

<http://doi.org/10.22133/mtlj.2023.389496.1184>

## Feasibility of AI-powered war weapons criminal liability and the challenge of impunity in the International Criminal Court

Mehrnoosh Abouzari<sup>1</sup>, Mohammad Reza Barzegar<sup>2\*</sup>, Zahra Naderi<sup>3</sup>

<sup>1</sup> Assistant Professor, Faculty of Law and Political Science, University of Tehran, Tehran, Iran.

<sup>2</sup> Ph. D. Candidate in Criminal Law and Criminology, Faculty of Law and Political Science, University of Tehran, Tehran, Iran

<sup>3</sup> Ph.D. student in Criminal law and Criminology university of Tehran, Tehran, Iran

### Article Info

### Abstract

#### Original Article

**Received:**  
13-3-2023  
**Accepted:**  
23-05-2023

#### Keywords:

Artificial Intelligence  
Command  
Responsibility  
Autonomous  
Weapons  
Impunity

The issue of delegating the authority to kill humans to artificial intelligence in wartime is one of the most challenging emerging disciplines, as it raises numerous legal issues. This is the first time that legal knowledge has encountered a phenomenon that, despite being created by humans, is independent of humans in thinking and decision-making, and it may be possible after multiple stages that even the creator of the system cannot comprehend or control its operation. Therefore, this article discusses the obstacles that the International Criminal Court must overcome to apply direct criminal responsibility and the doctrine of commander's responsibility in cases where artificial intelligence weapons have led to the killing of civilians and war crimes, as well as the steps that should be taken to overcome these obstacles. Following a documentary approach and a descriptive-analytical analysis, this article concludes that the current provisions of the statute regarding artificial intelligence need to be revised due to the emergence of a concept called the gap in responsibility, which is caused by the unique nature of artificial intelligence. To adapt to the new conditions, the international community must require manufacturers to provide maximum transparency and prohibit producing certain types of artificial intelligence. Regarding the possible killings caused by the use of artificial intelligence weapons with the ability to self-learning after production, the court statute must explicitly and exclusively specify the rules of criminal responsibility so that the existence of AI weapons does not lead to impunity.

#### \*Corresponding author

**e-mail:** [Mrbarzegar@ut.ac.ir](mailto:Mrbarzegar@ut.ac.ir)

#### How to Cite:

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*, 4(8), 119-135.

Published by University of Science and Culture <https://www.usc.ac.ir>  
Online ISSN: 2783-3836



<http://doi.org/10.22133/mtlj.2023.389496.1184>

## امکان‌سنجی مسئولیت کیفری سلاح‌های جنگی مبتنی بر هوش مصنوعی و مسئله بی‌کیفرمانی در دادگاه کیفری بین‌المللی

مهرنوش ابوذری<sup>۱</sup>، محمدرضا برزگر<sup>۲</sup>، زهرا نادری<sup>۳</sup>

<sup>۱</sup> استادیار حقوق جزا و جرم‌شناسی، دانشکده حقوق و علوم سیاسی، دانشگاه تهران، تهران، ایران  
<sup>۲</sup> دانشجوی دکتری حقوق کیفری و جرم‌شناسی، دانشکده حقوق و علوم سیاسی، دانشگاه تهران، تهران، ایران  
<sup>۳</sup> دانشجوی دکتری حقوق جزا و جرم‌شناسی دانشگاه تهران، تهران، ایران

### چکیده

### اطلاعات مقاله

یکی از چالش‌برانگیزترین حوزه‌های نوظهور، بحث و آگاهی اختیاری کشتن انسان با هوش مصنوعی در شرایط جنگی است که به مباحث حقوقی متعددی می‌انجامد. این نخستین بار است که دانش حقوق با پدیده‌ای مواجه می‌شود که هرچند ساخته دست بشر است، اما برعکس تمامی پیشرفت‌های علمی - صنعتی تحقق یافته در طول تاریخ، در تفکر و تصمیم‌گیری مستقل از بشر است و ممکن است پس از طی مراحل حتی سازنده و خالق سیستم نیز قادر به درک چگونگی عملکرد یا مهار آن نباشد؛ از این رو، نوشته حاضر، به این بحث می‌پردازد که در شرایطی که اسلحه هوش مصنوعی منجر به قتل غیرنظامیان و جنایت جنگی شده، دیوان کیفری بین‌المللی برای اعمال مسئولیت کیفری مستقیم و همچنین دکتترین مسئولیت فرمانده با چه مسائلی مواجه است و برای تفوق بر این مشکلات، چه اقداماتی باید انجام شود. پژوهش حاضر، به روش اسنادی و با رویکردی توصیفی - تحلیلی پیش رفته و به این نتیجه رسیده است که با توجه به ظهور مفهومی به نام شکاف در مسئولیت، که ناشی از ذات منحصر به فرد هوش مصنوعی است، مقررات فعلی اساسنامه در خصوص آن نارسایی دارد و جامعه بین‌المللی باید برای تطبیق با شرایط جدید، سازندگان را ملزم به ارائه شفافیت حداکثری کند؛ همچنین تولید برخی از انواع هوش مصنوعی را ممنوع کند؛ و در نهایت درباره کشتار احتمالی ناشی از به‌کارگیری سلاح هوش مصنوعی با قابلیت خودبهبودی پس از تولید، اساسنامه دیوان را اصلاح کند و صراحتاً و انحصاراً قواعد مسئولیت کیفری را طوری مشخص کند که وجود سلاح هوش مصنوعی منجر به بی‌کیفرمانی نشود.

### مقاله پژوهشی

### تاریخ دریافت:

۱۴۰۱/۱۲/۲۲

### تاریخ پذیرش:

۱۴۰۲/۳/۰۲

### واژگان کلیدی:

هوش مصنوعی  
مسئولیت فرمانده  
اسلحه خودمختار  
بی‌کیفرمانی

\*نویسنده مسئول

رایانامه: [Mrbarzegar@ut.ac.ir](mailto:Mrbarzegar@ut.ac.ir)

### نحوه استناددهی:

ابوذری، مهرنوش، برزگر، محمدرضا، و نادری، زهرا (۱۴۰۲). امکان‌سنجی مسئولیت کیفری سلاح‌های جنگی مبتنی بر هوش مصنوعی و مسئله بی‌کیفرمانی در دادگاه کیفری بین‌المللی. حقوق فناوری‌های نوین، ۴(۸)، ۱۱۹-۱۳۵.

ناشر: دانشگاه علم و فرهنگ <https://www.usc.ac.ir>

شاپای الکترونیکی: ۲۷۸۳-۳۸۳۶

## مقدمه

یکی از تأثیرگذارترین اتفاقات دهه اخیر، پیشرفت چشمگیر در حوزه هوش مصنوعی بوده؛ چنان‌که این فناوری تمامی نقش‌های انسانی را برعهده گرفته است؛ شاعر، نویسنده، مترجم، راننده، سرباز ارتش، متصدی استخدام، خلبان و... این‌ها همه را بهتر از انسان متعارف انجام می‌دهد و همچنان در حال پیشرفت و تغییرات روزافزون است. در حال حاضر، گونه‌هایی از هوش مصنوعی در حال توسعه یا استفاده هستند که قادرند تصمیم بگیرند و بدون دخالت انسان عمل کنند. دستورالعمل‌هایی که اساس کار هوش مصنوعی است در طول فرایند تولید ثابت نیستند و می‌توانند در حین کار هم از طریق خود برنامه تغییر کنند.

برای اولین بار در طول تاریخ، دانش حقوق با پدیده جدیدی مواجه است که ممکن است پس از طی مراحل حتی سازنده و خالق سیستم قادر به درک چگونگی عملکرد یا مهار آن نباشد (ابوذری، ۱۴۰۰، ص ۲۲۸) و این آغاز مشکلی عظیم در علوم انسانی و به‌ویژه در حقوق کیفری است.<sup>۱</sup> درباره این‌که چرا این موضوع امروزه اهمیت بسیاری دارد، یکی از پیشگامان علوم رایانه بیان داشته است که «یادگیری ماشینی»<sup>۲</sup> و «یادگیری عمیق»<sup>۳</sup> در طول چند سال گذشته، بسیار سریع‌تر از آن چیزی پیشرفت کرده که پیش‌بینی می‌کردیم. این دلیلی است که چرا این مسئله امروز به موضوعی چنین بزرگ تبدیل شده است. اندیشمندان می‌بینند که حوزه فناوری در حال پیشرفت است و گام‌های بعدی به دغدغه فکری آن‌ها تبدیل می‌شود (صالح‌آبادی، ۱۳۹۸، ص ۵۱)؛ چرا که هوش مصنوعی افزون‌بر ایجاد صدمات بدنی، مرتکب طیف وسیعی از جرائم رایانه‌ای هنوز ناشناخته<sup>۴</sup> می‌شود (Mosechkin, 2019, p. 475).

حقوق‌دانان و قانون‌گذاران سراسر دنیا نیز از مسائل مرتبط با این فناوری آگاه‌اند و درباره آن هشدار داده‌اند. در نوامبر ۲۰۱۲، سازمان حقوق بشر در همکاری با کلینیک بین‌المللی حقوق بشر دانشکده حقوق هاروارد، گزارشی با عنوان «تلف‌شدن انسانیت: دعوایی علیه ربات‌های قاتل» منتشر کرده و بیان داشته است که سلاح خودمختار، نه‌فقط قادر نیست استانداردهای قانونی را برآورده کند، بلکه به دلیل فقدان محافظت مؤثر از غیرنظامیان، آن‌ها را به خطر می‌اندازد. بنابراین، چنین تسلیحاتی باید ممنوع شوند و دولت‌ها باید فوراً این موضوع را پیگیری کنند.<sup>۵</sup> به دلیل این گرایش، سازمان‌های بین‌المللی، مانند سازمان ملل متحد، تأثیر سامانه‌های هوش مصنوعی و رباتیکی مسلح در امنیت بین‌المللی و چارچوب قانونی موجود در مورد درگیری‌های مسلحانه را بررسی کرده‌اند (Barber, 2020, p. 7). کمیته بین‌المللی صلیب سرخ در اعلام موضع رسمی خود درباره سلاح هوش مصنوعی در سال ۲۰۲۱، اعلام کرد که باید به‌کارگیری سلاح خودمختار با رفتار پیش‌بینی‌ناپذیر و همچنین به‌کارگیری سلاح خودمختار برای هدف قراردادن هر موجود انسانی از جمله نظامیان ممنوع شود و در غیر این موارد، استفاده از سلاح خودمختار، باید با اعمال اقدامات احتیاطی مانند محدودیت زمانی و مکانی در به‌کارگیری سلاح خودمختار همراه شود و از همه مهم‌تر این‌که باید تمهیداتی اندیشیده شود که کنترل مؤثر انسانی بر سلاح خودمختار تضمین شود.<sup>۶</sup> اگرچه ممنوعیت استفاده راه‌حلی آسان و در دسترس است، اما مسئله مهم‌تر این است که چنانچه کشورها به‌رغم این توصیه‌ها از سلاح خودمختار استفاده کنند، آیا مقررات فعلی دیوان برای اعمال مسئولیت کیفری پاسخ‌گو هست یا خیر؟

<sup>۱</sup> تحلیل، بررسی و نظریه‌پردازی در خصوص جنایاتی که هوش مصنوعی در آن دخیل است، نه‌فقط از نظر حقوق کیفری داخلی اهمیت دارد، بلکه از منظر حقوق کیفری بین‌المللی نیز این موضوع در مواردی که کشورها در جنگ‌ها از سلاح مجهز به هوش مصنوعی استفاده می‌کنند مهم تلقی می‌شود.

