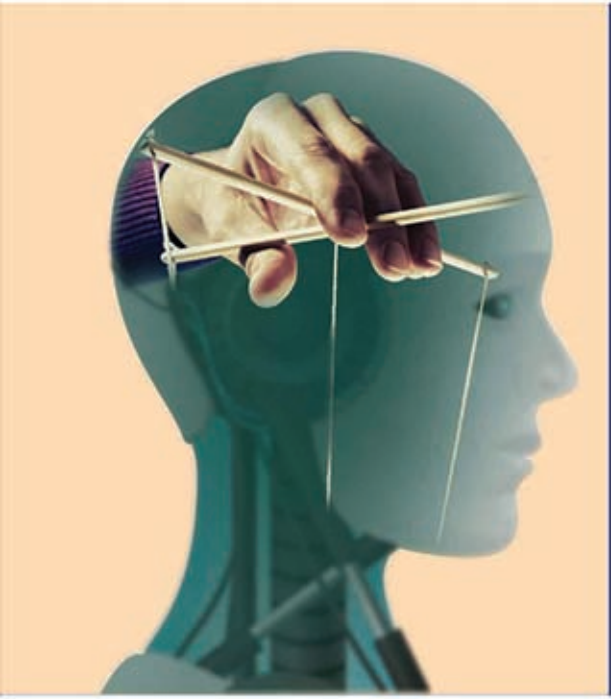
مرکز ملی فضای مجازی
پژوهشگاه فضای مجازی

عصر فضای مجازی
نود و چهارم



هوش مصنوعی و شکاف مسئولیت

Artificial intelligence and responsibility gap

بسم الله الرحمن الرحیم

عصر فضای مجازی

گزارش شماره ۹۴

بهمن ۱۴۰۰

مرکز ملی فضای مجازی
پژوهشگاه فضای مجازی

هوش مصنوعی و شکاف مسئولیت

محتوای انتشار یافته در این اثر
الزاماً بیانگر دیدگاه مرکز ملی فضای مجازی نیست

تهیه شده در پژوهشگاه فضای مجازی
(گروه مطالعات بنیادین فضای مجازی)

تهیه کننده: دکتر امیرحسین زادیوسفی (عضو هیئت علمی گروه فلسفه دانشگاه تربیت مدرس)

ناظر علمی: دکتر حسین مطلبی کربکندی و دکتر علیرضا کاظمی

نشانی: تهران، میدان آرژانتین، خیابان بیهقی، نبش خیابان ۱۶ غربی، پلاک ۳۰
تلفن: ۸۶۱۵۱۰۶۱-۰۲۱
کد پستی: ۱۵۱۵۶۷۴۳۱۱

فهرست

سخن نخست

سخن نخست

فضای مجازی با شتاب شگرف و رو به تزایدی که در حال بسط و گسترش است تمام ساحات اجتماعی، اقتصادی، سیاسی و فرهنگی زندگی بشر را درنوردیده و هر روز بخش بزرگی از زندگی واقعی را در خود فرو برده و حیات متفاوت و جدیدی به آن می‌دهد. لذا به نظر می‌رسد دو نگاه کلان به فضای مجازی وجود دارد: نگاه اول که بالاخص در ابتدای رشد و تکوین فضای مجازی مسلط شده بود، آن را همچون ابزاری کنار سایر ابزارهای بشری تصویر می‌کرد که تنها طریقیت داشت. اما نگاه دوم، در نتیجه رشد تحولات خیره‌کننده فضای مجازی و سایه گستری آن در حوزه‌ها و شئون بشر در یک دهه اخیر آن را چون سکویی می‌داند که بسیار فراتر از شأن ابزاری حیات انسان‌ها را سامان جدیدی داده و ادعای تمدن نوینی را دارد. رویکردی که از قضا از چشمان بصیر رهبر انقلاب نیز دور نمانده و انتظاری تمدنی از فضای مجازی در ایران را مطالبه داشته‌اند.

در همین راستا گزارش‌های عصرفضای مجازی تلاش می‌کند تا فهم سازمان‌ها و دستگاه‌های مرتبط با حوزه فضای مجازی را ارتقاء بخشیده و آن‌ها را برای مواجهه فعال و خردمندانه با تحولات این عرصه مهیا سازد.

سید ابوالحسن فیروزآبادی
دبیر شورای عالی و رئیس مرکز ملی فضای مجازی

چکیده

چکیده

اساساً چه کسی یا چه چیزی مسئول اقدامات و افعال هوش مصنوعی است؟ این پرسش، مسئله‌ای را در ادبیات هوش مصنوعی به وجود آورده است که از آن به عنوان «مسئله شکاف مسئولیت» یاد می‌شود. این نوشتار قصد دارد به بررسی ابعاد مختلف مسئله شکاف مسئولیت بپردازد. در ابتدا سعی می‌کنیم به معرفی مسئله شکاف مسئولیت بپردازیم. در این راستا، در ابتدا هوش مصنوعی خودمختار در حوزه‌های مختلف از جمله، اتومبیل‌های خودران، سیستم‌های تسلیحاتی خودمختار و سیستم‌های خودمختار در پزشکی را معرفی کرده و سعی خواهیم کرد تصویری را از مسئله شکاف مسئولت ارائه دهیم. در گام بعد، به پاسخ‌های مختلفی که به مسئله شکاف مسئولیت داده شده است اشاره می‌کنیم. در گام بعد و از آنجا که مسئله مسئولیت رابطه وثیقی با جنبه‌های حقوقی دارد، از نگاه حقوقی و مجازات کیفری به مسئله شکاف مسئولیت نگاه خواهیم کرد و در این راستا به بررسی سه کشور آمریکا، آلمان و ایران خواهیم پرداخت. سپس، ملاحظاتی فلسفی را در باب مسئله شکاف مسئولیت مطرح می‌کنیم. در انتهای کار، پیشنهادهایی راهبردی برای

مدیـران و تصمیم‌گیـران عرصـه هـوش مصنوعـی جمهـوری اسـلامی ایـران ارائـه خواهـد شـد.

واژگان کلیــدی: هـوش مصنوعی، هـوش مصنوعی خودمختـار، اتومبیل‌های خودران، مسـئله شـکاف مسئولیت

مقدمه

مقدمه

اگرچه گسترش و پیشرفت روزافزون سیستم‌های هوشمند امکانات فراوانی را در اختیار انسان‌ها قرار داده است، ولی خود مستلزم به‌وجودآمدن مسائل فلسفی، اجتماعی، اقتصادی و اخلاقی مختلفی شده است. برای مثال، اینکه آیا بشر قادر است در آینده‌ای نه چندان دور به ساخت سیستم‌های هوشمندی که مانند انسان‌ها دارای خودآگاهی باشند نائل آید یا خیر نمونه‌ای از سؤالات و چالش‌های فلسفی است. و یا اینکه توسعه سیستم‌های هوشمند چه تأثیراتی بر روی مشاغل خواهند گذاشت و اینکه آیا توسعه این سیستم‌ها در آینده سبب بیکاری مفرط در سطح جهان خواهد شد نمونه‌ای از چالش‌هایی است که در سطح اقتصادی مطرح می‌شود. و یا برای مثال، اینکه هوش مصنوعی چه تأثیری بر روابط بینافردی انسان در اجتماع دارد نمونه‌ای از سوالات و چالش‌ها در حوزه اجتماعی است. در این بین، یکی از چالش‌ها و پرسش‌های مهم فلسفی-اخلاقی (و مرتبط با مسائل حقوقی) که ظهور سیستم‌های هوشمند پیش می‌کشد، مسئله‌ای است که به «شکاف مسئولیت»[1] موسوم است. در این گزارش قصد داریم به این مسئله بپردازیم.

1. Responsibility Gap

ساختار گزارش ازاین‌قرار است. در بخش اول، سیستم‌های هوشمند خودمختار معرفی می‌شوند. در بخش دوم و با کمک از بخش اول، مسئله شکاف مسئولیت را طرح می‌کنیم. در بخش سوم به رویکردهای مختلف در باب مسئله شکاف مسئولیت اشاره خواهیم کرد. پس از آن، سعی خواهیم کرد از نظر حقوقی به مسئله شکاف مسئولیت نگاهی بیندازیم. در پایان، پس از ذکر برخی ملاحظات پیرامون مسئله شکاف مسئولیت، چند پیشنهاد راهبردی برای استفاده مسئولان و تصمیم‌گیران این عرصه در کشور ارائه خواهد شد. در نهایت، گزارش را با یک جمع‌بندی به پایان می‌بریم.

بخش اول

هوش مصنوعی خودمختار

بخش اول
هـوش مصنوعـی خودمختار

۱-۱- ماشین‌های خودران

اتومبیل‌هـای خـودران[1] کـه بعضـاً بـا نام‌هـای دیگـری ماننـد وسـایل نقلیـه خودمختـار،[2] اتومبیل‌هـای بـدون راننـده[3] نیـز از آنهـا یـاد می‌شـود، اتومبیل‌هایـی هسـتند کـه بـدون نیـاز بـه راننـده بـه مقصـد مـورد نظـر رانندگـی کـرده و بـه آنجـا می‌رسـند. اتومبیل‌هـای خـودران وسـایل نقلیه‌ای هسـتند کـه دارای سنسـورهای متعـددی بـرای درک محیـط پیرامـون خـود هسـتند. بـرای مثـال، ایـن وسـایل نقلیـه مجهـز بـه سیسـتم‌های رادار، جـی. پـی. اس، سیسـتم‌های اندازه‌گیـری مسـافت[4] و ... هسـتند. ایـن خودروهـا بـا داشـتن مقصـد مـورد نظـر بهتریـن مسـیر بـرای رسـیدن بـه آن را تشـخیص داده، و بـا داشـتن نقشـه خیابان‌هـا بـه سـمت آنجـا حرکـت می‌کنـد. ایـن خودروهـای مبتنـی بـر هـوش مصنوعـی ماننـد یـک انسـان، موانـع پیش‌رویشـان را تشـخیص داده و از ایـن رو در صـورت وجـود یـک مانـع بـر سـر راهشـان توقـف می‌کننـد. اولیـن ماشـین خـودران در دهـه ۱۹۸۰ توسـط محققـان دانشـگاه کارنگـی ملـون[5] سـاخته شـد و سـاخت آنهـا در طـول زمـان بـا پیشـرفت‌های خوبـی مواجـه شـده اسـت. اگرچـه ایـن اتومبیل‌هـا امـروزه به‌صـورت گسـترده مـورد اسـتفاده نیسـتند، ولـی

1. Self-driving car
2. autonomous vehicle
3. driverless car
4. measurement units
5. Carnegie Mellon

دور نخواهد بود آن روزی که از آنها در سرتاسر دنیا به‌صورت گسترده استفاده شود. برای مثال، شهرهایی مانند پکن،[1] شانگهای[2] و شنزن[3] در چین آزمایش‌های ماشین‌های خودران را در برخی از جاده‌های خاص آغاز کرده‌اند. نکته مهمی که سبب شده است به این اتومبیل‌ها، خودمختار (خودران) اطلاق شود، این است که این وسایل می‌توانند بر اساس الگوریتم‌های یادگیری عمیق، از تجربه بیاموزند.

۱-۲- سیستم‌ها تسلیحاتی خودمختار

سیستم‌های تسلیحاتی خودمختار به سیستم‌هایی اطلاق می‌شود که خود به‌صورت مستقل و بدون دخالت و کنترل عامل انسانی، بر اساس برنامه‌هایی که به آن داده شده است، اهداف نظامی را جستجو، شناسایی و در نهایت به سمت آنها شلیک کند. امروزه، بسیاری از سیستم‌های تسلیحاتی نیمه خودمختار در عملکرد خود وابستگی‌ای به عامل انسانی دارند. در واقع، در این سیستم‌های نیمه خودمختار این عامل انسانی است که باید شلیک به سمت هدف را تایید کند. اما در مقابل، سیستم‌های تسلیحاتی خودمختار بدون دخالت عامل انسانی می‌توانند از طریق برنامه‌هایی که از پیش برای آن نوشته شده است اهداف خود را شناسایی کرده و وارد عمل شود. در این سیستم‌ها از تکنولوژی‌هایی از قبیل تشخیص چهره، مسیریابی خودکار و ... استفاده شده است. سیستم‌های تسلیحاتی خودمختار، همانند ماشین‌های خودران، از الگوریتم‌های یادگیری عمیق بهره برده و لذا می‌توانند از تجربه بیاموزند. از این سیستم‌های تسلیحاتی خودکار می‌توان در نبردهای هوایی، نبردهای زمینی و نبردهای

1. Beijing
2. Shanghai
3. Shenzhen

دریایی استفاده کرد. برای مثال، پهپادهایی که می‌توانند اهداف نظامی دشمن را شناسایی کرده و آنها را مورد هدف قرار دهند، از این دست سیستم‌های تسلیحاتی خودمختار هستند.

