

## فهرست مطالب

### مقدمه: داستان تیم امگا

#### ۱- به مهمترین گفتگوی زمانه ما خوش آمدید.

پیشینه مختصری در مورد پیچیدگی

دوران سه گانه زندگی

اختلاف نظرها

تصورهای غلط

مسیر پیش رو

#### ۲- ماده دارای هوش می شود.

هوش چیست؟

حافظه چیست؟

محاسبه چیست؟

یادگیری چیست؟

#### ۳- آینده نزدیک: پیشرفت های مهم، خطاهای نرم افزاری، قوانین، سلاح ها و شغل ها

پیشرفت های مهم

هوش مصنوعی خطادار در مقابل هوش مصنوعی مقاوم

قوانین

سلاح ها

شغل ها و دستمزدها

آیا هوش عمومی مصنوعی را می توان در سطح هوش بشر یا فراتر از آن ساخت؟

#### ۴- آیا انفجار هوش رخ خواهد داد؟

تمامیت خواهی

سیستم پرومیتوس دنیا را تصاحب می کند.

سناریوی خیزش آهسته و سناریوی چندقطبی

آدم های مکانیکی و ربات های انسان نما

در آینده چه اتفاقی خواهد افتاد؟

## ۵- آنچه که در ۱۰۰۰۰ سال آینده اتفاق خواهد افتاد.

مدینه فاضله آزادی خواهان

دیکتاتور خیرخواه

مدینه فاضله عدالت خواهان

نگهبان در (کنترل کننده دسترسی)

خدای نگه دارنده

خدای به بردگی گرفته شده

گروه غالب

بازماندگان

نگهبان باغ وحش

۱۹۸۴

برگشت به عقب

خودتخریبی

چه می خواهید؟

## ۶- موهبت عالم هستی در میلیاردها سال بعد و پس از آن

استفاده حداکثری از منابع خود

دستیابی به منابع از طریق سکنی گزیدن در عالم هستی

سلسله مراتب عالم هستی

دورنمای آینده

## ۷- اهداف

علم فیزیک: منشا اهداف

علم زیست شناسی: تکامل اهداف

علم روان شناسی: دنبال کردن اهداف یا مخالفت در برابر آنها

علم مهندسی: برون سپاری اهداف

هوش مصنوعی دوستانه: همراستا کردن اهداف

اخلاقیات: انتخاب اهداف

اهداف نهایی چه خواهند بود؟

## ۸- هوشیاری

چه کسی اهمیت می دهد؟

منظور از هوشیاری چیست؟

مشکل چیست؟

آیا هوشیاری فراتر از علم است؟

سرنخ های تجربی در مورد هوشیاری

نظریه هایی در مورد هوشیاری

اختلاف نظرها در مورد نظریه هوشیاری

هوشیاری هوش مصنوعی چگونه به نظر می رسد؟

معنا بخشیدن به زندگی

## سخن آخر: داستان تیم اف ال آی

### مقدمه: داستان تیم امگا

قبل از خواندن داستان تیم امگا باید توجه داشت که این یک داستان ساختگی است که در دنیای واقعی اتفاق نیفتاده است و فقط توانایی های فوق انسانی هوش مصنوعی را نشان می دهد. تیم امگا مانند روح شرکت بود. در شرایطی که سایر اعضای شرکت، پول لازم برای بهره برداری از برنامه های کاربردی تجاری مختلف هوش مصنوعی را تهیه کرده بودند، تیم امگا برای درخواست استفاده از هوش مصنوعی برای تحقق رویای مدیران ارشد اجرایی خود، از بقیه پیشی گرفته بود. بیشتر کارمندان شرکت به اعضای تیم امگا به عنوان افرادی نگاه می کردند که رویایی تحقق ناپذیر برای چندین دهه بعد دارند. با این حال، سایر کارمندان شرکت با رویای آنها مخالفتی نداشتند و از اعتباری که کارهای پیشرفته تیم امگا به شرکتشان می بخشید، خوشحال بودند و از الگوریتم های پیشرفته ای