<sup>۲</sup> Machine Learning

<sup>۳</sup> Deep Learning

<sup>۴</sup> هوش مصنوعی قادر است جرایم بسیاری انجام دهد؛ به دلیل این‌که برای برخی از این موارد هنوز معادل فارسی به‌وجود نیامده است از عبارات اصلی استفاده می‌شود:

Market bombing, Tricking face recognition, Learning-based cyber attacks, Bias exploitation, AI-authored fake reviews, AI-assisted stalking, Evading AI detection, AI-authored fake news

<sup>۵</sup> HUMAN RIGHTS WATCH, LOSING HUMANITY: THE CASE AGAINST KILLER ROBOTS 42 (2012), <https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots>

<sup>۶</sup> available at: <https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems>

نحوه عملکرد سیستم‌های تسلیحاتی خودمختار یعنی انتخاب مستقل و اعمال نیرو بر اهداف بدون دخالت انسان سؤالاتی را در مورد امکان عملی اعمال مسئولیت کیفری به دلیل پیامدهای استفاده از هوش مصنوعی، از جمله نقض قوانین بین‌المللی حقوق بشر یا به طور خاص کشتار غیرنظامیان مطرح می‌کند. اکنون پرسشی که مطرح می‌شود این است که اگر در یک مخاصمه مسلحانه غیرنظامیانی از طریق اسلحه هوش مصنوعی به قتل برسند، مقررات کنونی دیوان برای اعمال مسئولیت بر به‌کارگیرنده سلاح هوش مصنوعی با چه مشکلاتی مواجه است؟ متعاقباً ضمن پاسخ به این پرسش، پیشنهادهایی ارائه می‌شود تا در صورت استفاده از سلاح هوش مصنوعی در مخاصمات، مقررات دیوان به طور مقتضی بتواند پاسخ‌گوی کسانی که از این نوع سلاح استفاده می‌کنند باشد. در حالی که چندین کشور اطمینان خود را ابراز کرده‌اند که هرگز سلاح‌هایی را که نمی‌توانند مطابق با حقوق بین‌الملل حقوق بشر عمل کنند مستقر نخواهند کرد، اما تاریخ نشان می‌دهد که دولت‌ها ممکن است از این نظر کاملاً قابل اعتماد نباشند<sup>۱</sup> و همچنان تصمیم بگیرند که از سلاح‌های هوش مصنوعی استفاده کنند (Acquaviva, 2022, p. 110).

از این رو، با مسئله مهم جهانی روبه‌رو هستیم که نیازمند پاسخ‌های مقتضی پژوهشگران، اعم از حقوق‌دانان و فلاسفه اخلاق است؛ چراکه موضوع حیات انسان‌هاست و نقش حقوق کیفری در قبض و بسط استفاده از سلاح هوش مصنوعی از نقش خود دانشمندان این حوزه نیز مهم‌تر است؛ زیرا افراد زیادی در حال طراحی، ساخت و استفاده سیستم می‌توانند به طور قانونی و احتمالاً سامان‌مند از سرزنش اعمال خود شانه خالی کنند و این‌گونه با نبود خطر مجازات، دلیلی برای استفاده نکردن وجود ندارد. در این راستا و برای پاسخ به پرسش مقاله، در نوشته حاضر ابتدا مسئولیت کیفری استفاده‌کنندگان و به‌کارگیرندگان انواع سلاح‌های جنگی مبتنی بر هوش مصنوعی بر پایه قواعد و مقررات بین‌المللی موجود امکان‌سنجی می‌شود و سپس خلأها و ابهام‌های موجود در اساسنامه رم، که در این خصوص دیوان کیفری بین‌المللی را در قبال این نوع مرتکبان جنایات جنگی با مسئله بی‌کیفرمانی بزهکاران بین‌المللی مواجه می‌کند، تبیین می‌شود و در نهایت راهکارهای حقوقی برای برون‌رفت از این مشکل مطرح می‌شود.

در خصوص پیشینه پژوهش، نویسندگان، نوشتار یا پژوهشی را نیافتند که به مسئله بی‌کیفرمانی ناشی از خودفرمان بودن هوش مصنوعی پرداخته باشد، اما می‌توان پژوهش‌های ذیل را به‌منزله پیشینه پژوهش حاضر معرفی کرد. عزیزی بساطی و سکوتی (۱۳۹۴) در مقاله‌ای به این سؤال پرداخته‌اند که آیا استفاده از هوش مصنوعی در تسلیحات نظامی، صلح و امنیت بین‌المللی را به خطر می‌اندازد و به این نتیجه رسیده‌اند که استفاده از هوش مصنوعی در سلاح‌های نظامی، به تهدید و نقض صلح و امنیت بین‌المللی منجر می‌شود. سیفی (۱۳۹۹) در مقاله‌ای به مسئله مسئولیت کیفری سلاح‌های بدون سرنشین خودکار پرداخته است و به این نتیجه رسیده است که در صورت وقوع نقض قواعد بنیادین حقوق بشردوستانه، مسئولیت کیفری فردی مرتبط با این نوع وسایل محرز خواهد بود. موارد دیگری نیز وجود دارند که عبارت‌اند از «مقدمه‌ای بر چالش‌های هوش مصنوعی در حوزه مسئولیت مدنی» (تخشید، ۱۴۰۰)، «مسئولیت مدنی ناشی از تولید ربات‌های مبتنی بر هوش مصنوعی خودمختار» (واتقی و همکاران، ۱۳۹۸) و «ضمان در هوش مصنوعی» (رجبی، ۱۳۹۸) که با موضوع پژوهش حاضر متفاوت‌اند. در کتاب حقوق و هوش مصنوعی (ابوذری، ۱۴۰۱)، به طور کلی مباحثی در خصوص مبانی و قواعد شناخت مسئولیت در سیستم‌های هوش مصنوعی مطرح شده است، اما مطالبش با مقاله حاضر هم‌پوشانی ندارد.

علاوه بر موضوع پژوهش حاضر، آنچه این پژوهش را منحصر به فرد می‌کند نگاه ویژه و متمرکزی است که به بحث شکاف مسئولیتی برآمده از سرشت منحصر به فرد هوش مصنوعی می‌اندازد. در حالی که سازنده وسیله یا سلاح هوش مصنوعی پیش از به‌کارگیری آن، برنامه‌ها و دستورالعمل‌هایی را برای این تجهیزات مقرر می‌دارد، هوش مصنوعی با تجاربی که خود از محیط کسب می‌کند قادر است از دستورالعمل‌های

۱. دیپلماسی بین‌المللی هنوز درباره به‌کارگیری سلاح‌های خودمختار به اجماع نرسیده است. در سال ۲۰۱۸، کشورها در کنوانسیون ملل متحد در خصوص بررسی سلاح‌های متعارف‌گردد آمدند. چین زمانی که دیپلمات‌ها برای ممنوعیت جزئی استفاده از سلاح‌های خودمختار، نه توسعه و تولید آن‌ها، موضوع را مطرح کردند، آن را به تعویق انداخت. اما فرانسه و آلمان برای یک قطعنامه الزام‌آور سیاسی، که نظریاتشان را درباره تسلیحات خودمختار بیان می‌کند، استدلال آوردند. ایالات متحده و روسیه نیز با هرگونه تلاش نظارتی جدید مخالفت کردند و در عوض استدلال کردند که قوانین بین‌المللی موجود کافی است (Scharre, 2018, p. 365).

در حین راه‌اندازی فراتر رود و به منزله‌ی ماهیتی مستقل از سازنده، دست به اقداماتی بزند که برای سازنده هم پیش‌بینی‌ناپذیر است. همگام با آخرین پیشرفت‌های هوش مصنوعی، توجه این مقاله به این واقعیت اساسی، یعنی بررسی سلاح هوش مصنوعی به‌مثابه وجودی خودفرمان و نه ابزاری خودکار<sup>۱</sup>، وجهی است که این مقاله را نوآمد و منحصر به فرد می‌کند.

## ۱. مسئولیت کیفری استفاده‌کنندگان و به‌کارگیرندگان انواع سلاح‌های جنگی مبتنی بر هوش مصنوعی

هوش مصنوعی فتاوری سه‌کاره‌ای در نظر گرفته می‌شود که می‌توان از آن برای مقاصد غیرنظامی، نظامی و مجرمانه استفاده کرد (Larina & Ovchinsky, 2018, p. 374). سامانه‌های سلاح کاملاً خودکار باید از سامانه‌های نیمه‌خودکار، که در جنگ‌های معاصر متداول‌اند، متمایز شوند. در سامانه‌های نیمه‌خودکار، انسان‌ها هدف‌های خاص یا دسته‌های خاصی از هدف‌ها را انتخاب می‌کنند و سامانه برای شناسایی و حمله به آن‌ها دست‌به‌کار می‌شود. برای مثال، موشک «آتش و رها»<sup>۲</sup> در یک هواپیما در برابر هدفی که خلبان شناسایی کرده است قفل می‌شود و سپس بدون دخالت انسان حمله می‌کند (Schmitt, 2013, p. 5). این نوع سلاح‌ها موضوع پژوهش حاضر نیست؛ بلکه بر سلاح خودمختار، که خودش هدف را انتخاب می‌کند و به آن شلیک می‌کند، متمرکز است. با توجه به این موضوع، دیوان در راه برای اعمال مسئولیت کیفری در خصوص سلاح هوش مصنوعی دارد: یکی مبتنی بر مسئولیت مستقیم و دیگری مبتنی بر دکتترین مسئولیت فرمانده. در ادامه، طی دو بند مسئولیت کیفری براساس مسئولیت مستقیم (بند اول) و مسئولیت فرمانده (بند دوم) بررسی می‌شود.

