Accountability of Artificial Intelligence Systems for International Crimes and the Attributability of Combatant Status or E-Personhood: Necessities, Obstacles, and Solutions

Heidar Piri\* 

Assistant Professor of Public International Law, Law  
Department, Faculty of Humanities and Social  
Sciences, University of Kurdistan, Sanandaj, Iran.

### Abstract

### Introduction

The rapid development and proliferation of Artificial Intelligence (AI) systems, particularly in military and security domains, represent a paradigm shift with profound implications for international law. While offering potential benefits in efficiency and capability, the autonomous nature of advanced AI systems raises acute legal and ethical challenges, especially concerning accountability for serious violations of international law. The deployment of AI in armed conflict, such as through Lethal Autonomous Weapon Systems (LAWS), and its use in contexts that may facilitate international

---

\* Corresponding Author: H.piri@uok.ac.ir

**How to Cite:** Piri, H. (2025). Criminal Accountability of Artificial Intelligence Systems for International Crimes and the Attributability of Combatant Status or E-Personhood: Necessities, Obstacles, and Solutions. *Journal Of Criminal Law Research*, 13 (52), 151 - 196. doi: 10.22054/jclr.2025.80888.2687

crimes, like genocide, war crimes, and crimes against humanity, necessitates a fundamental re-examination of traditional legal frameworks. Current international criminal law is fundamentally anthropocentric, built upon the principle of individual criminal responsibility which requires both a physical act (*actus reus*) and a guilty mind (*mens rea*). This framework struggles to accommodate entities that can act independently, learn from their environments, and cause significant harm without direct, predictable human intervention at the moment of the act. This article delves into the core legal dilemma: who can and should be held accountable when an AI system commits an act that constitutes an international crime? It explores the feasibility and necessity of attributing criminal liability directly to AI systems themselves, potentially by granting them a novel legal status such as *electronic personality* (e-person), as opposed to, or in conjunction with, holding the various human actors in their chain of development and deployment responsible.

### **Research Question(s)**

This research is guided by the following primary questions:

- ✓ In the context of international crimes, who can be held accountable for proscribed acts or crimes committed by AI systems? Are there any grounds for labeling them as criminals or granting the status of e-person?
- ✓ What are the legal and philosophical grounds for, and obstacles against, granting autonomous AI systems a form of legal personality (e-personhood) to bear rights and obligations, including criminal liability?
- ✓ Can the existing requirements for establishing criminal responsibility under international law, particularly the *actus reus* and *mens rea*, be satisfied by AI systems in their current or foreseeable state of development?

- ✓ What alternative models of liability, such as holding programmers, manufacturers, military commanders, or states responsible, are available and effective under current international law, and what are their limitations?
- ✓ What are the potential solutions and necessary legal reforms, including at the level of the ICC Statute, to address the accountability gap posed by AI systems capable of committing international crimes?

### **Methodology**

The article has been performed based on the descriptive and analytical research method. The necessary data has been collected by library method. The research adopts a critical and forward-looking approach, analyzing the coherence and sufficiency of existing legal doctrines, identifying conceptual gaps, and proposing normative solutions based on logical reasoning, comparative analysis of analogous legal constructs (e.g., corporate criminal liability), and the functional demands of international justice.

### **Results**

The investigation yields several key findings:

- **The Case for Electronic Personality and Direct AI Liability:** Arguments for granting AI systems a form of legal personality is compelling, drawing parallels with the historical extension of legal personhood to corporations. Proponents argue that highly autonomous AI, capable of independent analysis, decision-making, and learning, possesses a functional equivalence to the rational agency required for responsibility. The concept of "electronic responsibility" is presented as a necessary tool to prevent human actors from evading liability by hiding behind the complexity and autonomy of machines.
- **Significant Legal Obstacles:** The path to direct AI criminal liability is fraught with major hurdles under current law:

- **Anthropocentric Foundations:** The Rome Statute and the general principles of international criminal law are firmly rooted in human agency. Terms like "person" are interpreted as natural persons.
- **The *Mens Rea* Requirement:** The most formidable barrier is the mental element. While an AI system can arguably satisfy the *actus reus* (the physical act), attributing *intent, knowledge, or recklessness*—subjective mental states tied to consciousness, moral understanding, and foresight—to a machine remains deeply problematic both legally and philosophically.
- **Lack of Legal Personality:** AI systems currently lack recognized legal personality in international law, a prerequisite for being a subject of rights and duties, including criminal liability.
- **Punishment Incommensurability:** Traditional penal theories (retribution, deterrence, rehabilitation) lose meaning when applied to non-human entities that cannot feel guilt, suffer, or be morally reformed.
- **Analysis of Alternative Human Liability:** In the absence of direct AI liability, the focus shifts to human actors. However, attributing responsibility to programmers, manufacturers, operators, or military commanders is often hampered by practical and legal difficulties: the problem of many hands, the challenge of proving individual *mens rea* for unforeseeable autonomous actions, and the potential lack of "effective control" required for command responsibility when dealing with learning systems.

The article concludes that a multi-pronged approach is needed:

- ✓ **Strict Liability Models:** The ICC and international criminal law may need to embrace forms of strict or no-fault liability for situations involving autonomous systems, moving away from *mens rea* as an absolute central pillar for certain contexts.
- ✓ **Regulation and Prohibition:** Strengthening IHL compliance through rigorous legal reviews of new weapons (Article 36, AP I),

enhancing precautionary measures, and potentially negotiating treaties to limit or ban certain types of autonomous weapons.

✓ **Statutory Reform:** For direct AI liability to become viable, the Rome Statute would require amendment. Articles 1 and 25(1) could be revised to explicitly extend the Court's personal jurisdiction to *legal persons* or *electronic persons*, and a new framework for *electronic responsibility* would need to be codified.

✓ **Ethical and Technical Safeguards:** Implementing robust ethical guidelines for developers, incorporating IHL rules directly into AI training (law encoding), and creating reliable fail-safe mechanisms for deactivation.

### **Conclusion**

The advent of AI systems with significant autonomy presents one of the most profound challenges to the international criminal justice system. While the theoretical appeal of holding AI directly accountable, grounded in its capacity for autonomy, independent analysis, decision-making, and a functional approximation of intentionality, is as compelling as arguments for extending legal personality to other non-human entities, it is currently precluded by foundational legal principles. The core impediments are the irreconcilable absence of a *mens rea* in AI and the lack of an established legal personality, making the imposition of direct criminal responsibility neither theoretically coherent nor practically feasible under the extant anthropocentric framework. Consequently, in the near term, the most viable accountability mechanisms must focus on reinforcing the responsibility of the myriad natural and juridical persons involved in the construction, programming, training, and deployment of AI systems for their malfunction or unlawful outcomes. However, this human-centric approach is itself severely constrained by traditional *mens rea* requirements, creating a significant liability gap and risking an anomaly where international criminal law becomes

ineffective in addressing harms caused by this technology, thereby undermining human rights protection and international justice.

To bridge this gap, a dual-path strategy is essential. Immediately, it necessitates strengthening preventive International Humanitarian Law (IHL) regulations and rigorously applying existing models of human responsibility across the AI lifecycle. Simultaneously, for the future, the international legal community must engage in proactive and principled reform. As AI evolves towards greater sophistication, the pressure to reconceptualize legal personhood will intensify. The ICC will only be able to effectively prosecute crimes involving autonomous AI agents if it embraces legal innovations such as strict liability and other alternative models of fault-based liability, which have so far been marginalized. Ultimately, ensuring accountability and preventing impunity for the gravest crimes may require international criminal law to transcend its strict anthropocentrism and incorporate a functional, graduated model of *electronic responsibility*. This represents a seismic shift in legal philosophy, demanding careful preparation and reasoned debate to develop the unique formula necessary for a future where the law keeps pace with technological agency.

**Keywords:** Electronic responsibility, AI, Autonomous weapons, International Criminal Court, Legal personality, International crimes.

## مسئولیت کیفری سیستم‌های مبتنی بر هوش مصنوعی برای ارتکاب جرائم بین‌المللی و اعطای وضعیت رزمنده یا شخصیت الکترونیکی به آن‌ها؛ ضرورت‌ها، موانع و راه‌حل‌ها

استادیار حقوق بین‌الملل عمومی، گروه حقوق، دانشکده علوم انسانی و اجتماعی، دانشگاه کردستان، سنندج، ایران.

حیدر پیری \*  ID

### چکیده

توسعه و استفاده از فناوری هوش مصنوعی، با وجود افزایش بهره‌وری و کارایی در بخش‌های مختلف زندگی، با ایجاد ضرر و زیان و ارتکاب جرم، مشکلات حقوقی بی‌شماری را در زمینه مسئولیت کیفری فردی ایجاد کرده و نظام مسئولیت کیفری بین‌المللی را به چالش کشانده است. این مقاله در صدد پاسخ به این پرسش است که چه کسی می‌تواند در قبال جرائم ارتكابی سیستم‌های مبتنی بر هوش مصنوعی مسئول شناخته شود؟ آیا امکان اعطای وضعیت «شخص الکترونیکی» به این سیستم‌ها وجود دارد تا بتوان مسئولیت کیفری را بر آنان بار کرد؟ نگارنده معتقد است اعطای شخصیت الکترونیکی و مسئولیت به سیستم‌های مبتنی بر هوش مصنوعی، با توجه به خودمختاری، توانایی تجزیه و تحلیل، تصمیم‌گیری مستقل و درجات خاصی از هوشیاری به‌عنوان رکن معنوی، به اندازه سایر افراد غیر انسانی قانع‌کننده است. با این حال، مبتنی بودن نظام عدالت کیفری بین‌المللی بر مسئولیت کیفری فردی، شرط تقصیر در احراز مسئولیت کیفری و یافتن مجازات مناسب برای هوش مصنوعی، به صورت‌بندی منحصربه‌فرد نیاز دارد؛ زیرا شخصیت هوش مصنوعی را نمی‌توان با شخصیت شخص حقیقی یا نهاد حقوقی یکسان تلقی کرد. از آنجاکه تحمیل مسئولیت کیفری بر خود هوش مصنوعی، به دلیل فقدان شخصیت حقوقی و ذهن گناهکارانه، هنوز از نظر عملی و تئوری امکان‌پذیر نیست، می‌توان افراد حقیقی و حقوقی متعددی را که در ساخت، طراحی، برنامه‌نویسی، آموزش و استقرار آن‌ها دخیل هستند، مسئول عملکرد نادرست این سیستم‌ها دانست. با وجود این، الزامات مربوط به رکن معنوی، مسئولیت کیفری فردی را به شدت محدود می‌کند و این احتمال را به وجود می‌آورد که در تطبیق قوانین بین‌المللی کیفری با این فناوری نوعی ناهنجاری پدید آید و کارایی آن‌ها در حفظ حقوق بشر و اجرای عدالت بین‌المللی از دست برود.

**کلیدواژه‌ها:** جنایات بین‌المللی، دیوان بین‌المللی کیفری، سلاح‌های خودمختار، شخصیت حقوقی، مسئولیت الکترونیکی، هوش مصنوعی.

## مقدمه

هوش مصنوعی به مثابه جدیدترین نوآوری در عصر دیجیتال شناخته می‌شود. این فناوری در سال‌های اخیر پیشرفتی بسیار سریع داشته و به طیف گسترده‌ای از کاربردهای نظامی و غیرنظامی منجر شده است (پیری، ۱۴۰۴: ۱۷۵). هوش مصنوعی که از آن به‌عنوان چهارمین انقلاب صنعتی و سومین انقلاب در امور نظامی یاد می‌شود، امروزه به‌ویژه در حوزه نظامی به واقعیتی انکارناپذیر تبدیل شده است (Ozdemir, 2019: 7) و این قابلیت را دارد که بر همه عرصه‌ها (زمین، دریا، هوا، فضا و اطلاعات) و تمامی سطوح جنگ (سیاسی، استراتژیک، عملیاتی و تاکتیکی) تأثیر بگذارد.

اگرچه امکانات فراوانی از کاربرد هوش مصنوعی حاصل می‌شود، اما خطرات و تهدیدات ناشی از آن برای وقوع آسیب و ارتکاب جنایت نیز چشمگیر است. فناوری‌های مبتنی بر هوش مصنوعی، ارتکاب جنایاتی را ممکن ساخته‌اند که پیش از این هرگز تصور نمی‌شد. سیستم‌های خودمختار هوش مصنوعی نه تنها می‌توانند به‌عنوان عاملان غیرمستقیم (با کمک، مشارکت، تسهیل یا تحریک) در ارتکاب جنایات بین‌المللی عمل کنند؛ بلکه قادرند مستقلاً مرتکب جنایات ذاتی در صلاحیت دیوان بین‌المللی کیفری و نیز جرائم خارج از صلاحیت این دیوان، از جمله قاچاق انسان، شکنجه، قاچاق مواد مخدر و سایر جرائم بین‌المللی، شوند (Swart, 2023: 591). سلاح‌های مبتنی بر هوش مصنوعی می‌توانند جنایات جنگی همچون کشتار کورکورانه غیرنظامیان را مرتکب شوند (عرب‌چادگانی و مرادیان، ۱۴۰۲: ۲۰۳۸). هنگامی که توسعه‌دهندگان، قابلیت‌های هوش مصنوعی را در یک سیستم بازجویی ادغام می‌کنند، می‌توان گفت عوامل خودمختار قادر با ارتکاب شکنجه هستند. هوش مصنوعی همچنین می‌تواند با تسهیل و تسریع در گسترش نفرت‌پراکنی، شایعات و بدنام‌سازی گروه‌های خاص (Albadi et al, 2019)، شکاف و دوگانگی میان گروه‌ها و افراد را براساس نژاد، قومیت، مذهب یا ملیت تشدید کند (Stanton, 2016). برای نمونه برخی بر این باورند که فیس‌بوک در تشدید نسل‌زدایی در میانمار نقش داشت. اعضای ارتش میانمار از این پلتفرم برای تبلیغات ضد روهینگایی و هدف‌گیری اقلیت مسلمان روهینگیا استفاده کردند (Mozur, 2018). گزارش هیئت

حقیقت‌یاب مستقل سازمان ملل نیز اعلام کرد که فیس‌بوک ابزار سودمندی برای کسانی بوده که در پی گسترش نفرت علیه روئینگیها بوده‌اند (HRC, 2018). بنابراین، توانایی هوش مصنوعی در انجام اقدامات مستقلی که در حکم جرم شناخته می‌شوند، ضرورت تنظیم قوانین کیفی حاکم بر این موضوع را اجتناب‌ناپذیر ساخته است. در سال‌های آینده، شتاب توسعه هوش مصنوعی، بازنگری اساسی در تمامی شاخه‌های حقوق، به‌ویژه حقوق کیفی، و نهادهای مرتبط را در پی خواهد داشت؛ نیازی که حل مسئله مفهومی اعطای حقوق و تعهدات به هوش مصنوعی را می‌طلبد.

