



United Nations  
Educational, Scientific  
and Cultural Organization



UNESCO Chair on Studies in Bioethics (UCSB),  
Medical Ethics and History of Medicine Research  
Center, Tehran University of Medical Sciences



## توصیه نامه یونسکو

### در خصوص اخلاق در هوش مصنوعی

#### مقدمه

در چهل و یکمین نشست کنفرانس عمومی سازمان آموزشی، علمی و فرهنگی ملل متحد (یونسکو) در تاریخ ۹ تا ۲۴ نوامبر در پاریس به موارد زیر پرداخته شد:

بررسی اثرات عمیق و دینامیک مثبت و منفی هوش مصنوعی بر جوامع، محیط زیست، اکوسیستم‌ها، زندگی و اندیشه انسانی، که از نتایج تاثیرات نوین هوش مصنوعی بر تفکرات، تعاملات و تصمیم‌گیری‌های انسان‌ها بوده و علوم آموزشی، انسانی، اجتماعی و طبیعی و همچنین فرهنگ، ارتباطات و اطلاعات را تحت تأثیر قرار می‌دهد.

یادآوری این نکته مهم است که بر اساس نظامنامه یونسکو، هدف این سازمان گسترش صلح و امنیت از طریق همکاری میان کشورهای دنیا در زمینه‌های آموزشی، علمی، فرهنگی، ارتباطات و اطلاعات و برقراری عدالت جهانی، حاکمیت قانون، حقوق بشر و آزادی‌های بنیادین شناخته‌شده توسط ملت‌های دنیا، می‌باشد.

توصیه‌نامه ارائه شده در این نوشتار، ابزاری برای تعیین استانداردهایی بر پایه رویکردی جهانی مبتنی بر قوانین بین‌المللی و متمرکز بر کرامت انسانی، حقوق بشر، برابری جنسیتی، عدالت و توسعه اجتماعی و اقتصادی، سلامت جسمی و روحی، گوناگونی، همبستگی، فراگیری، و حفاظت از محیط زیست و اکوسیستم می‌باشد که

می‌تواند هوش مصنوعی را به شکلی مسئولانه هدایت کند. این توصیه‌نامه منطبق بر اهداف و اصول منشور حقوق بشر می‌باشد.

علیرغم این که هوش مصنوعی می‌تواند برای تمامی ملت‌ها و کل بشریت فواید بسیاری به ارمغان بیاورد، ممکن است موجب پدید آمدن نگرانی‌های اساسی اخلاقی نیز باشد. به عنوان نمونه، هوش مصنوعی می‌تواند باعث ایجاد و یا تشدید تعصباتی گردد که به تبعیض، نابرابری، شکاف‌های دیجیتال و حذف بیانجامد، و گوناگونی فرهنگی، اجتماعی و بیولوژیک را در معرض تهدید قرار دهد، یا شکاف‌های اجتماعی و اقتصادی پدید بیاورد. در این راستا، شفافیت و مفهوم بودن سازوکار الگوریتم‌ها و داده‌های تشکیل‌دهنده هوش مصنوعی و اثرات احتمالی آنها بر کرامت انسانی، حقوق بشر، آزادی‌های بنیادین، برابری جنسیتی، دموکراسی، فرآیندهای اجتماعی، اقتصادی، سیاسی و فرهنگی، شیوه‌های مرسوم علمی و مهندسی، رفاه حیوانات، محیط زیست و اکوسیستم از اهمیت ویژه‌ای برخوردار هستند.

لازم به ذکر است که فن‌آوری هوش مصنوعی می‌تواند به شکاف‌ها و نابرابری‌های موجود در جهان، چه در داخل کشورها و چه در بین ملل مختلف دامن بزند. برقراری و حفظ اعتماد، عدالت و انصاف برای تمامی افراد و ملت‌ها و بدون استثناء از طریق دسترسی عادلانه به فن‌آوری هوش مصنوعی و برخورداری از مزایا، و یا حفاظت در برابر پیامدهای منفی آن، در نظر گرفتن تفاوت‌های کشورهای مختلف و احترام به عدم تمایل برخی افراد به استفاده از این فن‌آوری از جمله موارد حائز اهمیت در این زمینه است.

واضح است که همه کشورهای جهان با افزایش کاربرد فن‌آوری‌های اطلاعات، ارتباطات و هوش مصنوعی، و نیاز روزافزون به ارتقاء سطح سواد رسانه و اطلاعات مواجه هستند. اقتصاد دیجیتال، چالش‌ها و فرصت‌های مهم اجتماعی، اقتصادی و زیست‌محیطی را در ارتباط با تشریح مزایا ایجاد کرده است، به ویژه برای کشورهای با درآمد پایین و متوسط، از جمله کشورهای توسعه‌نیافته، کشورهای در حال توسعه محصور در خشکی، و کشورهای جزیره‌ای کوچک در حال توسعه، که برای رشد اقتصاد دیجیتال پایدار، نیازمند به رسمیت شناختن، حفظ و ارتقاء فرهنگ‌ها، ارزش‌ها و دانش بومی هستند.

فن‌آوری هوش مصنوعی دارای این پتانسیل است که برای محیط زیست و اکوسیستم‌ها سودمند باشد، و برای محقق شدن این پتانسیل، صدمات و اثرات منفی احتمالی بر محیط زیست و اکوسیستم نباید نادیده گرفته شوند، و بلکه باید مورد بررسی قرار بگیرند.

پرداختن به مخاطرات و نگرانی‌های اخلاقی نباید مانعی بر سر راه پیشرفت و نوآوری باشد، بلکه باید فرصت‌های جدید ایجاد کرده و منجر به انجام تحقیقات و ابداعات اخلاق‌مدار گردد، که به نوبه خود باعث تشبیت فن‌آوری هوش مصنوعی در حقوق بشر و آزادی‌های بنیادین، اصول و ارزش‌ها، و اندیشه‌های اخلاقی خواهد شد.

لازم به یادآوری است که در چهلمین نشست کنفرانس عمومی یونسکو مورخ نوامبر ۲۰۱۹، قطعنامه C/Resolution ۳۷۴۰ تصویب شده بود، که به موجب آن دبیر کل یونسکو ملزم به «تدوین ابزار تعیین استاندارد بین‌المللی در خصوص اخلاق هوش مصنوعی در قالب یک توصیه‌نامه» و ارائه آن در چهل و یکمین نشست کنفرانس عمومی در سال ۲۰۲۱ شد.

توسعه فن‌آوری هوش مصنوعی، افزایش متناسب سطح سواد داده، رسانه و اطلاعات، و همچنین دسترسی به منابع مستقل، کثرت‌گرا و قابل اطمینان اطلاعات را ایجاد می‌کند. این موارد بخشی از تلاش برای کاهش خطرات اشاعه اخبار نادرست و یا دروغ، نفرت‌پراکنی و آسیب‌های ناشی از سوءاستفاده از داده‌های شخصی را تشکیل می‌دهند.

طبق مشاهدات، یک چهارچوب هنجاری برای فن‌آوری هوش مصنوعی و نتایج اجتماعی آن، ریشه در ساختارهای حقوقی ملی و بین‌المللی، حقوق بشر و آزادی‌های بنیادین، اخلاقیات، نیاز به دسترسی به داده‌ها، دانش و اطلاعات، آزادی تحقیق و نوآوری، رفاه انسانی و سلامت محیط زیست و اکوسیستم دارد. چنین چهارچوبی قادر است با تکیه بر درک متقابل و اهداف مشترک، اصول و ارزش‌های اخلاقی را با چالش‌ها و فرصت‌های مرتبط با فن‌آوری هوش مصنوعی پیوند دهد.

به نظر می‌رسد که اصول و ارزش‌های اخلاقی می‌توانند از طریق راهنمایی و با در نظر گرفتن سرعت بالای رشد تکنولوژی، به روند توسعه و اجرای سیاست‌های مبتنی بر حقوق و هنجارهای قانونی کمک کنند. در این میان مطمئناً استانداردهای جهانی فن‌آوری هوش مصنوعی با رعایت کامل قوانین بین‌المللی و به ویژه قوانین حقوق بشر، نقش کلیدی را در توسعه هنجارهای مرتبط با هوش مصنوعی در سراسر دنیا ایفا خواهند کرد.

در این اجلاس، اعلامیه جهانی حقوق بشر مصوب سال ۱۹۴۸، ابزارهای چهارچوب حقوق بشر بین‌المللی، از جمله کنوانسیون مربوط به وضعیت پناهندگان (۱۹۵۱)، کنوانسیون مربوط به منع تبعیض در امور استخدام و اشتغال (۱۹۵۸)، کنوانسیون بین‌المللی رفع هر نوع تبعیض نژادی (۱۹۶۵)، میثاق بین‌المللی حقوق مدنی و سیاسی (۱۹۶۶)، میثاق بین‌المللی حقوق اقتصادی، اجتماعی و فرهنگی (۱۹۶۶)، کنوانسیون رفع هرگونه تبعیض علیه زنان (۱۹۷۹)، کنوانسیون حقوق کودک (۱۹۸۹)، کنوانسیون حقوق افراد دارای معلولیت (۲۰۰۶)، کنوانسیون مبارزه با تبعیض در آموزش (۱۹۶۰)، کنوانسیون حفظ و ترویج تنوع مظاهر فرهنگی (۲۰۰۵)، و سایر ابزارها، توصیه‌نامه‌ها و بیانیه‌های بین‌المللی مرتبط مورد توجه قرار گرفتند.

سایر موارد مورد بحث در این اجلاس به قرار زیر می‌باشند:

بیانیه سازمان ملل متحد در مورد حق توسعه (۱۹۸۶)، بیانیه در مورد مسئولیت‌های نسل حاضر در قبال نسل‌های آینده (۱۹۹۷)، اعلامیه جهانی اخلاق زیستی و حقوق بشر (۲۰۰۵)، اعلامیه سازمان ملل متحد

درباره حقوق مردمان بومی (۲۰۰۷)، قطعنامه مجمع عمومی سازمان ملل متحد در مورد بررسی اجلاس جهانی سران در مورد جامعه اطلاعاتی (۲۰۱۵) (A/RES/۷۰/۱۲۵)، قطعنامه مجمع عمومی سازمان ملل متحد در مورد تغییر جهان ما: دستور کار ۲۰۳۰ برای توسعه پایدار (۲۰۱۵) (A/RES/۷۰/۱)، توصیه‌نامه در مورد حفظ و دسترسی به میراث مستند، از جمله اسناد دیجیتال (۲۰۱۵)، اعلامیه اصول اخلاق در رابطه با تغییرات اقلیمی (۲۰۱۷)، توصیه‌نامه در مورد علم و پژوهشگران علمی (۲۰۱۷)، شاخص‌های فراگیری اینترنت (تایید شده توسط برنامه بین‌المللی یونسکو برای توسعه ارتباطات در سال ۲۰۱۸)، از جمله اصول ROAM (مخفف حقوق، آزادی، دسترسی برای همگان، و مشارکت چندجانبه، تایید شده توسط کنفرانس عمومی یونسکو در سال ۲۰۱۵)، قطعنامه شورای حقوق بشر درباره «حق بر حریم خصوصی در عصر دیجیتال» (۲۰۱۹) (A/HRC/RES/۴۲/۱۵)، و قطعنامه شورای حقوق بشر درباره «فن‌آوری‌های دیجیتال نوین و نوظهور و حقوق بشر» (۲۰۱۹) (A/HRC/RES/۴۱/۱۱).

از موارد دیگری که در این اجلاس بر آن تأکید شد می‌توان به ضرورت توجه ویژه به کشورهای دارای درآمد پایین و متوسط، از جمله کشورهای توسعه‌نیافته، کشورهای در حال توسعه محصور در خشکی، و کشورهای جزیره‌ای کوچک در حال توسعه اشاره کرد، زیرا این کشورها ظرفیت‌های خاص خود را دارند، اما در بحث اخلاق هوش مصنوعی به آنها کمتر توجه شده است. این مسئله نگرانی‌هایی را در مورد نادیده گرفتن دانش محلی، کثرت‌گرایی فرهنگی، سیستم‌های ارزشی و ضرورت‌های عدالت جهانی برای مقابله با اثرات مثبت و منفی فن‌آوری هوش مصنوعی ایجاد می‌کند.

در زمینه هوش مصنوعی باید به سیاست‌های ملی بشمار دیگر، سایر چهارچوب‌ها و ابتکارات از جانب نهادهای مرتبط سازمان ملل نیز آگاه بود، همچنین سازمان‌های بین‌دولتی از جمله سازمان‌های منطقه‌ای، خصوصی، حرفه‌ای، غیردولتی، و جامعه علمی، که در ارتباط با اخلاق و نظارت بر فن‌آوری هوش مصنوعی می‌باشند. فن‌آوری هوش مصنوعی می‌تواند مزایای بسیاری داشته باشد، اما برخورداری از این مزایا ممکن است به تشدید تنش‌ها پیرامون نوآوری و دسترسی نابرابر به دانش و فن‌آوری بیانجامد، که شامل کم‌سواد دیجیتال و مدنی نیز هست، و توانایی عموم برای مشارکت در موضوعات مرتبط با هوش مصنوعی را محدود می‌سازد. برخی از موارد مهم دیگر در ارتباط با موانع دسترسی به اطلاعات و شکاف در ظرفیت، توانایی‌های انسانی و سازمانی، موانع دسترسی به ابداعات تکنولوژی، فقدان زیرساخت‌های فیزیکی و دیجیتال کافی و چهارچوب‌های قانونی مرتبط با داده‌ها می‌باشند، که لازم است به همه آنها پرداخته شود.

تقویت همکاری و همبستگی جهانی از طریق مشارکت میان کشورها نه تنها دسترسی عادلانه به فن‌آوری هوش مصنوعی را ممکن می‌سازد، بلکه به رفع چالش‌هایی که بر سر راه گوناگونی و ارتباط میان فرهنگ‌ها و سیستم‌های اخلاقی وجود دارند کمک می‌کند. این همکاری و همبستگی احتمال سوءاستفاده را کاهش داده،

تحقق ظرفیت کامل هوش مصنوعی را به ویژه در زمینه توسعه فراهم می‌کند، و همچنین تضمینی است برای اخلاق مدار بودن استراتژی‌های ملی هوش مصنوعی.

توسعه سریع فن‌آوری هوش مصنوعی، اجرا و حاکمیت اخلاقی این فن‌آوری، و همچنین احترام به گوناگونی فرهنگی و حفاظت از آن را به چالش می‌کشد، و به صورت بالقوه می‌تواند ارزش‌ها و استانداردهای اخلاقی محلی و منطقه‌ای را مختل کند.

این اجلاس (۱) در تاریخ ۲۳ نوامبر ۲۰۲۱، توصیه‌نامه فعلی در خصوص اخلاق هوش مصنوعی را به رسمیت می‌شناسد، (۲) پیشنهاد می‌کند که کشورهای عضو، به شکل داوطلبانه مفاد این توصیه‌نامه را با روش‌های مناسب اعمال کنند، از جمله از طریق هرگونه اقدام قانونی لازم و غیره در راستای قانون اساسی و ساختارهای حاکمیتی هر کشور، تا در چهارچوب صلاحیت‌های خود و همگام با قوانین بین‌المللی مانند قانون بین‌المللی حقوق بشر، اصول و هنجارهای ذکر شده در این توصیه‌نامه را به اجرا دریاورند، و (۳) توصیه می‌کند که کشورهای عضو، همه ذینفعان، از جمله شرکت‌های تجاری را به کار گیرند تا در رابطه با اجرای این توصیه‌نامه، نقش خود را ایفا کنند؛ همچنین کشورهای عضو می‌بایست توجه مقامات، نهادها، مؤسسات پژوهشی و دانشگاهی، ارگان‌ها و مؤسسات بخش دولتی، خصوصی و جامعه مدنی درگیر با هوش مصنوعی را به این توصیه‌نامه جلب کنند، تا توسعه و استفاده از این فن‌آوری به وسیله تحقیقات درست علمی و تحلیل و ارزیابی اخلاقی هدایت شود.

### الف) دامنه کاربرد

یک - این توصیه‌نامه به مسایل اخلاقی مربوط به هوش مصنوعی در حیطه اختیارات یونسکو می‌پردازد. رویکرد آن به هوش مصنوعی به شکل یک بازتاب هنجاری سیستماتیک، و مبتنی بر چهارچوبی همه جانبه، جامع، چندفرهنگی و در حال تکامل از ارزش‌ها، اصول و اقدامات وابسته به یکدیگر می‌باشد که قادر است جوامع را در برخورد مسئولانه با اثرات شناخته شده و ناشناخته هوش مصنوعی بر انسان، جامعه، محیط زیست و اکوسیستم هدایت کند، و مبنایی برای قبول یا رد هوش مصنوعی به آنها ارائه نماید. توصیه‌نامه حاضر، اخلاق را مبنایی دینامیک برای ارزیابی هنجاری و هدایت فن‌آوری هوش مصنوعی می‌داند که ریشه در کرامت و رفاه انسان و پیشگیری از آسیب‌ها دارد، و در عین حال مانند یک قطب‌نمای مبتنی بر اخلاق علمی و فن‌آوری نیز عمل می‌کند.

