

## تحلیلی اخلاقی از رابطه احساسی انسان و روبات

سعیده بابایی<sup>۱</sup>، نرجس صابری<sup>۲</sup>، منیره بحرینی<sup>۳</sup>

<sup>۱</sup> محقق پسادکتری، مرکز اخلاق در علم و علوم انسانی، دانشگاه توبینگن، توبینگن، آلمان  
saeedeh.babaii@izew.uni-tuebingen.de

<sup>۲</sup> دانشجوی کارشناسی ارشد، دانشکده روانشناسی و علوم تربیتی، دانشگاه علامه طباطبایی، تهران  
n\_saberi@atu.ac.ir

<sup>۳</sup> دانشجوی کارشناسی ارشد، گروه فلسفه علم و فناوری، دانشگاه صنعتی امیر کبیر، تهران  
bahreini@aut.ac.ir

### چکیده

روابطی که بین انسان‌ها و روبات‌های اجتماعی شکل می‌گیرند، ویژگی‌های متفاوت قابل توجهی دارند که آنها را از روابط میان انسان‌ها با یکدیگر و با سایر تکنولوژی‌ها متمایز می‌کند. مطالعات مربوط به تعامل انسان/روبات، ذیل حوزه روان‌شناسی سایبری که یک شاخه‌ی علمی نوظهور است مورد بحث قرار می‌گیرد. تعاملات رو به گسترش انسان و روبات‌های اجتماعی می‌تواند منجر به شکل‌گیری نوعی پیوند احساسی میان کاربران و روبات‌ها شود. روبات‌های اجتماعی به دلیل ماهیت منحصر به فرد خود و انتظاراتی که شرکت‌ها و تبلیغات از آنها برای مردم ایجاد می‌کنند، این قابلیت را دارند که کاربر، سطحی از عاملیت و احساسات را به آنها نسبت و پیوند عاطفی یک‌سویه‌ای را با آنها شکل دهد. این امر می‌تواند مخاطرات اخلاقی گسترده‌ای برای کاربر و جوامع انسانی در پی داشته باشد. این مقاله به بررسی برخی چالش‌های اخلاقی در رابطه انسان و روبات‌های اجتماعی و تعدادی از راه‌حل‌های پیشنهادی آنها می‌پردازد.

**کلمات کلیدی:** تعامل انسان و روبات، روبات اجتماعی، روان‌شناسی سایبری، اخلاق روباتیک.

## ۱ مقدمه

ادعای گزافی نیست اگر هوش مصنوعی و تکنولوژی‌های مبتنی بر آن را از جمله داغ‌ترین و پرمناقشه‌ترین حوزه‌های تکنیکی جهان امروز تلقی کنیم که شاهد رشد سریع و بی‌سابقه آن هستیم. هر فرد در طول شبانه‌روز ساعت‌های زیادی را در تعامل و استفاده از ابزارهای هوشمند مصنوعی، همچون چت‌بات، لوازم خانگی هوشمند و روبات‌ها صرف می‌کند. روبات هوشمند بر خلاف تجهیزات مکانیکی غیر هوشمند که اغلب پیش از این، در صنعت کاربرد داشتند، موجودی مکانیکی و خودمختار<sup>۱</sup> است که در عین اینکه می‌تواند محیط

<sup>۱</sup>Autonomously

پیرامون خویش را تغییر دهد، می‌تواند خود را با محیط تطبیق دهد؛ چنین عملکردی نیازمند مجهز بودن روبات به هوش مصنوعی است (Friedenberg & Silverman 2005). در دهه‌ی گذشته، افزایش گسترده‌ای در کاربرد روبات‌های اجتماعی و خدماتی صورت گرفته است. در سال ۲۰۰۶ تعداد روبات‌های اجتماعی در قالب روبات‌های خدماتی و شخصی از تعداد روبات‌های صنعتی پیشی گرفته است که شامل روبات‌های شستشو مثل رومبا<sup>۲</sup> تا روبات‌های سرگرمی مثل پلیو<sup>۳</sup>، روبات‌های عروسک مثل بیبی الایو<sup>۴</sup>، روبات‌هایی مثل پارو<sup>۵</sup>، نائو<sup>۶</sup>، تگا<sup>۷</sup>، در درمان کودکان مبتلا به اتیسم یا درمان افسردگی و آموزش و خدمات عمومی می‌شوند (Schultz & Cominelli, Mazzei & Emilio De Rossi 2018).