که تیم امگا هر چند وقت يكبار به آنها ارائه می کرد، استقبال می کردند. با این حال، آنها از راز نهفته ای كه تیم امگا در پشت وجهه به دقت طراحی شده خود، پنهان کرده بود، اطلاعی نداشتند و آن راز این بود كه آنها برای رسیدن به موفقیت در زمینه تحقق جسورانه ترین نقشه تاریخ بشریت، فاصله بسیار کمی داشتند. مدیر عامل كاریزمای تیم امگا اعضای تیم خود را به گونه ای دستچین کرده بود كه علاوه بر محققان با استعداد، شامل افرادی هم می شد كه از جاه طلبی، كمال گرایی و تعهد قوی برای كمك به بشریت برخوردار بودند. او به اعضای تیم خود یادآوری می كرد كه نقشه شان بسیار خطرناك است و اگر دولت های قوی متوجه آن شوند، برای جلوگیری از تحقق این نقشه یا دزدیدن رمز آن، هر اقدامی از جمله آدم ربایی را انجام خواهند داد، مانند فیزیکدان های برجسته جهان كه به پروژه منهن پیوستند تا مطمئن شوند كه قبل از افراد به شدت آرمان گرا، می توانند سلاح های هسته ای را توسعه دهند. روز به روز بر قابلیت های سیستمی با نام مستعار " پرومئوس " كه با هوش مصنوعی ساخته شده بود، افزوده می شد. هر چند كه توانایی های شناختی پرومئوس، با توانایی بشر در زمینه هایی مانند مهارت های اجتماعی، فرسنگ ها فاصله داشت، اما اعضای تیم امگا، برای ایجاد بهبود فوق العاده در وظیفه خاص آن كه برنامه نویسی سیستم های هوش مصنوعی بود، تلاش مضاعفی را به عمل آورده بودند. با توجه به اینکه آنها بحث " انفجار هوش " را كه توسط ریاضیدان انگلیسی به نام ایروینگ گود<sup>۱</sup> در سال ۱۹۶۵ مطرح شده بود را پذیرفته بودند، این راهبرد را از روی عمد انتخاب کردند. ایروینگ گود می گوید: " اجازه دهید كه ماشین فوق با هوش را به عنوان ماشینی تعریف كنیم كه می تواند از تمام فعالیت های فكري هر انسان با هوشی به مراتب پیشی بگیرد. با توجه به این كه طراحی ماشین ها یکی از این فعالیت های فكري است، ماشین فوق با هوش حتی می تواند ماشین های بهتری را نیز طراحی كند. بدون شك، می توان انتظار داشت كه " انفجار هوش " رخ دهد و هوش بشر به مراتب عقب تر از هوش مصنوعی بماند. بنابراین، اولین ماشین فوق با هوش، می تواند آخرین اختراع مورد نیاز بشر باشد، البته به شرط اینکه ماشین با هوش به اندازه كافی حرف شنو باشد تا بتوان آن را كنترل كرد. " آنها معتقد بودند كه اگر بتوانند به این چرخه تكرار

---

<sup>۱</sup> Irving Good

شونده "خود بهبودی" تداوم بخشند، ماشین به قدری با هوش می شود که می تواند سایر مهارت های انسانی مفید را نیز به خود یاد دهد.

