



فصلنامه علمی پژوهشی اخلاق پژوهی

سال پنجم • شماره اول • بهار ۱۴۰۱

Quarterly Journal of Moral Studies
Vol. 5, No. 1, Spring 2022

علمی پژوهشی



استدلال‌هایی اخلاقی له و علیه ربات‌های قاتل

سید محسن اسلامی*

چکیده

ربات‌های قاتل یا سامانه‌های تسلیحاتی خودآیین (AWS) از نمونه‌های بارز کاربست هوش مصنوعی هستند، به نحوی که قابلیت‌ها و نگرانی‌ها درباره گسترش کاربست هوش مصنوعی در حوزه‌های مختلف را به خوبی نشان می‌دهند. طبیعتاً این موضوع واکنش‌های بسیاری را برانگیخته است. این سامانه‌ها از هوش مصنوعی در عملکرد اصلی خود استفاده می‌کنند، یعنی در انتخاب هدف و اقدام درباره او. طی سال‌های اخیر این بحث از حیث نظری در عرصه اخلاق، حقوق و سیاست اهمیت داشته است، با این نگاه که اگر چنین سلاح‌هایی در کار باشند باید چه موضعی درباره آنها (در سطح توسعه، ساخت و استفاده) اتخاذ کرد؟ گرچه این مسئله فارغ از امکان یا عدم امکان چنین سلاح‌هایی مهم است، پیشرفت‌های تکنولوژیک نیز به چنین ایدئالی نزدیک شده است و اهمیت مسئله را دوچندان کرده‌اند. در این نوشته به استدلال‌هایی اخلاقی له و علیه این سامانه‌ها می‌پردازم. به نظر نمی‌رسد هیچ‌یک از این استدلال‌ها برای اتخاذ موضع در این باره کافی باشند. بنابراین، یا باید استدلال‌های دیگری را در این باره جست‌وجو کرد، یا این وضع را دال بر نتیجه‌ای کلی‌تر درباره چارچوب این بحث دانست که نظریه جنگ عادلانه است. در ضمن این بررسی، بعضی از ابعاد اخلاقاً مهم این عرصه مرور خواهند شد.

کلیدواژه‌ها

اخلاق جنگ، اخلاق هوش مصنوعی، استدلال پیامدگرایی، کرامت انسانی، ربات قاتل.

* استادیار گروه فلسفه، دانشکده علوم انسانی، دانشگاه تربیت مدرس، تهران، ایران. | s.eslami@modares.ac.ir

تاریخ پذیرش: ۱۴۰۱/۰۴/۲۶ □ تاریخ تأیید: ۱۴۰۱/۰۵/۰۹

■ اسلامی، سید محسن. (۱۴۰۱). استدلال‌هایی اخلاقی له و علیه ربات‌های قاتل. فصلنامه اخلاق پژوهی. ۵(۱۴)، ۵۶-۳۱.

doi 10.22034/ethics.2022.50372.1551

۱. مقدمه

شنیدن اخباری با مضمون «رونمایی از سلاح خودآیین جدید» روز به روز عادی تر می‌شود. در عین حال، کسانی روز به روز این موقعیت را ترسناک‌تر ارزیابی می‌کنند. بسیاری معتقدند که این نسل جدید سلاح‌ها اساساً متفاوت با سلاح‌های قبلی هستند، چه اینکه نقش انسان در این سلاح‌ها بسیار کم است. این سلاح‌ها خود تصمیم می‌گیرند که چه کسی یا چه چیزی را هدف بگیرند و سپس علیه آن اقدام می‌کنند. این است که دولت‌هایی در پی توسعه و ساخت و البته، فروش این سلاح‌ها و طبعاً مخالف منع قانونی آنها هستند و دولت‌هایی در تلاش برای ایجاد ممنوعیت ساخت آنها؛^۱ گروه‌هایی در پی تأکید بر مزیت‌های این سلاح‌ها هستند و انجمن‌هایی در پی محکومیت و هشدار دادن درباره عواقب نابودگر آن.

با وصفی که گذشت، طبیعی است که بحث‌ها در این باره تنها محدود به محافل دانشگاهی نیست؛ اگر چه ادبیات مفصلی در حوزه‌های مربوط (به ویژه در فلسفه و حقوق) شکل گرفته است، سازمان‌هایی مانند «دیدبان حقوق بشر» و «صلیب سرخ» نیز پیوسته مطالب این حوزه را پوشش می‌دهند.^۲ به همین ترتیب، در رسانه‌های عمومی تر مانند نیویورک تایمز و گاردین نیز می‌توان مطالبی در ستایش یا مذمت این سلاح‌ها یافت، اما این دعوا درباره چیست؟ و طرفین دعوا چه می‌گویند؟ در اینجا هدف مرور بعضی ملاحظات اخلاقی در این باره است.

۲. از خودکار به خودآیین

آنچه امروزه با عنوان «سامانه سلاح خودآیین»^۳ شناخته می‌شود، به نوعی مبتنی بر تفکیکی میان خودآیین^۴ و خودکار^۵ است. دیرزمانی است که با مراحل «خودکار» در ماشین‌ها آشناییم. همه ماشین‌های پیچیده‌ای که اطراف ما هستند، مراحل «خودکار» دارند که بدون دخالت ما انجام می‌شوند. خودرویی که می‌رانیم بسیار مراحلی دارد که بی‌دخالت ما کار می‌کنند، اما، در نهایت،

۱. طبق گزارشی به تاریخ ۹ فروردین ۱۳۹۸ در گاردین، انگلستان، آمریکا و روسیه از جمله مخالفان منع قانونی این سلاح‌ها هستند (Gayle, 2019).

2. see <https://www.hrw.org/topic/arnas/killer-robots>
<http://blogs.icrc.org/new-delhi/tag/aws/>
 3. autonomous weapon system (AWS)
 4. Autonomous
 5. Automated

بی حضور و فعالیت انسان خودرو کار نمی‌کند. در سلاح‌ها نیز این مقوله آشناست. کافی است فرد ماشه‌ای را فشار دهد تا تفنگ شلیک کند و پوکه را بیرون بیندازد و فشنگ جدید را جایگزین کند و سلاح دوباره آماده شلیک باشد. این مراحل «خودکار» و بدون دخالت انسان انجام می‌شوند، اما انسان و فعالیت او بخشی اساسی از فرایند عملکرد این سلاح‌هاست.

حال، در نظر بگیرید - با بهره‌گیری از هوش مصنوعی - قابلیت‌های ماشین افزایش یابد و در مقابل، دخالت انسان در ماشین کم‌تر و کم‌تر شود. بدین ترتیب، می‌توانیم خودرویی داشته باشیم که «خودران» باشد؛ نه تنها تک‌تک تصمیم‌های چنین خودرویی را انسان نمی‌گیرد، بلکه خودرو در مهم‌ترین عملکردها (سرعت یا جهت حرکت) مستقل از انسان عمل می‌کند. به بیانی، انسان داخل فرایندهای عملکرد خودرو نیست، بلکه خارج از آن است.^۱ بدین ترتیب، اگر خودرو از راه دور کنترل شود، خودروی صرفاً بدون سرنشین است. در این صورت، بسیاری از عملکردها «خودکار» هستند، اما انسان همواره نقش خود را در عملکردهای خودرو ایفا می‌کند - همه تصمیم‌ها و دستورها از جانب انسان است، اما خودروی خودران یا «خودآیین» اساساً مستقل از انسان کار می‌کند و عملکردهایش منتظر دستورات انسانی نیست یا دست‌کم این چیزی است که حالت آرمانی خودروهای خودران به نظر می‌رسد.

به همین ترتیب، «سامانه سلاح خودآیین» آن سامانه و سلاحی است که افزون بر فرایندهای «خودکار»، از قابلیت‌هایی مضاعف برخوردار و بدین ترتیب «خودآیین» است.^۲ باید توجه کنیم که بحثی که در این حوزه درگرفته راجع به «سلاح» خودآیین است، به این معنا که خودآیینی در حوزه عملکردهای اساسی سلاح باشد، مانند انتخاب هدف و تشخیص اقدام علیه او.^۳ بدین ترتیب، نزاع حاضر درباره روایی یا ناروایی استفاده از هوش مصنوعی در بخش پرواز یا نحوه عبور از موانع نیست، بلکه چالش آنجاست که خود سلاح هدف را برگزیند و برای مثال، تصمیم به کشتن او بگیرد. این است که گاه به این سلاح‌ها «ربات قاتل» می‌گویند؛ عنوانی که به ویژه

۱. این تفاوت را با تفکیک *in the loop* و *out of the loop* تفکیک می‌کنند. گاه، در میانه، به *on the loop* نیز اشاره می‌شود؛ حالتی که انسان هرگاه بخواهد می‌تواند در فرایندها مداخله کند (see Leveringhaus, 2016, pp. 3-4).

۲. در ادامه، نکات بیشتری درباره کیفیت «خودآیینی» بیان خواهد شد.

۳. برای نمونه، نظر به اهمیت این مطلب است که گزارش ۲۰۱۶ صیلب سرخ از نشست‌هایی که در این باره برگزار کرده، در عنوان به «عملکردهای اساسی» (critical functions) اشاره کرده است:



مخالفتان این سلاح‌ها از آن استفاده می‌کنند.^۱ گویی ما صرفاً این سلاح را می‌نویسیم و می‌سازیم و روانه جنگ می‌کنیم. باقی به عهده خود سلاح است.

۳. ابعاد فلسفی

در کنار مسائل فنی و حقوقی، بخشی از سؤالات راجع به این موضوع فلسفی هستند و از آن میان قسم مهمی را سؤالات مرتبط به فلسفه اخلاق تشکیل می‌دهند. در ادامه، به نمونه‌هایی از این سؤالات می‌پردازیم. پیش از آن، بد نیست به چند دسته سؤال دیگر نیز اشاره شود.