روبات اجتماعی، روباتی در تعامل نزدیک با انسان است که علاوه بر آنکه مانند سایر روبات‌ها (روبات‌های صنعتی) رفتارهایی از پیش‌برنامه‌نویسی شده و یا آموخته‌شده دارد، می‌تواند اطلاعات معناداری از محیط اجتماعی دریافت نموده و بر اساس آن به کاربر واکنش نشان دهد. یک روبات اجتماعی می‌تواند به نحوی شخصی‌شده با کاربر ارتباط برقرار نماید. چنین روبات‌هایی اگر چه منافعی را برای انسان به همراه می‌آورند، خطرات و چالش‌هایی نیز به همراه دارند که مهم‌ترین این چالش‌ها تأثیرات روانی مخربی است که می‌تواند بر کاربر وارد نمایند. بنابراین بررسی تأثیرات روانی ناشی از تعامل انسان با روبات اهمیت زیادی پیدا می‌کند و گسترش گسترده‌ی تحقیقات در حوزه تعاملات انسان و روبات<sup>۸</sup> ذیل روان‌شناسی سایبری در دهه‌های اخیر گویای این مساله است. به هر میزان که این مطالعات سریع‌تر صورت گیرد می‌تواند اثربخشی بیشتری داشته باشد؛ زیرا قبل از عمومیت یافتن تکنولوژی‌ها، بهترین فرصت برای پرداختن به جوانب احتمالی استفاده از آن‌ها است.

یکی از مهم‌ترین چالش‌های اخلاقی تعامل انسان و روبات، وابستگی احساسی انسان نسبت به روبات و به صورت ویژه روبات اجتماعی است. اینگونه وابستگی به روبات‌های اجتماعی تفاوت بنیادینی با وابستگی به سایر تکنولوژی‌های تجربه‌شده توسط بشر دارند. روبات‌ها می‌توانند انسان‌ها را نسبت به خود از منظر عاطفی وابسته نمایند و همدلی آنان را نسبت به خود برانگیزانند در حالیکه این رابطه‌ی احساسی یک‌طرفه، و صرفاً از سوی کاربر است. این مساله می‌تواند عواقب امنیتی و احساسی برای کاربر در پی داشته باشد. یک پیشنهاد برای رفع این مشکل، القاء احساسات در روبات است. این مقاله به بررسی مساله رابطه انسان و روبات و برخی از پیشنهادات و سوالات اخلاقی پیرامون آن می‌پردازد.

<sup>2</sup>Roomba

<sup>3</sup>Pleo

<sup>4</sup>baby alive

<sup>5</sup>Paro

<sup>6</sup>Nao

<sup>7</sup>Tega

<sup>8</sup>Human Robot Interaction (HRI)

## ۲ ماهیت و کارکرد احساسات در انسان

احساسات<sup>۹</sup> پدیده‌ای روان‌شناختی، پیچیده، شخص‌محور و حاصل تجارب ذهنی حالات هیجانی<sup>۱۰</sup> انسان است. احساسات در عین اینکه ذهنی هستند، از آنجا که وابسته به هیجاناتی هستند که در بدن اتفاق می‌افتند جهان‌شمول و بین‌الذهانی نیز هستند. می‌توان گفت احساسات، تجربیات حسی<sup>۱۱</sup> هستند که در برابر تجربیات حسی پیشین، بررسی شده و دوباره برچسب‌گذاری می‌شوند؛ و از آنجا که هر فرد دارای احساسات قبلی متمایز و منحصر به فردی است که هنگام تفسیر و برچسب‌زدن احساسات خود از آنها استفاده می‌کند، بنابراین احساسات، شخصی، بیوگرافیک و پویا هستند و بر اساس تجربیات جدید ناشی از تغییر محیط و تغییر روابط می‌توانند تغییر کنند<sup>(۱۲)</sup>. بر خلاف دیدگاه‌های پیشین که تصور می‌شد احساسات و هیجان‌ها نقش در شناخت ندارند، مطالعات جدید نشان می‌دهد که احساسات تأثیر زیادی بر شناخت، رفتار و جهت‌گیری انسان دارند و فرایندهای شناختی مهمی مانند یادگیری، حافظه، توجه، تصمیم‌گیری و غیره نیازمند و متأثر از عواطف هستند (Bechara, Damasio & Damasio March 2000). کارکرد احساسات در انسان می‌تواند در سه سطح بررسی شود. اولین سطح مربوط به کارکردهای درون‌فردی احساسات است که در این سطح عواطف به عنوان یک سیستم پردازشی سریع، می‌تواند منجر به عمل با حداقل نیاز به تفکر و آگاهی، اقدام فوری در موقعیت خطر با انگیزه‌ی حفظ بقا و همچنین با اثرگذاری بر افکار و خاطرات در تصمیم‌گیری و شناخت موثر باشند. احساسات در این سطح می‌توانند بزرگترین موتور محرکه رفتار در انسان باشند و رفتارهای آینده‌ی فرد را جهت دهند. دومین سطح مربوط به کارکردهای بین‌فردی احساسات است که به نقش عواطف بین افراد اشاره دارد. در این سطح عواطف ادراک فرد را از رفتارها و موقعیت‌های بین‌فردی تسهیل می‌کند و انگیزه‌ی شکل‌گیری روابط بین فردی را ایجاد می‌نماید. سومین سطح مربوط به کارکردهای اجتماعی احساسات است که نقش احساسات را در شکل‌گیری ساختار اجتماعی و حفظ نظم اجتماعی در جامعه بررسی می‌کند (Matsumoto and Hwang 2019). به عنوان مثال همدلی به افراد اجازه می‌دهد تا روابط اجتماعی رضایت‌بخشی با سایر افراد ایجاد کنند و این روابط را حفظ نمایند. با این توضیحات می‌توان به نقش برجسته و کارکرد وسیع احساسات در انسان پی برد. حال این پرسش مطرح می‌شود که آیا انسان‌ها می‌توانند در رابطه با یک تکنولوژی هوشمند فاقد احساسات دچار حالات احساسی شوند؟

