

# توسعه هوش مصنوعی خودآگاه چه مزایا و معایبی برای جامعه به همراه می‌آورد؟ - دیجیاتو

شایان ضیایی | چهارشنبه، ۰۴ تیر ۱۳۹۹

ربات‌های قاتل به نظر از دل داستان‌های علمی-تخیلی درآمده‌اند، اما ظاهراً دارند به واقعیت نزدیک می‌شوند. اخیراً ۱۰۰ متخصص حوزه رباتیک، از جمله ایلان ماسک، موسس تسلا و اسپیس اکس، به سازمان ملل متحد نسبت به خطرات احتمالی سلاح‌های مرگبار خودکار هشدار داده و خواستار این شده‌اند که استفاده از هوش مصنوعی در زمینه‌های تسلیحاتی، به لیست سلاح‌های ممنوعه اضافه شود.

سوال اینجاست که چرا اکنون چنین اتفاقی افتاده؟ اکنون که هوش مصنوعی از هر زمان دیگر برای بشریت مفیدتر به حساب می‌آید. چرا این متخصصان نگران خطرات هوش مصنوعی هستند؟ برای پاسخ به این سوالات، باید نخست درک کنیم که چطور به این نقطه رسیده‌ایم و ماشین‌های خودآگاه چطور در حال ظهور هستند.

جمع‌آوری هوشمند



در تابستان سال ۱۹۵۶ میلادی، پدران هوش مصنوعی به یک گردهمایی در کالج دارتماوث واقع در نیو همپشایر رفتند تا به گفتگو درباره این علم نوظهور پرداخته و اهداف کلی آن را ترسیم کنند. مفهوم مد نظر آن‌ها از «هوش انسانی» بسیار دقیق و مشخص بود. کامپیوترها باید کاری را انجام دهند که یک مرد بالغ، تحصیل کرده و منطقی انجام می‌دهد - چرا که در ذهن این پدران، هوش مصنوعی یک مرد بود و نمی‌توانست زن باشد. کامپیوتر باید از دانش و منطق خود برای حل کردن مسائل پیچیده استفاده می‌کرد. این هدفی بود که پا را فراتر از پردازش‌های عددی کامپیوترهای آن زمان می‌گذاشت.

دانش جدید هوش مصنوعی نیازمند کامپیوتری متفاوت بود که بتواند دانش بسازد و آن را ذخیره‌سازی و قابل دسترسی کند، در تحلیل اطلاعات و حقایق از قوانین منطقی پیروی نماید، سوال بپرسد، به حقایق تازه دست یابد، تصمیم بگیرد و پشت تمام کارها و منطق‌هایش، توضیحی قابل پذیرش باشد.

سال‌ها پیش‌تر آلن تورینگ، ریاضی‌دان انگلیسی، ماشینی هوشمند را متصور شده بود که به زبان ما صحبت کرده و به این ترتیب با شباهت‌هایش به انسان، ما را متقاعد نسبت به توانایی‌های خود می‌سازد. با این همه، منابع الهام بنیادین هوش مصنوعی هیچ ارتباطی به احساسات، اخلاف و خودآگاهی نداشتند. اگرچه درک زبان یکی از نخستین اهداف هوش مصنوعی بود، قرار نبود ذهن انسان درون یک ماشین بازسازی شود، بلکه قرار بود صرفاً از برخی ابعاد آن تقلید شود. این را هم فراموش نکنیم که اواخر دهه ۱۹۵۰ میلادی هنوز دانش ما راجع به ذهن و مغز در دوران طفولیت بود.