ویژگی تعیین‌کننده هوش مصنوعی خودمختار، توانایی آن در انتخاب و درگیری اهداف بدون نیاز به مداخله بیشتر اپراتور انسانی است. سیستم‌های پیشرفته‌تر همچنین از ویژگی یادگیری ماشینی برخوردارند؛ به این معنا که می‌توانند به‌طور مستقل، توانایی خود را در انجام وظایف محوله با استنتاج از تجربیات گذشته بهبود بخشند (Malik, 2018: 611). همین ویژگی‌های مستقل، شناسایی شخص پاسخگو را دشوار می‌سازد. برنامه‌نویسان و توسعه‌دهندگان چنین سیستم‌هایی ممکن است قادر به پیش‌بینی و جلوگیری از عملکرد نادرست آن باشند یا نباشند؛ فرماندهی که دستور عملیات را صادر می‌کند، ممکن است توانایی پیش‌بینی عواقب را داشته یا نداشته باشد و همین امر در مورد اپراتوری که این سیستم‌های خودآموز را به کار می‌گیرد نیز مصداق دارد. اما حتی اگر مسلم باشد که «فردی» در عملکرد سلاح مبتنی بر هوش مصنوعی مرتکب اشتباه شده باشد، شناسایی فرد واحدی که علت خطای کشنده بوده، اغلب دشوار یا حتی غیرممکن است. در بسیاری موارد، فرد پشت دستگاه یا قابل شناسایی نیست یا شناسایی او بسیار دشوار است. هویت مرتکب جرم احتمالاً عمداً پنهان یا مبهم نگه داشته می‌شود. در نتیجه، بازیگران انسانی از مسئولیت کیفی بین‌المللی مبرا خواهند ماند. و برای دشوارتر کردن وضعیت، این احتمال وجود دارد که هیچ‌یک از انسان‌های درگیر، مقصر نباشند تا بتوان آنان را از نظر کیفی مسئول عواقب وخیم حمله دانست، اما عملکرد خودآموز دستگاه، به‌گونه‌ای غیرقابل پیش‌بینی، موجب شلیک به‌سوی یک هدف غیرنظامی محافظت‌شده، شده باشد (Weigend, 2023: 1137).

حقوق کیفری عموماً گذشته‌نگر و فردمحور است و برای هدایت توسعه فناوری‌ها مناسبی ندارد (Gordon, 2021: 260). قوانین و دستورالعمل‌های کیفری کنونی، به جرائمی که توسط هوش مصنوعی انجام می‌شود و از جرائم ارتکاب‌یافته توسط افراد غیرقابل تمایز است، نمی‌پردازند. بنابراین، این پرسش مطرح می‌شود که آیا ممکن است (یا باید) سیستم‌های مبتنی بر هوش مصنوعی را از منظر نظریه‌های سنتی حقوق کیفری، به‌عنوان مسئول حقوقی رفتارها و جرائم ارتکابی‌شان به رسمیت شناخت؟ این پرسش به‌ویژه با توجه به دگرگونی‌های اجتماعی اخیر و ماهیت متحول درک جامعه از جرم و رابطه با آن، مرتبط است. آیا هوش مصنوعی از شخصیت حقوقی مستقل به‌مثابه نوعی جدید، به‌عنوان شخصیت حقوقی الکترونیکی، برخوردار است که در قبال اعمال خود مسئولیت کیفری داشته باشد؟ آیا هوش مصنوعی نظامی قادر است ارکان مادی و معنوی لازم را محقق سازد؟ اگرچه برای تحمیل مسئولیت کیفری در جرائم عمدی، وجود هر دو رکن مادی و معنوی ضروری است، اما چالش اصلی، احراز رکن معنوی است؛ چگونه می‌توان قصد سیستم‌های مبتنی بر هوش مصنوعی یا یک موجود غیرانسانی را، به دلیل فقدان آگاهی، شناسایی کرد؟ رکن معنوی عمیقاً با رکن مادی پیوند خورده، چراکه این رکن معنوی است که قصد لازم برای مسئولیت کیفری را تثبیت می‌کند؛ امری که در مورد فناوری‌های مبتنی بر هوش مصنوعی دشوار می‌نماید (Dimitrova, 2022: 1-5). البته، شرط تقصیر، مانعی غیرقابل حل برای یافتن مسئول در حقوق بین‌الملل کیفری نیست. مسئولیت مبنی بر عدم تقصیر می‌تواند مبنایی قابل قبول برای مسئول شناختن ماشین‌ها در قبال جنایات بین‌المللی باشد. اعطای شخصیت حقوقی به ماشین‌ها و عوامل مستقل و سپس انتساب مسئولیتی مبتنی بر عدم تقصیر در زمینه جرائم بین‌المللی، مستلزم بازنگری در محوریت مسئولیت کیفری فردی است و بر عاملیت انسان در حقوق بین‌الملل کیفری تأکید دارد. اگر زنجیره‌ای از تابعان بالقوه حقوق بین‌الملل کیفری وجود داشته باشد، استدلال برای شخصیت الکترونیکی و مسئولیت ماشین‌ها، به همان اندازه قانع‌کننده است و دست کم به اندازه استدلال مربوط به سایر موجودات غیرانسانی مانند نهادهای شرکتی و حیوانات،

قانع‌کننده خواهد بود. اگر مسئولیت کیفری سلاح‌های مجهز به هوش مصنوعی را بپذیریم، مجازات کیفری قابل اعمال برای این سیستم‌ها چیست؟ از آنجا که تعیین مسئولیت کیفری برای تسلیحات مبتنی بر هوش مصنوعی دیدگاهی نسبتاً آینده‌نگرانه است، در ادامه به طور خلاصه به راه‌حل‌های جایگزین سنتی، یعنی تحمیل مسئولیت کیفری بر انسان‌ها، و مدل مناسب مسئولیت در قبال آنان خواهیم پرداخت.

### ۱. ضرورت مسئولیت الکترونیکی؛ موافقان تحمیل مسئولیت کیفری بر

#### سیستم‌های مبتنی بر هوش مصنوعی

ایده گسترش مسئولیت کیفری به سیستم‌های خودمختار و هوشمند، دیگر به قلمرو داستان‌های علمی تخیلی تعلق ندارد؛ بلکه از سوی شمار زیادی از اندیشمندان این حوزه پیشنهاد شده است (Hallevy, 2010: 171). استدلال اصلی برای مسئولیت کیفری و مجازات هوش مصنوعی بر توانایی آن‌ها در تصمیم‌گیری هوشمندانه نهفته است. اگر سلاح‌های مبتنی بر هوش مصنوعی بتوانند به طور مستقل هدف دشمن را شناسایی کنند و بدون دخالت انسان مرتکب اعمال مجرمانه شوند و دارای درجه مشخصی از استقلال و هوش باشند، ممکن است سطح مشخصی از مسئولیت کیفری داشته باشند. برخی از «عوامل مصنوعی» با تغییر مقادیر، ویژگی‌ها یا حالات درونی خود به محرک‌ها پاسخ می‌دهند و قادرند این حالت‌ها را بدون محرک‌های بیرونی تغییر دهند و قواعدی را که از طریق آن‌ها همان حالات تغییر می‌کنند، بهبود یا ارتقا دهند (Abbott & Sarch, 2019: 357). بنابراین، برخی از هوش مصنوعی‌ها دارای عملکردی معادل توانایی استدلال و تصمیم‌گیری فرض می‌شوند. در پرتو این ویژگی‌ها، این استدلال مطرح شده که به نهادهای هوشمند شخصیت حقوقی و به تبع آن حقوق و تعهداتی اعطا شود. این ایده برای مدت کوتاهی در اتحادیه اروپا مورد توجه قرار گرفت (European Parliament, 2017: 8).

اگرچه در حال حاضر هیچ منبای حقوقی برای انتساب مسئولیت کیفری به سیستم‌های مبتنی بر هوش مصنوعی وجود ندارد، اما از نظر اصولی هیچ مانعی بر سر راه تدوین مقرراتی

جدید نیست که این تسلیحات را در قبال آسیب به افراد تحت حمایت بین‌المللی، مسئول -حتی به صورت کیفری- بشناسد. هیچ چیز در حقوق بین‌الملل مانع جرم‌انگاری رفتار چنین اشخاصی نمی‌شود، نمونه بارز این وضعیت مسئولیت کیفری دولت‌ها در ماده ۱۹ طرح مسئولیت بین‌المللی دولت‌ها در سال ۱۹۷۶ بود، پیشنهادی که نهایتاً از طرح نهایی کمیسیون حقوق بین‌الملل حذف شد. با وجود این، در قوانین کیفری داخلی اکثر کشورها، افراد غیرانسانی (شرکت‌ها و سایر اشخاص حقوقی) مجازات می‌شوند. در واقع، یکی از دلایل تحمیل مسئولیت کیفری بر سیستم‌های هوش مصنوعی، بر پایه ترسیم شباهت بین این سیستم‌ها و شرکت‌ها است (Mazzacuva, 2021: 146-147). مسئولیت کیفری شرکت‌ها و سیستم‌های مبتنی بر هوش مصنوعی، واکنشی به معضل مشابهی است. در هر دو مورد، ردیابی مسئولیت به یک فرد دشوار است. اگرچه شرکت‌ها در شکل مدرن خود از قرن چهاردهم وجود داشته‌اند، صدها سال طول کشید تا تابع قوانین کیفری شوند (Hallevy, 2016: 200). ترفند حقوقی که منجر به این نتیجه مطلوب می‌شود این است که با شرکت‌ها به مثابه نهاد مسئول تحت قوانین کیفری رفتار شود، حتی با علم به اینکه شرکت‌ها نه جسم دارند، نه روح و محصول صرف قوانین حقوقی هستند. بنابراین، نمی‌تواند از نظر فیزیکی عمل کند، قصد و نیت داشته باشد و بفهمد که مجازات شدن به چه معناست (Weigend, 2023: 1138). آیا می‌توان از این رویکرد نظری به منظور توجیه مسئولیت در قبال فعالیت‌های مضر سیستم‌های تسلیحاتی خودمختار استفاده کرد؟ نهادهای مبتنی بر هوش مصنوعی، مانند شرکت‌ها، در بخش‌های بزرگ‌تری از فعالیت‌های انسانی نقش ایفا می‌کنند. جرائم ممکن است مستقیماً توسط نهادهای مبتنی بر هوش مصنوعی یا با واسطه‌گری آنان انجام شود. نهادهای مبتنی بر هوش مصنوعی روح ندارند و برخی از آن‌ها نه جسم دارند و نه روح. بنابراین، هیچ تفاوت حقوقی ماهوی میان ایده مسئولیت کیفری شرکت‌ها و نهادهای مبتنی بر هوش مصنوعی وجود ندارد. ظالمانه خواهد بود که آن‌ها را تابع قوانین انسانی قرار ندهیم، همان‌طور که شرکت‌ها چنین بوده‌اند.

اثر پیشگیری از جرم ناشی از احراز مسئولیت کیفری برای رفتار معین ممکن است مستقل از این باشد که آیا نهاد مسئول یک انسان است یا خیر. از این رو، برخی نظریه

کارکرد گرایانه حقوق کیفری را تأیید می‌کنند و مسئولیت کیفری را بر اساس «انتساب آزادی به عنوان یک واقعیت اجتماعی» می‌دانند (Simmler & Markwalder, 2019: 15-17). طبق این نظریه، قانون ممکن است «مفهومی از شخصیت را بپذیرد که به اهلیت عامل برای نادیده گرفتن انتظارات هنجاری بستگی دارد». سیستم‌های مبتنی بر هوش مصنوعی می‌توانند هنجارها را بی‌ثبات و چالشی ایجاد کنند که مستلزم واکنش حقوق کیفری باشند. آن‌ها به عنوان مخاطب انتظارات هنجاری در روابط متقابل اجتماعی مانند انسان‌ها یا اشخاص حقوقی قلمداد می‌شوند (Simmler & Markwalder, 2019: 25,30). وقتی (و اگر) سیستم‌های مبتنی بر هوش مصنوعی به جایگاه عضویت در جامعه برسند، منطقی است آن‌ها را به یک جرم محکوم کنیم - حتی اگر مجازات کیفری عملاً مفیدی برای سیستم‌های هوش مصنوعی وجود نداشته باشد.

یکی دیگر از رویکردها برای تثبیت مسئولیت کیفری نهادهای دارای هوش مصنوعی، بر مفهوم اصلی انتساب رفتار بر خود سیستم‌ها و -نه منحصرأ به انسان‌هایی که آن‌ها را تولید یا مورد استفاده قرار می‌دهند- استوار است. این رویکرد روشی برای مقابله با وضعیت‌هایی است که کنترل معنادار انسانی وجود ندارد (Acquaviva, 2022: 32). چنین رویکردی می‌تواند بر اساس این شناخت باشد که اگر سیستم‌های مبتنی بر هوش مصنوعی به درجه خاصی از خودمختاری برسند، وظایف پیچیده مربوط به استقرار خود و در نهایت شناسایی، درگیر شدن، و تصمیم برای حمله به اهداف خاص را انجام می‌دهند که از نظر کیفی با عملیات‌های مشابهی که انسان‌ها انجام می‌دهند، متفاوت نیستند. بنابراین، باید درجاتی از شخصیت حقوقی برای آن‌ها قائل شد. کافی نیست افراد یا برنامه‌نویسان پشت عوامل خودمختار را مسئول بدانیم، بلکه باید بتوان عوامل مستقلی را که مرتکب جنایات بین‌المللی می‌شوند، مسئول دانست (Swart, 2023: 590). بنابراین، اگر هوش مصنوعی مجهز به الگوریتم‌هایی باشد که تصمیمات اخلاقی اتخاذ کند، قادر به انتقال تصمیمات اخلاقی خود به انسان‌ها باشد و اجازه داشته باشد بدون نظارت انسان روی محیط خود عمل کند، مسئولیت کیفری آن‌ها محتمل است.

یک قانون کیفری برای هوش مصنوعی، مزایای آموزشی را برای افرادی فراهم می‌کند که مسئولیت آن‌ها را برعهده دارد (Ying, 2019:509). علاوه بر این، تحمیل مسئولیت کیفری بر هوش مصنوعی ممکن است ارزش ابزاری داشته باشد و به شناسایی افراد مقصر کمک کند که واقعاً مسئول آسیب وارد شده و عملکرد نادرست آن‌ها هستند. تحمیل چنین مسئولیتی احتمالاً بر همه افراد مرتبط با آن هوش مصنوعی تأثیر منفی می‌گذارد. اگر هوش مصنوعی به جرمی محکوم شود، سازندگان و دارندگان آن‌ها ممکن است دچار آسیب‌های قابل توجهی بشوند. در نتیجه، تحمیل مسئولیت کیفری بر هوش مصنوعی می‌تواند انگیزه بیشتری برای این افراد ایجاد کند تا مکانیسم‌هایی برای جلوگیری از رفتارهای غیرمسئولانه و اشتباهات هوش مصنوعی ایجاد کنند. به عبارت دیگر، مسئولیت کیفری هوش مصنوعی به عنوان یک وسیله خود پلیسی عمل می‌کند.