## ۳ نمونه‌هایی از رابطه احساسی یک‌طرفه انسان و روبات

چنانچه پیش از این اشاره شد، روبات اجتماعی روباتی در تعامل نزدیک با انسان است که همچون روبات‌های صنعتی رفتارهایی از پیش برنامه‌نویسی شده و یا آموخته شده دارد؛ اما علاوه بر آن می‌تواند اطلاعات معناداری از محیط دریافت کرده و بر اساس آن به رفتار کاربر واکنش نشان دهد. این روبات باید بتواند اطلاعات دریافتی

<sup>9</sup>Feelings

<sup>10</sup>Emotion

<sup>11</sup>Sensation

<sup>12</sup>Shouse 2005

از محیط را درونی و در سطحی بالاتر درباره‌ی آنها استدلال نماید و نظرات خود را به نحو مستقل بسازد، سپس به طور خودکار و با توجه به تجربه‌ی منحصر به فرد خود تصمیم‌گیری کند (Cominelli, Mazzei & Emilio De Rossi 2018). این توانایی روبات‌های هوشمند اجتماعی در تطبیق رفتارهایشان با محیط و یادگیری از تجربه و همچنین ارتباط نزدیک و شخصی با کاربر می‌تواند عواقب مهمی به همراه داشته باشد؛ برای مثال، می‌تواند انسان‌ها را به سمتی سوق دهد که برای درک رفتار روبات‌ها به آن‌ها اراده، قصدمندی، خودمختاری و مهمتر از آن، احساسات نسبت دهند و از این جهت، آنها را متمایز از سایر تکنولوژی‌های هوشمند و یا غیر هوشمند که تاکنون تجربه کرده‌اند ادراک نمایند. شواهد زیادی وجود دارد که شکل‌گیری نوعی رابطه احساسی عمیق با روبات‌ها را از سوی کاربران تأیید می‌کند. به عنوان مثال نظامیان در مواجهه با روبات‌های مین‌یاب رفتارهایی که نشان از شکل‌گیری نوعی تعلق عاطفی است از خود نشان می‌دهند. روبات خنثی‌کننده‌ی مین روباتی است که با راه رفتن روی مین‌ها، هر بار یک پا از چندین پای خود را از دست می‌دهد و از این طریق مین‌ها را خنثی می‌کند. اظهارات تکنسین‌های این روبات‌ها، که گاهی برای روبات‌هایشان اسم نیز می‌گذارند، حاکی از وابستگی احساسی آنها به روبات‌هاست. آنها تمایلی ندارند روبات دیگری را جایگزین روباتشان کنند و از روبات‌ها به مثابه قهرمان و موجوداتی دارای احساس و گاهی مانند عضوی از خانواده خود یاد می‌کنند. تبیینی که می‌توان برای اینگونه از رفتارها ارائه کرد این است که انسان نوعی عاملیت را روی روبات فرافکنی می‌کند و به او حیات و احساسات نسبت می‌دهد. (Scheutz 2012) نمونه‌ی دیگر در تأیید این مساله، رومبا است که یک جاروبرقی هوشمند است و دیده شده که با گذشت زمان کاربران حس قدردانی زیادی را به خاطر تمیزکردن خانه‌شان به رومبا پیدا می‌کنند. ایشان همچنین از شرم اینکه توسط رومبا دیده نشوند، در هنگام تعویض لباس، در اتاق خود را می‌بندند. در حالی که رومبا حتی نمی‌داند دارای مالک است و با مالکان خود به مثابه موانعی مثل صندلی و میز برخورد می‌کند، مالکان رومبا مایلند که به قصد قدردانی از او کار خوبی برایش انجام دهند؛ برخی از کاربران بخشی از کار نظافت را انجام می‌دهند تا رومبا کمی استراحت کند یا او را با خود به سفر می‌برند (Sung 2007). کاربران در حالی رابطه احساسی یک‌طرفه‌ای را با رومبا شکل می‌دهند که او هیچ درکی از آن ندارد. به عنوان یک نمونه دیگر، مالکان آیبو که یک سگ روباتیک مبتنی بر هوش مصنوعی است، تمایل زیادی دارند که باورهای (اشتباهی) درباره‌ی حالات ذهنی او داشته باشند. در مطالعه‌ای روی مالکان روبات‌های حیوان خانگی، چهار دسته ویژگی ماهیت، عاملیت، هویت اجتماعی و هویت اخلاقی از سوی مالکان به آیبو نسبت داده شده است (Kahn, Friedman and Hagman 2002) (Scheutz 2012). در سال ۲۰۱۵، یک معبد بودایی در ژاپن که به طور معمول برای حیوانات خانگی مرده مراسم برگزار می‌کند، مراسم مشابهی را برای یک روبات آیبو برگزار کرد. برگزاری چنین مراسمی برای یک روبات نشان‌دهنده‌ی وابستگی قابل توجه بین کاربران و روبات‌هایشان است (Tony J. Prescott 2021).