و با این حال، تمایل به بزرگان‌اندیشی از همان ابتدا مشاهده می‌شد. در سال ۱۹۴۳ میلادی، یک عصب‌شناس پیشرو به نام وارن مک کولاک و همینطور والتر پیترس منطقدان از شباهت‌های میان الکترونیک‌ها و نورون‌های عصبی پرده برداشتند. چه می‌شد اگر می‌توانستیم تمام مغز انسان و سیم‌کشی‌های ظریف‌اش را درون یک کامپیوتر الکترونیکی بازسازی کنیم؟ چه می‌شود اگر به جای تشریح اینکه کامپیوتر باید به چه چیزی فکر کند، اجازه دهیم خودش فکر کرده و «ذهن» خودش را به مرور شکل دهد؟ چه می‌شود اگر هوش مصنوعی را بیشتر شبیه به انسان کنیم؟

بعد در سال ۱۹۵۶ میلادی، همان سالی که کنفرانس کالج دارتماوث برگزار شد، فرانک روزنبلات، روان‌شناس آمریکایی از «Perceptron» رونمایی کرد، الگوریتمی که روی سخت‌افزارهای مقلد سیستم عصبی به اجرا در می‌آمد و می‌توانست شبیه به شبکه‌های عصبی، موضوعات مختلف را فرا بگیرد و این کار را از طریق تقویت یا تضعیف ارتباطات میان نورون‌های همسایه انجام می‌داد. Perceptron پدر بزرگ شبکه‌های عصبی مصنوعی یا یادگیری عمیق است که امروز -یعنی ۶۰ سال بعد- تمام آن‌ها را جزو ایده بزرگی به نام «هوش مصنوعی» دسته‌بندی می‌کنیم.

## تفکری شبیه به انسان



رویکرد منطقی ابتدایی در زمینه هوش مصنوعی، طی چند سال نتایجی جالب با خود به همراه آورد اما در نهایت به بن‌بست خورد. ابتکار عمل روزنبلات، رویکردی جایگزین به ارمغان آورد، رویکردی که باعث می‌شد کامپیوترها پا را فراتر از منطق گذاشته و به سراغ حل مشکلی بسیار دشوار بروند: ادراک. کار او برای مدتی نسبتاً طولانی به فراموش سپرده شد، اما در دهه ۱۹۹۰ میلادی، نسل جدیدی از محققان نابغه دوباره آن را احیا کردند. با توسعه سخت‌افزاری که قادر به پردازش موازی بود، ساخت الگوریتمی که ذهن انسان را شبیه‌سازی کند امکان‌پذیر شد. این یک انقلاب تکنولوژیک بود که هوش مصنوعی و اهدافش را از نو تعریف می‌کرد.

ما اکنون در دورانی زندگی می‌کنیم که ماشین‌های هوشمند تقریباً هر روز رکوردشکنی می‌کنند. درحالی که میلیاردها دلار روی تحقیقات مرتبط با هوش مصنوعی سرمایه‌گذاری می‌شود، ماشین‌ها دارند شکلی هوشمندتر به خود می‌گیرند. کلید دستیابی به این هوشمندی سریع، توانایی آن‌ها در یادگیری است. یک شبکه عصبی مصنوعی (درست مانند شبکه عصبی طبیعی درون مغز ما) می‌تواند تشخیص حقایق را با پردازش دیتای درون اتصالات عصبی بیاموزد.

برای مثال، شبکه عصبی مصنوعی می‌تواند پیکسل‌های یک تصویر را پردازش کرده و چهره انسان، یا یک حیوان یا یک شی را شناسایی کند. و به محض اینکه هوش مصنوعی قادر به استخراج حقایق از اطلاعات باشد، می‌تواند همان کار را دوباره و دوباره انجام دهد. چنین ماشین‌هایی برای یادگیری نیازمند حجم عظیمی از اطلاعات هستند و معمولاً هم انسان‌ها بر روند یادگیری نظارت دارند.