هیچ‌یک از اثرات سودمند کارکرد سرزنشی و محکومیتی حقوق کیفری را نمی‌توان با تحمیل صرف مسئولیت مدنی بر سیستم‌های مبتنی بر هوش مصنوعی، به اندازه کافی به دست آورد. مطالعات نشان می‌دهد که مردم هوش مصنوعی را به خاطر اعمالشان سرزنش می‌کنند و با انسان‌سازی تر شدن آن‌ها تمایل بیشتری به این کار دارند. در واقع، برخورد با هوش مصنوعی صرفاً به عنوان محصولی که استانداردهای مربوط به ایمنی را برآورده نمی‌کند، آسیب عاطفی قربانی را بی‌اهمیت جلوه می‌دهد. در مقابل، برچسب زدن به آن‌ها به عنوان جنایتکار، این باور قربانی را مبنی بر اینکه به طور غیرقابل توجهی آسیب دیده، اثبات می‌کند. این نشان‌دهنده شناخت جامعه از شدت ظلمی است که علیه او مرتکب شده است.

## ۲. موانع مسئولیت کیفری سیستم‌های خودمختار مبتنی بر هوش مصنوعی

اگرچه اندیشمندان و متخصصان درباره هوش مصنوعی و پاسخگویی آن‌ها در قبال اعمالشان بحث می‌کنند، اما مسئولیت خود آنان را نفی می‌کنند (Osmani, 2020: 58). مسئولیت کیفری حداقل مستلزم خودآگاهی تابع، تصمیم‌گیری آگاهانه و امکان درک ماهیت ممنوعیت‌های قانونی است. سیستم‌های هوش مصنوعی در شکل کنونی خود، براساس اهداف از پیش برنامه‌ریزی شده و شاخص‌های فنی که می‌توانند از طریق سنسورها

و نرم‌افزار خود تشخیص دهند، کار می‌کنند (Bo et al, 2022: 4). آن‌ها تصمیم‌گیری نمی‌کنند، بلکه به برنامه‌نویسان و اپراتورهای خود برای آنچه که ظاهراً «تصمیمات» خود به نظر می‌رسد وابسته هستند (Lior, 2020:1070-1071). بنابراین، رفتار یکسان با انسان‌ها و هوش مصنوعی به‌عنوان نهادهای مسئول در حقوق کیفری، نه تنها به تضعیف مبنای هنجاری ضروری مسئولیت کیفری انسان‌ها منجر می‌شود، بلکه همچنین به‌طور بالقوه انسان‌های مسئول را تشویق می‌کند تا با انتقال مسئولیت خود به دستگاه‌های هوش مصنوعی، از تعقیب کیفری شانه خالی کنند (Bryson et al, 2017: 286-287).

کسانی که مسئولیت کیفری شرکت را دلیلی بر مسئولیت کیفری غیرانسان‌ها می‌آورند استدلالشان بر بنیان ضعیفی استوار است، چراکه پذیرش مسئولیت کیفری شرکت‌ها در بسیاری از حوزه‌های قضایی محصول عمل‌گرایی پیشگیرانه از جرم است تا تفکر صحیح حقوقی (Dafni, 2018: 688). افزون‌بر این، این تردید وجود دارد که آیا ملامت اخلاقی ذاتی در مجازات کیفری می‌تواند به اندازه کافی متوجه یک شخص حقوقی شود که وجدان ندارد و نمی‌تواند احساس گناه کند. اما حتی اگر مسئولیت کیفری شرکت را به‌عنوان یک راه‌حل کیفری بپذیریم، تفاوت‌های قابل‌توجهی میان شرکت‌ها و هوش مصنوعی وجود دارد. هوش مصنوعی در مرحله کنونی توسعه خود نمی‌تواند تصمیمات واقعاً مستقل بگیرد. درحالی‌که، شرکت‌ها می‌توانند این کار را از طریق ارگان‌های خود انجام دهند. در سطح مجازات، شرکت‌ها می‌توانند - از طریق مدیران خود - از ارتکاب جرائم بیشتر خودداری کنند و می‌توانند عملکرد خود را بهبود بخشند و به «شهروندان شرکتی» بهتری تبدیل شوند. هوش مصنوعی هیچ‌کدام از این توانایی‌ها را ندارد (Osmani, 2020: 58-61). تفاوت‌های بین اشخاص حقوقی و دستگاه‌های مبتنی بر هوش مصنوعی لزوماً امکان برخورد با چنین دستگاه‌هایی را به‌عنوان مجرم‌ان جنایی در آینده منع نمی‌کند. با وجود این، این واقعیت که شرکت‌ها را می‌توان به‌عنوان تابع حقوق کیفری به رسمیت شناخت، به تنهایی، دلیلی کافی برای اعطای وضعیت مشابه به سیستم‌های مبتنی بر هوش مصنوعی امروزی، محسوب نمی‌شود. بنابراین، مهم‌ترین موانع در باب انتساب مسئولیت کیفری به هوش مصنوعی و شناسایی مسئولیت الکترونیکی

عبارتند از: مبتنی بودن نظام سنتی حقوق بین الملل کیفری بر مسئولیت فردی انسان‌ها، فقدان شخصیت حقوقی برای ماشین‌ها، شرط تقصیر در احراز مسئولیت کیفری و اتکای مسئولیت کیفری بر ارکان مادی و معنوی، عدم امکان اعطای وضعیت رزمنده به هوش مصنوعی به دلیل فقدان عقل سلیم و درک، و دشواری در تعیین و اجرای مجازات مناسب.

## ۱-۲. مسئولیت کیفری فردی و شروط آن؛ رکن مادی و معنوی و قابلیت‌های

### شناختی

یکی از مهم‌ترین کارکردهای حقوق بین الملل کیفری انتساب مسئولیت کیفری به افراد بود. قوانین کیفری بین المللی موجود را «اساساً انسان‌محور» توصیف کرده‌اند (Prosperi & Terrosi, 2017: 510). دادگاه‌های کیفری بین المللی نیز صلاحیت خود را بر همین مبنا قرار دادند. در نتیجه، انسان‌ها هنوز معیار همه چیز در حقوق بین الملل کیفری هستند. دیوان بین المللی کیفری «به وضوح از نظریه استثناگرایی انسانی حمایت کرده است، همان‌طور که تفسیر متنی، سیستمی و عملکرد بعدی پیرامون مفهوم شخص نشان می‌دهد» (Swart, 2023: 595). چارچوب‌های هنجاری با اشاره به اشخاص، انسان را به ذهن متبادر می‌سازند. ماده ۱ اساسنامه رم مدعی است که دیوان قدرت اعمال صلاحیت نسبت به «اشخاص»<sup>۱</sup> مرتکب خواهد داشت. ماده ۲۶ به صلاحیت نسبت به «اشخاص»<sup>۲</sup> اشاره دارد. ماده ۲۵ بیان می‌کند که دیوان در مورد اشخاص حقیقی صلاحیت دارد و هرکس مرتکب جرمی در صلاحیت دیوان شود، طبق این اساسنامه شخصاً مسئول است. مسئولیت کیفری فردی در مواردی قابل اعمال است که فردی مستقیماً مرتکب جرم شود یا مستقیماً از طریق دستور، برنامه‌ریزی، تحریک، اغوا، همکاری، مشارکت مجرمانه،

---

<sup>۱</sup>persons

<sup>۲</sup> اگرچه مفهوم «شخص» تاکنون به عنوان «شخص انسانی» تفسیر شده است، متن اساسنامه رم برای تفسیر فراگیرتری باز است. همان‌طور که شخص در سایر حوزه‌های حقوقی (بین المللی) اغلب به عنوان اشخاص حقیقی و حقوقی تعبیر می‌شود، ممکن است شامل اشخاص مصنوعی نیز شود. بنابراین، می‌توان با اتخاذ تفسیری گسترده‌تر از هنجارهای موجود، بر انسان‌محوری حقوق بین الملل کیفری غلبه کرد. با این حال، با توجه به پیامدهای چشمگیر گسترش این مفهوم به بازیگران هوش مصنوعی، اساسنامه رم باید اصلاح شود تا به طور صریح شامل هوش مصنوعی شود.

کمک و معاونت در آن مشارکت داشته باشد. جرائم بین‌المللی همواره با مشارکت یک کنشگر انسانی (یک شخص حقیقی) محقق می‌شود که رفتار خاصی را انجام می‌دهد (رکن مادی) و به منافع حمایت شده، از جمله حیات با عنصر ذهنی لازم (رکن معنوی) آسیب وارد می‌کند. اما در مورد سیستم‌های مبتنی بر هوش مصنوعی که به‌طور کامل توسط هوش مصنوعی کنترل می‌شود و منجر به مرگ غیرنظامیان می‌شود، با درک سنتی از رکن مادی و معنوی قابل انتساب به یک انسان نمی‌توان حل و فصل کرد. بند ۱ ماده ۳۰ اساسنامه رم تصریح می‌کند: «جز در مواردی که به نحو دیگری مقرر شده باشد، هیچ کس دارای مسئولیت کیفری نیست و نمی‌تواند به دلیل ارتکاب یکی از جرائمی که رسیدگی به آن‌ها در صلاحیت دیوان است مجازات شود مگر اینکه عنصر مادی جرم همراه با قصد و علم نیز باشد». براساس مقررۀ فوق، تحمیل مسئولیت کیفری بر یک شخص مستلزم وجود دو عنصر اصلی است: نخست، رکن خارجی یا رفتار کیفری (عنصر مادی)، و دوم، رکن داخلی یا عنصر روانی (علم یا قصد). در صورت فقدان هر یک از این عناصر، مسئولیت کیفری تحمیل نمی‌شود. هیچ معیار یا قابلیت دیگری، نه برای انسان‌ها و نه برای هیچ نهاد دیگری از جمله شرکت‌ها و هوش مصنوعی، برای تحمیل این مسئولیت مورد نیاز نیست.

## ۲-۱-۱. رکن مادی جرائم کیفری ارتكابی توسط سیستم‌های هوش مصنوعی

رکن مادی معمولاً به‌عنوان جزء خارجی یعنی انجام جرم درک می‌شود. ساختار آن برای هر نوع جرمی (عمدی یا از روی غفلت) یکسان است. بنابراین، رکن مادی مشخص می‌کند که متهم باید چه کاری انجام بدهد (فعل) یا از انجام چه چیزی خودداری کند (ترک فعل). نمی‌توان رکن مادی را به حرکات بدنی عضلانی یا حرکات بدنی ارادی سنتی تقلیل داد. در نتیجه یک عمل غیرارادی و ناخواسته همچنان می‌تواند به‌عنوان رکن مادی در نظر گرفته شود (Lagioia; Sartor, 2020: 448). باتوجه به این توصیف از رکن مادی، هم سیستم‌های هوش مصنوعی مسئول کنترل اشیای فیزیکی (روبات‌ها) و هم آن‌هایی که حضور فیزیکی ندارند (عوامل نرم‌افزاری) می‌توانند الزامات رفتاری مورد نیاز برای رکن مادی جرم را محقق کنند. این قابلیت نه تنها زمانی صادق است که فعل مورد بحث، نتیجه محاسبات درونی انجام‌شده به‌وسیله سیستم‌های هوش مصنوعی باشد، بلکه همچنین در مواقعی که این

سیستم‌ها دستورالعمل‌هایی را اجرا می‌کنند که از سوی یک اپراتور انسانی از راه دور تغذیه می‌شوند، نیز صادق است. به‌طور کلی، تحقق شرط عنصر خارجی یک جرم به‌راحتی به نهادهای مبتنی بر هوش مصنوعی نسبت داده می‌شود. تا زمانی که یک موجودیت مبتنی بر هوش مصنوعی یک مکانیزم مکانیکی یا مکانیزم دیگری را برای حرکت دادن قطعات متحرک خود کنترل می‌کند، هر فعلی ممکن است به‌عنوان عمل انجام‌شده توسط سیستم‌های مبتنی بر هوش مصنوعی در نظر گرفته شود. بنابراین، هنگامی که هوش مصنوعی بازوی الکتریکی یا هیدرولیکی خود را فعال می‌کند و به حرکت درمی‌آورد و مرتکب جرم می‌شود، ممکن است این یک فعل در نظر گرفته شود. وقتی جرمی ممکن است به دلیل ترک فعل ارتکاب یابد، احراز آن درخصوص سیستم‌های خودمختار ساده‌تر است. اگر تکلیف به اقدامی به عامل مبتنی بر هوش مصنوعی تحمیل شود، و او عمل نکند، شرط رکن مادی مربوط به جرم خاص از طریق ترک فعل برآورده می‌شود.

## ۲-۱-۲. رکن معنوی جرائم کیفری ارتكابی توسط سیستم‌های هوش

### مصنوعی؛ الزامات شناختی و اراده

حقوق کیفری قبل از اینکه بتوان گفت ارتکاب جرمی انجام شده است، مستلزم حضور رکن معنوی است. رکن معنوی به وضعیت روانی قابل سرزنش اشاره دارد. این عنصر ذهنی است که نقطه بن‌بست را از نظر مسئول نگه داشتن سیستم‌های مبتنی بر هوش مصنوعی تحت حقوق بین‌الملل کیفری نشان می‌دهد. تنها الزامات ذهنی مورد نیاز برای تحمیل مسئولیت کیفری، علم، قصد، سهل‌انگاری و غیره است که در جرم خاص و براساس نظریه عمومی حقوق جزا لازم است (Hallevy, 2010: 187).

علم به‌عنوان دریافت حسی داده‌های واقعی و درک آن داده‌ها تعریف می‌شود. اکثر سیستم‌های مجهز به هوش مصنوعی به‌خوبی برای چنین دریافتی مجهز هستند. گیرنده‌های حسی مناظر، صداها، تماس فیزیکی، لمس و غیره در اکثر سیستم‌های مبتنی بر هوش مصنوعی رایج هستند. این گیرنده‌ها داده‌های واقعی دریافت‌شده را به واحدهای پردازش مرکزی که داده‌ها را تجزیه و تحلیل می‌کنند، منتقل می‌کنند. فرایند تجزیه و تحلیل در

هوش مصنوعی به موازات درک انسان است. مغز انسان با تجزیه و تحلیل آن داده‌های دریافتی را به وسیله چشم، گوش، دست و غیره درک می‌کند. الگوریتم‌های پیشرفته هوش مصنوعی در تلاش هستند تا فرایندهای شناختی انسان را تقلید کنند. این فرایندها چندان متفاوت نیستند. نیت خاص قوی‌ترین شرط مورد نیاز از عنصر معنوی است. قصد خاص، وجود غایت یا مقصودی است که یک رویداد واقعی رخ خواهد داد. در نتیجه وجود چنین قصدی، عنصر خارجی جرم ارتکاب می‌یابد. این وضعیت منحصر به انسان نیست. یک موجودیت مبتنی بر هوش مصنوعی ممکن است طوری برنامه‌ریزی شود که هدف یا هدفی داشته باشد و اقداماتی را برای دستیابی به آن هدف انجام دهد.