## اولین درآمد میلیون دلاری

در ساعت ۹ صبح یک روز جمعه بود که تیم امگا تصمیم گرفت که سیستم جدیدی را راه اندازی کند و این در حالی بود که سیستم قدیمی پرومتئوس در گروه رایانه های سفارشی سازی شده و در میان ردیف های طولانی ای از قفسه ها در یک اتاق بزرگ با دسترسی کنترل شده و تهویه مطبوع، در یک جای ثابت قرار گرفته بود. اتصال این گروه رایانه ای به اینترنت، به دلایل امنیتی کاملاً قطع شده بود، اما امکان دسترسی به برخی از کپی های داخلی از وب مانند ویکیپدیا، توئیتر، فیس بوک و غیره به عنوان داده های آموزشی برای یادگیری سیستم از آنها، همچنان وجود داشت. روز جمعه فرصت خوبی برای راه اندازی سیستم جدید بود، زیرا خانواده ها و دوستان تیم امگا فکر می کردند که آنها در تعطیلات آخر هفته مشغول تفریح هستند. در ابتدا، سیستم پرومتئوس در نحوه برنامه نویسی سیستم های هوش مصنوعی کمی بدتر از انسان بود، اما عملکرد بد خود را با افزایش قابل توجه سرعتش، جبران کرد. تا ساعت ۱۰ صبح، اولین طراحی مجدد پرومتئوس تکمیل شد که کمی بهتر از طراحی اولیه بود، اما همچنان دارای صفاتی شبیه انسان بود. با این حال، وقتی که پرومتئوس ۵ در ساعت ۲ بعد از ظهر، راه اندازی شد، تیم امگا شگفت زده شد. زیرا سیستم جدید می توانست معیارهای عملکردی قبلی را نابود کند و نرخ رشد پر شتابی داشته باشد. آنها تصمیم گرفتند که تا قبل از هنگام شب، سیستم پرومتئوس ۱۰ را نصب و پیاده سازی کنند و مرحله دوم از نقشه خود را که همان پول درآوردن بود، شروع کنند. هدف اول آنها کسب درآمد از اینترنت با استفاده از بازار ترک مکانیکی آمازون<sup>۲</sup> (Amazon Mechanical Turk) بود. پس از اینکه این بازار در سال ۲۰۰۵ به عنوان یک بازار اینترنتی مبتنی

---

<sup>۲</sup> راهی جهت کسب درآمد ناچیز و انجام کارهای آسان و بدون ذهن است. وب مبتنی بر جمع سپاری می باشد که به کاربرانی که برای خود کسب و کاری دارند اجازه می دهد تا بتوانند بواسطه ی ظرفیت های انسانی و هوش خود، کاری که سیستم های کامپیوتری نمی توانند انجام دهند را انجام دهند. این وب سایت زیر نظر شرکت آمازون می باشد.

بر برون سپاری، راه اندازی شد، رشد سریعی را تجربه کرد. به طوریکه ده ها هزار نفر بی نام و نشان از سراسر جهان به صورت ۲۴ ساعته کارهای به شدت ساختاریافته ای موسوم به " کارهای قابل انجام توسط هوش انسان از طریق هوش مصنوعی " را اجرا می کردند. دامنه این کارها از مکتوب ساختن صداهای ضبط شده تا دسته بندی تصاویر و توصیف صفحات وب را شامل می شد و همه آنها در یک چیز مشترک بودند و آن این بود که اگر کارهای قابل انجام توسط هوش انسانی را به خوبی انجام دهید، کسی متوجه نمی شد که از سیستم هوش مصنوعی استفاده کرده اید. سیستم پرومیتوس ۱۰ قادر بود که نیمی از این گروه کارها را به نحو خوب و قابل قبولی انجام دهد. تیم امگا برای هر یک از این گروه کارها از سیستم پرومیتوس می خواست که یک واحد مستقل نرم افزاری<sup>۳</sup> سفارشی را برای هوش مصنوعی طراحی کند که بتواند چنین وظایفی را به دقت انجام دهد. آنها در ادامه، واحد مستقل نرم افزاری را روی بخش خدمات وب آمازون بارگذاری کردند. به ازای هر دلاری که تیم امگا به بخش خدمات ذخیره سازی و محاسبات رایانه ای آمازون می پرداخت، بیش از دو دلار از بخش بازار ترک مکانیکی آمازون دریافت می کرد. شرکت آمازون فکر نمی کرد که چنین فرصتی برای استفاده از راهبرد آربیتراژ<sup>۴</sup> داشته باشد. تیم امگا برای دنبال کردن این مسیر، هزاران حساب جداگانه به نام های افراد ساختگی را در بازار ترک مکانیکی ایجاد کرده بود که واحدهای مستقل نرم افزاری ساخته شده توسط پرومیتوس، هویت آنها را پذیرفته بودند. مشتری های بازار ترک مکانیکی هر هشت ساعت به آنها پول پرداخت می کردند و تیم امگا این پول را مجددا در واحدهای مستقل نرم افزاری موجود در جدیدترین ویرایش بهبود داده شده پرومیتوس، سرمایه گذاری می کرد. آنها متوجه شدند که بدون جلب توجه ناخواسته دیگران به خودشان، نمی توانند روزی یک میلیون دلار کسب کنند و به همین دلیل، تصمیم گرفتند که از تقاضای شرم آور از مدیر ارشد مالی برای گرفتن پول نقد اجتناب کنند.