۳.۱. ماشین «هوشمند»

«ربات‌های قاتل»، در تصویری اولیه، نوعی ماشین هوشمند هستند. در اینجا بحثی کلی این است که (الف) منظور از هوشمندی چیست و (ب) آیا ماشین هوشمند امکان دارد؟



۳۴

الف) هوشمندی

چارچوب و شکل امروزی بحث‌ها درباره ذهن مدیون آثار و آرای رنه دکارت - فیلسوف فرانسوی قرن هفدهم - است. او بر آن بود که دو جوهر نفس و ماده را باید کاملاً از هم تفکیک کرد؛ نفس که سرشت آن فکر است و ماده که مقابل آن است. بدین ترتیب، اساساً شیء مادی توان فکر (به معنای عام) ندارد (دکارت، ۱۳۹۱، تأمل دوم).

کسانی آنچه را امروزه «هوشمندی» می‌دانیم به آنچه دکارت «فکر» می‌خواند، مرتبط می‌کنند. دکارت، توانایی فکر در انسان را به سبب جوهر نفس می‌دانست.^۲ حال، اگر این تلقی از هوشمندی را مد نظر داشته باشیم، با این سؤال مواجه هستیم که منشأ این قابلیت چیست و آنگاه

۱. به عنوان نمونه‌ای از این نوع سلاح‌ها، گاه به «هاروپ» (Harop) اشاره می‌شود؛ هوایم‌ای بدون سرنشینی که می‌تواند به پرواز در آید، به تصمیم خودش هدف را منفجر کند یا بدون اقدام به آشیانه خود بازگردد. این هوایم‌ها از سال‌ها پیش به بعضی کشورها فروخته شده است و حتی در مواردی از آن استفاده کرده‌اند (see Plain, 2019).

۲. بنا به روایت رایج، دکارت بر آن بود که فقط انسان «نفس» دارد. تمام حیوانات دیگر فاقد نفس هستند و واکنش‌هایی که در آنها مشاهده می‌کنیم به سبب آن است که ماشین‌هایی بسیار پیچیده‌اند (نک: دکارت، ۱۳۴۴، ج ۱، بخش پنجم، به ویژه، ص ۲۲۷-۲۲۸).

باید با استدلال‌های دکارت درگیر شویم. بالاخره، چگونه ممکن است چیزی که فاقد نفس است، توانایی فکر داشته باشد و اگر تلقی دیگری مدّ نظر داشته باشیم، باید نخست آن را توضیح دهیم.^۱

ب) ملاک تعیین هوشمندی

فارغ از جزئیات تعیین مراد از «هوش»، امروزه بسیاری معتقدند که «تست تورینگ» ملاکی عرضه می‌کند تا به این سؤال پاسخ دهیم که چه ماشینی هوشمند است، اگر ماشینی بتواند به نحوی عمل کند که نتوان آن را از فاعل انسانی تشخیص داد، آن ماشین هوشمند است. این ایده‌ای است که در میان متخصصان علوم کامپیوتر متداول است. در نقد این رویکرد، جان سِرل چالش «اتاق چینی» را طرح کرده است.^۲

کسی را در نظر بگیرید که هیچ آشنایی با زبان چینی ندارد. او را در اتاقی قرار داده‌اند که کاربران آن اتاق چینی زبان هستند. افراد درخواست خود را روی کاغذ می‌نویسند و به داخل اتاق می‌فرستند. فرد داخل اتاق اگر چه معنای آن نوشته‌ها را نمی‌داند، کتابچه راهنمایی دارد که در آن به تفصیل تعیین شده است که در مقابل هر نوشته/علامت باید چه پاسخی بدهد؛ پاسخ می‌تواند هر چیزی باشد، اعم از یک نوشته/علامت باشد یا تحویل یک کالا. بدین ترتیب، فرد با استفاده از کتابچه راهنمای خود، به درخواست‌ها و گفته‌های افراد بیرون اتاق پاسخ می‌دهد، به نحوی که افراد بیرون اتاق باور می‌کنند که فرد درون اتاق، زبان چینی می‌داند.

حال، تست تورینگ را به یاد بیاورید. در آنجا، ایده آن است که اگر ماشین عملکردی مشابه انسان داشته باشد، می‌توان گفت که ماشین هوشمند است. طبق آن ملاک، به نظر می‌رسد باید بگوییم که فرد داخل اتاق نیز چیزی بلد است؛ چون عملکرد او مانند کسی است که چینی بلد است.

با این حال، فردی می‌تواند عملکردی مانند فرد چینی دان داشته باشد، متفاوت است از کسی که چینی بلد است. فرد مورد نظر، اصولاً هیچ درکی ندارد که درخواست افراد چه بوده است و هیچ درکی ندارد که به آنها چه پاسخی می‌دهد، اما کسی که چینی بلد است، نه تنها به افراد پاسخ

۱. از سوءتفاهم‌ها درباره «هوش مصنوعی» یکی از این است که آنچه اصطلاحاً «هوشمند» نامیده می‌شود عملکردی مانند ذهن انسان دارد. این مطلبی است که مایک وولدریج - مدیر دانشکده علوم کامپیوتر در دانشگاه آکسفورد - از آن به عنوان «بزرگ‌ترین بدفهمی» مردم درباره پیشرفت‌ها در هوش مصنوعی نام برده است (Butterfield, 2018).
۲. برای بحثی مختصر درباره این چالش و پاسخی به آن، نک: ریونزکرافت، ۱۳۸۷، ص ۱۰۱-۱۰۶. بحث تفصیلی درباره صورت‌بندی، لوازم، و پاسخ‌ها را می‌توان در اینجا یافت: Cole, 2019.



می دهد، بلکه درکی دارد از آنچه در جریان است. به نحو متناظر، استدلال بر اساس اتاق چینی این است که گذراندن تست تورینگ را نمی توان معادلِ هوشمندی دانست. آنچه در هوشمندی مهم است؛ نه عملکردی مشابه انسان، بلکه فهم است، اما این چیزی نیست که بتوان با تست تورینگ آن را سنجید. جزئیات و ادامه این بحث موضوع صحبت ما نیست. در اینجا، همین مقدار کافی است که اگر ماشینی عملکرد مشابه انسان داشته باشد، همچنان نمی توان گفت که ماشین دارای هوش یا ذهنی مشابه انسان است. البته، این لزوماً نقضی برای ماشین های به اصطلاح هوشمند نیست. چه بسا انتظار ما از آنها انجام بعضی عملکردها باشد و آنها نیز از پس آن بر بیایند، اما این موضوع در چارچوب کلان تر اهمیت می یابد.

چنانکه خواهیم دید، این نگرانی درباره سلاح های خودآیین جدی است. در مباحث جاری درباره این سلاح ها، وصف «خودآیینی» در مورد آنها استفاده می شود و به نحوی برای آنها توانایی مشابه انسان فرض شده است. با این حال، از طرفی، با این پرسش مهم مواجهیم که چگونه می توانیم در این باره قضاوت کنیم و اصولاً بدانیم که آنها قابلیت های ذهنی مشابه انسان دارند یا خیر. از طرف دیگر، باید توجه داشته باشیم که چگونه این امر در استدلال های مختلف نقش ایفا می کند.

۲.۳. ایضاح مفاهیم

از جمله وجوه دیگر بحث حاضر که سمت و سوی فلسفی دارد، ایضاح مفاهیم اصلی است. برای مثال - چنان که دیدیم - نخستین سؤال این است که «هوش» چیست و ملاک «هوشمندی» چیست. به همین ترتیب، در بحث ها شاهد استدلال هایی هستیم که در آنها برخی مفاهیم - مثل کرامت - نقش محوری دارند. در این موارد، از سویی، آن مفاهیم از آرای فیلسوفان مهمی در تاریخ وام گرفته شده اند و در سیاق تاریخ فلسفه می توان درباره جزئیات آنها بیشتر دانست. از سوی دیگر، آن مفاهیم در چارچوب های نظری - مثلاً درباره منشأ اخلاق - طرح شده اند که شاخه ای از فلسفه هستند. افزون بر اینها، اصولاً تحلیل مفاهیم از حیث روشی بیشتر به حوزه کار فلسفی نزدیک می شود.

۳.۳. مسائل روشی

معرفت شناسی، یکی از شاخه های اصلی فلسفه است و به این سؤال می پردازد که «چه

می‌دانیم؟» یا «چه می‌توانیم بدانیم؟». این سؤال کلی در بحث حاضر مصادیق زیادی دارد. برای مثال – چنان‌که دیدیم – این سؤال همچنان باقی است که چگونه می‌توانیم بدانیم ماشینی دارای «درک» است و آیا دانستن چنین چیزی اساساً ممکن است؟

به همین ترتیب – در بحث از ربات‌های قاتل – با مقوله‌ای مواجه هستیم که جدید است، درباره آن مجهولات زیادی داریم و آینده آن و نتایج آن برای ما روشن نیست. پس با این سؤال مواجهیم که درباره چنین پدیده‌هایی تا چه اندازه می‌توانیم بدانیم. یا در تصمیم‌گیری درباره پدیده‌ای که مجهولات زیادی درباره آن داریم، باید از چه قواعد راهنمایی پیروی کنیم؟

باری، در این مختصر، صرفاً به بخشی از ملاحظات اخلاقی درباره به‌کارگیری ربات‌های قاتل در جنگ می‌پردازیم و دیگر مسائل را دنبال نمی‌کنیم.^۱

۴. ابعاد اخلاقی به‌کارگیری ربات‌های قاتل

اصل بر پرهیز از جنگ است؛ چنان‌که اصل بر آسیب نزدن است. بنابراین، برای آنکه به دیگران آسیب نزنیم در پی دلیل اخلاقی نیستیم، اما ممکن است در شرایطی و به ملاحظات اخلاقاً مجاز باشیم به کسی آسیب بزنیم. برای مثال، اگر کسی به ما حمله کند و قصد جان ما را داشته باشد، می‌توانیم از خود دفاع کنیم. به همین ترتیب، ممکن است در شرایطی و با ملاحظات مجاز باشیم وارد جنگ شویم. معمولاً تعیین آن شرایط و ملاحظات در چارچوب نظریه جنگ عادلانه^۲ طرح و بحث می‌شود. اگر چه این تصویر، یگانه رویکرد به جنگ نیست، چارچوب غالب در مباحث جاری درباره جنگ است. بخش عمده‌ای از بحث‌ها درباره سلاح‌ها نیز ذیل اخلاق جنگ قرار دارد.