### ۱-۱. مسئولیت مستقیم در خصوص به‌کارگیرنده هوش مصنوعی

چنانچه فرد یا افرادی مرتکب جنایت شوند، در صورتی که از هوش مصنوعی استفاده کرده باشند یا این‌که از نتایج ناشی از به‌کارگیری آن آگاهی کامل داشته باشند، در این موارد اعمال مسئولیت مستقیم گزینه مناسبی است. در این نوع مسئولیت، عناصر جرم - اعم از مادی و معنوی - به‌صراحت و مستقیم به شخصی که مسئول است استناد داده می‌شود؛ البته این بدان معنا نیست که شخص مسئول همواره باید بی‌واسطه کنش فنی و فیزیکی یا اصطلاحاً عنصر مادی جرم را مرتکب شود، بدین معناست که وی مسئولیت خویش را از دیگران (مثلاً خدمه یا زیردستانش همانند آنچه در مفهوم مسئولیت فرمانده به چشم می‌خورد) عاریه نمی‌گیرد؛ خواه ارتکاب با واسطه باشد یا بی‌واسطه. بر همین اساس، در اینجا مفهوم مسئولیت مستقیم، که در ماده (الف)(۳) ۲۵ اساسنامه رم مطرح شده است، بیان می‌شود.

آغاز قسمت (الف) از بند (۳) ماده ۲۵ اساسنامه درباره مباشرت مستقیم است؛ یعنی جایی که ارتکاب جرم به‌صورت فردی و بدون مشارکت یا واسطه دیگری بوده است. در چنین حالتی، به مرتکب، «عامل اصلی به‌تنهایی» گفته می‌شود (کسمتی، ۱۳۸۷، ص ۱۷۱). البته در این فرض، لزوماً فرد به‌تنهایی جرم را به سرانجام نمی‌رساند؛ بلکه ممکن است در این زمینه، افراد دیگری جایگاه فرعی و معاونتی داشته باشند. باین حال، فقط مباشر است که عناصر اصلی جرم را تحقق می‌بخشد و مسئول مستقیم شمرده می‌شود.

همچنین ماده (الف)(۳) ۲۵ بر ارتکاب جرم به شیوه مشترک با اشخاص دیگر نیز توجه می‌کند. نگرش اساسنامه به این نوع از مباشرت، که باز هم به مسئولیت مستقیم منجر می‌شود، نگرش به شرکا در جایگاه عاملان حیاتی وقوع جرم است؛ اگرچه این عاملان فقط جزئی از جرم را انجام داده باشند. به‌عبارت‌دیگر، اگر هریک از شرکت‌کنندگان در ارتکاب جنایت از طریق هوش مصنوعی، فقط مرتکب قسمتی از عملیات اجرایی جرم شوند، اما چنانچه آن‌ها آن اقدام مشخص را انجام نمی‌دادند و جرم تحقق نمی‌یافت، از شرکت‌کنندگان بی‌واسطه بودند و مسئول مستقیم کل عمل مجرمانه به‌شمار می‌رفتند. مبنای این اندیشه، توافق و طرح مشترکی است که مرتکبان را به هم پیوند می‌زند (عبداللهی و

<sup>۱</sup> در ابزار خودکار، تمامی اعمال منتسب به سازنده یا به‌کارگیرنده است و آن ابزار همچون وسیله‌ای در دست آن‌هاست که اقدامات مدنظر آن‌ها را، که از قبل به آن وسیله خودکار داده شده است، انجام می‌دهد. اما در سلاح هوش مصنوعی خودفرمان، اعمال با دقت و از قبل برای هوش مصنوعی تعریف نشده است. برای مطالعه بیشتر، به کتاب زیر رجوع شود:

Mueller, J. P., & Massaron, L. (2021). *Artificial intelligence for dummies*. New Jersey: John Wiley & Sons

<sup>۲</sup> Fire and Forget

همکاران، ۱۳۹۶، ص ۱۹۶)؛ به‌گونه‌ای که این طرح یا توافق مشترک، پایه اسناد متقابل اعمال به شرکت‌کنندگان شده است. در نتیجه هریک از آنان را مسئول کل جرم می‌شناسد.

چنانچه هوش مصنوعی قوی را عاملی مستقل بدانیم، اما کنترل انسانی بر آن چنان قوی باشد که به‌مثابه ابزاری در دست وی تلقی شود، می‌توان به قسمت اخیر ماده (الف)(۳) ۲۵(۳) استناد کرد که به ارتکاب جرم «از طریق شخصی دیگر» می‌پردازد. در این شکل از ارتکاب جرم، با آن‌که بحث از مسئولیت مستقیم به میان می‌آید، اما عنصر غالب عامل غیر مستقیمی است که از دیگری به‌شکل فیزیکی برای انجام جرم استفاده می‌کند. به دیگر سخن، کشگر مستقیم به‌مثابه ابزاری در دست عامل غیر مستقیم است (عبداللهی و همکاران، ۱۳۹۶، ص ۲۰۰؛ کسمتی، ۱۳۸۷، ص ۱۷۷) یا همان کسی که می‌توانیم از وی با عنوان مسئول مستقیم یاد کنیم. منطق شکل‌دهنده این رویکرد، به این دلیل «مرتکب پشت پرده»<sup>۱</sup> را مسئول مستقیم می‌شناسد که وی اراده مرتکب بی‌واسطه را کنترل می‌کند. بنابراین، مسئولیت مستقیم مباشر با واسطه، مستلزم وجود کنترلی قوی بر عامل مستقیم است تا وی را به حد ابزار تقلیل دهد و در این راستا، اقدامی بیش از ترغیب یا تحریک فرد به ارتکاب جرم ضرورت دارد.

در برخی موارد، قابلیت مجازات با مقررات کنونی دیوان، اما در غالب مسائل جدی، وجود دارد؛ چراکه هوش مصنوعی، به‌منزله فناوری‌ای پیشرفته، قادر است به‌صورت خودفرمان و بدون دخالت انسانی، تصمیم‌گیری و در بسیاری از زمینه‌های کاربردی عمل کند؛ اما در شرایطی که مربوط به جنگ است خودمختاری هوش مصنوعی با متغیرهای بسیاری مواجه است که باعث می‌شود پیش‌بینی حمله واقعی و نحوه پاسخ به آن بسیار دشوار باشد؛ برای مثال، در مواجهه با دشمن، متغیرهایی مانند موقعیت جغرافیایی، نوع و تعداد سلاح‌های دشمن، نحوه حرکت آن‌ها و همچنین شناخت عملکرد سلاح‌های خودمختار تأثیر بسیاری در تصمیم‌گیری هوش مصنوعی خواهند داشت. این متغیرهای بسیار و متنوع، به همراه نامعلوم بودن نیازهای واقعی در زمان حمله، باعث می‌شوند که پیش‌بینی حمله در شرایط جنگ، که شامل متغیرهای بسیار است، پیچیده و دشوار باشد. درنهایت، با توجه به دانش اندک از خودمختاری، رباتیک یا برنامه‌نویسی پیچیده، رفتار یک سلاح خودمختار به‌قدری پیش‌بینی‌ناپذیر است که نتوان رکن روانی مذکور در ماده ۳۰ را به‌کاربر نسبت داد و در این صورت، مجازات کاربر با مشکلی جدی مواجه است.

## ۱-۲. اعمال دکتین مسئولیت مافوق درخصوص به‌کارگیرنده هوش مصنوعی

در حقوق کیفری بین‌المللی، مسئولیت غیر مستقیم - که به‌منزله مسئولیت فرماندهی نیز شناخته می‌شود - آن است که یک فرمانده نظامی یا مافوق غیرنظامی به دلیل کوتاهی در جلوگیری از اقدامات زبردستان یا مجازات‌کردن ایشان از لحاظ کیفری مسئول شناخته می‌شود.<sup>۲</sup> به‌طور خاص، مسئولیت در چارچوب مسئولیت فرماندهی زمانی رخ می‌دهد که مافوق اقدامات لازم و معقولانه‌ای برای جلوگیری یا مجازات اعمال مجرمانه زبردستانی که مافوق کنترل مؤثری در آن‌ها دارد انجام ندهد؛ آن‌هم در صورتی که مافوق از اعمال مجرمانه آن‌ها اطلاع داشته باشد.<sup>۳</sup> مسئولیت فرماندهی مافوق را در قبال ترک وظیفه و ترک فعل مجرمانه مسئول می‌داند.<sup>۴</sup> اساسنامه دیوان علاوه بر مسئولیت کیفری فرمانده نظامی، مسئولیت کیفری شخصی که عملاً وظایف فرمانده نظامی را انجام می‌دهد نیز پذیرفته است.

اگر بپذیریم که هوش مصنوعی، با توجه به تصمیم‌گیری مستقل پس از تولید، فاعلی مستقل و نه عاملی ابزاری به‌شمار می‌رود و با توجه به پیش‌بینی‌ناپذیر بودن آن، ضوابط مرتبط با مسئولیت کیفری مستقیم برای به‌کارگیرنده فراهم نمی‌شود و از این مجرا، دیوان نمی‌تواند در قبال

1. Perpetrator Behind Perpetrator

۲. از لحاظ تاریخی، «مسئولیت فرماندهی» به فرماندهان نظامی محدود می‌شد، اما از زمان جنگ جهانی دوم این دکتین برای مافوق غیرنظامی نیز اعمال شد و برخی منابع را به استفاده از اصطلاح فراگیرتر «مسئولیت مافوق» سوق داد. «مسئولیت فرماندهی» و «مسئولیت مافوق» امروزه گاهی به‌جای هم استفاده می‌شوند.

3. Prosecutor v. Zdravko Mucic aka "Pavo", Hazim Delic, Esad Landzo aka "Zenga", Zejnil Delalic (Trial Judgment), IT-96-21-T, International Criminal Tribunal for the former Yugoslavia (ICTY), 16 November 1998, available at: <https://www.refworld.org/cases/ICTY,41482bde4.html> [accessed 25 June 2023]

4. Prosecutor v. Halilovic (Trial Judgment), IT-01-48-T, International Criminal Tribunal for the former Yugoslavia (ICTY), 16 November 2005, available at: <https://www.refworld.org/cases/ICTY,48ac3a6c2.html> [accessed 25 June 2023]

به‌کارگیرنده مسئولیت کیفری اعمال کند، لاجرم راه دیگر اعمال مسئولیت، مسئولیت کیفری غیرمستقیم براساس ماده ۲۸ اساسنامه است. برای این کار، باید در مجموع شرایط ذیل فراهم باشد:

(۱) هوش مصنوعی تحت اقتدار و کنترل واقعی به‌کارگیرنده باشد و شخص مذکور با توجه به اوضاع می‌دانسته یا باید می‌دانسته که هوش مصنوعی تحت امر او، جرم مذکور را مرتکب شده یا قریباً مرتکب خواهد شد و اقدامات لازم و معقولی را که در توانش بوده برای جلوگیری از جرم یا منع از انجام آن اعمال نکرده باشد.