۱-۳- سیستم‌های پزشکی

سیستم‌های هوشمند در حوزه پزشکی نیز کاربردهای فراوانی دارند. در یک تقسیم‌بندی می‌توان سیستم‌های هوشمندی را که در حوزه پزشکی به کار می‌رود را به دو گروه سیستم‌های خبره و سیستم-های مبتنی بر یادگیری عمیق تقسیم کرد. سیستم‌های خبره سیستم‌هایی هستند که خود یاد نمی‌گیرند بلکه بر اساس پایگاه دانشی که برای آنها تعریف شده است و با استفاده از منطق اگر-آنگاه به پزشکان در تشخیص و درمان بیماری‌ها کمک می‌کنند. اما در مقابل، سیستم‌های مبتنی بر یادگیری عمیق در پزشکی خود می‌توانند از تجربه بیاموزند. از این سیستم‌ها برای خواندن عکس‌های رادیولوژی، تشخیص سرطان و ... استفاده می‌شود. نکته قابل توجه در ارتباط با سیستم‌های هوشمند پزشکی مبتنی بر یادگیری عمیق، برخلاف سیستم‌های خبره، این است که برنامه‌نویسان نمی‌توانند دریابند که سیستم هوشمند چگونه به تصمیمی نائل شده است. این مسئله در واقع امری ذاتی برای همه سیستم‌های مبتنی بر یادگیری عمیق است.

بخش دوم

مسئله شکاف مسئولیت

بخش دوم

مسئله شکاف مسئولیت

حال که در بخش قبل نمونه‌هایی از سیستم‌های هوشمند خودمختار را ارائه کردیم، اجازه دهید به چند سناریوی مهم اشاره کنیم. برای مثال، اتومبیل خودران را در نظر بگیرد. مثلا یک نفر در یک روز بهاری برای خرید نان از خانه خارج شده و در هنگام عبور از خیابان به یک خودروی خودران که نتوانسته است او را به‌عنوان یک عابر پیاده تشخیص دهد برخورد کرده و در اثر شدت ضربه‌ای که به مغزش وارد شده است، فوت می‌کند.

حال، یک سیستم تسلیحاتی خودمختار را در نظر بگیرید. در جنگی که میان دو کشور وجود دارد، یکی از طرفین نزاع، از پهپادهای خومختار استفاده می‌کند. این پهپاد طوری ساخته شده است که می‌تواند نیروهای دشمن را شناسایی، ردگیری و منهدم نماید. برای مثال، فرض کنید این پهپاد هوشمند می‌تواند افرادی که در دست خود تفنگ حمل می‌کنند را به‌عنوان دشمن تشخیص داده و منهدم نماید. حال، فرض کنید یکی از روستاییانی که خانه‌اش در منطقه درگیری قرار دارد، صبح زود برای کار در مزرعه کشاورزی-اش از خواب برخاسته و بیل خود را بر روی دوش قرار داده و به سمت مزرعه

حرکت می‌کند. اما متاسفانه، یکی از پهپادهای دشمن او را به‌عنوان سربازی که سلاح خود را بر دوش قرار داده است شناسایی کرده و او را، که فردی غیرنظامی است، به قتل می‌رساند.

حال یک سیستم هوشمند در پزشکی را در نظر بگیرید. فرض کنید یک فرد که مدتی است که دردهایی را در بدن خود احساس می‌کند برای مداوا به پزشک مراجعه می‌کند. از قضا، پزشکی که وی به او مراجعه کرده برای تشخیص و ارائه راهی برای درمان، از سیستم‌های هوشمند بهره می‌برد. ازاین‌رو، سیستم هوشمند پس از اسکن بدن وی، و بر اساس الگوریتم‌هایی که برای ما نامعلوم است، به این نتیجه می‌رسد که او به سرطان بدخیمی مبتلا است. این سیستم هوشمند به پزشک پیشنهاد می‌کند که تنها راه درمان و جلوگیری از پیشروی سرطان وی، قطع عضو سرطانی است. از این‌رو، پزشک باتوجه‌به اینکه اعتماد بالایی به سیستم هوشمند دارد، به توصیه آن گوش فرا داده و عضو سرطانی را قطع می‌نماید. اما متاسفانه در واقع سیستم هوشمند به اشتباه تشخیص داده است که آن فرد به سرطان بدخیم مبتلاست. آنچه در واقع موجب احساس درد آن فرد شده بود، تنها چند غده خوش‌خیم بوده که می‌شد به‌راحتی و بدون قطع کردن عضوی، از بدن وی خارج شوند.

حال، باتوجه‌به مثال‌های فرضی فوق این سؤال اساسی مطرح است که در هر یک از سناریوهای فوق، مسئولیت نتایج حاصل آمده برعهده چه کسی است؟ به بیان دیگر، در سناریوی اول که در آن فردی توسط یک ماشین خودران کشته شده است، چه کسی مسئول کشته شدن وی است؟ و یا در سناریوی دوم که در آن یک

کشاورز ساده به اشتباه به جای یک نیروی نظامی به قتل رسیده است، چه کسی مسئول کشته شدن وی است؟ و یا در سناریوی سوم، چه کسی مسئول قطع شدن بدون وجه عضوی از بدن است (درحالی‌که فی‌الواقع نیازی به قطع آن نبوده است)؟ این سوال، سوالی است که اندریاس ماتیاس[1] آن را مسئله «شکاف مسئولیت» می‌نامد (Matthias, 2004). از نظر ماتیاس ماشین‌های خودمختار قادر به یادگیری هستند، به این معنی که اقدامات آنها غیر قابل پیش‌بینی و غیر قابل کنترل می‌شوند. ماتیاس پس از ارائه مثال‌هایی بیان می‌کند که همه این مثال‌ها در این ویژگی مشترک هستند که: یک الگوریتم به دستگاه اجازه می‌دهد تا از تجربه بیاموزد. این بدان معنی است که رفتار دستگاه با گذشت زمان تغییر می‌کند. این به نوبه خود به این معنی است که برنامه‌نویس نمی‌تواند کارهایی را که ماشین در یک موقعیت خاص انجام می‌دهد کنترل کند، زیرا اساساً رفتار ماشین باتوجه‌به تجربه آن تعیین می‌شود. تجارب مختلف می‌تواند به رفتارهای مختلف و حتی فاجعه‌بار منجر شود ازآنجاکه این رفتار به تجربه بستگی دارد و برنامه‌نویسان نمی‌توانند از قبل آنها را بدانند، پیش‌بینی رفتار امکان‌پذیر نیست. در واقع باید گفت، سیستم‌های مبتنی بر یادگیری عمیق به گونه‌ای هستند که حتی خود طراحان این سیستم‌ها درنخواهند یافت که سیستم هوشمند چگونه به نتیجه‌ای خاص رسیده است. به بیان دیگر، می‌توان سیستم‌های مبتنی بر یادگیری عمیق را به یک جعبه سیاه تشبیه کرد.

از نظر ماتیاس معمولاً، هنگامی‌که یک عامل[2] از واقعیت‌های خاص

1. Andreas Matthias
2. agent

مربوط به یک اقدام یا نتیجه اطلاع ندارد، یا عامل، کنترل محدودی بر یک موقعیت دارد، در این صورت آن عامل مسئولیت کمتری نیز در قبال آنچه اتفاق افتاده است، دارد. بنابراین، ماتیاس یک معضل پیش روی ما قرار می‌دهد. یا از ماشین‌های یادگیری خودمختار استفاده نکنیم که پیشنهاد جذابی نیست زیرا استفاده از این سیستم‌ها انواع مزایا را برای ما انسان‌ها فراهم می‌کنند. یا اینکه با مسئله شکاف مسئولیت روبرو خواهیم بود. باتوجه‌به اینکه اتومبیل‌های بدون راننده و همچنین برخی از ابزارهای تشخیص پزشکی در گروه ماشین‌های یادگیری خودمختار قرار می‌گیرند و ما به طور معمول می‌خواهیم کسی را در هنگام اشتباه این ماشین‌ها مسئول بدانیم، معضلی که ماتیاس پیش روی ما قرار می‌دهد یک مسئله مهم است (Matthias 2004).

بخش سوم

رویکردهای مختلف به مسئله شکاف مسئولیت

بخش سوم

رویکردهای مختلف به مسئله شکاف مسئولیت

در مواجهه با مسئله شکاف مسئولیت میان فیلسوفان و متفکران اقوال مختلفی وجود دارد. شاید بتوان در یک تقسیم ابتدایی، این‌طور بیان کرد که برخی بر این باورند که شکاف مسئولیت به‌راحتی قابل پر شدن نیست. از نظر این گروه استفاده از چیزی به نام سیستم هوشمند خودمختار سبب می‌شود که نتوانیم کسی را مسئول خطاهای سیستم هوشمند بدانیم. این گروه پیشنهاد می‌کنند که باید از استفاده از سیستم‌های هوشمند دست کشید. برخی دیگر بر این باورند که مسئله شکاف مسئولیت قابل پاسخگویی بوده و می‌توان جوابی را برای آن مهیا کرد. این گروه در باب اینکه چه کسی مسئول است به دو بخش تقسیم می‌شوند. برخی بر این باورند که انسان، اعم از برنامه‌نویس و یا کاربر، مسئول است. برخی دیگر بر این عقیده‌اند که خود سیستم هوشمند مسئول است. در ادامه قصد داریم نگاهی به هر یک از این آراء بیندازیم.

۳-۱- دیدگاه انکار مسئول

شاید در بادی امر به نظر برسد که این برنامه‌نویسان و طراحان

سیستم‌های هوشمند هستند که باید در قبال خطاهای پیش‌گفته مسئول شناخته شوند. اما این ایده مخالفانی دارد. یکی از مخالفان این ایده رابرت اسپارو[1] است (Sparrow, 2007). اجازه دهید ببینیم که اسپارو چگونه استدلال می‌کند؛ از نظر اسپارو برنامه‌نویسان کنترلی بر روی اینکه سیستم‌های هوشمند ساخته دستشان چگونه عمل خواهند نمود ندارند. همان‌طور که در بخش‌های قبل توضیح دادیم، این سیستم‌های هوشمند مبتنی بر یادگیری عمیق بوده که خود به این معنی است که حتی طراحان و برنامه‌نویسان آنها نیز از اینکه سیستم هوشمند چگونه به تصمیمی رسیده است بی اطلاع‌اند. از نظر اسپارو نمی‌توان فردی را که نمی‌تواند کنترلی بر روی عملی داشته باشد را مسئول آن عمل دانست. اسپارو در ابتدا سناریوی مشابهی را که در بالا ارائه شد طراحی می‌کند. سناریو این است: یک سیستم تسلیحاتی هوشمند خودمختار دسته‌ای از سربازان دشمن را که سلاح‌های خود را بر روی زمین گذاشته و تسلیم شده‌اند به گلوله بسته و همه آنها را می‌کشد. فرقی نمی-کند دلیل سلاح خودمختار برای کشتن سربازان چه بوده است، مسئله این است که در اینجا سربازانی کشته شده‌اند که اعلام کرده‌اند قصد دارند تسلیم شوند و بر اساس اخلاق جنگ نباید سربازانی که قصد تسلیم شدن را دارند هدف قرار داد و آنها را کشت.

