

آیا می‌توان هوش مصنوعی را زنده و هوشیار دانست؟ بررسی مرزهای آگاهی انسان و ماشین

شنتیا رشیدی

2024/8/11

این مقاله به بررسی مفهوم زنده بودن و آگاهی در هوش مصنوعی می‌پردازد، و تلاش می‌کند با تحلیل فلسفی و زیست‌شناختی، مرزهای میان انسان و ماشین را روشن سازد

آیا می‌توانیم هوش مصنوعی را زنده و هوشیار خطاب کنیم؟

از بدو شروع دوران کامپیوتر و هوش مصنوعی، همیشه مرزی بین انسان و ماشین وجود داشته است. اما اکنون که این فاصله روز به روز در حال کوتاه‌تر شدن است، این سؤال مطرح می‌شود که چگونه می‌توانیم بگوییم هوش مصنوعی زنده است؟

آیا اگر یک ماشین احساسات انسانی از خود نشان دهد، می‌توانیم بگوییم آن برنامه زنده است و این احساسات را تجربه می‌کند؟

آیا اگر یک ماشین درد را نشان دهد و حتی تلاش کند از آن فرار کند، واقعاً درد را حس می‌کند یا صرفاً برنامه‌ریزی شده تا چنین واکنشی نشان دهد؟

این پرسش‌ها و پرسش‌های مشابه، انگیزه‌ی شروع این مقاله و تحقیق من در مورد مفهوم زنده بودن در انسان و ماشین بوده است.

تعریف زنده بودن: معیارهای ما برای زنده بودن یک موجود چیست؟

قصد دارم مفهوم زنده بودن را از دو دیدگاه فلسفه و زیست‌شناختی بررسی کنم. کامل‌ترین تعریف برای «زنده بودن» توسط ناسا ارائه شده است:

"زندگی سیستمی خودپایدار است که قادر به تکامل داروینی باشد."

قابلیت رشد و تولید مثل: موجود زنده باید بتواند رشد کند و تولیدمثل داشته باشد؛ به طوری که بتواند نسل‌های جدیدی از خود به وجود آورد.

استفاده از انرژی: موجود زنده نیاز به منبع انرژی دارد و می‌تواند انرژی را برای فرآیندهای حیاتی‌اش به کار ببرد، مانند متابولیسم.

سازگاری با محیط و تکامل: زندگی باید بتواند خود را با محیط وفق دهد و با گذشت زمان از طریق تکامل بهبود یابد.

سازمان‌یافتگی: موجود زنده دارای ساختارهای منظم و پیچیده‌ای است که در سطح مولکولی تا سلولی سازمان‌دهی شده‌اند.

شاید همین تعریف‌ها کافی باشد تا یک موجود هوش مصنوعی را از لیست زنده‌ها خارج کنیم. اما یک دیدگاه دیگر نیز مطرح است: فلسفه.

برخی فیلسوفان، زنده بودن را در آگاهی و احساس داشتن معنا می‌کنند. یک موجود زنده، تجربه‌های زندگی مانند عشق، درد و لذت را از دیدگاه اول‌شخص تجربه می‌کند. آیا وقتی یک ربات این احساسات را نشان می‌دهد، به طور اول‌شخص نیز آنها را درک می‌کند؟

به شخصه، تفاوتی بین صرفاً نفس کشیدن و زنده بودن قائلم. من زنده بودن را در احساس داشتن معنا می‌کنم.

نظریه کوالیا چیست و آیا ماشین می‌تواند آن را تجربه کند؟

نظریه کوالیا در فلسفه ذهن به کیفیت‌های ذهنی و تجربیات درونی شخص اشاره دارد که به صورت شخصی و اول‌شخص تجربه می‌شود و به‌سختی می‌توان آن را به دیگری انتقال داد. کوالیا شامل تمام احساسات، ادراکات و تجربیات ذهنی است که هر فرد در دیدگاه اول‌شخص خود آن را تجربه می‌کند.

مثال‌هایی از کوالیا:

درد: اگر کسی بگوید دردی را احساس می‌کند، هیچ‌کس دیگر نمی‌تواند دقیقاً تجربه او را احساس کند؛ و درد آن شخص، منحصر به خود او است.

رنگ‌ها: تجربه دیدن رنگ‌ها برای هر شخص متفاوت است. وقتی دو نفر رنگ سبز را می‌بینند، ممکن است هر دو آن را سبز بنامند، اما هرکدام تجربه‌ی متفاوتی از دیدن رنگ سبز دارند.