در جرائم عمدی، رکن معنوی دارای دو جزء است: شناخت و اراده. شناخت عبارت است از آگاهی عامل از واقعیت‌ها و شامل تمام اجزای رکن مادی است. اراده عبارت است از قصد انجام فعل و رسیدن به نتیجه آن و هرگز نمی‌تواند تنها باشد، همیشه با آگاهی همراه است. الزامات شناختی یعنی آگاهی عامل معمولاً شامل فهم و درک است. برای آگاهی انسان، دریافت اطلاعات در مورد حقایق خاص (از طریق احساس) و تشکیل تصویری جامع از این حقایق لازم است. در ارزیابی امکان انتساب چنین فرایندهای شناختی به سیستم‌های هوش مصنوعی، باید مکانیسم‌های شناختی را جدا کنیم که این سیستم‌ها را قادر می‌سازد اطلاعات را به دست آورند و تصاویر جامع قابل استفاده بسازند. آگاهی به عنوان درک عناصر موجود در محیط در یک محدوده زمانی و مکانی، درک معنا و طرح وضعیت آن‌ها در آینده نزدیک تعریف می‌شود (Endsley & Garland, 2000). سیستم‌های هوش مصنوعی، در اصل، می‌توانند نیازهای شناختی را برآورده کنند؛ آن‌ها می‌توانند با گرفتن داده‌های واقعی و ایجاد تصاویر کلی، گذراندن سه مرحله فهم، درک و فرافکنی (برآورد)، به آگاهی از موقعیت دست یابند.

در خصوص شرط اراده، می‌توان دو سطح قصد و سهل‌انگاری را تشخیص دهیم. شرط رکن معنوی در جرائم عمدی شامل قصد تحقق عمل مجرمانه و آثار آن با آگاهی از تمامی شرایط مربوط است. بنابراین، قصد مجرمانه به سمت آینده هدایت می‌شود. قصد مجرمانه شامل یک فرایند ذهنی است که اصولاً تحت کنترل فاعل است و ممکن است به آگاهی

برسد، برخلاف امیال، غرایز یا تکانه‌های غیرقابل کنترل/ناآگاه (Lagioia; Sartor, 2020: 445). برای اثبات وجود یک جنایت بین‌المللی، یک عمل باید عمداً و با قصد و علم انجام شده باشد. مسئولیت کیفری یک نرم‌افزار ممکن است رخ ندهد زیرا پذیرفته شده است که توانایی ذهنی در هیچ سیستم مبتنی بر هوش مصنوعی یافت نمی‌شود. بنابراین، فرض وجود قصد یا علم که با عقل و منش ذهنی تعیین می‌شود در اعمال سیستم‌های مبتنی بر هوش مصنوعی دشوار است، زیرا مانند انسان دارای ویژگی‌های ذهنی یا آگاهانه نیستند و از پیامدهای اقدامات خود آگاه نیستند (Hallevy, 2015: 33) که قادر به تفکر و اتخاذ اقدامات مناسب باشند. اگر بپذیریم هوش مصنوعی توانایی جمع‌آوری دانش و تجزیه و تحلیل آن و اتخاذ تصمیمات مناسب دارد، توانایی مسئولیت‌پذیری و تثبیت رکن مادی و معنوی برای هوش مصنوعی در حقوق کیفری عملی است. توانایی تجزیه و تحلیل و تعیین تصمیمات صحیح می‌تواند به عنوان عنصر معنوی یا سوءنیت قلمداد شود.

از آنجایی که قصد یک «وضعیت ذهنی» داخلی است، برای تعیین اینکه آیا یک سیستم هوش مصنوعی دارای چنین حالت داخلی است، باید بر ساختار داخلی آن و عملکرد آن تمرکز کنیم. وضعیت داخلی یک سیستم هوش مصنوعی (مثلاً رشته‌های کاراکتر درگیر در ضبط داده و توابع ذخیره‌سازی داده‌ها که برخی از اطلاعات محیطی خاص را می‌گیرند) نشان‌دهنده اعتقاد به وجود موقعیت‌های خاصی است که آن نهاد (۱) وضعیت مربوطه را براساس چنین موقعیت‌هایی می‌پذیرد و (۲) آن وضعیت در ایجاد آن کمک می‌کند تا واحد مورد نظر مطابق این موقعیت‌ها رفتار کند. در نتیجه، حالت‌های شناختی مربوط به حقوق کیفری (شناخت/آگاهی، از جمله پاسخگویی به دلیل، و اراده/نیت)، تحت شرایط خاص، می‌توانند به سیستم‌های هوش مصنوعی خاصی نیز نسبت داده شوند که در معماری و پیکربندی آن‌ها قابل تشخیص یا از رفتارشان قابل پیش‌بینی است. به طور خاص، می‌توان گفت که برخی سیستم‌های هوش مصنوعی، از حقایق مربوطه آگاهی دارند و انتخاب‌های عمدی انجام می‌دهند (Lagioia; Sartor, 2020: 448).

همیشه این احتمال وجود دارد که سیستم‌های هوش مصنوعی به دلایل غیرقابل پیش‌بینی خارج از وظایف برنامه‌ریزی‌شده خود عمل کنند. چنین اعمال ناخواسته‌ای ممکن است

به وسیله خطاهای نرم‌افزاری و نقص سیستم، سوگیری شناختی انسان، سوگیری الگوریتمی داخلی، تصادفات عادی که در آن هیچ خطای واضحی وجود ندارد، هک عمدی و جعل یا حملات سایبری ایجاد شود. واضح است که قصد در چنین شرایطی نمی‌تواند وجود داشته باشد. از سوی دیگر، می‌توان فرض کرد که تسلیحات مبتنی بر هوش مصنوعی باعث ایجاد عملی به دلیل غفلت یا قصور در زمینه جنایت جنگی شوند. با این حال، از آنجا که جنایت جنگی عمداً ارتکاب می‌یابد و شرط قصد و علم با غفلت همراه نخواهد بود، عملی که از این طریق ایجاد می‌شود جنایت جنگی محسوب نمی‌شود. به همین ترتیب، در اساسنامه عنصر اشتباه موضوعی به عنوان عامل دیگری که می‌تواند مسئولیت فردی را رفع یا محدود کند در نظر گرفته شده است. بر اساس بند ۱ ماده ۳۲ اساسنامه رم، اشتباه در موضوع موجب معافیت از مسئولیت کیفری نیست مگر اینکه عنصر معنوی جرم را از بین برده باشد. در این خصوص، اگر یک پهپاد خودمختار با قابلیت هوش مصنوعی بسیار توسعه یافته با حمله اشتباه به هدف زمینی، هواپیما یا کشتی یک کشور دیگر باعث آسیب شود، این اقدام به عنوان جنایت جنگی تلقی نخواهد شد زیرا قصد و آگاهی وجود ندارد.

## ۲-۲. شخصیت حقوقی یا الکترونیکی سیستم‌های هوش مصنوعی بر اساس

### قوانین کیفری

حتی اگر بتوان بر انسان محوری حقوق بین‌الملل کیفری غلبه کرد، مسئولیت فقط زمانی به هوش مصنوعی نسبت داده می‌شود که شخصیت حقوقی داشته باشند. از این رو، ایجاد یک شخصیت حقوقی مجزا و تحمیل مسئولیت مستقیم بر خود سیستم هوش مصنوعی ممکن است راه‌حلی ظریف و عملی برای برخی چالش‌ها باشد. اگر یک عامل هوش مصنوعی را نتوان یک شخص تلقی کرد، در اصل نمی‌تواند مانند انسان‌ها دارای حقوق و ملزم به تعهدات باشد (Bonfim, 2022: 49). سؤالی که مطرح است این است که آیا دولت‌ها نوعی شخصیت به سیستم‌های مبتنی بر هوش مصنوعی اعطا می‌کنند؟

هانس کلسن اظهار می‌دارد شخصیت در مفهوم حقوقی مجموعه‌ای از هنجارها، حقوق و تکالیف است (Kelsen, 2000: 178). شخصیت حقوقی شامل چیزی است که در

نظام‌های حقوق نوشته اهلیت قانونی یا صلاحیت حقوقی<sup>۱</sup> می‌نامند. اهلیت حقوقی صرفاً با شخص حقیقی ارتباط ندارد (Naučius, 2018: 114).

انسان بودن (اشخاص حقیقی) یا اداره شدن توسط انسان (اشخاص حقوقی)، احساس، ایفای نقش اجتماعی، داشتن منافع مستقل و زنده بودن از ویژگی‌های اشخاص دارای اهلیت هستند. به جز دو ویژگی نخست، در حال حاضر یا در آینده نزدیک این ویژگی‌ها در هوش مصنوعی یافت می‌شوند. چنین شخصیت حقوقی به این معنی است که روایات‌ها و عوامل نرم‌افزاری بسیار پیچیده مخاطب تعهدات قانونی و دارندگان حقوق قانونی هستند (Wendehorst, 2020: 155). این بحث با قطعنامه تصویب‌شده پارلمان اروپا در مورد قوانین مدنی در زمینه رباتیک تقویت شده است. این قطعنامه توصیه‌هایی به کمیسیون اروپا ارائه کرد که براساس آن، در درازمدت قانونگذاران باید به برخی از سیستم‌های هوش مصنوعی بسیار پیشرفته شخصیت حقوقی اعطا کنند (European Parliament Resolution, 2017: (C 252).

ایده اصلی این است که اگر انتساب آسیب‌های ناشی از هوش مصنوعی به انسان دشوار باشد، پس می‌توان خود هوش مصنوعی را مسئول دانست. همان‌طور که نفین استدلال می‌کند: «هر چیزی می‌تواند یک شخص حقوقی باشد، زیرا اشخاص حقوقی به این صورت تصریح شده‌اند یا وجود دارند» (Naffine, 2003: 351). طبق این درک از شخصیت، حیوانات، جنین، مردگان، محیط زیست، شرکت‌ها یا هر چیز دیگری که قانون در جامعه افراد می‌پذیرد، می‌توانند اشخاص حقوقی باشند. بسیاری از نهادهای غیرانسانی از جمله معابد در هندوستان، کل اکوسیستم در اکوادور نیز شخصیت حقوقی به آن‌ها اعطا شده است (Chesterman, 2020: 22). در نیوزلند تحولاتی در رابطه با اعطای شخصیت حقوقی به افراد غیرانسانی مانند زمین‌ها و رودخانه‌ها صورت گرفته است (Rousseau, 2016) و در زمینه حقوق حیوانات، شخصیت حقوقی محدود گونه‌های غیرانسانی به رسمیت شناخته شده است (Platvoet, 2020: 49). در سال ۲۰۱۵، یک قاضی در آرژانتین به یک اورانگوتان شخصیت حقوقی اعطا کرد (GUARDIAN, 2019). این

---

<sup>۱</sup>Legal capacity

تحولات نشان می‌دهد که با وجود اختلافات موجود، مفهوم حقوقی شخصیت اکنون در حال گسترش است. بنابراین، نسبت دادن شخصیت حقوقی به هوش مصنوعی با این روند مطابقت دارد.

این احتمال وجود دارد که سیستم‌های مبتنی بر هوش مصنوعی که در آینده حقیقتاً از انسان‌ها غیرقابل تمییز باشند روزی جایگاهی مشابه با افراد حقیقی کسب کنند. بسیاری از نویسندگان با توجه به ویژگی‌های هوش مصنوعی از جمله هوش، خودمختاری و تصمیم‌گیری مستقل، توانایی‌های شناختی پیشرفته، توانایی یادگیری از تجربه، حافظه، برنامه‌ریزی، پیچیدگی و قابلیت دستکاری ساختارها از ایده‌ی اعطای شخصیت حقوقی به هوش مصنوعی حمایت می‌کنند (Novelli et al, 2022: 200; Paulius et al, 2017: 685). هوش مصنوعی هم با حیوانات و هم با اشخاص حقوقی تفاوت اساسی دارد. یک ربات مانند حیوانات زنده نیست، اما مانند شرکت‌ها صرفاً یک فرض یا داستان خیالی صرف نیز نیست. با وجود این، می‌توان وجودشان را مستقل و بدون دخالت انسان‌ها تصور کرد. آن‌ها را با تفکر، هم از اشخاص حقوقی در وهله‌ی اول و هم از حیوانات در بعد دوم متمایز می‌کنند. سیستم‌های هوش مصنوعی اشیایی صرف نیستند که عملکرد آن‌ها تحت تأثیر دیگران باشد. به دلیل منحصربه‌فرد بودنشان با سایر الگوریتم‌های معمولی رایانه‌ها (برنامه‌ها) متفاوت هستند زیرا قادر به یادگیری مستقل، جمع‌آوری و تجزیه و تحلیل داده‌ها و ارائه‌ی راه‌حل‌های متفاوت براساس تجزیه و تحلیل موقعیت‌های مختلف مستقل از اراده‌ی توسعه‌دهندگان آن‌ها (برنامه‌نویس) عمل می‌کنند. به دلیل توانایی آن‌ها در تصمیم‌گیری مستقل، فناوری‌های مبتنی بر چنین سیستم‌هایی نمی‌توانند صرفاً به‌عنوان اشیا در نظر گرفته شوند.

خلاً در پاسخگویی و مسئولیت سیستم‌های مبتنی بر هوش مصنوعی با اعطای شخصیت حقوقی و ایجاد مسئولیت الکترونیکی برای ماشین‌ها می‌تواند رفع شود. علاوه بر این، مسئولیت الکترونیکی باعث می‌شود که مرتکبان جرائم با تکیه بر ویژگی‌های انسان‌محوری حقوق کیفری از مسئولیت فرار نکنند (Swart, 2023: 594). مهم است که از ایجاد انگیزه برای افراد به‌منظور پنهان شدن پشت عوامل خودمختار یا استفاده از این عوامل برای پنهان کردن هویت واقعی مجرمان جلوگیری شود. با این حال، باید توجه داشت که تحمیل

مسئولیت الکترونیکی به این معنی نیست که افرادی که ماشین را هدایت می‌کنند، نمی‌توانند یا نباید مسئول شناخته شوند.

اعطای شخصیت حقوقی به عوامل مبتنی بر هوش مصنوعی پیچیده است. از یک سو، وقتی به روبات‌ها یا سایر عوامل مصنوعی خودمختار شخصیت نسبت داده می‌شود، آن‌ها تابع حقوق می‌شوند (برخلاف اشیا) و وارد قلمرو اشخاص حقوقی می‌شوند. با وجود این، مفهوم شخصیت شرکتی یک تخیل حقوقی است، ابزاری آسان برای ساده‌تر کردن امور. از این رو، یک «جعبه ابزار» نظارتی برای هوش مصنوعی نیز ممکن است جایگزینی برای مقابله با کمبود شخصیت روبات‌های مستقل پیچیده باشد. تئوری شخصیت الکترونیکی، حقوق و مسئولیت‌های قانونی خاصی را برای عوامل هوش مصنوعی ایجاد می‌کند، اما قصد ندارد به روبات‌ها حقوق بشر اعطا کند یا با آن‌ها به عنوان یک انسان رفتار شود، و همچنین بهانه‌ای برای افراد نخواهد بود تا رایانه‌ها را برای اعمال خود مقصر بدانند (Abbott, 2020: 205). نظارت اتحادیه اروپا بر شخصیت حقوقی برای سیستم‌های خودمختار می‌تواند به مثابه نقطه شروعی برای تخصیص حقوق و مسئولیت‌ها به سیستم‌های مبتنی بر هوش مصنوعی عمل کند.