دو - این توصیه‌نامه به دنبال ارائه یک معنای واحد برای هوش مصنوعی نیست، زیرا معنای این عبارت باید با گذشت زمان و به اقتضای پیشرفت تکنولوژی، تغییر کند. هدف این توصیه‌نامه، پرداختن به آن قسمت از ویژگی‌های سیستم‌های هوش مصنوعی است که عمدتاً به اخلاق مربوط می‌شوند. بنابراین، رویکرد این توصیه‌نامه به هوش مصنوعی مانند سیستم‌هایی با قابلیت پردازش داده‌ها و اطلاعات به شکلی شبیه به

رفتارهای هوشمندانه می‌باشد که معمولاً دارای ابعادی از توانایی استدلال، یادگیری، مشاهده، پیش‌بینی، برنامه‌ریزی یا کنترل هستند. در این نوع رویکرد، سه عنصر اصلی دارای نقش بسیار مهمی هستند:

۱) سیستم‌های هوش مصنوعی، فن‌آوری‌های پردازش اطلاعاتی هستند که الگوها و الگوریتم‌ها را با یکدیگر ادغام میکنند که قابلیت یادگیری و انجام عملیات شناختی را ایجاد می‌کنند، که نتیجه آن می‌تواند پیش‌بینی یا تصمیم‌گیری در محیط‌های واقعی و مجازی باشد. این سیستم‌ها به نحوی طراحی می‌شوند که با درجات مختلف استقلال و از طریق مدل‌سازی و بازنمایی دانش، استفاده از داده‌ها و محاسبه روابط، عمل کنند. سیستم‌های هوش مصنوعی روش‌های مختلفی را به کار می‌گیرند، از جمله: ۱- یادگیری ماشینی، شامل یادگیری عمیق و یادگیری تقویتی، و ۲- استدلال ماشینی، شامل برنامه‌ریزی، زمان‌بندی، بازنمایی دانش و استدلال، جستجو و بهینه‌سازی.

سیستم‌های هوش مصنوعی ممکن است در نظام‌های فیزیکی سایبری از قبیل اینترنت اشیا، سیستم‌های رباتیک، رباتیک اجتماعی، و اینترفیس‌های انسان و کامپیوتر استفاده شوند، که شامل کنترل، مشاهده، پردازش داده‌های جمع‌آوری شده توسط حسگرها، و اداره فعال‌کننده‌ها در محیط کار هوش مصنوعی می‌باشد.

۲) سؤالات اخلاقی در مورد سیستم‌های هوش مصنوعی به تمامی مراحل چرخه زندگی این سیستم‌ها مربوط می‌شوند که می‌تواند شامل تحقیق، طراحی و توسعه تا استقرار و استفاده، نگهداری، بهره‌برداری، تجارت، تامین مالی، نظارت و ارزیابی، اعتبارسنجی، پایان استفاده، جداسازی (اجزاء) و ختم عملیات باشد. به علاوه، عنوان «عاملین هوش مصنوعی» می‌تواند به هر یک از افراد حقیقی یا حقوقی که دست‌کم در یک مرحله از چرخه زندگی هوش مصنوعی شرکت داشته اند اطلاق شود، از جمله پژوهشگران، برنامه‌نویسان، مهندسان، دانشمندان داده، کاربران نهایی، کسب و کارها، دانشگاه‌ها، نهادهای دولتی و خصوصی و غیره.

۳) سیستم‌های هوش مصنوعی انواع جدیدی از مسائل اخلاقی را به وجود می‌آورند، از جمله تحت تاثیر قرار دادن تصمیم‌گیری، اشتغال و کار، تعاملات اجتماعی، خدمات درمانی، آموزش، رسانه، دسترسی به اطلاعات، شکاف دیجیتالی، حفاظت از اطلاعات شخصی و مصرف‌کننده، محیط زیست، دموکراسی، حاکمیت قانون، امنیت و نیروی پلیس، کاربرد دوگانه تکنولوژی، و حقوق بشر و آزادی‌های بنیادین، از قبیل آزادی بیان، حریم خصوصی و منع تبعیض. علاوه بر این، پتانسیل الگوریتم‌های هوش مصنوعی برای بازتولید و تقویت سوگیری‌های موجود باعث ایجاد چالش‌های اخلاقی جدید می‌شود، که این امر می‌تواند به اشکال موجود تبعیض، تعصب و تفکر قالبی دامن بزند. بعضی از این مشکلات به قابلیت‌های سیستم‌های هوش مصنوعی برای انجام کارهایی مربوط می‌شوند که در گذشته مختص موجودات زنده و بعضاً نوع بشر بودند. این ویژگی‌ها به سیستم‌های هوش مصنوعی نقشی جدید و مهم در مرادوات انسانی، جامعه، و رابطه هوش مصنوعی با محیط زیست و اکوسیستم اعطا می‌کنند. این موضوع زمینه‌ای برای کودکان و جوانان فراهم می‌آورد تا بتوانند در آن

رشد کنند، به درک مناسبی از جهان و خودشان برسند، رسانه‌ها و اطلاعات را با دید انتقادی بررسی و فهم کنند، و مهارت تصمیم‌گیری کسب کنند. در طولانی‌مدت، سیستم‌های هوش مصنوعی ممکن است حس خاص تجربه و عاملیت انسان را به چالش بکشند، که خود موجب نگرانی‌هایی درباره مسائلی از قبیل درک انسان از خود، تعاملات اجتماعی، فرهنگی و زیست‌محیطی، استقلال و خودمختاری، عاملیت، ارزش و شأن و منزلت انسان می‌باشد.

سه - این توصیه‌نامه به پیامدهای اخلاقی گسترده‌تر سیستم‌های هوش مصنوعی در رابطه با حوزه‌های اصلی فعالیت یونسکو توجه ویژه دارد، به عنوان نمونه، آموزش، دانش، فرهنگ، ارتباطات و اطلاعات. موارد زیر در مطالعه مقدماتی درباره اخلاق هوش مصنوعی در سال ۲۰۱۹ توسط کمیسیون جهانی اخلاق علم و تکنولوژی یونسکو (COMEST)<sup>۱</sup> بررسی شده بودند:

(۱) آموزش - زیرا زندگی در جوامعی که در مسیر دیجیتال شدن هستند، نیازمند شیوه‌های آموزشی، تفکر اخلاقی، اندیشه انتقادی، شیوه‌های طراحی مسئولانه و مهارت‌های نوین می‌باشد. از جمله دلایل این امر می‌توان به پیامدهای بازار کار، اشتغال‌پذیری و مشارکت مدنی اشاره کرد.

(۲) دانش - در اینجا اگر ساده‌ترین مفهوم دانش را که شامل رشته‌های دانشگاهی مانند علوم طبیعی و پزشکی تا جامعه‌شناسی و علوم انسانی می‌شود در نظر بگیریم، می‌توان گفت که سیستم‌های هوش مصنوعی از یک سو ظرفیت‌ها و شیوه‌های تحقیقاتی نوینی در اختیار ما می‌گذارند، و از سوی دیگر پیامدهایی برای تعبیر ما از درک و توضیح علمی به همراه داشته، مبنای جدیدی برای تصمیم‌گیری‌ها خواهند بود.

(۳) هویت و گوناگونی فرهنگی - سیستم‌های هوش مصنوعی می‌توانند باعث غنای صنایع فرهنگی و خلاقه بشوند، اما ممکن است به افزایش تمرکز محتوای فرهنگی، داده، بازار و درآمد در دست افراد معدودی نیز بیانجامند. این امر می‌تواند به صورت بالقوه بر روی گوناگونی و کثرت‌گرایی زبان، رسانه، مظاهر فرهنگی، مشارکت و برابری تأثیر منفی بگذارد.

(۴) ارتباطات و اطلاعات - سیستم‌های هوش مصنوعی به شکل فزاینده‌ای نقش پراهمیتی در پردازش، ساختار و تهیه اطلاعات بازی می‌کنند. مسائل مربوط به روزنامه‌نگاری الگوریتمی و کلا استفاده از الگوریتم در تهیه اخبار و تعدیل و نظارت بر محتوای رسانه‌های اجتماعی و موتورهای جستجو تنها چند نمونه از مسائلی هستند که در ارتباط با اطلاعات (درست، نادرست و دروغ)، نفرت‌پراکنی، پیدایش روایات اجتماعی جدید، تبعیض، آزادی بیان، حریم خصوصی، سواد رسانه و اطلاعات مطرح می‌شوند.

چهار - مخاطبان این توصیه‌نامه، کشورهای عضو هستند، هم در مقام عاملین هوش مصنوعی، و هم به عنوان مقاماتی که عهده‌دار تعیین چهارچوب‌های قانونی و نظارتی حاکم بر کل چرخه زندگی سیستم‌های هوش

1. Commission on Ethics of Scientific Knowledge and Technology

مصنوعی، و همچنین ارتقاء مسئولیت تجاری آن می‌باشند. ضمناً این توصیه‌نامه به ارائه راهنمایی اخلاقی به کلیه عاملین هوش مصنوعی در بخش‌های خصوصی و دولتی، از طریق ارائه مبنایی برای ارزیابی اخلاقی اثرات سیستم‌های هوش مصنوعی در طول چرخه زندگی آنها می‌پردازد.

## ب) اهداف و مقاصد

پنج - یکی از اهداف این توصیه‌نامه، ایجاد مبنایی برای به کار گرفتن سیستم‌های هوش مصنوعی در جهت منافع بشریت، افراد، جوامع، محیط زیست و اکوسیستم، و پیشگیری از آسیب‌هاست. هدف دیگر، ترغیب عاملین به استفاده صلح‌آمیز این سیستم‌ها می‌باشد.

شش - علاوه بر چهارچوب‌های اخلاقی موجود برای سیستم‌های هوش مصنوعی در سراسر جهان، این توصیه‌نامه قصد دارد ابزاری هنجاری و مورد قبول ارائه دهد که تمرکز آن نه تنها بر ابراز اصول و ارزش‌ها، بلکه بر تحقق عملی آنها از طریق توصیه‌های سیاستی مستحکم استوار باشد. چنین ابزاری می‌بایست بر مسایل مربوط به شمول و برابری جنسیتی و حفاظت از محیط زیست و اکوسیستم، تأکید داشته باشد.

هفت - از آنجایی که پیچیدگی مسائل اخلاقی پیرامون هوش مصنوعی، همکاری تعداد زیادی از افراد و سازمان‌های ذینفع در سطوح و بخش‌های مختلف جوامع بین‌المللی، منطقه‌ای و ملی را ایجاب می‌کند، این توصیه‌نامه قصد دارد به ذینفعان این امکان را بدهد که بر پایه یک گفتگوی جهانی و بین‌فرهنگی، در قبال فن‌آوری هوش مصنوعی مسئولیت مشترکی بر عهده بگیرند.

هشت - اهداف این توصیه‌نامه به شرح زیر می‌باشند:

۱) ارائه یک چهارچوب جهانی متشکل از ارزش‌ها، اصول و اقدامات با هدف هدایت کشورها در امر تدوین قوانین، سیاست‌ها و سایر ابزار مرتبط با هوش مصنوعی و منطبق بر قوانین بین‌المللی

۲) هدایت اقدامات افراد، گروه‌ها، جوامع، مؤسسات و شرکت‌های بخش خصوصی به منظور اطمینان از القاء اخلاقیات در تمامی مراحل چرخه زندگی سیستم‌های هوش مصنوعی

۳) حفاظت، ترویج و احترام به حقوق بشر و آزادی‌های بنیادین، کرامت انسانی و برابری، از جمله برابری جنسیتی؛ محافظت از منافع نسل حاضر و نسل‌های آینده؛ حفاظت از محیط زیست، تنوع زیستی و اکوسیستم؛ و احترام به گوناگونی فرهنگی در تمامی مراحل چرخه زندگی سیستم‌های هوش مصنوعی

۴) حمایت از گفتگوهای چندجانبه، چندزمینه‌ای و کثرت‌گرایانه، و ایجاد اجماع درباره مسائل اخلاقی مربوط به سیستم‌های هوش مصنوعی.

۵) افزایش دسترسی عادلانه به پیشرفت‌ها و دانش در زمینه هوش مصنوعی و اشتراک منافع، همراه با توجه ویژه به نیازها و کمک‌های کشورهای دارای درآمد پایین و متوسط، از جمله کشورهای توسعه‌نیافته، کشورهای در حال توسعه محصور در خشکی، و کشورهای جزیره‌ای کوچک در حال توسعه.



### ج) اصول و ارزش‌ها

نه - همه عاملین چرخه زندگی سیستم‌های هوش مصنوعی موظفند در وهله اول به اصول و ارزش‌هایی که در ادامه این متن خواهد آمد احترام بگذارند، و در صورت نیاز و صلاحدید، از طریق ویرایش قوانین موجود و تدوین قوانین، مقررات و دستورالعمل‌های تجاری جدید، به ترویج آنها بپردازند. این موضوع باید منطبق بر قوانین بین‌المللی از جمله منشور ملل متحد و تعهدات حقوق بشری کشورهای عضو صورت بگیرد، و همچنین در راستای اهداف پایداری اجتماعی، سیاسی، زیست‌محیطی، آموزشی، علمی و اقتصادی مورد توافق بین‌المللی همچون اهداف توسعه پایدار سازمان ملل متحد (SDG<sup>2</sup>) باشد.

ده - ارزش‌ها به عنوان آرمان‌های انگیزه بخش، نقش مهمی در شکل‌دادن به سیاست‌ها و هنجارهای قانونی دارند. مجموعه ارزش‌های ذکر شده در این توصیه‌نامه، رفتارهای مطلوب را القا می‌کنند و نشانگر مبانی اصول می‌باشند، و اصول نیز ارزش‌هایی که بر مبنای آن استوار هستند را به شکلی ملموس نمایان می‌سازند. به این ترتیب، فرآیند عملیاتی شدن ارزش‌ها می‌تواند با سهولت بیشتری در بیانیه‌ها و امور مربوط به سیاست‌گذاری انجام گیرد.

یازده - تمامی اصول و ارزش‌های ذکر شده در این توصیه‌نامه به خودی خود مطلوب می‌باشند، اما در عمل ممکن است بین این ارزش‌ها و اصول تنش‌هایی به وجود بیاید. در هر موقعیت قابل فرض لازم است ارزیابی زمینه‌ای برای کنترل تنش‌های احتمالی به عمل بیاید، ضمن این که در نظر گرفتن اصل تناسب و رعایت حقوق بشر و آزادی‌های بنیادین ضروری می‌باشد. در کلیه موارد، اعمال هرگونه محدودیتی بر حقوق بشر و آزادی‌های بنیادین باید مبنای قانونی داشته و منطقی، ضروری و متناسب بوده و در راستای تعهدات کشورها تحت قوانین بین‌المللی باشد. هدایت عاقلانه چنین سناریوهایی قاعداً نیازمند تعامل با طیف گسترده‌ای از ذینفعان و گفتگوی اجتماعی، و همچنین تأملات اخلاقی، دقت کافی و ارزیابی پیامدها می‌باشد.

دوازده - برای اطمینان از این که فن‌آوری هوش مصنوعی در جهت منافع بشریت، افراد، جوامع، محیط زیست و اکوسیستم کار می‌کند و نمایانگر اصول و ارزش‌های ذکر شده در این توصیه‌نامه می‌باشد، لازم است چرخه زندگی سیستم‌های هوش مصنوعی قابل اعتماد و درست باشد. مردم باید اطمینان داشته باشند که سیستم‌های هوش مصنوعی برای افراد و کل جامعه سودمند هستند و در عین حال اقدامات کافی برای کاهش مخاطرات صورت می‌گیرند. یکی از شروط لازم برای قابل اعتماد بودن سیستم‌های هوش مصنوعی این است که در طول چرخه زندگی خود و به اقتضای نیاز، تحت نظارت دقیق ذینفعان قرار داشته باشند. از آنجایی که قابل اعتماد بودن هوش مصنوعی از نتایج عملیاتی شدن اصول ذکر شده در این توصیه‌نامه می‌باشد، کلیه سیاست‌های

<sup>2</sup> Sustainable Development Goals

پیشنهادی در اینجا در راستای افزایش این ویژگی در تمامی مراحل چرخه زندگی سیستم‌های هوش مصنوعی قرار گرفته اند.

### ج - (۱) ارزش‌ها

#### - احترام، حفاظت و ترویج حقوق بشر و آزادی‌های بنیادین و کرامت انسانی:

سیزده - کرامت هر انسان یک امر ذاتی و غیر قابل تعرض است که اساس نظام جهانی تفکیک ناپذیر، غیرقابل انکار، و دارای همبستگی و ارتباط درونی حقوق بشر و آزادی‌های بنیادین را تشکیل می‌دهد. از این رو احترام، حفاظت و ترویج کرامت و حقوق انسان که توسط قوانین بین‌المللی از جمله قانون بین‌المللی حقوق بشر تبیین شده، در طول چرخه زندگی سیستم‌های هوش مصنوعی ضروری است. کرامت انسانی با ارزش ذاتی و برابر افراد مرتبط است، فارغ از نژاد، رنگ پوست، اصل و نسب، جنسیت، سن، زبان، مذهب، عقاید سیاسی، خاستگاه ملی، قومی و اجتماعی، شرایط اقتصادی یا اجتماعی خانوادگی، معلولیت و غیره.

چهارده - هیچ فرد یا جامعه بشری نباید از نظر فیزیکی، اقتصادی، اجتماعی، سیاسی، فرهنگی یا روانی در هیچ یک از مراحل چرخه زندگی سیستم‌های هوش مصنوعی آسیب ببیند یا تحت سلطه قرار بگیرد. کیفیت زندگی افراد باید در طول این چرخه زندگی بهبود یابد. تعریف «کیفیت زندگی» باید برای افراد و گروه‌ها باز گذاشته شود، مشروط بر این که حقوق بشر و آزادی‌های بنیادین و کرامت انسانی به واسطه این تعریف، مورد نقض و یا سوءاستفاده قرار نگیرند.