۱. توجه به تمایز بررسی حقوقی و اخلاقی نیز مهم است که در اینجا به آن نمی‌پردازیم. شاید کسی فکر کند – در نهایت – قوانین در این باره مهم هستند و بنابراین، اخلاق مهم نیست. فعلاً کافی است در نظر بگیریم که اساساً بنا بر اصل مارتنز – دستکم در این مباحث – اخلاق یکی از سرچشمه‌های مهم حقوق بین‌الملل بشردوستانه است (Sparrow, 2017). گذشته از آن، ما در نبود قانون یا برای تصمیم درباره قوانین و نیل به قوانین خوب، ناگزیر به ملاحظات اخلاقی متوسل می‌شویم. ضمن آنکه مباحث اخلاقی می‌توانند افراد را تحت تأثیر قرار بدهند و آنها نیز بر سازمان‌ها تأثیر دارند. یک نمونه عینی این موضوع، قطع همکاری گوگل در پروژه‌ای است که ابعاد نظامی داشت و این به سبب نگرانی‌های اخلاقی کارکنان بوده است (see Wakabayashi and Shane, 2018).

2. Just war



۱.۴. جنگ عادلانه و سلاح‌ها

دو بخش عمده نظریه جنگ عادلانه، عبارتند از عدالت پیش از جنگ و عدالت در جنگ.^۱ مقتضیات عدالت پیش از جنگ^۲ مواردی است نظیر اینکه هدف و انگیزه عادلانه در کار باشد (چنان که کشورگشایی نمی تواند جنگ عادلانه باشد) یا راه بدیل یا چاره دیگری جز جنگ در کار نباشد. این موارد کمتر به طور مستقیم به اخلاق سلاح‌ها به طور عام یا سامانه سلاح خودآیین به طور خاص مربوط می شوند.

مقتضیات عدالت در جنگ^۳ در دو اصل اساسی بیان می شود: تمایز و تناسب (برای نمونه، نک: Frowe, 2014, xiv). نظر به «اصل تمایز»، هر اقدامی در جنگ که مرز میان نظامیان و غیر نظامیان را نادیده بگیرد، محکوم است. «اصل تناسب» نیز شرطی راجع به اقدامات مجاز است. فرض کنید کسی به من حمله کرده و اگر اقدامی نکنم مرا می کشد. در این شرایط، دفاع از خود ضرورت دارد، اما باید با ابزاری متناسب از خودم دفاع کنم - اگر با هل دادن نجات می یابم، نیازی به صدمه جدی یا کشتن او نیست. پس باید ابزار مناسب و متناسب را برگزینم. ضمن آنکه ابزار مناسب بر گستره ابزارها حدی می گذارد؛ چنانکه تناسب بعضی رفتارها (مثل شکنجه) را کلاً نفی می کند. به عبارتی، این اصل نافی «هدف وسیله را توجیه می کند» است (Leveringhaus, 2016, p. 17).

روشن است که در چارچوب مقتضیات عدالت حین جنگ، می توان درباره سلاح‌ها نیز بحث کرد. برای مثال، سلاح کشتار جمعی به سبب نقض «تمایز» محکوم است و استفاده از آن غیر اخلاقی است. به همین ترتیب، استفاده از سلاحی که رنج نالازم ایجاد کند یا تبعات غیر ضروری داشته باشد، محکوم است. نمونه چنین منعی درباره سلاح‌های شیمیایی دیده می شود. حال باید دید در این چارچوب اخلاقی، درباره سامانه سلاح خودآیین چه می توان گفت.^۵

۱. گاه از عدالت پس از جنگ (jus post bellum) نیز بحث شده است. بررسی نسبت سامانه سلاح خودآیین با عدالت پس از جنگ بی راه نیست، اما می توان برای بحث اولیه از آن صرف نظر کرد.

2. jus ad bellum

3. jus in bello

۴. گاه به «ضرورت» به عنوان اصل سوم اشاره می شود (see Leveringhaus, 2016, p. 17).

۵. از قضا، چارچوب جنگ عادلانه در سیاق حقوق بین الملل بشردوستانه نیز مورد استناد و توجه است و اصول آن با آنچه در «اخلاق جنگ» بحث می شود، مشابه است. با این حال، حتی مفاهیم مشابه جنگ عادلانه در سیاق اخلاقی و حقوقی متفاوت عمل می کنند. برای مثال، از نظر اخلاقی، قابل فهم است که در بهترین حالت یکی از طرفین جنگ بر حق باشد و طرف دیگر برخطا، اما در چارچوب قانونی، مادام که طرفین به قوانین مربوط پایبند باشند، یکسان هستند (see Leveringhaus, 2016, pp. 26-27).

۲.۴. سامانه سلاح خودآیین و عدالت در جنگ

فرض کنیم که تصویری که از سامانه سلاح خودآیین عرضه می‌شود، سازگار است و ساخت چنین سلاح‌هایی ممکن است. آیا توسعه، ساخت و کار بست این سلاح‌ها اخلاقاً رواست؟

۴. ۲. ۱. استدلال له

نظر به قابلیت‌های هوش مصنوعی در عرصه سلاح‌ها، استدلال‌هایی به نفع سامانه سلاح خودآیین طرح شده‌اند. این استدلال‌ها نه تنها به کارگیری این سلاح‌ها در جنگ را اخلاقاً بی‌اشکال می‌دانند، بلکه استفاده از آنها را اخلاقاً ارجح می‌دانند.

طبعاً مواردی از کاربردهای هوش مصنوعی در این فضا مربوط به (مثلاً) نجات مجروحان جنگی یا درمان آنهاست. این موارد، اگر چه ارزشمند، مستقیماً در حوزه سامانه سلاح خودآیین قرار نمی‌گیرد. چنان‌که پیشتر اشاره شد، سامانه سلاح خودآیین ناظر به کار بست هوش مصنوعی در عملکردهای اصلی سلاح‌هاست. با این توضیح، می‌توان به این دو استدلال به نفع سامانه سلاح خودآیین اشاره کرد.^۱

الف) کاهش تلفات انسانی

- ۱) باید در جهتی بکوشیم که تلفات انسانی در جنگ را کاهش می‌دهد.
- ۲) توسعه، ساخت و استفاده از سامانه سلاح خودآیین، تلفات انسانی در جنگ را کاهش می‌دهد. پس، باید در جهت توسعه، ساخت و استفاده از سامانه سلاح خودآیین بکوشیم.

۱. در دفاع از سامانه سلاح خودآیین، گاه استدلال شده است که این سلاح‌ها «جدید» نیستند، یعنی تفاوت ماهوی با دیگر سلاح‌ها ندارند، اما این استدلال‌ها اگر موفق باشند، صرفاً نشان می‌دهند که سامانه سلاح خودآیین مانند دیگر سلاح‌های بلااشکال است و از این رو، اخلاقاً ایراد خاصی به آن وارد نیست. به این معنا که جدید نبودن استدلالی له این سلاح‌ها نیست. اما به فرض جدید نبودن این سلاح‌ها، همچنان ممکن است به جهات دیگری محل اشکال باشند. برای مثال، تصور کنید که در مقام دفاع، از تفنگ‌های ساخته‌شده با پرینتر سه‌بعدی استدلال شود: این تفنگ‌ها همانند دیگر تفنگ‌ها هستند و تفاوت ماهوی خاصی ندارند. همچنان می‌توان در عین پذیرش اینکه تفاوت ماهوی خاصی میان این سلاح‌ها نیست، استدلال کرد که تفنگ‌های ساخته‌شده با پرینتر سه‌بعدی از جهت دسترس پذیری یا قیمت پایین مسائل خاص خود را پیش می‌کشند. برای نمونه‌ای از بحث در باره جدید بودن / نبودن سامانه سلاح خودآیین (see Plain, 2019).



توضیح مقدمه^۲ این است که اگر سربازهای رباتیک را جایگزین سرباز انسانی کنیم، آنگاه طبعاً تلفات انسانی کاهش خواهد یافت. این استدلال، معطوف به تلفات سربازان خودی است،^۱ اما ممکن است، مقدمه دوم را ناظر به عملکرد بهتر سامانه سلاح خودآیین و فایده آن را کاهش آسیب به غیرنظامیان بدانیم. در این صورت، می توان استدلال دیگری طرح و صورت بندی کرد.

ب) عملکرد بهتر از انسان

۱) سامانه سلاح خودآیین حتی اگر بی نقص نباشد، عملکردی (مثلاً در رعایت اصل تمایز) بهتر از انسان دارد.

۲) اخلاقاً باید در جنگ از ابزاری استفاده کنیم که (مثلاً در رعایت اصل تمایز) عملکردی بهتر از انسان داشته باشد.

پس، اخلاقاً باید در جنگ از سامانه سلاح خودآیین استفاده کنیم.

در دفاع از مقدمه^۱، می توان به قابلیت های مفید و متنوع در طراحی ربات ها اشاره کرد. برای مثال، ربات را می توان به نحوی «فداکار» طراحی کرد، یعنی «بقا» یا «حفظ خود» جزو اهدافش نباشد (Arkin, 2010). در این صورت، ربات در مقایسه با انسان در حین عملکرد خود ریسک کمتری می کند؛ به صرف اینکه نگران جان خودش باشد، اقدام به قتل رقیب نمی کند. نمونه دیگر امکان رهایی ربات از عواطف است. حذف عواطف (خشم یا ترس) یا کاهش خطا (تشخیص رقیب از خودی یا نظامی از غیرنظامی) موجب می شود تا جنگ، تلفات کمتری در بر داشته باشد.

