بسمه تعالی

نشست ملی

پیامدهای اخلاقی هوش مصنوعی: ملاحظات اخلاقی در توسعهٔ مسئولانه نظام‌های هوش مصنوعی

نشست ملی **پیامدهای اخلاقی هوش مصنوعی: ملاحظات اخلاقی در توسعه مسئولانه نظام‌های هوش مصنوعی،** ۱۶ اردیبهشت ۱۴۰۵ در تالار ملل دانشکده مطالعات جهان دانشگاه تهران، به همت کمیسیون ملی یونسکو در ایران و با همکاری کرسی یونسکو در فضای مجازی و فرهنگ: دوفضایی شدن جهان برگزار شد. هدف محوری این برنامه، ایجاد فرصت برای بررسی و تبیین ابعاد گوناگون «توصیه‌نامه اخلاق در هوش مصنوعی یونسکو» بود که در سال ۲۰۲۱ منتشر شده است. همچنین، دو هدف محوری دیگر این نشست، ایجاد امکانی برای ارتقای دانش ملی-بومی در این حوزه نوظهور و فراهم‌سازی بستری برای هماهنگ‌سازی آراء و دیدگاه‌های صاحب‌نظران حوزه‌های مختلف اعلام شدند.

این نشست با حضور و سخنرانی دکتر حسن فرطوسی، دبیرکل کمیسیون ملی یونسکو در ایران، دکتر سعیدرضا عاملی، رئیس دانشکده مطالعات جهان دانشگاه تهران، دکتر حمید شهریاری، دبیرکل مجمع جهانی تقریب مذاهب اسلامی، دکتر علی کرمانشاه، عضو هیئت علمی دانشکده مدیریت و اقتصاد دانشگاه صنعتی شریف، دکتر حسین حسنی از پژوهشگاه فرهنگ، هنر و ارتباطات برگزار شد.

فهرست

افتتاحیه: سخنرانی آقای دکتر فرطوسی (کمیسیون ملی یونسکو در ایران)

در ابتدای این نشست، آقای **دکتر حسن فرطوسی**، دبیرکل کمیسیون ملی یونسکو در ایران، ضمن خیرمقدم به سخنرانان و حاضران، به اهمیت فن‌آوری‌های نوظهور، به‌ویژه هوش مصنوعی، در جهان امروز اشاره کرده و اظهار داشتند: امروزه در جهانی زندگی می‌کنیم که فن‌آوری های نوظهور به‌ویژه هوش مصنوعی و آثار مترتب بر آن زندگی ما را دگرگون ساخته است. تحولات شتابان فن‌آوری در حوزه هوش مصنوعی و گسترش کاربردهای آن در عرصه‌های اجتماعی فرهنگی رسانه‌ای و حکمرانی، توجه نهادهای ملی و بین‌المللی را به پیامدهای اخلاقی این فن‌آوری معطوف کرده است. هوش مصنوعی در سال‌های اخیر از یک ابزار فناورانه فراتر رفته به عنصری اثرگذار در فرایندهای تصمیم گیری، تولید دانش و سیاست گذاری عمومی تبدیل شده است امری که ضرورت توجه نظام‌مند به ملاحظات اخلاقی آن را برجسته می‌سازد.

دبیرکل کمیسیون ملی یونسکو در ایران در ادامه افزودند: سازمان آموزشی علمی و فرهنگی ملل متحد، یونسکو، به عنوان نهاد پیشرو در تدوین اسناد بین‌المللی اخلاقی، از جمله اخلاق هوش مصنوعی، نقش مهمی در هدایت گفتگوهای جهانی پیرامون توسعه‌ی مسئولانه‌ی این فن‌آوری ایفا کرده و بر اصولی چون عدالت، شفافیت، مسئولیت‌پذیری، احترام به کرامت انسانی و شمول اجتماعی تأکید دارد. تحقق عملی این اصول البته مستلزم انطباق آن‌ها با شرایط و اولویت‌های ملی است. جمع‌بندی کلی این نگاه در یونسکو در سال ۲۰۲۱ در قالب **توصیه‌نامه اخلاق در هوش مصنوعی** منتشر شد. یونسکو ذیل این سند تلاش کرده است توصیه‌نامه‌ی اخلاقی جامعی تدوین کند که بر حفظ کرامت و حقوق انسان‌ها و دفاع از آزادی‌های اساسی استوار است. هوش مصنوعی فرصت‌های بی‌نظیری برای افراد و نهادها فراهم کرده است اما همزمان چالش‌های جدی در این حوزه وجود دارد؛ چالش‌هایی مانند حفظ حریم خصوصی، تأمین امنیت داده‌ها و موضوع اخلاق در بهره‌گیری از هوش مصنوعی و تضمین دسترسی برابر به فن‌آوری برای همه. یونسکو هشدار می‌دهد که اگر این چالش‌ها مدیریت نشوند به جای کاهش مسائل و معضلات، باعث تشدید آن‌ها خواهد شد. این سند از سویی بستری برای گفتگوهای میان‌رشته‌ای میان متخصصان فراهم می‌آورد تا امکان شناسایی ظرفیت‌ها و چالش‌ها در این حوزه فراهم آید و از سوی دیگر، بایسته‌ها و هنجارها را در سطح ملی و بین‌المللی تبیین شود.

دکتر فرطوسی در انتهای صحبت‌های خود اشاره کردند که این نشست در امتداد ماموریت های کمیسیون ملی یونسکو در ایران طراحی شده است و هدف آن عبارت است از ارتقای ظرفیت نهادی در سطح ملی از خلال بررسی این سند توسط صاحب‌نظران و کارشناسان برجسته‌ی ایرانی در این حوزه. نهایتاً، ایجاد خروجی‌های موثر در راستای بکارگیری آن‌ها در فرایندهای سیاست‌گذاری‌های ملی در این حوزه کلیدی که امروزه تمام وجوه حیات ما را در بر گرفته است، هدف دیگر این نشست اعلام شد.

سخنران اول: آقای دکتر سعید رضا عاملی (دانشگاه تهران)

آقای **دکتر سعید رضا عاملی**، رئیس دانشکده مطالعات جهان و رئیس کرسی یونسکو در فضای مجازی و فرهنگ: دوفضایی شدن جهان، سخنران اول این نشست بودند. دکتر عاملی در سخنرانی خود با عنوان **ظرفیت‌ها و مخاطرات هوش مصنوعی و فن‌آوری‌های نو، اخلاق جنگ و احترام به انسان‌ها** اظهار داشتند: موضوعی را که به نظر من آمد برجسته کنیم ظرفیت و مخاطرات هوش مصنوعی و فن‌آوری نو در جنگ است. ما در واقع داریم بحث را به سمتی می‌بریم که ببینیم هوش مصنوعی چه اخلاق جنگی را به ما توصیه می‌کند. متناسب با احترام به انسان‌ها، متناسب با پرهیز از تخریب اماکن طبیعی و انسان‌ها. چه ظرفیت هایی باید مورد توجه ما باشد؟ طرح مسئله این است: ببینید، هوش مصنوعی ظرفیتی خودکار است، یعنی به لحاظ ارتباطی، هم ارتباط‌گر را حذف می‌کند و هم ارتباط‌گیر را. به یک تعبیری، هم ارتباط‌گر را حذف می‌کند و هم ارتباط‌گیران را. به معنایی دیگر، هم ارتباط‌گرها را حذف می‌کند و هم ارتباط‌گیران را یعنی رابطه‌ی یکی با یکی و یکی با همه و رابطه‌ی همه با همه را خودکار می‌کند و در واقع، الگوواره‌هایی که تحت سیطره‌ی الگوریتم‌ها طراحی می‌شوند این مسیر را فراهم می‌کنند.

در ادامه، رئیس دانشکده مطالعات جهان و رئیس کرسی یونسکو در فضای مجازی و فرهنگ افزودند: وقتی که هوش مصنوعی در ظرفیت جنگ قرار می‌گیرد، در واقع یک ظرفیت بدون روح انسانی و بدون مواجهه‌ی انسانی، در بستر جنگ حاصل می‌شود. چنین اتفاقی مخاطرات جنگ و عملیات جنگ را بسیار پیچیده می‌کند که در ادامه به آن خواهم پرداخت. همچنین، خطای الگوریتمی این‌جا می‌تواند نتیجه‌ای به بار بیاورد از جنس فاجعه. ظرفیت الگوریتم‌ها قدرت همزمانی دارد یعنی می‌تواند براساس پارامترهای طراحی شده، ده‌ها عملیات را به‌صورت همزمان اجرایی کند و از این جهت، خطرات کار را بیشتر می‌کند. در واقع، در گذشته فرماندهان نظامی هزینه‌ی انسانی جنگ را رؤیت می‌کردند، در میدان جنگ بودند و کشتار را می‌دیدند، خرابی‌ها را مشاهده می‌کردند و بوی باروت به مشام‌شان می‌خورد. ولی در جنگ هوشمند، فضای دیگری حاکم می‌شود: جنگ از راه دور است و تلفات جانی برای کسانی که در بستر جنگ قرار می‌گیرند، محسوس نیست. در واقع ما این‌جا بحث‌مان راجع به اخلاق جنگ در پیوند با هوش مصنوعی است.