اسپارو سپس این سؤال را مطرح می‌کند که در سناریوی بالا، چه کسی را باید مسئول دانست. گزینه-هایی که وی مطرح می‌کند عبارتند از (۱) خود ربات، (۲) برنامه‌نویسان آن، (۳) افسری که دستور استفاده از آن را داده است یا اینکه هیچ‌کسی مسئول نیست. وی نشان

1. Robert Sparrow

می‌دهد که هر یک از این گزینه‌ها با مشکلاتی روبرو خواهند شد.

در سناریوی بالا، آیا برنامه‌نویسان سیستم تسلیحاتی خودمختار مسئول آنچه اتفاق افتاده هستند؟ پاسخ اسپارو به این سؤال خیر است. اسپارو بیان می‌کند تنها در صورتی می‌توان برنامه‌نویسان را مقصر و مسئول دانست که آنها در برنامه‌نویسی و انجام اقدامات لازم کوتاهی کرده باشند. این در حالی است که در فرض مسئله، این است که آنها کوتاهی‌ای در این باب انجام نداده‌اند. دوباره باید توجه کنیم که سیستم فوق، یک سیستم مبتنی بر یادگیری عمیق است که خود از تجربه می‌آموزد و در این فرایند برنامه‌نویسان دخالتی ندارند. اما چرا اسپارو برنامه‌نویسان را مسئول نمی‌داند. وی در این رابطه به دو نکته اشاره می‌کند. نکته اول اینکه، احتمال اینکه سیستم تسلیحاتی خودمختار به اهداف اشتباهی (همچون مورد بالا) شلیک کند ممکن است محدودیتی برای سیستم باشد که از قبل شناخته شده است. حال اگر سازندگان آن این محدودیت را برای کاربران آن از قبل مشخص کرده باشند، به نظر می‌رسد که دیگر نمی‌توان سازندگان و برنامه‌نویسان را مسئول دانست. در این حالت، مسئولیت برعهده کسانی است که از این سیستم تسلیحاتی خودکار استفاده کرده‌اند.

نکته دوم اینکه، احتمال اینکه یک سیستم خودمختار تصمیماتی را برخلاف آنچه برنامه‌نویسان آن پیش‌بینی کرده و طالب آن بوده‌اند را اتخاذ کند، امری است که ذاتی یک سیستم خودمختار است. به بیان دیگر، ازآنجاکه این سیستم خودمختار است و می‌تواند از تجربه یاد بگیرد، بنابراین می‌توان انتظار داشت تصمیماتی را برخلاف

خواسته برنامه‌نویسانش اتخاذ کند. اسپارو بیان می‌کند هرچه یک سیستم خودمختارتر باشد، قابلیت بیشتری دارد تا بتواند تصمیماتی را برخلاف نظر و خواست سازندگانش اتخاذ کند. از این‌رو، ازآنجاکه برنامه‌نویسان نه بر تصمیمی که سیستم خودمختار اتخاذ می‌کند کنترلی دارند و نه می‌توانند آن را پیش‌بینی کنند، بنابراین آنان مسئول نخواهند بود. در واقع، اسپارو بر روی دو شرط برای مسئولیت‌پذیری تأکید می‌کند: ۱) داشتن کنترل بر روی سیستم و ۲) پیش‌بینی‌پذیری سیستم. از نظر اسپارو زنجیره علّی میان تصمیم سیستم خودمختار و برنامه‌نویسانش تنها به این دلیل که سیستم خومختار فرض شده است، از بین رفته و هیچ ارتباط علّی‌ای را نمی‌توان میان آنها برقرار کرد. اسپارو از این مثال جالب استفاده می‌کند که اگر بتوان برنامه‌نویسان را مسئول سیستم تسلیحاتی خودمختار دانست، آنگاه می‌باید، به طریق مشابه، پدرها و مادرها را نیز مسئول اعمال اشتباه فرزندانشان بدانیم.

اگرچه اسپارو صورت‌بندی‌ای از استدلالش ارائه نکرده است، به نظر می‌رسد می‌توان استدلال وی را مانند زیر صورت‌بندی کرد:

۱) اگر فرد S بر روی عمل x (۱) کنترلی نداشته و (۲) نتواند آن را پیش‌بینی کند، آنگاه فرد S مسئول عمل x نیست.

۲) در سناریوی بالا، ازآنجاکه سیستم هوشمند خودمختار بوده و از طریق یادگیری عمیق ار تجربه می‌آموزد، طراحان و برنامه‌نویسان نمی‌توانند تصمیمات آن را پیش‌بینی کرده و بر روی آن کنترلی داشته باشند.

۳) بنابراین، در سناریوی بالا، طراحان و برنامه‌نویسان مسئول آنچه

سیستم انجام داده است نیستند.

اسپارو در مورد اینکه آیا فرماندهانی که از این سلاح‌ها استفاده می‌کنند مسئول هستند، استدلال مشابهی را بکار می‌برد. در مورد فرماندهان نیز می‌توان گفت ازآنجاکه آنها رفتارهای سیستم خودمختار را نه می‌توانند پیش‌بینی و نه کنترل کنند، بنابراین آنها را نیز نمی‌توان مسئول دانست.

اما آیا می‌توان خود سیستم هوشمند را مسئول دانست؟ پاسخ اسپارو به این پرسش نیز منفی است. اسپارو در ابتدا بیان می‌کند این ایده که سیستم هوشمند را در سناریوی مطرح شده در بالا مسئول بدانیم ایده‌ای است که به‌سختی می‌توان به آن تن داد. اما چرا اینگونه است؟ پاسخ اسپارو این است که مسئله این است که اساساً چگونه باید یک سیستم و ماشین را مسئول دانست. در واقع، اینکه ما فردی را مسئول فعلی می‌دانیم به این معنی است که آنها را شایسته‌ی سرزنش یا ستایش و در نتیجه شایسته‌ی تنبیه یا پاداش می‌دانیم. بنابراین، برای اینکه بتوانیم یک ماشین و سیستم هوشمند را مسئول کارش بدانیم، باید بتوانیم تنبیه و پاداش را نسبت به آن تصور کنیم. اما به نظر می‌رسد که این امر مسئله‌ای خلاف شهود بوده و بالبداهه پیداست که تنبیه یک ماشین هوشمند و یا پاداش به آن امری بی‌معنا است.

اما ممکن است کسی با این شهود همراهی نکرده و در عوض قائل شود که تنبیه و پاداش برای سیستم‌های هوشمند نیز کاملاً معنادار است. در واقع، ممکن است کسی چنین بگوید که ازآنجاکه

این سیستم‌ها، هوشمند هستند، بنابراین می‌توان برای آنها تنبیه و پاداش را تصور نمود. در واقع، این فرد بیان می‌کند اگر ما سیستم‌های هوشمند را موجوداتی فرض کنیم که دست‌کم ظرفیت عقلانی‌ای شبیه به انسان‌ها دارند، آنگاه معقول به نظر می‌رسد که فرض کنیم آنها می‌توانند تنبیه شوند. برای مثال، اگر سیستم هوشمندی عمل اشتباهی را انجام داد می‌توانیم آن را به زندان بیاندازیم، می‌توانیم به آن آسیب فیزیکی بزنیم، مثلا برخی از قطعات آن را از آن جدا کنیم یا مثلا می‌توانیم برخی مشکلات نرم‌افزاری برای آن ایجاد کنیم، و از این طریق آن را تنبیه کنیم.

اسپارو در پاسخ به نکته مهمی اشاره می‌کند. وی بیان می‌کند که اگرچه اقدامات فوق ممکن است به لحاظ روانی برای خانواده فرد بی‌گناهی که توسط سیستم تسلیحاتی خودمختار کشته شده است مؤثر بوده و حس آنها برای انتقام را تسلی دهد، اما به نظر می‌رسد که اقدامات فوق را نمی‌توان به‌عنوان تنبیه در نظر گرفت. زیرا به نظر می‌رسد که بر اساس یک نظریه از تنبیه می‌توان گفت:
(*) اگر فرد S تنبیه شود، آنگاه وی باید در نتیجه آن تنبیه رنج[1] ببرد.

اسپارو بیان می‌کند که باتوجه‌به (*) به‌سختی می‌توان پذیرفت که اقدامات بالا سبب شوند که ماشین هوشمند رنج ببیند. برای مثال، اگر یکی از بستگان قربانی حادثه‌ی فوق از ما بپرسد که آیا ماشین هوشمند به اندازه کافی تنبیه شد، عجیب به نظر می‌رسد که در جواب بگوییم که بله، ما با روغن‌کاری نکردن چرخ‌دنده‌هایش، آن را رنج دادیم. در واقع، رنج کشیدن ماشین باید به‌گونه‌ای باشد

1. suffer

که اگر ما بعد از اینکه ماشین را تنبیه کردیم متوجه شدیم که او بی‌گناه بوده است، آنگاه این احساس به ما دست دهد که کار بسیار اشتباهی کرده‌ایم. اما به نظر می‌رسد که چنین نیست.

خلاصه آنکه، از نگاه اسپارو، نه برنامه‌نویسان سیستم‌های تسلیحاتی خودمختار، نه افسرانی که از این سیستم‌ها استفاده می‌کنند و نه خود این سیستم‌ها را نمی‌توان در قبال آنچه در سناریوی بالا مطرح شد، مسئول دانست. از این‌رو، وی این نگاه را مطرح می‌کند که ازآنجاکه نمی‌توان کسی را مسئول دانست شاید باید استفاده از سیستم‌های هوشمند خودمختار را متوقف کرد.

خوب است به این نکته اشاره کنیم که اگرچه تمرکز استدلال و بحث اسپارو بر روی سیستم‌های تسلیحاتی خودمختار است، ولی می‌توان تمام آنچه وی بیان می‌کند را بر دیگر سیستم هوشمند خودمختار نیز سرایت داد. برای مثال، بحث‌های مشابهی را می‌توان در مورد خودروهای خودران و سیستم‌ها هوشمند در پزشکی نیز جاری کرد. به‌طورکلی، آنچه وی بیان می‌کند در مورد هر سیستم هوشمند خودمختاری که بر اساس یادگیری عمیق طراحی شده است صادق و قابل تکرار است.

حال اجازه دهید به استدلال دیگری که توسط هِولک[1] و نیدارواملین[2] ارائه شده است، اشاره کنیم (Hevelke & Nida-Rümelin, 2015). این استدلال در مورد ماشین‌های خودران که در بالا به آنها اشاره شد می‌باشد. اگرچه ایشان خود صورت‌بندی‌ای از استدلال‌شان ارائه نداده‌اند، ولی شاید بتوان استدلال آنها را مانند زیر صورت‌بندی کرد:

1. Hevelke
2. Nida-Rümelin

۱) اگر بخواهیم سازندگان ماشین‌های خودران را مسئول حوادثی که توسط این ماشین‌های خودران به وجود می‌آید بدانیم، آنگاه این امر سازندگان را از تولید چنین ماشین‌هایی دلسرد می‌کند.
۲) اما این نتیجه نامطلوب است. (زیرا قابلیت‌های امنیتی ماشین‌های خودران دلیل بسیار مهمی است که سبب می‌شود ما علاقه‌مند به تولید این ماشین‌ها باشیم.)
۳) بنابراین، سازندگان ماشین‌های خودران را نباید مسئول دانست.

همان‌طور که پیداست استدلال هِولک و نیدارواملین را می‌توان یک استدلال عمل‌گرایانه دانست نه یک استدلال دقیق فلسفی و اخلاقی. اما به نظر نگارنده می‌توان نکتهٔ ظریفی را از استدلال بالا استخراج کرد. اینکه اگر منفعت سیستم‌های هوشمند خودمختار بیش از زیان آنها باشد، آنگاه نباید سازندگان این سیستم‌ها را مسئول نقایص و رفتارهای بعضاً خطای این سیستم‌ها دانست. شاید بتوان استدلالی را مانند زیر ترتیب داد:

۱) اگر منافع یک سیستم هوشمند بیش از زیان‌های آن باشد، آنگاه سازندگان و برنامه‌نویسان سیستم هوشمند مسئول خطاهای سیستم نخواهند بود.
۲) منافع سیستم‌های هوشمند بسیار بیش از مضرات آنهاست.
۳) بنابراین، سازندگان و برنامه‌نویسان این سیستم‌ها مسئول خطاهای این سیستم‌ها نیستند.