به نظر می رسد که هر دو استدلال پیش گفته در جهت واحدی حرکت می کنند. بنابراین، در ارزیابی آنها به نگرانی های مشترکی اشاره می کنم. نخست این که به نظر می رسد ممکن است مزیت های مذکور را بدون توسل به سامانه سلاح خودآیین تأمین کنیم. برای مثال، چه بسا بدون فرستادن سرباز انسانی به میدان جنگ و با هدایت از راه دور بتوان مزیت مطرح در استدلال اول را برآورده کنیم.^۲ در این صورت، نیازی نیست تصمیم درباره برخورد با افراد انسانی (مثلاً قتل) را به

۱. شاید این استدلال در راستای مقتضیات عدالت پیش از جنگ فهم شود. بدین ترتیب و با این اقدام، گویی به معنایی ورود به جنگ صورت نگرفته است. به عبارت دیگر، جنگیدن با این ربات ها نوعی «بدلیل» برای جنگ متعارف است و از همین رو، مادام که این گزینه ممکن باشد، جنگ متعارف «آخرین چاره» (last resort) نیست.

۲. در اینجا پیچیدگی دیگری هم وجود دارد که بررسی آنها موضوع مستقلی است. برای مثال، چه بسا تولید صنعتی این

عهدهٔ سلاح خودآیین بگذاریم، بلکه کافی است هوش مصنوعی را در عملکردهای غیر اساسی سلاح‌ها به کار ببندیم.

شاید مدافعان سامانهٔ سلاح خودآیین بتوانند این گونه به این اشکال جواب بدهند که اولاً، اگر چه کاربست هوش مصنوعی در عملکردهای غیر اساسی فواید بسیاری دارد، این موارد جایگزین هوش مصنوعی در عملکردهای اساسی نیستند. توضیح آنکه، برای مثال، سامانهٔ سلاح خودآیین می‌تواند مانع از قتل‌های مبتنی بر خشم یا انتقام شود. حال آنکه اگر از سلاح‌های کنترل از راه دور استفاده کنیم، همچنان فرد می‌تواند از سر انتقام یا خشم تصمیم به قتل بگیرد. ثانیاً، کنترل سلاح‌های از راه دور نیازمند نیروی انسانی توانمند (دارای توانایی فنی برای کاربری از سلاح و همچنین دارای توانایی قضاوت مناسب برای اتخاذ تصمیم‌های اخلاقی) است و فراهم کردن این نیرو ساده نیست. حال آنکه کاربست هوش مصنوعی در عملکردهای اساسی سبب می‌شود به این نیروی کار نیازمند نباشیم. ثالثاً، یکی از مزیت‌های استفاده از هوش مصنوعی «سرعت» است. بدین ترتیب، سامانهٔ سلاح خودآیین در میدان جنگ تصمیم‌های سریع می‌گیرد. حال اگر بنا باشد تصمیم‌های اساسی را به نیروی انسانی محوّل کنیم، احتمالاً زمان تصمیم‌گیری افزایش قابل توجه پیدا می‌کند و این امر ممکن است تبعاتی به دنبال داشته باشد.^۱

دو استدلال پیش‌گفته به نفع سامانهٔ سلاح خودآیین از حیث دیگری هم محل نقد هستند. به نظر می‌رسد این استدلال‌ها در چارچوب پیامدگرایانه طرح شده‌اند. حال آنکه اگر ملاحظاتی علیه سامانهٔ سلاح خودآیین داشته باشیم، آنگاه کاربرد آنها موجه نیست؛ حتی اگر به کاهش تلفات انسانی منجر شود. ما نمی‌توانیم برای نیل به هدفی خوب (کاهش تلفات انسانی از نیروهای خودی یا غیرنظامیان) به هر وسیله‌ای دست بزنیم. توضیح این اشکال بدین شرح است. دربارهٔ ارزیابی اخلاقی اعمال چند دیدگاه عمده وجود دارد. دسته‌ای مهم از این نظریات ذیل پیامدگرایی می‌گنجد. طبق پیامدگرایی، یگانه امری که در ارزیابی اخلاقی عمل دخیل است پیامدهای آن است. برای مثال، درست‌پیش‌روی شما دو قطار در حال حرکت هستند و در جلوی هر دوی آنها مانعی وجود دارد. اگر قطارها به مانع برخوردند، تمام سرنشینان آنها از بین

سلاح‌ها در بلندمدت، به سبب آسیب به محیط زیست، تلفاتی انسانی بیشتری به دنبال داشته باشد. این ملاحظات را کنار می‌گذارم؛ چون موارد محاسباتی و روش‌شناسانه دیگری را پیش می‌کشند.
۱. در عین حال، خواهیم دید که کسانی «فرصت تجدید نظر» را محوری می‌دانند. در این باره، استدلال چهارم علیه این سلاح‌ها را ببینید.



می روند. شما نمی توانید مانع را برطرف کنید، اما ممکن است بتوانید به یک لوکوموتیوران اطلاع دهید تا قطار را متوقف کند. حال، شما می دانید که یکی از قطارها ۱۰ سرنشین دارد و قطار دیگر ۱۰۰ سرنشین دارد و این تمام چیزی است که می دانید. در این شرایط، برای نجات کدام قطار تلاش می کنید؟ خیلی از ما سعی می کنیم قطاری را نجات دهیم که ۱۰۰ سرنشین دارد. اگر از ما دلیل انتخاب را بخواهند، می گوئیم در این مورد افراد بیشتری نجات یافته اند و خیر بیشتری حاصل شده است. حال، پیامدگرایان این شکل استدلال را می پذیرند و آن را تعمیم می دهند و سپس پا فراتر می گذارند و ادعا می کنند که این یگانه ملاحظه مربوط است.

اما پیامدگرایی، فارغ از انواع و اقسامش، مخالفان زیادی دارد. یکی از استدلال‌ها علیه پیامدگرایی، توسل به مثال‌های نقض است. برای مثال، این وضعیت را در نظر بگیرید: پنج نفر که همه از افراد تأثیرگذار یک کشور هستند در بیمارستان بستری شده اند و نیاز به اهدای عضو دارند، اما امیدی نمی رود که عضو مناسب برای آنها پیدا شود. حال، جوانی در نزدیکی بیمارستان هست که اعضای بدنش سالم است. بدین ترتیب، اگر پزشک‌ها او را بکشند و از اعضای بدنش استفاده کنند، می توانند جان پنج نفر را نجات دهند. آیا چنین اقدامی اخلاقی است؟

اگر پیامدگرایی درست باشد، یگانه ملاحظه مربوط این است که نتایج بهتری حاصل شود. اگر پزشک‌ها جوان را بکشند، یک نفر (جوان) کشته می شود و پنج نفر نجات می یابند، اما اگر جوان را نکشند، یک نفر (جوان) زنده می ماند و پنج نفر می میرند. اگر پیامدگرایی درست باشد، آنگاه پزشک‌ها باید جوان را بکشند و جان پنج نفر دیگر را نجات دهند (برای مثال، نک: Foot, 1967, p. 24)، اما، شهوداً، به نظر می رسد این کار اشتباه است و پزشک‌ها نباید جوان را بکشند. بنابراین، پیامدگرایی نادرست است.

منظور از نفی پیامدگرایی این است که افزون بر پیامدها ملاحظات دیگری هم دخیل هستند. نیل به نتایج بهتر ترجیح دارد، اما اخلاقاً روا نیست که برای نیل به نتایج بهتر به هر کاری دست بزنیم؛ چه اینکه بعضی افعال - مثل کشتن انسان بیگناه - فارغ از نتایج آن بد هستند. بنابراین، مدعای استدلال‌های فوق له سامانه سلاح خودآیین باید به این شکل تعدیل یافته در بیابند: در شرایط برابر، استفاده از راه‌ها یا وسایلی که تلفات انسانی را کاهش دهد ترجیح دارد، اما معلوم نیست در این باره مجاز به استفاده از هر وسیله‌ای باشیم. بعضی از مخالفان سامانه سلاح خودآیین استدلال می کنند که استفاده از سامانه سلاح خودآیین در معرض اشکالات اخلاقی است (که در زیر به مواردی از آنها اشاره شده است) و بدین ترتیب استفاده از سامانه سلاح

خودآیین اخلاقاً نارواست، حتی اگر به نتایج خوبی مانند کاهش تلفات انسانی منجر شود. بنابراین، استدلال له سامانه‌ها مستلزم بررسی بیشتر است.

۴. ۲. ۲. استدلال علیه

نگرانی‌ها درباره «سامانه سلاح خودآیین» را می‌توان از جهتی به دو دسته تقسیم کرد: دسته‌ای از اشکالات که موضوعی عام‌تر مربوط هستند و سامانه سلاح خودآیین نیز در ذیل آن قرار می‌گیرد و دسته‌ای از اشکالات که خاص سامانه سلاح خودآیین است. در اینجا هدف اشاره به چند نمونه از ملاحظات دسته دوم است. با این وصف، برای مثال، به این نگرانی نمی‌پردازیم که کاربست هوش مصنوعی در حوزه تسلیحات ممکن است موجب کاهش شغل شود یا این اشکال که به سبب استفاده از اینترنت یا شبکه‌ای که همه سلاح‌های خودآیین را متصل می‌کند، سامانه سلاح خودآیین در معرض تهدیدهای جدی است و این امر سبب می‌شود اتکا به آنها بسیار خطرناک باشد (Plain, 2019)، اما نگرانی‌هایی نیز در کارند که به طور خاص درباره سامانه سلاح خودآیین طرح می‌شوند و در آنها بناس است فرض بر امکان و توفیق ساخت سلاح‌های خودآیین باشد.

یک: فقدان قابلیت‌های انسانی

۱) اگر موجودی فاقد قابلیت‌های انسانی (مثل تجربه درد یا درک عواطف) باشد، ناتوان از برخورد انسانی^۱ است.

۲) سامانه سلاح خودآیین فاقد قابلیت‌های انسانی (مثل تجربه درد یا درک عواطف) است. پس، سامانه سلاح خودآیین ناتوان از برخورد انسانی است.