ایشان اشاره کردند: اگرچه جنگ از بنیان نوعی تبعیض انسانی و تبعیض اجتماعی است، ولی با فرض این‌که جنگ آغاز می‌شود، باید جنگِ عادلانه صورت بگیرد. جنگی که متناسب با ضوابطی که در منابع بین‌المللی هم بر آن تأکید شده است و در حوزه تفکیک تناسب و کرامت انسانی مورد ملاحظه قرار گرفته است. شما ملاحظه کردید که وزیر جنگ آمریکا به وضوح از عدم رعایت اصول اخلاقی و منصفانه جنگ صحبت می‌کند. نقل قول از ایشان است که «ارتش کشورمان همیشه نوک نیزه برنامه‌هایمان بوده است و شش حوزه فن‌آوری حیاتی مورد توجه ما بوده است». معاون این وزیر تضمین می‌کند

که «سربازان ما هرگز وارد یک نبرد منصفانه نمی‌شوند مگر این‌که بهترین سیستم‌ها را برای حداکثر کشتار در دست داشته باشند». او ادامه می‌دهد که «وزارت جنگ متعهد است تا مرگبارترین نیروی جنگی روی سیاره زمین باقی بماند». منطق جنگ امریکایی، کشتار بزرگ است و در این سال‌ها هم نسل‌کشی‌هایی که توسط آمریکا صورت گرفته نسل‌کشی‌های بزرگی بوده‌اند؛ در کامبوج ۲۵ سال طول کشید، کشتار جمعی ویتنام ۲۱ سال طول کشید و از ۱۹۵۵ تا ۱۹۷۶ درگیر جنگ بودند. و تقریباً در این تاریخ ۲۵۰ ساله امریکا، به جز ۱۶ سال، بقیه سال‌ها را جنگیده‌اند و کشتار کرده‌اند. این فضایی است که بر جنگ غالب است.

دکتر عاملی در ادامه افزودند: اجازه بدهید چند ملاحظه را راجع به هوش مصنوعی عرض کنم. هوش مصنوعی اولاً بر بستر ماهیت رُقومی عمل می‌کند. ماهیتش ماهیت آنالوگ نیست. دیجیتال و آنالوگ ماهیت شیئی‌اند، مثل جنگ فیزیکی که ماهیت شیئی دارد. ولی ماهیت پدیده‌ی هوش مصنوعی یک ماهیت عمومی است منطبق بر منطق رُقومی. وقتی ماشینی با پارامترهای رُقومی سر و کار دارد، با الگوریتم های رُقومی سر و کار دارد. نکته دومی که راجع به ماهیت هوش مصنوعی باید گفت این است که از عنصر شبکه‌ای بودن در مقابل منفرد بودن برخوردار است. شبکه‌ای عناصر را به هم وصل می‌کند. شبکه ها را به‌هم متصل می‌کند و به تعبیری ماهیت موژولار دارد. ماهیت موژولار، ماهیتی است که می‌تواند ماژول‌ها را که هر کدام از چهار-پنج بخش ارتباطی برخوردار هستند به صورت یکی و یکی، یکی با همه و همه با همه مرتبط کند تا مفهوم پویای ماژولاریتی تحقق پیدا کند. نکته‌ی سوم این است که ماهیت همزمانی و پردازش موازی دارد. در هوش مصنوعی، پدیده جای پدیده را نمی‌گیرد. در آنالوگ پدیده جای پدیده را می‌گیرد. مثلاً شما در صف باید بایستید، باید یکی برود تا نفر بعدی نوبتش بشود. مثل زمان، فیزیکی است. زمان فیزیکی همانند موضوع مرگ و حیات است که یک نفر آنی می‌میرد و یک آن دیگری به وجود می‌آید، لذا گذشته و حال و آینده در آن معنا دارد. این‌جا ما با مفهومی به نام «وجود همه‌جا حاضر» مواجه هستیم که می‌تواند همزمانی را برای پدیده‌ها خلق کند. لذا قدرت پویایی کار در این‌جا قابل مقایسه با قدرت کار آنالوگ نیست. نکته چهارم با توجه به ریز فن‌آوری‌ها مانند ریزتراشه‌هایی که به وجود آمده‌اند (برای مثال نانو) مطرح می‌شود. ما امروز صحبت از کوانتوم می‌کنیم یعنی نانو یک ماهیت کوانتومی هم پیدا کرده است: معلق بودن و عدم قطعیت و ریزتراشه‌هایی که می‌تواند علم مقدار بین صفر و یک به وجود بیاورند، این پدیده را پیچیده‌تر می‌کنند. بهره‌مندی از زمان مجازی به نظرم عنصر مهمی است، این‌جا مسیر خطی نیست. همان‌گونه که اشاره شد، این‌جا در زمان مجازی، ما در سه تا عنصر فاصله و حرکت و سرعت، فاصله را حذف می‌کنیم. وقتی فاصله را حذف می‌کنیم، حرکت و سرعت هم بی‌معنا می‌شوند. همه در یک جا قرار می‌گیرند. مفهوم پویایی همه جا تحقق پیدا می‌کند و قابلیت کار مجازی را دارد.

ایشان با مروری تاریخی اشاره کردند: من عرایضم را در پنج بخش تنظیم کرده‌ام. ابتدا باید دوره‌بندی جنگ را بر اساس فن‌آوری‌های جنگ داشته باشیم تا ببینیم در دوره‌ی هوش مصنوعی چه اتفاقی افتاده است. دوره‌ی اولی که در جنگ داریم که آن را تعبیر به دوره صفر می‌کنیم یعنی ماقبل فن‌آوری. عصر عضله؛ یعنی هرکسی که زور بازوی قوی‌تری دارد و در

جنگ قدرتمندتر است می‌تواند مسیر بیشتری پیش ببرد. وقتی در جنگ وارد دوره‌ی دوم می‌شویم که از آن تعبیر به عصر شیمیایی، کشتار خطی و صنعتی یاد می‌شود. این‌جا ما با یک گسست بزرگ مواجه می‌شویم: با اختراع مسلسل. این‌جا برای اولین بار در تاریخ می‌بینیم که یک سرباز با سرعت ۶۰۰ گلوله در دقیقه و برد موثر بیش از هزار متر می‌تواند عملیات انجام دهد. این از اولین اقدامات صنعتی دوره جنگ بود که توانست کشتارهای بزرگی را به وجود بیاورد. هرچه فن‌آوری جنگ جلوتر می‌رود، امکان کشتار جمعی فراگیرتر می‌شود. دوره سوم، عصر دیجیتال و شبکه‌ای و در واقع جنگ از راه دور و عاری از رنج قربانی است. زمانی که جنگ چهره به چهره و فیزیکی است، هر دو طرف کشته می‌دهند و رنج قربانی وجود دارد، چه در طرف حمله‌کننده و چه در طرف حمله‌شده.

دکتر عاملی در ادامه افزودند: جنگ دوره‌ی سوم، جنگ از راه دور است یعنی جنگی که اساساً محنت شرایطی را که در میناب برای خانواده‌های ۱۶۸ کودک به وجود آمده را درک نمی‌کند. دکمه‌ای است که با یک اشاره، سه بار موشک را در یک مکان خاص می‌زند و جمعی را عزادار و دچار بحران می‌کند. دوره‌ی سوم را بعد از ۱۱ سپتامبر تعریف می‌کنند. بعد از حملات ۱۱ سپتامبر در واقع اصطلاح جنگ علیه ترور به وجود آمد. این نیز از امور جنگ شناختی است که سعی می‌شود غلبه‌ی معنایی ایجاد شود. دشمنان می‌توانند معنایی کاملاً دروغین را به یک سخن به‌ظاهر درست تبدیل کنند. کسانی که خودشان هزار کلاهک هسته‌ای آماده‌ی پرتاب دارند و سالیانه ۹۰ میلیارد دلار برای نگهداری این هزار کلاهک هسته‌ای هزینه می‌کنند (یعنی امریکایی‌ها)؛ آمده‌اند در تهران و اصفهان و در شهرهای دیگر ما، برای نابود کردن ظرفیت هسته‌ای ایران اقدام نظامی کنند؛ آن‌هم ظرفیت هسته‌ای صلح‌آمیزی که برای استفاده در بخش انرژی، پزشکی و مصالح صلح‌آمیز استفاده می‌شود. آن‌ها جنگ می‌کنند و جنگ افروزی می‌کنند که در نتیجه‌ی این رویکرد، صنایع زیادی به وجود آمده‌اند که من وارد این بحث نمی‌شوم. بعداً مقاله‌ای در اختیار حضار قرار می‌دهم. دوره چهارم، دوره‌ی عصر الگوریتم و خودکار شدن جنگ یا کشتار رایانه‌ای یا کشتار هوش مصنوعی است. وقتی از چت‌جی‌پی‌تی سؤال می‌کنیم که از چند پارامتر برای استخراج داده استفاده می‌کند، می‌گوید یک تریلیون پارامتر. هوش مصنوعی پارامتریک است و کاملاً محیط پارامتریک قابلیت الگوریتمی شدن دارد. الگوریتم چیست؟ الگوریتم می‌تواند هزاران متغیر را در یک بسته‌ی قطعی قرار بدهد. سیستم‌های خودکار بسته‌های قطعی هستند که در آن رحم و ترحم، اخلاق و حس و ادراک انسانی وجود ندارد. صنعتی است که اگر صنعت آنالوگ بود کسی باید پای ارابه‌ی آن صنعت می‌ایستاد و دکمه‌ها را فشار می‌داد و بر فرایند نظارت می‌کرد. در آن دوره باید ارابه را مدام تعمیر می‌کردند، استهلاک وجود داشت و مواردی از این دست. حال، در دنیایی بدون استهلاک است که همه‌چیز اتفاق می‌افتد. این‌جا دنیای ریاضی است که بر مبنای منطق پارامتریک عملیاتش را انجام می‌دهد. آغاز این دوره را باید ۲۴ فوریه ۲۰۲۲ و در قالب تهاجم روسیه به اوکراین دانست. این جنگ برای اولین بار در تاریخ، مقیاس‌پذیر بودن هوش مصنوعی در میدان نبرد واقعی را به نمایش گذاشت. باز اما تصمیم نهایی شلیک همچنان با یک اپراتور انسانی بود، منتها اپراتوری که در اتاق با کولر و امکانات و در بطن فضایی بدون جنگ تصمیم می‌گرفت که چه اقدامی صورت گیرد.