پانزده - افراد می‌توانند با سیستم‌های هوش مصنوعی در طول چرخه زندگی آنها به تعامل بپردازند و از آنها کمک بگیرند. نمونه‌هایی از این کمک‌ها می‌توانند شامل مراقبت از افراد آسیب‌پذیر یا کسانی که در شرایط حساس قرار دارند از جمله کودکان، سالمندان، افراد دارای معلولیت یا بیماران باشند. در چنین تعاملاتی افراد نباید به شکل اشیاء دیده شوند، یا به شأن و منزلتشان خدشه وارد شود، و یا حقوق انسانی و آزادی‌های بنیادین آنها نقض یا مورد سوءاستفاده قرار گیرد.

شانزده - در طول چرخه زندگی سیستم‌های هوش مصنوعی، احترام، حفاظت و ترویج حقوق بشر و آزادی‌های بنیادین ضروری است. دولت‌ها، بخش خصوصی، جامعه مدنی، سازمان‌های بین‌المللی و انجمن‌های فنی و دانشگاهی موظفند ابزارها و چهارچوب‌های حقوق بشر را در مداخلاتی که پیرامون چرخه زندگی سیستم‌های هوش مصنوعی انجام می‌دهند، در نظر بگیرند. فن‌آوری‌های نوین می‌بایست ابزار جدیدی برای حمایت، دفاع و اعمال حقوق بشر فراهم کنند و از نقض آن بپرهیزند.

### - شکوفایی محیط زیست و اکوسیستم:

هفده - در طول چرخه زندگی سیستم‌های هوش مصنوعی، به رسمیت شناختن، حفاظت و ارتقاء محیط زیست و اکوسیستم ضروری است. مضاف بر این که در امر بهره‌مندی انسان و سایر موجودات زنده از منافع پیشرفت در زمینه هوش مصنوعی، محیط زیست و اکوسیستم از اهمیت حیاتی برخوردار هستند.

هجده - تمامی عاملین در چرخه زندگی سیستم‌های هوش مصنوعی موظفند از قوانین بین‌المللی و مقررات، استانداردها و رویه‌های معمول داخلی پیروی کنند. از این میان می‌توان به اقدامات احتیاطی در جهت حفاظت و احیای محیط زیست و اکوسیستم، و توسعه پایدار اشاره کرد. عاملین همچنین باید تلاش کنند تا اثرات زیست‌محیطی سیستم‌های هوش مصنوعی (مانند ردپای کربنی) را کاهش بدهند. این کار احتمال تغییرات آب و هوایی و ریسک فاکتورهای زیست‌محیطی را به حداقل می‌رساند و از بهره‌برداری ناپایدار، استفاده و دگرگونی منابع طبیعی و نتیجتاً تخریب محیط زیست و اکوسیستم جلوگیری می‌کند.

### - حصول اطمینان از گوناگونی و فراگیری:

نوزده - در تمامی مراحل چرخه زندگی سیستم‌های هوش مصنوعی لازم است گوناگونی و فراگیری طبق قوانین بین‌المللی از جمله قانون حقوق بشر مورد احترام، حفاظت و ارتقاء قرار بگیرد. این امر با ترویج مشارکت فعال همه افراد و گروه‌ها، فارغ از نژاد، رنگ پوست، اصل و نسب، جنسیت، سن، زبان، مذهب، عقاید سیاسی، خاستگاه ملی، قومی و اجتماعی، شرایط اقتصادی یا اجتماعی خانوادگی، معلولیت و غیره تحقق می‌یابد.

بیست - دامنه انتخاب‌های مربوط به سبک زندگی، باورها، نظرات، نحوه بیان یا تجربیات شخصی، از جمله استفاده اختیاری از سیستم‌های هوش مصنوعی و مشارکت در طراحی این ساختارها، نباید در هیچ یک از مراحل چرخه زندگی سیستم‌های هوش مصنوعی دچار محدودیت شوند.

بیست و یک - تلاش‌هایی از قبیل همکاری‌های بین‌المللی برای غلبه بر فقدان زیرساخت‌های تکنولوژیک، آموزشی و مهارتی لازم، چهارچوب‌های قانونی و همچنین منع سوءاستفاده از این کمبودها ضروری است؛ به ویژه در کشورهای دارای درآمد پایین و متوسط، از جمله کشورهای توسعه‌نیافته، کشورهای در حال توسعه محصور در خشکی، و کشورهای جزیره‌ای کوچک در حال توسعه.

### - زندگی در جوامع صلح‌طلب، عدالت‌محور و به هم پیوسته:

بیست و دو - تمامی عاملین هوش مصنوعی باید نقشی مبتنی بر مشارکت و توانمندسازی در پدید آوردن جوامع صلح‌طلب و عدالت‌محور ایفا کنند. اساس این مشارکت بر آینده‌ای به هم پیوسته در راستای منافع همگان و همگام با حقوق بشر و آزادی‌های بنیادین می‌باشد. ارزش زیستن در چنین جوامعی به پتانسیل سیستم‌های هوش مصنوعی برای کمک به همبستگی تمامی موجودات زنده با یکدیگر و محیط زیست طبیعی اشاره دارد.

بیست و سه - ایده همبسته بودن انسان‌ها با یکدیگر از علم به این موضوع سرچشمه می‌گیرد که همه انسان‌ها جزیی از یک کل هستند که رشد آن در گرو رشد تک تک اجزای آن است. زندگی در جوامع صلح‌طلب، عدالت‌محور و به هم پیوسته نیازمند نوعی پیوند همبستگی ارگانیک، بلاواسطه و محاسبه‌نشده است که مدام در تلاش برای رسیدن به روابط صلح‌آمیز و مراقبت از دیگر افراد و محیط زیست طبیعی در مفهوم کلی آن می‌باشد.

بیست و چهار - بر اساس این ارزش، صلح، فراگیری، عدالت، برابری و به هم پیوستگی باید در طول چرخه زندگی سیستم‌های هوش مصنوعی ارتقاء بیابند. فرآیندهای این چرخه نباید رویکردی تبعیض‌آمیز داشته و یا آزادی و استقلال رأی افراد را مورد استفاده ابزاری و یا تحت‌الشعاع قرار دهند، امنیت افراد و جوامع را به مخاطره بیندازند و باعث جدایی و دشمنی آنها بشوند، و یا همزیستی میان انسان‌ها و موجودات زنده و محیط زیست طبیعی را تهدید کنند.

## ج - (۲) اصول

### - اصل تناسب و پرهیز از آسیب:

بیست و پنج - توجه به این نکته ضرورت دارد که فن‌آوری هوش مصنوعی لزوماً و به خودی خود متضمن شکوفایی بشر، محیط زیست و اکوسیستم نمی‌باشد. به علاوه، هیچ یک از فرآیندهای مربوط به چرخه زندگی سیستم‌های هوش مصنوعی نباید از حدود آنچه برای دستیابی به اهداف و مقاصد مشروع لازم است فراتر رود، و باید با موقعیت متناسب باشد. در صورت وقوع احتمالی هرگونه آسیبی به افراد، حقوق بشر و آزادی‌های بنیادین، اجتماعات و کل جامعه یا محیط زیست و اکوسیستم، باید روند ارزیابی ریسک و انجام اقدامات لازم جهت جلوگیری از چنین آسیبی اجرا شود.

بیست و شش - انتخاب برای استفاده از سیستم‌های هوش مصنوعی و این که چه روشی به کار گرفته شود، باید از سه طریق توجیه شود: (۱) روش انتخابی هوش مصنوعی باید مناسب و متناسب برای دستیابی به یک هدف مشروع باشد؛ (۲) روش انتخابی هوش مصنوعی نباید در تناقض با ارزش‌های بنیادین مندرج در این سند باشد، و به ویژه استفاده از آن نباید حقوق بشر را نقض کند؛ و (۳) روش انتخابی هوش مصنوعی باید متناسب با موقعیت و بر پایه‌های علمی دقیق استوار باشد. در سناریوهایی که عواقب تصمیمات، به سختی قابل برگشت بوده و یا غیرقابل برگشت هستند و ممکن است تبعات جانی داشته باشند، تصمیم‌گیری نهایی باید بر عهده عامل انسانی باشد؛ ضمناً هوش مصنوعی به هیچ عنوان نباید برای سیستم‌های امتیازدهی اجتماعی یا نظارت انبوه مورد استفاده قرار بگیرد.

**- ایمنی و امنیت:**

بیست و هفت - برای تضمین ایمنی و امنیت بشر، محیط زیست و اکوسیستم لازم است از آسیب های ناخواسته (خطرات ایمنی) و آسیب پذیری در مقابل حملات (خطرات امنیتی) اجتناب شود و در طول چرخه زندگی هوش مصنوعی، این آسیب ها شناسایی، متوقف و حذف شوند. ایمنی و امنیت هوش مصنوعی با توسعه چهارچوب های دسترسی به داده های پایدار و حافظ حریم خصوصی امکان پذیر است. در قالب چنین چهارچوب هایی، آموزش و اعتبارسنجی مدل های هوش مصنوعی با استفاده از داده های باکیفیت امکان پذیر خواهد بود.

**- انصاف و منع تبعیض:**

بیست و هشت - عاملین هوش مصنوعی موظف به ترویج عدالت اجتماعی، حفاظت از انصاف و جلوگیری از هرگونه تبعیض در راستای قوانین بین المللی می باشند. این به معنی رویکردی فراگیر برای اطمینان از دسترسی به خدمات هوش مصنوعی برای همگان است. در این میان باید نیازهای ویژه گروه های سنی و نظام های فرهنگی و زبانی مختلف، افراد دارای معلولیت، دختران و زنان، افراد محروم و گروه هایی که به آنها کمتر توجه می شود، اقشار آسیب پذیر و افرادی که در شرایط حساس قرار دارند، در نظر گرفته شود. کشورهای عضو باید برای ارتقاء دسترسی همگان، از جمله گروه های محلی، به سیستم های هوش مصنوعی حاوی خدمات و محتوای بومی تلاش کنند، و به چندزبانی و گوناگونی فرهنگی احترام بگذارند. کشورهای عضو همچنین ملزم به مقابله با شکاف های دیجیتال و فراهم آوردن دسترسی همگانی و مشارکت در توسعه هوش مصنوعی می باشند. در سطح ملی، کشورهای عضو باید به ترویج برابری دسترسی و مشارکت در چرخه زندگی سیستم های هوش مصنوعی بین مناطق روستایی و شهری و در میان مردم بپردازند، فارغ از نژاد، رنگ پوست، اصل و نسب، جنسیت، سن، زبان، مذهب، عقاید سیاسی، خاستگاه ملی، قومی و اجتماعی، شرایط اقتصادی یا اجتماعی خانوادگی، معلولیت و غیره. در سطح بین المللی، کشورهای پیشرفته مسئولند که در زمینه بهره مندی کشورهای توسعه نیافته از منافع فن آوری هوش مصنوعی با آنان همبستگی نشان دهند، زیرا دسترسی و مشارکت آنها در چرخه زندگی هوش مصنوعی به نظم جهانی عادلانه تری در زمینه اطلاعات، ارتباطات، فرهنگ، آموزش، تحقیقات و ثبات اجتماعی، اقتصادی و سیاسی کمک می کند.

بیست و نه - در طول چرخه زندگی سیستم های هوش مصنوعی و به منظور برقراری عدالت، عاملین هوش مصنوعی باید در حد معقول تلاش کنند تا برنامه ها و نتایج تبعیض آمیز یا مغرضانه را به حداقل رسانده، از تقویت و تداوم آنها جلوگیری کنند. در این راستا باید راه حل مؤثر در برابر تبعیض و تصمیمات الگوریتمی تعصب آمیز وجود داشته باشد.

سی - شکاف دیجیتال و دانش در داخل و میان کشورها باید در طول چرخه زندگی هوش مصنوعی از دو منظر مورد توجه قرار گیرد: از نظر دسترسی و کیفیت دسترسی به فن‌آوری و داده‌ها و منطبق بر چهارچوب‌های ملی، منطقه‌ای و بین‌المللی مرتبط، و همچنین از نظر ارتباط، دانش، مهارت‌ها و مشارکت معنادار جوامع درگیر، به نحوی که با همه افراد عادلانه برخورد شود.

#### - پایداری:

سی و یک - توسعه جوامع پایدار متکی بر دستیابی به مجموعه پیچیده‌ای از اهداف در زنجیره‌ای از ابعاد انسانی، اجتماعی، فرهنگی، اقتصادی و زیست‌محیطی می‌باشد. ظهور فن‌آوری هوش مصنوعی می‌تواند برای اهداف مربوط به پایداری جلوبرنده یا بازدارنده باشد، و این موضوع به چگونگی استفاده از این فن‌آوری در کشورهای با سطوح مختلف توسعه بستگی دارد. به همین علت ارزیابی اثرات انسانی، اجتماعی، فرهنگی، اقتصادی و زیست‌محیطی فن‌آوری هوش مصنوعی باید دائماً و با آگاهی کامل از تبعات این فن‌آوری برای پایداری به عنوان مجموعه‌ای از اهداف در حال رشد در ابعاد مختلف در جریان باشد، و این اهداف در حال حاضر در اهداف توسعه پایدار سازمان ملل متحد (SDG) مطرح شده‌اند.

#### - حریم خصوصی و حفاظت از داده‌ها:

سی و دو - در طول چرخه زندگی سیستم‌های هوش مصنوعی، حریم خصوصی باید به عنوان یک حق اساسی برای حفظ کرامت، خودمختاری و عاملیت انسان مورد احترام قرار بگیرد. جمع‌آوری، استفاده، اشتراک‌گذاری، بیگانی و حذف داده‌های سیستم‌های هوش مصنوعی باید منطبق بر قوانین بین‌المللی و ارزش‌ها و اصول مندرج در این سند، و در چهارچوب قوانین مرتبط ملی، منطقه‌ای و بین‌المللی صورت بگیرد.

سی و سه - چهارچوب‌های مناسب برای حفاظت و مکانیسم‌های حاکمیت داده‌ها باید در سطح ملی یا بین‌المللی و با رویکردی مبنی بر چندین ذینفع ایجاد شود. این چهارچوب‌ها باید توسط نظام قضایی محافظت و در طول چرخه زندگی سیستم‌های هوش مصنوعی تضمین شوند. چهارچوب‌های حفاظت از داده‌ها و سایر مکانیسم‌های مرتبط باید مبتنی بر اصول و استانداردهای بین‌المللی حفاظت داده در خصوص جمع‌آوری، استفاده و افشای اطلاعات شخصی و حقوق صاحبان اطلاعات باشند. در عین حال برای پردازش اطلاعات شخصی، تضمین مشروعیت هدف و مبنای قانونی معتبر و همچنین رضایت آگاهانه ضروری است.

سی و چهار - در استفاده از سیستم‌های الگوریتمی، ارزیابی مناسب از تأثیر بر حریم خصوصی ضروری است. این ارزیابی شامل ملاحظات اجتماعی و اخلاقی در خصوص استفاده از سیستم‌های الگوریتمی، و استفاده خلاقانه از رویکرد حفظ حریم خصوصی در طراحی می‌شود. عاملین هوش مصنوعی باید نشان بدهند که درقبال طراحی و اجرای سیستم‌های هوش مصنوعی پاسخگو هستند، و قادر به محافظت از اطلاعات شخصی افراد در طول چرخه زندگی سیستم‌های هوش مصنوعی می‌باشند.

### - نظارت و تصمیم‌گیری انسانی:

سی و پنج - کشورهای عضو باید اطمینان حاصل کنند که همیشه می‌توان برای هر مرحله‌ای از چرخه زندگی سیستم‌های هوش مصنوعی و در موارد جبران خسارت، مسئولیت اخلاقی و قانونی را به اشخاص واقعی یا نهادهای حقوقی نسبت داد. به این ترتیب نظارت انسانی نه فقط به معنای نظارت یک فرد، بلکه در صورت لزوم به معنای نظارت عموم می‌باشد.

سی و شش - در برخی موارد ممکن است انسان‌ها به منظور بالا بردن راندمان از سیستم‌های هوش مصنوعی استفاده کنند، اما تصمیم‌گیری در مورد واگذاری کنترل در موارد محدود همچنان در اختیار انسان است، زیرا انسان می‌تواند برای تصمیم‌گیری و انجام کارها از هوش مصنوعی بهره بگیرد، اما این فن‌آوری هرگز نمی‌تواند جایگزین مسئولیت انسانی بشود. در مجموع، تصمیم‌گیری‌های مربوط به مرگ و زندگی نباید به سیستم‌های هوش مصنوعی محول شوند.

### - شفافیت و توضیح پذیری:

سی و هفت - از جمله شروط ضروری برای احترام، حفاظت و ارتقاء حقوق بشر، آزادی‌های بنیادین و اصول اخلاقی می‌توان به شفافیت و توضیح پذیری سیستم‌های هوش مصنوعی اشاره کرد. شفافیت برای کارکرد مؤثر سیستم‌های مسئولیت ملی و بین‌المللی ضروری است، و فقدان آن می‌تواند احتمال به چالش کشیدن مؤثر تصمیمات به دست آمده از نتایج حاصل از سیستم‌های هوش مصنوعی را تضعیف کند. به این ترتیب، عدم شفافیت ممکن است حق داشتن یک محاکمه عادلانه یا راه حلی مؤثر را نقض کند، و دامنه استفاده قانونی از چنین سیستم‌هایی را محدود کند.