بنا به این استدلال، سامانه سلاح خودآیین اصولاً فاقد توانایی رفتار انسانی است، حال آنکه می‌دانیم از اصول پذیرفته‌شده در اخلاق جنگ این است که برخوردها و رفتارها انسانی باشد. بنابراین، کاربست سامانه سلاح خودآیین در جنگ، مصداق رفتار غیر انسانی است (HRW, 2018).

دو: فقدان قضاوت اخلاقی

۱) اگر موجودی ناتوان از قضاوت اخلاقی درست باشد، شرکت او در جنگ مجاز نیست.

1. Humane



۲) سامانه سلاح خودآیین ناتوان از قضاوت اخلاقی درست است. پس، شرکت سامانه سلاح خودآیین در جنگ مُجاز نیست.

پذیرش مقدمه (۱) دشوار نیست، گرچه می‌توان دربارهٔ چند و چون آن پرسید - بالاخره، مادام که مسئلهٔ ماروایی یا ناروایی اخلاقی ساخت و استفادهٔ سامانهٔ سلاح خودآیین باشد، پذیرفته‌ایم که اخلاق برایمان مهم است. در این صورت، انتظار داریم که سامانهٔ سلاح خودآیین، مطابق اخلاق رفتار کند و بنابراین، لازم است قادر به قضاوت اخلاقی باشد. برای مثال، مهم است که در تصمیم‌گیری تمایز میان انسان و شیء لحاظ شود، یا مهم است که در واکنش به مهاجم نوعی تناسب در کار باشد و اگر می‌شود بدون آسیب اساسی او را متوقف کرد از اقدامات سهمگین‌تر خودداری شود، اما چرا مقدمه (۲) را بپذیریم؟ از جمله توضیحاتی که اخیراً در بیانیهٔ تأثیرگذار دیدبان حقوق بشر (HRW, 2018) آمده این است که «سلاح‌های کاملاً خودآیین» در بعضی حوزه‌ها بسیار محدودیت دارند، از جمله در «تفسیر رفتارهای ظریف انسانی» یا فهم پیچیدگی‌های سیاسی، اجتماعی و زیست‌محیطی. حال آنکه قضاوت اخلاقی بدون این موارد امکان ندارد.

دربارهٔ دو استدلال اخیر چه می‌توان گفت؟ به نظر می‌رسد ضعف مهم این استدلال‌ها از این است که هر دو استدلال، مدعایی دربارهٔ وضع بالفعل سامانهٔ سلاح خودآیین طرح می‌کنند و به بعضی مقدمات کلان دربارهٔ امکان و امتناع ماشین‌هوشمند یا ماشین اخلاقی اشاره دارند.

به تعبیری، هر دو استدلال بالا را می‌توان در این جهت فهمید که اصولاً سامانهٔ سلاح خودآیین ممتنع است و هر سلاحی که به عنوان سلاح خودآیین، روانهٔ جنگ شود، عملاً فاقد خودآیینی است. این استدلال می‌تواند جذاب و قابل توجه باشد، اما به نوعی وابسته به مدعیاتی تجربی یا دست‌کم بحث‌های مستقل دربارهٔ هوش مصنوعی است. حال آنکه در اینجا بناست به ملاحظاتی پردازیم که می‌توانند با فرض سامانهٔ سلاح خودآیین سازگار باشند. بدین ترتیب، این استدلال‌ها، استدلالی علیه سامانهٔ سلاح خودآیین نیستند، بلکه علیه سلاح‌هایی هستند که ادعا شده است دارای خودآیینی هستند. از این حیث، ایدهٔ نهفته در این استدلال‌ها را می‌توان به شکل دیگری بیان کرد. در حال حاضر، سلاح‌هایی که خودآیین دانسته می‌شوند، در واقع، خودآیین نیستند. بنابراین، خطاست که تصمیم مرگ و زندگی انسان‌ها را به ماشین‌هایی واگذار کنیم که امکان تحلیل و تصمیم درست ندارند. صد البته، این استدلال - از حیث عملی - حائز اهمیت است.

اما اشاره به دو نکتهٔ دیگر هم مفید است. اولاً، شاید کسی بگوید آنچه مهم است الگوریتم انتخاب هدف است و در آن مرحله برنامه‌نویس متأثر از احساسات انسانی و عواطف آن الگوریتم

را نوشته است. چه نیازی است که خود سلاح هم چنان قابلیت داشته باشد؟ البته، باید توجه کرد که در این صورت، شاید دیگر موضوع صحبت سامانه سلاح خودآیین به معنای دقیق کلمه نباشد، بلکه سلاح‌هایی باشد که اصطلاحاً خودآیین نامیده می‌شوند. ثانیاً، می‌توان پرسید آیا انسان‌ها همواره در جنگ رفتار انسانی دارند و ظرافت‌ها را درک می‌کنند؟ خیر. البته، از نظر استدلال‌های پیش‌گفته، انسان امکان رفتار انسانی و درک ظرافت‌ها را دارد، اما همچنان اگر به ضعف‌های انسان آگاه باشیم، لازم است به این سؤال نیز توجه کنیم که آیا ممکن است ماحصل عملکرد سامانه سلاح خودآیین در مقایسه با عمل انسان‌ها بیشتر مطابق اخلاق باشد؟ در اینجا نظرم به تکنیک بنا بر وظیفه / مطابق وظیفه است (Kant, 2002, AK4 398). می‌توان با مثال معروف کانت این تکنیک را توضیح داد.

فرض کنید فروشنده اخلاقی آن است که برای رعایت انصاف و پرهیز ظلم قیمت مناسب و منصفانه‌ای برای کالاها تعیین کند. عمل این فروشنده (تعیین قیمت منصفانه) اخلاقی است و از روی وظیفه (نظر به ملاحظات اخلاقی) است. حال، ممکن است فروشنده دیگری نیز همین عمل (تعیین قیمت منصفانه) را انجام دهد، اما این بار ملاحظات او اخلاقی نباشند، بلکه او برای آنکه مشتریان بیشتری جلب کند و سود بیشتری ببرد، قیمت منصفانه‌ای برای کالاها تعیین می‌کند. در این مورد، اگر چه عمل او مطابق وظیفه است، اما بنا به وظیفه نیست. از نمودهای تفاوت این دو آن است که وقتی کسی عمل درستی را بنا به وظیفه انجام می‌دهد، او را می‌ستاییم، اما اگر کسی عمل خوبی را صرفاً مطابق وظیفه و مثلاً، به انگیزه‌های خودخواهانه انجام می‌دهد، او را نمی‌ستاییم. حال، اگر تنها در صورتی استفاده از سامانه سلاح خودآیین مجاز باشد که بتواند بنا به وظیفه عمل کند، آنگاه استدلال‌های بالا (به فرض صحت مقدمات) نشان می‌دهند که استفاده از سامانه سلاح خودآیین اخلاقاً مجاز نیست، اما این امری است که نیازمند بررسی است. بالاخره، مدافعان سامانه سلاح خودآیین، مدعی هستند که با چنین وضعی مواجه هستیم؛ افراد انسانی که امکان عمل بنا به وظیفه را دارند، اما در بسیاری از موارد چنین نمی‌کنند و ماشین‌هایی که امکان عمل بنا به وظیفه را ندارند، اما در بسیاری از موارد مطابق وظیفه عمل می‌کنند. بنابراین، دفاع از این دو استدلال علیه سامانه سلاح خودآیین، نیازمند تلاش بیشتری است.

سه: کرامت انسانی

۱) تصمیم مرگ و زندگی انسان از سوی ماشین (غیرانسان) ناقض کرامت انسانی و بنابراین، غیر اخلاقی است.



۲) سامانه سلاح خودآیین تصمیم درباره مرگ و زندگی انسان را به عهده می‌گیرد. پس، سامانه سلاح خودآیین ناقض کرامت انسانی و بنابراین، غیراخلاقی است.

در این استدلال پذیرش مقدمه (۲) دشوار نیست و بنا به توضیحی که درباره موضوع ارائه شد، «خودآیینی» ویژگی اصلی سامانه‌های مورد بحث است. البته، یکی از دشواری‌ها در بحث از سامانه سلاح خودآیین، تعیین مراد از «خودآیینی» است. توضیح این نکته دشوار است که چرا قابلیت‌های سامانه سلاح خودآیین را صرفاً فرایندهای بسیار پیچیده «خودکار» ندانیم که الگوریتم‌های آن را انسان نوشته است. در عین حال، به سبب ابعاد کلان این موضوع، نمی‌توان تا روشن شدن مطلب و توافق درباره منظور از «خودآیینی» بحث‌های بعدی را تعلیق کنیم، اما چرا باید بپذیریم که این امر می‌تواند به نقض کرامت انسان منجر شود؟ از ملاحظات که در بیانیه دیدبان حقوق بشر بر آن تأکید شده همین استدلال است (HRW, 2018). در آنجا ایده این است که «فاعل باید پیش از گرفتن جان کسی حقیقتاً ارزش زندگی انسان و اهمیت فقدان آن را بفهمد» و اگر جز این باشد، کرامت انسانی نقض شده است.

دوباره این استدلال را باید طوری بفهمیم که راجع به اصل امکان سامانه سلاح خودآیین نباشد. چه بسا سامانه سلاح خودآیین بتواند به چنان فهمی برسد، اما همچنان این استدلال مشکلاتی دارد. نخست این که مشخص نیست چه تفاوت معناداری بین گلوله، مین و ربات هست. چه چیزی سبب می‌شود کشته شدن به واسطه ربات را ناقض کرامت انسانی می‌کند؟ دوم، در اینجا چالش جدی‌تری نیز هست و آن توضیح مراد از «کرامت انسانی» است. آریادنا پاپ، استدلال کرده است که «کرامت» را به دو شکل می‌توان فهمید و هیچ کدام به توفیق این استدلال کمک نمی‌کند. بگذریم از اینکه استدلال بر اساس کرامت انسانی بسیار پرتکرار است، اما کمتر برای توضیح و توجیه آن تلاش شده است (Pop, 2018).