درباره این پیشنهاد در بخش «ملاحظاتی در باب شکاف مسئولیت» بیشتر سخن خواهیم گفت.

۳-۲- دیدگاه وجود مسئول

در مقابلِ کسانی که قائل‌اند مسئولیت سیستم‌های هوشمند خودمختار نه برعهده برنامه‌نویسان و طراحان است و نه برعهده خود سیستم هوشمند (به طور کل مسئولیت برعهده هیچ‌کس نیست)، دیدگاهی قرار دارد که مسئولیت خطاهایی که از سیستم‌های هوشمند سر می‌زند را به فرد یا یک شیء منتسب می‌کنند. در این بین برخی مسئولیت را برعهده برنامه‌نویسان و طراحان و برخی دیگر مسئولیت را برعهده خود سیستم هوشمند می‌گذارند. در ادامه به این دو دیدگاه اشاره می‌کنیم.

۳-۲-۱- مسئول انسان (برنامه‌نویس/کاربر) است

یکی از دیدگاه‌هایی که قائل است برای خطاهای سیستم‌های هوشمند خودمختار مسئولی وجود دارد دیدگاهی است که توسط توربن اسووبودا[1] ارائه شده است (Swoboda, 2017). وی برخلاف اسپارو بحث می‌کند که می‌توان برنامه‌نویسان و طراحان سیستم‌های تسلیحاتی خودمختار را مسئول خطاهای برآمده از این سیستم‌ها دانست. قبل از اینکه دیدگاه اسووبودا را بیان کنیم، خوب است به این نکته توجه کنیم که اگرچه اسووبودا بحث خود را در باب سیستم‌های *تسلیحاتی* خودمختار ارائه می‌کند، ولی می‌توان حاصل سخن وی را در مورد هر سیستم هوشمند خودمختاری پیاده کرد.

اسووبودا بحث خود را در تقابل با استدلال اسپارو که در بخش قبل آن را بیان کردیم ارائه می‌کند. همان‌طور که دیدیم از نظر اسپارو ازآنجاکه سیستم هوشمند خودمختار دو شرط کنترل و

1. Torben Swoboda

پیش‌بینی‌پذیری را برآورده نمی‌کند، بنابراین نمی‌توانیم برنامه‌نویسان و طراحان آن را مسئول بدانیم. اما اسووبودا برخلاف اسپارو قائل است که برنامه‌نویسان و طراحان می‌توانند هم بر روی سیستم هوشمند کنترل داشته باشند و هم اینکه می‌توانند رفتارهای آن را پیش‌بینی کنند. اسووبودا بیان می‌کند سیستم‌های هوشمند بر اساس یادگیری عمیق از تجربه می‌آموزند. ولی باید به این نکته توجه داشت که برنامه‌نویسان و طراحان این سیستم‌ها بر روی اینکه آیا سیستم یادگیری را ادامه دهد یا نه کنترل دارند. به بیان دیگر، از نظر اسووبودا اگر برنامه‌نویسان سیستم هوشمند ملاحظه کنند و یا احتمال دهند که سیستم هوشمند یادگیری را در مسیری غیر از آنچه مورد خواست آنهاست ادامه می‌دهد، آنها می‌توانند از یادگیری سیستم ممانعت به عمل آورند. اسووبودا پیشنهاد می‌کند که می‌توان در ابتدا سیستم هوشمند را بر اساس آنچه مورد خواست برنامه‌نویسان و طراحان است آموزش داد و سپس آموزش را متوقف کرد و از سیستم هوشمند استفاده نمود. در واقع، اسووبودا پیشنهاد می‌کند که می‌توان از آموزشِ در حین عمل سیستم هوشمند ممانعت نمود.

اسووبودا در گام بعد به سراغ شرط پیش‌بینی‌پذیری سیستم هوشمند می‌رود. وی در ابتدا دو نوع از پیش‌بینی را از یکدیگر متمایز می‌کند: پیش‌بینی محلی و پیش‌بینی غیر محلی.[1] وی برای توضیح این دو مورد در ابتدا توضیحی را میان رفتار محلی و رفتار غیرمحلی می‌دهد.[2] وی برای توضیح رفتار محلی و غیرمحلی از مثال شطرنج استفاده می‌کند. در بازی شطرنج حرکت مهره فیل از خانه e2 به

1. Local and Non-local Prediction
2. local and non-local behaviour

خانه b5 توسط شطرنج‌باز یک رفتار محلی محسوب می‌شود. در مقابل، مجموع رفتارها و استراتژی‌های وی برای بردن بازی، رفتار غیرمحلی وی محسوب می‌شود. وی با استفاده از دو مفهوم رفتار محلی و رفتار غیرمحلی تمایز میان پیش-بینی محلی و پیش‌بینی غیرمحلی را روشن می‌کند. منظور از پیش‌بینی محلی، پیش‌بینی رفتار محلی و منظور از پیش‌بینی غیرمحلی پیش‌بینی رفتار غیرمحلی است. با استفاده از این تمایز وی بیان می‌کند نمی‌توان پیش‌بینی کرد که یک سیستم تسلیحاتی خودمختار در یک لحظه و موقعیت خاص قصد حمله به یک شیء خاص را دارد یا خیر (این همان پیش‌بینی محلی است). اما از نظر اسووبودا درست است که برنامه‌نویسان و طراحان نمی‌توانند رفتار محلی سیستم تسلیحاتی خودمختار را پیش‌بینی کنند، ولی آنها کماکان می‌توانند رفتارهای غیر محلی آن را پیش‌بینی کنند (این همان پیش‌بینی غیرمحلی است). وی بیان می‌کند که اولا، برنامه‌نویس دستگاه را به گونه‌ای طراحی کرده است که خطر طبقه‌بندی کاذب حداکثری اهداف غیرمجاز را به حداقل برساند. به بیان دیگر، برنامه‌نویس در ابتدای کار خود سیستم را طوری طراحی کرده است که سیستم قادر باشد اهداف مجاز، مانند سربازان دشمن، تانک‌های دشمن و ...، را از اهداف غیرمجاز، مانند انسان‌های غیرنظامی، خانه‌های مسکونی و ...، متمایز کند. دوماً، در مرحله آزمایش برنامه‌نویس دریافته است که ماشین از داده‌های جدید چقدر با دقت استفاده می‌کند. این یک انتظار اولیه را برای برنامه‌نویس فراهم می‌کند که ماشین در دنیای واقعی چقدر خوب عمل می‌کند. هر دو عامل به برنامه‌نویس اجازه می‌دهد

پیش‌بینی‌های غیر محلی خوبی از رفتار دستگاه داشته باشد.

اسووبودا سپس این سؤال را مطرح می‌کند که برنامه‌نویس و طراح برای اینکه وی را مسئول بدانیم، به کدام نوع از پیش‌بینی (محلی یا غیرمحلی) نیاز دارد. پاسخ وی به این سوال این است که برنامه‌نویسان و طراحان تنها نیاز دارند که پیش‌بینی غیرمحلی از رفتار سیستم هوشمند داشته باشند. از نظر وی، آنها نیازی به پیش‌بینی محلی رفتارهای سیستم برای اینکه مسئول شناخته شوند ندارند. وی برای روشن شدن مطلب از مثال کارگر ساختمانی استفاده می‌کند. فرض کنید که یک کارگر ساختمانی با بی‌احتیاطی از بالای یک ساختمان در حال ساخت، آواری به خیابان می‌اندازد. عابر پیاده‌ای که اتفاقاً در خیابان راه می‌رود، زیر آوار مانده و آسیب می‌بیند. حال، کارگر ساختمانی صادقانه می‌تواند بگوید که او نمی‌دانست کسی در خیابان قدم می‌زند. اما واضح است که این بیان او را معذور نمی‌دارد، زیرا به طور معقول می‌توان دانست که ممکن است عابران پیاده از آنجا عبور کنند و از ریختن آوار آسیب ببینند. وی بیان می‌کند ندانستن اینکه چه کسی به طور خاص قرار است از پیاده‌رو عبور کند و از آمدن آوار بر رویش مورد ظلم واقع شود، به این معنی نیست که نمی‌توان مسئولیتی را برعهده گرفت. حال، اگر این نوع بهانه را در مورد کارگر ساختمانی قبول نکنیم، پس نباید اجازه دهیم برنامه‌نویس سیستم هوشمند نیز معذور باشد. اسووبودا سپس بیان می‌کند آنچه مهم است این است که باتوجه‌به نتایج مرحله‌ی آزمون سیستم هوشمند، برنامه‌نویس می‌تواند دریابد که چند درصد از اهداف توسط سیستم هوشمند به‌عنوان اهداف

غیرمجاز و چند درصد از اهداف به‌عنوان اهداف مجاز طبقه‌بندی شده است. حال، می‌توان گفت برنامه‌نویس مسئول مرگ‌های ناشی از خطاهای سیستم تسلیحاتی خودمختار است، زیرا او با کمال میل پذیرفته است که درصد معینی از اهداف غیرمجاز به اشتباه به‌عنوان اهداف مجاز دسته‌بندی می‌شوند.

اما دراین‌بین سؤالی وجود دارد که باید به آن پاسخ گفت. سوال این است که نسبت درصد تشخیص اهداف مجاز نسبت به اهداف غیرمجاز چه مقدار باید باشد تا برنامه‌نویس را مسئول بدانیم. به بیان دیگر، دقت سیستم تسلیحاتی خودمختار باید چقدر باشد تا برنامه‌نویس را مسئول دانست؟ آیا این دقت باید ۱۰۰درصد، یا ۹۹ درصد و یا ۵۰ درصد باشد؟ اسووبودا این مسئله را مسئله آستانه دقت نام می‌نهد و در جواب آن چنین می‌نویسد:

«قبل از استفاده از سیستم تسلیحاتی خودمختار در صحنه نبرد، این سیستم چقدر باید دقیق باشد؟ واکنش اولیه می‌تواند این باشد که نباید هیچ اشتباهی انجام دهد. دستیابی به این امر از یک سو عملاً غیرممکن است. از طرف دیگر، یک استدلال نتیجه‌گرایانه نیز علیه این وجود دارد. سربازان انسانی اشتباه می‌کنند و افراد غیرمجاز [مثلاً غیرنظامیان] را هدف قرار می‌دهند. وقتی باید تصمیم بگیریم که آیا از سیستم تسلیحاتی خودمختار استفاده کنیم، با هزینه‌های فرصت روبرو می‌شویم. سیستم تسلیحاتی خودمختار دارای نرخ مثبت کاذب مشخصی است و انسان نیز دارای یک نرخ مثبت خواهد بود.

اگر چنین باشد که سیستم تسلیحاتی خودمختار به اهداف غیرمجاز کمتری از سربازان انسانی حمله کند، استفاده از سیستم تسلیحاتی خودمختار دست‌کم باید از نظر نتیجه‌گرایی مجاز باشد. این به این دلیل است که با استفاده از سیستم تسلیحاتی خودمختار زندگی افراد بیشتری را حفظ می‌کنیم. مولر[1] یک نکته مرتبط را بیان می‌کند. ما انتظار نداریم که داروسازان در زندگی یا مرگ هیچ اشتباهی نکنند. در عوض ما انتظار «مراقبت لازم»[2] توسط داروسازان را داریم. بحث من در اینجا این است که اگر سیستم تسلیحاتی خودمختار اشتباهات مثبت کاذب کمتری نسبت به انسان داشته باشد، برنامه‌نویس از آن مراقبت کرده است.» (Swoboda, 2017)

همان‌طور که پیداست، پاسخ اسووبودا این است که صرفاً کافی است سیستم تسلیحاتی خودمختار در مقایسه با انسان‌ها، اشتباهات کمتری انجام دهد. حال اگر انسان‌ها، برای مثال، در ۳۰ درصد مواقع اهداف غیرمجاز را مجاز تلقی می‌کنند، آنگاه سیستم تسلیحاتی هوشمند می‌باید وضعیت بهتری نسبت به انسان داشته باشد؛ برای مثال، در ۲۰ درصد مواقع اهداف غیرمجاز را مجاز تلقی کند و در ۸۰ درصد دیگر واقعاً اهداف را درست تشخیص دهد. حال اگر برنامه‌نویس سیستمی را طراحی کند که عملکرد بدتری نسبت به انسان داشته باشد، آنگاه برنامه نویس در برابر حوادث ناشی از سیستم هوشمند مسئول خواهد بود و اگر سیستم طراحی شده توسط وی عملکرد بهتری نسبت به عملکرد انسان داشته باشد، آنگاه

1. Mueller
2. due care

برنامه نویس معذور خواهد بود.