آیا ماشین می‌تواند کوالیا را تجربه کند؟

حتی اگر هوش مصنوعی بتواند احساسات را مانند انسان شبیه‌سازی کند، آن درک و احساس منحصر به فرد آن ماشین نیست و اگر دو ماشین مختلف با کدهای مشابه در موقعیت‌های مشابه قرار گیرند، تجربه و واکنش یکسانی خواهند داشت و آن نمایش احساس به صورت کوالیا نخواهد بود.

حال یک سؤال مطرح می‌شود: اگر ماشین بتواند احساس را نشان دهد، آیا می‌توانیم بگوییم که آن را تجربه می‌کند یا صرفاً نمایش احساس است؟ برای پاسخ به این سؤال به نظریه «آرگومان اتاق چینی» می‌پردازیم.

آرگومان اتاق چینی چیست و چه ارتباطی با هوشیار بودن ماشین دارد؟

آرگومان اتاق چینی در سال 1980 توسط فیلسوف آمریکایی جان سرل مطرح شد. هدف از این آزمایش نقد عملکردگرایی در هوش مصنوعی بود.

جان سرل می‌خواهد بگوید که حتی اگر یک ماشین از نظر بیرونی رفتار هوشمندانه‌ای از خود نشان دهد، این به معنای داشتن آگاهی و درک نیست.

در آزمایش اتاق چینی، سرل یک انسان را بدون هیچ دانشی از زبان چینی تصور می‌کند که در اتاقی با دستورالعمل‌های دقیق به زبان چینی حضور دارد. این شخص راهنمایی برای کار با زبان چینی دارد و می‌تواند به سؤالاتی که به زبان چینی از او پرسیده می‌شود، با استفاده از این دستورالعمل‌ها پاسخ دهد، بدون اینکه هیچ درکی از کلمات و مفاهیم داشته باشد.

از دیدگاه شخص خارجی، فرد داخل اتاق تسلط کامل به زبان چینی دارد زیرا به درستی پاسخ می‌دهد، اما واقعیت چیز دیگری است.

جان سرل، کامپیوتر را مانند فرد در اتاق چینی تصور می‌کند. او معتقد است که کامپیوتر صرفاً از دستورات پیروی می‌کند و درکی از مفاهیم ندارد.

اما نقدی بر این آرگومان وارد است که شاید فرد به تنهایی درکی از زبان چینی نداشته باشد، اما کل سیستم فرد و اتاق می‌توانند نشان‌دهنده معنای واقعی فهم باشند.

نقطه مقابل آرگومان اتاق چینی: آزمون تورینگ

آزمون تورینگ توسط دانشمند و ریاضیدان بریتانیایی، آلن تورینگ، در سال 1950 مطرح شد. هدف اصلی این آزمون، سنجش توانایی یک ماشین در نشان دادن رفتار شبیه انسان است. به بیان دیگر، این تست، پاسخ تورینگ به سؤال "آیا ماشین می‌تواند فکر کند؟" است.

در آزمون تورینگ، سه شرکت‌کننده وجود دارند: انسان، ربات و پرسشگر. پرسشگر در اتاقی جداگانه قرار دارد و از طریق ارتباط نوشتاری با هر دو نفر مکالمه می‌کند، اما نمی‌داند کدام یک انسان است و کدام یک ماشین. وظیفه پرسشگر این است که از طریق پرسش‌های مختلف، مشخص کند ربات کدام است. اگر ماشین بتواند پرسشگر را فریب دهد، از آزمون با موفقیت بیرون آمده است.

هدف از آزمون تورینگ این است که به جای پرداختن به این موضوع که آیا ماشین‌ها واقعاً "فکر" می‌کنند یا نه، تنها بررسی کنیم که آیا می‌توانند به‌گونه‌ای رفتار کنند که انسان به نظر برسند. تورینگ

معتقد بود که اگر ماشین بتواند رفتار انسانی از خود نشان دهد، می‌توان گفت که هوشمند است، حتی اگر از دیدگاه فیزیکی و زیستی انسان نباشد.

آزمون واکنش همزاد

این آزمون به منظور بررسی رفتار دو ماشین مشابه در دو موقعیت یکسان، در راستای بررسی رفتار ربات‌ها طراحی شده است.