از یک سو برخی معتقدند که اعطای شخصیت و وضعیت حقوقی به روبات‌ها با برخی از حقوق بشر مانند حق بر کرامت، حق شهروندی و حق بر مزد در تضاد است (Osmani, 2020: 60). در مقابل، در حال حاضر، در نظریه حقوق کیفری، تنها عاملان انسانی از قوه تعقل برخوردارند که اساس مسئولیت کیفری را تشکیل می‌دهد. اشخاص حقوقی که تنها نمونه بارز گسترش مسئولیت کیفری فراتر از کنشگران انسانی هستند، همچنان بر عاملیت انسانی مبتنی هستند. علاوه بر این، در زمینه حقوق کیفری، شخصیت حقوقی ارتباط تنگاتنگی با سرزنش دارد، زیرا تنها فردی که می‌تواند حق را از نادرست تشخیص دهد و در موقعیت انتخاب قرار دارد، می‌تواند به دلیل انتخاب اشتباه سرزنش شود (Dafni, 2018: 684). سرزنش، توانایی درک این است که هر انتخابی مستلزم آن است و توانایی انتخاب آزادانه را پیش فرض می‌گیرد. در نهایت، آیا ممکن است در آینده شکلی از وجدان و حتی ظرفیت اهلیت اخلاقی و استدلالی را ایجاد کند که ممکن است به او اجازه دهد

که به اندازه عوامل انسانی مورد سرزنش قرار گیرد که در مورد اشخاص حقوقی یا حیوانات صدق نمی‌کند. اما تا زمانی که هم درک ما و هم عملی بودن سرزنش با خودآگاهی و تصمیمات آگاهانه که ریشه در تجربه انسانی دارد، همراه باشد، عوامل هوش مصنوعی نمی‌توانند مشارکت داشته باشند (Dafni, 2018: 687). از این منظر، هوش مصنوعی دارای شخصیت حقوقی نیست، زیرا فاقد بسیاری از ویژگی‌های انسانی است. در شناخت حق از باطل تشخیص ندارد، فاقد همدلی، دانش، احساس، و فاقد قصد یا عنصر عمد است. بنابراین، اثبات مسئولیت کیفری در زمانی که سلاح‌های مبتنی بر هوش مصنوعی مرتکب جرمی می‌شوند، بسیار دشوار است زیرا این تسلیحات رکن مادی جرم را مرتکب می‌شوند اما اثبات رکن معنوی جرم آن‌ها بسیار دشوار تلقی می‌شود. بنابراین، به نظر نمی‌رسد که اکثر نظام‌های حقوقی بتوانند به سیستم‌های مبتنی بر هوش مصنوعی شکلی از شخصیت را اعطا کنند. سیستم‌های مبتنی بر هوش مصنوعی تنها می‌توانند حقوق و تعهداتی داشته باشند که از سوی قانونگذاران به طور دقیق برای آن‌ها تعریف شده باشد.

## ۲-۳. انتساب وضعیت رزمنده به تسلیحات مبتنی بر هوش مصنوعی و مسئولیت

### کیفری مستقیم آن‌ها

برخی مسئول دانستن انسان را به دلیل نقض حقوق بشردوستانه و ارتکاب جنایت جنگی توسط تسلیحات مبتنی بر هوش مصنوعی، عادلانه نمی‌دانند (Gaeta, 2016: 44)، به همین دلیل و با توجه به تصمیم‌گیری مستقل این سلاح‌ها، از مسئولیت کیفری خود آن‌ها سخن می‌رانند (Roff, 2015) که می‌توان آن را الگوی مسئولیت مستقیم نامید. برخی نیز معتقدند که هیچ جایگزین عملی دیگری برای نحوه تعامل با این سیستم‌ها وجود نخواهد داشت، باید موضع رزمنده را برای آن‌ها اتخاذ کرد (Ohlin, 2016: 19).

هرچند تسلیحات مبتنی بر هوش مصنوعی در بیشتر موارد جایگزین انسان‌ها می‌شوند، با این حال، رزمنده یک اصطلاح حقوقی با مرزهای مشخص است که به صورت قانونی و با بهره‌مندی از امتیازاتی در مخاصمات مسلحانه بین‌المللی شرکت می‌کند. رزمنده در بند ۲

---

۱ وضعیت رزمنده دامنه کاربردی محدودی دارد و فقط در مخاصمات مسلحانه بین‌المللی قابل اجراست.

ماده ۴۳ پروتکل اول الحاقی به تمام اعضای نیروهای مسلح یک طرف محاصمه (به جز کارکنان بهداری و مذهبی) اطلاق شده است. هر چند قوانین داخلی برخی کشورها از جمله آلمان و آمریکا از واژه «شخص» به جای «اعضای نیروهای مسلح» استفاده کرده‌اند اما مقررۀ فوق صراحتاً به انسان اشاره نمی‌کند تا به عنوان رزمنده قلمداد شود. در عوض، معیاری را برای عضویت نیروهای مسلح در یک محاصمه مسلحانه تعیین می‌کند. اما تفاسیر کمیته بین‌المللی صلیب سرخ مفهوم اعضای نیروهای مسلح را از نظر ماهیت فقط به اشخاص محدود می‌داند (Sandoz et al., 1987: para.1672). بنابراین، بازیگران غیرانسانی به عنوان رزمنده در بند ۲ ماده ۴۳ قلمداد نمی‌شوند.

از آنجا که طرح چنین مسئولیتی نیازمند دو عنصر مادی و معنوی است و تاکنون سلاح‌های خودمختاری تولید نشده است که دارای عنصر دوم (رکن روانی) باشد. بنابراین، چنین تسلیحاتی از درک هرگونه مسئولیت در ارتباط با خود و درک یا اقدام به رفتار مؤثری در آینده ناتوان هستند. هنوز هم انسان در فرایند تصمیم‌گیری حضور دارد. بدون شک تسلیحات خودمختاری که فاقد عقل سلیم، درک یا شعور باشند، نمی‌توانند مسئول اقدامات خود شناخته شوند. بنابراین، تلقی هرگونه مسئولیت مجازات‌محور برای این تسلیحات، از آن جهت نادرست است که نمی‌توان آن‌ها را براساس مقررات طراحی شده برای انسان‌ها، مورد رسیدگی و مجازات قرار داد (رنجریان و بذار، ۱۳۹۷، ۷۷). بنابراین، اعطای وضعیت رزمنده به سیستم‌های تسلیحاتی نظامی مبتنی بر هوش مصنوعی بر پایه اخلاقیات و تفاوت‌های معنادار رفتاری انسان و ماشین حتی در هوشمندانه‌ترین و مستقل‌ترین شکل‌های آن‌ها، ناکافی و نامناسب است (Can Sati, 2024: 122). مفهوم وسایل و روش‌های جنگی در ماده ۳۶ پروتکل اول الحاقی به اندازه کافی گسترده است که بسیاری از برنامه‌های کاربردی مبتنی بر هوش مصنوعی را در زمینه نظامی تحت پوشش قرار دهد.

## ۲-۴. هوش مصنوعی کاملاً خودمختار به عنوان تابع حقوق کیفری

تابع حقوق، صلاحیت برخورداری از حقوق و برعهده گرفتن تعهدات (تکالیف مقرر در قانون) است. بحث در مورد تابع حقوق کیفری را نمی‌توان از بحث مسئولیت کیفری جدا کرد. در نظام حقوق بین‌الملل کیفری، افراد یا انسان‌ها تابع حقوق کیفری قلمداد می‌شوند.

این موضوع را از شماری از مواد اساسنامه رم که با عبارات «اشخاص»، «هرکسی»، «هر شخصی» و «افراد» شروع می‌شود، می‌توان استنباط کرد. اما در نظام‌های حقوقی داخلی، شرکت‌ها و اشخاص حقوقی نیز به‌عنوان تابع حقوق کیفری پذیرفته شده‌اند.

سیستم‌های مبتنی بر هوش مصنوعی تمامی عناصر اشخاص و نهادهای حقوقی شناخته شده را به‌عنوان تابع حقوق -از جمله هوش، تصمیم‌گیری مستقل، توانایی یادگیری از تجربیات خود، حافظه، برنامه‌ریزی، پیچیدگی و توانایی دستکاری ساختارها- دارا هستند. همچنین از شرایط ضروری برای برخورداری از شخصیت حقوقی -همچون اهلیت قانونی (دارا بودن حقوق و بر عهده گرفتن تعهدات) و اهلیت اقدام (اهلیت استیفاء) بهره‌مند هستند (Paulius et al, 2017: 697). اما صرف نظر از عملکرد استثنایی و منحصر به فرد این سیستم‌ها، هیچ‌یک از نظام‌های حقوقی تاکنون تابع حقوق بودن آن‌ها را به‌طور کلی به رسمیت نشناخته است.

در تئوری، برخی به دلیل مداخله فعال در روابط انسانی و تفاوت قابل توجه آن‌ها با سایر ابزارها، از امکان اعطای شخصیت حقوقی به سیستم‌های مجهز به هوش مصنوعی، به‌عنوان نهادهای حقوقی یا تابعان مصنوعی حقوق، صحبت کرده‌اند (Paulius et al, 2017: 699). اما در عمل، مسئله هوش مصنوعی به‌عنوان تابع حقوق را می‌توان در دادگاه فدرال استرالیا در قضیه دابوس (استفان تالر علیه اداره ثبت اختراع استرالیا) در ۳۰ ژوئیه ۲۰۲۱ مشاهده کرد. با تصمیم دادگاه فدرال استرالیا، هوش مصنوعی به نام «دابوس» که توسط تالر ایجاد شده بود چون قادر به تولید یا خلق اختراع (محصول) جدید بود، به‌عنوان مخترع یا تابع حقوق مطابق قانون ثبت اختراع استرالیا شناخته شد (Mahardhika et al, 2023: 3). در سطح بین‌المللی نیز، حقوق روبات‌ها در حال حاضر اهمیت یافته است. عربستان سعودی در ۲۵ اکتبر سال ۲۰۱۷ برای اولین بار به روباتی به نام «سوفیا»<sup>۱</sup> که ظاهر و رفتاری مشابه با انسان داشت، تابعیت اعطا کرد (Aulia Rahman; Habibulah, 2019, 150). کشور پیشرو در این زمینه -ژاپن- مسئله حقوق روبات‌ها را حل کرده است.

در ۷ نوامبر ۲۰۱۰ یک روبات درمانی به نام «پارو»<sup>۱</sup> مجوز اقامت ویژه<sup>۲</sup> که در جامعه ژاپن حقوق قانونی -از جمله خانواده، ملیت و شهروندی- را در برمی گیرد، دریافت کرد (Osmani, 2020: 60). در همین سال پارلمان اروپا قطعنامه‌ای تحت عنوان حقوق مدنی روبات‌ها تصویب کرد که در چارچوب آن، با اعطای وضعیت حقوقی خاص به روبات‌ها به مثابه اشخاص الکترونیکی<sup>۳</sup>، آن‌ها را مسئول هرگونه آسیب ناشی از تصمیمات مستقل و بدون دخالت انسان قلمداد می‌کند (European Parliament Resolution, 2017). پارلمان اروپا با تصویب قطعنامه فوق از کمیسیون خواست که در درازمدت ایجاد یک وضعیت حقوقی خاص برای روبات‌ها را بررسی کند تا حداقل پیچیده‌ترین روبات‌های خودمختار به عنوان افراد الکترونیکی شناسایی شده و مسئول جبران هرگونه خسارتی شوند که ایجاد می‌کنند (European Parliament, 2017: para 59(f)).

شرکت روسی سبربانک<sup>۴</sup> یک وکیل روبات راه‌اندازی کرده است که می‌تواند علیه افراد طرح دعوا کند. شرکت گلاواستراچ<sup>۵</sup> روباتی برای کمک به حل و فصل اختلافات مربوط به بیمه ایجاد کرد. در سال ۲۰۱۵ پارلمان روسیه پیش‌نویس قانون گرشین<sup>۶</sup> را تهیه کرد که در چارچوب آن قانون مدنی فدراسیون روسیه را اصلاح کرد و مسئولیت حقوقی را برای توسعه‌دهندگان، اپراتورها یا سازندگان روبات مقرر می‌کند و همچنین قواعد جدیدی وضع شد که موضوع نمایندگی روبات‌ها در دادگاه‌ها را تحت پوشش قرار می‌دهد (Mahardhika et al, 2023: 2).

## ۲-۵. مجازات سیستم‌های تسلیحاتی مبتنی بر هوش مصنوعی

---

<sup>۱</sup>Paro

<sup>۲</sup>Koseki

<sup>۳</sup>Electronic Person

<sup>۴</sup>Sberbank

<sup>۵</sup>Glavstrah Control

<sup>۶</sup>Grishin Act

مهم‌ترین ویژگی اعطای شخصیت حقوقی به سیستم‌های مبتنی بر هوش مصنوعی، قابلیت مجازات شدن است. با فرض دارا بودن شخصیت حقوقی و در نتیجه مسئولیت کیفری، پرسشی که مطرح می‌شود این است که دادگاه‌ها چگونه می‌توانند هوش مصنوعی را مجازات کنند؟ آیا آن‌ها می‌توانند سرزنش را تجربه کنند؟ سیستم عدالت کیفری بین‌المللی سنتی توسط انسان و برای انسان طراحی شده است. این امر سبب می‌شود که مجازات سیستم‌های مبتنی بر هوش مصنوعی نامناسب به نظر برسد. قوانین کیفری بین‌المللی با مجازات مجرمان به دنبال بازدارندگی خاص برای مجرم، بازدارندگی عمومی برای سایر افراد جامعه، تحقق عدالت، برقراری امنیت و نظم در جامعه بین‌المللی است. در این راستا، نقش حقوق بین‌الملل کیفری اصلاح متخلفان و ادغام مجدد آن‌ها در جامعه است. واضح است این امر در مورد سیستم‌های مبتنی بر هوش مصنوعی صادق نیست (Osmani, 2020: 61). برخی معتقدند مجازات هوش مصنوعی، مضحک و بی‌معنا به نظر می‌رسد (Acquaviva, 2022: 35). علاوه بر این، با فرض برخورداری از رکن معنوی، تعقیب و مجازات چنین تسلیحات خودمختاری کاری عبث است، چراکه مجازات آن‌ها نه تنها به احساس پشیمانی منتهی نمی‌گردد، بلکه حساسیت اخلاقی را نیز به منظور جلوگیری از ارتکاب مجدد جنایت در آن‌ها ایجاد نمی‌کند. بنابراین، آخرین نگرانی در باب انتساب مسئولیت کیفری به هوش مصنوعی، یافتن یک مجازات مناسب است.