البته اسووبودا خود ملاحظاتی را درباره معیار فوق مطرح می‌کند. اول اینکه چقدر طراحی سیستم تسلیحاتی‌ای که بتواند معیار فوق را برآورده کند در عمل شدنی است. وی بیان می‌کند مسئله تعیین آستانهٔ دقت کاملا به این بستگی دارد که سیستم تسلیحاتی خودمختار برای چه کاری طراحی شده است. مثلاً، سیستم‌هایی که برای انهدام تانک‌ها، قایق‌ها یا هواپیماهای خاص طراحی شده‌اند به اندازه کافی خوب عمل می‌کنند. اما طراحی سیستم‌هایی که برای انهدام انسان‌ها به کار می‌روند کار بسیار پیچیده‌تری است. یک مشکل دیگر مربوط به معیار فوق توافق بر سر یک خط مبنا است. برای مثال، ارتش‌ها ازآنجاکه تمایل دارند از این سیستم‌های تسلیحاتی خودمختار استفاده کنند، ممکن است در بیان درصد خطای سربازان انسانی اغراق کنند تا از این طریق بتواند شرایط را برای به کارگیری این سیستم‌های تسلیحاتی راحت‌تر جلوه دهند.

اسووبودا استدلال دیگری را نیز برای نشان‌دادن اینکه برنامه‌نویسان و طراحان، مسئول خطاهای سیستم هوشمند هستند ارائه کرده است. استدلال وی بطور خلاصه این است که برنامه‌نویسان خود تعیین می‌کنند که از چه ملاک‌هایی برای متمایز کردن افراد نظامی و افراد غیرنظامی استفاده کنند. حال، اگر برنامه‌نویس در باب تعیین شرایط درست و اخلاقی برای سیستم جهت تمایز افراد نظامی از غیرنظامی کوتاهی و قصور کرده باشد، آنگاه می‌توان او را مسئول خطاهای ناشی از عملکرد بد سیستم تسلیحاتی هوشمند دانست.

در پایان، اسووبودا به ملاحظاتی کلی در باب استفاده از سیستم‌های

تسلیحاتی خودمختار اشاره می‌کند که از این قرارند؛ اول اینکه، نمی‌توان کاملا مطمئن شد که آیا سیستم هوشمند واقعاً یادگرفته است که اهداف نظامی را به خوبی از اهداف غیرنظامی تمییز دهد یا خیر. در این راستا وی سیستم شبکه عصبی مصنوعی‌ای را مثال می‌زند که برای تفکیک تصاویری که در آنها تانک وجود دارد از تصاویری که در آنها مخازن استتار وجود دارد طراحی شد. اما در نهایت معلوم شد که این سیستم هوشمند آموخته است که تصاویر آفتابی را از تصاویر ابری تفکیک کند. اسووبودا بیان می‌کند که این مشکل را می‌توان با آزمایش دقیق کاهش داد، اما نمی‌توان آن را برطرف کرد.

دوم اینکه، سیستم تسلیحاتی خودمختار در معرض هک است. اگر یک سیستم تسلیحاتی خودمختار دارای یک روزنه امنیتی باشد، آنگاه تمام سیستم‌های تسلیحاتی خودمختار با همان نرم‌افزار و سخت‌افزار نیز همان نقطه‌ی ضعف را خواهند داشت. این مسئله این خطر را دارد که کل ارتش سیستم تسلیحاتی خودمختار از کار بیافتد یا در مقابل سربازان خودی قرار بگیرد.

سوم اینکه، درست است که بیان شد مسئولیت خطاهای سیستم تسلیحاتی خودمختار برعهدهٔ برنامه-نویس است، اما ازآنجاکه برای طراحی یک سیستم هوشمند چندین (و نه تنها یک) برنامه‌نویس نقش دارند، بنابراین، ما با یک مسئله‌ی مسئولیت جمعی و نه فردی روبرو هستیم. ملاحظاتی در باب استدلال اسووبودا به نظر نگارنده می‌رسد که آنها را در بخش «ملاحظاتی در باب شکاف مسئولیت» مطرح خواهد کرد.

یکی دیگر از رویکردها برای پاسخگویی به مسئله شکاف مسئولیت دیدگاهی است که بر نظریهٔ «ابزارگرایی»[1] مبتنی است. این دیدگاه نیز بر این عقیده است که مسئولیت خطاهای ناشی از سیستم‌های هوشمند خودمختار برعهده‌ی انسان، اعم از برنامه‌نویسان، سازندگان و کاربران، است. در ابتدا اجازه دهید بیان کنیم که منظور از ابزارگرایی چیست. دیدگاه ابزارگرایی در باب تکنولوژی دیدگاهی است که در میان انسان‌های عادی رایج است. مطابق این دیدگاه تکنولوژی ابزاری است که به وسیله کاربران انسانی برای اهدافی خاص به خدمت گرفته شده است. مارتین هایدگر که خود از مخالفین این دیدگاه است، این رویکرد به تکنولوژی را «توصیف ابزاری»[2] می‌نامد و در این باب می‌نویسد:

> «هنگامی‌که ما درباره تکنولوژی سوال می‌پرسیم در واقع می‌پرسیم که تکنولوژی چیست. هر فردی به دو جمله‌ای که سوال ما را پاسخ می‌دهد علم دارد. یکی که می‌گوید: تکنولوژی ابزار و وسیله‌ای است برای یک هدف و دیگری که می‌گوید: تکنولوژی یک فعالیت انسانی است. این دو تعریف از تکنولوژی به یکدیگر مربوطاند. زیرا تعیین کردن اهداف و به کار بردن ابزار برای آن هدف‌ها یک فعالیت انسانی است» (Heidegger, 1977, pp. 4–5).

بر اساس دیدگاه فوق، نظریه‌ای شکل گرفته است که به آن «نظریه ابزارگرایانه»[3] اطلاق شده است. بر اساس این دیدگاه،

1. Instrumentalism
2. the instrumental definition
3. Instrumentalist theory

تکنولوژی تنها یک وسیله و ابزاری است که انسان برای رسیدن به اهدافی خاص به کار می‌برد. ازآنجاکه یک ابزار به خودی خود هیچ ویژگی هنجاری و ارزشی‌ای ندارد، یک مصنوع تکنولوژیکی را نباید با در نظر گرفتن خودش (به خودی خود) ارزیابی کرد. بلکه در عوض، آن را باید بر اساس نحوه استفاده طراحان انسانی و کاربران آن ارزیابی نمود (Feenberg, 1991, p. 5).

حال، برای پاسخ به مسئلهٔ شکاف مسئولیت برخی از ایده‌ی ابزارگرایی استفاده کرده‌اند. برای مثال، یوآنا بریسان[1] بیان کرده است که می‌توان از ربات‌ها صرفا به‌عنوان برده استفاده کرد؛ نه به‌عنوان همتایان و همراهانی برای انسان (Bryson 2010, p. 63). این ایده بر دو امر تأکید می‌کند. اول اینکه انسان را از هر چیز دیگری مستثنی می‌کند و جایگاه ویژه‌ای به انسان داده و بیان می‌کند که این تنها انسان است که دارای حق و مسئولیت است. تکنولوژی هر چقدر هم که می‌خواهد پیچیده باشد تنها می‌تواند به‌عنوان ابزاری برای افعال و اهداف انسان‌ها باشد و نه چیزی بیشتر از آن. دوم اینکه، بریسان بیان می‌کند که مسئولیت هر گونه خطا و آسیبی که از جانب تکنولوژی به وجود آید، تماماً متوجه انسان است. این ما انسان‌ها هستیم که بطور مستقیم یا غیرمستقیم اهداف و رفتار تکنولوژی را تعیین می‌کنیم. مثلاً این ما هستیم که تعیین می‌کنیم هوشمندی را برای آن‌ها قرار دهیم و یا حتی بالاتر، این ما هستیم که برای آنها تعیین می‌کنیم که چگونه هوشمندی خود را کسب کنند (اشاره به یادگیری عمیق دارد). در این راستا، برخی بیان می‌کنند که مسئولیت مربوط به خطرات و آسیب‌های ناشی از

1. Joanna Bryson

سیستم‌های هوشمند و ربات‌ها برعهده، تولیدکنندگان، کاربران و یا حتی (در سطح سیاسی) مسئولان امری که اجازه تولید آن ربات‌ها را داده‌اند، می‌باشد (Gladden 2016, p. 184; Datteri 2013; Calverley 2008, p. 533). به نظر می‌رسد بتوان استدلال این گروه را مانند زیر بیان کرد:

۱) تکنولوژی صرفاً ابزاری است که انسان‌ها برای اهداف خود از آنها استفاده می‌کنند [نظریه ابزارگرایی].

۲) بنابراین، هرگونه مسئولیت استفاده از تکنولوژی برعهده انسان (اعم از کاربران و سازندگان) است.

۳) سیستم‌های هوشمند خودمختار مصادیقی از تکنولوژی هستند.

۴) بنابراین، هرگونه مسئولیت استفاده از سیستم‌های هوشمند خودمختار برعهده‌ی انسان (اعم از کاربران و سازندگان) است.

یکی از مشکلاتی که در برابر این دیدگاه مطرح شده است این است که چنین رویکردی به مسئلهٔ مسئولیت، اینکه طراحان سیستم‌های هوشمند را تماماً مسئول خطاهای آن بدانیم، یک لازمه ناخوشایند دارد و آن اینکه ما بطور ناخواسته خلاقیت و توسعه سیستم‌های هوشمند را توسط طراحان آن محدود کرده‌ایم (Gunkel 2017). در واقع، اگر طراحان سیستم‌های هوشمند خود را مسئول همه عواقب ناشی از مصنوعاتشان بدانند، آنگاه بسیار محافظه‌کار می شوند و اساساً میلی به تولید تکنولوژی‌های برتر و پیش-رو نخواهند داشت. این خود سبب عدم پیشرفت بشر خواهد گشت. به بیان عامیانه، این امر سبب خواهد شد که جرأت و شجاعت از برنامه‌نویسان و طراحان برای ایجاد پیشرفت در تکنولوژی سلب شود.

۳-۲-۲- مسئول، خود سیستم هوشمند است

حال اجازه دهید به دیدگاه افرادی اشاره کنیم که خود سیستم هوشمند را مسئول می‌دانند (و نه کاربران و برنامه‌نویسان آن را). دوباره تأکید می‌کنیم که اگرچه این استدلال در باره ماشین‌های خودران ارائه خواهد شد، ولی می‌توان به آن کلیت داده و در مورد هر سیستم هوشمند خودمختاری بکار برد.

برخی در باب ماشین‌های خودمختار بر این عقیده هستند که می‌توان عاملیت[1] را به ماشین‌های خودران نسبت داد. برای مثال، مارک کولبرگ[2] قائل است که هنگامی که یک انسان از ماشین خودران استفاده می‌کند، عاملیت بطور کل به ماشین انتقال می‌یابد (Coeckelbergh, 2016, p. 754). با استفاده از این ادعای کوکلبرگ، می‌توان استدلالی را مانند زیر استخراج کرد:

۱) اگر فرد S در انجام عمل x عاملیت داشته باشد، آنگاه فرد S مسئول عمل x است.

۲) در هنگام استفاده از سیستم‌های خودران، همه عاملیت از انسان به ماشین منتقل می‌شود.

۳) بنابراین، ماشین مسئول است.