سی و هشت - به منظور پشتیبانی از حاکمیت دموکراسی و از جمله در مواردی که تأثیرات سیستم‌های هوش مصنوعی فرمانطقه‌ای هستند، لازم است در جهت افزایش شفافیت و توضیح پذیری آنها در طول چرخه زندگی تلاش شود. اما در مواردی ممکن است لازم باشد بین شفافیت و توضیح پذیری، و اصول دیگری مانند حفاظت از حریم خصوصی، ایمنی و امنیت، تعادل برقرار شود. از این رو، سطح شفافیت و توضیح پذیری همواره باید متناسب با موقعیت و اثرات این فن‌آوری باشد. در مواردی که تصمیم‌گیری بر اساس الگوریتم‌های هوش مصنوعی صورت می‌گیرد و ایمنی یا حقوق بشر تحت تأثیر قرار می‌گیرد ضروری است که به افراد اطلاع‌رسانی دقیق انجام شود. در چنین مواردی افراد باید بتوانند از عامل مسئول هوش مصنوعی یا نهادهای دولتی اطلاعات و توضیحات درخواست کنند. همچنین افراد حق دارند به دلایل تصمیماتی که حقوق و آزادی‌هایشان را تحت تأثیر قرار می‌دهند، دسترسی داشته باشند و به یکی از کارکنان مؤسسه دولتی یا خصوصی که اختیار بررسی و تغییر تصمیمات را دارد، مطالبی را ارائه دهند. اگر در فرآیند تولید محصولات یا ارائه خدمات به طور مستقیم

یا غیرمستقیم از هوش مصنوعی استفاده شده باشد، عاملین موظفند این موضوع را به شکلی درست و به موقع به کاربران اطلاع دهند.

سی و نه - از دید فنی و اجتماعی، شفافیت به تحقق صلح و آرامش، عدالت، دموکراسی و فراگیری در جوامع کمک می‌کند، و بررسی عمومی را امکان پذیر می‌سازد. نتیجه این بررسی، کاهش فساد و تبعیض، و شناسایی و جلوگیری از تأثیرات منفی بر حقوق بشر خواهد بود. هدف از شفافیت، ارائه اطلاعات مناسب به مخاطبان و بالا بردن درک آنها از اطلاعات و افزایش اعتماد است. به خصوص در مورد هوش مصنوعی، شفافیت می‌تواند کمک کند تا افراد بفهمند چگونه هر یک از مراحل هوش مصنوعی با توجه به زمینه و حساسیت سیستم در جای خود قرار می‌گیرند. همچنین شفافیت می‌تواند درکی از عوامل تأثیرگذار بر یک پیش‌بینی یا تصمیم خاص را فراهم آورد یا نشان بدهد که اقدامات مناسب در جهت افزایش ایمنی، عدالت و غیره انجام شده‌اند یا خیر. در موارد نقض جدی حقوق بشرممکن است در راستای ایجاد شفافیت، اشتراک گذاری کدها و داده‌ها نیز لازم باشد.

چهل - توضیح پذیری به معنی روشن کردن و ارائه بینش در خصوص پیامدهای سیستم‌های هوش مصنوعی می‌باشد. توضیح پذیری همچنین به قابل فهم بودن ورودی، خروجی و کارکرد هر یک از اجزای سازنده الگوریتم و نقش آن در پیامد سیستم اطلاق می‌شود. به این ترتیب، بین توضیح پذیری و شفافیت ارتباطی تنگاتنگ وجود دارد، زیرا نتایج و فرآیندهای فرعی که به نتایج منتهی می‌شوند باید قابل فهم و ردیابی بوده، با موقعیت تناسب داشته باشند. عاملین هوش مصنوعی باید در قبال توضیح پذیری الگوریتم‌های مورد استفاده متعهد باشند. در مورد اپلیکیشن‌هایی که تأثیر آنها بر کاربر نهایی موقتی نبوده، به راحتی قابل برگشت نیستند و ریسک پایینی هم ندارند، باید اقدامات لازم صورت بگیرند تا همزمان با هر تصمیمی که به چنین اثراتی منجر می‌شود، توضیح معناداری ارائه شود، تا نتیجه از شفافیت لازم برخوردار باشد.

چهل و یک - رابطه نزدیکی بین شفافیت و توضیح پذیری از یک سو، و اقدامات مربوط به مسئولیت پذیری و پاسخگویی از سوی دیگر، و همچنین قابل اعتماد بودن سیستم‌های هوش مصنوعی وجود دارد.

#### **- مسئولیت پذیری و پاسخگویی:**

چهل و دو - عاملین هوش مصنوعی و کشورهای عضو موظف به احترام، حفظ و ارتقاء حقوق بشر و آزادی‌های بنیادین و حفاظت از محیط زیست و اکوسیستم می‌باشند. آنها می‌بایست مسئولیت اخلاقی و قانونی خود را منطبق بر قوانین ملی و بین‌المللی و خصوصاً تعهدات حقوق بشری کشورهای عضو، و راهنمایی‌های اخلاقی در طول چرخه زندگی هوش مصنوعی از جمله در ارتباط با عاملین هوش مصنوعی در حیطه اختیارات خود تقبل کنند. مسئولیت و تعهد اخلاقی برای تصمیمات و اقداماتی که به هر شکل مبتنی بر سیستم هوش



مصنوعی هستند، در نهایت همیشه متوجه عاملین هوش مصنوعی و متناسب با نقش آنها در چرخه زندگی هوش مصنوعی خواهد بود.

چهل و سه - به منظور پاسخگو کردن سیستم‌های هوش مصنوعی در قبال تأثیراتشان در طول چرخه زندگی، لازم است مکانیسم‌های نظارت، ارزیابی اثرات، ممیزی و ارزیابی بایسته (کنکاش موشکافانه) مناسب، از جمله حفاظت از افشاگران، ایجاد شوند. طراحی فنی و سازمانی باید متضمن قابلیت ممیزی و ردیابی (کارکرد) سیستم های هوش مصنوعی باشند، به ویژه به منظور رسیدگی به هرگونه تضاد با هنجارها و استانداردهای حقوق بشر و تهدیدات محیط زیست و اکوسیستم.

#### - آگاهی و سواد:

چهل و چهار - آگاهی و درک عمومی از فن‌آوری هوش مصنوعی و ارزش داده‌ها باید از طریق آموزش باز و در دسترس، مشارکت مدنی، مهارت‌های دیجیتال، تعلیمات مربوط به اخلاق هوش مصنوعی، سواد رسانه و اطلاعات، و آموزش‌هایی که مشترکاً توسط دولت‌ها، سازمان‌های بین‌دولتی، جوامع مدنی، دانشگاه‌ها، رسانه، رهبران جامعه و بخش خصوصی ارائه می‌شوند، ارتقاء یابد. این امر باید با توجه به گوناگونی زبانی، اجتماعی و فرهنگی موجود، و با مشارکت مؤثر عموم صورت گیرد تا همه افراد جامعه بتوانند در مورد استفاده از سیستم های هوش مصنوعی به شکلی آگاهانه و مصون از نفوذ نابجا تصمیم‌گیری کنند.

چهل و پنج - یادگیری درباره اثرات سیستم‌های هوش مصنوعی باید شامل یادگیری درباره‌ی، از طریق، و برای حقوق بشر و آزادی‌های بنیادین باشد. این یعنی رویکرد و درک این سیستم‌ها باید براساس تأثیر آنها بر حقوق بشر و دسترسی به حقوق، و همچنین بر محیط زیست و اکوسیستم باشد.

#### حاکمیت و همکاری چندجانبه و انطباقی:

چهل و شش - قوانین بین‌المللی و حاکمیت ملی باید در استفاده از داده‌ها رعایت شوند. این یعنی دولت‌ها می‌توانند در راستای قوانین بین‌المللی، داده‌های تولیدشده در داخل یا در حال عبور از کشور خود را کنترل کنند، و اقداماتی در جهت مدیریت مؤثر داده‌ها، از جمله حفاظت داده را براساس احترام به حق حریم خصوصی و منطبق با قوانین بین‌المللی و سایر هنجارها و استانداردهای حقوق بشر به عمل آورند.

چهل و هفت - مشارکت ذینفعان در طول چرخه زندگی سیستم‌های هوش مصنوعی برای رویکردهای فراگیر به حاکمیت هوش مصنوعی، تشریک منافع و توسعه پایدار ضروری است. از جمله ذینفعان می‌توان به دولت‌ها، سازمان‌های بین‌دولتی، جامعه فنی، جامعه مدنی، پژوهشگران و دانشگاهیان، رسانه‌ها، بخش آموزشی، سیاست‌گذاران، شرکت‌های بخش خصوصی، بنیادهای حقوق بشر و نهادهای برابری، نهادهای ناظر بر منع تبعیض، و گروه های کودکان و جوانان اشاره کرد. به کارگیری استانداردهای باز و قابلیت تعامل با سایرین برای سهولت همکاری الزامی است. باید تدابیری اتخاذ شوند تا تحولات تکنولوژی و پیدایش ذینفعان جدید در نظر گرفته

شود، ضمن این که گروه ها، جوامع و افراد به حاشیه رانده شده و مردم بومی بتوانند در حاکمیت داده‌های خود به شکلی معنی‌دار مشارکت داشته باشند.

#### **(د) حوزه اقدامات سیاستی**

چهل و هشت - اقدامات سیاستی ذکر شده در حوزه‌های زیر، ارزش‌ها و اصول برشمرده در این توصیه‌نامه را اجرایی می‌کنند، و مهمترین اقدامی که کشورهای عضو می‌توانند انجام دهند، اجرای مؤثر مکانیسم‌ها یا چهارچوب‌های مربوط به سیاست و الزام ذینفعان از جمله شرکت‌های خصوصی، مؤسسات تحقیقاتی و دانشگاهی و جامعه مدنی به رعایت آنها می‌باشد؛ ضمناً کشورهای عضو می‌توانند همه ذینفعان را به توسعه حقوق بشر، حاکمیت قانون، دموکراسی، و ابزارهای ارزیابی اثرات اخلاقی و ارزیابی بایسته منطبق بر راهنمایی‌هایی از قبیل اصول راهنمای سازمان ملل متحد در خصوص تجارت و حقوق بشر تشویق کنند. روند توسعه چنین سیاست‌ها و مکانیسم‌هایی باید در برگزیده همه ذینفعان بوده، موقعیت و اولویت‌های هر یک از کشورهای عضو را در نظر بگیرد. یونسکو نیز می‌تواند به عنوان شریک، از کشورهای عضو در ایجاد، نظارت و ارزیابی مکانیسم‌های سیاستی حمایت کند.

چهل و نه - یونسکو به این موضوع توجه دارد که کشورهای عضو از نظرابعد علمی، تکنولوژی، اقتصادی، آموزشی، قانونی، نظارتی، زیرساخت، اجتماعی، فرهنگی و غیره به درجات مختلف آماده اجرای این توصیه‌نامه خواهند بود. آمادگی در اینجا یک وضعیت دینامیک است و بنابراین یونسکو به منظور اجرای مؤثر این توصیه‌نامه، اقدامات زیر را انجام می‌دهد: (۱) ایجاد روشی برای ارزیابی میزان آمادگی کشورهای عضو برای شناسایی موقعیت خود در زمان‌های مشخصی و در ابعاد مختلف، و (۲) حمایت از کشورهای عضو علاقه مند به ایجاد متدولوژی یونسکو برای ارزیابی تاثیر اخلاقی هوش مصنوعی، اشتراک‌گذاری روش‌های موفق، دستورالعمل‌های ارزیابی، و سایر مکانیسم‌ها و تحلیل‌ها.

#### **حوزه سیاستی ۱: ارزیابی اثرات اخلاقی**

پنجاه - کشورهای عضو باید چهارچوب‌هایی برای ارزیابی اثرات اخلاقی و غیره ارائه کنند تا بتوان مزایا، نگرانی‌ها و ریسک‌های سیستم‌های هوش مصنوعی را شناسایی کرد، و همچنین تدابیر مناسبی جهت پیشگیری از خطرات، اقدامات نظارتی و سایر مکانیسم‌های حصول اطمینان اندیشید. منظور از چنین ارزیابی‌هایی، شناسایی اثرات هوش مصنوعی بر حقوق بشر و آزادی‌های بنیادین است، به خصوص حقوق افراد کمتر مورد توجه و اقشار آسیب‌پذیر یا کسانی که در شرایط حساس قرار دارند، حقوق کارگران، محیط زیست و اکوسیستم و تبعات اخلاقی و اجتماعی. این ارزیابی‌ها همچنین می‌توانند باعث سهولت مشارکت شهروندان در راستای ارزش‌ها و اصول مندرج در این توصیه‌نامه بشوند.

پنجاه و یک - کشورهای عضو و شرکت‌های خصوصی باید مکانیسم‌های نظارتی و ارزیابی بایسته‌ی مناسب ایجاد کنند تا در جهت شناسایی، پیشگیری، کاهش و توضیح در مورد چگونگی برخورد با اثرات سیستم‌های هوش مصنوعی بر حقوق بشر، حاکمیت قانون و جوامع فراگیر، گام بردارند. کشورهای عضو همچنین باید قادر به ارزیابی اثرات اجتماعی/اقتصادی سیستم‌های هوش مصنوعی بر فقر بوده، متعهد شوند که اکنون و در آینده، با استفاده گسترده از این فن‌آوری، فاصله طبقاتی بین فقرا و ثروتمندان افزایش نیافته و شکاف دیجیتال در داخل و میان کشورها عمیق‌تر نخواهد شد. به این منظور باید پروتکل‌های شفافیت قابل اجرا متناسب با دسترسی به اطلاعات ایجاد شوند، از جمله، اطلاعاتی که جزو منافع عمومی هستند ولی در اختیار بخش خصوصی قرار دارند. کشورهای عضو، شرکت‌های خصوصی و جامعه مدنی باید اثرات جامعه‌شناختی و روانشناختی توصیه‌هایی که براساس هوش مصنوعی صورت می‌گیرند را بر استقلال تصمیم‌گیری افراد بررسی کنند. سیستم‌های هوش مصنوعی که به صورت بالقوه برای حقوق بشر تهدیدآمیز باشند باید پیش از عرضه به بازار و به عنوان بخشی از ارزیابی اثرات اخلاقی و در صورت لزوم در شرایط واقعی توسط عاملین هوش مصنوعی به طور گسترده مورد آزمایش قرار بگیرند.

پنجاه و دو - کشورها و شرکت‌های تجاری عضو باید به عنوان بخشی از ارزیابی اثرات اخلاقی، اقدامات لازم را در جهت نظارت بر تمامی مراحل چرخه زندگی سیستم‌های هوش مصنوعی به عمل آورند، از جمله کارکرد الگوریتم‌های مورد استفاده در تصمیم‌گیری، داده‌ها، و عاملین هوش مصنوعی درگیر در این فرآیند، به ویژه در خدمات عمومی و هر جایی که تعامل مستقیم با کاربر نهایی لازم است. تعهدات کشورهای عضو در رابطه با قانون حقوق بشر باید بخشی از جنبه‌های اخلاقی ارزیابی سیستم‌های هوش مصنوعی را تشکیل بدهند.

پنجاه و سه - دولت‌ها باید یک چهارچوب نظارتی برای تنظیم پروسه ارزیابی اثرات اخلاقی، به ویژه توسط مقامات دولتی، داشته باشند. با اجرای چنین پروسه‌ای می‌توان پیامدهای سیستم‌های هوش مصنوعی را پیش‌بینی و از پیامدهای مخرب جلوگیری کرد، ریسک‌ها را کاهش داد، مشارکت شهروندان را تسهیل کرد و به چالش‌های اجتماعی پرداخت. این ارزیابی همچنین باید مکانیسم‌های نظارتی مناسب دارای قابلیت حسابرسی، ردیابی و توضیح‌پذیری را ایجاد کند، تا به این وسیله امکان ارزیابی الگوریتم‌ها، داده‌ها و روند طراحی، و همچنین بررسی خارجی سیستم‌های هوش مصنوعی فراهم شود. ارزیابی‌های اثرات اخلاقی باید در صورت امکان، شفاف و در دسترس عموم باشند، و در چندین زمینه و برای ذینفعان متعدد قابل اجرا بوده، چندفرهنگی، کثرت‌گرا و فراگیر نیز باشند. مقامات دولتی باید از طریق ارائه مکانیسم‌ها و ابزار مناسب، ملزم به نظارت بر سیستم‌های هوش مصنوعی که خود اجرا و یا مستقر کرده‌اند، باشند.

### حوزه سیاستی ۲: حاکمیت و سرپرستی اخلاقی

پنجاه و چهار - کشورهای عضو باید اطمینان حاصل کنند که مکانیسم‌های حاکمیت هوش مصنوعی فراگیر، شفاف، قابل اجرا در چندین زمینه و برای ذینفعان متعدد، و چندجانبه باشند، که این آخری شامل امکان کاهش و جبران آسیب‌ها در آن سوی مرزها نیز می‌شود. حاکمیت باید به‌ویژه شامل ابعاد پیش‌بینی، حفاظت مؤثر، نظارت بر اثرات، اجرا و جبران آسیب‌ها باشد.

پنجاه و پنج - کشورهای عضو باید اطمینان حاصل کنند که آسیب‌های ناشی از سیستم‌های هوش مصنوعی از طریق اعمال مکانیسم‌های اجرایی و اقدامات اصلاحی قوی، بررسی و جبران می‌شوند. به این ترتیب می‌توان مطمئن شد که حقوق بشر، آزادی‌های بنیادین و حاکمیت قانون هم در دنیای دیجیتال و هم در دنیای فیزیکی رعایت می‌شوند. چنین مکانیسم‌ها و اقداماتی باید شامل سازوکارهای اصلاحی ارائه شده توسط شرکت‌های بخش خصوصی و دولتی باشند. ارتقاء قابلیت ممیزی و ردیابی سیستم‌های هوش مصنوعی باید با در نظر داشتن این اهداف انجام بگیرد. به علاوه، کشورهای عضو باید ظرفیت‌های سازمانی خود را برای اجرای این تعهد افزایش بدهند و با همکاری پژوهشگران و سایر ذینفعان به بررسی، پیشگیری و تخفیف اثر هرگونه استفاده مخرب احتمالی از سیستم‌های هوش مصنوعی بپردازند.