اصل اساسی حقوق کیفری این است که یک شخص تنها در صورتی می‌تواند مجازات شود که بتواند مسئول ارتکاب جرم شناخته شود. تحمیل «مجازات» بر سیستم‌هایی که این قابلیت را ندارند به هیچ وجه به عنوان مجازات در نظر گرفته نمی‌شوند. همچنین نمی‌توان انتظار داشت که مجازات سیستم‌های مبتنی بر هوش مصنوعی دیگران را از ارتکاب جرم باز دارد (Weigend, 2023: 1145). همین ملاحظات در مورد تأثیر مجازات «انتقام‌جویانه» بر این سیستم‌ها نیز اعمال می‌شود. تحمیل مجازات به عنوان نوعی تلافی برای تخلف قبلی، پیش فرض می‌گیرد که سیستم‌های تسلیحاتی خودمختار می‌توانند پیامدهای تحمیل شده بر آن‌ها را با برخی تخلفات قبلی از سوی خود مرتبط کنند و بنابراین، «احساس مجازات» کنند. دستگاه‌های هوش مصنوعی امروزی فاقد این ویژگی کلیدی

مجازات پذیری هستند، این دستگاه‌ها نمی‌توانند (از نظر فکری و احساسی) ایده مجازات را به‌عنوان بیان‌کننده سرزنش اخلاقی درک کنند (Weigend, 2023: 1145). هرچند دانشمندان در حال تلاش برای توسعه و ایجاد هوش مصنوعی هستند تا احساس درد کند، اما تاکنون چنین سیستمی به وجود نیامده است تا مجازات بر آن اعمال شود (Ellamey; Elwakad, 2023: 98).

مولیگان به نفع مجازات هوش مصنوعی برای اهداف انتقام‌جویانه استدلال کرده است، زیرا انجام این کار می‌تواند مزایای روانی را برای قربانیانی که خواستار انتقام هستند، فراهم کند. او معتقد است که رضایت‌بخش‌ترین نتیجه برای قربانیان این است که به قربانی اجازه دهند تا از نظر فیزیکی به یک هوش مصنوعی آسیب برسانند یا آن را نابود کنند (Mulligan, 2018: 593). ابداع مجازات‌هایی که برای نشان دادن عدم تأیید جامعه از اقدامات سیستم‌های مبتنی بر هوش مصنوعی طراحی شده‌اند قابل تصور است. همان‌طور که ابوت و سارچ استدلال کرده‌اند ناتوانی یا از بین رفتن برای یک هوش مصنوعی «بد» است، همان‌گونه که خودداری از تغذیه یا ایجاد آسیب فیزیکی برای موجودات بیولوژیکی مانند گیاهان یا حیوانات «بد» است (Abbott & Sarch, 2019: 365). بنابراین، حتی اگر یک سیستم تسلیحاتی نتواند نارضایتی اخلاقی ناشی از تخریب یا غیرفعال شدن را احساس کند، معنای نمادین چنین عملی - پس از رسیدگی شبه‌کیفری مناسب - می‌تواند انسان‌هایی را که شاهد آن هستند تحت تأثیر قرار دهد و حتی ممکن است باعث رضایت قربانیان از اقدامات مضر سیستم‌های خودمختار شود.

از جمله مجازات‌های عملی برای سیستم‌های مبتنی بر هوش مصنوعی می‌توان به برنامه‌ریزی مجدد، تخریب و از بین بردن آن‌ها را اشاره کرد. با الهام از اثربخشی مجازات‌های بدنی در ناتوان‌سازی مجرمان انسانی، می‌توان معادل این اقدام را برای هوش مصنوعی، حذف نرم‌افزاری دانست که هسته کنترل‌کننده سیستم تسلیحاتی را از کار می‌اندازد. پس از اجرای حکم حذف، نهاد متخلف قادر به ارتکاب جرم دیگری نیست. این حذف و از بین بردن فیزیکی وجود مستقل سیستم تسلیحاتی را از بین می‌برد و برابر با مجازات اعدام است. در مورد مجازات حبس که مستلزم محرومیت یا محدودیت در حقوق

و آزادی‌های فرد است باید اظهار داشت که آزادی سیستم‌های مبتنی بر هوش مصنوعی شامل آزادی عمل به‌عنوان یک موجودیت مبتنی بر هوش مصنوعی در حوزه‌ی مربوطه است. با در نظر گرفتن ماهیت حکم حبس، اقدام عملی که ممکن است به همان تأثیرات حبس در هنگام تحمیل بر یک هوش مصنوعی دست یابد این است که نهاد مبتنی بر هوش مصنوعی را برای مدت معینی از حالت استفاده خارج کرد. در طول این دوره، هیچ‌گونه استفاده یا بهره‌برداری از آن نهاد مجاز نخواهد بود.

در صورت فقدان گزینه‌ی حبس، تشخیص «مسئولیت الکترونیکی» با اعمال جزای نقدی و پرداخت غرامت قابل مجازات است. دستور پرداخت جریمه از صندوق بیمه، معادل جزای نقدی است. اهمیت پرداخت جزای نقدی برای انسان، محرومیت از مقداری از مال است. با این حال، اکثر نهادهای مبتنی بر هوش مصنوعی هیچ پول، دارایی، یا حساب بانکی ندارند. اگر یک سیستم مبتنی بر هوش مصنوعی دارایی یا پول خود را داشته باشد، اعمال جزای نقدی بر آن با اعمال جریمه برای افراد یا شرکت‌ها یکسان خواهد بود. در نتیجه، جزای نقدی که هوش مصنوعی تحمیل می‌شود ممکن است به‌عنوان پول، دارایی یا نیروی کار به نفع جامعه جمع‌آوری شود. بنابراین، بیشتر مجازات‌های متداول برای سیستم‌های مبتنی بر هوش مصنوعی نیز قابل اعمال است. اعمال مجازات‌های خاص برای این سیستم‌ها مصنوعی، ماهیت این مجازات‌ها را در مقایسه با اعمال آن‌ها بر انسان‌ها نفی نمی‌کند.

### ۳. راه‌حل‌های جایگزین مسئولیت کیفری سیستم‌های خودمختار مبتنی بر هوش

#### مصنوعی

در صورتی که اعمال مجازات کیفری بر سیستم‌های مبتنی بر هوش مصنوعی منطقی نباشد، به نظر می‌رسد که خلأ مسئولیت نتیجه‌ی اجتناب‌ناپذیر استفاده از چنین تسلیحاتی باشد. سؤال دیگری که مطرح می‌شود این است آیا راه‌های دیگری برای رفع این خلأ وجود دارد؟ حقوق بین‌الملل بشردوستانه قابل تقلیل به مسئولیت کیفری فردی یا حتی مسئولیت اخلاقی نیست. برای درک این موضوع، ارتشی متشکل از کودکان را در نظر بگیرید. آن‌ها در حقوق بین‌الملل کیفری دارای مسئولیت کامل نیستند، و از آنجاکه اساسنامه رم

صلاحیت کیفری بر کودکان را مستثنی کرده، دیوان بین‌المللی کیفری صلاحیت رسیدگی به اعمال ارتكابی آن‌ها را ندارد. با این حال، در مخاصمه با ارتشی از کودکان اصول اصلی حقوق بشردوستانه همچنان حاکم خواهد بود و طرفین درگیری موظف به رعایت محدودیت‌های آن هستند. همین موضوع در مخاصمات مسلحانه علیه ارتشی از روایات‌ها یا سیستم‌های مبتنی بر هوش مصنوعی صدق می‌کند. این واقعیت که یکی از طرفین درگیری یک عامل مصنوعی است سایر رزمندگان را از درگیر شدن با چنین تسلیحاتی و پیروی از مقررات حقوق بین‌الملل بشردوستانه معاف نمی‌کند. بنابراین، یکی از راه‌حل‌های جایگزین مسئولیت کیفری هوش مصنوعی نظامی، ممنوعیت<sup>۱</sup> یا محدودیت تولید و کاربست سلاح‌های خودمختار مبتنی بر هوش مصنوعی در چارچوب حقوق بین‌الملل بشردوستانه است.

اگرچه در حال حاضر هیچ معاهده‌ای وجود ندارد که به‌طور خاص به استفاده از سلاح‌های خودمختار اشاره کند یا بر استفاده از آن‌ها نظارت داشته باشد اما اجماع جهانی وجود دارد که حقوق بشردوستانه بر همه سیستم‌های تسلیحاتی از جمله سلاح‌های مبتنی بر هوش مصنوعی اعمال می‌شود (Sehrawat, 2017: 38). التزام دولت‌ها در زمینه توجه به آثار مختلف ناشی از کاربرد تسلیحات خودمختار مبتنی بر هوش مصنوعی از منظر حقوق بین‌الملل بشردوستانه از تعهدات عام دولت‌ها برای تضمین اجرای حقوق بشردوستانه مقرر در ماده ۱ مشترک کنوانسیون‌های چهارگانه ژنو و ماده ۱ پروتکل اول الحاقی و سه نوع تعهد خاص آن‌ها یعنی اصل بشریت، تعهد به حفاظت از غیرنظامیان و محدودیت در به‌کارگیری هر نوع تسلیحات، روش و شیوه جنگی و همچنین تعهد به انجام بررسی‌های حقوقی سلاح‌های جدید در کلیه مراحل مطالعه، به‌کارگیری و تحصیل انواع تسلیحات مقرر در ماده ۳۶ پروتکل اول الحاقی ناشی می‌شود. تعهد به انجام اقدامات احتیاطی در

۱ فیزیکدان معروف استیون هاوکینگ اظهار می‌دارد: «توسعه هوش مصنوعی کامل می‌تواند پایانی برای نسل بشر باشد» (Hawking, 2018: 184). قانون هوش مصنوعی اتحادیه اروپا مصوب ۲ فوریه ۲۰۲۴ در ماده ۵ کاربردهای خاصی از هوش مصنوعی را ممنوع کرده است (Artificial Intelligence Act, 2024: Art.5). این ممنوعیت‌ها گامی مهم در جهت تضمین استفاده مسئولانه و اخلاقی از این فناوری است.

حمله (مواد ۵۱، ۵۷ و ۵۸ پروتکل اول الحاقی) و تعهد به کنترل صادرات از دیگر تعهدات حفاظتی دولت‌ها در استفاده از تسلیحات مبتنی بر هوش مصنوعی است (پیری، ۲۰۱۴: ۴). این تعهدات که برای اجتناب از خطرات غیرقابل پیش‌بینی تکنولوژی‌های نوظهور اهمیت ویژه‌ای دارند نه تنها منسوخ نشده‌اند، بلکه دولت‌ها را ملزم می‌کنند که به انواع آثار به‌کارگیری سلاح‌های مبتنی بر هوش مصنوعی توجه کنند.

شاید در نگاه اول، خودمختاری سیستم‌های هوش مصنوعی این تصور را ایجاد کند که نمی‌توان فرد دیگری را در قبال اعمال آن‌ها مسئول دانست. اما بدیهی است که پذیرش یک نرم‌افزار یا سخت‌افزار هوش مصنوعی به‌عنوان یک انسان و زندانی کردن آن یا تلاش برای رمزگذاری مجدد آن برای ایجاد بازدارندگی برای سایر سیستم‌های نرم‌افزاری منطقی نخواهد بود. بنابراین، باید یک عامل اخلاقی مسئول خسارات هوش مصنوعی باشد. از آنجاکه افراد و گروه‌های متعددی در توسعه، تولید یا کاربست هوش مصنوعی نظامی همکاری دارند، مسئول دانستن یک فرد خاص با استفاده از قوانین موجود در قبال رفتارهایی که منجر به جنایات بین‌المللی می‌شود، اگر نگوئیم غیرممکن اما دشوار است (McFarland & McCormack, 2014: 384). طراحان یا تولیدکنندگان، توسعه‌دهندگان، اپراتورها یا کاربران، برنامه‌نویسان، فرماندهان نظامی، رهبران شرکت‌ها و مقامات ارشد وزارت دفاع یا ترکیبی از این بازیگران ممکن است مسئول شناخته شوند. یک سیستم تسلیحاتی، صرف‌نظر از سطح پیچیدگی آن، یک جسم بی‌جان است. وجود و توانایی آن برای ایجاد آسیب ناشی از تلاش سازندگان سلاح است. توسعه‌دهندگان نه تنها بر دامنه اقداماتی که سیستم تسلیحاتی قادر به انجام آن است، بلکه بر اقدامات خاصی که در واقع پس از استقرار یا زمان استفاده واقعی از سیستم انجام می‌دهد، کنترل بیشتری خواهند داشت. بر این اساس، مسئله پاسخگویی توسعه‌دهندگان موضوعیت پیدا می‌کند. مسئولیت سازنده در مواردی است که سیستم‌های تسلیحاتی خودمختار عمداً برای نقض حقوق بین‌الملل بشر، حقوق بشردوستانه یا سایر قوانین مرتبط طراحی شده باشد (Casey- Maslen, 2014: 508). اگر سازنده عمداً سیستم تسلیحاتی طراحی کند که در ذات خود

غیرقانونی است و قادر به تفکیک میان غیرنظامیان و رزمندگان نباشد یا باعث رنج غیرضروری باشد (Chengeta, 2016: 40)، مسئول خواهد بود.