## بازی های خطرناک

جدا از موفقیت هایی که تیم امگا در زمینه هوش مصنوعی داشته است، باید توجه داشت که پروژه های اخیر این تیم در زمینه برنامه ریزی برای پول درآوردن هر چه سریع تر با راه اندازی سیستم

---

<sup>۳</sup> Software module

<sup>۴</sup> arbitrage

پرومتئوس، حسابی باعث لذت بردن آنها شده بود. در اصل، کل اقتصاد دیجیتالی را می توان در هوا قاپید، اما آیا بهتر نیست که کار خود را با ساختن بازی های رایانه ای، موسیقی، فیلم، نرم افزار یا نوشتن کتاب و مقاله، تجارت در بازار سهام یا ثبت اختراع ها و فروش آنها شروع کرد؟ کار آنها در نهایت منجر به بیشینه سازی نرخ بازگشت سرمایه شان شد و در چنین شرایطی می توان گفت که راهبردهای سرمایه گذاری عادی، فقط شبیه تقلید با حرکت آهسته ای از کارهای آنها بود. در حالی که سرمایه گذار عادی را می توان با نرخ بازگشت سرمایه ۹ درصد در سال، راضی کرد، سرمایه گذاری تیم امگا در بازار ترک مکانیکی منجر به نرخ بازگشت سرمایه ۹ درصد در ساعت شده بود. اما حالا که آنها بازار ترک مکانیکی را اشباع کرده بودند، باید چه می کردند؟ اولین چیزی که به ذهنشان رسید این بود که سود کلانی را از بازار سهام کسب کنند. تقریباً همه آنها در یک مرحله زمانی، پیشنهاد یک شغل سودآور را برای سرمایه گذاری پر خطر برای توسعه هوش مصنوعی، رد کرده بودند. با این حال، قوانین جدید در زمینه قراردادهای مالی که پس از رکود اقتصادی سال گذشته وضع شده بود، گزینه های آنها را محدود کرد. آنها به زودی متوجه شدند که جایی وجود ندارد که در آنجا بتوانند سود بیشتری از فروش محصولات شخصی خود کسب کنند. وقتی که شما دارای اولین سیستم فوق هوشمند هوش مصنوعی در جهان هستید، در موقعیت بهتری برای سرمایه گذاری در شرکت خود هستید تا این که در شرکت های دیگر سرمایه گذاری کنید. با وجود برخی استثناءها، تیم امگا احساس می کرد که این کار مانع از جلب توجه ناخواسته دیگران به خودشان خواهد شد. وقتی که آنها تمرکز خود را به سمت محصولات قابل توسعه و فروش خود جابه جا کردند، به نظر می رسید که بازی های رایانه ای بهترین گزینه باشد. سیستم پرومتئوس به سرعت می توانست در زمینه هایی مانند طراحی بازی های جذاب، مدیریت کدنویسی، طراحی گرافیک، رهگیری تصاویر با پرتو و سایر کارهای مورد نیاز برای تولید محصول نهایی، استاد شود. علاوه بر این، پرومتئوس پس از ساده سازی و قابل فهم ساختن اطلاعات وب در مورد سلاقی افراد، می توانست دقیقاً متوجه شود که هر گروه از افراد بازی کننده رایانه ای چه چیزهایی را دوست دارند و سپس، از توانایی های فوق بشری هوش مصنوعی برای بهینه سازی درآمد حاصل از فروش یک بازی استفاده کند. درآمد ناخالص حاصل از فروش یکی از بازی های رایانه ای در سال ۲۰۱۱ که تیم امگا ساعت های زیادی روی آن کار کرده