اما دیدگاه فوق مخالفانی نیز دارد که با مقدمهٔ دوم استدلال بالا مخالفت می‌کنند. برای مثال، فیلیپ بری[3] در ابتدا عاملیت را این‌طور تعریف می‌کند: عمل بر اساس باورها و امیال.[4] وی سپس بیان می‌کند که ربات‌ها فاقد باور و امیال هستند و لذا فاقد عاملیت می‌باشند (Brey, 2013) و یا دونکن پوروس[5] بر این عقیده است که عاملیت شامل عمل بر اساس دلایل[6] است. حال آنکه ربات‌ها فاقد

1. Agency
2. Mark Coeckelbergh
3. Philip Brey
4. beliefs and desires
5. Duncan Purves
6. reasons

ایـن قابلیـت هسـتند کـه بتواننـد بـر اسـاس دلایـل عملـی را انجـام دهنـد (Purves et al., 2015).

شـکل (۱) ایـن سـه دیـدگاه در مـورد اسـناد مسـئولیت سیسـتم‌های هوشـمند خودمختـار را خلاصـه می‌کنـد.

شکل ۱- نگاه‌های موجود درباب اسناد مسئولیت به سیستم‌های هوشمند خودمختار

بخش چهارم

سیستم‌های هوشمند خودمختار از منظر حقوقی

بخش چهارم

سیستم‌های هوشمند خودمختار از منظر حقوقی

در بخش‌های قبل به نزاعی که در باب مسئولیت اخلاقی مربوط به خطاهای ناشی از سیستم‌های هوشمند خودمختار وجود داشت اشاره کردیم. حال، در این بخش قصد داریم که از جنبه حقوقی نیز به این بحث بپردازیم. منظور از نگاه حقوقی به این مسئله این است که در نظام حقوقی چگونه مجازات کیفری برای حوادث ناشی از سیستم‌های هوشمند خودمختار تعیین شده است. در این راستا، با استفاده از تحقیقی که توسط سعید عطازاده و جلال انصاری (عطازاده و انصاری، ۱۳۹۸) انجام شده است، مسئله شکاف مسئولیت از منظر حقوقی را در سه کشور ایالات متحده آمریکا، آلمان و ایران بررسی خواهیم کرد.

۴-۱- قوانین حقوقی ایالات متحده آمریکا در باب مسئولیت کیفری

سیستم حقوقی ایالات متحده آمریکا برای مسئولیت کیفری دو شرط قرار داده است. این شروط عبارت‌اند از: الف) عمل ممنوعه‌ای توسط شخص انجام شده باشد و ب) فرد قصد مجرمانه برای انجام آن عمل داشته باشد. البته باید توجه کرد که این شروط نه فقط برای افراد

حقیقی، بلکه برای افراد حقوقی نیز اعمال می‌شود. البته این قانون افرادی را که نمی‌توانند درکی از قصد مجرمانه داشته باشند (مانند مجانین و اطفال) مستثنی کرده است. در واقع، از نظر نظام حقوقی آمریکا قاضی برای اینکه اثبات کند فرد مسئولیت کیفری عملی را برعهده دارد یا نه، باید دو چیز را اثبات کند. اول اینکه فرد مرتکب جرم شده است و دوم اینکه قصد انجام آن را جرم را نیز داشته است (عطازاده و انصاری، ۱۳۹۸).

حال که دانستیم مسئولیت کیفری در نظام ایالات متحده آمریکا چگونه تعریف می‌شود اجازه دهید ببینیم در نظام حقوقی آمریکا مسئولیت کیفری سیستم‌های هوش مصنوعی خودمختار برعهده چه کسی گذاشته شده است. بر اساس نظام حقوقی ایالات متحده آمریکا اگر قربانی حادثهٔ ایجاد شده توسط سیستم هوشمند خودمختار (در صورت حیات) یا یکی از بستگان وی علیه سازنده‌ی سیستم هوشمند اقامه دعوی نماید وی باید ثابت کند که حادثه‌ی ایجاد شده به خاطر یکی از سه مورد زیر یا ترکیبی از آنها ایجاد شده است:

مشکل ایجاد شده ناشی از نقص طراحی است

مشکل ایجاد شده ناشی از نقص ساخت است

مشکل ایجاد شده ناشی از هشدارهای ناکافی در ارتباط با خطرات صدمات ناشی از ربات است (عطازاده و انصاری، ۱۳۹۸)

همچنین، در نظام حقوقی آمریکا اگر مشخص شود «تولیدکننده نقش موثری در ایجاد آسیب عهده‌دار بوده، می‌توان شرکت تولید کننده را با اتهام سهل‌انگاری کیفری یا جنایی مواجه کرد.» (عطازاده و انصاری، ۱۳۹۸)

۴-۲- قوانین حقوقی آلمان در باب مسئولیت کیفری

مسئولیت کیفری در نظام حقوقی آلمان دارای دو مبناست: اول اینکه، فرد بتواند عمل کند، یعنی، قادر باشد اهدافی را برای خود تعیین کند و برای رسیدن به آنها اقداماتی را انجام دهد و دوم اینکه، افراد قادر به درک خطا و اشتباه در رفتار خود باشند (عطازاده و انصاری، ۱۳۹۸). حال، ازآنجاکه ربات‌ها و سیستم‌ها هوشمند نمی‌توانند این دو شرط را برآورده کنند، بنابراین خود این سیستم‌های هوشمند نمی‌توانند مسئول شناخته شوند. همچنین، در کشور آلمان چیزی به نام «اصل شایستگی سرزنش» تعریف شده است که بر اساس آن فرد باید به‌خاطر خطایی که انجام داده است، شایستگی سرزنش را داشته باشد. از این‌رو، بر اساس این قانون کودکان (مجانین) ازآنجاکه شایستگی سرزنش ندارند، نمی‌توانند مقصر شناخته شوند. حال باتوجه‌به این اصل، عطازاده و انصاری بیان می‌کنند که:

«باتوجه‌به این استدلال، شایستگی سرزنش، توانایی عمل‌کننده در تصمیم‌گیری بین درست و غلط بودن را به‌عنوان پیش‌فرض لحاظ می‌نماید، یا به عبارت دیگر، توانایی عمل‌کننده برای اجتناب از ارتکاب یک عمل نادرست را یک اصل می‌داند. به طور قطعی به نظر می‌رسد که این تعریف، استحقاق سرزنش یا مقصر بودن را حتی در مورد ربات‌های بسیار هوشمند حذف می‌کند.» (عطازاده و انصاری، ۱۳۹۸)

عطازاده و انصاری بیان می‌کنند که تاکنون نظام حقوقی آلمان، بر اساس،

دو مبنـای فـوق، مسـئولیت کیفـری افـراد حقوقـی (و نـه حقیقـی) را بـه رسـمیت نشـناخته اسـت؛ چـرا کـه آنهـا دو شـرط ذکـر شـده را بـرآورده نمی‌کننـد. همیـن اسـتدلال در مـورد ربات‌هـا نیـز بـکار رفتـه و مسـئولیت کیفـری آنهـا نیـز بـه رسـمیت شـناخته نشـده اسـت.

بـر اسـاس نظـام حقوقـی آلمـان فـرد در صورتـی مقصـر شـناخته می‌شـود کـه اولاً، رابطـه‌ای علّـی میـان حادثـه و عمـل شـخص برقـرار باشـد و دومـاً، فـرد می‌بایـد توانایـی پیش‌بینـی آسـیب را داشـته باشـد ولـی بـرای جلوگیـری از آسـیب قابل‌پیش‌بینـی قصـور کـرده باشـد. از نظـر عطـازاده و انصـاری چنیـن معیارهایـی نمی‌تواننـد بـه ربات‌هـا منتقـل شـوند؛ چراکـه آنهـا نمی‌تواننـد عواقـب کارهـا را پیش‌بینـی نماینـد. اگرچـه در کشـور آلمـان ایـن قانـون را نمی‌تـوان بـر ربات‌هـا اطـلاق کـرد، ولـی در ایـن کشـور قوانیـن بسـیار سـختگیرانه‌ای در بـاب تولیـد محصولات ناامـن وجـود دارد بـه طـوری کـه نقـض ایـن قوانیـن می‌توانـد مسـئولیت کیفـری را برعهـده‌ی سـازندگان و طراحـان بگـذارد. بـه بیـان دیگـر، نظـام حقوقـی آلمـان در مـواردی کـه حوادثـی از سـوی سیسـتم‌های هوشـمند خودمختـار ایجـاد شـود، سـازندگان و طراحـان را پاسـخگو قلمـداد کـرده اسـت. از ایـن‌رو در آلمـان، «قبـل از بازاریابـی یـک محصـول، تولیدکننـده بایـد ثابـت کنـد کـه اسـتانداردهای علمـی و فنـی فعلی را بـرآورده می‌سـازد و بـا آنهـا مطابقـت داشـته و ایمنـی آن بـرای مشـتریان بـه میـزان کافـی مـورد آزمایـش قـرار گرفتـه اسـت و یـا بـرای مثـال زمانـی کـه محصـول در بـازار اسـت، بـرای جلوگیـری از آسـیب بیشـتر تولیدکننـده بایـد هشـدارها را صـادر کنـد، بـه محصـولات معیـوب بـرای تعمیـر فراخـوان دهـد یـا حتی بازاریابـی و فـروش آنهـا را متوقـف سـازد. اگـر تولیدکننـده نتوانـد بـا ایـن

استانداردها همخوانی داشته باشد ممکن است از لحاظ کیفری مسئول هرگونه آسیب ناشی از محصول باشد.» (عطازاده و انصاری، ۱۳۹۸)

۴-۳- قوانین حقوقی ایران در باب مسئولیت کیفری

عطازاده و انصاری پیش از بیان مفهوم مسئولیت کیفری در نظام حقوقی ایران در ابتدا دیدگاه اسلام را در باب مسئولیت کیفری معرفی می‌کنند. از نظر آنها، مسئولیت کیفری در اسلام بر سه عنصر مبتنی است که عبارت‌اند از: «۱) انسان عمل ممنوعه‌ای را انجام داده باشد ۲) عمل وی همراه با اختیار باشد و ۳) مرتکب، به ماهیت و چیستی عمل خودآگاهی داشته باشد». بر اساس این سه عنصر آنها مسئولیت کیفری از نگاه اسلام را این‌طور تعریف می‌کنند: «انسان نتایج آن دسته از اعمال ممنوعه‌ای را که با اختیار و آگاهی از محتوا و نتایج‌شان ارتکاب نموده تحمل کند.» (عطازاده و انصاری، ۱۳۹۸) اما در مورد مسئولیت کیفری در نظام حقوقی ایران ایشان بیان می‌کنند که تعریفی مشخصی از مسئولیت کیفری در حقوق جمهوری اسلامی ایران ارائه نشده است و تنها شرایط آن به‌صورت پراکنده نام برده شده است. این شرایط عبارتند از بلوغ، عقل، عدم وجود ضرورت ارتکاب جرم و عدم وجود اجبار. ایشان پس از بیان اینکه در میان حقوقدانان داخلی توافقی بر روی تعریف مسئولیت کیفری وجود ندارد و پس از بیان چند دیدگاه مختلف بیان می‌کنند که «مسئولیت کیفری در حقوق اسلام و ایران مبتنی بر عقل، بلوغ و اختیار می‌باشد که این موضوع در ماده ۱۴۰ قانون مجازات اسلامی ۱۳۹۲ و کتب فقهی به طور موردی و پراکنده آمده و همچنین به

این موضوع اشاره شده که صغار و مجانین مسئولیت کیفری ندارد.» (عطازاده و انصاری، ۱۳۹۸)

باتوجه‌به تعریف فوق، ایشان بر این عقیده‌اند که انتساب مسئولیت کیفری به هوش مصنوعی خودمختار، در زمان حاضر، امری غیرممکن است چراکه هوش مصنوعی خودمختار هیچ یک از شروط اشاره شده در بالا را برآورده نمی‌کند و از این‌رو نمی‌توان هوش مصنوعی را مسئول دانست. ایشان در پاسخ به این سؤال که پس در نهایت، بر اساس نظام حقوقی ایران، چه کسی مسئول حوادث ناشی از سیستم‌های هوشمند خودمختار خواهد بود بیان می‌کنند که بر اساس مواد ۵۰۱ و ۵۲۶ قانون مجازات اسلامی می‌توان شخص برنامه‌نویس یا کاربر یا شرکت سازنده را مسئول حوادث ناشی از سیستم‌های هوشمند خودمختار دانست و مجازات را بر آنها بار کرد. شکل (۲)، مواجهه نظام‌های حقوقی ایران، آلمان و آمریکا را با سیستم‌های هوشمند خودمختار به تصویر می‌کشد.