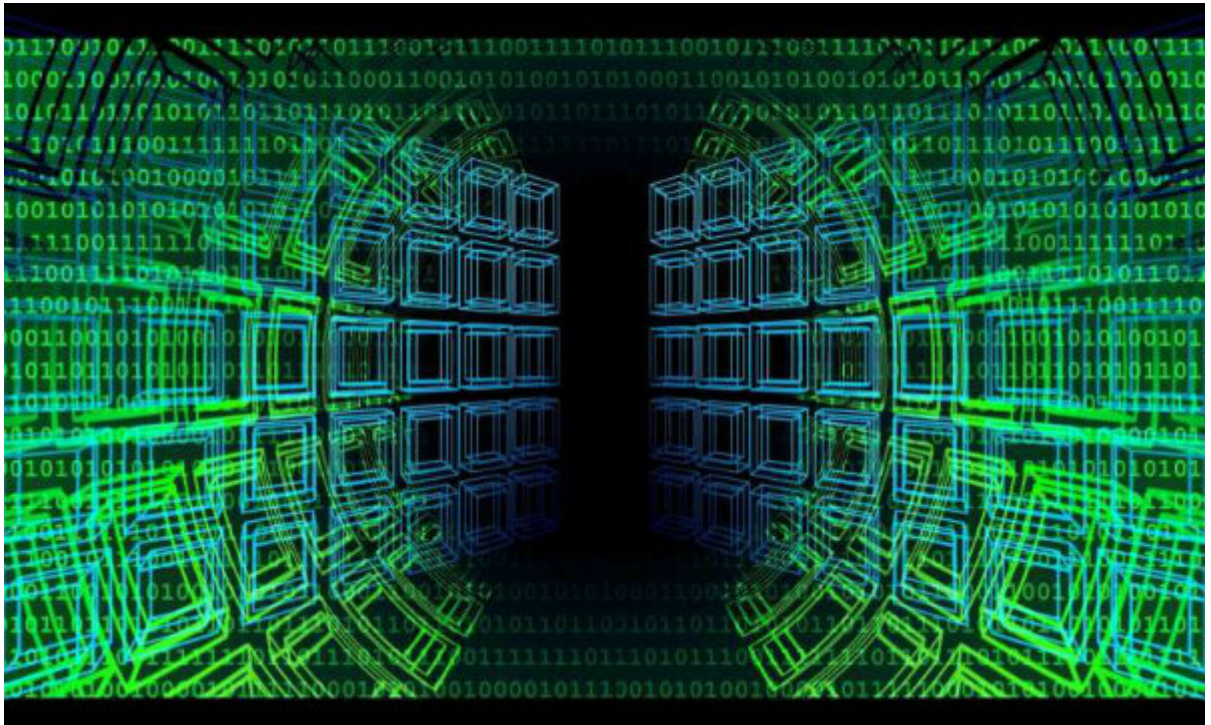
اما ماشین‌ها به تنهایی هم قادر به آموزش دیدن هستند و این کار از طریق پروسه‌ای به نام «یادگیری تقویتی» انجام می‌شود. آلفا گو، الگوریتم توسعه یافته توسط شرکت انگلیسی DeepMind، به همین روش توانست به قهرمان جهانی بازی «گو» تبدیل شود. در بازی‌ای که به پیچیدگی بسیار زیاد شهرت داد، هوش مصنوعی با رقابت مداوم با خودش توانست استاد کمربند-مشکی دان ۹ را شکست دهد.

آلفاگو نخستین گام در تکامل هرچه بیشتر هوش مصنوعی بود چون توانست تصویری هرچند کوچک از هدف غایی این تکنولوژی ترسیم کند: ساخت هوشی عمومی و شبیه به انسان. برای پیروزی در بازی پیچیده‌ای مانند گو، باید خلاقانه فکر کرده و از «شهود» بهره ببرید. این یعنی باید بتوانید آموخته‌های قبلی را در مشکلات تازه و غیرمنتظره به کار بگیرید.