نکته‌ی دومی که دکتر عاملی در این بخش سخنرانی مورد بحث و بررسی قرار دادند موضوع جنگ و احترام به انسان بود. ایشان در این بخش اشاره کردند: هسته‌ی مرکزی حقوق بشردوستانه‌ی بین‌المللی بر این فرض بنیادین استوار است که حتی در خشونت سازمان‌یافته‌ی جنگ نیز انسانیت خط قرمز است و نمی‌توان از آن عبور کرد. کنوانسیون‌های ژنو ۱۹۴۹ و پروتکل الحاقی ۱۹۷۷، این خط قرمزها را در قالب اصولی چون تفکیک تناسب و احتیاط صورت‌بندی کرده‌اند. اما با ورود هوش مصنوعی به زنجیره‌ی تصمیم‌گیری نظامی، این پرسش اساسی مطرح است: آیا ماشینی که فاقد درک بافت انسانی، وجدان اخلاقی و مسئولیت‌پذیری حقوقی است می‌تواند احترام به این اصول را تضمین کند؟ پاسخ منفی است. در این‌جا سه اصل وجود دارد: اول اصل تفکیک است که در ماده ۵۱ پروتکل اول ژنو و ماده ۱۳ پروتکل دوم رمزگذاری‌شده و می‌گوید که طرفین درگیر باید همواره میان نظامیان و غیرنظامیان و همچنین اهداف نظامی و اشیای غیرنظامی که شامل ساختمان‌ها، پل‌ها، جنگل‌ها و طبیعت می‌شود، تفاوت قائل شوند. این پروتکل به حمله به اهداف غیرنظامی به طور مطلق اهمیت می‌دهد. ترامپ آن‌جا نشسته است و اولاً جنگ غیر مشروع (یعنی جنگ ابتدایی که بر اساس همه‌ی قوانین بین‌المللی ممنوع است) را آغاز می‌کند. طرفین می‌توانند از خودشان دفاع کند که آن هم دفاع مشروع است ولی جنگ ابتدایی نباید انجام شود و می‌دانیم که در فرهنگ تشیع، جنگ ابتدایی اساساً حرام است و جنگ ابتدایی فقط در دوران امام معصوم می‌تواند اتفاق بیفتد. بنابراین، این‌ها با هم تطبیق دارد یعنی دیدگاه الهی و قرآنی و دیدگاه نهادهای بین‌المللی، هر دو، جنگ ابتدایی را مشروع نمی‌دانند. دوم اصل تناسب است که در بند پنج ب ماده ۵۱ پروتکل اول ژنو آمده که حمله‌ای را ممنوع می‌کند که خسارتی به جان غیرنظامیان، از جمله جراحت آنان، آسیب به اشیای غیرنظامی و یا ترکیبی از این‌ها وارد آورد که در مقایسه با مزیت نظامی مستقیم و ملموس مورد انتظار زیاده‌روی نشده باشد.

دکتر عاملی چالش هوش مصنوعی در جنگ را ناتوانی آن در درک بافت انسانی دانسته و اشاره کردند: هوش مصنوعی نظامی دچار شکاف معنایی است به این معنا که الگوریتم، داده را مجموعه‌ای از پیکسل‌ها، سیگنال‌ها یا مختصات جغرافیایی می‌بیند اما معنای انسانی آن داده را درک نمی‌کند. کما این‌که بعضی از مکان‌ها را به اعتبار اسم آن بمباران کردند و اسم مدنظر فاقد بار معنایی لازم بود. دومین چالش در این زمینه مربوط به اصل تناسب است. در نظریه جنگ عادلانه و حقوق بشردوستانه‌ی بین‌المللی، تناسب ارزیابی، کیفی و مبتنی بر زمینه است. یک فرمانده‌ی انسانی باید عواملی چون اطمینان از اطلاعات، ابهامات موجود، ارزش‌های انسانی در خطر و حتی وجدان عمومی را در تصمیم خود دخیل کند. برای مثال یک فرمانده ممکن است خطر مرگ پنج غیرنظامی را برای کشتن ده سرباز دشمن بپذیرد اما اگر آن پنج غیرنظامی کودک باشند ممکن است همان خطر را برای کشتن ۲۰ سرباز نپذیرد. این موضوع مرتبط با بحث زمینه است، این‌که چه زمینه‌ای وجود دارد و در چه زمینه‌ای است که این ریسک پذیرفته می‌شود. در سیستم‌های خودکار، اراده انسانی حذف می‌شود. نکته سوم پرهیز از کشتار جمعی هوشمندانه است. بایدها و نبایدهای بسیاری در مورد کشتار جمعی وجود دارد. در دوره‌ای، بشر در هیجان و عصبیت جاهلیت زندگی می‌کرد که این دستکاری ذهن خودش بر او حکم می‌کرد که اگر میخواهی قدرت داشته باشی، بکُش. امپراتوری‌ها با همین کشتارها جلو رفتند ولی امروز که ما در قرن بیست و یکم میلادی قرار داریم و هزاره‌هایی از عمر بشر گذشته است، توقع می‌رود دیگر شاهد کشتارهای جمعی نباشیم.

دکتر عاملی اشاره کردند که کارشناسان دولتی ذیل کنوانسیون سلاح‌های متعارف از سال ۲۰۱۴ تاکنون مشغول بحث بر سر این مفهوم‌اند اما با وجود گذشت بیش از یک دهه هنوز به یک تعریف الزام‌آور دست نیافته‌اند. مفهوم «نظارت معنادار» یعنی در جنگ باید نظارت معنادار برای پرهیز از کشتار جمعی تحقق پیدا کند و باید تضمین‌هایی برای پرهیز از کشتار انسانی داده شود. با این حال، اجماع در حال رشدی در میان حقوقدانان بین‌المللی، جامعه مدنی و بسیاری از دولت‌ها شکل گرفته است که بر اساس آن، نظارت معنادار مستلزم آن است که یک انسان مسئول بتواند در زمان واقعی بر فرایند شناسایی، انتخاب و درگیری با هدف نظارت کند و این نظارت را نباید به نوعی تأیید نمادین تبدیل کرد. لذا زمانی که از نظارت‌پذیر کردن جنگ هوشمند و جنگ کشتار جمعی صحبت می‌شود، سه عنصر پُررنگ می‌شود:

(۱) اطلاع‌رسانی: به این معنا که تصمیم‌گیرنده‌ی انسانی باید به تمام اطلاعات مرتبط با هدف دسترسی داشته باشد و این اطلاعات نباید توسط الگوریتم فیلتر یا تحریف شود.

(۲) تسخیرناپذیری: یعنی مسئولیت تصمیم به کشتن هرگز نمی‌تواند به ماشین واگذار بشود. هوش مصنوعی ماشینی است که شناسایی را انجام می‌دهد و پس از پایش محیط به سوی هدف شلیک می‌کند. این تصمیم نباید به ماشین واگذار شود؛ به این معنا که انسان باید فرصت کافی برای بررسی توصیه‌ی الگوریتم داشته باشد و به الگوریتم تکیه نکند. لذا توصیه می‌شود یکی از توصیه‌هایی که شاید در جمع‌بندی این نشست بتوانیم بر آن تأکید کنیم و در واقع پیشنهاد علمی مشخصی است در راستای مهار رفتار ضداخلاقی هوش مصنوعی در جنگ، ضرورت ایجاد «قفل اخلاقی» است. ایجاد قفل اخلاقی در مدارهای تصمیم‌گیری نظامی صورت می‌گیرد. بر اساس این پیشنهاد اگر سیستم‌های هوش مصنوعی نتوانند با اطمینان ۹۹,۹ درصد غیرنظامی بودن یا نظامی بودن یک هدف را اثبات کنند، حمله به طور خودکار ممنوع می‌شود و باید مورد بررسی انسانی قرار گیرد.