(۲) هوش مصنوعی در خصوص مصادیق بسیاری می‌تواند قوانین و قواعد کیفری کنونی را با مشکل مواجه کند. در خصوص استفاده از دکترین مسئولیت فرمانده برای اعمال مسئولیت کیفری بر به‌کارگیرنده هوش مصنوعی، در میان حقوق‌دانان اختلاف نظر وجود دارد؛ زیرا برخی معتقدند زیردست مدنظر ماده ۲۸ اساسنامه صرفاً موجود انسانی است و نمی‌توان این ماده را برای زیردست غیرانسانی جاری دانست. همچنین برخی مخالفانند که مسئولیت فرمانده منحصر به زیردست انسانی شود و معتقدند با یک منطق حقوقی پویا می‌توان مفاهیم و محصولات جدید را بر قواعد و قوانین از پیش موجود تطبیق داد. به نظر می‌رسد در مواردی که محصول یا سلاح صرفاً ابزاری محض در دست فرمانده باشد، استناد به دکترین فرمانده قابل اعتنا نیست و قطعاً اگر مسئولیتی وجود داشته باشد، براساس قسمت اخیر ماده (الف)(۳) ۲۵ اساسنامه، مسئولیت مستقیم است؛ اما در مواردی مانند هوش مصنوعی - که صرفاً ابزار نیست و رفتارش به طور کامل خواست به‌کارگیرنده نیست و چه بسا به‌کارگیرنده حتی درک نکند که چرا هوش مصنوعی این رفتار را انجام داده - می‌توان به مسئولیت غیرمستقیم استناد کرد، اما حتی اگر این نظر را بپذیریم که مسئولیت فرمانده منحصر به زیردست انسانی نیست، باز هم قواعد کنونی دیوان دربارهٔ مقابله و مجازات به‌کارگیرندگان هوش مصنوعی کافی نیست.

یکی از مواردی که می‌تواند به حقوق‌دانان در درک و تصمیم‌گیری براساس دکترین فرمانده کمک کند، بررسی رویه‌های سابق است؛ اگر چه مطابق بند ۲ ماده ۲۱ اساسنامه دیوان، تصمیمات قبلی و آرای دیگر محاکم الزام‌آور نیستند، اما تصمیم‌های دادگاه‌های موقتی در این باره نقش مهمی در تفسیر و اعمال مقررات اساسنامه ایفا می‌کنند. در همین راستا، شعبهٔ رسیدگی به پروندهٔ سلیمی در یوگسلاوی شرایط مسئولیت فرماندهان در حقوق عرفی را تبیین کرده است. این شرایط عبارت‌اند از: وجود رابطهٔ مافوق - مادون و همچنین عنصر روانی و ناتوانی در اتخاذ تصمیم مناسب برای پیشگیری یا مجازات زبردستان (عبداللهی و همکاران، ۱۳۹۶، ص ۲۲۰). دکترین مسئولیت مافوق فرض می‌کند که مافوق انسانی از طرح‌ها یا اقدامات مجرمانهٔ زبردستان خود آگاه است یا دللیلی برای اطلاع از آن دارد. اگر اسلحهٔ هوش مصنوعی در اجرای وظایفش پیش‌بینی ناپذیر باشد، به سختی می‌توان مسئولی را پیدا کرد که مقصر آسیب خاصی باشد و در نتیجه چنین امر پیش‌بینی ناپذیری رخ می‌دهد؛ دست‌کم در بیشتر مواردی که یک الگو قابل مشاهده نیست (Acquaviva, 2022, p. 112). بنابراین موانع مهمی برای پاسخ‌گویی در قبال اعمال مجرمانهٔ ارتكابی سلاح‌های کاملاً خودکار در چارچوب دکترین مسئولیت فرماندهی وجود خواهد داشت. برخی حقوق‌دانان، در دیدگاهی پویا و با ارائهٔ تحلیلی منسجم و البته طولانی بر این باورند که هوش مصنوعی همانند انسان قادر است رکن روانی مدنظر برای جرایم را داشته باشد (هالوی، ۱۳۹۸، ص ۱۰۷-۱۰۸). در مقابل، برخی معتقدند با توجه به این که ربات‌ها قادر نیستند رکن معنوی مورد نیاز یک جرم را برآورده سازند، بنابراین آن‌ها هیچ‌گاه نمی‌توانند مرتکب جرم می‌شوند و وقتی آن‌ها نتوانند مرتکب جرم شوند، در نتیجه، نمی‌توانند در نقش مأموران زبردست مدنظر دیوان تلقی شوند تا براین اساس، فرماندهان در چارچوب مسئولیت فرماندهی مسئول شناخته شوند. بنابراین، این دکترین زمانی که زیردست یک اسلحهٔ هوش مصنوعی است و با هوش مصنوعی کنترل می‌شود اعمال‌شدنی نیست (Acquaviva, 2022, p. 112). مانع دیگر برای اعمال مسئولیت فرماندهی این است که مطابق ماده ۲۸ اساسنامه، فرمانده یا آمر با هدف مجازات زیردست، به منظور تحقیق و تعقیب نزد مقامات صلاحیت‌دار طرح مسئله نکرده باشد. اعمال مجازات در خصوص هوش مصنوعی معنایی ندارد و همین اصل، مسئولیت فرماندهی در خصوص اعمال مجرمانه را در موارد استفاده از سلاح‌های کاملاً خودمختار، خارج از دایرهٔ شمول موضوع قرار می‌دهد.

## ۲. سلاح هوش مصنوعی و ایجاد خلأها و ابهام در اجرای اساسنامه رم

سلاح هوش مصنوعی به صورت خودکار و بدون نیاز به دخالت انسانی، تصمیم‌هایی را اتخاذ می‌کند؛ قابلیت انطباق با تغییرات محیطی را دارد؛ با استفاده از الگوریتم‌های یادگیری قادر به بهبود عملکرد خود از طریق تجربیات گذشته است و از همه مهم‌تر، به دلیل پیچیدگی و شفاف نبودن، قادر نیست دقیقاً توضیح دهد که چگونه به تصمیمات اتخاذی خود رسیده است. همه این موارد نوظهور به مسائلی در اجرای قوانین قبلی منجر می‌شود که در اینجا این موارد طی دو دسته‌بندی کلی بررسی می‌شوند.

### ۲-۱. سلاح هوش مصنوعی؛ مسئله‌ای برای مسئولیت مستقیم

چنان‌که در ماده ۲۵ اساسنامه رم آمده است، دکترین مسئولیت مستقیم افراد را به خاطر نقش مؤثر مستقیم در ارتکاب جرم مسئول می‌داند؛ خواه این نقش مؤثر مستقیم به شکل بی‌واسطه و خواه با واسطه اعمال شود. از آنجاکه دیوان دو راه برای اعمال مسئولیت کیفری دارد (یکی مبتنی بر مسئولیت مستقیم و دیگری مبتنی بر دکترین مسئولیت فرمانده)، باید بررسی شود که چرا درباره سلاح هوش مصنوعی مقررات فعلی در برخی موارد کافی نیستند. یکی از مواردی که باید به منظور اعمال مسئولیت کیفری مستقیم مدنظر قرار گیرد، تحقق رکن روانی مصرح در ماده ۳۰ اساسنامه است. مطابق ماده مذکور، هیچ‌کس مسئولیت کیفری ندارد 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 باشد. مطابق این ماده، وجود قصد مفروض است، موقعی که شخص نسبت به اصل عمل، قصد ارتکاب آن را داشته باشد، نسبت به نتیجه عمل، قصد ایجاد نتیجه را داشته باشد یا آگاه باشد که این نتیجه در مسیر طبیعی حوادث پیش خواهد آمد.

به طور سنتی در اکثر نظام‌های حقوقی، «قصد» نه فقط جنبه «ارادی»<sup>۱</sup> (جایی که نتیجه مطلوب است)، بلکه جنبه «شناختی»<sup>۲</sup> (جایی که این آگاهی وجود دارد که نتیجه به وقوع می‌پیوندد یا موجب به وقوع پیوستن می‌شود) را نیز دربرمی‌گیرد. این مفهوم شناختی قصد در چندین نظام حقوقی، با عباراتی همچون قصد غیر مستقیم<sup>۳</sup>، قصد ضمنی<sup>۴</sup> یا عبارات دیگر<sup>۵</sup> شناخته می‌شود. این نگرش، سؤال مهمی را در خصوص این‌که یک پیامد چقدر باید قطعی یا محتمل باشد تا در سنجه مدنظر ماده ۳۰ اساسنامه (روی دادن در جریان عادی وقایع) بگنجد، مطرح می‌کند. رویه قضایی دیدگاه‌های متفاوتی درباره درجه احتمال لازم ارائه می‌دهد. در یک سر طیف، چنین می‌نمایاند که قصد غیر مستقیم شامل پیامدهایی است که نزدیک به قطعیت اند و اغلب با عباراتی مانند «تقریباً قطعی»<sup>۶</sup>، «عملاً قطعی»<sup>۷</sup> یا «اساساً قطعی»<sup>۸</sup> توصیف می‌شوند.<sup>۹</sup> در آن سوی طیف، چنین می‌نمایاند که خطرات یا احتمالات بعید معیار «روی دادن» را برآورده نمی‌کند. در این باره، عمدتاً تأکید می‌شود که مفاهیم «بی‌پروایی»<sup>۱۰</sup> و «قصد احتمالی»<sup>۱۱</sup>، که احتمالات صرف را دربر می‌گیرد، در اساسنامه گنجانده نشده است<sup>۱۲</sup>؛ بنابراین اعمال مسئولیت کیفری توسط دیوان برای به کارگیرنده هوش مصنوعی مبتنی بر این مفاهیم حقوقی ناممکن است؛ زیرا با توجه به پیچیدگی و تعداد بالای پارامترهایی که می‌تواند رفتار هوش مصنوعی را تحت تأثیر قرار دهد، توانایی پیش‌بینی رفتار هوش مصنوعی در بسیاری از حالت‌ها محدود

1. Volitional

2. Cognitive

3. Indirect Intent

4. Oblique Intent

5. Dolus Indirectus or Dolus Directus in the Second Degree

6. Virtually Certain

7. Practically Certain

8. Substantially Certain

9. Bemba (PTC II Decision), para. 362.

10. Recklessness

11. dolus eventualis

12. Ambos (2009). 22 LeidenJIL 715, 718; The Rome Statute of the International Criminal Court: A Commentary (2002) 902, 915-916.; Lubanga (TC I Judgment), para. 1011.