افزون بر اینها، نگرانی سومی هم در کار است و آن این است که باید دو خوانش از استدلال را تفکیک کرد: روایت «علی الاصول» و روایت «موردی». اگر اشکال علی الاصول باشد، بحث درباره این است که اصولاً سامانه سلاح خودآیین امکان چنین و چنان تأملاتی را ندارد؛ یا توسل به سامانه سلاح خودآیین اصولاً نمی‌تواند ناقض کرامت انسانی نباشد، ولی اگر مسئله را موردی نگاه کنیم، باید ببینیم چه تعداد از رزمنده‌های انسانی میدان جنگ به این ملاحظات توجه دارند؛ چه بسا عملاً انگیزه‌های انسانی نادرست (مثل خشم و نفرت) بیشتر ناقض کرامت انسانی باشد. یا مشابه نکته آخر درباره دو استدلال قبل، ترجیح داشته باشد که در این موارد گزینه‌ای را ترجیح بدهیم که بیشتر مطابق

وظیفه است، فارغ از این که نگران این باشیم که سلاح بنا بر وظیفه عمل کند. همچنین می توان گفت شاید تأمل مربوط درباره ارزش جان انسان در سطح طراحی الگوریتم و از سوی انسان ها لحاظ شده باشد. این مطلب مشابه فضای استدلال چهارم بر اساس شکاف مسئولیت است.

چهار: شکاف مسئولیت

(۱) بنیاد جنگ عادلانه بر پذیرش مسئولیت است.

(۲) استفاده از سامانه سلاح خودآیین سبب شکاف مسئولیت و پس نافی بنیاد مسئولیت است. پس، سامانه سلاح خودآیین بنیاد جنگ عادلانه را زیر سؤال می برد.

این یکی از معروف ترین استدلال ها علیه سامانه سلاح خودآیین است و از شاخص ترین مدافعان آن اسپارو (۲۰۰۷) است. مشابه این استدلال در سیاق حقوقی نیز طرح و به تفصیل بحث شده است. استدلال از ما می خواهد که فرض کنیم اقدام نادرستی صورت گرفته است. آیا باید خود ربات قاتل را مسئول دانست و احیاناً مجازات کرد؟ آیا مسئولیت به عهده سازنده است؟ آیا مسئولیت به عهده طراح است؟ آیا مسئولیت به عهده فرمانده جنگ است؟ مسئله این است که اگر نتوان کسی را مسئول دانست، زمینه بی اخلاقی فراهم می شود. و طبق استدلال، اصولاً ربات های قاتل موجب می شوند شکافی در اسناد مسئولیت ایجاد شود و نتوان مسئول اقدامات انجام شده را مشخص کرد. بالاخره، بنا به تعریف، این ربات ها خودآیین هستند؛ چنین نیست که از جانب کاربر انسانی به آنها فرمان داده شده باشد.

این استدلال اگر چه جذاب است و نکات مهمی درباره سامانه سلاح خودآیین را برجسته می کند، با چالش هایی مواجه است.^۱ نخست، باید بررسی شود که چه کسان و چیزهایی می توانند مسئول باشند. برای مثال، آیا سازمان یا گروه می تواند مسئول باشد؟ به نظر می رسد بررسی این سؤالات به پل زدن بر شکاف مسئولیت کمک می کنند. دوم، باید پیچیدگی های مربوط به شراکت^۲ یا عمل و مسئولیت جمعی^۳ را نیز در حساب آوریم. گویی در استدلال فرض

۱. لورینگهاوس، استدلال بر اساس شکاف مسئولیت را نقد می کند و سپس استدلال دیگری را علیه سامانه سلاح خودآیین پیش می نهد. به نظر او، اگرچه موجهی از مباحث راجع به ربات قاتل بر محور «مسئولیت» بوده است، باید توجه خود را به ملاحظات دیگری (یعنی، خطرات) معطوف کنیم (see Leveringhaus, 2016, ch. 3).

2. Complicity
3. Collective



شده است که در صورتی مسئولیت برقرار است که با یک فرد مشخص طرف باشیم. حال آنکه محتمل است که شکاف مسئولیت ادعایی در واقع، ناشی از تصویر نادقیق از مسئولیت باشد. بدین ترتیب، اگر به پیچیدگی‌ها و حالت‌های دیگر مسئولیت توجه کنیم، ممکن است مسئولیت در عملکردهای سامانه سلاح خودآیین مشابه مسئولیت در دیگر موارد باشد. نه این است که مسئولیت اقدامات سرباز در جنگ پیچیدگی‌های مشابهی را به میان می‌آورد؟ برای مثال، تصور کنید مافوقی به سرباز خود فرمان داده که در زمین مین کار بگذارد. کودکی به واسطه این مین دچار معلولیت می‌شود. آیا مسئول این اقدام صرفاً شخص سرباز است؟ آیا مافوق او نیز در این امر مسئول است؟ آیا فرماندهان اصلی جنگ نیز در این امر سهیم هستند؟ آیا کارخانه سازنده مین بخشی از مسئولیت را به عهده دارد؟ آیا طراحان آن کارخانه نیز دخیلند؟ آیا ممکن است کارگران آن کارخانه نیز شریک این اقدام باشند؟ البته، می‌دانیم قوانینی علیه استفاده از مین زمینی وضع شده‌اند، اما این نکته تأثیری در مثال ندارد. نظر به ادبیات گسترده مسئولیت جمعی، حتی ناظر به مسائلی مانند گرمایش زمین، به سادگی می‌توان وجوه جمعی مسئولیت در حوزه جنگ را دید.

سرانجام، کسانی بر آن هستند که شکافی در کار نیست و می‌توان در مورد سامانه سلاح خودآیین تعیین مسئولیت کرد. برای مثال، مسئولیت به عهده طراحان و برنامه‌نویسان است، یا به عهده کسانی است که سلاح را به جنگ می‌فرستند.

شاید مدافعان استدلال مذکور علیه سامانه سلاح خودآیین چنین پاسخ دهند: همه ملاحظات بالا یک نکته اصلی را نادیده گرفته‌اند و آن اینکه فرض بر آن بود که سامانه سلاح خودآیین دارای خودآیینی است و به معنایی خود مسئول اعمالی است که انجام می‌دهد.

در مقابل این دفاع، دو نکته می‌توان گفت. نخست، به نظر می‌رسد خودآیینی سامانه سلاح خودآیین با پیچیدگی وضعیت سازگار است. فرض کنید کشوری در جنگ از رزمنده‌هایی استفاده کند که قوانین جنگ را نمی‌دانند و از همین رو، به کشتار غیرنظامیان دست می‌زنند. در این وضعیت، اگر چه رزمنده‌ها خودآیینی دارند، همچنان درباره مسئولیت آنها سوالات مختلفی مطرح است؛ چقدر از مسئولیت این جنایت به عهده آنهاست و چقدر به عهده فرماندهان؟ آیا در این میان مریبی در آموزش قوانین جنگ به این رزمنده‌ها کاهلی کرده‌اند؟ آیا فرماندهان از جهل رزمنده‌ها به قوانین جنگ آگاه بوده‌اند؟ بدین ترتیب، ممکن است در عین این که سامانه سلاح خودآیین دارای خودآیینی است، سوالات خاصی را درباره مسئولیت به بار بیاورد و البته، با کاوش و بررسی بتوان به آن سوالات پاسخ گفت.

دوم، شاید منظور این باشد که این سلاح‌ها خودآیین هستند و خود مسئولیت عمل را به عهده دارند، اما در سیاق جنگ این کافی نیست که یک ماشین را مسئول عملش بدانیم و او را مجازات کنیم. برای مثال، گویی مجازات ماشین سبب نمی‌شود احساس کنیم که به نحوی جنایت انجام شده پاسخ داده شده است. برای مثال، تصور کنید که ربّاتی موجب قتل فردی بی‌گناه شده باشد. عجیب نیست اگر احساس انتقام در خانواده مقتول با از بین بردن آن ربّات فروکش نکند. در این مثال، منظور این نیست که حس انتقام یا موارد مشابه آن موجه هستند، بلکه بناست با استفاده از این مثال‌ها متوجه نوعی تفاوت میان سامانه سلاح خودآیین و انسان شویم که در جنگ لوازم اخلاقی به دنبال دارد، اما دوباره مسئولیت ماشین سبب نمی‌شود که مسئولیت از دیگر افراد / سازمان‌ها (مثل فرماندهان یا برنامه‌نویسان) سلب شود. بدین ترتیب، ممکن است در اینجا چالشی پیچیده‌تر علیه سامانه سلاح خودآیین باشد، اما نخست باید دید که آن چالش چیست.

پنج: فرصت تجدید نظر

(۱) استفاده از ابزاری که فرصت تجدید نظر ندهد اخلاقاً نارواست.

(۲) سامانه سلاح خودآیین فرصت تجدید نظر را می‌گیرد.

پس، استفاده از سامانه سلاح خودآیین اخلاقاً نارواست.

لورینگهاس (۲۰۱۶) معتقد است فاصله‌ای که به واسطه سامانه سلاح خودآیین ایجاد می‌شود، زمینه‌ساز حذف فرصت تجدید نظر است. برای مثال وقتی فرد در حین تصمیم به کشتن رقیب خود را نمی‌بیند، نوعی فاصله در میان است. یا وقتی فرد یکسره به لحاظ فرهنگی و روانی از رقیب خود بیگانه باشد و نتواند با او همدلی کند، نوعی فاصله در میان است. این فاصله سبب می‌شود فرد در تصمیم خود هیچ تردیدی نکند و اگر تصمیمی (مثلاً به قتل) گرفت، فرصتی برای تجدید نظر نداشته باشد. البته، همیشه نیازی به تجدید نظر نیست، اما نفس این فرصت خوب است. این چیزی است که در تجربه انسانی داریم. فردی که در جنگ مشارکت می‌کند همواره در معرض تجدید نظر است و می‌توانیم تصویر کنیم که در خیلی از موارد تجدید نظر فرد خوب است. بدین ترتیب، سامانه سلاح خودآیین این فرصت ارزشمند را حذف می‌کند و از این جهت اخلاقاً بد است.