دکتر عاملی در جمع بندی افزودند: تقویت اخلاقی الهی در جنگ و تلاش برای از بین بردن ذهنیت جنگ‌طلبی و سلطه‌گرایی شاید از تمام این قوانین بین‌المللی اهمیت بیشتری داشته باشد. این یعنی مهار سلطه‌گری در جهان. می‌بینید، ترامپ نه یونسکو و نه سازمان ملل را می‌پذیرد. در مقابل، نگاه ما چیست؟ اخلاق جنگ در سنت الهی که در زبان پاپ لیو هم شجاعانه طرح شده است، مقابله با این روحیه سلطه‌طلبی است که میراث عیسی و موسی و ابراهیم و انبیا الهی است. پیامبر ما هرگاه جنگ می‌خواست صورت بگیرد، می‌فرمودند: «از شما می‌خواهم به نام خدا و در راه خدا و بر روش پیامبر او حرکت کنید. به دشمنان خویش خیانت نکنید». اولین توصیه‌ی ایشان این است که: «به دشمن‌تان خیانت نکنید، به‌وقت مذاکره جنگ را آغاز نکنید، آن‌ها را مثله نکنید و مکر نورزید». پیامبر تأکید می‌کنند: «پیرمرد ضعیف و زن و کودک را نکشید؛ درختان را قطع نکنید مگر این‌که ناچار شوید؛ هرکس از مسلمین از کوچک و بزرگ توجهی به یکی از مشرکین داشته باشد و او را پناه دهد در امان است تا کلام خدا را بشنود. اگر از شما تابعیت کرد از برادران دینی شما محسوب می‌شود و اگر امتناع کرد او را به منزلگاه خود برسانید». رویکرد جنگ در اندیشه اسلامی، احترام به انسان، طبیعت و یک جنگ عادلانه است.

سخنران دوم: آقای دکتر علی کرمانشاه (دانشگاه صنعتی شریف)

در ادامه‌ی این نشست، آقای **دکتر علی کرمانشاه**، عضو هیئت علمی دانشکده مدیریت و اقتصاد دانشگاه صنعتی شریف، سخنرانی خود را با عنوانِ **طرح موضوع و چارچوب مقدماتی برای هوش مصنوعی اخلاقی و مسئولانه** ارائه دادند. ایشان تأکید کردند که بحث هوش مصنوعی باید در امتداد مفهوم تاریخی مسئولیت اجتماعی بنگاه‌ها بررسی شود. دکتر کرمانشاه اشاره کردند که در بحث از هوش مصنوعی که مربوط به یکی از تکنولوژی‌های دیجیتال است و به تحول کشور از یک پارادایم و عصر توسعه صنعتی به یک عصر جدید مربوط می‌شود، چارچوب نظری و عملیاتی که می‌تواند کمک‌کننده باشد، باید چارچوبی باشد که در واقع به زمینه‌های تاریخی بحث مسئولیت و عدم مسئولیت بپردازد و این تحول و پویایی را که در این یکی دو دهه‌ی اخیر اتفاق افتاده و از عصری به عصر دیگری به نام عصر دیجیتال، تبیین کرده و چارچوبی برای آینده پیشنهاد کند. البته قطعاً در این فرصت کم یک طرح اولیه‌ی بسیار ناقص قابل ارائه است اما با فضای فرهیخته‌ای که در جمع است انشاالله بتوانم در فرصتی مناسب، آن چارچوبی را که فکر می‌کنم می‌تواند تا حدی به این بحث کمک کند را تحت عنوان مسئولیت اجتماعی بنگاه و یا مسئولیت اجتماعی سازمان‌ها Corporate social responsibility (CSR) ارائه دهم.

دکتر کرمانشاه افزودند: در این‌جا مفاهیم مختلفی من‌جمله Business Ethics و بنگاه‌های پایدار مطرح می‌شود که کم و بیش ناظر به همان مفهوم کلیدی مسئولیت اجتماعی بنگاه‌ها است. این مفهوم از دهه‌های ۱۹۵۰ و ۱۹۶۰ وارد ادبیات مدیریت و اقتصاد شد و همچنان مورد استفاده قرار می‌گیرد و تحولات خودش را داشته است. این مفهوم ناظر به توجه و نگاه همزمان بنگاه‌ها و سازمان‌ها به مسئله اقتصاد و ملاحظات سهامداران و سود همزمان با توجه به فاکتورها و عوامل اجتماعی مثل انسان، محیط زیست، کرامت، عدالت و هر آن‌چیزی است که در واقع اجتماعی، غیراقتصادی، مرتبط به محیط زیست و مواردی از این دست است. یک نکته‌ی ظریف در بحث سی‌اس‌آر است که برخلاف بسیاری از زمینه‌های مدیریت که ما معمولا نمی‌گوییم سازمان‌های ناکارآمد، سازمان‌های بی‌کیفیت و سازمان‌های غیربهره‌ور، بلکه همه بیشتر سازمان‌ها را بهره‌ور، کارا و چابک می‌بینیم. سازمان تنبل خیلی در این زمینه جا نیفتاده است اما عجیب است که به همان اندازه که مسئولیت‌پذیری مطرح است، رفتار غیرمسئولانه هم مطرح است، یعنی تمام ساختارها، نهادها، مدل‌ها، شاخص‌ها و رتبه‌بندی‌ها، همزمان که همه ظرفیت‌ها را مورد نظارت قرار می‌دهند، مسئولیت‌پذیری را نیز پُررنگ می‌کنند و بهترین شرکت‌های مسئول را ردیف می‌کنند و با همان نگاه به رفتارهای غیرمسئولانه نیز توجه دارند و این نکته‌ی جالبی است.

این استاد دانشگاه در ادامه افزودند: وقتی سی‌اس‌آر و نهاد مسائل اجتماعی با این پدیده‌ی جدید برخورد می‌کند، دو نوع رابطه بین سی‌اس‌آر و اقتصاد دیجیتال مطرح می‌شود: (۱) سی‌اس‌آر و محصولات اجتماعی چه نقشی و چه برخوردی با این پدیده جدید دارند، چه مقدار در چارچوب و جعبه ابزار سی‌اس‌آر و نگاه نظری دیده می‌شوند و چه ادراکی از آن

می‌شود؟ به چه میزان مسئولانه و غیرمسئولانه شمرده می‌شوند؟ البته مثل هر پدیده‌ای که قدرت و ظرفیت‌هایی دارد این پدیده نیز هم در جهت مثبت و هم در جنبه‌ی منفی‌اش باید جدی و مهم تلقی شود. از منظر جنبه‌های غیرمسئولانه‌ای که مورد اشاره قرار گرفت، این بنگاه‌ها به همان نسبت که درواقع از طریق پلتفرم‌هایشان به ظرفیت میلیاردی مشتری دسترسی دارند اما اکثراً ساختارهای انحصاری دارند و شاید برخلاف بنگاه‌های سنتی که این منافع را فقط چند نفر نمی‌بردند، الان از کلیک میلیاردها انسان چند نفر در گوگل و فیسبوک و سایر شبکه‌های اجتماعی دارند منتفع می‌شوند. (۲) آن‌چه که امروزه تحت عنوان سرمایه‌داری پلتفرمی و در قالب نوعی نقد اقتصاد سیاسی مطرح شده است که به شاخص غیرمسئولانه عمل کردن این بنگاه ها به صورت جدی توجه می‌کند و در مقابل، دیدگاه بنگاه‌های پلتفرمی است. در این‌جا سوالاتی از این دست مطرح می‌شود که چرا کاربران، چه در بخش تولید و چه در بخش مصرف، نباید مالک آن پلتفرم باشند؟ این یعنی اتخاذ یک رویکرد انسانی. شمول و عدالت مبتنی بر دربرگیری همگان یکی از ظرفیت‌های جهان دیجیتال است اما این اصول جاهایی توسط این پلتفرم‌ها نقض می‌شوند. پرسش اساسی این است که چگونه می‌توان به جای راه‌حل‌های غیرمسئولانه، راه‌حل‌های همگانی و مسئولانه داشته باشیم؟ و برای پاسخ به این پرسش‌ها از پنجره‌ی دیجیتال و با لنز سی‌اس‌آر یک نگاه است، آن هم با ۴۰-۵۰- سال سابقه‌ی کار. چه نگاهی وجود دارد و به چه میزان این بروز دیجیتال را محصول کار خودش می‌داند؟ آیا می‌توان گفت این بنگاه‌ها دارای مدل کسب‌وکار مسئولانه هستند؟ بنابراین، اولین نقش و آورده‌ی سی‌اس‌آر در این پلتفرم‌ها عبارت است از تمام آن ظرفیت‌ها و تجربیاتی که می‌تواند در خدمت رتبه‌بندی و مواردی از این دست قرار گیرد. لذا یک بحثی داریم که طبق آن پرسیده می‌شود که سی‌اس‌آر به تعهد دیجیتال چه کمکی می‌کند؟ در ادامه‌ی آن، کاهش ریسک‌ها و رفتارهای غیرمسئولانه و افزایش بهره‌وری و توجه به حریم خصوصی داده مطرح می‌شوند.

دکتر کرمانشاه با اشاره به شیوع الگوریتم‌های سودمحور و جانبداری در داده‌ها تأکید کردند که امروزه در تحلیل داده‌ها یک حق جدید به نام «حق برداشت معقول از داده‌های افراد» ایجاد شده است. این یعنی الگوریتم‌ها حق ندارند با استفاده از داده‌ها اقدام به برداشت‌هایی کنند که غیرمنصفانه است. مثلا شهری‌ها یک چیزی بگویند راجع به روستایی‌ها و آن‌ها برداشتی داشته باشند و و یا در مورد زن‌ها، بچه‌ها و افریقایی‌ها. یعنی این داده‌ها درواقع اگر دارای سوگیری باشد، الگوریتم‌ها این حق را ضایع می‌کنند. بنابراین این‌ها ظرفیت‌های نهادهای مسئولانه است. حالا بعضاً این‌که در ساختارهای بین‌المللی این حد عملکرد غیرمسئولانه در برخی جاها رواج دارد باعث شده است که این حقوق مطرح شود.