است. از این رو، ممکن است در اینجا خلأ ایجاد شود؛ زیرا در واقع یکی از ویژگی‌های توصیف‌شده اسلحه هوش مصنوعی این است که توانایی یادگیری و محاسبه داده‌های ورودی‌های مختلف در راستای پارامترهای تنظیم‌شده اولیه را دارد و حتی در برخی موارد، چنین پارامترهایی را اصلاح می‌کند. این امر باعث می‌شود که پیامدهای چنین فعل و انفعالی برای انسان‌ها درک‌پذیر نباشد؛ چه رسد به پیش‌بینی آن‌ها. موردی واقعی که به فهم موضوع کمک می‌کند، موضوع ربات گفت‌وگوی آنلاین مبتنی بر یادگیری تقویتی است. در سال ۲۰۱۶، شرکت مایکروسافت جدیدترین محصول هوش مصنوعی خود به نام «تای»<sup>۱</sup> را رونمایی کرد. یک برنامه هوش مصنوعی که مانند نوجوانان صحبت می‌کرد و قرار بود در شبکه‌های مجازی گوناگون با کاربران ۱۸ تا ۲۴ ساله گفت‌وگو کند. قرار بود تای از طریق مکالمات خود با کاربران، یاد بگیرد و زبان آموخته‌شده خود را در ارتباطات بعدی به کار گیرد. پالایه‌های متعددی در نظر گرفته شد و آزمایش‌های بسیاری با هدف عملکرد صحیح ربات انجام شد. پس از راه‌اندازی، هوش مصنوعی در مکالماتی که با افراد در شبکه‌های اجتماعی انجام داد از عبارات به شدت توهین‌آمیز و حاوی مطالب نژادپرستانه استفاده کرد (Swanson, 2018, p. 1205). ممکن است به هوش مصنوعی، که صرفاً اختیار مکالمه با انسان را دارد، اختیار هدف‌گیری و کشتن انسان سپرده شود؛ در این صورت معیارهای رایج برای مسئولیت کیفری نمی‌تواند منجر به مجازات شود؛ زیرا رفتار یا تصمیم‌های هوش مصنوعی ممکن است به هیچ‌وجه بیانگر قصد طراح یا کاربر آن نباشد. هوش مصنوعی قادر است واکنش‌هایی مشابه رفتارهای هوشمند انسانی داشته باشد؛ از جمله درک شرایط پیچیده، شبیه‌سازی فرایندهای تفکری و شیوه‌های استدلالی انسانی و پاسخ موفق به آن‌ها، یادگیری و توانایی کسب دانش و استدلال برای حل مسائل.<sup>۲</sup> این امر بدین معناست که ضوابط مربوط به قصد مبتنی بر قصد اشخاص برای دستیابی به نتیجه غیرقانونی اجراشدنی نیستند. همچنین از آنجاکه تعیین دلایل تصمیم یا پیش‌بینی هوش مصنوعی ناممکن است، ضوابط مربوط به دلایل رفتار را بررسی می‌کنند اجراشدنی نیستند (Bathae, 2017, p. 94). در واقع پیش‌بینی ناپذیر بودن هوش مصنوعی اعمال مسئولیت کیفری را با مشکل مواجه می‌کند. اگر در موردی خاص ثابت شود که رفتار مجرمانه اسلحه هوش مصنوعی پیش‌بینی ناپذیر بوده است، این امر به طور مؤثری از مسئولیت کیفری پیشگیری می‌کند (Acquaviva, 2022, p. 111). چنانچه پیش‌بینی ناپذیر بودن هوش مصنوعی موجب شود که مجازات به کارگیرنده آن امکان‌پذیر نباشد، در این صورت ممکن است به کارگیرنده این عمل را تکرار کند؛ بدون این که خطر محاکمه و مجازات تهدیدش کند. همین موضع در خصوص اعمال دکترین مسئولیت فرمانده نیز صادق است.

## ۲-۲. سلاح هوش مصنوعی؛ مسئله‌ای برای مسئولیت فرمانده

مطابق ماده ۲۸ اساسنامه، شرط اولیه برای اعمال مسئولیت کیفری مبتنی بر دکترین مسئولیت فرمانده این است که او می‌دانسته افراد تحت امرش مرتکب جرم خواهند شد. باید توجه شود که به دلیل ماهیت خاص هوش مصنوعی، به احتمال بسیار زیاد، هر جنایتی که هوش مصنوعی مرتکب شود با استناد به تحقق نیافتن این شرط براساس دکترین مذکور، مجازات‌شدنی نیست. مسئولیت فرماندهی فقط در صورتی ایجاد می‌شود که فرمانده از جرم آگاهی مسلم یا مفروض داشته باشد. دستیابی به سطح آگاهی مقتضی برای فرماندهان/آمران دشوار خواهد بود. آگاهی مسلم از یک عمل مجرمانه قریب‌الوقوع فقط در صورتی رخ می‌دهد که سلاح خودکار انتخاب هدف خود را قبل از شروع حمله اعلام کند. مشخص نیست که فرماندهان تا چه حد از یک حمله مجرمانه قریب‌الوقوع، که زمان رویدادش بسیار کمتر از مجال واقعی برای خنثی کردن آن است، اطلاع مسلم داشته باشند؛ زیرا بیشترین اطلاعات ما از شیوه کار و شیوه طراحی ماشین‌ها و برنامه‌های رایانه‌ای این است که کاربران یا سازندگان آن‌ها به دنبال چه اهدافی هستند و برای رسیدن به آن اهداف، به ماشین یا برنامه خود اجازه استفاده از چه ابزارهایی را می‌دهند. در برخی موارد،

<sup>۱</sup>. TAY

<sup>۲</sup>. فیلسوفی انگلیسی به نام آلن ماتیسون تورینگ، که از او با عنوان پدر علم محاسبه نوین و هوش مصنوعی یاد می‌شود، پیشنهاد داد که معیار هوش مصنوعی محسوب‌شدن را یک آزمودن رفتاری بگذاریم؛ بدین ترتیب که به صورت ناشناس یک ماشین را در مقابل یک انسان قرار دهیم، آن انسان با آن ماشین یکسری مکالمات تایی انجام می‌دهد. اگر انسان نتوانست تشخیص دهد جواب‌هایی که می‌گیرد از یک ماشین است، آن ماشین یک سیستم هوش مصنوعی به‌شمار می‌رود (Hallevy, 2015, p.7) و اگر آن انسان بتواند تشخیص دهد که در مقابلش یک ماشین قرار دارد، آن ماشین یک عامل هوشمند به‌شمار نمی‌رود.

با بررسی کدهای برنامه‌نویسی شده می‌توان تشخیص داد که طراح برنامه در تلاش برای دستیابی به چه چیزی بوده است و برنامه ممکن است از چه ابزارهایی برای تحقق آن اهداف استفاده کرده باشد، اما در نهایت جعبه سیاه هوش مصنوعی به‌گونه‌ای عمل می‌کند که سازندگان برنامه قادر به پیش‌بینی آن نباشند یا کارهایی انجام دهد که سازندگان هوش مصنوعی ممکن است درک نکنند یا قادر به پیش‌بینی آن‌ها نباشند ( Bathaee, 2017, pp. 906-7).

تفاوت هوش مصنوعی با دیگر محصولات یا سلاح‌ها این است که عملکرد دیگر سلاح‌ها به راحتی نشان‌دهنده نیت سازنده و یا به‌کارگیرنده آن است و به راحتی می‌توان تشخیص داد که به‌کارگیرنده به دنبال چه نتیجه‌ای بوده است، اما در خصوص هوش مصنوعی این‌گونه نیست و نمی‌توان همه اعمال و رفتار هوش مصنوعی را نماینده قصد سازنده یا به‌کارگیرنده آن دانست. چه بسا هوش مصنوعی رفتاری انجام دهد که به ذهن سازنده خطور هم نکرده باشد یا رفتاری انجام دهد که به‌کارگیرنده به هیچ‌وجه قادر به پیش‌بینی آن نباشد؛ بنابراین چنانچه دیوان نتواند نشان دهد که به‌کارگیرنده هوش مصنوعی در خصوص نتیجه قصد تحقق داشته است یا مطابق سیر عادی امور نتیجه محقق می‌شود، در این صورت نمی‌توان مسئولیتی اعمال کرد. مهم‌ترین مسئله در مسیر اعمال مسئولیت مربوط به پیش‌بینی وقوع نتیجه براساس سیر طبیعی امور است؛ به این معنا که چنانچه مرتکب یا مرتکبان فاقد قصد نتیجه باشند، آیا می‌توان براساس پیش‌بینی وقوع نتیجه مسئولیت کیفری را اعمال کرد؟ برای پاسخ به این پرسش، نخست باید به این پرسش اساسی پاسخ داد که آیا رفتار اسلحه هوش مصنوعی اصولاً پیش‌بینی‌پذیر است که این پیش‌بینی به منزله‌ی مستند اعمال مسئولیت کیفری قرار بگیرد؟

برای پاسخ به پرسش قبلی، باید به این واقعیت توجه داشت که اسلحه هوش مصنوعی به‌طور مؤثری مسیر خود را در مقابله با شرایط گوناگون نوع رفتار خود انتخاب می‌کند و دائماً با سازوکارهایی که برای انسان نامفهوم باقی می‌ماند، یاد می‌گیرد و رفتار خود را بهینه می‌کند و در این مسیر، هرچقدر کنترل انسانی زیاد باشد، باز هوش مصنوعی به نحو کاملاً پیش‌بینی‌ناپذیر رفتار می‌کند؛ به این دلیل است که هوش مصنوعی در اصل براساس الگوریتم‌هایی است که به‌صورت خودکار و بدون دخالت انسانی طراحی شده‌اند. این الگوریتم‌ها ممکن است با شرایط ورودی گوناگون به‌طور متفاوت عمل و نتایج متفاوتی را تولید کنند. به‌علاوه، هوش مصنوعی با تکیه بر یادگیری ماشین و شبکه‌های عصبی، با استفاده از داده‌های بزرگ و تنوع‌پذیر، به‌طور مستمر و با تغییراتی در طول زمان بهبود می‌یابد؛ به این معنی که پیش‌بینی دقیق و قطعی عملکرد آینده هوش مصنوعی، به دلیل پیچیدگی و متغیر بودن آن، ممکن نیست و همواره نیاز به تحلیل و بررسی دقیق برای شرایط خاص دارد. مثال‌هایی مستند وجود دارند که به خوبی این واقعیت را نشان می‌دهند.

سازمان حقوق بشر در گزارش خود، درباره سلاح هوش مصنوعی اشاره کرد که در نظر گرفتن مسئولیت کیفری برای فرماندهان در خصوص عملکرد هوش مصنوعی، که نمی‌توان بر آن کنترل کافی داشت، منصفانه نیست؛ از این رو خودمختاری این اسلحه‌ها باعث ایجاد مشکلاتی در مسیر اعمال مسئولیت می‌شود.<sup>۱</sup> اکنون می‌توان نشان داد که فرایند فزاینده‌ای از اقدامات ماشینی وجود دارد که در آن هیچ‌کس کنترل کافی بر اعمال ماشین ندارد تا بتواند مسئولیت آن را برعهده گیرد (Matthias, 2004, p. 177). ما می‌توانیم فرایندی را شناسایی کنیم که در آن تولیدکننده یک هوش مصنوعی به‌طور فزاینده‌ای کنترل خود را بر آن از دست می‌دهد و به تدریج این کنترل را به خود ماشین منتقل می‌کند. در پیشرفتی ثابت، نقش برنامه‌نویس از کدگذار به خالق موجودات نرم‌افزاری تغییر می‌کند. اساساً برنامه‌نویس بخشی از کنترل خود بر محصول را به محیط منتقل می‌کند (Matthias, 2004, p. 182) و به همان میزان که تأثیر سازنده در ماشین کاهش می‌یابد، تأثیر محیط نیز افزایش می‌یابد.