بر خلاف این نمونه‌ها که در آن کاربران غیر متخصص به شمار می‌آیند، به نظر می‌رسد سازندگان روبات‌ها هم از این مساله مستثنی نیستند. سینتیا بریزیل<sup>۱۳</sup> سازنده‌ی کیسمت<sup>۱۴</sup> که اولین روبات هوشمند اجتماعی

<sup>13</sup>Cynthia Breazeal<sup>14</sup>Kismet

است نیز رابطه‌ای شخصی را با روبات خویش شکل داده بود؛ پیوندی مادرانه که روبات را فقط یک ماشین نمی‌دید و وقتی پروژه‌اش به پایان رسید و مجبور به ترک روبات شد، فقدانی شدید را تجربه کرد. همچنین دانشجویان ام.آی.تی که در آزمایشگاه با روبات کیسنت کار می‌کردند، گاهی پرده‌ای را میان خود و روبات می‌کشیدند، چون خیره شدن روبات باعث برهم خوردن تمرکز آن‌ها می‌شد (Scheutz 2012).

## ۴ مخاطرات رابطه احساسی یک طرفه‌ی انسان با روبات

تجربه‌های یادشده از وابستگی‌های عاطفی یک طرفه انسان و روبات این مساله را روشن می‌نماید که صرف تقلید و شبیه‌سازی برخی از رفتارهای انسان یا حیوان واقعی، می‌تواند روبات را آماده‌ی ساختن رابطه‌ای عاطفی یک‌سویه از سمت کاربر نموده و از او یک همدم بسازد. فرافکنی ذهن‌مندی و عاملیت به روبات و تشخیص بخشی به روبات از سوی کاربر، می‌تواند منجر به نسبت‌دادن ویژگی‌های زیست‌شناختی انسانی به روبات و تجربه‌ی احساساتی نظیر دلسوزی، مسئولیت‌پذیری، شرم، عذاب وجدان، قدردانی، مراقبت و حفاظت از روبات‌ها شود. علاوه بر وجود این فرایند احساسی ناآگاهانه از سوی کاربران انسانی، شرکت‌های سازنده‌ی روبات‌های اجتماعی با وجود آگاهی از عدم توانایی روبات‌های خود در درک و بروز احساسات، در راستای مقاصد تبلیغاتی سعی می‌کنند این تصور را در کاربران خود به وجود بیاورند که روبات‌ها احساسات آن‌ها را درک می‌کنند و یا احساسات واقعی از خود بروز می‌دهند؛ در حالیکه به این امر آگاهند که روبات‌ها تنها نمایشی صوری از احساسات بروز می‌دهند و این به معنی فریب کاربر است. پژوهشگران دانشگاهی نیز در ارائه پژوهش‌های خود به قدر کافی مراقب نیستند و از تعبیری مانند «احساس داشتن» برای توصیف ساخته‌های خود استفاده می‌کنند و استفاده از این ادبیات در نوشتن مقالات و گزارش‌های علمی موجب می‌شود افراد غیر متخصص فرایندهای کنترل را در این مصنوعات با فرایندهای کنترل در ارگانیسم‌های طبیعی مثل حیوانات و انسان‌ها یکسان بیانگرند و در حالی که در سطح آگاهانه این ادعاها را ممکن است رد کنند، باورهای غلطی را در سطح عمیق ناخودآگاه نسبت به امکان برقراری رابطه عاطفی با روبات‌ها پیدا می‌کنند (Scheutz 2012).

رابطه عاطفی یک طرفه با تکنولوژی‌های هوشمند و یا تصور امکان برقراری رابطه عاطفی با تکنولوژی‌هایی که از چنین امکانی برخوردار نیستند می‌تواند عواقب جدی برای جوامع انسانی داشته باشد؛ چرا که امکان سوء استفاده‌های اقتصادی، سیاسی و امنیتی در مقیاس بزرگ را از سوی روبات‌ها و یا شرکت‌های سازنده‌ی آنها فراهم می‌کند. نمایش صوری احساسات از سوی روبات‌ها و رفتار سطحی آنها می‌تواند آسیب جسمی و روانی کاربر را به دنبال داشته باشد. همچنین اعتیاد ناآگاهانه به بات به دلیل همانندسازی رابطه انسانی با آنها می‌تواند اثرات مخربی بر تعاملات انسانی داشته باشد (Scheutz 2012). ایجاد رابطه تعاملی با روبات‌ها می‌تواند به توانایی و تمایل ما برای معاشرت با سایر انسان‌ها آسیب وارد نماید. تعامل با یک روبات اجتماعی (همدم) می‌تواند با تضعیف ظرفیت ما برای دلبستگی ایمن، تشویق رابطه با روبات به دلیل سهولت و ماهیت غیر چالش‌برانگیز همراهی مصنوعی در مقابل رابطه با یک انسان و با پرکردن زمان و ظرفیت کاربر، به روابط انسانی آسیب وارد نماید (Tony J. Prescott 2021). بنابراین می‌توان گفت علی‌رغم همه‌ی کاربردهایی مفیدی که روبات‌های اجتماعی می‌توانند برای کاربران داشته باشند، چالشی جدی ناشی از شکل‌گیری رابطه