بحث دیگری که دکتر کرمانشاه در این‌جا مطرح کردند این است که خودِ جهان دیجیتال چه کمکی به سی‌اس‌آر می‌کند. یعنی رابطه‌ی مقابل دیجیتال و ظرفیت‌های سی‌اس‌آر چیست؟ ظرفیت‌های سی‌اس‌آر به عنوان یک نهاد شصت‌هفتاد ساله چیست؟ ایشان در پاسخ به این پرسش‌ها اشاره کردند: این هم بحث مهمی است که معمولاً در پرداختن به آن دو رویکرد وجود دارد: این تکنولوژی های دیجیتال و بیزینس دیجیتال در برخی ابعاد بهبودهای تدریجی به سی‌اس‌آر می‌دهد که کارهایش را بهتر انجام بدهد. اگر می‌خواهد شفاف سازی، ریپورتینگ و رابطه بین بنگاه و جامعه را تنظیم کند، آن چه که

در سی‌اس‌آر است، به دیجیتال کمک می‌کند. این البته تا جایی می‌رسد که مفروضات سی‌اس‌آر بتواند بماند، حفظ بشود و دیجیتال را هم تأمین کند اما جایی می‌رسد که آنقدر ناکارآمدی در ساختارهای سی‌اس‌آر است که ناشی از نوع نگاه و تعریف بنگاه، ناشی از نوع نگاه به جامعه و ناشی از نوع نگاه یک طرفه از منظر بنگاه به جامعه است. بخش قابل توجهی از ادبیات سی‌اس‌آر از منظر بنگاه شکل گرفته است، چون سی‌اس‌آر را به هر حال بیزینس‌من‌ها شکل داده‌اند. آن‌ها بعضاً برای پوشاندن عملکرد غیرمسئولانه‌ی خود بسیاری از آن بنگاه‌هایی که در رتبه‌های بالایی قرار دارند را نادیده می‌گیرند. بسیاری از آن ها عملکرد غیرمسئولانه دارند و خیلی‌ها علاقه‌مندند که سیاست‌هایی خاص را تقویت کنند؛ برای این‌که سی‌اس‌آر بتواند روابط عمومی این‌ها را بهتر انجام دهد تا این بنگاه‌ها بتوانند همچنان عملکرد غیرمسئولانه‌شان را ادامه دهند. بنابراین، فرض این‌که از منظر بنگاه به جامعه نگاه کنیم، مفروضی است قابل قبول.

دکتر کرمانشاه تأکید کردند که بسیاری از ناکارآمدی‌های این مدل‌ها ناشی از نوع نگاه آن‌ها به جامعه و بخش قابل‌توجهی از آن ناشی از کهنه بودن آن مدل بود. بنابراین در منطق جدید می‌گویند بنگاه‌ها باید باز شوند و روی پلتفرم‌ها ارزش‌آفرینی کنند. می‌بینیم که در واقع در این اقتصاد جدید دیگر آن دوگانگی و جدایی و مرز تفکیک بین بنگاه و جامعه وجود ندارد. اساساً جامعه با بنگاه و جامعه یکپارچه شده است و ما این را به عنوان نمونه در دانشگاه‌های نسل چهارم و پنجم می‌بینیم. آنچه در این راستا در قالب مفهوم فرم مطرح می‌شود بدین معناست که فضای مجازی در واقع حقیقی یکپارچه با جامعه است. بنابراین، در چنین دنیا و فضایی است که مفهوم مسئولیت دیجیتال مطرح می‌شود که البته اشتراکاتی با سی‌اس‌آر داد اما راه خودش جدا است.

دکتر کرمانشاه در انتها اشاره کردند: این خلاصه‌ی بسیار مقدماتی بود که به هر حال فکر کردم به عنوان چارچوبی مطرح کنم درباره‌ی این‌که ما چگونه می‌توانیم سی‌اس‌آر را در مراکزی مثل یونسکو و این دانشکده برای توسعه‌ی ظرفیت‌های جدی امکان‌پذیر کنیم. متاسفانه در کشور ما بر خلاف خیلی از کشورهای حتی در حال توسعه‌ی دنیا از ظرفیت‌های بین‌المللی سی‌اس‌آر خیلی کم استفاده شده است. با همین مشکلات و ساختارهای ناقص دنیا ظرفیت چشمگری وجود دارد. اگر مثلاً ما خودمان را با هندی‌ها مقایسه کنیم متوجه می‌شویم چقدر ما در واقع از این ظرفیت‌ها کم استفاده می‌کنیم. امیدواریم در گفتمان مسئولیت دیجیتال بتوانیم هم‌پای دنیا جلو برویم و تعامل داشته باشیم و از ظرفیت‌هایش استفاده کنیم. انشالله بهتر بتوانیم راه تحولات دیجیتال را که مقوله‌ی پیچیده‌ای است بپیماییم. من نگرانم که مقوله‌های تکنولوژی‌های دیجیتال ارائه‌شده توسط ای‌آی و بلاک‌چین و مواردی از این دست را هم در جهت محکم‌تر کردن ناکارآمدی نهادی و ادامه دادن همان روندهای پیشین استفاده کنیم، یعنی مثلاً می‌دانیم که منطق بلاک‌چین نهادها را متحول می‌کند. نگرانی من این است که ما بلاک‌چین را هم به‌نحوی بومی کنیم که در نتیجه‌ی آن به این ساختارها دست نزنیم و اتفاقی نیفتد. بنابراین عملکرد مسئولانه دیجیتال این‌جا باید بسیار هوشیار عمل کند و به ماموریت مسئولانه‌اش عمل کند.

سخنران سوم: آقای دکتر حسین حسنی (پژوهشگاه فرهنگ، هنر و ارتباطات)

آقای دکتر حسنی سخنران سوم این نشست بودند و سخنرانی خود را با عنوان **مطالعه‌ی تطبیقی سند ملی هوش مصنوعی ایران با توصیه‌نامه اخلاق هوش مصنوعی یونسکو** ارائه دادند. ایشان اشاره کردند که به‌رغم کوتاه بودن فرصت برای آماده کردن و آماده شدن برای این سخنرانی، تلاش شده است خوانشی ابتدایی از دو سند مهم در حوزه هوش مصنوعی ارائه دهند. ایشان این دو سند را توصیه‌نامه‌ی اخلاق هوش مصنوعی یونسکو و سند ملی هوش مصنوعی ایران معرفی کردند که در سال ۱۴۰۳ توسط شورای عالی انقلاب فرهنگی تصویب شد.

دکتر حسنی ادامه دادند: یکی از ابعاد مثبت سند ملی هوش مصنوعی این است که اخلاق هوش مصنوعی در آن امده است. در ماده یک آن تعریف شده است و برای شاید اولین بار در یک سند رسمی و عملیاتی تعریفی از اخلاق هوش مصنوعی ارائه می‌شود که شامل مجموعه اصول اخلاقی برای هدایت توسعه و استفاده مسئولانه از هوش مصنوعی است، مبتنی بر ارزش‌های اسلامی هم است و در آن برخی از مسائل اخلاقی مانند حریم خصوصی، حقوق فردی و اجتماعی، امنیت اجتماعی، انصاف، شفافیت، عدم تبعیض، پاسخگویی، همسویی با هنجارهای جامعه اسلامی، مسئولیت پذیری، اعتماد و جلوگیری از سوءاستفاده از ابزارهای هوش مصنوعی ذکر شده است. در این تعریف، بهینه‌سازی، تاثیر سودمند هوش مصنوعی بر جامعه و کاهش مخاطرات و پیامدهای ناخواسته است. این است یک تعریف تقریباً جامع و مانع که تلاش کرده هم آن مضامین و مفاهیم جهانی اخلاقی را در نظر بگیرد و هم پس‌زمینه‌ی دینی و فرهنگی ما را مدنظر قرار دهد. در اصول و مبانی این سند هم به شکل‌های مختلف به بحث‌های اخلاقی پرداخته شده است از جمله التزام و ارزش‌های اسلامی، پرهیز از فساد و منکرات و غفلت‌زدایی انسان‌محوری (بر اساس فلسفه اسلامی)، توجه به ابعاد وجودی انسان، عدالت‌محوری، توجه به کرامت انسانی، سلامت جسمی و روانی، بحث حریم خصوصی و امنیت داده‌ها. همین‌طور بحثی در آن مطرح شده است با عنوان جلوگیری از سلطه که مراد عبارت است از سلطه‌ی انسان بر انسان و هوش مصنوعی بر انسان که بخشی از آن در بحث استفاده از هوش مصنوعی در ساخت جنگ افزارهای نوین هوشمند مطرح است.