شکل (۲) مواجهه نظام‌های حقوقی ایران، آلمان و آمریکا با سیستم‌های هوشمند خودمختار

بخش پنجم

ملاحظاتی درباب شکاف مسئولیت

بخش پنجم

ملاحظاتی درباب شکاف مسئولیت

در بخش‌های قبل نزاعی را که در باب شکاف مسئولیت میان متفکران وجود دارد، به تصویر کشیدیم. در این بخش قصد داریم ملاحظاتی را در باب شکاف مسئولیت و ادبیات شکل گرفته پیرامون آن بیان کنیم.

همان‌طور که دیدیم، از نظر برخی نمی‌توان هیچ‌کس را مسئول حوادث ناشی از سیستم‌های هوشمند خودمختار دانست. از نظر این گروه، ربات‌ها و سیستم‌های هوشمند نیز نمی‌توانند مسئول تلقی شوند. در مقابل این دیدگاه، دیدگاهی قرار داشت که قائل بود خود سیستم هوشمند می‌باید مسئول شناخته شود. به نظر می‌رسد اگرچه دیدگاهی که ربات‌ها و سیستم‌های هوشمند را مسئول می‌داند در بادی امر عجیب به نظر می‌رسد، ولی شاید بتوان با نگاهی عمیق‌تر آن را قابل دفاع دانست. شاید یک مدافع این نظریه بتواند این طور از نظر خود دفاع کند. فرض کنیم سیستم‌های هوشمند به حدی از هوشمندی رسیده‌اند که یا دست‌کم به اندازه انسان‌ها هوشمند شده‌اند و یا هوشمندی آنها فراتر از انسان‌هاست. ممکن است کسی بگوید اگر چه این سیستم‌ها به یک معنا

هوشمند خوانده می‌شوند ولی نمی‌توان آنها را عامل[1] خواند. چرا که آنها دارای خودآگاهی نیستند. اما به نظر می‌رسد در برابر این اشکال می‌توان چنین استدلال کرد که چرا معیار عاملیت را خودآگاهی قرار دهیم. ممکن است بتوان معیار عاملیت را این‌طور تعریف کرد که از نگاه ناظر بیرونی یک شیء دارای عاملیت تشخیص داده شود؛ دقیقا همان معیاری که برای هوشمندی توسط تورینگ ارائه شده بود. به بیان دیگر، شاید همان‌طور که می‌توان هوشمندی را کارکردگرایانه تعریف نمود، عاملیت را نیز بتوان کارکردگرایانه توصیف کرد. حال اگر این‌طور باشد، می‌توان ربات هوشمندی را تصور کرد که از نگاه ناظران بیرونی هم هوشمند و هم دارای عاملیت تلقی شود. بنابراین، به نظر می‌رسد که با اتخاذ موضعی متفاوت نسبت به عاملیت (دیدگاهی که خودآگاهی را شرطی برای عاملیت نمی‌داند) بتوان از مسئول بودن خود ربات‌های هوشمند دفاع کرد. البته، همان‌طور که در بخش‌های قبلی نیز دیدیم، ممکن است کسی چنین اشکال کند که اینکه ربات را مسئول خطاها و حوادث ناشی از سیستم‌های هوشمند بدانیم، در صورتی امری معنادار است که بتوان آن را مجازات کرد و مجازات تنها در صورتی معنادار است که سیستم هوشمند رنج ببیند. اما ازآنجاکه سیستم هوشمند رنج نمی‌بیند بنابراین مجازات وی بی معنا بوده و لذا مسئول دانستن ربات امری بی‌فایده و گزاف است. اما همان‌طور که پیداست، این اشکال فرض گرفته است که مجازات یک فرد/شیء تنها در صورتی معنا دارد که آن فرد/شیء رنج ببیند. اما به نظر می‌رسد که می‌توان دیدگاه متفاوتی را نسبت به مجازات اتخاذ کرد. برای مثال، اگر فرض کنیم که هدف از مجازات

1. Agent

فرد خاطی این است که قربانی حادثه و یا بستگانش به لحاظ روانی به نوعی آرامش دست یابند، آنگاه به نظر می‌رسد که حتی در صورتی که خاطی رنج نبینند نیز مجازات محقق شده است. برای تقویت کردن این شهود شاید این مثال جالب باشد. بسیاری از ما این تجربه را داریم که کودکانمان در اثر برخورد با اشیاء پیرامون دردی را احساس کرده‌اند. برای مثال، هنگام راه رفتن، پای خود را به میزی کوبیده‌اند. بسیاری از والدین در این شرایط و برای تسلی خاطر کودکشان شیءای که کودک با آن برخورد کرده است را مجازات کرده و آن را مؤاخذه می‌کنند. برای مثال، به‌عنوان تنبیه، میز را با چند ضربه بر روی آن تنبیه می‌کنند. شاید آنچه ذکر شد، در مورد بزرگسالان هم صادق باشد. این نشان می‌دهد اینکه ایده و فلسفه پشت تنبیه آرامش روانی فرد آسیب‌دیده باشد، ایده بیراهی نیست. بنابراین، به نظر می‌رسد در صورتی که سیستم هوشمندی در یک سانحه مقصر شناخته شود می‌توان با از بین بردن وی که سبب آرامش روانی حادثه‌دیده (در صورت بودن در قید حیات) و یا خانواده وی می‌شود امر مجازات را محقق نمود. بنابراین، به نظر می‌رسد که اگر کسی دیدگاهی کارکردگرایانه نسبت به هوشمند دانستن یک شیء و نیز عامل دانستن یک شیء اتخاذ کند و نیز فلسفه مجازات را آرامش روانی حادثه دیده (یا بستگان وی) معرفی کند، وی به خوبی می‌تواند از ایده‌ی مسئول دانستن خود سیستم هوشمند دفاع کند. نگارنده در اینجا قصد ندارد که از این ایده دفاع کند. بلکه آنچه در بالا پیشنهاد شد، تنها به این معنی است که می‌توان کماکان پرونده مسئول دانستن سیستم هوشمند را باز نگه داشت.

در بخش‌های قبل بیان شد که از نظر برخی این برنامه‌نویسان هستند که باید مسئول شناخته شوند. چراکه اگرچه آنها مسئول رفتارهای محلی سیستم هوشمند نیستند ولی ازآنجاکه می‌توانند رفتارهای غیرمحلی آنها را پیش‌بینی کنند، می‌توان آنها را مسئول دانست. از نظر این گروه، اگر برنامه‌نویس سیستمی را طراحی کند که عملکرد آن نسبت به انسان بدتر باشد و با این وجود استفاده از آن سیستم را روا بدارد، آنگاه برنامه‌نویس مسئول خواهد بود. اما اگر سیستم طراحی شده توسط برنامه‌نویس نسبت به عملکرد انسان وضعیت بهتری داشته باشد، آنگاه وی در حوادث ناشی از سیستم هوشمند معذر خواهد بود.

به نظر می‌رسد در رابطه با این راه حل، یک ملاحظه بسیار اساسی وجود دارد که مدافعان آن باید به آن پاسخ گویند. ملاحظه این است که در این راه‌حل فرض گرفته شده است که برنامه‌نویسان در ابتدا سیستم را طراحی، آن را آزمایش و پس از گرفتن بازخوردهای مناسب آن را برای استفاده کاربران تجویز می‌کنند. اما به نظر می‌رسد که اساساً این نوع سیستم هوشمند دیگر شباهتی با سیستم هوشمند خودمختاری که قرار است در اثر تجربه بیاموزد ندارد. در واقع، در این راه‌حل فرض گرفته شده است که پس از مرحله‌ی آزمایش سیستم، سیستم دیگر نمی‌تواند یادگیری را ادامه دهد. یعنی، تنها یک یادگیری ابتدایی وجود دارد و بس. اما، مسئله‌ی اصلی شکاف مسئولیت درباره سیستم‌های هوشمندی است که با استفاده از تکنولوژی یادگیری عمیق می‌توانند از تجربه بیاموزند و در مرور زمان عملکرد خود را بهینه کنند. ولی اگر قرار است که یک

سیستم پیش از استفاده گسترده از آن و در مرحله‌ی آزمایشگاهی همه آنچه را که نیاز داریم بدانیم به ما نشان دهد، به نظر می‌رسد که دیگر با سیستم هوشمند خودمختار که از تجربه می‌آموزد سر و کاری نداریم. در واقع، سیستم‌هایی که مسئله شکاف مسئولیت درباره آنها سخن می‌گوید سیستم‌هایی پویا هستند که به مرور زمان عملکرد خود را بهینه می‌کنند. بنابراین، به نظر می‌رسد برنامه‌نویس واقعا نتواند حتی عملکرد غیرمحلی سیستم هوشمند را از قبل پیش‌بینی نما ید.

ملاحظه دیگری نیز درباره این راه‌حل وجود دارد. فرض کنیم راه‌حل فوق درست بوده و یک برنامه-نویس همه وظایف خود برای طراحی یک سیستم هوشمند را انجام داده و از قبل این مسئله را نیز امتحان نموده است که این سیستم هوشمند عملکرد بهتری از انسان‌ها دارد. حال، اگر سیستم هوشمند سبب بروز حادثه‌ای شود، از نگاه این راه‌حل برنامه‌نویس مقصر نخواهد بود. ازاین‌رو، به نظر می‌رسد این راه‌حل در چنین مواردی باید بر این مسئله صحه بگذارد که مواردی ازاین‌دست، می‌باید همچون حوادث طبیعی مانند سیل، زلزله و ... تلقی شود. این مسئله لازمه مهمی است که باید چنین راه‌حلی صراحتاً بر روی آن تاکید کند.

شاید بتوان با الهام از راه‌حل موردبحث در بالا و ملاحظه‌ی دوم که در بالا اشاره شد، پاسخ بدیعی را برای مسئله شکاف مسئولیت ارائه کرد. این راه‌حل ارتباط نزدیکی با یکی از نظریات در باب توجیه در حوزه معرفت‌شناسی دارد. نظریه اتکاگرایی[1] در معرفت‌شناسی بیان می‌کند اینکه یک باور به لحاظ معرفت‌شناختی موجه است تابعی

1. Relaibilism

است از اینکه آیا آن باور محصول یک فرایند شناختی قابل اتکاست یا خیر (Lemos, 2007, p. 85). بر اساس این دیدگاه، باور من به اینکه این گل قرمز است (هنگامی که به گل قرمزی نگاه می‌کنم) موجه است زیرا این باور حاصل یک فرایند شناختی (یعنی ادراک بصری) است که در اغلب موارد باورهای صادق تولید می‌کند (و لذا قابل اتکاست). این دیدگاه بر این مسئله تکیه دارد که عملکرد یک قوه شناختی چگونه است. اگر عملکرد آن به‌گونه‌ای باشد که در اکثر موارد باور صادق تولید کند آنگاه می‌توان بر باورهای تولید شده توسط آن قوه اتکا نمود و آنها را موجه دانست. حال، با الهام از این ایده در معرفت‌شناسی شاید بتوان در باب مسئله شکاف مسئولیت چنین گفت که اگر عملکرد سیستم هوشمند در زمینه‌ای خاص بسیار بهتر از عملکرد انسان باشد، آنگاه استفاده از آن سیستم نه تنها بلامانع است بلکه حوادث احتمالی برآمده از آن، ازآنجاکه کمتر از حوادثی است که به دست انسان‌ها به وجود می‌آید، همانند حوادث طبیعی مانند سیل و زلزله تلقی شده و نباید به دنبال مسئولی برای آن بود. در واقع، اگر یک سیستم هوشمند در اکثر مواقع بهتر از انسان عمل کند، قابل اتکا بوده و می‌توان از آن استفاده کرد. برای مثال، فرض کنیم که درصد حوادث رانندگی‌ای که در آنها رانندگان انسانی کنترل اتومبیل‌ها را برعهده دارند ۳۰ درصد باشد. یعنی، از هر ۱۰۰ سفر در طول جاده ۳۰ درصد از آنها به سبب خطاهای انسانی منجر به حادثه شده است. حال اگر عملکرد خودروهای خودران بهتر باشد، مثلا، از هر ۱۰۰ سفری که این خودروها در جاده داشته‌اند تنها ۵ درصد منجر به سانحه شده

است، آنگاه می‌توان نتیجه گرفت استفاده از خودروهای خودران به مراتب بهتر از استفاده از رانندگان انسانی است. بنابراین، هم می‌توان از این خودروها استفاده نمود و هم از خطاهای احتمالی ناشی از آن، تنها به خاطر این منفعت عظیم، چشم‌پوشی کرد و این حوادث را همچون حوادث طبیعی تلقی نمود.