احساسی یک طرفه با روبات وجود دارد. پیشنهاداتی برای پیشگیری از اثرات مخرب این مساله مطرح شده است. این پیشنهادات می‌توانند موردی باشند و یا اینکه همه‌ی انواع روبات‌های اجتماعی را شامل شوند. بسته به نوع و کاربرد روبات می‌توان این پیشنهادات را بررسی کرد.

## ۵ چند پاسخ احتمالی به مخاطرات رابطه یک طرفه انسان با روبات

با توجه به فواید و کاربردهایی که تکنولوژی‌های هوشمند اجتماعی می‌توانند داشته باشند، مانند درمان‌های توانبخشی، تشویق به رفتارهای مثبت در راستای سلامت و بهزیستی، مراقبت از سالمندان و آموزش‌های شخصی و غیره، پیشنهادهایی مثل توقف تولید روبات‌های اجتماعی منطقی به نظر نمی‌رسد؛ اما مخاطرات رابطه یک طرفه‌ی انسان و روبات به قدری جدی و همه‌گیر است که نمی‌توان از آن نیز صرف نظر کرد.

یکی از پیشنهادهای ممکن برای پیشگیری از آسیب‌های رابطه احساسی یک طرفه‌ی روبات و انسان، مجهز کردن روبات به قدرت استدلال اخلاقی در قالب الگوریتم‌های تصمیم‌گیری است که مانع از امکان سوء استفاده از کاربر بشود؛ اما باید دانست که تعداد کنش‌های قابل پیش‌بینی در قالب الگوریتم‌های تصمیم‌گیری محدود است در حالی که دلالت‌های اخلاقی کنش‌ها و موقعیت‌های محتمل، بسیار پیچیده و پرشمارند و همه‌ی تلاش‌ها در طول تاریخ فلسفه نتوانسته است روی مجموعه‌ای کامل از قواعد اخلاقی توافق کند. علاوه بر این، ترجمه قواعد اخلاقی به زبان محاسباتی نیز ممکن است با محدودیت‌های عملی مواجه باشد. حتی اگر این مجموعه مشخص از قواعد اخلاقی (مثل قواعد آسیموف) قابل تبدیل شدن به کد باشند، به این خاطر که اولویت‌بندی قواعد و حل تناقض میان قواعد، وابسته به موقعیت است و موقعیت‌مندی مساله‌ای بسیار دشوار است، می‌توان گفت این سیستم در عمل کار نمی‌کند (Scheutz 2012).

علیرغم اهمیت شباهت ظاهری روبات با انسان در روبات‌های انسان‌نما، ممکن است در ابتدا تصور کنیم که صرف شباهت ظاهری ماشین به انسان است که موجب می‌شود خودمختار نبودن روبات را فراموش کنیم و به آن احساسات نسبت بدهیم؛ اما روباتی مثل رومبا که شباهتی به انسان یا حیوان ندارد نیز ایده‌ی عاملیت‌داشتن را به کاربرش القاء کرده و حتی می‌تواند حس قدردانی و دلسوزی را در او فعال کند. با این حال به عنوان راه حلی دیگر می‌توان این پیشنهاد را مطرح کرد که طراحی روبات به گونه‌ای باشد که دائماً به کاربر این اخطار را بدهد که یک ماشین است، احساسات ندارد و نمی‌تواند رابطه‌ای متقابل داشته باشد. هر چند ممکن است این اخطار، احتمال و شدت پیوندهای عاطفی با روبات را کاهش دهد اما همانند پیام‌های اخطار روی پاکت سیگار ضمانتی وجود ندارد که انسان‌ها درگیر رابطه با روبات نشوند (Scheutz 2012).

دیدگاه دیگر این است که برای حل چالش رابطه یک‌سویه‌ی انسان و روبات، روبات‌ها را به احساساتی مشابه انسان مجهز کنیم. بدین ترتیب روبات‌ها نیز می‌توانند با داشتن توانایی احساسی مشابه انسان، در جوامع انسانی شبیه انسان‌ها عمل کنند و در صورت برقراری رابطه بین انسان و روبات، رابطه‌ای دو سویه برقرار شود. عملی کردن این پیشنهاد در عین اینکه می‌تواند آورده‌هایی برای کاربر داشته باشد، چالش‌های جدیدی را پیش روی ما قرار می‌دهد.