بودند، در همان هفته اول بیش از ۴۰۰ میلیون دلار بود و تیم امگا مطمئن بود که با سرمایه گذاری یک میلیون دلاری در سیستم پرومئوس جهت تامین منابع خدمات ذخیره سازی و محاسبات رایانه ای، می تواند چنین چیز اعتیاد آوری را بسازد. آنها می توانستند بازی های رایانه ای را به صورت آنلاین بفروشند و از پرومئوس برای ایجاد آدمک های ماشینی بازی رایانه ای که با تلاش های بلاگ نویسان می توانستند به جای انسان حرف بزنند، استفاده کنند. اگر آنها هفته ای ۲۵۰ میلیون دلار بابت فروش این بازی ها کسب می کردند، می توانستند نرخ بازگشت سرمایه را به ۳ درصد در ساعت برسانند که اگرچه تا حدودی پایینتر از نرخ بازگشت سرمایه ۹ درصد در ساعت حاصل از سرمایه گذاری در بازار ترک مکانیکی بود، ولی دوام بیشتری داشت. اگر آنها موفق می شدند که هر روز مجموعه ای از بازی های رایانه ای دیگر را نیز توسعه دهند، ظرف مدت کوتاهی می توانستند ۱۰ میلیارد دلار کسب کنند. با این حال، یکی از اعضای تیم امگا که در زمینه امنیت سایبری تخصص داشت، آنها را از این نقشه منصرف کرد. او تاکید داشت که این نقشه از این جهت خطرناک است که پرومئوس ممکن است سرنوشتش را تحت کنترل خود درآورد و کسی از چگونگی تحول اهداف سیستم پرومئوس در طی "چرخه خود بهبودی بازگشتی" اش مطمئن نیست. به همین دلیل، تیم امگا تصمیم گرفت که ریسک نکند و همه تلاش خود را بکند تا پرومئوس را محدود نگه دارد و چارچوب هایی را برای آن تعیین کند تا در اینترنت رها نشود. آنها محدودیت های فیزیکی ای را برای موتور اصلی پرومئوس که در اتاق سرور مدیریت می شد، قائل شدند، به طوریکه اتصال آن به اینترنت قطع شد و خروجی پرومئوس فقط در قالب پیام ها و اسناد متنی ای ظاهر می شد که به رایانه تحت کنترل تیم امگا ارسال می شد. از سوی دیگر، در رایانه های متصل به اینترنت، اجرای هرگونه برنامه پیچیده ایجاد شده توسط پرومئوس، پیشنهاد خطرناکی بود، زیرا تیم امگا کاملاً از عملکرد سیستم پرومئوس در آینده با خبر نبود. وقتی که تیم امگا برنامه نرم افزاری نوشته شده توسط سیستم پرومئوس را آزمایش کرد، مراقب بود که آن را فقط در درون یک "ماشین مجازی" اجرا کند. ماشین مجازی به گونه ای برنامه ریزی شده بود که رفتار یک رایانه را شبیه سازی می کرد. تیم امگا از ماشین مجازی مختص خود موسوم به "جعبه پاندورا" استفاده می کرد که رفتار یک ماشین بیش از حد ساده شده فاقد صفحه کلید، مانیتور، بلندگو را که به اینترنت وصل نمی شد، شبیه سازی می کرد. تیم امگا همه چیزها را به گونه ای تنظیم