اما کامپیوترها هنوز به چنین نقطه‌ای نرسیده‌اند: کامپیوترها می‌توانند با تشخیص چهره، صدا یا زبان انسانی، حقایق را از اطلاعات استخراج کرده و بسته به درکی که از آن اطلاعات دارند، دست به پیش‌بینی‌های گوناگون بزنند. برای اینکه گام بزرگ بعدی در رسیدن به هوش شبیه به انسان برداشته شود، ماشین‌ها باید از ادراک خود برای تصمیمات و فعالیت‌های بلندرنگ استفاده کنند.

برای مثال، کافی نخواهد بود اگر یک اتومبیل خودران تشخیص دهد پیکسل‌هایی که می‌بیند در واقع ونی سفید رنگ هستند که دارد سرعتش را کم می‌کند. هوش مصنوعی باید این منطق را داشته باشد که تصمیمی در راستای این اتفاق بگیرد. و چنین تصمیمی، تفاوت میان مرگ و زندگی خواهد بود. به عبارت دیگر، فاز بعدی در تکامل هوش مصنوعی اینست که ماشین‌ها وارد حوزه دردرساز اخلاقیات انسانی شوند.





تلاش برای اینکه هوش مصنوعی شبیه به انسان شود کاری بسیار دشوار است. ما به «هوش» به عنوان سنگ بنای تکامل‌مان نگاه می‌کنیم و برای آن ارزش قائلیم. این ارزش عمیقا در تمام فرهنگ انسانی رسوخ کرده است. آنچه باعث شد گونه انسان قادر به بقا در مواجهه با شکارچیان مجهزتر باشد، ریشه در توانایی ما در یادگیری، ابتکار و تطبیق‌پذیری دارد. هوش مصنوعی، خلاقیت و هوش انسانی را توربوشارژ خواهد کرد. دلایل اقتصادی بسیاری هم برای تلاش بیشتر در راستای ساخت ماشین‌های هوشمندتر وجود دارد.

اگرچه هوش مصنوعی بسیاری از مشاغل مانند کارهای کارخانه‌ای و خرده‌فروشی را از بین خواهد برد، هیچ تکنولوژی دیگری چنین پتانسیلی برای رشد مداوم اقتصادی و تامین نسل‌های آتی ندارد. در زمین علم هم ماشین‌ها می‌توانند انبوهی از اطلاعات را پردازش و دانش‌های جدیدی کشف کنند که مشابه‌شان هیچوقت دیده نشده.

تمام حوزه‌های علمی از مزیت‌های هوش مصنوعی در مدیریت دیتا بهره‌مند خواهند شد. فیزیک‌دانان می‌توانند روی قوانین کلی حاکم بر طبیعت پژوهش کنند، زیست‌شناسان می‌توانند داروهای تازه برای درمان بیماری‌ها بیابند، پزشکان هم می‌توانند تشخیص‌های بهتری داشته باشند. به دلایل متعدد فرهنگی، اقتصادی و علمی، تلاش برای توسعه هرچه بیشتر هوش مصنوعی کاملا منطقی به نظر می‌رسد. اما هرچه ماشین‌هایمان بیشتر شبیه انسان می‌شوند، هوش مصنوعی راهش را به اپلیکیشن‌های بیشتری باز می‌کند و به این ترتیب، مشکلاتی تازه هم ظهور می‌کنند.



شبیه‌سازی ذهن انسان با شبکه‌های عصبی مصنوعی به این معناست که کامپیوترهایی مرموز و نافرمان‌بردار خواهیم داشت. این همان مشکلی در حوزه هوش مصنوعی است که از آن به عنوان «جعبه سیاه» یاد می‌شود. در ذهن و در ماشین، اطلاعات درون شبکه پخش شده است. وقتی ما یک شماره تلفن را از حافظه‌مان به یاد می‌آوریم، اینطور نیست که به بخش خاصی از ذهن که درون گوشت و پوست قرار گرفته دسترسی یابیم.