### ۳. راهکارهای حقوقی مقابله با مسائل هوش مصنوعی برای تحقق مقابله با بی‌کیفرمانی

دیوان کیفری بین‌المللی ساختار حقوقی نوینی برای تعقیب مرتکبان جنایات بین‌المللی و ناقضان حقوق انسان‌هاست. دیوان مزبور با ایجاد پیوند بین صلح و امنیت بین‌المللی و عدالت کیفری با مقوله حقوق بشر و حقوق بشردوستانه ارتباط یافته است که در پرتو اساسنامه‌اش، نقش مهمی

<sup>۱</sup>. HUMAN RIGHTS WATCH, LOSING HUMANITY: THE CASE AGAINST KILLER ROBOTS 42 (2012), <https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots>

در پیشبرد سیاست‌های حقوق بشردوستانه خواهد داشت. تحول و تکامل حقوق کیفری بین‌المللی به سمت حمایت سازمان‌یافته از حقوق بشردوستانه و برخورد کیفری با نقض‌های بنیادین حقوق بشر دوستانه است (فروغی‌نیا، ۱۳۹۸، ص ۳۷). همچنین پیدایش سلاح هوش مصنوعی هدف اصلی دیوان، یعنی مقابله با بی‌کیفرمانی، را با مشکل مواجه می‌کند. برای مقابله با این شرایط، جامعه بین‌المللی باید اقداماتی انجام دهد. در ادامه، طی چهار قسمت به این موارد پرداخته می‌شود.

### ۳-۱. مقابله با بی‌کیفرمانی و مسئله هوش مصنوعی

در مواجهه با ارتکاب گسترده جنایات توصیف‌ناپذیر پس از جنگ جهانی دوم، جامعه بین‌المللی به شیوه بی‌سابقه‌ای به لزوم رعایت عدالت توجه نشان داد. از جمله از طریق تصویب کنوانسیون پیشگیری و مجازات جرم نسل‌کشی، کنوانسیون‌های چهارگانه ژنو و اصول نورنبرگ، به دنبال پایان‌بخشیدن به چنین جرائمی بود (Sang-Hyun Song, 2012). در واقع، از زمان دادگاه‌های نورنبرگ و توکیو تاکنون، مفهوم مسئولیت کیفری بین‌المللی فردی، نقش مهمی در تحول بنیادین حقوق بین‌الملل ایفا کرده است. در این تحول، تأکید بر فرد/انسان در تبدیل حقوق بین‌الملل از حقوق بین‌دولتی به حقوق جامعه بین‌المللی، اهمیت فزاینده‌ای یافته است (Chao Yi, 2019, p. 115). اهمیت این گذار در این است که تا پیش از آن، فقدان سازوکارهای اجرایی معتبر، نقض قوانین بین‌المللی بشردوستانه با مصونیت آشکاری که عمدتاً دولتمردان و صاحبان قدرت از آن برخوردار بودند با شدت جریان داشت و «بی‌کیفرمانی»<sup>۱</sup> جنبه بسیار خطرناکی در این پهنه به‌شمار می‌رفت که مانع از تحقق عدالت می‌شد. بنابراین، در پاسخ به این وضعیت، جامعه بین‌المللی تصمیم گرفت با ایجاد یک سیستم به‌هم‌پیوسته عدالت بین‌المللی، اقدام مشترکی انجام دهد تا از بی‌کیفرمانی مرتکبان بدترین جنایات‌های شناخته‌شده علیه بشریت پیشگیری کند.

در راستای تحقق عدالت کیفری بین‌المللی، پس از لازم‌الاجرا شدن اساسنامه رم در یکم ژوئیه ۲۰۰۲، اولین دادگاه کیفری بین‌المللی دائمی تشکیل شد (Lansford & Muller, 2012, p. 1688). چنانچه در مقدمه این اساسنامه آمده است، تأسیس دیوان کیفری بین‌المللی برای پایان‌بخشیدن بی‌کیفرمانی<sup>۲</sup> و مجازات مرتکبان جدی‌ترین جرائمی است که ارتکاب آن‌ها کلیت جامعه بین‌المللی را نگران می‌کند، تا از این طریق از بروز چنین جرائمی پیشگیری شود (Gaskins, 2022, p. 35). به بیان دقیق‌تر، دولت‌های عضو در مقدمه اساسنامه دیوان کیفری بین‌المللی تأکید کردند: «فجیع‌ترین جنایاتی که موجب نگرانی جامعه بین‌المللی می‌شود، نباید بدون مجازات بماند و این که آن‌ها مصمم‌اند که به بی‌کیفرماندن مرتکبان این جنایات پایان دهند.»<sup>۳</sup>

هم‌زمان با تأسیس دیوان، با هدف پایان‌دادن به بی‌کیفرمانی مجرمان، پیشرفت‌های چشمگیر در فناوری و از جمله هوش مصنوعی، موجب به وجود آمدن نگرانی‌هایی در خصوص مسئولیت ناشی از به‌کارگیری آن‌ها شد. اولین بار اندیشمندی به نام آندرس ماتئاس<sup>۴</sup> در مقاله‌ای مهم<sup>۵</sup> به این مسئله اشاره کرد که دستگاه‌های یادگیرنده خودکار، که براساس یادگیری تقویتی<sup>۶</sup> توسعه یافته‌اند، به دلیل ماهیتی که از آن با عنوان جعبه سیاه<sup>۷</sup> یاد می‌شود، باعث ایجاد معضلی<sup>۸</sup> جدی در خصوص سرزنش‌پذیری می‌شوند<sup>۹</sup> که از آن با عنوان شکاف مسئولیت<sup>۱۰</sup> یاد شده است. صورت‌بندی ماتئاس از شکاف مسئولیت، به‌ویژه درباره دستگاه‌های تسلیحاتی خودکار مجهز به هوش مصنوعی<sup>۱۱</sup>، بسیار تأثیرگذار بود (International

1. Impunity

2. Ending Impunity

3. ICC Statute, Preamble, paras 4 – 5: "That the Most Serious Crimes of Concern to the International Community as a Whole Must not Go Unpunished", and that They are "Determined to Put an End to Impunity for the Perpetrators of These Crimes".

4. Andreas Matthias

5. Matthias, A. (2004). The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata. Ethics and Information Technology

6. Reinforcement Learning

7. Black Box

8. Dilemma

9. این موضوع و مسائل جانبی آن، به‌قدری مهم است که مجلس اعیان بریتانیا برای بررسی آن کمیته هوش مصنوعی تشکیل داده است.

10. Responsibility Gap

11. در ترور دانشمند ایرانی، محسن فخری‌زاده، از اسلحه مجهز به هوش مصنوعی استفاده شد.

(Human Rights Clinic, 2015). او در این مقاله اشاره می‌کند که استفاده از هوش مصنوعی و ماشین‌های یادگیری مبتنی بر داده، نوع جدیدی از ابهام فناوری‌ها و به بیان دیگر، شرابیطی با فقدان توضیح برای رفتار را به وجود آورده است که فراهم‌شدن عناصری مثل قصد، قابلیت پیش‌بینی و کنترل را برای سازنده ناممکن کرده است. در واقع، اشاره او به این موضوع است که حقوق‌دانان امروز در نقطه چرخشی تاریخی در حقوق کیفری قرار گرفته‌اند؛ چراکه حقوق کیفری مدرن انسان‌محور است و فاعلی به جز انسان در نظر ندارد و آنچه تاکنون به دست انسان ساخته شده، از منظر حقوق کیفری در حکم وسیله‌ای در اختیار انسان تلقی شده است و هر رفتاری از ساخته دست بشر سر بزند، فرض اصلی این بوده است که آن رفتار منتسب به سازنده است و او پاسخ‌گوی آسیب ایجادشده ناشی از محصول است؛ اما هوش مصنوعی این فرض غالب را از جهات گوناگون تغییر داده است.

### ۲-۳. الزام سازندگان به ایجاد شفافیت حداکثری در هوش مصنوعی

الزام سازندگان به ایجاد شفافیت حداکثری در هوش مصنوعی، یکی از راه‌های مقابله با بی‌کیفرمانی ناشی از سلاح هوش مصنوعی است؛ زیرا شفاف نبودن و پیش‌بینی ناپذیر بودن جزء ذات هوش مصنوعی است و بدون توانایی نشان‌دادن فرایندهای فکری هوش مصنوعی، نمی‌توان شواهدی را برای وجود قصد یا ارتباط بین اقدامات هوش مصنوعی و پیامدهای آن‌ها به دست آورد و این پنهان‌کاری یک معضل است. هوش مصنوعی خود یادگیرنده ماهیت بسیار پیچیده و خاصی دارد. به بیان دیگر، سیستم هوش مصنوعی به جعبه سیاهی می‌ماند که نمی‌دانیم داخل آن چه اتفاقاتی رخ می‌دهد و این شفاف نبودن نیز خود باعث به وجود آمدن مسائل حقوقی دیگر می‌شود. از جمله در مواردی که دو شرکت در ساخت یک سلاح هوش مصنوعی همکاری می‌کنند و آن سلاح موجب جنایت می‌شود، حقوق‌دانان باید بدانند که منشأ آسیب عملکرد کدام شرکت است و در برخی موارد، حتی نیاز است که میزان تقصیر احتمالی سازندگان یا به کارگیرندگان مشخص شود؛ در حالی که ماهیت جعبه سیاه هوش مصنوعی این اجازه را به حقوق‌دانان نمی‌دهد و یکی از پیچیدگی‌های هوش مصنوعی این است که تقریباً محال است بتوان میزان تأثیرات مداخله هر فرد در عملکرد و نتایج هوش مصنوعی را تعیین کرد (ولی پور و اسماعیلی، ۱۴۰۰، ص ۹). شفافیت در هوش مصنوعی ممکن است شامل ارائه دقیق الگوریتم‌ها، داده‌های ورودی و خروجی، نحوه پردازش داده‌ها و روش تصمیم‌گیری هوش مصنوعی باشد. شفافیت در هوش مصنوعی به دادگاه‌ها اجازه می‌دهد تا تعیین کنند که تصمیمات هوش مصنوعی چگونه گرفته شده و به چه عواملی توجه شده است. شفافیت به معنای ارائه اطلاعات دقیق درباره عملکرد و تصمیم‌های هوش مصنوعی است که باعث می‌شود کاربران بتوانند بهتر درک کنند که هوش مصنوعی چگونه تصمیم‌گیری می‌کند و چگونه به این تصمیم‌ها می‌رسد. شفافیت برای تضمین کارایی نظارت، مسئولیت‌پذیری انسانی و اعتماد به کارکرد سلاح بسیار مهم است.