## ۶ رابطه احساسی دوطرفه انسان و روبات

با هوشمندتر شدن هوش مصنوعی و تعاملات رو به گسترش انسان‌ها با تکنولوژی‌های هوشمند از جمله روبات‌های اجتماعی، این دیدگاه مطرح شد که از طریق القاء احساسات به روبات‌ها علاوه بر آن که می‌توان آن‌ها را برای کاربر ایمن‌تر، جذاب‌تر و کارا تر کرد، این امید وجود دارد که از این طریق بتوانیم روبات‌های هوشمند اجتماعی را اخلاقی نیز بکنیم (Lagrandeur 2015). در ادامه به بررسی اثرات مثبت و چالش‌های اخلاقی این دیدگاه می‌پردازیم.

تسهیل تعامل انسان و روبات یکی از مهمترین اثرات مثبت القاء احساسات در هوش مصنوعی و روبات‌هاست. روبات‌های اجتماعی برای به‌عهده گرفتن نقش‌های اجتماعی خود نیاز به توانایی درک اجتماعی دارند. درک اجتماعی نیازمند حساسیت، یا توانایی شناخت هیجانات و احساسات در دیگران و برقراری رابطه همدلانه با آنان است (Castellano, et al. 2012). به عبارتی روبات اجتماعی برای اجتماعی بودن، نیازمند توانایی تفسیر و پاسخ به احساسات انسان‌ها است؛ به عنوان مثال یک روبات حساس<sup>۱۵</sup>، به احتمال بیشتری با کاربر انسانی درگیر می‌شوند و مدت طولانی‌تری با او تعامل قابل قبول و مطلوب برقرار می‌کند؛ به این معنی که سعی می‌کند اطمینان حاصل کند که کاربر به حفظ تعامل علاقه‌مند است و در صورت ناراحت بودن یا عدم تمایل به تعامل از سوی کاربر، با او همدلانه رفتار می‌کند. چنین روباتی با تلاش برای درگیر کردن فرد در فعالیت‌ها، مزاحمتی برای کاربر ایجاد نمی‌کند. به عبارتی درک اجتماعی و توانایی شناخت هیجانات کاربر از سوی روبات، احتمال مزاحمت کمتر برای کاربر و تعامل باورپذیر و بلندمدت‌تر و در نهایت جذابیت روبات برای کاربر را در پی دارد. برخورداری روبات از احساسات قابل درک برای انسان، آگاهی از نحوه تأثیرپذیری هوش مصنوعی از محیط را برای انسان در پی دارد. این مساله باعث پیش‌بینی رفتارهای روبات و دستیابی به نحوه‌ای از مکانیسم کنترل می‌شود که در نهایت امنیت را برای کاربران به ارمغان می‌آورد.

## ۷ جای‌دهی احساسات در هوش مصنوعی و چالش‌های آن

بیشتر ایده‌های موجود درباره‌ی ایجاد احساسات در هوش مصنوعی، دور از دسترس و مربوط به آینده‌ی خیلی دور است که ممکن است هیچگاه محقق نشود. به نظر می‌رسد تلاش برای رسیدن روبات به احساسات کاملاً شبیه انسان و تحقق هوش مصنوعی قوی دارای احساسات از نظر فنی و از نظر فلسفی ممکن نباشد؛ اما در صورت امکان با چالش‌های فلسفی و اخلاقی زیادی روبروست.

مساله اول این است که مبنای اخلاق صحیح چیست؟ وقتی حرف از ارزش‌ها یا مقولاتی مثل خیرخواهی یا وجدان می‌زنیم کدام مکتب اخلاقی مد نظر است؟ اخلاق کانتی؟ وظیفه‌گرایی (قاعده صرف)؟ سودگرایی (نفع بزرگ جمع)؟ بودیسم (ایثار و شفقت)؟ (Lagrandeur 2015) مساله بعد به این موضوع می‌پردازد که این مسیر آزمایش‌ها و توسعه‌ی ایجاد روبات‌های هوشمند اجتماعی درست نیست؛ به این دلیل که تا قبل از مجهز کردن روبات‌ها به احساسات، دلبستگی عاطفی یک‌طرفه بین انسان و روبات وجود دارد و این مساله

<sup>15</sup>Affect-sensitive robot

می‌تواند انسان را در مقابل برخی مخاطرات آسیب پذیر کند (Lagrandeur 2015). برای تولید احساسات مصنوعی مبتنی بر احساس گناه و همدلی، باید موجودی تولید کرد که رنج دیده بوده و قادر به رنج کشیدن باشد تا بتواند رنج دیگران را درک کند. مساله سوم به سوال می‌پردازد که آیا تولید موجودی رنجور اخلاقی است؟ آیا ما حق داریم که موجودی را خلق کنیم که آسیب پذیر است و ممکن است مورد سوء استفاده قرار گیرد؟ در هر صورت، این پیشنهاد را نمی‌توان و لازم نیست در هر نوعی از روبات اجتماعی عملی کرد، بلکه بر حسب نیاز و کاربرد می‌توان از این قابلیت استفاده کرد (Scheutz 2012). مساله دیگر به این سوال توجه دارد که آیا این گونه اقدامات منجر به تقلیل احساسات انسانی به یکسری الگوهای تقلیدی و رفتارهای سطحی خشک نیست؟ آیا صرف تقلید رفتار از جایگاه و ارزش احساسات پیچیده‌ی انسانی مخصوصاً نوع غیر کلامی‌اش نمی‌کاهد؟ و مساله آخر این است که آیا این شبیه‌سازی منحصربه‌فرد بودن ارتباطات انسانی را تهدید می‌کند؟ در جواب باید گفت، بله ممکن است (Lagrandeur 2015). علاوه بر ایده‌ی دستیابی به هوش مصنوعی قوی مبتنی بر احساسات، راهکارها و ایده‌های کوتاه‌مدت برای نزدیک‌تر شدن هوش مصنوعی به احساسات انسانی نیز وجود دارد که هر کدام مزیت‌ها و کاربردهایی دارند ولی نقدهای اخلاقی جدی به هر یک از آنها نیز وارد است.