پنجاه و شش - توصیه می‌شود که کشورهای عضو، استراتژی‌های ملی و منطقه‌ای در قبال هوش مصنوعی اتخاذ کنند و اشکالی از حکمرانی نرم مانند مکانیسم اعطای گواهینامه هوش مصنوعی، و به رسمیت شناختن گواهینامه‌های یکدیگر را در پیش بگیرند. این موضوع می‌تواند با در نظر گرفتن حساسیت حوزه اپلیکیشن و تأثیر مورد انتظار آن بر حقوق بشر، محیط زیست، اکوسیستم و سایر ملاحظات اخلاقی مندرج در این توصیه نامه صورت بگیرد. چنین مکانیسمی می‌تواند شامل سطوح مختلف بازرسی سیستم‌ها و داده‌ها، و پایبندی به دستورالعمل‌های اخلاقی و اصول و روش کار و با در نظر گرفتن ابعاد اخلاقی باشد. در عین حال، چنین مکانیسمی نباید مانع نوآوری شود، یا در نتیجه اعمال فشار بیش از حد امور اداری، به کسب و کارها و شرکت‌های کوچک و متوسط یا استارت‌آپ‌ها، جامعه مدنی و مؤسسات علمی و تحقیقاتی آسیب برساند. این مکانیسم‌ها همچنین باید دارای یک بخش نظارت دائم برای اطمینان از استحکام سیستم، انسجام پایدار و پیروی از دستورالعمل‌های اخلاقی در تمام طول چرخه زندگی سیستم‌های هوش مصنوعی و در صورت لزوم، تجدید اعتبار گواهینامه باشند.

پنجاه و هفت - کشورهای عضو و مقامات دولتی باید خودارزیابی شفافیت از سیستم‌های هوش مصنوعی موجود و پیشنهادی انجام دهند، که به ویژه مشخص شود که آیا استفاده از هوش مصنوعی درست هست یا خیر، و در صورت درست بودن، در مورد روش مناسب تصمیم‌گیری انجام بدهند. مورد دیگری که باید مشخص شود

این است که آیا استفاده از هوش مصنوعی منجر به نقض یا سوءاستفاده از تعهدات حقوق بشری کشورهای عضو می‌شود، که در آن صورت باید استفاده از هوش مصنوعی را ممنوع کنند.

پنجاه و هشت - کشورهای عضو باید نهادهای دولتی، شرکت‌های بخش خصوصی و سازمان‌های جامعه مدنی را تشویق کنند تا ذینفعان مختلف را در حاکمیت هوش مصنوعی به کار گیرند، و همچنین یک مأمور مستقل یا مکانیسم دیگری را به منظور نظارت بر ارزیابی اثرات اخلاقی، بازرسی و پایش مستمر، و اطمینان از هدایت اخلاقی سیستم‌های هوش مصنوعی به کار گیرند. از کشورهای عضو، شرکت‌های بخش خصوصی و سازمان‌های جامعه مدنی دعوت می‌شود تا با حمایت یونسکو شبکه‌ای متشکل از مأموران مستقل اخلاقیات هوش مصنوعی را به منظور پشتیبانی از این پروسه در سطوح ملی، منطقه‌ای و بین‌المللی ایجاد کنند.

پنجاه و نه - کشورهای عضو باید از توسعه و دسترسی به یک اکوسیستم دیجیتال برای توسعه اخلاقی و فراگیر سیستم‌های هوش مصنوعی در سطح ملی حمایت کنند، تا شکاف‌های موجود در روند دسترسی به چرخه زندگی هوش مصنوعی را برطرف، و به همکاری‌های بین‌المللی کمک کنند. چنین اکوسیستمی به ویژه شامل فن‌آوری و زیرساخت دیجیتال، و مکانیسم‌هایی به منظور اشتراک‌گذاری دانش هوش مصنوعی در صورت لزوم می‌باشد.

شصت - کشورهای عضو ملزم هستند مکانیسم‌هایی را با همکاری نهادهای بین‌المللی، سازمان‌های فراملیتی، مؤسسات دانشگاهی و جامعه مدنی ایجاد کنند تا مشارکت فعال همه کشورهای عضو در گفتگوهای بین‌المللی در مورد حاکمیت هوش مصنوعی تضمین شود، به ویژه کشورهایی با درآمد پایین و متوسط، خصوصاً کشورهای توسعه‌نیافته، کشورهای در حال توسعه محصور در خشکی، و کشورهای جزیره‌ای کوچک در حال توسعه. این مشارکت می‌تواند از طریق تأمین بودجه، تضمین مشارکت منطقه‌ای برابر، یا هر مکانیسم دیگری صورت بگیرد. به علاوه، به منظور اطمینان از فراگیر بودن این مجامع، کشورهای عضو باید رفت و آمد عاملین هوش مصنوعی را به بیرون مناطق خود و بالعکس به منظور شرکت در این مجامع تسهیل کنند، به خصوص در مورد کشورهای در حال توسعه محصور در خشکی، و کشورهای جزیره‌ای کوچک در حال توسعه.

شصت و یک - اصلاحات انجام گرفته در مورد قوانین ملی موجود یا توضیحات قوانین ملی جدید در خصوص سیستم‌های هوش مصنوعی باید منطبق بر تعهدات حقوق بشری کشورهای عضو باشد، و حقوق بشر و آزادی‌های بنیادین را در طول چرخه زندگی سیستم‌های هوش مصنوعی ارتقاء دهند. ترویج این قوانین نیز باید در قالب ابتکارات حاکمیت، نمونه‌های مطلوب شیوه‌های مشارکت مربوط به هوش مصنوعی، و دستورالعمل‌های فنی و مرتبط با متدولوژی ملی و بین‌المللی در حین پیشرفت فن‌آوری هوش مصنوعی صورت بگیرد. بخش‌های مختلف از جمله بخش خصوصی باید در اقداماتشان در برخورد با سیستم‌های هوش

مصنوعی، در جهت احترام، حفاظت و ارتقاء حقوق بشر و آزادی‌های بنیادین، و با استفاده از ابزارهای موجود و نوین در کنار عمل به این توصیه نامه گام بردارند.

شصت و دو - کشورهای عضوی که از سیستم‌های هوش مصنوعی در زمینه‌های حساس حقوق بشر مانند اجرای قانون، رفاه، اشتغال، رسانه و ارائه اطلاعات، بهداشت و درمان و سیستم قضایی مستقل استفاده می‌کنند، موظفند مکانیسم‌هایی فراهم آورند تا مقامات ناظر مناسب از قبیل مقامات مستقل حفاظت از داده‌ها، ناظرین بخش‌ها و نهادهای دولتی ناظر بتوانند اثرات اجتماعی/اقتصادی این سیستم‌ها را رصد کنند.

شصت و سه - کشورهای عضو باید ظرفیت قوه قضاییه را برای تصمیم‌گیری‌های مرتبط با سیستم‌های هوش مصنوعی منطبق بر حکم قانون و قوانین و استانداردهای بین‌المللی افزایش دهند، تا این قوه بتواند از سیستم‌های هوش مصنوعی در بررسی‌های خود با رعایت اصل نظارت عامل انسانی بهره‌بردارد. در صورت به کارگیری سیستم‌های هوش مصنوعی توسط قوه قضاییه لازم است تدابیر کافی برای تضمین حفاظت از حقوق بنیادین بشر، حاکمیت قانون، استقلال قوه قضاییه، اصل نظارت عامل انسانی و غیره به کار گرفته شود، تا توسعه و استفاده از هوش مصنوعی در این قوه به شکلی قابل اعتماد و با محوریت انسان و عموم مردم انجام پذیرد.

شصت و چهار - کشورهای عضو باید اطمینان حاصل کنند که دولت‌ها و سازمان‌های چندوجهی از طریق مشارکت ذینفعان متعدد، نقشی پیشرو در حفاظت از امنیت و ایمنی سیستم‌های هوش مصنوعی ایفا می‌کنند. کشورهای عضو، نهادهای بین‌المللی و سایر مقامات مرتبط باید استانداردهای بین‌المللی برای توصیف سطوح قابل اندازه‌گیری و آزمایش ایمنی و شفافیت ایجاد کنند، تا بتوان با ارزیابی بی‌طرفانه، میزان پیروی سیستم‌ها را از قوانین مربوطه معین کرد. کشورهای عضو و شرکت‌های تجاری باید به طور مستمر از تحقیقات استراتژیک در مورد ریسک‌های ایمنی و امنیتی احتمالی سیستم‌های هوش مصنوعی حمایت کنند. همچنین آنها باید تحقیق در مورد شفافیت و توضیح‌پذیری، فراگیری و سوادآموزی را از طریق اختصاص بودجه بیشتر در زمینه‌ها و سطوح مختلف مانند زبان تخصصی و طبیعی، تشویق کنند.

شصت و پنج - کشورهای عضو باید سیاست‌هایی اتخاذ کنند تا اطمینان حاصل شود که اقدامات عاملین هوش مصنوعی در طول چرخه زندگی آن با قوانین، استانداردها و اصول بین‌المللی حقوق بشر سازگار است. در این مورد، کشورهای عضو باید ضمن در نظر گرفتن تقدم و جهان‌شمولی حقوق بشر، به گوناگونی فرهنگی و اجتماعی از قبیل سنت‌های محلی و مذهبی توجه کامل داشته باشند.

شصت و شش - کشورهای عضو باید مکانیسم‌هایی را اجرایی کنند که عاملین هوش مصنوعی را ملزم به افشا و مبارزه با هر نوع تفکر قالبی آگاهانه یا غیرآگاهانه در ارتباط با نتایج سیستم‌های هوش مصنوعی و داده‌ها بکنند. کشورهای عضو همچنین باید تضمین کنند که داده‌های آموزشی سیستم‌های هوش مصنوعی باعث

ترویج نابرابری‌ها و تعصبات فرهنگی، اقتصادی یا اجتماعی و اطلاعات نادرست یا دروغ نشده، به آزادی بیان و دسترسی به اطلاعات خدشه‌ای وارد نمی‌کنند. در این میان لازم است به مناطقی که داده‌ها در مورد آنها محدود است توجه ویژه بشود.

شصت و هفت - کشورهای عضو باید سیاست‌هایی را در جهت ارتقاء و افزایش گوناگونی و فراگیری اجرا کنند که منعکس‌کننده‌ی این جمعیت‌ها در تیم‌های توسعه و داده‌های آموزشی هوش مصنوعی بوده، ضامن دسترسی برابر به این فن‌آوری و مزایای آن باشند، به ویژه برای گروه‌های به حاشیه رانده شده، چه در مناطق روستایی و چه در شهرها.

شصت و هشت - به منظور افزایش تعهد و مسئولیت‌پذیری در قبال محتوا و نتایج سیستم‌های هوش مصنوعی در طول چرخه زندگی آنها، کشورهای عضو می‌بایست چهارچوب‌های نظارتی را بنا به اقتضا ایجاد، بازبینی، و سازگار نمایند. همچنین کشورهای عضو باید در صورت لزوم چهارچوب‌هایی برای تعهد ایجاد کنند و یا چهارچوب‌های موجود را به روشنی تفسیر کنند تا به این وسیله (افراد و سازمان‌ها را) در قبال نتایج و کارکرد سیستم‌های هوش مصنوعی پاسخگو سازند. ضمناً در زمان ایجاد چهارچوب‌های نظارتی، کشورهای عضو باید به ویژه این موضوع را در نظر بگیرند که مسئولیت و پاسخگویی نهایی باید متوجه اشخاص حقیقی یا حقوقی باشد، و نباید به سیستم‌های هوش مصنوعی شخصیت حقوقی نسبت داده شود. به این منظور، چهارچوب‌هایی از این دست باید منطبق بر اصل نظارت عامل انسانی باشند و از رویکردی جامع و متمرکز بر عاملین هوش مصنوعی و فرآیندهای تکنولوژیک چرخه زندگی سیستم‌های هوش مصنوعی برخوردار باشند.

شصت و نه - در مواردی که هنجاری وجود ندارد و باید هنجارهای جدید تعیین شوند، یا قرار است چهارچوب‌های قانونی موجود پس از سازگاری مورد استفاده قرار بگیرند، کشورهای عضو ملزم به مشارکت دادن عاملین هوش مصنوعی (از جمله پژوهشگران، نمایندگان جامعه مدنی و مجریان قانون، بیمه‌گران، سرمایه‌گذاران، تولیدکنندگان، مهندسان، وکلا، کاربران و غیره) می‌باشند. این هنجارها می‌توانند به روش‌های موفق، قوانین و مقررات تبدیل شوند. بهتر است کشورهای عضو از مکانیسم‌هایی همچون نمونه اولیه سیاست و سندباکس (جعبه‌شنی)‌های نظارتی استفاده کنند تا روند وضع قوانین، مقررات، سیاست‌ها و بازنگری منظم آنها را در راستای توسعه سریع فن‌آوری‌های نوین تسریع کنند، و پیش از اجرای رسمی قوانین بتوانند آنها را در محیطی امن آزمایش کنند. کشورهای عضو باید از دولت‌های محلی در توسعه سیاست‌ها، قوانین و مقررات محلی منطبق بر چهارچوب‌های قانونی ملی و بین‌المللی حمایت کنند.

هفتاد - کشورهای عضو باید الزامات روشنی را برای شفافیت و توضیح‌پذیری سیستم‌های هوش مصنوعی ارائه دهند تا از قابل اعتماد بودن همه مراحل چرخه زندگی این سیستم‌ها اطمینان حاصل شود. این الزامات باید

در برگیرنده طراحی و اجرای مکانیسم های اثرگذاری باشند که ماهیت حوزه عملکرد اپلیکیشن، استفاده موردنظر، مخاطبان و امکان پذیر بودن هر سیستم هوش مصنوعی خاص را در نظر بگیرند.

### حوزه سیاستی ۳: سیاست داده‌ها

هفتاد و یک - کشورهای عضو باید در جهت توسعه استراتژی‌های حاکمیت داده‌ها که متضمن ارزیابی مستمر کیفیت داده‌های آموزشی برای سیستم‌های هوش مصنوعی هستند، تلاش کنند. منظور از کیفیت، کافی بودن پروسه جمع‌آوری و انتخاب داده‌ها، استفاده از اقدامات امنیتی و حفاظتی مناسب، و مکانیسم‌های بازخوردی برای درس گرفتن از اشتباهات و اشتراک گذاری روش‌های موفق در میان عاملین هوش مصنوعی می‌باشد.

هفتاد و دو - کشورهای عضو باید در راستای قوانین بین‌المللی تدابیر مناسبی را برای حفاظت از حریم خصوصی بیان‌دیشند و به نگرانی‌هایی در مورد مسائلی از قبیل زیر نظر گرفتن افراد پردازند. ضمناً کشورهای عضو باید چهارچوب‌های قانونی را اتخاذ و اجرا کنند که امکان حفاظت مناسبی را مطابق با قوانین بین‌المللی فراهم می‌سازند. کشورهای عضو می‌بایست همه عاملین هوش مصنوعی از جمله شرکت های تجاری را قویاً به این دو امر تشویق کنند: (۱) پیروی از استانداردهای بین‌المللی موجود، و به ویژه انجام ارزیابی کافی از اثرات بر حریم خصوصی به عنوان بخشی از ارزیابی اثرات اخلاقی، که به تأثیرات گسترده‌تر اجتماعی/اقتصادی پردازش داده‌های مورد نظر توجه می‌کنند، و (۲) تلاش برای حفظ حریم خصوصی در طراحی سیستم‌های خود. در طول چرخه زندگی سیستم‌های هوش مصنوعی، احترام، حفاظت و ارتقاء حریم خصوصی لازم است.

هفتاد و سه - کشورهای عضو باید اطمینان حاصل کنند که افراد حقوق خود را در رابطه با داده‌های شخصی خود حفظ می‌کنند و توسط چهارچوبی که موارد زیر را در بر می‌گیرد حمایت می‌شوند: شفافیت؛ حمایت‌های لازم در مورد پردازش داده‌های حساس؛ سطح مناسبی از حفاظت از داده‌ها؛ برنامه‌ها و مکانیسم‌های پاسخگویی مؤثر و معنادار؛ برخورداری تمام و کمال صاحبان داده‌ها از حقوقشان و امکان دسترسی و حذف داده‌های شخصی در سیستم‌های هوش مصنوعی، مگر در موارد خاص و مطابق با قوانین بین‌المللی؛ سطح قابل قبولی از حمایت در انطباق کامل با قوانین حفاظت از داده‌ها، در شرایطی که از داده‌ها برای مقاصد تجاری استفاده شود، از قبیل تبلیغات خرد و انتقال برون مرزی داده‌ها؛ و نظارت مؤثر و مستقل به عنوان بخشی از مکانیسم حاکمیت داده‌ها، که به افراد امکان کنترل داده‌هایشان را می‌دهد و مزایای جریان آزادانه داده‌ها در سطح بین‌المللی و دسترسی به داده‌ها را تقویت می‌کند.

هفتاد و چهار - کشورهای عضو باید سیاست‌های داده یا چهارچوب‌های مشابهی را وضع کنند و یا چهارچوب‌های موجود را اعمال کنند تا بتوانند امنیت کامل داده‌های شخصی و حساس که در صورت افشا می‌توانند آسیب‌ها و سختی‌های شدیدی برای افراد ایجاد کنند را حفظ کنند. به عنوان مثال، داده‌های مربوط به جرایم، دادرسی‌های کیفری، محکومیت‌ها و اقدامات امنیتی مرتبط؛ داده‌های بیومتریک، ژنتیک و سلامت؛



و داده‌های شخصی از قبیل نژاد، رنگ پوست، اصل و نسب، جنسیت، سن، زبان، مذهب، عقاید سیاسی، خاستگاه ملی، قومی و اجتماعی، شرایط اقتصادی یا اجتماعی خانوادگی، معلولیت و غیره. هفتاد و پنج - کشورهای عضو باید داده‌های باز را ترویج کنند، و در این راستا باید سیاست‌ها و چهارچوب‌های نظارتی خود را در مورد دسترسی به اطلاعات و دولت باز، بازنگری کنند، به نحوی که منعکس‌کننده‌ی الزامات و مکانیسم‌های ارتقاء مختص هوش مصنوعی باشد. مخازن باز برای داده‌هایی که تأمین مالی یا نگهداری آنها بر عهده عموم است، کد مبدأ و سپرده‌گذاری داده‌ها همگی نمونه‌هایی از این مکانیسم‌ها هستند که اشتراک‌گذاری ایمن، عادلانه، قانونی و اخلاقی داده‌ها را امکان‌پذیر می‌سازند.