حال اجازه دهید به ملاحظهٔ دیگری در باب شکاف مسئولیت بپردازیم. همان‌طور که دیدیم، مسئله‌ی شکاف مسئولیت می‌تواند برای قانون‌گذاران مسئله‌ی مهمی باشد. ازآنجاکه قانون‌گذران تا پیش از ظهور سیستم‌های هوشمند خودمختار قوانین را برای انسان‌های مختار تنظیم کرده بودند، به نظر می‌رسد نیاز است باتوجه‌به ظهور سیستم‌های هوشمند خودمختار، قوانین را بروز کرده و آنها را با شرایط امروز تطبیق دهند.

همان‌طور که دیدیم مسئولیت کیفری در نظام‌های حقوقی ایالات متحده آمریکا، آلمان و ایران طوری تنظیم شده است که بر اموری همچون ربات‌ها و سیستم‌های هوشمند صدق نمی‌کند. درست به همین خاطر است که کشورهایی مانند آمریکا و آلمان، همان‌طور که ملاحظه شد، در مواردی که سیستم‌های هوشمند دخالت دارند، مسئولیت کیفری را متوجه سازندگان و طراحان می‌کند.

همان‌طور که پیداست قانون‌گذار برای اینکه بتواند قوانین کیفری را در حوزه سیستم‌های هوشمند خودمختار وضع کند، در ابتدا نیاز دارد تا بداند که مسئولیت کیفری در حوادث ناشی از عملکرد بد سیستم‌های هوشمند خودمختار برعهده چه کسی است. این مسئله، همان‌طور که دیدیم، خود مبتنی بر اتخاذ موضعی فلسفی نسبت

به این مسئله است. این امر نشان می‌دهد که حوزه قانون‌گذاری تا چه اندازه به مواضع فلسفی وابسته است.

نکته جالب توجهی که در بخش «سیستم‌های هوشمند خودمختار از منظر حقوقی» وجود داشت این بود که کشورهایی مانند ایالات متحده آمریکا و آلمان در نظام حقوقی خود قوانینی را به طور خاص برای حوادث ناشی از سیستم‌های هوشمند خودمختار وضع نموده‌اند. اگرچه ممکن است این قوانین، باتوجه‌به مباحث فلسفی‌ای که در بالا اشاره شد، هنوز نوپا بوده و جای بحث داشته باشند، ولی مسئله این است که آنها جایی را برای سیستم‌های هوشمند خودمختار در نظام حقوقی خود باز کرده‌اند. اما این در حالی است که در نظام حقوقی ایران، هنوز خلاء قانونی درباره این مسئله به چشم می‌خورد. به بیان دیگر، در نظام حقوقی ایران هنوز قانونی بطور خاص و به‌صورت صریح برای حوادث ناشی از سیستم‌های هوشمند خودمختار وضع نشده است. از این‌رو به نظر می‌رسد حقوق‌دانان و قانون‌گذاران می‌باید در این رابطه، پیش از آنکه زمان را از دست بدهند، فکری کنند. از این‌رو، نیاز است تا حقوقدانان دست در دست فیلسوفان بتوانند قوانین درخوری را برای مسئله شکاف مسئولیت وضع نمایند. برای مثال، اگر راه‌حل اخیری که در بالا ارائه شد از نگاه فیلسوفان و حقوقدانان قابل قبول باشد، به نظر می‌رسد دیگر نیازی نیست که برنامه‌نویسان را مسئول بدانیم. در واقع، اگر برنامه‌نویسان و طراحان تمامی مسئولیت‌های اخلاقی، قانونی و شرعی را در ارتباط با سیستم هوشمندی که طراحی می‌کنند انجام داده باشند، و نیز اگر سیستم طراحی شده توسط آنها عملکردی به مراتب بهتر نسبت به

انسان‌ها داشته باشد، آنگاه به نظر می‌رسد می‌توان حوادث ناشی از سیستم‌های هوشمند خودمختار را همچون حوادث طبیعی تلقی نمود و برای آن مجازاتی را بر عهده‌ی برنامه‌نویسان قرار نداد.

جمع بندی

جمع‌بندی

در ایـن گـزارش در ابتـدا هـوش مصنوعـی خودمختـار را در حوزه‌هـای مختلـف از جملـه اتومبیل‌هـای خـودران، سیسـتم‌های تسـلیحاتی خودمختـار و سیسـتم‌های خودمختـار در پزشـکی معرفـی کردیـم. سـپس مسـئله‌ی شـکاف مسـئولیت را توضیـح داده و بیـان کردیـم کـه مسـئله شـکاف مسـئولیت بـه دنبـال پاسـخ بـه ایـن سـؤال اسـت کـه «مسـئولیت اقدامـات هـوش مصنوعـی بـر عهـده‌ی چـه کسـی اسـت». سـپس بـه پاسـخ‌های مختلفـی کـه بـه مسـئله‌ی شـکاف مسـئولیت داده شـده اسـت اشـاره کردیـم. راه‌حل‌هـای مختلـف در ابتـدا بـه دو دسـته‌ی کلـی تقسـیم می‌شـوند. برخـی بـر ایـن باورنـد کـه اساسـا هیـچ فـرد یـا چیـزی را نمی‌تـوان مسـئول اقدامات/حـوادث ناشـی از هـوش مصنوعـی دانسـت. در مقابـل دسـته‌ی دیگـر بـر ایـن باورنـد کـه فـرد یـا چیـزی را می‌تـوان مسـئول دانسـت. ایـن گـروه خـود بـه دو دسـته تقسـیم می‌شـوند. برخـی بـر ایـن باورنـد کـه مسـئولیت بـر عهـده‌ی انسـان (برنامه‌نویـس یـا کاربـر) اسـت. برخـی دیگـر بـر ایـن باورنـد کـه ایـن خـود هـوش مصنوعـی اسـت کـه بایـد مسـئول باشـد. در بخـش بعـد از نـگاه حقوقـی و مجـازات کیفـری بـه مسـئله‌ی شـکاف مسـئولیت نـگاه کردیـم و در ایـن راسـتا بـه بررسـی

سه کشور آمریکا، آلمان و ایران پرداختیم. در بخش بعد، ملاحظاتی فلسفی را در باب مسئله‌ی شکاف مسئولیت مطرح کردیم. از جمله‌ی این ملاحظات، ارائه‌ی راه‌حل بدیعی برای مسئله شکاف مسئولیت بود. بر اساس این راه‌حل (که بر گرفته از یکی از نظریات در باب توجیه در معرفت‌شناسی است) می‌توان گفت که اگر عملکرد سیستم هوشمند در زمینه‌ای خاص بسیار بهتر از عملکرد انسان باشد، آنگاه استفاده از آن سیستم نه تنها بلامانع است بلکه حوادث احتمالی برآمده از آن، ازآنجاکه کمتر از حوادثی است که به دست انسان‌ها به وجود می‌آید، همانند حوادث طبیعی مانند سیل و زلزله تلقی شده و نباید به دنبال مسئولی برای آن بود. در واقع، اگر یک سیستم هوشمند در اکثر مواقع بهتر از انسان عمل کند، قابل اتکا بوده و می‌توان از آن استفاده کرد.

باتوجه‌به مباحثی که در بالا مطرح شد، نگارنده پیشنهادهای راهبردی زیر را ارائه می‌کند:

- در نظام حقوقی ایران برخلاف ایالات متحده آمریکا و آلمان، قانون مستقلی برای حوادث ناشی از سیستم‌های هوشمند خودمختار وجود ندارد. ازاین‌رو، به نظر می‌رسد که قانون‌گذاران می‌باید فکری به حال این خلأ قانونی کنند.
- همان‌طور که دیدیم، مسئله قانون‌گذاری در حوزهٔ سیستم‌های هوشمند خودمختار، ربط وثیقی به مسئله‌ی شکاف مسئولیت، که مسئله‌ای فلسفی-اخلاقی است، دارد. باتوجه‌به اینکه فیلسوفان و نظریه‌پردازان داخلی حضور کمرنگی در این عرصه دارند، پیشنهاد

می‌شود که سیاستگذاران، پروژه‌های تحقیقاتی‌ای را برای مباحث نظری در این حوزه تعریف نمایند تا از خلال نقض و ابرام‌های فلسفی، حقوقدانان دست در دست فیلسوفان بتوانند به سمت ارائه قانونی شایسته قدم بردارند.

منابع

[۱] عطازاده، سعید؛ جلال انصاری. (۱۳۹۸). بازپژوهی مفهوم مسئولیت کیفری هوش مصنوعی (مطالعه موردی خودروهای خودران) در حقوق اسلام، ایران، آمریکا و آلمان. فصلنامه پژوهش تطبیقی حقوق اسلام و غرب. سال ششم، شماره چهارم.

[2] Brey, P. (2013). From moral agents to moral factors: The structural ethics approach. In P. Kroes, & P.-P. Verbeer (Eds.), The moral status of artifacts (pp. 125– 142). Springer.

[3] Bryson, J. J. (2010). Robots should be slaves. In Y. Wilks (Ed.), Close engagements with artificial companions: Key social, psychological ethical and design issues (pp. 63–74). Amsterdam: John Benjamins.

[1] Calverley, D. J. (2008). Imaging a non-biological machine as a legal person. AI & Society, 22(4), 523–537.

[2] Coeckelbergh, M. (2016). Responsibility and the moral phenomenonology of using self-driving cars. Applied Artificial Intelligence, 30(8), 748–757.

[3] Datteri, E. (2013). Predicting the long-term effects of human-robot interaction: A reflection on responsibility in medical robotics. Science and Engineering Ethics, 19(1), 139–160.

[4] Feenberg, A. (1991). Critical theory of technology. New York: Oxford University Press.

[5] Gladden, M. E. (2016). The diffuse intelligent other: An ontology of nonlocalizable robots as moral and legal actors. In M. Nørskov (Ed.), Social robots: Boundaries, potential, challenges (pp. 177–198). Burlington, VT: Ashgate.

[6] Gunkel, D. J. (2017). Mind the gap: responsible robotics and the problem of responsibility. Ethics and Information Technology, 1-14.

[7] Heidegger, M. (1977). The Question concerning technology and other essays (trans. by William Lovitt). New York: Harper and Row.

[8] Hevelke, A., & Nida-Rümelin, J. (2015). Responsibility for crashes of autonomous vehicles: An ethical analysis. Science andEngineering Ethics, 21, 619– 630.

[9] Lemos, N. (2007). An introduction to the theory of knowledge. Cambridge University Press.

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

[11] Purves, D., Jenkins, R., & Strawser, B. J. (2015). Autonomous machines, moral judgment, and acting for the right reasons. Ethical Theory and Moral Practice, 18(4), 851– 872.

[12] Sparrow, R. (2007). Killer robots. Journal of Applied Philosophy, 24(1), 62– 77.

[13] Swoboda, T. (2017, November). Autonomous Weapon Systems-An Alleged Responsibility Gap. In 3rd Conference on" Philosophy and Theory of Artificial Intelligence (pp. 302-313). Springer, Cham.

csri.majazi.ir

حوزه فضای مجازی به اندازه انقلاب اسلامی اهمیت دارد. این فضا مثل یک رودخانه پر از آب و خروشان است که می آید و دائماً هم بر آب آن افزوده و خروشان تر می شود. اگر ما بر این رودخانه تدبیر کنیم و برنامه داشته باشیم، زه‌کشی کنیم و هدایت کنیم این رودخانه را تا به سد بریزد، می‌شود فرصت. اگر رهایش کنیم و برنامه‌ای برای آن نداشته باشیم می‌شود یک تهدید.

csri.majazi.ir