ایشان بحث درباره‌ی این سند را در دو محور کلی ادامه دادند. در محور اول اشاره شد که در بحث سیاست‌های راهبردی در سند ملی هوش مصنوعی تلاش شده است به بحث‌های اخلاقی هم توجه بشود. بحث مردم-محوری و تعاون در آن مطرح است که سخنران قبلی، آقای دکتر کرمانشاه، هم به‌نوعی به آن اشاره کردند. یکی از مسئولیت‌ها یا اهداف پلتفرم‌های جایگزین این است که تأکید می‌شود خود کاربران نهایی پلتفرم‌ها بهتر از هر کسی می‌دانند. یعنی آن‌ها به جای این‌که برای مثلاً برای یک تاکسی اینترنتی مشهور کار کنند، خودشان یک سری تعاونی‌ها را شکل دهند و بتوانند از منافع آن بهره‌مند شوند. به جای این‌که بخش عمده‌ی این منافع نصیب دارندگان و صاحبان پلتفرم بشود و کاربران، بدون بیمه و خدمات دیگر، صرفاً کمیسیون اندکی دریافت کنند.

در محور دوم، ایشان افزودند: بحث هوش مصنوعی مسئولیت‌پذیر و صیانت از سرمایه انسانی نیز در این سند مطرح شده است؛ مسائلی مانند حافظه، اعتمادبه‌نفس و خلاقیت در مواجهه با هوش مصنوعی و مسائل دیگر حتی در بحث آموزش و پرورش و هم آموزش و پژوهش که در عمل بحث‌هایی اخلاقی‌اند که در این سند نیز به آن‌ها اشاره شده است. نمونه‌ای از این دست را نظام‌نامه‌ی اخلاقی دانست که وزارت علوم تدوین کرده است و به پژوهشگران و اساتید می‌گوید چگونه از ابزارهای هوش مصنوعی استفاده کنند. در این سند تأکید شده است که اگر ما برای انجام پژوهش از هوش مصنوعی استفاده می‌کنیم، حتما به آن اشاره کنیم. این سند بحث‌های دیگری هم دارد که من از آن‌ها می‌گذرم اما به طور خلاصه ابعاد اخلاقی اصلی این سندی که مصوب شورای عالی انقلاب فرهنگی است یکی همسویی با ارزش‌های اسلامی است؛ بحث توحید، عدالت، کرامت، اخلاق فردی و اجتماعی، حفظ حریم خصوصی و امنیت داده، شفافیت، انصاف و عدم تبعیض، مسئولیت پذیری، پاسخگویی، حفظ استقلال و منافع ملی در برابر سلطه فراملی، محافظت از انسان در برابر تهدیدهای هوش مصنوعی و توسعه‌ی اخلاق هوش مصنوعی به عنوان یک حوزه‌ی میان رشته‌ای. این خلاصه‌ای بود از ابعاد اخلاقی سند هوش مصنوعی جمهوری اسلامی ایران که مصوب شده و در دسترس همگان قرار دارد.

دکتر حسنی در ادامه به سند دوم، یعنی توصیه‌نامه‌ی اخلاق هوش مصنوعی یونسکو، پرداخته و اشاره کردند: سعی می‌کنم مقایسه‌ای بین این دو سند انجام می‌دهم. اول بحث مبانی فلسفی و در واقع منشأ ارزش‌های اخلاقی است. اگر شما به سند یونسکو نگاه کنید، مبتنی است بر نوعی قراردادگرایی سکولار و اسناد بین‌المللی حقوق بشر، از جمله همان اعلامیه‌ی معروف جهانی حقوق بشر، است که منشأ ارزش‌ها را اراده و توافق جمعی انسان‌ها می‌داند. فهم من این است که این سند کرامت ذاتی را مستقل از هر نوع جهت‌بینی دینی مطرح می‌کند. اصول کلیدی انسان را هم عدم تبعیض، مشارکت، شفافیت، مسئولیت‌پذیری و استفاده از نظارت انسانی می‌داند. حالا اگر ما به سند ملی هوش مصنوعی نگاه کنیم که مبتنی بر نوعی جهان‌بینی توحیدی است که در مقدمه و ماده دوی خود به آن اشاره شده است؛ منشأ ارزش‌ها را اراده‌ی خداوند، معارف اسلامی، فقه، حکمت و اخلاق اسلامی می‌داند. اصول کلیدی هم حیات طیبه، تقرب به خدا، کمال‌آفرینی، عدالت محوری و پرهیز از سلطه‌ی هوش مصنوعی بر انسان است. بنابراین ما شاهد یک تفاوت بنیادین در منبع ارزش‌ها هستیم؛ یک انسان قراردادی در برابر انسانی که در برابر شر و دین مسئولیت دارد. رویکرد اخلاقی سند یونسکو عمل‌محور است. ممنوعیت سیستم‌های امتیازدهی اجتماعی در آن پیشنهاد شده است، نوعی نظارت همگانی در آن دیده شده و طبقه‌بندی متریک معرفی شده است و تأکید بر قواعد و ممنوعیت‌های عملی برای جلوگیری از پیامدهای منفی در آن دیده می‌شود. حالا اگر ما به سندی که در ایران تصویب شد نگاه کنیم، هدف غایی‌اش کمال انسانی و قرب الهی در چارچوب تمدن نوین اسلامی و توجه به فضایل درونی مثل صداقت، وفای به عهد، بحث ساده زیستی، حسن ظن و اخلاق به‌عنوان هم‌سویی با ارزش‌های جامعه‌ی اسلامی تعریف شده است. این‌که با وجود اشتراک‌هایی که یک نوع دل‌مشغولی‌های ظاهری مثل حریم خصوصی، شفافیت و تبعیض وجود دارد؛ می‌بینیم که این دو تفاوت‌های بنیادینی دارند. هم در منبع ارزش‌ها و هم رویکرد اخلاقی آن‌ها و هم حوزه‌های اولویت آن‌ها. بنابراین پرسشی که برای من مطرح می‌شود، به‌عنوان کسی که ارتباطات خوانده است، این است که اگر قرار باشد یک سند مثل توصیه‌نامه‌ی اخلاق هوش مصنوعی را در چارچوب جمهوری اسلامی ایران پیاده

کنیم، چه باید کنیم؟ آیا ملزم هستیم که حتماً از این تبعیت کنیم؟ اگر این کار را انجام دهیم با تفاوت‌های مبنایی که داریم چه می‌کنیم؟ جمهوری اسلامی مبتنی بر یک نظام دینی است؛ اصلاً شورای عالی انقلاب فرهنگی چنین سندی را تصویب کرده و هدفش همین بوده است؛ در واقع این‌که بخواهد یک سری مبانی دینی را در این بحث پیاده کند. بسیاری از این ابعاد در سندی که من مطالعه کردم، علاوه بر آن اختلاف مبنایی، ما یک رویکرد مشخص را نمی‌بینیم. برای مثال اگر قوانین هوش مصنوعی رسانه‌های اجتماعی آمریکا را ببینید بحثِ آزادی فردی است و در اتحادیه اروپا بحث بیشتر درباره‌ی حقوق انسانی است.

دکتر حسنی افزودند: همه‌ی این‌ها برمی‌گردند به منشور حقوق بنیادین. به نظر می‌رسد که در این زمینه دچار چالشی اساسی هستیم. ما در خیلی از بخش‌های این سند می‌بینیم که بیشتر بحث صیانت از حکومت مطرح است؛ بحث، دفاع از نظارت دولت است و کمتر حقوق فردی مورد تأکید قرار گرفته است. در سنت ادبی‌مان مثلاً بحث کرامت انسانی را داریم که آیات و روایاتی در این‌باره مطرح است. به‌نظرم ما حتی می‌توانیم این را مبنای کارمان قرار دهیم. در متن مقدس‌مان به‌صراحت ذکر شده است که انسان خلیفه‌ی خدا یا جانشین خدا بر روی زمین است. به نظر من، مقام انسان خیلی فراتر از آن دیده شده است. ولی با وجود این می‌بینیم بسیاری از این قوانین به جای این‌که در واقعیت معطوف به استیفای حق فردی باشد، بیشتر توصیه‌ی جامعه نظارتی و محدود کردن اختیارات فرد است.

دکتر حسنی در انتها اشاره کردند: یک بحث عمده در هوش مصنوعی بحث حریم خصوصی است. تلاش برای تصویب یک قانون حریم خصوصی در ایران از سال ۱۳۸۴ آغاز شده و به‌رغم لایه‌ها و قوانین و تلاش‌های مختلفی که صورت گرفته است، هیچ قانون جامعی درباره‌ی حفاظت از حریم خصوصی و حریم داده‌ها نداریم آن‌هم به‌رغم این‌که ۲۰ سال از اولین تلاش‌ها در این زمینه سپری شده است. شما اگر به اصول اخلاقی پلتفرم‌ها و هوش مصنوعی نگاه کنید یک ارزش محوری، حریم خصوصی و همانطور متناسب با آن حریم داده است. انگار برای حاکمیت و برای خود فرد حریم خصوصی داده بی‌معناست. مفهوم حریم خصوصی و امنیت داده در ذهن ما به شکل تاریخی جاگیر نشده است تا مطالبه‌ی این حق را در واقع داشته باشیم. به نظر من حالا اگر بخواهیم بین آن چیزی که در زمینه اجتماعی و فرهنگی ما می‌گذرد و آن چیزی که در سطح بین‌الملل در قالب توصیه‌نامه‌هایی مثل توصیه‌نامه‌ی اخلاق هوش مصنوعی یونسکو می‌گذرد، باید یک راه دیگری پیدا کنیم که بتوانیم این‌ها را به هم پیوند بزنیم. این خیلی فراتر از آن نگاه دیستوپیایی یا اتوپیایی نسبت به تکنولوژی است. در کانتکس ما یک سری بحث‌ها مانند توحیدگرایی و بحث‌های دینی هم مطرح می‌شود که حتماً باید در مواجهه با فن‌آوری و به‌اصطلاح اقتباس آن مورد توجه قرار دهیم.