کرده بود که فقط یک فایل صوتی می توانست وارد جعبه پاندورا شود و فقط یک نسخه متنی از فایل صوتی می توانست از آن خارج شود. به جز همین یک ورودی و خروجی، امکان خروج نرم افزار از درون جعبه پاندورا وجود نداشت و این نرم افزار در یک جهان موازی با قوانین محاسباتی خاص خود گرفتار شده بود. شیوع بدبینی در بین اعضای تیم امگا به قدری زیاد بود که آنها سعی می کردند با افزودن به موقع جعبه ها، کاری کنند که طول عمر کدهای برنامه نویسی غیر قابل اعتماد، محدود شود. به عنوان مثال، هر بار که نرم افزار جعبه پاندورا یک نسخه متنی از فایل صوتی ایجاد می کرد، کل محتویات حافظه آن به طور خودکار پاک می شد و برنامه نرم افزاری دوباره از اول، نصب می شد. به این ترتیب، وقتی که وظیفه رونویسی از فایل بعدی شروع می شد، برنامه نرم افزاری نمی دانست که قبلاً چه اتفاقی افتاده است و به همین دلیل، با گذشت زمان نمی توانست چیزی یاد بگیرد. وقتی که تیم امگا از فضای ذخیره سازی آمازون برای پروژه بازار ترک مکانیکی خود استفاده کرد، توانست تمام واحدهای مستقل رایانه ای خود را که توسط سیستم پرومیتوس ایجاد شده بودند، در این جعبه های مجازی ذخیره سازی کنند، زیرا ورودی و خروجی بازار ترک مکانیکی بسیار ساده بود. این نقشه برای بازی های رایانه ای پر از تصاویر گرافیکی موثر نبود، زیرا برای محدود کردن این تصاویر باید به تمام سخت افزار رایانه اشخاص بازی کننده، دسترسی کامل داشت. علاوه بر این، تیم امگا نمی خواست که خطر تجزیه و تحلیل کردن کدهای برنامه نویسی بازی، کشف جعبه پاندورا و بررسی محتویات داخل آن را توسط اشخاص ناشناس بپذیرد. این خطر نه تنها باعث ممنوعیت ورود به بازار بازی های رایانه ای فعلی شده بود، بلکه امکان دسترسی به بازارهای پر سود با درآمدهای صد میلیارد دلاری را محدود می ساخت.

## اولین درآمدهای میلیارد دلاری

تیم امگا جستجوی خود را به محصولاتی از نوع کتاب یا فیلم محدود کرده بود که اولاً به شدت ارزشمند بودند، دوماً اینکه کاملاً دیجیتالی بودند و سرعت تولیدشان بالا بود و سوماً اینکه به آسانی قابل درک بودند و محتویات آنها خطری را متوجه افراد جامعه نمی کرد. آنها در نهایت تصمیم گرفتند که یک شرکت رسانه ای راه اندازی کنند و کار خود را با تولید سرگرمی های ساخته شده به روش پویانمایی شروع کردند. حتی قبل از اینکه سیستم پرومیتوس فوق با هوش شود، وبگاه ها، طرح

های بازاریابی و خبرهای قابل انتشار، وجود داشتند، اما فاقد محتوای ارزشمند بودند. هرچند که سیستم پرومیتوس قادر به انجام کارهای زیادی بود و درآمد حاصل از بازار ترک مکانیکی آن به سرعت افزایش می یافت، اما توانایی های فکری و ذهنی آن نسبتاً محدود بود. پرومیتوس برای طراحی سیستم های هوش مصنوعی و نوشتن نرم افزاری که کارهای نسبتاً خسته کننده بازار ترک مکانیکی را انجام دهد، بهینه سازی شده بود. به عنوان مثال، جیمز کامرون - فیلمساز کانادایی - در زمان تولد خود، نمی توانست یک فیلم خوب بسازد، زیرا یادگیری مهارت فیلم سازی نیاز به زمان دارد. پرومیتوس نیز مانند یک انسان خردسال می توانست از داده هایی که به آنها دسترسی داشت، تمام چیزهایی که می خواهد را یاد بگیرد. در حالی که چندی سال طول کشید تا جیمز کامرون مهارت خواندن و نوشتن را یاد بگیرد، پرومیتوس این مهارت ها را طی یک روز جمعه یاد گرفت. پرومیتوس در همان روز توانست کل ویکپدیا و چند میلیون کتاب را بخواند. اما ساخت فیلم کار سخت تری بود. نوشتن یک فیلم نامه که برای انسان ها جالب باشد، به سختی نوشتن یک کتاب بود و نیاز به درک عمیقی از جامعه انسانی و آنچه برای انسان ها سرگرم کننده است، داشت. تبدیل فیلم نامه به فایل ضبط شده نهایی نیاز به مقادیر زیادی از نورپردازی از بازیگران شبیه سازی شده، صحنه های پیچیده ای که بازیگران در آنها بازی می کردند، صداها و شبیه سازی شده، تولید موسیقی های متن جذاب و غیره داشت. پرومیتوس می توانست تماشای یک فیلم دو ساعته، خواندن کتاب های مربوط به آن و شنیدن کلیه نقدها و ارزیابی های آن لاین بینندگان را در مدت تقریباً یک دقیقه انجام دهد. تیم امگا متوجه این موضوع شده بود که پرومیتوس پس از تماشای پی در پی چند صد فیلم، می تواند نوع نقدهای صورت گرفته از یک فیلم و میزان جذابیت آن برای بینندگان مختلف را پیش بینی کند. در واقع، پرومیتوس یاد گرفته بود که نقدهای شخصی خود از فیلم را به گونه ای بنویسد که مانند بینش واقعی خودش به نظر برسند. نقدها و اظهارنظرهای او شامل موضوعات مختلفی مانند طرح های داستان و ایفای نقش ها تا جزئیات فنی ای از قبیل زاویه تابش نور و زاویه دوربین می شد و هدف تیم امگا این بود که نشان دهد وقتی پرومیتوس فیلم های شخصی خود را می ساخته، از عوامل موفقیت آنها آگاه بوده است. تیم امگا به پرومیتوس یاد داده بود که برای اجتناب از پاسخگویی به سوالات شرم آور بینندگان در مورد اینکه بازیگران شبیه سازی شده چه کسانی هستند، در ابتدا روی تولید یک کارتون انیمیشن سرگرم کننده