تصمیمات حیاتی در مورد رفتار و اثرات سلاح‌های مبتنی بر هوش مصنوعی در مرحله برنامه‌نویسی اتخاذ می‌شود؛ غیرممکن نیست که برنامه‌نویس نرم‌افزار را مسئول جنایات ارتكابی سلاح‌های مبتنی بر هوش مصنوعی بدانیم (Saria & Celik, 2021: 11). برنامه‌نویسان مطابق ماده ۳۶ پروتکل اول الحاقی، هنگامی که حقوق بشر دوستانه را در روند تحقیق، توسعه یا تملک تسلیحاتی مبتنی بر هوش مصنوعی برنامه‌ریزی می‌کنند. بنابراین، تدوین الگوریتمی که باعث نقض جدی و عمدی این تعهدات شود، یا برنامه‌نویسی که این تسلیحات را به گونه‌ای برنامه‌نویسی کرده باشد که ذاتاً قادر به تفکیک بین اهداف نظامی و غیرنظامی نباشد، مسلماً مسئولیت کیفری برنامه‌نویسان را به عنوان مجرم اصلی در پی خواهد داشت (رنجریان و بذار، ۱۳۹۷: ۷۸). برنامه‌نویسان این توانایی را دارند که با طراحی نرم‌افزاری مناسب از اجرا و عملکرد غیرقابل توجه این سیستم‌ها جلوگیری کنند. کاربر یا اپراتور با وجود اینکه کنترل مؤثری بر اقدامات سیستم‌های مبتنی بر هوش مصنوعی ندارد (Ford, 2017: 466; Beard, 2014: 652)، اما اگر از برنامه‌نویسی معیوب سیستم اطلاع داشته باشد و همچنان از سیستم خودمختار استفاده کند، از شرکای جنایت ارتكابی سیستم‌ها خواهد بود (Chengeta, 2016: 43). در واقع، اگر با یقین بتوان ثابت کرد که کاربر عمداً عمل کرده، از اجرای اعمال خود آگاه بوده است یا آگاه باشد که نتایج در جریان عادی رویدادها رخ خواهد داد یا اعمال خود را به منظور تسهیل جرم انجام می‌دهد می‌توان گفت که طبق مواد ۲۵ و ۳۰ (۲) الف و ب اساسنامه دیوان مسئولیت کیفری خواهد داشت. علاوه بر این، استدلال کمیته بین‌المللی صلیب سرخ مبنی بر اینکه وقتی یک اپراتور دستورالعملی را برای انجام عملی وارد سیستم‌های تسلیحاتی خودمختار می‌کند و منجر به جرم بین‌المللی می‌شود، پاسخگویی اپراتور تحت دکتترین مسئولیت فرماندهی مورد مناقشه نیست (U.N. Office, 2016). همچنین، رهبرانی که تصمیمات غیرمسئولانه می‌گیرند ممکن است مسئول شناخته شوند. فرماندهی که به طور هدفمند یک سیستم تسلیحاتی خودمختار را برای یک جنایت بین‌المللی به کار می‌گیرد،

دکترین مسئولیت فرماندهی به خوبی قابل اعمال است. بنابراین، بر اساس ماده ۲۸ اساسنامه رم و ماده ۸۶ پروتکل اول الحاقی، یک فرمانده نظامی در صورتی در قبال جنایات ارتكابی هوش مصنوعی نظامی مسئولیت کیفری دارد که اولاً کنترل واقعی و مؤثری بر این تسلیحات داشته باشد. ثانیاً اطلاعات و آگاهی کافی از برنامه‌نویسی این تسلیحات داشته باشد، ثالثاً انتساب عنصر مادی و معنوی جرم بر مبنای ماده ۲۵ و ۳۰ اساسنامه رم به فرمانده احراز شود (بصیری، ۱۳۹۸: ۹۹). اگر باوجود کنترل واقعی و آگاهی، اقدامات لازم و معقولانه‌ای که در حیطه قدرتشان بوده براساس ماده ۲۸ اساسنامه برای جلوگیری یا مجازات اعمال مجرمانه انجام نداده باشد، مسئولیت کیفری دارد (ابوذری، ۱۴۰۲، ۱۲۴). البته موضوع کنترل مؤثر نیز به سادگی در چارچوب سیستم‌های تسلیحاتی خودمختار تعیین نمی‌شود. همچنین دکترین مسئولیت مافوق یا فرماندهی با در نظر گرفتن رابطه بین دو انسان، یعنی فرمانده مافوق و زیردست، تدوین شده (Prosecutor v. Halilovid, 2005) و سیستم‌های مبتنی بر هوش مصنوعی را نمی‌توان به عنوان «نیروها/زیردستان» به مفهوم ماده ۲۸ اساسنامه رم دید.

در نهایت، از آنجا که سیستم‌های هوش مصنوعی قادر به برآوردن اصول حقوق کیفری موجود و الزامات عناصر واقعی و معنوی نیستند، برای غلبه بر این مشکلات، «مسئولیت بدون تقصیر» (محض) یکی از چارچوب‌های قانونی ممکن برای پاسخگویی هوش مصنوعی است. علی‌رغم اینکه تاکنون مقاومت زیادی در برابر مسئولیت ناشی از تقصیر به‌عنوان مبنایی برای مسئولیت کیفری وجود داشته است (Swart, 2023: 599)، شعبه مقدماتی دیوان بین‌المللی کیفری نیز در قضیه «بمبا» مفهوم مسئولیت بدون تقصیر را تأیید نمی‌کند چرا که «انتساب مسئولیت کیفری برای هر یک از جرائمی که در صلاحیت دیوان است منوط به وجود حالت ذهنی خاص یا درجه‌ای از تقصیر است» (Prosecutor v Bemba, 2009, 427). با این حال، مسئولیت مبتنی بر عدم تقصیر، بسیاری از مشکلات مرتبط با جست‌وجوی عوامل انسانی «پشت» عوامل خودمختار و پاسخگویی آن‌ها را حل می‌کند. رویکردی که طراحان، برنامه‌ریزان یا سازندگان هوش مصنوعی را مسئول می‌داند، لزوماً عادلانه نیست (King et al., 2020: 95). به همین دلیل، مسئولیت نباید

متکی بر علم یا قصد باشد زیرا ممکن است انگیزه‌ای برای عوامل انسانی ایجاد کند تا از یافتن آنچه که سیستم یادگیری ماشین دقیقاً انجام می‌دهد، اجتناب کنند. همچنین، پذیرش مسئولیت محض از یک سو به فقدان اطمینان حقوقی پایان می‌دهد (Wendehorst, 2020: 155). و از سوی دیگر، می‌تواند امکان توسعه بیشتر برنامه‌های سیستم‌های تسلیحاتی خودمختار را تضعیف کند، زیرا با توجه به مسئولیت کیفری احتمالی طراحان یا اپراتورها، از توسعه و کاربست آنها دلسرد می‌شوند.

### فرجام سخن

از مباحث مطروحه می‌توان دریافت که اگرچه فواید استفاده از هوش مصنوعی برای بشر انکارناپذیر است، اما نگرانی‌های جدی دربارهٔ مسئولیت کیفری ناشی از جنایات ارتكابی توسط هوش مصنوعی نظامی، موضوعی نیست که بتوان به سادگی از کنار آن گذشت. به نظر می‌رسد امکان رفع خلأ مسئولیت از طریق مسئول مستقیم دانستن خود سیستم‌های مبتنی بر هوش مصنوعی یا با تثبیت مسئولیت برای افراد، بدون در نظر گرفتن رفتار سزاوار سرزنش آنان، توسط اصول بنیادین حقوق کیفری مردود اعلام شده است. هر سیستم تسلیحاتی کاملاً خودمختار که بدون مداخله انسانی عمل کند، در صورت آسیب رساندن به انسان‌ها یا اموال، به این معنا نخواهد بود که مسئولیت انسان به طور کلی منتفی می‌شود. با استناد به مادهٔ ۸ اساسنامهٔ رم، آشکار است که یک سیستم مبتنی بر هوش مصنوعی، صرف نظر از سطح هوشمندی یا مرحلهٔ توسعهٔ آن، نمی‌تواند به خودی خود مسئول یک جنایت جنگی شناخته شود. سیستم‌های تسلیحاتی مجهز به پیشرفته‌ترین نرم‌افزارهای هوش مصنوعی، در چارچوب حقوق بین‌الملل، به ویژه اساسنامهٔ رم، فاقد مسئولیت کیفری هستند. این سیستم‌ها همچنین نمی‌توانند عمداً مرتکب جرم شوند، چراکه فاقد توانایی ذهنی لازم هستند و این بدان معناست که عاملان اصلی یک جنایت جنگی احتمالی، گردانندگان آنان خواهند بود. به بیان دیگر، توجیحات تحمیل مسئولیت کیفری بر سیستم‌های مبتنی بر هوش مصنوعی، کماکان بر محور انسان استوار است. حتی در صورت کنار گذاشتن این رویکرد در برخورد با جرائم هوش مصنوعی و نسبت دادن رکن معنوی به نهادهای هوش مصنوعی، از آنجا که این نهادها فاقد شخصیت حقوقی هستند، قابل تعقیب کیفری نخواهند بود.

در دوراهی عدم مسئولیت خود سیستم و مسئولیت اشخاص دخیل در تولید یا کاربست، تلاش برای ممنوعیت کلی توسعه و استفاده از سیستم‌های تسلیحاتی خودمختار خودآموز، چندان عاقلانه و منطقی به نظر نمی‌رسد؛ چراکه این سیستم‌ها در شرایط عادی، خطر کمتری برای غیرنظامیان و افراد بی‌گناه در مقایسه با سربازان انسانی مسلح به سلاح‌های سنتی ایجاد می‌کنند. در واقع، کارآمدترین پاسخ به مسئله مسئولیت، نظارت دقیق بر رفتار هوش مصنوعی به منظور جلوگیری از مشارکت آن در فعالیتهای غیرقانونی است. با توجه به آسیب‌های بالقوه‌ای که سلاح‌های مبتنی بر هوش مصنوعی می‌توانند بر افراد و اموال وارد کنند، بی‌تردید باید اقدامات جامعی برای پیشگیری از جنایات ناشی از این تسلیحات انجام پذیرد. نخست، دولت‌ها باید استفاده از سلاح‌های هوشمند را از طریق اعمال قوانین بین‌المللی موجود تنظیم کنند. دوم، باید قواعد حقوقی و اخلاقی، منطبق با اصول حقوق بشردوستانه و قواعد آمره مربوط به توسل به زور در حقوق بین‌الملل، به تمامی سیستم‌های مبتنی بر هوش مصنوعی آموزش داده شود. سوم، ایجاد یک چارچوب حقوقی برای این سیستم‌ها، از جمله یک کد اخلاقی برای سازندگان، ضروری است. درنهایت، ممکن است ایجاد یک مکانیسم ایمن برای از کاراندازی به‌عنوان اقدامی احتیاطی ضرورت داشته باشد تا در صورت لزوم، سیستم‌های هوش مصنوعی به‌طور ایمن غیرفعال شوند.

شایان ذکر است که حقوق کیفری ایجاب نمی‌کند که هر خسارتی لزوماً به مسئولیت کیفری (یا حتی مدنی) فردی بینجامد. برخی حوادث به‌عنوان «قوة قاهره» یا تحت تأثیر علل طبیعی رخ می‌دهند و گاهی انسان در وقوع آن‌ها دخیل است، اما به دلیل عدم توانایی در اجتناب از پیامدهای زیانبار، قابل سرزنش نیست. از این رو، باید بپذیریم که در هر مورد از جنایات مرتبط با تسلیحات مبتنی بر هوش مصنوعی، نمی‌توان مسئول (کیفری) مشخصی را تعیین کرد. هرگونه تلاش برای ایجاد یک دکترین حقوقی که خلأ مسئولیت این سیستم‌ها را یک‌بار برای همیشه برطرف کند، نه تنها به حقوق بین‌الملل کیفری آسیب می‌زند، بلکه استفاده از قوانین کیفری بین‌المللی به‌مثابه ابزاری برای رفع تمام تردیدهای جدی پیرامون هوش مصنوعی نظامی، انتظاری فراتر از ظرفیت آن است. دیوان بین‌المللی کیفری تنها در صورتی قادر خواهد بود جنایات بین‌المللی ارتکاب‌یافته توسط سیستم‌های

مبتنی بر هوش مصنوعی را تعقیب کند که از «مسئولیت بدون تقصیر» یا دیگر مدل‌های مسئولیت مبتنی بر تقصیر، که تاکنون در عدالت کیفری بین‌المللی مورد غفلت واقع شده‌اند، پیروی کند. برای محقق ساختن این امر، باید گام بلندی در جهت فاصله گرفتن از «تقصیر» به عنوان شرط محوری مسئولیت کیفری برداشت. اساسنامه رم از فلسفه‌ای تبعیت می‌کند که در نورنبرگ بیان شد: این انسان است، و نه نهادهای انتزاعی، که مرتکب جنایات علیه حقوق بین‌الملل می‌شود (I.M.T, 1946: 466). اکنون لازم است تا برای اصلاح برخی مقرراتی که پیش‌تر اساسی و حتی تخطی‌ناپذیر تلقی می‌شدند، گام برداریم. تغییر رویکرد فرد-محور حاکم بر حقوق بین‌الملل کیفری و پذیرش و ایجاد «مسئولیت الکترونیکی» ضروری می‌نماید. برای تحقق این شکل از مسئولیت، لازم است ماده ۱ و بند ۱ ماده ۲۵ اساسنامه رم اصلاح شود تا به طور صریح، صلاحیت شخصی دیوان را به «اشخاص حقوقی» گسترش دهد. اگر موجودیتی بدون شکل فیزیکی محسوس، مانند یک شرکت، می‌تواند دارای شخصیت حقوقی باشد، اعطای چنین شخصیتی به موجودی که وجود فیزیکی دارد، جهش مفهومی عظیمی محسوب نمی‌شود. بنابراین، باید مبانی اصرار بر محدود کردن شخصیت به «افراد» بار دیگر مورد بازبینی عمیق قرار گیرد.

هوش مصنوعی و توسعه شتابان آن در سال‌های آینده، بی‌تردید چالش‌های بزرگی را پیش روی حقوق کیفری قرار خواهد داد. اما نخستین و محوری‌ترین مسئله، همانا «مسئولیت کیفری» است. در این راستا، متخصصان باید مفاهیم بنیادین حقوق کیفری را بازتعریف و توسعه دهند. قانونگذاران موظفند مدل‌هایی ارائه کنند که به گونه‌ای کارآمد به تخصیص و تحمیل مسئولیت کیفری پردازد و قضات و دادرسان نیز باید این مدل‌ها را به بهترین شکل ممکن در عمل به کار گیرند. این اصلاحات حتی ممکن است روزی، ما را وادار به بازنگری در مبانی مسئولیت کیفری، اعمال متخلفانه، تصویب قوانین یا اصول جدید، یا پذیرش رویکردهای نوینی مانند «شخصیت الکترونیکی» کند. ممکن است زمانی چارچوب حقوق کیفری به طور کامل بر هوش مصنوعی قابل اعمال باشد. در آینده، اگر آبرهوش‌های مصنوعی پیچیده‌تر و مشابه انسان تولید گردند و به کار گرفته شوند، اعمال

مسئولیت کیفری بر آنان دور از ذهن نخواهد بود. با این حال، قابلیت‌های کنونی سیستم‌ها، معیارهای «آگاهی» و «اراده» مورد نیاز قانون کیفری را برآورده نمی‌سازد. اگرچه هوش مصنوعی در حال حاضر از لحاظ اهلیت فکری و عاطفی در سطح انسان‌ها قرار ندارد، اما ممکن است روزی به چنین جایگاهی دست یابد، همان‌گونه که آثار علمی - تخیلی تلاش داشته‌اند این هشدار را به ما بدهند. نگارنده این احتمال را رد نمی‌کند که اگر در آینده، سیستم‌های تسلیحاتی از انسان غیرقابل تشخیص شوند، بتوان تحمیل مسئولیت کیفری را، براساس منافی که برای خود این سیستم‌ها به ارمغان می‌آورد، توجیه کرد. بنابراین، اگر زمانی آن روز فرا برسد، ما باید آماده باشیم که مسئولیت کیفری را مستقیماً به آنان تسری دهیم و آنان را تا همان حد که قادر به انتخاب‌های آگاهانه اخلاقی و ارتکاب جرم هستند، به رسمیت بشناسیم. یک اصل حقوقی به ندرت یک شبه و به‌طور چشمگیر دگرگون می‌شود؛ بلکه در دوره‌ای طولانی تکامل می‌یابد و شاید این، بزرگ‌ترین چالش در مواجهه با سرعت سرسام‌آور توسعه فناوری‌های هوش مصنوعی باشد.