سخنران چهارم: آقای دکتر حمید شهریاری (سازمان مطالعه و تدوین کتب دانشگاهی)

سخنران آخر این نشست، **آقای دکتر حمید شهریاری**، رئیس سازمان مطالعه و تدوین کتب دانشگاهی و دبیرکل مجمع تقریب مذاهب اسلامی بودند که با عنوان **چارچوب اخلاقی در توسعه‌ی هوش مصنوعی** به سخنرانی خود پرداختند. ایشان در ابتدا اشاره کردند که ما اگر بخواهیم در حوزه اخلاق و ملاحظات اخلاقی در حوزه‌ی هوش مصنوعی سخن بگوییم، یک پرسش فلسفی پیشینی وجود دارد و آن ناظر به نسبت انسان با ماشین‌های هوشمند است. این پرسش فلسفی که آیا ماشین در نهایت می‌تواند جای انسان بنشیند یا خیر و آیا انسان به یک ماشین تبدیل می‌شود یا خیر، این‌ها پرسش‌هایی فلسفی‌اند که پاسخش طبق فلسفه‌ی اسلامی منفی است یعنی هیچ وقت ماشین جای انسان را نخواهد گرفت. هرچند هم که ماشین پیشرفته شود و هوش مصنوعی قوی‌تر باشد؛ در نهایت و در قالب آینده‌نگاری ممکن است از اینکه چه اتفاقی می‌افتد بتوانیم فیلم‌های هالیوودی بسازیم مثل آن‌که هوش مصنوعی بدل به انسان می‌شود، باز هم جای انسان را می‌تواند بگیرد و مواردی از این دست.

دکتر شهریاری افزودند: اولین دلیل فلسفی این مدعا آن است که انسان دارای روح است و فقط جسم نیست، فقط رفتارهای جسمانی ندارد و این روح یک وجه متمایز فلسفیِ ماشین است. البته انسان ممکن است در روحیه‌اش ماشین وارد شود؛ کما این‌که الان ممکن است در بعضی جاها روحیه‌اش سگ‌واره شده باشد، یعنی انسان با انسان ارتباط می‌گیرد و با حیوانات در درجه‌ی دوم ارتباط می‌گیرد، منتها این در بعضی جوامع به مرحله‌ی سقوط رسیده است یعنی انسان‌ها بیشتر دوست دارند با سگ‌ها انس بگیرند. ممکن است زمانی هم برسد و قابل پیش‌بینی باشد که انسان‌ها بیشتر دوست داشته باشند با ماشین‌ها ارتباط بگیرند تا با انسان‌های دیگر. این یک تنزل است که در هویت انسانی ممکن است رخ دهد. به‌خوبی اشاره فرمودند که پاسخ دوم به این سوال که آیا ماشین می‌تواند انسان بشود یا نه، که پاسخ منفی دارد، برمی‌گردد به یک وجه ماهیتی. ماهیت کرامت انسان شاید قوی‌ترین مفهوم اخلاقی باشد که ما در اخلاق می‌توانیم از آن استفاده کنیم.

ایشان در ادامه اشاره کردند: یکی از اشکالات کلی که بر همه‌ی سندهای وارد است، این است که ما نتوانستیم فضایل اخلاقی انسانی را طبقه‌بندی کنیم. علمای حوزه اخلاق نتوانستند بگویند بالاترین ارزش انسانی چیست که دیگر هیچ ارزشی را نباید فدای او کرد. بعضی‌ها گفتند تقوا. حالا ما می‌گوییم کرامت چیست؟ بعد در درجه‌ی دوم چه چیزی را قرار بدهیم؟ درجه سوم چه چیزی را قرار بدهیم؟ این درجات به چه درد می‌خورد؟ وقتی ما سندها را می‌نویسیم می‌توانیم اولویت‌گذاری کنیم. طبق این طبقه‌بندی ارزش نمی‌توانیم بفهمیم که اگر پولی آمد و ما خواستیم خرجش کنیم، خرج هوش مصنوعی کنیم یا خرج مثلاً رفتارهای اجتماعی کنیم؟ به نظر من اگر ما یک هرم ارزش‌ها داشته باشیم و در رأس آن بخواهیم یک فضیلت یا یک ارزش اخلاقی را قرار بدهیم که بالاتر از آن دیگر هیچ چیزی وجود ندارد، به نظر من، این‌جا با دو فضیلت روبه‌رو هستیم: اول کرامت که مربوط به هویت انسان است و دوم عدالت که مربوط به ارتباط انسان با انسان است. این دو ارزش در راس این هرم قرار می‌گیرند و مفاهیم عامی هستند که شما در همه‌ی مجالس بین‌المللی می‌توانید به عنوان ارزش از آن حساب ببرید. مبانی اسلامی هم دارد که در این جلسه به آن ورود نمی‌کنم.

دکتر شهریاری با اشاره به اثر خود با عنوان «اخلاق فن‌آوری اطلاعات» اشاره کردند: یکی از اصول مطرح‌شده مرتبط با اخلاق هوش مصنوعی است. در آنجا پنج نکته بیان شد که در این فرصت، دو مورد از آن‌ها را خدمتتان ارائه خواهم داد. ما اصول اخلاقی داریم که باید حاکم بر هوش مصنوعی شوند. مقدمتاً نکته‌ای دقیق را خدمت حضار محترم عرض می‌کنم که نکته‌ای بسیار مهم و ظریف است. هنگامی که از اصول اخلاقی در حوزه‌ای خاص بحث می‌کنیم، که اصطلاحاً در مباحث اخلاق کاربردی مطرح می‌شود، با دو دسته ارزش مواجهیم: نخست، ارزش‌های عامی که در هر حوزه‌ای وارد می‌شویم، وجود دارند؛ و دوم، ارزش‌هایی که مختص همان حوزه‌ی خاص هستند. به‌عبارت‌دیگر، هنگامی که می‌خواهیم ملاحظات اخلاقی هوش مصنوعی را تبیین کنیم، می‌توانیم از مفاهیمی چون صداقت، شجاعت، امنیت و عدالت سخن بگوییم. ما مجموعه‌ای از مفاهیم داریم که در هر حوزه‌ی کاربردی قابل‌اجرا هستند؛ این‌ها اصول اخلاقی عام محسوب می‌شوند. اگر بخواهیم ملاحظات اخلاقی در هوش مصنوعی را مطرح کنیم، نباید اخلاق اسلامی را پیش‌رو قرار دهیم و به شمارش فضایل اخلاقی بپردازیم. فضایل اخلاقی در همه‌جا باید رعایت شوند و هنگامی که هوش مصنوعی وارد هر حوزه‌ای می‌شود، تمامی این فضایل باید مورد توجه قرار گیرند.

دکتر شهریاری در ادامه اشاره کردند: پس هنگامی که می‌خواهیم ملاحظات اخلاقی در هوش مصنوعی را بررسی کنیم، چه باید بگوییم؟ دو نوع ملاحظه‌ی اخلاقی را باید مطرح کنیم: نخست، ملاحظات اخلاقی اختصاصی که ویژه‌ی آن حوزه است؛ یعنی در جای دیگری یا وجود ندارد یا موارد محدودی دارد. هنگامی که ملاحظات اخلاقی در هوش مصنوعی مدنظر باشد، اگر کتاب‌ها و منابع غربی را مطالعه کنید، مانند کتابی که توسط چهارده نفر تدوین شده و تمامی منابع را تحلیل کرده و کلیدواژه‌های ارزشی را استخراج و فهرست‌بندی کرده‌اند، درمی‌یابید که مفهوم «شفافیت» بیش از سایر مفاهیم تکرار شده است. مفهوم شفافیت از جمله مفاهیم خاص حوزه‌ی هوش مصنوعی و از ملاحظات ویژه‌ی اخلاقی این حوزه است که در بسیاری از منابع به آن اشاره شده است. بنابراین، این مفهوم از اولویت اول برخوردار است؛ زیرا از نظر آماری در ذهن تمامی پژوهشگران بوده که این مفهوم باید مورد توجه قرار گیرد. پس این می‌شود ملاحظه‌ی اخلاقی در هوش مصنوعی؛ اما «صداقت» چنین نیست. این دسته را «اختصاصی» می‌نامیم، زیرا «اختصاصی» با «انحصاری» تفاوت دارد؛ یعنی نمی‌گوییم انحصاراً در هوش مصنوعی وجود دارد، بلکه مزیت اخلاقی است که ممکن است در دو یا سه حوزه‌ی دیگر نیز شفافیت بسیار مهم باشد، اما شفافیت در همه‌جا جریان ندارد. برعکس، گاهی در حوزه‌ی امنیت سخن می‌گویید و گفته می‌شود که شفافیت در آنجا مطلوب نیست. این یک دسته است: ارزش‌ها و ملاحظات اختصاصی.