هفتاد و شش - کشورهای عضو باید استفاده از مجموعه داده‌های با کیفیت و قوی را به منظور آموزش، توسعه و استفاده از سیستم‌های هوش مصنوعی ترویج و تسهیل کنند و در امر نظارت بر جمع‌آوری و به کارگیری آنها هوشیار باشند. این می‌تواند در صورت امکان شامل سرمایه‌گذاری در تولید مجموعه داده‌های استاندارد طلا باشد، و یا مجموعه داده‌های باز و قابل اعتماد متنوع و مبتنی بر مبنای قانونی معتبری مانند کسب رضایت صاحبان داده‌ها، در صورتی که قانون آن را ایجاب کند. استانداردهای برچسب‌گذاری مجموعه داده‌ها، شامل تفکیک داده‌ها بر اساس جنسیت و غیره باید ترویج شود تا بتوان به آسانی مشخص کرد که یک مجموعه داده‌ها چگونه جمع‌آوری می‌شود و دارای چه خصوصیتی می‌باشد.

هفتاد و هفت - همان‌گونه که در گزارش هیأت عالی‌رتبه همکاری دیجیتال دبیرکل سازمان ملل متحد آمده است، کشورهای عضو باید با حمایت سازمان ملل و یونسکو، رویکرد دیجیتالی مشترکی اتخاذ کنند که تعامل ابزارها و مجموعه داده‌ها و اینترفیس داده‌های میزبانی سیستم‌ها را افزایش دهد. کشورهای عضو همچنین باید شرکت‌های بخش خصوصی را تشویق کنند که در صورت امکان، داده‌هایی را که جمع‌آوری می‌کنند با همه ذینفعان و با هدف پژوهش، نوآوری یا منفعت عموم، به اشتراک بگذارند. کشورهای عضو همچنین باید بخش دولتی و خصوصی را تشویق کنند که پلتفرم‌های مشترکی به منظور اشتراک‌گذاری داده‌های با کیفیت در فضاهای امن و مورد اعتماد ایجاد کنند.

#### حوزه سیاستی ۴: توسعه و همکاری بین‌المللی

هفتاد و هشت - کشورهای عضو و شرکت‌های فراملیتی باید اخلاق هوش مصنوعی را از طریق گنجاندن بحث‌های مرتبط در مجامع بین‌المللی، بین‌دولتی و مشترک بین چندین ذینفع، در اولویت قرار دهند. هفتاد و نه - کشورهای عضو باید اطمینان حاصل کنند که کاربرد هوش مصنوعی در زمینه‌های توسعه مانند آموزش، دانش، فرهنگ، ارتباطات و اطلاعات، بهداشت و درمان، کشاورزی و منابع غذایی، محیط زیست، مدیریت منابع طبیعی و زیرساخت، برنامه‌ریزی و رشد اقتصادی و غیره بر ارزش‌ها و اصول مندرج در این توصیه‌نامه منطبق است.

هشتاد - کشورهای عضو باید از طریق سازمان‌های بین‌المللی تلاش کنند پلتفرم‌هایی برای همکاری بین‌المللی در زمینه توسعه هوش مصنوعی فراهم آورند. این همکاری می‌تواند به صورت آوردن تخصص، بودجه، داده، دانش دامنه (domain)، زیرساخت، و تسهیل تشریک مساعی چندین ذینفع به منظور مقابله با مشکلات چالش‌برانگیز توسعه باشد، به‌ویژه برای کشورهایی با درآمد پایین و متوسط، و به‌خصوص کشورهای توسعه‌نیافته، کشورهای در حال توسعه محصور در خشکی، و کشورهای جزیره‌ای کوچک در حال توسعه.

هشتاد و یک - کشورهای عضو باید در جهت ترویج همکاری بین‌المللی در زمینه پژوهش و نوآوری هوش مصنوعی تلاش کنند، از جمله در جهت ارتقاء مراکز و شبکه‌های پژوهش و نوآوری که منجر به مشارکت و رهبری گسترده‌تری از جانب پژوهشگران کشورهای با درآمد پایین و متوسط و سایر کشورها می‌شود، از جمله کشورهای توسعه‌نیافته، کشورهای در حال توسعه محصور در خشکی، و کشورهای جزیره‌ای کوچک در حال توسعه.

هشتاد و دو - کشورهای عضو می‌بایست از طریق همکاری با سازمان‌ها و مؤسسات تحقیقاتی بین‌المللی و شرکت‌های فراملیتی، پژوهش‌های اخلاق هوش مصنوعی را ترویج کنند. این می‌تواند مبنایی برای استفاده اخلاقی از سیستم‌های هوش مصنوعی توسط نهادهای دولتی و خصوصی باشد، از جمله تحقیقات در زمینه قابلیت اجرای چهارچوب‌های اخلاقی خاص در موقعیت‌ها و فرهنگ‌های خاص، و امکان توسعه راه‌حلی‌هایی که از نظر فنی امکان‌پذیر و در راستای این چهارچوب‌ها باشند.

هشتاد و سه - کشورهای عضو باید مشارکت و همکاری بین‌المللی در زمینه هوش مصنوعی را ترویج کنند تا امکان پر کردن شکاف‌های ژئوتکنولوژیک را فراهم سازند. همچنین لازم است در زمینه تکنولوژی و با رعایت کامل قوانین بین‌المللی، بین کشورهای عضو و مردمانشان، بین بخش‌های خصوصی و دولتی، و میان پیشرفته‌ترین و توسعه‌نیافته‌ترین کشورها، تبادلات و مشاوره صورت بگیرد.

#### حوزه سیاستی ۵: محیط زیست و اکوسیستم

هشتاد و چهار - کشورهای عضو و شرکت‌های تجاری باید اثرات زیست‌محیطی مستقیم و غیرمستقیم سیستم‌های هوش مصنوعی را در طول چرخه زندگی آنها ارزیابی کنند. این موضوع شامل ردپای کربنی، مصرف انرژی، و پیامدهای زیست‌محیطی استخراج مواد خام به منظور حمایت از تولید فن‌آوری هوش مصنوعی و غیره می‌شود. کشورهای عضو همچنین ملزم هستند اثرات زیست‌محیطی سیستم‌های هوش مصنوعی و زیرساخت‌های داده‌ها را کاهش بدهند، و از پیروی همه‌عاملین هوش مصنوعی از قوانین، سیاست‌ها و رویه‌های زیست‌محیطی اطمینان حاصل کنند.

هشتاد و پنج - کشورهای عضو باید هر زمان که لازم و مناسب است، بابت ارائه و به‌کارگیری راه‌حل‌های اخلاقی، حق‌محور و مبتنی بر فن‌آوری هوش مصنوعی برای مقابله با بلایا و سوانح، نظارت، حفاظت و بازسازی

محیط زیست، اکوسیستم و کره زمین، پاداش‌هایی را در نظر بگیرند؛ سیستم‌های هوش مصنوعی باید در طول چرخه زندگی خود مشارکت جوامع محلی و بومی را به کار بگیرند و از رویکردهایی از جنس اقتصاد چرخشی و الگوهای تولید و مصرف پایدار حمایت کنند. نمونه‌هایی از این رویکردها، استفاده از سیستم‌های هوش مصنوعی در صورت لزوم و با اهداف زیر است:

- (۱) کمک به حفاظت، نظارت و مدیریت منابع طبیعی
- (۲) کمک به پیش‌بینی، پیشگیری، کنترل و کاهش شدت مشکلات مربوط به آب و هوا
- (۳) حمایت از یک اکوسیستم غذایی کارآمدتر و پایدارتر
- کمک به تسریع دسترسی و پذیرش انبوه انرژی پایدار
- (۴) ترویج زیرساخت‌ها و الگوهای تجاری و مالی پایدار به عنوان روند اصلی
- (۵) شناسایی آلاینده‌ها یا پیش‌بینی سطح آلودگی و از این طریق کمک به ذینفعان مربوطه برای تعیین، برنامه‌ریزی و اجرای مداخلات هدفمند به منظور جلوگیری و کاهش آلودگی و خطرقرار گرفتن در معرض آلودگی.

هشتاد و شش - برخی از روش‌های هوش مصنوعی به صورت بالقوه نیازمند حجم زیادی از داده‌ها یا منابع هستند، که باعث می‌شود محیط زیست را تحت تأثیر قرار دهند. به همین علت کشورهای عضو باید عاملین هوش مصنوعی را به انتخاب شیوه‌هایی ترغیب کنند که داده، انرژی و منابع کمتری استفاده می‌کنند. به علاوه باید الزاماتی ایجاد شوند تا نشان دهند که یک اپلیکیشن هوش مصنوعی از اثربخشی مطلوبی برخوردار است، و یا اینکه تدابیر حفاظتی کافی برای توجیه استفاده از آن اپلیکیشن موجود هستند. در غیر این صورت باید اصل احتیاط را در نظر گرفت و در مواردی که به کارگیری هوش مصنوعی تبعات منفی فراوان برای محیط زیست به همراه داشته باشد، نباید از این فن‌آوری استفاده کرد.

#### حوزه سیاستی ۶: جنسیت

هشتاد و هفت - کشورهای عضو باید اطمینان حاصل کنند که از تمامی ظرفیت فن‌آوری دیجیتال و هوش مصنوعی در جهت دستیابی به برابری جنسیتی استفاده می‌شود، و در هیچ مرحله‌ای از چرخه زندگی هوش مصنوعی به حقوق بشر، آزادی‌های بنیادین، ایمنی و تمامیت دختران و زنان خدشه‌ای وارد نمی‌شود. علاوه بر این، ارزیابی اثرات اخلاقی سیستم‌های هوش مصنوعی باید از دیدگاهی فراجنسیتی برخوردار باشد.

هشتاد و هشت - کشورهای عضو باید بخشی از بودجه عمومی خود را به تأمین مالی طرح‌های مرتبط با جنسیت تخصیص دهند، و اطمینان حاصل کنند که سیاست‌های دیجیتال ملی شامل یک برنامه عملیاتی جنسیتی هستند، و سیاست‌های مرتبطی را به عنوان مثال در رابطه با آموزش شغلی و با هدف حمایت از دختران و زنان وضع کنند تا آنها از اقتصاد دیجیتال مبتنی بر هوش مصنوعی کنار گذاشته نشوند. همچنین

ضروریست سرمایه‌گذاری ویژه به منظور ارائه برنامه‌های هدفمند و زبان جنسیتی خاص مورد توجه قرار گرفته و اجرا شود، تا برای دختران و زنان فرصت‌های بیشتری برای مشارکت در علوم، تکنولوژی، مهندسی و ریاضیات (STEM<sup>۳</sup>)، و فن‌آوری اطلاعات و ارتباطات (ICT<sup>۴</sup>) فراهم شود، و آمادگی، امکان اشتغال، برابری در پیشرفت شغلی و رشد حرفه‌ای آنان افزایش یابد.

هشتاد و نه - کشورهای عضو باید از تحقق یافتن پتانسیل سیستم‌های هوش مصنوعی برای دستیابی به برابری جنسیتی اطمینان حاصل کنند. کشورهای عضو همچنین باید مطمئن شوند که این فن‌آوری، نه تنها شکاف‌های جنسیتی موجود در دنیای آنالوگ در برخی زمینه‌ها را عمیق‌تر نمی‌کند، بلکه آنها را از بین می‌برد. نمونه‌هایی از این شکاف‌ها شامل موارد زیر هستند: اختلاف میان دستمزد مردان و زنان؛ حضور نابرابر در برخی مشاغل و فعالیت‌ها؛ عدم حضور در سطوح بالای مدیریتی، هیأت‌های مدیره، یا تیم‌های تحقیقاتی در زمینه هوش مصنوعی؛ شکاف تحصیلاتی؛ شکاف مربوط به دسترسی، پذیرش، استفاده و استطاعت مالی در ارتباط با هوش مصنوعی و فن‌آوری دیجیتال؛ و توزیع نابرابر کار بدون دستمزد و مسئولیت‌های مراقبتی در جوامع امروز.

نود - کشورهای عضو باید مراقب باشند که کلیشه‌های جنسیتی و تعصبات تبعیض‌آمیز وارد حیطه سیستم‌های هوش مصنوعی نمی‌شوند، بلکه باید آنها را شناسایی کرده و به صورت فعالانه اصلاح کنند. تلاش برای خنثی کردن تأثیر منفی شکاف‌های تکنولوژیک در دستیابی به برابری جنسیتی و جلوگیری از خشونت‌هایی مانند آزار و اذیت، زورگویی، قاچاق دختران، زنان و گروه‌هایی که به آنها کمتر توجه می‌شود در محیط‌های آنلاین ضروری است.

نود و یک - کشورهای عضو باید در همه مراحل چرخه زندگی هوش مصنوعی، زنان را به کارآفرینی و مشارکت تشویق کنند، که از راه‌های زیر امکان‌پذیر می‌باشد: ارائه و ترویج انگیزه‌های اقتصادی، نظارتی و غیره، و طرح‌های حمایتی، سیاست‌هایی با هدف ایجاد توازن در مشارکت در تحقیقات دانشگاهی مرتبط با هوش مصنوعی، و افزایش حضور زنان در سطوح بالای مدیریت، هیأت‌های مدیره و تیم‌های تحقیقاتی در شرکت‌های فن‌آوری دیجیتال و هوش مصنوعی. کشورهای عضو باید اطمینان حاصل کنند که بودجه عمومی (برای نوآوری، پژوهش و فن‌آوری) به سمت برنامه‌ها و شرکت‌های فراگیر با حضور مناسب جنسیتی هدایت می‌شوند، و سرمایه‌های خصوصی نیز به نحوی مشابه و با استفاده از اصول اقدام مثبت در همین جهت قرار می‌گیرند. لازم است سیاست‌هایی با هدف ایجاد محیط‌های عاری از آزار و اذیت وضع و اجرا شوند، و روش‌های موفق برای ارتقاء گوناگونی در طول چرخه زندگی سیستم‌های هوش مصنوعی به افراد آموزش داده شود.

3. Science, Technology, Engineering and Mathematics

4. Information and Communication Technologies

نود و دو - شرکت‌های عضو باید گوناگونی جنسیتی را در تحقیقات مربوط به هوش مصنوعی در دانشگاه‌ها و صنایع تشویق کنند. از جمله راهکارهای این موضوع می‌توان به این موارد اشاره کرد: دادن انگیزه به دختران و زنان برای ورود به این حوزه، ایجاد مکانیسم‌هایی برای مبارزه با کلیشه‌های جنسیتی و آزار و اذیت در جامعه پژوهشگران هوش مصنوعی، و ترغیب نهادهای دانشگاهی و خصوصی به اشتراک‌گذاری روش‌های موفق ارتقاء گوناگونی جنسیتی.

نود و سه - یونسکو می‌تواند به ایجاد مخزنی از روش‌های موفق ایجاد انگیزه در دختران، زنان و گروه‌هایی که به آنها کمتر توجه می‌شود برای مشارکت در همه مراحل چرخه زندگی سیستم‌های هوش مصنوعی کمک کند.

#### حوزه سیاستی ۷: فرهنگ

نود و چهار - از کشورهای عضو درخواست می‌شود که در هر کجا که مناسب باشد، سیستم‌های هوش مصنوعی را در حفاظت، غنی‌سازی، درک، ارتقاء، مدیریت و دسترسی میراث فرهنگی ملموس، مستند و ناملموس بگنجانند، از جمله در مورد زبان‌های در خطر انقراض و زبان‌ها و دانش بومی. به عنوان مثال کشورهای عضو می‌توانند برنامه‌های آموزشی مرتبط با کاربرد هوش مصنوعی را معرفی و یا به‌روزرسانی کنند، و رویکردی مشارکتی را در برخورد با نهادها و عموم مردم به کار گیرند.

نود و پنج - کشورهای عضو باید اثرات فرهنگی سیستم‌های هوش مصنوعی را بررسی و به آنها رسیدگی کنند، به ویژه نقش اپلیکیشن‌های پردازش طبیعی زبان (NLP) از قبیل نرم‌افزارهای ترجمه خودکار و دستیار صوتی در رابطه با تفاوت‌های ظریف زبان و شیوه بیان انسانی. چنین ارزیابی‌هایی باید ورودی‌هایی را برای طراحی و اجرای استراتژی‌هایی فراهم سازند که مزایای این سیستم‌ها را به حداکثر برسانند، از طریق پرکردن شکاف‌های فرهنگی و ارتقاء درک انسان، و پرداختن به پیامدهای منفی مانند کاهش استفاده، که می‌توانند منجر به از بین رفتن زبان‌های در خطر انقراض، گویش‌های محلی و تفاوت‌های مربوط به لحن و فرهنگ که به زبان و شیوه بیان انسان‌ها مربوط هستند، از بین خواهند رفت.