### تعارض منافع

تعارض منافع وجود ندارد.

### ORCID

Heidar Piri



<https://www.orcid.org/0000-0002-4302-7938>

### منابع

#### فارسی

- ابوذری، مهنوش؛ برزگر، محمدرضا، نادری، زهرا. (۱۴۰۲). «امکان‌سنجی مسئولیت کیفری سلاح‌های جنگی مبتنی بر هوش مصنوعی و مسئله بی‌کیفرمانی در دادگاه کیفری بین‌المللی»، [10.22133/mtlj.2023.389496.1184](https://doi.org/10.22133/mtlj.2023.389496.1184)، شماره ۸، دوره ۴، شماره ۸.
- بصیری، عباس. (۱۳۹۸). «مسئولیت‌پذیری در به‌کارگیری جنگ افزارهای رباتیک مستقل»، [10.22034/thdad.2019.240068](https://doi.org/10.22034/thdad.2019.240068)، شماره ۴.

پیری، حیدر. (۱۴۰۴). «استفاده نظامی از سیستم‌های تسلیحاتی خودمختار مبتنی بر هوش مصنوعی در مخاصمات مسلحانه؛ تعهدات حفاظتی دولت‌ها در چارچوب حقوق بین‌الملل بشردوستانه»، *دوفصلنامه حقوق بشر*، دوره ۱۹(۲). [10.22096/hr.2024.2019027.1645](https://doi.org/10.22096/hr.2024.2019027.1645)

پیری، حیدر. (۱۴۰۴). «مسئولیت بین‌المللی دولت‌ها در ارتباط با کاربرد سلاح‌های مبتنی بر هوش مصنوعی در درگیری‌های مسلحانه»، *حقوق فناوری‌های نوین*، شماره ۶(۱۲). [10.22133/mtlj.2025.449001.1353](https://doi.org/10.22133/mtlj.2025.449001.1353)

رنجبریان، امیر حسین؛ بذار، وحید. (۱۳۹۷). «رعایت حقوق بین‌الملل بشردوستانه از سوی روبات‌های نظامی خودفرمان و مسئولیت ناشی از اقدامات آن‌ها»، *مجله حقوقی بین‌المللی*، شماره ۵۹. [10.22066/cilamag.2018.31884](https://doi.org/10.22066/cilamag.2018.31884)

عرب‌چادگانی، رضا، و بهرام مرادیان. (۱۴۰۲). «مشروعیت به کارگیری سامانه‌های نظامی هوشمند در مخاصمات مسلحانه»، *فصلنامه مطالعات حقوق عمومی*، دوره ۵۳، شماره ۴. <https://doi.org/10.22059/jplsq.2021.326075.2807>

## References

### Books & Articles

Abbott, Ryan (2020), *The Reasonable Robot: Artificial Intelligence and the Law*, Cambridge University Press.

Abbott, Ryan; Sarch, A (2019), “Punishing Artificial Intelligence”, *UC Davis Law Review*, Vol.53.

Acquaviva, Guido (2022), “Autonomous Weapons Systems Controlled by Artificial Intelligence: A Conceptual Roadmap for International Criminal Responsibility”, *The Military Law and the Law of War Review*, Vol.60, Issue.1.

Albadi, Nuha; Kurdi, Maram & Shivakant Mishra (2019), *Hateful People or Hateful Bots? Detection and Characterization of Bots Spreading Religious Hatred in Arab Social Media*, *Proceedings of the ACM on Human-Computer Interaction*, Vol.3.

Beard, JM (2014), “Autonomous Weapons and Human Responsibilities”, *Georgetown Journal of International Law*, Vol.45(3).

- Bo, M; Bruun, L. & V. Boulanin (2022), “Retaining Human Responsibility in the Development and Use of Autonomous Weapon Systems: On Accountability for Violations of International Humanitarian Law Involving AWS”, Stockholm International Peace Research Institute.
- Bonfim, Tany Calixto (2022), *Criminal Liability of Artificial Intelligent Machines: Eyeing into AI's Mind*, Lund University.
- Casey-Maslen, Stuart (2014), *Weapons Under International Human Rights Law*, Cambridge University Press.
- Cerka, Paulius; Grigiene, Jurgita and Gintare Sirbikyte (2017), “Is It Possible to Grant Legal Personality to Artificial Intelligence Software Systems?”, *Computer Law & Security Review*, Vol.33(5).
- Chengeta, S (2016), “Accountability Gap. Autonomous Weapon Systems and modes of Responsibility in International Law”, *Denver Journal of International Law and Policy*, Vol.45.
- Chesterman, Simon (2020), “Artificial Intelligence and the Limit of Legal Personality”, *International and Comparative Law Quarterly*, Vol.69.
- Dafni, Lima (2018), “Could AI Agents be held Criminally Liable? Artificial Intelligence and the Challenges for Criminal Law”, *South Carolina Law Review*, Vol.69(3).
- Dimitrova, R (2022), “Criminal Liability Associated with Artificial Intelligence Entities under the Bulgarian Criminal Law”, In 2022 XXXI International Scientific Conference Electronics, IEEE
- Ellamey, Yasser; Elwakad, Amr (2023), “The Criminal Responsibility of Artificial Intelligence Systems: A Prospective Analytical Study”, *Corporate Law & Governance Review*, Vol.5(1).
- Endsley, M. R; Garland, D (2000), *Situation Awareness Analysis and Measurement*, CRC Press.
- Ford, CM (2017), “Autonomous Weapons and International Law”, *South Carolina Law Review*, Vol.69.

- Gordon, John-Stewart (2021), “Smart Technologies and Fundamental Rights”, Value Inquiry Book Series, Vol.350, Brill.
- Hallevy, Gabriel (2013), *When Robots Kill: Artificial Intelligence under Criminal Law*, UPNE.
- Hallevy, Gabriel (2015), *Liability for Crimes Involving Artificial Intelligence Systems*, Springer.
- Hallevy, Gabriel (2016), “The Criminal Liability of Artificial Intelligence Entities”, Akron Intellectual Property Journal, Vol.4(2), 2016.
- Hawking, Stephen (2018), *Brief Answers to the Big Questions*, John Murray.
- J. Bryson, M. Diamantis and T. Grant (2017), “Of, For, and By the People: The Legal Lacuna of Synthetic Persons”, Artificial Intelligence Law, Vol.25.
- Kelsen, Hans (2000), *Pure Theory of Law*, Lawbook Exchange.
- King, Thomas C., et al (2020), “Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions”, Science and Engineering Ethics, Vol.26(1).
- Lagioia, Francesca; Sartor, Giovanni (2020), “AI Systems Under Criminal Law: A Legal Analysis and a Regulatory Perspective”, Philosophy & Technology, Vol.33.
- Lior, Anat (2020), “AI Entities as AI Agents: Artificial Intelligence Liability and the AI Respondeat Superior Analogy”, Mitchell Hamline Law Review, Vol.46.
- M. Simmler and N. Markwalder (2019), ‘Guilty Robots? Rethinking the Nature of Culpability and Legal Personhood in an Age of Artificial Intelligence’, Criminal Law Forum, Vol.30.
- Mahardhika, Vita; Astuti, Pudji; Mustafa, Aminuddin (2023), “Could Artificial Intelligence be the Subject of Criminal Law?” Yustisia Jurnal Hukum, Vol.12, No.1.

- Malik, S (2018), “Autonomous Weapon Systems. The Possibility and Probability of Accountability”, *Wisconsin International Law Journal*, Vol.35.
- Mazzacuva, F (2021), “The Impact of AI on Corporate Criminal Liability”, *Revue Internationale de Droit Pe´nal*, Vol.92.
- McFarland/ McCormack (2014), Mind the gap. Can developers of Autonomous Weapon Systems be liable for war crimes?, *International Law Studies*, Vol.90.
- Mozur, Paul (2018), A Genocide Incited on Facebook, with Posts from Myanmar’s Military, *New York TIMES*,.
- Mulligan, Christina (2018), “Revenge Against Robots”, *S.C. Law Review*, Vol.69.
- Naučius, Mindaugas (2018), “Should Fully Autonomous Artificial Intelligence Be Granted Legal Capacity?” *Teises Apzvalga Law Review*, Vol.17(1).
- Ngaire, Naffine (2003), “Who are Law’s Persons? From Cheshire Cats to Responsible Subjects”, *Modern Law Review*, Vol.66.
- Novelli, Claudio; Giorgio Bongiovanni & Giovanni Sartor (2022), “A Conceptual Framework for Legal Personality and its Application to AI”, *Jurisprudence*, Vol.13.
- Ohlin Jens David (2016), “The Combatant’s Stance: Autonomous Weapons on the Battlefield”, *International Law Studies*, Vol.92.
- Osmani, Nora (2020), “The Complexity of Criminal Liability of AI Systems”, *Masaryk University Journal of Law and Technology*, Vol.14.
- Ozdemir, Gloria Shkurti (2019), *Artificial Intelligence Application in the Military: The Case of United States and China*, Istanbul: Seta.
- Pagallo, U (2011), “Killers, Fridges, and Slaves: A Legal Journey in Robotics”, *AI & Society*, Vol.26.

Platvoet, Veerle (2020), “The Attribution of Limited Legal Personality to Nonhuman Species”, *Journal of Animal Ethics*, Vol.10.

Prosperi, Luigi; Terrosi, Jacopo (2017), “Embracing the Human Factor”, *Journal of International Criminal Justice*, Vol.15(3).

Rousseau, Bryant (2016), In New Zealand, Lands and Rivers Can Be People (Legally Speaking), *N.Y. TIMES*.

Saria, Onur; Celik, Sener (2021), “Legal Evaluation of the Attacks Caused by Artificial Intelligence-based Lethal Weapon Systems within the Context of Rome Statute”, *Computer Law & Security Review*, Vol.42.

Sehrawat, Vivek (2017), “Autonomous Weapon System: Law of Armed Conflict (LOAC) And Other Legal Challenges”, *Computer Law & Security Review*, Vol.33(1).

Stanton, Gregory (2016), *The Ten Stages of Genocide*, Genocide Watch.

Swart, Bert (2009), “Modes of international Criminal Liability”, in *The Oxford Companion to International Criminal Justice*, Antonio Cassese(ed.), Oxford University Press.

Swart, Mia (2023), Constructing “Electronic Liability” for International Crimes: Transcending the Individual in International Criminal Law, *German Law Journal*, Vol.24.

Weigend, Thomas (2023), “Convicting Autonomous Weapons? Criminal Responsibility of and for AWS under International Law”, *Journal of International Criminal Justice*, Vol.21(5).

Wendehorst, Christiane (2020), *Strict Liability for AI and Other Emerging Technologies*, *J. EUR. Tort Law*.

Ying, Hu (2019), “Robot Criminals”, *University of Michigan Journal of Law Reform*, Vol.52.

## **Cases & Documents**

European Commission, 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 (2021).

Civil Law Rules on Robotics, European Parliament Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)).

France v. Goering, 22 I.M.T. 411, 466 (1946).

Human Rights Council, Report of the Detailed Findings of the Independent International Fact-Finding Mission on Myanmar, at 323, U.N. Doc. A/HRC/39/CRP.2 (Sept. 28, 2018)

International Criminal Court, Elements of Crimes art. 8 intro., U.N. Doc. PCNICC/2000/1/Add.2 (June 30, 2000).

Orangutan Sandra Granted Personhood Settles into New Florida Home, GUARDIAN, (Nov. 7, 2019),

U.N. Office, The C.C.W. Informal Meeting of Experts on L.A.W.S., The communication of the International Committee of the Red Cross to the Conference on Disarmament convened by United Nations 5 (Apr. 11-15, 2016).

Council of the European Union, Artificial Intelligence Act, Brussels, 2 February 2024.

Prosecutor v. Halilovid, IT-01-48-T T.Ch. 1, Judgment, T 61 (Nov. 16, 2005).

European Parliament Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics, Feb. 16, 2017, 2017 O.J. (C 252).

Prosecutor v Bemba, ICC-01/05-01/08, Pre-Trial Chamber II, Decision on the Confirmation of Charges, 15 June 2009.

### Translated References into English

Abouzari, M; Barzegar, M R & Naderi, Z (2023), “Feasibility of AI-powered war weapons criminal liability and the challenge of impunity in the International Criminal Court”, *Modern Technologies Law*, Vol.4, Issue.8, pp.119-134. [10.22133/mtlj.2023.389496.1184](https://doi.org/10.22133/mtlj.2023.389496.1184) [In Persian].

Basiri, A (2019), “Responsibility in the Use of Independent Robotic Weapons”, *Culmination of Law*, Vol.5, Issue.2, pp.73-106. [10.22034/thdad.2019.240068](https://doi.org/10.22034/thdad.2019.240068) [In Persian].

Chadegani, RA; Moradian, B (2024), “The Legitimacy of Using Intelligent Military Systems in Armed Conflict”, *Public Law Studies Quarterly*, Vol.53, Issue.4, pp.2033-2055. <https://doi.org/10.22059/jplsq.2021.326075.2807> [In Persian].

Piri, Heidar (2025), “Consequences of International Responsibility of States Regarding the Military Use of Artificial Intelligence in Armed Conflicts”, *Modern Technologies Law*, Vol.6, Issue.12, pp.169-193. [10.22133/mtlj.2025.449001.1353](https://doi.org/10.22133/mtlj.2025.449001.1353) [In Persian].

Piri, Heidar (2025), “The Military Use of Artificial Intelligence in Armed Conflict: States Protective Obligations Stemming from IHL”, *The Journal of Human Rights*, Vol.19 (2). [10.22096/hr.2024.2019027.1645](https://doi.org/10.22096/hr.2024.2019027.1645) [In Persian].

Ranjbarian, AH; Bazar, V (2018), “The Observance of International Humanitarian Law by Autonomous Military Robots and the Responsibility for Their Actions”, *International Law Review*, Vol.35, No.59, pp.63-84. [10.22066/cilamag.2018.31884](https://doi.org/10.22066/cilamag.2018.31884) [In Persian].

---

**استناد به این مقاله:** پیری، حیدر. (۱۴۰۴). مسئولیت کیفری سیستم‌های مبتنی بر هوش مصنوعی برای ارتکاب جرائم بین‌المللی و اعطای وضعیت رزمنده یا شخصیت الکترونیکی به آنها؛ ضرورت‌ها، موانع و راه‌حل‌ها. پژوهش حقوق کیفری، ۱۳ (۵۲)، ۱۹۶-۱۵۱.

Doi: 10.22054/jclr.2025.80888.2687