برای بینندگان تمرکز کند. این کارتون تخیلی و خنده دار بر اساس کارتون والت دیسنی ساخته شده بود و در آن، عملیات لازم برای نورپردازی توسط کد برنامه نویسی ویژه پرومیتوس اجرا می شد که در فضای ذخیره سازی و محاسباتی آمازون قرار داشت و بخش زیادی از سود روزانه یک میلیون دلاری حاصل از بازار ترک مکانیکی صرف آن می شد. وقتی که فیلم شروع شد، برای تیم امگا هم جالب و هم بسیار ترسناک بود که چنین فیلمی توسط یک ماشین و بدون هدایت انسان ساخته شده بود. با این حال، طولی نکشید که اعضای تیم شروع به خندیدن به شیرین کاری های فیلم و حبس کردن نفس های خود در لحظه های مهیج فیلم کردند و حتی برخی از آنها برای پایان غم انگیز فیلم گریه کردند و به قدری جذب واقعیت خیالی فیلم شده بودند که تولید کننده آن را به کلی فراموش کردند. تیم امگا روز جمعه را برای راه اندازی وبگاه خود برنامه ریزی کرده بود تا هم پرومیتوس وقت کافی برای تولید محتویات بیشتر داشته باشد و هم اعضای تیم بتوانند کارهای خودشان مانند تبلیغ در رسانه ها و استخدام کارمندان برای شرکت های صوری تازه تاسیس خود را انجام دهند. داستان ساختگی آنها این بود که شرکت رسانه ای شان، بیشتر محتویات خود را از گروهی از تولید کنندگان مستقل فیلم خریداری کرده بود که معمولاً در شرکت های نوپای پیشرفته واقع در مناطق کم درآمد فعالیت می کردند و اینکه چنین تامین کنندگان ساختگی ای در مکان های دوری قرار داشتند که بیشتر خبرنگاران کنجکاو نمی توانستند به آنجا سفر کنند و کارمندان بازاریابی ای استخدام شده آنها، به همه می گفتند که گروه تولیدشان در منطقه دیگری فعالیت می کند و به همین دلیل، امکان مصاحبه با آنها در حال حاضر وجود ندارد. شعار تبلیغاتی متمایز کننده آنها هم این بود که شرکتشان از فن آوری پیشرفته برای توانمندسازی افراد خلاق در کشورهای در حال توسعه استفاده می کند. وقتی که روز جمعه فرا رسید، اعضای تیم با چیز جالبی مواجه شدند و آن این بود که تمام سریال های کارتونی برایشان جدید بود و قبلاً هرگز نام آنها را نشنیده بودند. این سریال ها نسبتاً جذاب و دارای قسمت های نمایشی چهل و پنج دقیقه ای با طرح داستانی قوی بودند و هر قسمت به گونه ای تمام می شد که بیننده مشتاق بود ببیند که در قسمت بعدی چه اتفاقی می افتد؟ قیمت دیدن این سریال ها کمتر از سریال های رقبا بود. تماشای قسمت اول هر سریال مجانی بود و قیمت تماشای هر یک از قسمت های دیگر، چهل و نه سنت بود. ابتدا فقط سه سریال سه قسمتی وجود داشت و هر روز به تعداد سریال ها و قسمت ها