دکتر شهریاری با اشاره به ملاحظات عام اخلاقی اشاره کرده و افزودند: ملاحضات عام اخلاقی داریم، منتها با یک شرط: همان صداقتی که عرض شد، نباید به‌عنوان ملاحظه‌ی اخلاقی در هوش مصنوعی مطرح شود؛ مگر آنکه شرطی به آن افزوده شود که در آن صورت می‌توان در هوش مصنوعی درباره‌ی آن بحث کرد. آن شرط این است که آن مفهوم عام اخلاقی، خصوصیتی در این حوزه داشته باشد که آن خصوصیت، موجب ذکر آن شود؛ وگرنه تمامی ارزش‌های اخلاقی را هنگام ورود به هوش مصنوعی باید رعایت کنیم. مگر می‌شود انسان برخی فضایل را در هوش مصنوعی رعایت کند و برخی دیگر را نه؟

همه را باید رعایت کند. اما چه چیزی را به‌عنوان ملاحظات اخلاقی هوش مصنوعی مطرح کنیم که ویژگی خاصی داشته باشد؟ مثال می‌زنم: حریم خصوصی در حوزه‌ی فن‌آوری اطلاعات همه‌جا باید رعایت شود؛ یعنی یک ارزش انسانی است که انسان‌ها باید حریم خصوصی یکدیگر را محترم بشمارند.

دکتر شهریاری در انتهای سخنرانی خود افزودند هفت اصل یا هفت ارزش محوری را که ملاحظات اصلی در هوش مصنوعی هستند را این‌گونه طرح کردند: نخست آنکه هوش مصنوعی هنگام توسعه، نباید با وجدانیات خیر و شر بشر منافات داشته باشد؛ یعنی بشر دانشی راجع به خیر و شر دارد که نباید توسعه، منافات با وجدانیات بشر، یعنی کل بشر در امور خیر و شر، داشته باشد. اصل دوم، تأکید بر خودتنظیم‌گری است؛ یعنی هر قدر هم اصول و ملاحظات اخلاقی را تبیین کنید، تا زمانی که فرد فعال در این حوزه، الزامات درونی برای خود نداشته باشد، در نهایت اخلاق از آن حاصل نمی‌شود. همه باید خود متعهد باشند؛ این خودتنظیم‌گری نیز باید مورد اشاره قرار گیرد. مقررات بیرونی، چنانچه فاقد انگیزه‌های درونی باشد، به‌خودی‌خود نمی‌تواند این فن‌آوری را در مسیر خدمت به بشریت قرار دهد. سوم آنکه داده‌های انسانی خطاپذیرند که باید از آن پرهیز شود. یک الزام مهندسی ایجاد شود که این داده‌های خطاپذیر، منشأ دستگاه هوشمند قرار نگیرند. هوش مصنوعی نباید موجب ضرر و زیان شود؛ این به‌عنوان یک اصل مطرح است. حال این‌که می‌گوییم باید این ضرر و زیان را تبیین کنیم، باید بیان کنیم که اگر ضرر یکی به منفعت دیگری منجر شد، چگونه تعدیل کنیم؟ این موارد را اکنون فرصت نیست که تفصیل دهم. بحث ضرر و زیان—که چه کسی باید رعایت کند: مهندسان، سیاستمداران، حاکمان و سرمایه‌گذاران—این چهار رکن باید بررسی کنند که ضرر، اصول و قواعد پیشینی که مقبولیت همگانی دارد و ملاحظات خاص دارد، نیز باید رعایت شود؛ که عدالت، برابری، آزادی و احترام به دیگران و این موارد را با ملاحظه‌ی خاص باید مدنظر قرار داد. اصل ششم آن است که باید آینده‌نگاری داشته باشیم؛ یعنی باید مطالعه کنیم و ببینیم این فن‌آوری به کجا می‌رود؟ تا ندانیم به کجا می‌رود، نمی‌توانیم ملاحظات اخلاقی آن را بیان کنیم. پس ما نیازمند عالمان آینده‌نگار هستیم تا به ما بگویند که این فن‌آوری به کجا می‌رود. آنگاه می‌توانیم بگوییم ملاحظات تفکیکی آن چیست؟ این هم نکته‌ی اصلی آن است. اصل هفتم نیز پیشگیری از انحصاری‌بودن است.

ایشان در انتهای سخنرانی اشاره کردند: بر مبنای این هفت اصل، هفت ارزش را که بعضاً مجبور به شناخت آن‌ها شده‌ایم و برخی دیگر نیز بیان می‌کنند، عرض می‌کنم: اول، شفافیت است؛ یعنی در انتها، مرا به کجا می‌برد؟ چگونه طراحی شده؟ با من چه می‌کند؟ چرا گاه خطا می‌گوید و من می‌پرسم چرا خطا گفته است؟ هوش مصنوعی می‌گوید: «ببخشید»؛ آیا این خطاگویی، در دل آن نهفته است یا عامدانه انجام می‌دهد؟ آیا می‌خواهد مرا مطالعه کند و به جهتی سوق دهد یا نه؟ این‌که چه اصولی بر آن موتور هوش مصنوعی حاکم شده و چه اهدافی را دنبال می‌کند، باید شفاف شود. هرکس هر ماشینی می‌سازد، باید بگوید اهداف من از ساخت این ماشین چیست. آیا واقعاً فقط می‌خواهد واقعیت را به ما نشان دهد یا نه؟ آیا پشت این نشان‌دادن واقعیت، سوگیری، جهت‌گیری و سودجویی‌هایی وجود دارد؟ به من بگوید؛ نمی‌گویم سوگیری نداشته باشد، همه‌ی ما سوگیری داریم. این‌ها به ترتیب اهمیت: ارزیابی خطرها، مهارپذیری خطرها و جبران خسارات، سه زنجیره

هستند. یعنی نخست باید بتوانم خطر این ماشین را ارزیابی کنم که چقدر خطرناک است، خطر دارد یا ندارد، خطرش در کجاست؟ این باید معلوم شود تا بتوانم خطر را مهار کنم؛ یعنی ابزارهایی در اختیار من باشد. ماشین نباید عکس مرا بگیرد و سپس شلیک کند. من چه می‌توانم بکنم؟ سوم آنکه اگر خطایی از آن سر زد، چگونه می‌توانند آن ضرر و زیان و آن خطر را جبران کنند؟ این اصل دوم است. حال توضیحاتی دارد. سوم رعایت برابری و انصاف است، زیرا اکنون برخی معتقدند که ابزارهای هوشمند به نفع سرمایه‌داران جهانی فعال می‌شوند، زیرا اطلاعات را جمع‌آوری کرده و در اختیار سرمایه‌داران و استکبار جهانی قرار می‌دهند و درنهایت، آن‌ها حاکم جهان خواهند شد. این سخن پوتین است. پوتین گفته است: «هر کس در هوش مصنوعی در دهه‌های آینده رشد پیدا کند، او حاکم جهان خواهد شد.» نباید چنین شود. چهارم، حفظ حریم خصوصی که توضیح آن داده شد. پنجم، حمایت از حیات، سلامت و رفاه انسان‌ها؛ یعنی ما ابزارهای جنگی می‌سازیم، ابزارهای هوشمند جنگی می‌سازیم؛ باید به‌گونه‌ای ساخته شوند که کل محیط انسانی و جغرافیای زمین را به خطر نیندازند. جغرافیا یکی است؛ اکنون با چه دقتی نقطه‌زنی می‌کنند؟ این همه خسارت انسانی معین و مشخص به ما وارد کردند؛ این نیز مسئله‌ای است که باید در ملاحظات رعایت شود. توجه به عواطف انسانی در عرضه‌ی کالا و خدمات؛ یعنی با عواطف ما بازی نکنند، کالایی را بهانه نکنند، خدمتی را به ما تحمیل نکنند که استفاده کنیم. ما همچنان بتوانیم آزادانه و با عواطف تصمیم بگیریم. ماشین هرقدر هم رشد پیدا کند، درد را نمی‌بیند. درد، خاص انسان است؛ چه جسمی و چه روحی. هیچ ماشینی، هر چقدر هم پیشرفت کند، احساس درد نخواهد کرد؛ نه جسمی و نه روحی؛ اما انسان احساس درد دارد، هم جسمی و هم روحی. بنابراین، خدماتی که در حوزه‌ی کالا و خدمات عرضه می‌شود، باید به عواطف انسانی توجه کند. و در نهایت، یک سازوکار نظارتی باید بر رفتار ربات‌های هوشمند و ماشین‌های هوشمند حاکم باشد؛ زیرا بالاخره نمی‌دانیم چه می‌شود؛ نمی‌دانیم آینده‌اش چه خواهد شد. حال آینده‌پژوهان بیایند و بگویند چه می‌شود. اگر هم بگویند، به‌صورت احتمالی می‌گویند. احتمالاتی وجود دارد که بعداً اتفاقاتی رخ دهد؛ لذا ما نیازمند دستگاه ناظری، منظورم دستگاه جهانی ناظر با طبقه‌بندی، هستیم که مراعات کند. ببینیم به کجا می‌رود و احیاناً به‌موقع تصمیمات درست بگیریم تا آن مشکلات به وقوع نپیوندد.