در عوض، هر عدد درون سیناپس‌های متعددی پخش شده که نورون‌های مختلف را به یکدیگر متصل می‌کنند. ما دقیقاً «نمی‌دانیم» که چه می‌دانیم و چطور آن را می‌دانیم. موضوع صرفاً اینست که ما به خاطر خودآگاه بودن‌مان، می‌توانیم تمام ماجرا را برای خودمان منطقی کنیم و ایده‌ها را «توصیف».

همانطور که دیوید ایگلن، عصب‌شناس آمریکایی نشان داده، اکثر چیزهایی که ما از آن‌ها آگاه می‌شویم، پیش‌تر در سطح ناخودآگاه اتفاق افتاده‌اند. این موضوع مشکلی چندان بزرگ برای انسان‌ها نبوده، چون در سیستم‌های قانونی و اخلاقی‌مان فرض کرده‌ایم که هر شخص به تنهایی مسئول تفکرات و اعمالش است - حداقل زمانی که تعاملات شیمیایی ذهن‌مان، در رنج قابل پذیرش و «نرمال» جامعه قرار می‌گیرند.



وقتی نوبت به یک ماشین هوشمند غیر خودآگاه می‌رسد، «مشکل جعبه سیاه» به طرح این مسئله می‌پردازد که گرچه پیش‌بینی‌ها و پیشنهادات ماشین می‌توانند دقیق و کاربردی باشند، اما هوش مصنوعی قادر به توضیح منطق خود نیست. اتومبیلی بدون راننده را تصور کنید که مجبور به تصمیم میان مرگ و زندگی شده، تصادف می‌کند و چند انسان را به کشتن می‌دهد. با تکنولوژی‌هایی که امروز در اختیار داریم، تشخیص اینکه چرا اتومبیل دست به چنین تصمیمی زده غیرممکن خواهد بود.

مشکل جعبه سیاه زمانی جدی‌تر می‌شود که اطلاعاتی جانب‌دارانه به ماشین بدهیم. در این صورت رباتی خواهیم داشت که به نتیجه‌گیری‌های جانب‌دارانه خواهد رسید و رفتارهای غیر اجتماعی از خود نشان می‌دهد. در ماه مارس ۲۰۱۶ برای مثال مایکروسافت یک بات در توئیتر منتشر کرد که با تحلیل توئیتهای دیگران، زبان انسان را فرا می‌گرفت. در کمتر از ۲۴ ساعت، ربات مایکروسافت شروع به ساخت جملات نژادپرستانه و بیگانه‌هراسانه کرد.

محققان سخت در تلاش بوده‌اند که به طرق مختلف، با مشکل جعبه سیاه مقابله کنند. در ماه اکتبر ۲۰۱۶، محققان دیپ‌ماینند یک مقاله در نشریه نیچر منتشر کردند و به توصیف «ماشین پردازش عصبی تفکیک‌پذیر» پرداختند که می‌تواند شبکه عصبی را با حافظه اکسترنال ترکیب کند.

جدا کردن روند پردازش از دیتا، گامی مهم به سمت ماشین‌های هوشمندی است که از لحاظ اخلاقی هم قابل اتکا خواهند بود. در تئوری امکان‌پذیر است که ارزش‌های اخلاقی را درون هوش مصنوعی کدنویسی کنیم. اما این رویکرد هیبریدی برای توسعه هوش مصنوعی امن‌تر، شاید برای آینده کافی نباشد.