یک راهکار کوتاه‌مدت، نرم‌افزاری است برای طراحی یک «وجدان اخلاقی»<sup>۱۶</sup> که بتواند مثل یک فرمانده اخلاقی برای سیستم‌های هوشمند خودمختار عمل کند (Arkin 2007). این تجهیز اخلاقی بیشتر روی «احساس گناه» متمرکز است و سعی می‌کند با کدنویسی به نحوی گناه روباتیک را ایجاد کند؛ به این معنی که این کدنویسی منجر می‌شود که خود روبات با خودپایشی برخی پارامترهای قابل اندازه‌گیری، به حس گناه برسد. در این ایده، اگر اندازه‌ی فعلی حالت عاطفی گناه از یک مقدار از پیش تعیین شده‌ی حد آستانه بیشتر شود، نشانه‌ی شکسته شدن اخلاق است و روبات به نحو اتوماتیک غیر فعال می‌شود. سه نقد به این راهکار وارد است؛ اول اینکه بدون تجربه قبلی، تشخیص حد آستانه برای روبات ممکن نیست. به نظر می‌رسد طراحی چنین عملکرد عاطفی‌ای تنها بعد از وقوع یک فاجعه انسانی میسر می‌شود. دوم اینکه این احساسات روباتی تنها یک شبیه‌سازی مبهم از نوع انسانی‌اش هستند، زیرا مبتنی بر خود و پویا نبوده و در حد یک سیستم عیب‌یابی کامپیوتری عمل می‌کنند که در حال حاضر هم وجود دارد. سوم اینکه این احساساتی که قرار است در طراحی روبات‌ها مورد استفاده قرار گیرند، یک محاسبات بی‌روح و سرد از یک سری معیار ارزیابی هستند و شامل همدلی، همدردی و حتی ترحم نمی‌شوند. بنابراین نمی‌توان نام چنین کاری را محاسبات عاطفی یا احساسات بنیانی گذاشت؛ به این دلیل که نه کدگذاری آنها به اندازه‌ی کافی پیچیده است و نه هوش مصنوعی هنوز به اندازه کافی پیشرفته است (Lagrandeur 2015).

راهکار کوتاه‌مدت دیگر این است که ماشین را واجد نوعی «خیرخواهی»<sup>۱۷</sup> کنیم؛ به این صورت که از رابطه مادر و کودک الهام بگیریم و هنگامی که روبات برای اولین بار کاربر انسانی خود را دید، با او پیوند اجتماعی خودکار برقرار کند و این روند هر بار که انسان با روبات چیزی را به اشتراک می‌گذارد، تقویت شود؛ اما در نقد این ایده می‌توان گفت در واقع «خیرخواهی» فقط یک رفتار نیست. ممکن است ظهور و بروز

<sup>16</sup>ethical conscience<sup>17</sup>benevolence



خیرخواهی در رفتار نمایان شود ولی خود خیرخواهی در واقع یک موضع اخلاقی پیچیده است؛ یک تصمیم آگاهانه بر اساس مجموعه‌ای از احساسات و تجربیات و دلایل است که منجر به انجام عمل به نفع دیگری می‌گردد. بنابراین چیزی بیشتر از یک سیستم پاداش‌دهی ساده است (Lagrandeur 2015). تلاش برای احساسی کردن روبات‌ها اگرچه راهکار مناسبی به نظر می‌رسد ولی در عمل با چالش‌های فلسفی و اخلاقی و عملی زیادی روبروست. با این حال تلاش در جهت جای‌دهی احساسات در روبات پروژه‌ای در جریان است که می‌تواند در آینده ایده‌ها و نمونه‌های جدیدی را به ظهور برساند.