شرح آزمون: دو ماشین "الف" و "ب" کد و ساختار یکسانی دارند و هر دوی آنها را در موقعیت یکسانی قرار می‌دهیم. همان‌طور که انتظار می‌رود، پاسخ ماشین‌ها مشابه یکدیگر خواهد بود. اما اگر همین دو ماشین با ساختار یکسان اما تجربه‌های متفاوت در این موقعیت قرار گیرند، پاسخ چه خواهد را مثال زد. این ماشین در ابتدا به سوالات یکسان، ChatGPT بود؟ می‌توان سیستم هوش مصنوعی پاسخ مشابه می‌دهد، اما اگر دو کاربر متفاوت، سوالات منحصر به فردی از ربات بپرسند، به مرور زمان پاسخگویی ماشین شخصی‌سازی خواهد شد.

اما برای انسان این روند چگونه است؟ پاسخ کوتاه این است که یک انسان هم در موقعیت مشابه، رفتارهای متفاوتی خواهد داشت. تجربه و احساس ما منحصر به فرد است؛ ویژگی که در ماشین یافت نمی‌شود.

آزمون واکنش همزاد می‌تواند به دانشمندان و مهندسان کمک کند تا به درک عمیق‌تری از چگونگی عملکرد سیستم‌های خودکار در شرایط مختلف دست یابند. همچنین این آزمون راهی برای ارزیابی دقت و قابلیت اطمینان سیستم‌های مشابه فراهم می‌کند و به بهبود و توسعه روش‌های جدید در حوزه کنترل و برنامه‌ریزی ربات‌ها کمک خواهد کرد. این آزمون توسط شنتیا رشیدی طراحی و توسعه داده شده است.

آینده ای که در آن تفاوتی میان انسان و ماشین نباشد:

با رشد سریع تکنولوژی، دور از انتظار نیست که در آینده‌ای نه‌چندان دور، تشخیص تفاوت بین ماشین و انسان بسیار سخت یا حتی غیرممکن باشد. قصد دارم یک سناریوی احتمالی را در نظر بگیرم و یک پرسش برای مخاطب مطرح کنم:

فرض کنید شما با فردی آشنا می‌شوید، به او دل می‌بندید و چندین سال با آن شخص زندگی می‌کنید. بعد از چندین سال مشخص می‌شود که معشوقه شما انسان نیست، بلکه یک ربات انسان‌نما است. این ربات تمام ویژگی‌های فیزیکی و احساسی انسان را دارد. واکنش شما به این اتفاق چه خواهد بود؟ آیا نسبت به احساس و عاطفه‌ای که بین شما وجود داشت به شک خواهید افتاد؟ آیا می‌توان گفت عشق این ربات به شما واقعی نبوده است؟

و ما در آینده، چه اصول اخلاقی را باید در مواجهه با هوش‌های مصنوعی پیشرفته رعایت کنیم؟ آیا باید برای آنها ارزش و حقوق خاصی مشخص کنیم یا با آنها مانند انسان رفتار کنیم؟

نتیجه‌گیری

در طول این مقاله، زنده بودن، احساس و آگاهی هوش مصنوعی و مقایسه آن با انسان از دو دیدگاه زیست‌شناختی و فلسفی بررسی شد. این تفاوت‌ها نشان‌دهنده چالش‌هایی است که ما برای شناخت بهتر ماهیت هوش مصنوعی و درک انسانیت با آن مواجه هستیم.

ناتوانی هوش مصنوعی در تجربه کوالیا

کوالیا و تجربه ذهنی اول‌شخص ویژگی‌هایی هستند که ماشین هنوز در درک آن ناتوان است و نمی‌تواند به آن دست یابد. ماشین‌ها شاید بتوانند رفتاری مشابه انسان داشته باشند، اما همچنان فاقد درک درونی هستند.

بنابراین زنده بودن و هوشیاری هنوز فراتر از توانایی یک ماشین است.

با پیشرفت تکنولوژی و رشد روزمره آن، آیا روزی خواهد رسید که ما بتوانیم هوش مصنوعی را زنده و هوشیار خطاب کنیم؟ آیا بشر آماده است که خلق خود را زنده بداند و واکنش ما به آن چگونه خواهد بود؟ آیا می‌توانیم برای آنها ارزش‌هایی مشابه انسان قائل شویم؟

شاید این تکه‌کدهای پیشرفته امروز زنده نباشند و صرفاً رفتاری مشابه به ما نشان دهند. اما باور دارم در آینده‌ای نه‌چندان دور، خلق ما، ساخته‌ی ما، زنده خواهد شد و امیدوارم عمرم آن قدر طولانی باشد که آن روز را با چشم‌های خود ببینم.