در همین راستا، در قانون هوش مصنوعی اتحادیه اروپا (مصوب ۲۰۲۱)<sup>۱</sup>، که حاصل هزاران ساعت گفت‌وگوی علمی متخصصان هوش مصنوعی و حقوق‌دانان است، در ماده ۱۳ بیان شده است که سیستم‌های هوش مصنوعی پرخطر باید به گونه‌ای طراحی شوند و توسعه یابند که اطمینان حاصل شود عملکردشان به اندازه کافی شفاف است تا کاربران بتوانند خروجی سیستم را تفسیر و به شکل مناسب از آن استفاده کنند. سیستم‌های هوش مصنوعی پرخطر باید همراه با دستورالعمل‌هایی برای استفاده در قالب دیجیتالی مناسب یا حاوی اطلاعات مختصر، کامل، صحیح و واضحی باشند که برای کاربران مرتبط، در دسترس و درک‌پذیر باشد و در ادامه، در ماده ۱۴ بیان می‌دارد که سیستم‌های هوش مصنوعی پرخطر باید به گونه‌ای طراحی و توسعه داده شوند که بتوانند از طریق اشخاص حقیقی در طول دوره‌ای که سیستم هوش مصنوعی در حال استفاده است به طور مؤثر نظارت شوند. نظارت انسانی باید با هدف پیشگیری یا به حداقل رساندن خطرات سلامتی،

1. Anthropocentric

2. ARTIFICIAL INTELLIGENCE ACT 2021

ایمنی یا حقوق اساسی - که ممکن است هنگام استفاده از یک سیستم هوش مصنوعی پرخطر مطابق با هدف مدنظر خود یا تحت شرایط سوء استفاده منطقی پیش‌بینی‌پذیر پدیدار شوند - انجام شود.

### ۳-۳. ممنوعیت تولید سلاح‌های هوش مصنوعی با خطرات ناپذیرفتنی

ممنوع کردن تولید انواع خاصی از سلاح‌های هوش مصنوعی، که قابلیت خودبهبودی و فراترفتن از نظارت انسانی را دارند، راه‌حلی مناسب است. چنین ممنوعیتی براساس حقوق بشر و اصول بشردوستانه‌اند، که نه فقط اخلاقی، بلکه قانونی هستند. به‌ویژه، یک الزام ضمنی برای قضاوت انسانی را می‌توان در حقوق بین‌المللی بشردوستانه حاکم بر منازعات مسلحانه یافت. در واقع، این الزام در اصول تمایز، تناسب و ضرورت نظامی - که در معاهدات بین‌المللی، مانند کنوانسیون‌های ژنو ۱۹۴۹، یافت می‌شود - در حقوق عرفی بین‌الملل کاملاً هویداست. اصول مشابهی نیز در حقوق بین‌الملل حقوق بشر - که حقوق بشر معینی را برای همه مردم، صرف‌نظر از منشأ ملی یا قوانین محلی، در همه حال تضمین می‌کند - مستتر است. تأکید پژوهش حاضر بر این است که حقوق بشر در قبال طیف وسیعی از فناوری‌های خودکار و مستقل وظیفه‌ای خاص دارد؛ به‌ویژه، افراد و دولت‌ها در زمان صلح و همچنین رزمندگان، سازمان‌های نظامی و دولت‌های در شرایط درگیری مسلحانه وظیفه دارند که اختیار یا قابلیت شروع استفاده از نیروی کشنده را به‌طور مستقل به یک ماشین یا فرایند خودکار واگذار نکنند؛ زیرا لازم است در مورد مشروعیت اخلاقی و قانونی آن در هر مورد تصمیم‌های انسانی گرفته شود.

در حقوق بین‌الملل بشردوستانه، اصول تمایز و تناسب وجود دارد. اصل تناسب، حملات در درگیری را - که جمعیت غیرنظامی را در معرض آسیب بیش از حد نسبت به مزیت نظامی به‌دست‌آمده قرار می‌دهد - ممنوع می‌کند. اصل تمایز، تعهد طرف‌های درگیری مسلحانه را ایجاد می‌کند که بین اهداف غیرنظامی و نظامی تمایز قائل شوند و عملیات خود را فقط علیه اهداف نظامی هدایت کنند. اشاره شده است که «سلاح‌های کاملاً خودمختار با موانع مهمی برای رعایت اصول تمایز و تناسب مواجه خواهند بود»؛ برای مثال این سیستم‌ها فاقد قضاوت انسانی لازم برای تعیین این هستند که آیا آسیب غیرنظامی مدنظر بیشتر از مزیت نظامی پیش‌بینی‌شده در موقعیت‌های جنگی همیشه در حال تغییر و پیش‌بینی‌نشده است یا خیر.

علاوه بر این، استدلال شده است که سلاح‌های خودمختار مرگبار دستورات وجدان عمومی را نقض و «اصول بشریت را تضعیف می‌کنند؛ زیرا نمی‌توانند شفقت یا قضاوت انسانی را برای تصمیم‌گیری‌های استفاده از زور به‌کار ببرند. نکته دیگر این‌که در حذف انسان‌ها از حلقه، توانایی توضیح این‌که چرا تصمیم خاصی گرفته شده است نیز حذف می‌شود. کمیته بین‌المللی صلیب سرخ خاطرنشان کرد که «توضیح‌پذیری» مشکلی اساسی برای الگوریتم‌های یادگیری ماشینی است که در نحوه عملکردشان شفاف نیستند و هیچ توضیحی برای این‌که چرا خروجی مشخصی تولید می‌کنند که مشکلی مهم برای سیستم‌های تسلیحاتی خودمختار است، ارائه نمی‌دهند (Asaro, 2012, p. 689). اگر سلاح پیش‌بینی‌ناپذیری نامطمئن یا اشتباه رفتار کند، در حال حاضر نمی‌توان دلیل آن را تعیین کرد. الگوریتم‌ها قادر به درک یا مفهوم‌سازی ارزش زندگی انسان نیستند. سلاح‌های خودمختار کشنده خط قرمز اخلاقی واضحی را نشان می‌دهند که به الگوریتم‌ها اجازه می‌دهد تصمیم بگیرند که چه کسی می‌میرد و آن آسیب‌کشنده را اعمال کنند. نقشه راه سازمان ملل متحد برای همکاری دیجیتال به‌صراحت اشاره کرده است که: «تصمیم‌های مرگ و زندگی نباید به ماشین‌ها واگذار شود.»<sup>۱</sup> همچنان‌که کنوانسیون سازمان ملل متحد در مورد برخی سلاح‌های متعارف<sup>۲</sup> و ماده ۱۴ قواعد دهللی نو<sup>۳</sup> استفاده از سلاح‌هایی که قادرند از کنترل افرادی که آن‌ها را به‌کار می‌گیرند خارج شوند و در نتیجه جمعیت غیرنظامی را به خطر بیندازند ممنوع می‌کند؛ بنابراین قبل از ظهور طیف وسیعی از سیستم‌های تسلیحاتی خودمختار، که احتمالاً تهدیدی جدی برای حقوق اولیه‌اند، ایجاد این وظیفه در حکم هنجاری بین‌المللی و بیان آن با یک معاهده مفید خواهد بود.

<sup>1</sup>. [https://www.un.org/techenvoy/sites/www.un.org.techenvoy/files/Update\\_on\\_Roadmap\\_implementation\\_July\\_2022.pdf](https://www.un.org/techenvoy/sites/www.un.org.techenvoy/files/Update_on_Roadmap_implementation_July_2022.pdf)/ 04/28/2023

<sup>2</sup>. The UN Convention on Certain Conventional Weapons

<sup>3</sup>. New Delhi Rules

### ۳-۴. اصلاح اساسنامه و تعیین صریح مسئولیت در موارد سلاح هوش مصنوعی

در موارد پیشین، به این اشاره شد که باید شفافیت حداکثری و حتی ممنوعیت تولید برخی از انواع را برای سلاح‌های هوش مصنوعی در نظر گرفت. پرسشی که مطرح می‌شود این است که چنانچه به‌رغم تلاش‌های جامعه بین‌المللی این موارد رعایت نشد و استفاده از این نوع سلاح‌ها موجب ارتکاب جنایت شد، چه راهکاری پیشنهاد می‌شود؟ در این موارد، جامعه جهانی با شرایطی مواجه است که سلاح هوش مصنوعی مثلاً منجر به کشتن غیرنظامیان می‌شود؛ اما از آنجاکه نمی‌توان صرف به‌کارگیری هوش مصنوعی - که ذاتاً قابلیت کشتن دارد و خطرناک است - کسی را مجازات کرد، جرم‌بودن به‌کارگیری این سلاح باید به اساسنامه دیوان الحاق شود. سلاح‌های کاملاً خودمختار ظرفیت عمل مستقل را دارند و بنابراین می‌توانند مستقل و پیش‌بینی‌ناپذیر حمله‌ای بدون تمایزگذاری علیه غیرنظامیان انجام دهند. همچنین با توجه به ذات هوش مصنوعی، که در فرایند تصمیم‌گیری شفافیت ندارد، در بیشتر موارد نیز نمی‌توان جنایات هوش مصنوعی را درخصوص به‌کارگیرنده یا سازنده جنایات عمد دانست؛ از این رو با شرایطی مواجه‌ایم که نوعی ایجاد خطر برای بشریت است؛ بدون امکان پاسخ‌گو کردن و مجازات شخص خاصی. از همین منظر است که می‌توان چرایی شکل‌گیری جنبش‌های مردمی در راستای ممنوعیت استفاده از سلاح هوش مصنوعی را درک کرد.

همچنین توضیح‌پذیری<sup>۱</sup> و تفسیرپذیری<sup>۲</sup> نبودن دو مورد از ویژگی انحصاری و متمایزکننده هوش مصنوعی است. در سال ۲۰۱۶، یکی از شرکت‌های زیرمجموعه گوگل در لندن، هوش مصنوعی تولید کرد و بازی سختی به نام GO را در برابر حریف انسانی قرار داد. او به راحتی توانست بهترین بازیکن این بازی را شکست دهد، اما نکته شایان توجه این بود که هوش مصنوعی در حرکت ۳۷ بازی دوم، حرکتی کرد که از تصور همگان خارج بود و حتی متخصص حاضر گمان کرد در عملکرد هوش مصنوعی اشتباهی رخ داده است که این‌گونه حرکت کرده است. این مورد به خوبی پیش‌بینی‌ناپذیری رفتار هوش مصنوعی را نشان می‌دهد (Selbst, 2020, p. 1343). حال اگر تصور شود هوش مصنوعی به جای بازی کردن، وظیفه بمب‌باران مواضع غیرنظامی را بر عهده داشت، کشتار وسیع غیرنظامیان رخ می‌داد؛ بدون آن‌که کسی چنین چیزی را پیش‌بینی می‌کرد. زمانی که رفتار اسلحه هوش مصنوعی پیش‌بینی‌ناپذیر باشد، فجایع ناشی از آن رفتار نیز پیش‌بینی‌ناپذیر است و هنگامی که قابلیت پیش‌بینی وجود نداشته باشد، طبق اساسنامه دیوان، مجرای برای اعمال مسئولیت کیفری باقی نمی‌ماند و اگر مسئولیت وجود نداشته باشد، مبنایی برای استفاده نکردن از سلاح هوش مصنوعی، که به‌طور بالقوه قادر است موجب کشتار شود، وجود ندارد. از این رو، دیوان باید به‌صراحت صرف سپردن اختیار کشتن انسان به هوش مصنوعی را موجب مسئولیت کیفری بشمارد و به‌صورت شفاف و صریح در مورد نحوه اعمال مسئولیت کیفری در ارتباط با این موارد به‌کار گرفته شود. در این الحاقیه یا اصلاحیه، به‌خصوص به این موضوع پرداخته می‌شود که اگر عامل انسانی از هوش مصنوعی ای استفاده کند که به دلیل شفاف نبودن منشأ تصمیم‌گیری آن را مشخص کرد و موجب کشتار شود صرف به‌کارگیری موجب مسئولیت کیفری شخص به‌کارگیرنده می‌شود. علاوه بر این، اگر به‌کارگیرنده‌ای اجازه داشته باشد که از سلاح هوش مصنوعی با قابلیت خودبه‌بودی پس از تولید بدون ناظر انسانی استفاده کند و این خودبه‌بودی سبب تصمیم هوش مصنوعی در ارتکاب جنایت شود، این شخص باید در قبال هرگونه جنایت هوش مصنوعی مسئولیت کیفری داشته باشد.