## معضل اخلاقی واقعی



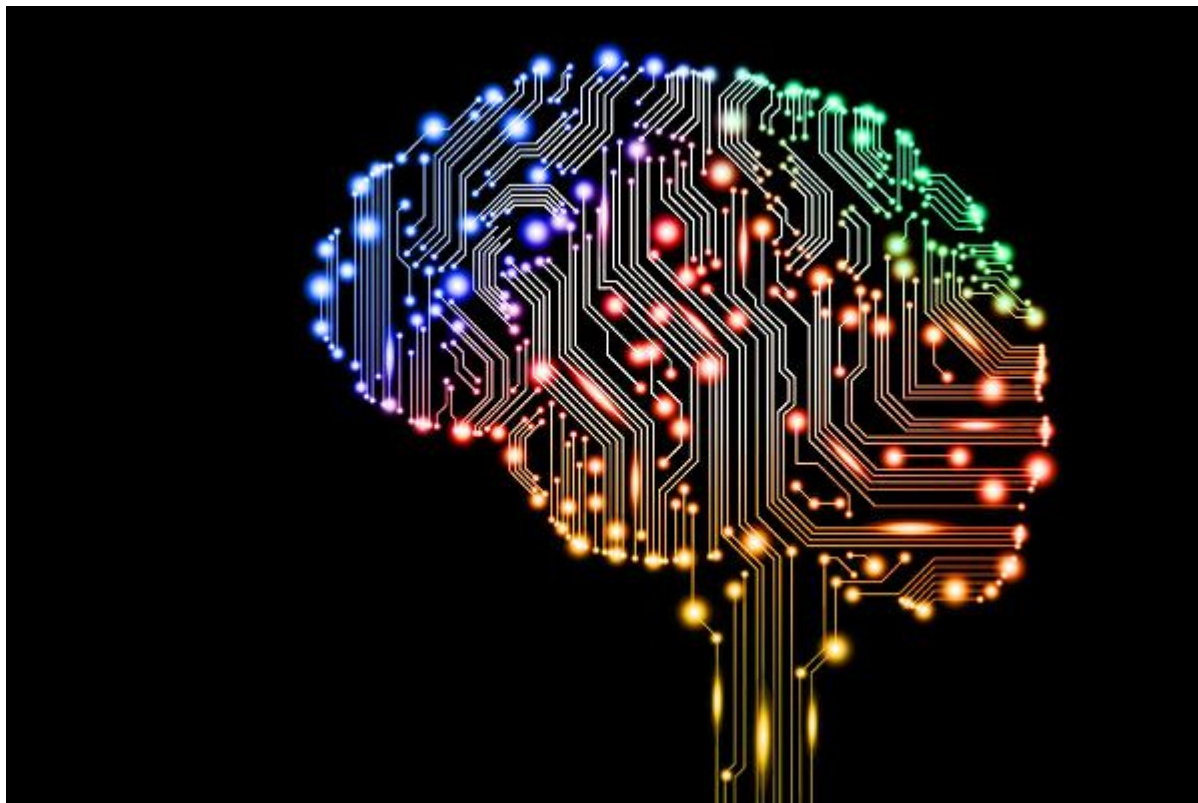
بعد از شروعی سخت و ناامیدی‌های فراوان در طول چندین سال، بشریت بالاخره به تکنولوژی‌ای دست یافته که پتانسیل بازسازی همه‌چیز را دارد. انگیزه‌های اقتصادی و فرهنگی پشت توسعه هرچه بیشتر هوش ماشینی به این معنا خواهند بود که در نهایت هوشی عمومی و کاملاً شبیه به انسان خواهیم داشت، احتمالاً تا ۱۰ تا ۲۰ سال دیگر. پیش‌بینی‌ها می‌تواند متفاوت باشد، اما اگر پیشرفت‌های اخیر را معیار قرار دهیم، فاصله چندانی با از راه رسیدن هوش مصنوعی عمومی نخواهیم داشت.

وقتی این اتفاق بیفتد، ما در واقع ماشین‌هایی ساخته‌ایم که قادر به دریافت حجم وسیعی از اطلاعات بوده و پیش‌بینی‌ها و تحلیل‌هایی فرابشری ارائه می‌کنند. طنز ماجرا اینجاست که اگر مشکل جعبه سیاه حل نشده باقی بماند، احتمالاً دوباره خودمان را در شرایطی مشابه به یونانیان باستان بیابیم که به سراغ غیب‌گوی نیایشگاه دلفی می‌رفتند و از آپولو می‌خواستند آینده‌شان را پیش‌بینی کند. زبان خدایان برای انسان کاملاً غیر قابل درک بود و بنابراین یک میانجی -یعنی پیتیا- احضار می‌شد تا بیانات مرموز خدایان -که آن زمان توسط مبلغان مذهبی به هر شکلی پیچ و تاب داده می‌شدند- را ترجمه کند.