نود و شش - کشورهای عضو باید آموزش دیجیتال و هوش مصنوعی برای هنرمندان و افراد حرفه‌ای خلاق را ترویج کند تا مشخص شود فن‌آوری هوش مصنوعی تا چه حد می‌تواند در حرفه آنان به کار بیاید. کشورهای عضو همچنین باید ضمن توجه به اهمیت حفظ میراث فرهنگی، گوناگونی و آزادی هنرمندان، به طراحی و اجرای فن‌آوری هوش مصنوعی مناسب کمک کنند، زیرا این فن‌آوری در آفرینش، تولید، توزیع، مخابره و مصرف انواع کالاها و خدمات فرهنگی استفاده می‌شوند.

نود و هفت - کشورهای عضو باید آگاهی و ارزیابی ابزارهای هوش مصنوعی را بین صنایع فرهنگی محلی و کسب و کارهای کوچک و متوسط که در زمینه‌های فرهنگی فعال هستند ارتقاء دهد تا از تمرکز در بازار فرهنگی جلوگیری کند.

نود و هشت - کشورهای عضو باید شرکت‌های فعال در زمینه فن‌آوری و سایر ذینفعان را ترغیب کنند تا طیف متنوعی از تولیدات فرهنگی و دسترسی چندگانه به آنها را ارتقاء دهند، و به ویژه اطمینان حاصل کنند که توصیه‌های الگوریتمی امکان کشف شدن و دیده شدن محتوای محلی را زیادتیر می‌کنند.

نود و نه - کشورهای عضو باید از تحقیقات جدید در نقطه تلاقی هوش مصنوعی و مالکیت فکری (IP) حمایت کنند، به طور مثال برای تعیین این که آیا/چگونه می‌توان با استفاده از حق مالکیت فکری از آثاری که به وسیله هوش مصنوعی تولید شده اند محافظت کرد. کشورهای عضو همچنین باید اثرات فن‌آوری هوش مصنوعی بر حقوق و منافع صاحبان مالکیت فکری که از آثارشان برای تحقیق، توسعه، آموزش یا اعمال اپلیکیشن‌های هوش مصنوعی استفاده میشوند را ارزیابی کنند.

صد - کشورهای عضو می‌بایست در سطح ملی، موزه‌ها، گالری‌ها، کتابخانه‌ها و بایگانی‌ها را به استفاده از سیستم‌های هوش مصنوعی تشویق کنند تا آنها بتوانند ضمن نمایش مجموعه‌ها و توسعه کتابخانه‌ها و پایگاه‌های داده و دانش خود، دسترسی کاربران را افزایش دهند.

#### حوزه سیاستی ۸: آموزش و پژوهش

صد و یک - کشورهای عضو باید با سازمان‌های بین‌المللی، مؤسسات آموزشی و نهادهای خصوصی و غیردولتی همکاری کنند تا سواد هوش مصنوعی کافی را در تمام سطوح و در همه کشورها به عموم آموزش دهند. هدف از این کار توانمندسازی مردم و کم کردن شکاف‌های دیجیتال و نابرابری دسترسی دیجیتال در نتیجه استفاده گسترده از سیستم‌های هوش مصنوعی می‌باشد.

صد و دو - کشورهای عضو باید کسب «مهارت‌های پیش‌نیاز» را در آموزش هوش مصنوعی ترویج کنند، از جمله سواد پایه، سواد عددی (حسابی)، کدگذاری و مهارت‌های دیجیتال، سواد رسانه و اطلاعات، و همچنین تفکر انتقادی و خلاق، کار گروهی، ارتباطات و مهارت‌های اخلاقی اجتماعی/عاطفی و هوش مصنوعی، مخصوصاً در کشورها و مناطقی که کمبود شدیدی در آموزش این مهارت‌ها به چشم می‌خورد.

صد و سه - کشورهای عضو باید برنامه‌های آگاهی عمومی در مورد پیشرفت‌های هوش مصنوعی را ترویج دهند، از جمله در زمینه داده و فرصت‌ها و چالش‌های ناشی از فن‌آوری هوش مصنوعی، اثرات سیستم‌های هوش مصنوعی بر حقوق بشر و پیامدهای آن، از قبیل حقوق کودک. این برنامه‌ها باید برای گروه‌های فنی و غیرفنی قابل دسترسی باشند.

صد و چهار - کشورهای عضو باید مشوق ابتکارات پژوهشی در مورد استفاده مسئولانه و اخلاقی از فن‌آوری هوش مصنوعی در آموزش، تربیت معلم، آموزش الکترونیک و غیره باشند تا در این زمینه فرصت‌ها را تقویت کنند و چالش‌ها و ریسک‌ها را کاهش دهند. ابتکارات باید با ارزیابی کافی از کیفیت آموزش و تأثیرات استفاده از فن‌آوری هوش مصنوعی بر دانش‌آموزان و معلمان همراه باشد. کشورهای عضو همچنین باید اطمینان حاصل کنند که فن‌آوری هوش مصنوعی، دانش‌آموزان و معلمان را توانمند کند و تجربه باکیفیت‌تری را در اختیار آنها قرار دهد، ضمن این که ابعاد رابطه‌ای و اجتماعی و ارزش اشکال سنتی آموزش در روابط بین معلم و شاگرد حیاتی هستند و در بحث به‌کارگیری فن‌آوری هوش مصنوعی در آموزش باید مورد توجه قرار گیرند. وقتی سخن از نظارت، ارزیابی توانایی‌ها، یا پیش‌بینی رفتارهای دانش‌آموزان در میان باشد، سیستم‌های هوش مصنوعی مورد استفاده در آموزش باید از الزامات سخت‌گیرانه‌ای پیروی کنند. هوش مصنوعی باید به فرآیند یادگیری کمک کند، بدون کاهش توانایی‌های شناختی و بدون حذف استخراج اطلاعات حساس، و منطبق بر استانداردهای حفاظت از اطلاعات شخصی مرتبط. داده‌های ارائه شده به منظور کسب دانش که در حین تعامل دانش‌آموز با سیستم هوش مصنوعی جمع‌آوری شده‌اند نباید مورد سوءاستفاده یا بهره‌برداری مجرمانه، از جمله برای اهداف تجاری قرار بگیرند.

صد و پنج - کشورهای عضو باید دختران و زنان، نظام‌های فرهنگی قومی و گوناگون، افراد دارای معلولیت، کسانی که به آنها کمتر توجه می‌شود و اقشار آسیب‌پذیر یا افرادی که در شرایط حساس قرار دارند، اقلیت‌ها و همه کسانی که از تمام مزایای تکنولوژی دیجیتال برخوردار نیستند را برای رهبری و مشارکت در برنامه‌های آموزش هوش مصنوعی در تمام سطوح تشویق کنند، و آنها را به نظارت و اشتراک‌گذاری روش‌های موفق مرتبط با هوش مصنوعی با سایر کشورهای عضو ترغیب نمایند.

صد و شش - کشورهای عضو باید برای همه سطوح و منطبق با برنامه‌ها و روال آموزشی کشور خود، برنامه آموزشی اخلاق هوش مصنوعی تنظیم کنند، و به همکاری متقابل در زمینه آموزش مهارت‌های فنی هوش مصنوعی و ابعاد انسانی، اخلاقی و اجتماعی آموزش این فن‌آوری بپردازند. دوره‌های آنلاین و منابع دیجیتال آموزش هوش مصنوعی باید به زبان‌های محلی و بومی ارائه شوند، ضمن این که لازم است به گوناگونی محیط و به ویژه امکان دسترسی فرمت‌های مختلف برای افراد دارای معلولیت توجه شود.

صد و هفت - کشورهای عضو باید تحقیقات هوش مصنوعی و به ویژه اخلاق مرتبط با این فن‌آوری را ترویج و حمایت کنند، از جمله از طریق سرمایه‌گذاری یا ایجاد انگیزه در بخش دولتی و خصوصی برای سرمایه‌گذاری در این زمینه. در این رابطه توجه به این نکته لازم است که تحقیقات به توسعه و بهبود هرچه بیشتر فن‌آوری هوش مصنوعی با در نظر گرفتن ترویج قوانین بین‌المللی و ارزش‌ها و اصول مندرج در این سند کمک شایانی

خواهد کرد. همچنین کشورهای عضو باید به طور گسترده، روش‌های موفق پژوهشگران و شرکت‌هایی را که با رعایت اخلاقیات در زمینه توسعه هوش اخلاقی فعال هستند و همچنین همکاری با آنها را ترویج دهند. صد و هشت - کشورهای عضو باید اطمینان حاصل کنند که پژوهشگران هوش مصنوعی در زمینه اخلاق پژوهش آموزش می‌بینند، و ضمناً از آنها بخواهند که اخلاقیات را در طراحی، تولیدات، انتشارات، و به ویژه آنالیز مجموعه داده‌هایی که استفاده و برچسب‌گذاری می‌کنند، و کیفیت و دامنه نتایج با استفاده از اپلیکیشن‌های موجود رعایت کنند.

صد و نه - کشورهای عضو باید شرکت‌های خصوصی را تشویق کنند تا دستیابی به داده‌های خود را برای جامعه علمی و به منظور تحقیقات، تسهیل کنند، به ویژه در کشورهایی با درآمد پایین و متوسط و به خصوص کشورهای توسعه‌نیافته، کشورهای در حال توسعه محصور در خشکی، و کشورهای جزیره‌ای کوچک در حال توسعه. این دستیابی باید منطبق بر استانداردهای مرتبط در زمینه حریم خصوصی و حفاظت از داده‌ها باشد. صد و ده - به منظور اطمینان از ارزیابی نقادانه تحقیقات هوش مصنوعی و نظارت صحیح بر سوءاستفاده‌های احتمالی یا اثرات نامطلوب، کشورهای عضو باید مطمئن شوند که هر گونه پیشرفتی در آینده در رابطه با فن‌آوری هوش مصنوعی بر پایه تحقیقات علمی دقیق و مستقل صورت خواهد گرفت. در همین راستا کشورهای عضو باید انجام تحقیقات بین‌رشته‌ای در مورد هوش مصنوعی را با در نظر گرفتن رشته‌هایی به جز علوم، تکنولوژی، مهندسی و ریاضیات (STEM) ترویج کنند. به این ترتیب رشته‌هایی از قبیل مطالعات فرهنگی، آموزش، اخلاق، روابط بین‌الملل، حقوق، زبان‌شناسی، فلسفه، علوم سیاسی، جامعه‌شناسی و روانشناسی در تحقیقات هوش مصنوعی لحاظ خواهند شد.

صد و یازده - با توجه به این که فن‌آوری هوش مصنوعی فرصت‌های خوبی را برای کمک به پیشرفت دانش و عملکرد علمی به ویژه در رشته‌هایی که به شکل سنتی مدل‌رانده هستند فراهم می‌سازد، کشورهای عضو باید جوامع علمی را تشویق کنند که از مزایا، محدودیت‌ها و مخاطرات استفاده از این فن‌آوری آگاه باشند؛ این شامل تلاش برای اطمینان از دقت و صحت نتایج حاصل از رویکردها، مدل‌ها و برخوردهای داده‌محور می‌باشد. علاوه بر این، کشورهای عضو باید از نقش جامعه علمی در کمک به سیاست‌گذاری و بالا بردن سطح آگاهی درباره نقاط قوت و ضعف فن‌آوری هوش مصنوعی استقبال و حمایت کنند.

#### حوزه سیاستی ۹: ارتباطات و اطلاعات

صد و دوازده - کشورهای عضو باید از سیستم‌های هوش مصنوعی به منظور بهبود دسترسی به اطلاعات و دانش استفاده کنند. این می‌تواند شامل حمایت از پژوهشگران، دانشگاهیان، خبرنگاران، توسعه‌دهندگان و عموم مردم باشد، با هدف ارتقاء آزادی بیان، آزادی آکادمیک و علمی، دسترسی به اطلاعات، و اعلام یا افشای فعالانه داده‌ها و اطلاعات رسمی.



صد و سیزده - کشورهای عضو باید اطمینان حاصل کنند که عاملین هوش مصنوعی در رابطه با تولید، بازبینی و بازپخش محتوا به صورت خودکار، آزادی بیان و دسترسی به اطلاعات را در نظر گرفته و ترویج می‌کنند. چهارچوب‌های مناسب از جمله مقررات باید شفافیت اپراتورهای ارتباطات و اطلاعات آنلاین را افزایش دهند، و برای کاربران امکان دسترسی به نظرات گوناگون را فراهم سازند. این چهارچوب‌ها همچنین باید فرآیندهایی را ایجاد کنند که برای کاربران اخطار فوری حاوی دلایل برای حذف محتوا یا هر اقدام دیگری فرستاده شود، و مکانیسم‌هایی جهت تقاضای جبران خسارت در اختیار آنها بگذارند.

صد و چهارده - کشورهای عضو باید از طریق سرمایه‌گذاری و ارتقاء مهارت‌های سواد دیجیتال، رسانه و اطلاعات، تفکر انتقادی و شایستگی‌های لازم برای درک کاربردها و پیامدهای سیستم‌های هوش مصنوعی را تقویت کنند، و به این وسیله با اطلاعات نادرست یا دروغ و نفرت‌پراکنی مقابله و یا آنها را کم‌اثر کنند. درک بهتر و ارزیابی اثرات مثبت یا منفی بالقوه سیستم‌های توصیه‌گر نیز باید در دستور کار قرار بگیرد.

صد و پانزده - کشورهای عضو باید برای رسانه‌ها محیط‌های توانمندسازی ایجاد کنند که در آن حق و منابع ارائه گزارش مؤثر در مورد مزایا و معایب سیستم‌های هوش مصنوعی را داشته باشند، و همچنین باید رسانه‌ها را تشویق کنند تا از سیستم‌های هوش مصنوعی در عملیات خود به شکلی اخلاقی استفاده کنند.

#### حوزه سیاستی ۱۰: اقتصاد و نیروی کار

صد و شانزده - کشورهای عضو باید اثرات سیستم‌های هوش مصنوعی را بر بازار کار و پیامدهای آن را برای الزامات آموزش، در همه کشورها و با تأکید ویژه بر کشورهایی که اقتصاد آنها نیازمند نیروی کار بیشتری است، ارزیابی و رسیدگی کنند. این میتواند شامل معرفی طیف گسترده تری از مهارت‌های مرکزی و بین‌رشته‌ای در تمامی سطوح آموزشی باشد که به کارگران فعلی و نسل‌های بعدی شانس خوبی برای کارایی در بازار به سرعت در حال تغییر، می‌دهد و آگاهی آنان را از جنبه‌های اخلاقی سیستم‌های هوش مصنوعی تضمین می‌کند. لازم است در کنار مهارت‌های تخصصی، فنی و همچنین کارهای مهارت پایین، مهارت‌هایی از جمله «یادگیری چگونه یادگرفتن»، ارتباطات، تفکر انتقادی، کارگروهی، همدلی و توانایی انتقال دانش به دامنه‌های دیگر آموزش داده شود. شفافیت درباره این که چه مهارت‌هایی مورد نیاز هستند و به روز رسانی برنامه آموزش بر حول این محور، کلیدی است.

صد و هفده - کشورهای عضو باید از توافقنامه‌های همکاری میان دولت‌ها، مؤسسات آکادمیک، آموزشگاه‌های فنی و حرفه‌ای، صنعت، سازمان‌های کارگری و جامعه مدنی حمایت کنند تا بتوانند شکاف بین مجموعه مهارت‌های لازم برای همسو کردن برنامه‌ها و استراتژی‌های آموزشی با آینده‌ی کار و نیازهای صنایع، شامل کسب و کارهای کوچک و متوسط را پرکنند. رویکردهای پروژه‌محور آموزش و یادگیری هوش مصنوعی باید

ترویج شوند و امکان مشارکت میان مؤسسات دولتی، شرکت های بخش خصوصی، دانشگاه ها و مراکز تحقیقاتی فراهم آورده شود.

صد و هجده - کشورهای عضو باید با شرکت های بخش خصوصی، جامعه مدنی و سایر ذینفعان از جمله کارگران و اتحادیه ها همکاری کنند تا برای کارکنان در معرض خطر، دوره گذار منصفانه ای را ایجاد کنند. این شامل ارائه برنامه های ارتقاء مهارت و مهارت آموزی مجدد، یافتن مکانیسم های مؤثر برای ادامه اشتغال کارکنان در طول دوره گذار، و بررسی برنامه های «تورایمنی» در صورت عدم ادامه اشتغال می باشد. کشورهای عضو باید برنامه هایی را برای تحقیق و رسیدگی به چالش های شناسایی شده ایجاد و اجرا کنند، از قبیل ارتقاء مهارت و مهارت آموزی مجدد، افزایش حمایت های اجتماعی، سیاست ها و مداخلات صنعتی فعالانه، مزایای مالیاتی، فرم های مالیاتی جدید و غیره. کشورهای عضو باید اطمینان حاصل کنند که برای حمایت از این برنامه ها بودجه عمومی کافی وجود داشته باشد. مقررات مرتبط مانند نظام های مالیاتی باید به دقت بررسی و در صورت نیاز و به منظور مقابله با پیامدهای بیکاری ناشی از اتوماسیون مربوط به هوش مصنوعی، تغییر داده شوند.

صد و نوزده - به منظور پیش بینی روندها و چالش های آینده، کشورهای عضو باید پژوهشگران را به تجزیه و تحلیل اثرات سیستم های هوش مصنوعی بر محیط کار محلی تشویق و از آنها حمایت کنند. این مطالعات باید رویکردی بین رشته ای اتخاذ کنند و اثرات سیستم های هوش مصنوعی را بر بخش های اقتصادی، اجتماعی و جغرافیایی، و همچنین بر تعاملات بین انسان ها و ربات ها، و روابط میان انسان ها بررسی کنند تا بتوانند در مورد روش های موفق مهارت آموزی مجدد و نقل و انتقال به شغل جدید، توصیه هایی ارائه کنند.