این استدلال نیز همچون استدلال چهارم متمرکز بر سامانه سلاح خودآیین است و در امکان و امتناع آن چند و چون نمی‌کند. از این رو، استدلال جذابی است و نکاتی درباره سامانه سلاح



خودآیین را برجسته می‌کند. با این حال، دفاع از این استدلال نیز دشواری‌هایی دارد. نخست، باید توضیح داد که «فاصله» از کجا آغاز می‌شود. بجاست که پرسیم مگر استفاده از سلاح گرم (در مقایسه با سلاح سرد) نوعی فاصله نیست؟ و مگر استفاده از سلاح سرد (در مقایسه با خفه‌کردن با دست، چشم در چشم با مقتول) نوعی فاصله نیست؟ به عبارت دیگر، می‌توان پرسید که چه چیز سبب می‌شود فاصله‌ای که سامانه سلاح خودآیین ایجاد می‌کند مشکل‌ساز باشد؟ به عبارت دیگر، چرا باید مقدمه (۲) را بپذیریم؟ بالاخره، این استدلال بناست درباره سامانه سلاح خودآیین باشد، نه درباره «هر سلاحی». حال آنکه هر سلاحی به نوبه خود نوعی فاصله را به میان می‌آورد. مدافعان سامانه سلاح خودآیین می‌توانند بپرسند که چرا فاصله‌ای که سامانه سلاح خودآیین ایجاد می‌کند، اخلاقاً نارواست و فاصله‌ای که دیگر سلاح‌ها ایجاد می‌کنند، اخلاقاً ناروا نیست.^۱

دوم، به نظر می‌رسد این مسئله معرفتی است، یعنی، روشن نیست چرا باید مقدمه^۱ را بپذیریم و بر آن باشیم که تجدید نظر امر ارزشمندی است. طبعاً تجدید نظر برای ما انسان‌ها خوب و لازم است؛ چون خطا می‌کنیم و دنبال فرصت‌هایی برای کاهش خطا هستیم، اما اگر مشکل معرفتی در کار نباشد، چه نیازی به فرصت تجدید نظر است؟ پس باید دو حالت را تفکیک کنیم: الف) اخلاق سامانه سلاح خودآیین آرمانی؛ ب) اخلاق سامانه سلاح خودآیین خطاکار. استدلال درباره حالت «الف» صدق نمی‌کند. منشأ ارزشمندی تجدید نظر برای انسان به سبب خطاست؛ از آنجا که تجدید نظر سبب می‌شود که فرد متوجه بعضی خطاها شود و در نهایت، احتمال خطا کاهش یابد، تجدید نظر امر ارزشمندی است. حال، اگر فرض کنیم که انسانی در فعالیتی بسیار ماهر است و هیچ خطایی در کار او راه نمی‌یابد (یا مقدار خطا بسیار ناچیز و بی‌اهمیت است)، آنگاه نیازی به تجدید نظر ندارد. در واقع، اگر چنین کسی پس از هر تصمیم دوباره در آن تجدید نظر کند و از ابتدا این فرایند را تکرار کند، ممکن است وقت و انرژی

۱. بد نیست تصریح کنم که مدافع سامانه سلاح خودآیین نمی‌تواند از این ادعا نتیجه بگیرد که «بنابراین، سامانه سلاح خودآیین و فاصله‌ای که به دنبال دارد، اخلاقاً بی‌اشکال است». برای مثال، به نظر می‌رسد شهوداً کشتن جنین یک‌روزه چندان قبیح اخلاقی نداشته باشد، اما کشتن جنین نه‌ماهه یا نوزاد یک‌روزه اخلاقاً قبیح است. در اینجا، می‌توان پرسید که دقیقاً در چه مقطعی وضعیت جنین تفاوت می‌کند. این سؤال ارزش بررسی دارد. با این حال، اینکه بتوانیم نشان دهیم که دقیقاً از چه زمانی کشتن جنین قبیح می‌شود، نشان نمی‌دهد که کشتن جنین نه‌ماهه تفاوتی با کشتن جنین یک‌روزه ندارد. با این توضیح، مدافع سامانه سلاح خودآیین فقط می‌تواند مدعی شود که آن استدلال علیه سامانه سلاح خودآیین موفق نیست؛ نه اینکه سامانه سلاح خودآیین اخلاقاً رواست.

خود را تلف کرده باشد. به همین ترتیب، اگر سامانه سلاح خودآیین دچار خطا نشود، آنگاه به تجدید نظر نیاز ندارد. چه بسا، در این صورت، فقدان تجدید نظر، امر مثبتی باشد و بتوان له سامانه سلاح خودآیین از آن استفاده کرد. (بدین ترتیب، این ملاحظه نیز ذیل استدلال دوم له سامانه سلاح خودآیین قرار می‌گیرد). پس اگر استدلال درباره سامانه سلاح خودآیین آرمانی باشد، مقدمه (۱) ناپذیرفتنی است.

اما استدلال درباره حالت «ب» نیز کارآمد نیست، یعنی حتی اگر استدلال راجع به سامانه سلاح خودآیین خطا کار باشد، همچنان بیشتر نکته‌ای درباره چگونگی طراحی چنان سلاح‌هایی است. لازمه استدلال این است که اگر سامانه سلاح خودآیین خطا کار باشد، باید بکوشیم در آن امکان تجدید نظر را تعبیه کنیم. بدین ترتیب، به نظر می‌رسد استدلال پنجم بیشتر درباره مقتضیات اخلاقی طراحی است، نه درباره استفاده از سامانه سلاح خودآیین در جنگ. در آن صورت، اگر سامانه سلاح خودآیین به واقع توانایی‌های بالایی داشته باشد، چرا نتوان در آن قابلیت تجدید نظر را تعبیه کرد؟



۵. استدلال‌های ضعیف و شهودهای قوی

به نظر می‌رسد استدلال‌ها علیه توسعه، ساخت و کاربست سامانه سلاح خودآیین به سادگی قابل دفاع نیستند یا، دست‌کم، دفاع از آنها دشواری‌های خود را دارد. به علاوه، استدلال‌ها له سامانه سلاح خودآیین نیز با موانعی جدی روبه‌رو هستند. در عین حال، شهود بسیاری حکایت از ناروایی اخلاقی این سلاح‌ها دارد. چگونه می‌توان این وضعیت را تبیین کرد؟ یک پیشنهاد این است که منشأ ناروایی سامانه سلاح خودآیین را نه بر اساس اصول عدالت حین جنگ که در مواضع دیگری بجوییم.

پیشنهاد این است که هنگام تأمل درباره سامانه سلاح خودآیین، صرفاً در پی یافتن یا تحلیل پدیده‌ای خاص و جدید نباشیم، بلکه به این توجه کنیم که شاید سلاح‌های خودآیین همان پدیده‌های آشنا در حوزه جنگ را شفاف‌تر و بزرگ‌تر نشانمان می‌دهند. برای مثال، شکاف مسئولیت که درباره سامانه سلاح خودآیین بسیار برجسته شده، در سلاح‌های معمول نیز وجود دارند. سرباز مسئولیت کار خود را به مافوق خود ارجاع می‌دهد و این سلسله همین‌طور بالا می‌رود. چالش فاصله نیز در اینجا دیده می‌شود و دیرزمانی است که جنگ‌ها «از نزدیک»

نیستند. مسئله کرامت انسانی نیز وضع مشابهی دارد، اما همه این ملاحظات در مورد سامانه سلاح خودآیین به نحو روشن تری به چشم می‌آیند.

بدین ترتیب، اگر نکته محوری در استدلال‌ها علیه توسعه، ساخت و کاربست سامانه سلاح خودآیین یا درباره اصل امکان سامانه سلاح خودآیین باشد، یا اگر درباره سامانه سلاح خودآیین است درباره دیگر سلاح‌های متداول نیز صدق کند، آنگاه شاید منشأ ناروایی سامانه سلاح خودآیین امر دیگری باشد. شاید نیز بتوان از وضع فعلی به استدلالی برای حک و اصلاح نظریه جنگ عادلانه نقب زد. مطلب اخیر را می‌توان در قالب این استدلال بیان کرد:

(۱) اگر دیدگاه ما درباره اخلاق جنگ درست باشد، سامانه سلاح خودآیین را محکوم می‌کند.

(۲) دیدگاه ما درباره اخلاق جنگ، سامانه سلاح خودآیین را محکوم نمی‌کند.

پس، پس، دیدگاه ما درباره اخلاق جنگ درست نیست.

مرور انتقادی استدلال‌ها علیه سامانه سلاح خودآیین ما را به پذیرش مقدمه (۲) ترغیب می‌کند، اما چرا باید مقدمه (۱) را بپذیریم؟ چنین استدلالی می‌تواند کمک کند: اگر دیدگاه ما درباره اخلاق جنگ درست باشد، شهودهای ما را برآورده می‌کند و یکی از شهودهای ما این است که سامانه سلاح خودآیین باید محکوم شود. در اینجا از طرفی به واکنش افراد آشنا نظر دارم که البته، بیش از حد محدود است. در مقیاس کلان‌تر نیز توجهم به نظرسنجی‌هایی است که در مقیاس کلان انجام شده‌اند. جز چند کشور معدود، در همه کشورها اکثریت مردم معتقدند باید سلاح‌های خودآیین ممنوع باشند.^۱

البته، قصد این قسمت دفاع از استدلال اخیر نیست. برای دفاع از چنین استدلالی دو گام اساسی باید برداشته شود. نخست، باید دفاع مفصل تری از این ارائه داد که اخلاق جنگ سامانه

۱. در نظرسنجی‌ای از دانشمندان گفته شده است که دانشمندان علیه ربات‌های قاتل نوعی اجماع دارند. در سطح وسیع‌تر نیز - در مجموع - سه نفر از هر پنج نفر مخالف ربات‌های کُشنده هستند. این تحقیق در ۲۰۱۸ م انجام شده و بنا بر آن ۶۱ درصد مردم مخالف این سلاح‌ها هستند و ۲۲ درصد موافق؛ ۱۷ درصد نیز در این باره نامطمئن هستند. قبل‌تر هم در تحقیق دیگری این نتیجه به دست آمده بود که ۵۶ درصد مردم مخالف این سلاح‌ها هستند و ۲۴ درصد نامخالف؛ ۱۹ درصد هم نامطمئن. بدین ترتیب، با فرض اعتبار هر دو نظرسنجی، مخالفت با این سلاح‌ها در مجموع افزایش یافته است (HRW, 2019). گفتی است فقط در چهار مورد مخالفان این سلاح‌ها اکثریت نیستند: هند، اسرائیل، برزیل، و ژاپن. این نکته و دیگر جزئیات این نظرسنجی شایسته بررسی، تحلیل و تبیین است (see Semeniuk, 2019).