با هوش مصنوعی، ما در حال ساخت خدایانی هستیم که بسیار برتر از ما ظاهر خواهند شد. منطق آن‌ها فراتر از درک بشری خواهد بود. بنابراین با یک معضل غیر قابل پذیرش مواجه خواهیم شد. آیا باید با ایمانی کورکورانه به این خدایان سیلیکونی اعتماد کنیم؟ بسیار بعید است که یک هوش مصنوعی سطح انسانی که قادر به توصیف منطق خود نیست مورد پذیرش جامعه قرار بگیرد.

به این ترتیب تنها یک انتخاب داریم: اینکه به توسعه هرچه بیشتر هوش مصنوعی ادامه داده و مشخصه‌های انسانی مانند هوش عاطفی، همدلی و خودآگاهی را هم به آن تزریق کنیم.





این تنها راه برای حل مشکل ارتباط میان ماشین‌های هوشمند و ما خواهد بود. ماشین‌های آینده باید به صورت درونی قادر به درک احساسات ما بوده و خود را با مود و شرایط روان‌شناختی ما تطبیق دهند. آن‌ها باید با تحلیل صدا و چهره ما به احساساتمان پی ببرند و براساس انبوه اطلاعاتی که طی چند سال راجع به ما دست آورده‌اند، دست به نتیجه‌گیری بزنند. در این صورت است که می‌توانیم به ربات‌ها اعتماد کرده و آن‌ها را به عنوان عضوی از جامعه بشری بپذیریم.

برخلاف دوست‌های «انسانی»مان، این ماشین‌های هوشمند از لحاظ عاطفی، همه‌چیز را راجع به ما می‌دانند. نخواهیم توانست چیزی را از آن‌ها پنهان کنیم. آن‌ها احتمالاً بیشتر از خود ما، ما را بشناسند و به همین طریق است که می‌توانند ما را در تصمیمات و انتخاب‌های بهتر در زندگی راهنمایی کنند. دسترسی به چنین چیزی در زندگی بسیار ارزشمند است و برای همین، انسانی‌تر کردن هوش مصنوعی مورد استقبال قرار می‌گیرد.

اما این اتکا بر ماشین‌های هوشمند، چند سوال اخلاقی هم به وجود می‌آورد. اگر ماشینی داشته باشیم که همواره از ما در برابر خطا محافظت کند -مثل یک مادر یا پدر بسیار مهربان و خردمند- انسان‌های آینده چطور از اشتباهات درس خواهند گرفت و خود به بلوغ بیشتر می‌رسند؟ آیا دسترسی به یک هوش مصنوعی همه‌چیزدان باعث نمی‌شود که مردم و فرهنگ به سمت نابخردی حرکت کنند؟

در مورد خودآگاهی هوش مصنوعی چه باید کرد؟ آیا باید آنقدر محدودیت‌های تکنولوژیک را کنار بزنیم تا ماشین‌هایی خودآگاه بسازیم؟ آیا کاری معقولانه است که در مجموعه‌ای از سیم‌ها، فن‌های خنک‌کننده و چیپ‌ها زندگی بدمیم؟ شاهکار گوتیک مری شلی، یعنی فرانکشتاین، و همینطور فیلم Balde Runner ریدلی اسکات، هر دو به زیبایی نشان می‌دهند که حرکت به سمت

چنین آینده‌ای می‌تواند چه تبعاتی داشته باشد.

دیجیاتو