## ۸ نتیجه‌گیری

در مقاله حاضر کوشیدیم تا تحلیلی از رابطه احساسی یک‌طرفه میان انسان و روبات‌های اجتماعی ارائه دهیم و به مثال‌هایی اشاره کنیم که شکل‌گیری چنین روابط احساسی‌ای را تصدیق می‌کنند. این رابطه احساسی یک‌سویه می‌تواند مخاطراتی را برای کاربران انسانی ایجاد کرده و امکان سوء استفاده سیاسی، اقتصادی و امنیتی از آن‌ها را فراهم کند. در حالی که خود روبات‌ها در حال حاضر توانایی درک احساسات و همدلی با کاربر را ندارند و صرفاً توهم داشتن ویژگی‌های شبه انسانی را در کاربر ایجاد می‌کنند. در همین رابطه، تا حدی به این موضوع پرداختیم که راه‌حل‌های ممکن برای پیشگیری یا کمینه کردن چنین مخاطراتی در تعامل انسان/روبات چه می‌تواند باشد. برای مثال، اخلاقی کردن تصمیم‌های روبات‌ها و به نوعی، جای‌دهی کردن اخلاقیات در طراحی روبات‌ها تا چه میزان امکان‌پذیر است و آیا می‌تواند مانع سوء استفاده‌های غیر اخلاقی از کاربر شود. چه راهکارهایی را می‌توان تدبیر کرد تا کاربر کمتر دچار انسان‌پنداری و نسبت‌دادن ویژگی‌های انسانی به روبات شده و فاصله احساسی خود از آن را کنترل کند.

بررسی کردیم که آیا راه مقابله با مخاطرات این رابطه یک‌سویه این است که در روبات‌ها احساسات و توانایی شناخت هیجانات و همدلی را شبیه‌سازی کنیم یا خیر. اما این راهکار خود با چالش‌هایی فلسفی و اخلاقی روبروست از جمله اینکه آیا ماهیت احساسات یا فضایی نظیر خیرخواهی به گونه‌ای است که شبیه‌سازی آن امکان‌پذیر باشد؟ آیا ساختن موجودی که توانایی رنج کشیدن و آسیب دیدن را داشته باشد، امری اخلاقی است؟ آیا ساختن چنین موجوداتی منجر به تهدید روابط انسانی نخواهد شد؟ پرسش‌هایی از این دست، همچنان پرسش‌های بازی هستند که نیازمند تأملات بین‌رشته‌ای میان علوم انسانی و علوم فنی است و هر علمی از زاویه نگاه خود می‌تواند تحلیلی از مسئله ارائه دهد؛ لیکن فارغ از اینکه در عمل طراحی روبات‌های دارای توانایی‌های احساسی ممکن باشد یا خیر، نیازمند تأملات اخلاقی عمیق و آینده‌نگرانه‌ای هستیم که عواقب اخلاقی تلاش در جهت توسعه این روبات‌ها را بر ما نمایان‌تر کند و ما را برای روبرویی با آینده‌های ممکن مختلف آماده کند.

## مراجع

- [1] Arkin, R. 2007. "Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture." Technical report, Georgia Institute of Technology.

- [2] Bechara, Antoine , Hanna Damasio, & Antonio R. Damasio. March 2000. "Emotion, Decision Making and the Orbitofrontal Cortex." *Cerebral Cortex*, Volume 10, Issue 3, 295–307.
- [3] Castellano, Ginevra, Iolanda Leite, Ana Paiva, & Peter W. McOwan. 2012. "Affective teaching: learning more effectively from empathic robot." *Awareness magazine: Self-Awareness in Autonomic Systems* doi: 10.2417/3201112.003948.
- [4] Cominelli, Lorenzo , Daniele Mazzei, & Danilo Emilio De Rossi. 2018. "SEAI: Social Emotional Artificial Intelligence Based on Damasio's Theory of Mind." *Front. Robot. AI*.
- [5] Friedenberg, Jay, & Gordon Silverman. 2005. *Cognitive Science: An Introduction to the Study of Mind*. California, USA: SAGE Publications.
- [6] Kahn, P. H. 1999. *The human relationship with nature: Development and culture*. MIT Press.
- [7] Lagrandeur, Kevin. 2015. "Emotion Artificial Intelligence and Ethics."
- [8] Matsumoto, David, & Hyisung Hwang. 2019. "Functions of Emotions." In: *Introduction to Psychology*, By: Jordan Cummings & Lee Sanders. University of Saskatchewan.
- [9] Peter H. Kahn, J., Friedman, B., & Hagman, J. 2002. "I care about him as a pal: Conceptions of robotic pets in online aibo discussion forums." *CHI EA '02: CHI '02 Extended Abstracts on Human Factors in Computing Systems*, 632–633.
- [10] Scheutz, Matthias. 2012. "The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots." *Robot Ethics: The Ethical and Social Implications of Robotics* (MIT Press), 205-221.
- [11] Shouse, Eric. 2005. "Feeling, Emotion, Affect." *M/C Journal* 8.6.
- [12] Sung, JY., Guo, L., Grinter, R.E., & Christensen, H.I. 2007. "My Roomba Is Rambo": Intimate Home Appliances". *UbiComp 2007: Ubiquitous Computing*. Berlin: Springer, 145-162.
- [13] Tony J. Prescott, & Julie M. Robillard. 2021. "Are friends electric? The benefits and risks of human-robot relationships." *iScience* Volume 24, Issue 1.