سلاح خودآیین را محکوم نمی‌کند. در اینجا - به اختصار - تنها به چند استدلال محدود اشاره شد که همه آنها در بخش اخلاق حین جنگ بودند. چه بسا استدلال‌های دیگری در کار باشند. دوم، باید از این دفاع کرد که نظرسنجی‌ها راجع به سامانه سلاح خودآیین را می‌توان در سیاق بحث حاضر نیز معتبر دانست. بالاخره، یک نگرانی مهم این است که - به فرض صحت داده‌ها و تحلیل آنها - تلقی عموم افراد از سامانه سلاح خودآیین نادقیق و آمیخته به افسانه باشد.^۱

در مقابل پیشنهاد پیش‌گفته، این نگرانی نیز وجود دارد که چه بسا بتوان استدلال‌ها علیه سامانه سلاح خودآیین را تقویت کرد یا استدلال‌های دیگری طرح کرد یا شاید شهود ما مشکل داشته باشد؛ نه سامانه سلاح خودآیین. در پاسخ، می‌توان گفت: ادعا این نیست که استدلال قاطعی علیه دیگر حالت‌های ممکن یا به سود رویکرد پیشنهادی عرضه شده است، بلکه هدف این است که رویکرد پیشنهادی را محتمل‌تر از دیگر رویکردها نشان دهیم. رویکردهای بدیل همان هستند که در اشکال آمده‌اند، یعنی ممکن است شهودهای ما مشکل داشته باشند و سامانه سلاح خودآیین بی‌اشکال باشد، اما این امکان همواره درباره شهودها صادق است و مادام که شاهد خوبی نداشته باشیم که در موردی خاص شهود مشکل دارد، چرا بنا را بر اشتباه بودن شهودها بگذاریم؟ همچنین ممکن است استدلال‌های دیگر یا تقویت‌شده‌ای علیه سامانه سلاح خودآیین داشته باشیم، اما بار دیگر باید چنان استدلالی عرضه شود و از آن در مقابل اشکال‌ها دفاع شود. به هر ترتیب، شاید بد نباشد در کنار استدلال‌های له و علیه به این رویکرد هم توجه شود که می‌توانیم آن را «رویکرد ذره‌بینی» بنامیم. در این صورت، «رَبات‌های قاتل» صرفاً ابعاد معمول جنگ را آشکارتر و صریح‌تر می‌سازند، بی‌آنکه چیز خاصی به آن اضافه کنند؛ آنها تنها ذره‌بین هستند.

۶. جمع‌بندی

بدین ترتیب، به نظر می‌رسد در کنار رویکردهای فنی و حقوقی به سامانه‌های تسلیحاتی خودآیین، رویکردی فلسفی (شامل رویکرد اخلاقی) ضرور و مفید است. از قضا، چه بسا بعضی مواضع فلسفی (مثلاً ملاک هوش یا شأن اخلاقی ربات‌ها) مقدم بر بحث‌های فنی یا حقوقی باشد. همچنین، چنان‌که دیدیم، کسانی به ملاحظاتی اخلاقی به این سامانه‌های

۱. برای بحثی موجز درباره ابعاد و مسائل روش‌شناختی رجوع به شهود افراد و به ویژه در چارچوب «اصل مارتنز» که اصل مهمی در حقوق بین‌الملل بشردوستانه است، نک: Sparrow, 2017.



فهرست منابع

تسلیحاتی خوش بین هستند و دیگرانی بدبین. همچنین، دیدیم که چطور با استدلال‌های مختلفی مواجه هستیم که بعضی از آنها دارای روایت‌های متفاوت هستند. هیچ‌یک از این استدلال‌ها به سادگی به نتیجه دلخواه خود نمی‌رسند و هر یک نیازمند دفاع‌های بیشتر هستند. هدف این بخش نه دفاع یا رد قاطع این استدلال‌ها، بلکه بیشتر توجه دادن به تعدادی از آنها بوده است تا وجوهی از چشم‌انداز اخلاقی به این پدیده را مرور کنیم.

گفتنی است که بحث حاضر در چارچوب نظریه جنگ عادلانه صورت گرفته است. ممکن است چند و چون در قواعد جافاتاده این چارچوب به نتایج متفاوتی درباره سامانه‌های تسلیحاتی خودآیین منتهی شود. در نهایت، به نظر می‌رسد تصمیم درباره‌ی روایی توسعه، ساخت و استفاده از این سلاح‌ها همچنین به بررسی‌های چندوجهی و میان‌رشته‌ای مورد به مورد این سلاح‌ها بستگی دارد.^۱

دکارت، رنه. (۱۳۴۴). گفتار در روش راه بردن عقل. ضمیمه در: فروغی، محمد علی. سیر حکمت در اروپا. ۱۳۴۴. تهران: انتشارات نیلوفر.

دکارت، رنه. (۱۳۹۱). تأملات در فلسفه اولی. (ترجمه: احمد احمدی، چاپ دهم). تهران: انتشارات سمت.

ریونزکرافت، ایان. (۱۳۸۷). فلسفه ذهن؛ یک راهنمای مقدماتی. (ترجمه حسین شیخ‌رضایی). تهران: انتشارات صراط.

Arkin, R. (2010). The Case for Ethical Autonomy in Unmanned Systems. *Journal of Military Ethics*, 9(4), 332-341.
<https://doi.org/10.1080/15027570.2010.536402>

Butterfield, L. (2018). *Leading academics reveal: What are we getting wrong about AI?*. Retrieved 2, 2022, from <https://www.research.ox.ac.uk/article/2018-10-15-leading-academics-reveal-what-are-we-getting-wrong-about-ai>

۱. این نوشته از نتایج کاری گروهی است که عمدتاً طی ۱۳۹۷-۱۳۹۸ با حمایت «بنیاد ملی نخبگان» در قالب یکی از طرح‌های شهید احمدی روشن با عنوان «سلاح‌های رباتیک خودگردان از منظر حقوق، فناوری و اخلاق» انجام شد. از همه اعضای این گروه بابت گفت‌وگوهای مفید ممنونم: دکتر پوریا عسکری (دانشیار گروه حقوق عمومی و بین‌الملل، دانشگاه علامه طباطبائی)، مدیر علمی و راهنمای این پروژه، دکتر وحید بذار و ناصر ندادف (دانش‌آموختگان حقوق) و نوید گلچین (دانش‌آموخته علوم کامپیوتر). ماحصل این همکاری در قالب یک کتاب تنظیم شده است و منتشر خواهد شد.

- Foot, P. (1967). The Problem of Abortion and the Doctrine of the Double Effect. *Oxford Review*, 5, 28-41.. Reprinted in Foot, Philippa. (2002). *Virtues and Vices and Other Essays in Moral Philosophy*. Oxford: Oxford University Press.
- Frowe, H., & Lang, G. (Eds.). (2014). *How We Fight: Ethics in War* (1st edition). Oxford: Oxford University Press.
- Gayle, Damienne. (2019). *UK, US and Russia among those opposing killer robot ban*. *The Guardian*, 29 March 2019: <https://www.theguardian.com/science/2019/mar/29/uk-us-russia-opposing-killer-robot-ban-un-ai>
- HRW. (2018). Heed the Call: A Moral and Legal Imperative to Ban Killer Robots. *Human Rights Watch*, <https://www.hrw.org/report/2018/08/21/heed-34call/moral-and-legal-imperative-ban-killer-robots>
- HRW. (2019). Poll Shows Strong Opposition to Killer Robots. *Human Rights Watch*, <https://www.hrw.org/news/2019/01/22/poll-shows-strong-opposition-killer-robots>
- Kant, I. & Schneewind, J. B. (2002). *Groundwork for the Metaphysics of Morals*. New Haven and London: Yale University Press.
- Leveringhaus, A. (2016). *Ethics and Autonomous Weapons*. London: Palgrave MacMillan.
- Plain, S. (2019). Autonomous weapons and the new laws of war. *The Economist*, 1 January 2019: <https://www.economist.com/briefing/2019/01/19/autonomous-weapons-and-the-new-laws-of-war>
- Pop, A. (2018). Autonomous weapon systems: A threat to human dignity? *ICRC blog*, 10 April 2018: <https://blogs.icrc.org/law-and-policy/2018/04/10/autonomous-weapon-systems-a-threat-to-human-dignity/>
- Semeniuk, I. (2019). Scientists call for ban on lethal, autonomous robots. *The Globe and Mail*, 2019: <https://www.theglobeandmail.com/business/technology/science/article-scientists-call-for-ban-on-lethal-autonomous-robots/>
- Sparrow, R. (2007). Killer robots. *Journal of applied philosophy*. 24(1), 62-77
- Sparrow, R. (2017). Ethics as a source of law: The Martens clause and autonomous weapons", *ICRC blog*, 14 November 2017: <https://blogs.icrc.org/law-and-policy/2017/11/14/ethics-source-law-martens-clause-autonomous-weapons/>
- Wakabayashi, D. & Shane, S. (2018). Google Will Not Renew Pentagon Contract That Upset Employees. *The New York Times*, 1 June 2018: <https://www.nytimes.com/2018/06/01/technology/google-pentagon-project-maven.html>