### نتیجه‌گیری

فلسفه اصلی تشکیل دیوان کیفری بین‌المللی، مقابله با بی‌کیفرمانی افرادی است که مرتکب نقض فاحش حقوق بشر می‌شوند. مقررات دیوان انسان‌محورانه و برپایه این واقعیت تدوین شده است که هر سلاح کشتاری در نهایت وسیله‌ای در دست مرتکبی انسانی به‌شمار می‌رود و هر نتیجه‌ای از آن منتج شود، منتسب به به‌کارگیرنده آن است. اما هوش مصنوعی موجب تغییر این وضعیت شده است و می‌توان حالاتی را متصور شد که اسلحه هوش مصنوعی مرتکب کشتار غیرنظامیان شود و کسی هم به‌خاطر این کشتار سزانش نشود؛ از این رو نظام عدالت کیفری

1. Explainability

2. Interpretability

بین‌المللی دوباره با شرایطی مواجه می‌شود که نتیجه آن بی‌کیفرماندن کسانی است که از سلاح هوش مصنوعی استفاده می‌کنند؛ چراکه معیارهای ذکر شده در اساسنامه دیوان از جمله رکن روانی لازم برای مسئولیت کیفری مستقیم و غیرمستقیم به دلیل ماهیت خاص، بی‌بدیل و غیر مسبوق به سابقه هوش مصنوعی نمی‌تواند برآورده شود؛ از این رو، نظام عدالت کیفری بین‌المللی به زودی شاهد بازپیدایش شرایط نامطلوبی است که دیوان برای مقابله با آن شرایط تأسیس شد و این‌گونه هوش مصنوعی و سلاح مجهز به آن، به مهم‌ترین مسئله دیوان برای پاسخ‌گویی کیفری به هرگونه کشتار بالقوه‌ای است که ممکن است در آینده‌ای نزدیک رخ دهد. پژوهش حاضر، پس از تبیین و تحلیل تعامل هوش مصنوعی و مقررات اساسنامه دیوان، به این نتیجه دست یافت که استفاده از هوش مصنوعی در مخاصمات بین‌المللی، مقررات کنونی اساسنامه دیوان در خصوص اعمال مسئولیت کیفری و مقابله با بی‌کیفرماندن را بی‌اثر می‌کند و با شرایط جدیدی مواجه‌ایم که قوانین انسان‌محورانه قبل پاسخ‌گو نیست. برای جبران این نارسایی، پیشنهاد می‌شود که جامعه بین‌المللی سازندگان را ملزم به ایجاد شفافیت حداکثری در سلاح هوش مصنوعی کند؛ تولید سلاح‌های هوش مصنوعی با خطرات ناپذیرفتنی را ممنوع کند و از همه مهم‌تر، نحوه اعمال مسئولیت کیفری را در ارتباط با مواردی که عامل جنایت هوش مصنوعی است به‌صراحت و شفافیت به اساسنامه الحاق کند. همچنین پیشنهاد می‌شود که محتوای این الحاقیه با این مضمون باشد که چنانچه اختیار کشتار انسان به هوش مصنوعی ای سپرده شد که به دلیل استفاده گسترده از جعبه سیاه نتوان مشخص کرد که عامل کشتار غیرنظامیان به کدام مرحله توسعه هوش مصنوعی یا کدام شخص انسانی مرتبط است و به‌کارگیرنده از این وضعیت آگاه باشد، صرف به‌کارگیری آن موجب شود شخص در قبال هرگونه جنایت هوش مصنوعی مسئولیت کیفری داشته باشد؛ همچنین است در مواردی که به‌کارگیرنده از اسلحه هوش مصنوعی‌ای که دارای قابلیت خودبهبودی پس از تولید بدون ناظر انسانی است استفاده کند و خودبهبودی موجب تصمیم هوش مصنوعی در قبال ارتکاب جنایت شود.

## منابع

- ابوذری، مهرنوش (۱۴۰۰). حقوق و هوش مصنوعی، تهران: نشر میزان.
- تخشید، زهرا (۱۴۰۰). مقدمه‌ای بر چالش‌های هوش مصنوعی در حوزه مسئولیت مدنی. مطالعات حقوق خصوصی، ۱۸(۱)، ۲۲۷-۲۵۰.
- رجبی، عبدالله (۱۳۹۸). ضمان در هوش مصنوعی. مطالعات حقوق تطبیقی، ۱۰(۲)، ۴۴۹-۴۶۶.
- سیفی، بهزاد (۱۳۹۹). مسئولیت کیفری به‌کارگیری وسایل بدون سرنشین خودکار از دید حقوق بین‌الملل بشردوستانه. آفاق امنیت، ۳(۴۹)، ۵۴-۲۹.
- صالح‌آبادی، رامین (۱۳۹۸). تعیین حدود تحقیقات هوش مصنوعی از منظر حق و مصلحت عمومی. پایان‌نامه کارشناسی ارشد، دانشکده حقوق دانشگاه شهید بهشتی.
- عبداللهی، محسن، شهریاری، عبدالنعیم، بابایی، یوسف و یعقوبی، اسماعیل (۱۳۹۶). دیوان کیفری بین‌المللی؛ درنگی در الحاق جمهوری اسلامی ایران، تهران: انتشارات دبیرخانه مجمع تشخیص مصلحت نظام، چاپ اول.
- عزیزی بساطی، مجتبی و سکوتی، مرضیه (۱۳۹۴). واکاوی تأثیر سلاح‌های خودکار بر صلح و امنیت بین‌المللی. سیاست خارجی، ۲۹(۳)، ۵۶-۳۵.
- فروغی‌نیا، حسین (۱۳۹۸). موانع و چالش‌های اجرایی دیوان کیفری بین‌المللی در تقویت ضمانت اجرای حقوق بین‌الملل بشردوستانه، فصل‌نامه تعالی حقوق، ۳۵(۱)، ۳۷-۷۴.
- کسمتی، زهرا (۱۳۸۷). اصول انتساب مسئولیت فردی کیفری در چارچوب اساسنامه دیوان بین‌المللی کیفری. پایان‌نامه دکتری، دانشکده حقوق و علوم سیاسی، دانشگاه علامه طباطبایی.

هالوی، گابریل (۱۳۹۸). مسئولیت کیفری ربات‌ها: هوش مصنوعی در قلمرو حقوق کیفری. ترجمه فرهاد شاهیده و طاهره قوانلو، تهران: نشر میزان.

وائقی، محسن، حکمت‌نیا، محمود و محمدی، مرتضی (۱۳۹۸). مسئولیت مدنی ناشی از تولید ربات‌های مبتنی بر هوش مصنوعی خودمختار. حقوق اسلامی، ۶۰(۱۶)، ۲۳۱-۲۵۸.

ولی‌پور، علی و اسماعیلی، محسن (۱۴۰۰). امکان سنجی مسئولیت مدنی هوش مصنوعی عمومی ناشی از ایجاد ضرر در حقوق مدنی. فصل‌نامه اندیشه حقوقی، ۶(۲)، ۱-۱۶.

Acquaviva, G. (2022). Autonomous weapons systems controlled by Artificial Intelligence: a conceptual roadmap for international criminal responsibility. *The Military Law and the Law of War Review*, 60(1), 89-121.

Asaro, P. (2012). On banning autonomous weapon systems: human rights, automation, and the dehumanisation of lethal decision-making. *International Review of the Red Cross*, 94(886), 687-709.

Barber, I. A. (2020). Autonomous Weapons Systems & Accountability: Rethinking Criminal Responsibility for War Crimes at the ICC. *SOAS LJ*, 7, 5.

Bathae, Y. (2017). The artificial intelligence black box and the failure of intent and causation. *Harv. JL & Tech*, 31(20), 889 – 938.

Chao Yi (2019). The Concept of International Criminal Responsibility for Individuals and the Foundational Transformation of International Law. In *Philosophical Foundations of International Criminal Law: Foundational Concepts*, Brussels: Torkel Opsahl Academic EPublisher (pp. 65 – 139).

Hallevey, G. (2015). *Liability for crimes involving artificial intelligence systems* (Vol. 257). New York, NY, USA: Springer International Publishing.

International Human Rights Clinic (2015). *Mind the Gap: The Lack of Accountability for Killer Robots*, Zurich: IHRC.

Lansford, T., & Muller, T. (2012). *Political Handbook of the World*, sage publisher.

Larina, E. S., & Ovchinsky, V. S. (2018). *Artificial intelligence. Big data. Crime*. Moscow: Knizhny Mir.

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and information technology*, 6(3), p. 175-183.

Mosechkin, I. N. (2019). Artificial intelligence and criminal liability: problems of becoming a new type of crime subject. *Bulletin of St. Petersburg university law*, 10(3), p. 461-476.

Mueller, J. P., & Massaron, L. (2021). *Artificial intelligence for dummies*. New Jersey: John Wiley & Sons.

Scharre, P. (2018). *Army of none: Autonomous weapons and the future of war*. WW Norton & Company.

Schmitt, M. N. (2013). Autonomous weapon systems and international humanitarian law: A reply to the critics,” *Harv. Nat’l Sec. J. Features*, 1(4). p. 1-37. Available at <http://centaur.reading.ac.uk/89864/>

Selbst, A. D. (2020). Negligence and AI's human users. *BUL Rev.*, 100, 1315.



Song, S. H. (2012). The role of the International Criminal Court in ending impunity and establishing the rule of law. *UN Chronicle*, 49(4).

Swanson, G. (2018). Non-autonomous artificial intelligence programs and product liability: how new AI products challenge existing liability models and pose new financial burdens. *Seattle UL Rev.*, 42(3), 1201-1222.