صد و بیست - کشورهای عضو باید با در نظر گرفتن اقدامات و مکانیسم های امکان پذیر در سطح ملی، منطقه ای و بین المللی، گام های لازم را در جهت تضمین بازارهای رقابتی و حمایت از مشتری بردارند، تا جلوی سوءاستفاده بازارهای غالب را (به طور مثال از طریق ایجاد انحصار در مورد داده ها، تحقیقات، تکنولوژی یا بازار) در رابطه با سیستم های هوش مصنوعی و در طول چرخه زندگی آنها بگیرند. کشورهای عضو باید مانع از نابرابری های حاصله بشوند، بازارهای مرتبط را ارزیابی کنند و بازارهای رقابتی را ارتقاء دهند. در این میان ضروری است که به کشورهای دارای درآمد پایین و متوسط، به ویژه کشورهای توسعه نیافته، کشورهای در حال توسعه محصور در خشکی، و کشورهای جزیره ای کوچک در حال توسعه توجه لازم معطوف شود، زیرا در نبود زیرساخت، ظرفیت انسانی، مقررات و عوامل دیگر، این کشورها بیشتر در معرض خطر سوءاستفاده های بازارهای غالب می باشند. عاملین هوش مصنوعی که در کشورهایی فعالیت می کنند که استانداردهای اخلاقی در آنها ایجاد یا پذیرفته شده اند باید این استانداردها را در مورد صادرات، توسعه یا استفاده از سیستم های هوش

مصنوعی در کشورهایی که فاقد چنین استانداردهایی هستند، رعایت کنند و در عین حال به قوانین بین‌المللی مرتبط و رویه‌ها، استانداردها و قوانین داخلی این کشورها احترام بگذارند.

### حوزه سیاستی ۱: سلامت و رفاه

صد و بیست و یک - کشورهای عضو می‌بایست تلاش کنند تا سیستم‌های هوش مصنوعی مؤثری را برای بهبود سلامت بشر و حفاظت از حق زندگی به کار گیرند، از قبیل تلاش برای کاهش شیوع بیماری‌ها و ایجاد و حفظ همبستگی بین‌المللی برای مقابله با خطرات و ابهامات مربوط به سلامت. این کشورها باید اطمینان حاصل کنند که استقرار سیستم‌های هوش مصنوعی در حوزه مراقبت سلامت منطبق بر قوانین بین‌المللی و تعهدات حقوق بشری آنها می‌باشد. کشورهای عضو همچنین باید مراقبت کنند که عاملین سیستم‌های هوش مصنوعی حوزه سلامت به اهمیت رابطه بیماران با خانواده‌هایشان و کادر درمان واقف باشند.

صد و بیست و دو - کشورهای عضو باید اطمینان حاصل کنند که توسعه و استقرار سیستم‌های هوش مصنوعی مرتبط با سلامت به طور کلی و سلامت روان به طور خاص، و توجه لازم به کودکان و جوانان به گونه‌ای تنظیم می‌شود که ایمن، اثربخش، کارآمد و از نظر علمی و پزشکی اثبات‌شده باشد، و امکان نوآوری و پیشرفت پزشکی را بر مبنای شواهد فراهم آورد. به علاوه در حوزه مداخلات پزشکی دیجیتال مرتبط، به کشورهای عضو توصیه می‌شود که بیماران و نمایندگانشان را به طور فعال در همه مراحل توسعه سیستم مشارکت دهند.

صد و بیست و سه - کشورهای عضو باید به نظارت بر راه‌حل‌های پیش‌بینی، شناسایی و درمان برای اپلیکیشن‌های هوش مصنوعی توجه ویژه داشته باشند، از طریق:

- ۱) تضمین نظارت برای به حداقل رساندن و کاهش تعصب
- ۲) حصول اطمینان از این که در زمان توسعه الگوریتم‌ها، افراد حرفه‌ای، بیماران، مراقبان (پرستاران) یا کاربران خدمات به عنوان «متخصص دامنه» در همه مراحل مرتبط و به عنوان اعضای تیم حضور داشته باشند
- ۳) توجه لازم به حریم خصوصی به دلیل احتمال نیاز به نظارت پزشکی، و اطمینان از رعایت همه الزامات ملی و بین‌المللی حفاظت از داده‌ها
- ۴) اجرای مکانیسم‌های مؤثر به گونه‌ای که افراد از استفاده و تجزیه و تحلیل داده‌های شخصی خود مطلع بوده و رضایت آگاهانه بدهند، بدون این که از مراقبت‌های سلامت محروم شوند
- ۵) حصول اطمینان از این که مراقبت‌های انسانی و تصمیم‌گیری در مورد تشخیص و درمان نهایتاً توسط عامل انسانی صورت می‌گیرد، هر چند سیستم‌های هوش مصنوعی می‌توانند در این موارد به عنوان دستیار انسان‌ها عمل کنند
- ۶) تضمین بازبینی سیستم‌های هوش مصنوعی توسط کمیته تحقیقات اخلاقی پیش از استفاده بالینی و در صورت لزوم

صد و بیست و چهار - کشورهای عضو باید درباره اثرات و تعدیل آسیب‌های احتمالی سیستم‌های هوش مصنوعی به سلامت روان، تحقیقات به عمل آورند. این آسیب‌ها شامل افزایش میزان افسردگی، اضطراب، انزوای اجتماعی، اعتیاد، قاچاق، افراط‌گرایی، اطلاعات نادرست و غیره می‌شوند.

صد و بیست و پنج - کشورهای عضو ملزم هستند دستورالعمل‌هایی برای تعاملات بین انسان و ربات و تأثیر آن بر روابط میان انسان‌ها بر پایه تحقیقات، و با هدف توسعه ربات‌ها در آینده و توجه ویژه به سلامت جسمی و روانی انسان‌ها ایجاد کنند. در این زمینه لازم است به کاربرد ربات‌ها در حوزه سلامت و مراقبت از سالمندان و افراد دارای معلولیت، آموزش، و همچنین ربات‌های مورد استفاده کودکان، اسباب‌بازی‌های رباتیک، بات‌های مکالمه و همراهی مخصوص کودکان و بزرگسالان توجه خاص مبذول شود. علاوه بر این باید از فن‌آوری هوش مصنوعی برای افزایش ایمنی و استفاده ارگونومیک از ربات‌ها، از جمله در محیط‌های کاری مشترک بین ربات و انسان استفاده شود. ضمناً باید به این موضوع توجه ویژه داشت که ممکن است هوش مصنوعی برای کنترل و سوءاستفاده از جهت‌گیری‌های شناختی انسان‌ها به کار گرفته شود.

صد و بیست و شش - کشورهای عضو باید اطمینان حاصل کنند که تعاملات بین انسان‌ها و ربات‌ها تابع ارزش‌ها و اصولی می‌باشند که سایر سیستم‌های هوش مصنوعی از آن پیروی می‌کنند، از جمله حقوق بشر و آزادی‌های بنیادین، ترویج گوناگونی، و حمایت از افشار آسیب‌پذیر یا افرادی که در شرایط حساس قرار دارند. به منظور حفظ کرامت و خودمختاری انسان لازم است به پرسش‌های اخلاقی درمورد سیستم‌های مبتنی بر هوش مصنوعی برای نورو تکنولوژی و اینترفیس بین کامپیوتر و مغز انسان پاسخ داده شود.

صد و بیست و هفت - کشورهای عضو باید مطمئن شوند که کاربران می‌توانند به آسانی تشخیص دهند که در حال تعامل با یک موجود زنده هستند یا یک سیستم هوش مصنوعی مقلد ویژگی‌های انسان یا حیوان، و قادرند به شکل مؤثر از ادامه چنین تعاملی سر باز بزنند و مداخله انسانی درخواست کنند.

صد و بیست و هشت - کشورهای عضو باید با هدف افزایش آگاهی درباره انسان‌انگاری فن‌آوری هوش مصنوعی و تکنولوژی‌هایی که عواطف انسانی را شناسایی و تقلید می‌کنند سیاست‌هایی اعمال کنند، از جمله به زبانی که برای اشاره به آنها استفاده می‌شود، و تظاهرات، پیامدهای اخلاقی و محدودیت‌های احتمالی انسان‌انگاری را ارزیابی کنند، به ویژه در مورد تعامل ربات و انسان، و مخصوصاً در شرایطی که کودکان درگیر باشند.

صد و بیست و نه - کشورهای عضو باید پژوهش‌های مشترک درباره اثرات تعامل طولانی‌مدت افراد با سیستم‌های هوش مصنوعی را تشویق و ترویج کنند، و مخصوصاً به تأثیر شناختی و روانشناختی این سیستم‌ها بر کودکان و جوانان بپردازند. این امر باید با استفاده از هنجارها، اصول، پروتکل‌ها و رویکردهای انضباطی گوناگون، و ارزیابی اصلاح رفتارها و عادات، و همچنین برآورد دقیق اثرات پایین‌دستی فرهنگی و اجتماعی

صورت بگیرد. علاوه بر این، کشورهای عضو باید تحقیق درباره تأثیر فن‌آوری هوش مصنوعی بر عملکرد سیستم سلامت و نتایج سلامت را ترویج کنند.

صد و سی - کشورهای عضو و همه ذینفعان باید مکانیسم‌هایی را برای مشارکت معنادار کودکان و جوانان در مکالمات، مناظرات و تصمیم‌گیری‌های مرتبط با اثرات سیستم‌های هوش مصنوعی بر زندگی و آینده آنها اعمال کنند.

### ه) نظارت و ارزیابی

صد و سی و یک - هر یک از کشورهای عضو باید با توجه به ساختارهای حاکم، مفاد قانون اساسی و شرایط خاص خود، نظارت و ارزیابی شفاف و معتبری را از سیاست‌ها، برنامه‌ها و مکانیسم‌های مرتبط با اخلاق هوش مصنوعی از طریق ترکیبی از رویکردهای کمی و کیفی به عمل آورند. به منظور حمایت از کشورهای عضو، یونسکو می‌تواند اقدامات زیر را انجام دهد:

۱) ایجاد یک متدولوژی مناسب توسط یونسکو برای ارزیابی اثرات اخلاقی فن‌آوری هوش مصنوعی بر پایه تحقیقات علمی دقیق و مبتنی بر قوانین حقوق بشر بین‌المللی، راهنمایی برای اعمال این متدولوژی در همه مراحل چرخه زندگی سیستم‌های هوش مصنوعی، و ارائه مواد لازم برای ظرفیت‌سازی به منظور پشتیبانی از تلاش‌های کشورهای عضو برای آموزش این متدولوژی به مقامات دولتی، سیاست‌گذاران و سایر عاملین هوش مصنوعی مرتبط

۲) ایجاد یک متدولوژی برای ارزیابی میزان آمادگی یونسکو برای کمک به کشورهای عضو در جهت شناسایی موقعیت خود در زمان‌های مشخصی از مسیر آمادگی در ابعاد مختلف

۳) ایجاد یک متدولوژی توسط یونسکو برای ارزیابی پیش از وقوع و پس از وقوع میزان اثربخشی و کارآمدی سیاست‌های اخلاقی و انگیزه‌های مربوط به هوش مصنوعی

۴) تقویت تجزیه و تحلیل مبتنی بر تحقیقات و شواهد و ارائه گزارش درباره سیاست‌های مرتبط با اخلاق هوش مصنوعی

۵) جمع‌آوری و نشر پیشرفت‌ها، نوآوری‌ها، گزارش تحقیقات، نشریات، داده‌ها و آمار علمی در رابطه با سیاست‌های اخلاقی هوش مصنوعی از طریق ابتکارات موجود، به منظور به اشتراک‌گذاری روش‌های موفق و یادگیری متقابل، و پیشبرد عملی این توصیه نامه

صد و سی و دو - فرآیندهای نظارت و ارزیابی باید مشارکت گسترده همه ذینفعان را تضمین کنند، از جمله اقشار آسیب‌پذیر یا افرادی که در شرایط حساس قرار دارند. گوناگونی اجتماعی، فرهنگی و جنسیتی باید با در نظر گرفتن بهبود روند آموزش و تقویت روابط بین یافته‌ها، تصمیم‌گیری، شفافیت و مسئولیت در قبال نتایج تضمین شود.

صد و سی و سه - به منظور ترویج سیاست‌ها و روش‌های موفق در رابطه با اخلاق هوش مصنوعی باید ابزارها و شاخص‌های مناسب جهت ارزیابی اثربخشی و کارآمدی آنها در برابر استانداردها، اولویت‌ها و اهداف مورد توافق (از جمله اهداف خاص برای افراد محروم و به حاشیه رانده شده، اقشار آسیب‌پذیر و افرادی که در شرایط حساس قرار دارند)، و همچنین برای ارزیابی اثرات سیستم‌های هوش مصنوعی در سطوح فردی و اجتماعی ایجاد شوند. نظارت و ارزیابی اثرات سیستم‌های هوش مصنوعی و سیاست‌ها و روش‌های اخلاقی هوش مصنوعی باید به طور پیوسته و سیستماتیک و متناسب با ریسک‌های مرتبط انجام شود. این موضوع باید براساس چهارچوب‌های مورد توافق بین‌المللی صورت بگیرد و شامل ارزیابی مؤسسات خصوصی و دولتی، ارائه‌دهندگان خدمات هوش مصنوعی و برنامه‌ها باشد، از جمله خود ارزیابی، مطالعات ردیابی و ایجاد مجموعه‌ای از شاخص‌ها. جمع‌آوری و پردازش داده‌ها باید مطابق با قوانین بین‌المللی، قوانین ملی درباره حفاظت و حریم خصوصی داده‌ها، و ارزش‌ها و اصول مندرج در این توصیه‌نامه صورت بگیرند.

صد و سی و چهار - کشورهای عضو می‌توانند مکانیسم‌های زیر را برای نظارت و ارزیابی به کار ببرند: کمیسیون اخلاقی، رصدخانه اخلاق هوش مصنوعی، مخزنی شامل بر توسعه اخلاقی و منطبق بر حقوق بشر سیستم‌های هوش مصنوعی، کمک به ابتکارات موجود از طریق پرداختن به رعایت اصول اخلاقی در حوزه‌های صلاحیت یونسکو، مکانیسم اشتراک‌گذاری تجربیات، سندباکس‌های نظارتی هوش مصنوعی، و راهنمای ارزیابی برای عاملین هوش مصنوعی جهت سنجش میزان پایداری آنان به سیاست‌های توصیه شده در این سند.

### **(و) استفاده و بهره‌برداری از این توصیه‌نامه**

صد و سی و پنج - کشورهای عضو و همه ذینفعان مشخص شده در این توصیه‌نامه باید به ارزش‌ها، اصول و استانداردهای اخلاقی مرتبط با هوش مصنوعی که در این توصیه‌نامه ذکر شده اند احترام بگذارند و به حفظ و ترویج آنها بپردازند، و تمام اقدامات ممکن را برای اعمال سیاست‌های توصیه شده در این سند انجام دهند. صد و سی و شش - کشورهای عضو باید تلاش کنند تا اقدامات خود را در رابطه با این توصیه‌نامه از طریق همکاری با کلیه سازمان‌های دولتی و غیردولتی ملی و بین‌المللی مربوطه، و شرکت‌های فراملیتی و سازمان‌های علمی که فعالیت‌هایشان مشمول محدوده و اهداف این توصیه‌نامه می‌شود توسعه داده و تکمیل کنند. تنظیم یک متدولوژی توسط یونسکو جهت ارزیابی اثرات اخلاقی و همچنین ایجاد کمیسیون‌های ملی برای اخلاق هوش مصنوعی می‌توانند از جمله ابزارهای مهم این امر باشند.

### ز) ترویج توصیه‌نامه حاضر

صد و سی و هفت - یونسکو این وظیفه را دارد که آژانس اصلی سازمان ملل متحد برای ترویج و نشر این توصیه‌نامه باشد، و از این رو با سایر نهادهای مرتبط سازمان ملل، ضمن احترام به فرامین آنها و با اجتناب از اصل دوباره کاری، همکاری می‌کند.

صد و سی و هشت - یونسکو و نهادهای آن، از جمله کمیسیون جهانی اخلاق علم و تکنولوژی یونسکو (COMEST)، کمیته بین‌المللی اخلاق زیستی (IBC)<sup>۷</sup> و کمیته بین‌دولتی اخلاق زیستی (IGBC) با سایر سازمان‌های دولتی و غیردولتی بین‌المللی، منطقه‌ای و ناحیه‌ای همکاری می‌کنند.

صد و سی و نه - با این که در یونسکو وظیفه ترویج و حفاظت در حیطه اختیارات دولت‌ها و نهادهای بین‌دولتی است، جامعه مدنی از مدافعان اصلی منافع بخش دولتی به شمار می‌آید و بنابراین یونسکو می‌بایست مشروعیت آن را تضمین و ترویج کند.

### ح) شروط نهایی

صد و چهل - این توصیه‌نامه باید به عنوان یک کل در نظر گرفته شود، و ارزش‌ها و اصول بنیادین اجزاء مکمل و مرتبط با یکدیگر می‌باشند.

صد و چهل و یک - هیچ یک از بخش‌های این توصیه‌نامه نمی‌تواند جایگزین تعهدات یا حقوق کشورها بر اساس قوانین بین‌المللی باشد، یا آنها را تغییر دهد و یا باعث هر نوع پیش‌داوری در این زمینه باشد. همچنین هیچ بخشی از این توصیه‌نامه نباید به عنوان تاییدی برای هیچ کشور یا عامل، گروه یا فرد سیاسی، اقتصادی یا اجتماعی برای شرکت در هر فعالیتی یا انجام هر عملی که با حقوق بشر، آزادی‌های بنیادین، کرامت انسانی و دغدغه‌های مربوط به محیط زیست و اکوسیستم، زنده یا غیرزنده، مغایرت داشته باشد، تفسیر شود.