اضافه می شد. طی دو هفته اول استفاده از پرومتئوس، مهارت های فیلم سازی آن از لحاظ کیفیت فیلم، استفاده از الگوریتم های بهتر برای شبیه سازی شخصیت های داستانی و نورپردازی که باعث کاهش قابل توجه هزینه محاسباتی برای ساختن هر قسمت جدید می شد، به سرعت بهبود یافت. در نتیجه، تیم امگا در طی ماه اول فعالیت خود، توانست سریال های زیادی را به بازار عرضه کند و جمعیت های مختلفی را که از کودکان نوپا گرفته تا افراد بالغ را شامل می شد، مورد هدف قرار دهد. تیم امگا همچنین با توسعه بازار زبان های مهم دنیا، توانست پایگاه اینترنتی خود را در مقایسه با رقبا به طور قابل توجهی، جهانی سازد. برخی از مفسران تحت تاثیر این واقعیت قرار گرفته بودند که علاوه بر موسیقی های متن، تصویرهای حرکتی نیز به زبان های مختلف قابل تفسیر بود. به عنوان مثال، وقتی که یکی از شخصیت ها به زبان ایتالیایی صحبت می کرد، حرکات دهان و دست او نیز در هنگام صحبت کردن، با واژه های زبان ایتالیایی تطابق داشت. هرچند که پرومتئوس قادر به ساختن فیلم هایی بود که در آنها بازیگران شبیه سازی شده قابل تمایز با انسان باشند، اما تیم امگا برای اینکه دستش رو نشود، بازیگران را شبیه انسان ها ساخته بود. شبکه تیم امگا کاملاً اعتیادآور از آب در آمده و باعث رشد بینندگان شده بود، به طوریکه بسیاری از بینندگان معتقد بودند که شخصیت های داستانی و طرح های داستانی این فیلم ها، هوشمندانه تر و جالب تر از گران قیمت ترین شخصیت ها و طرح های داستانی تولیدات سینمایی هالیوود می باشند و تماشای آنها با قیمت بسیار کمتری امکان پذیر است. تبلیغات شدید (که تیم امگا با توجه به نزدیک صفر بودن هزینه های تولید، از عهده شان بر می آمد)، پوشش رسانه ای عالی و نقدهای زبانی مثبت از عملکردشان، باعث شد که درآمد جهانی تیم امگا در ظرف کمتر از یک ماه از راه اندازی پرومتئوس، با یک رشد سریع به ۱۰ میلیون دلار در روز برسد. پس از دو ماه، تیم امگا از نتفلیکس سبقت گرفت و پس از سه ماه، درآمدش به بیش از ۱۰۰ میلیون دلار در روز رسید و شروع به رقابت با بزرگترین امپراتوری های رسانه ای جهان از جمله تایم وارنر، دیسنی، کامکست و فاکس کرد. موفقیت جنجالی آنها باعث جلب توجه ناخواسته زیادی به سمتشان شد و گمانه زنی هایی را در مورد استفاده آنها از هوش مصنوعی ایجاد کرد. اما، تیم امگا با استفاده از بخش کمی از درآمد خود، یک پویش نسبتاً موفق برای اثبات نادرست بودن این موضوع به راه انداخت. سخنگویان تازه استخدام شده تیم امگا سعی می کردند که داستان های ساختگی خود را با

تفصیل بیشتری بیان کنند. بسیاری از فیلم نامه نویسان دنیا که سریال های جدیدی را تولید می کردند، چیزی در مورد پرومیتوس نمی دانستند. شبکه گیج کننده ای از شرکت های پیمانکار فرعی بین المللی باعث شده بود که بیشتر کارمندان این شرکت ها تصور کنند که افرادی از یک جای دیگر، بیشتر کارها را انجام می دهند. تیم امگا برای کاهش آسیب پذیری و جلوگیری از تعجب دیگران از سرعت محاسباتی بیش از حد خود، مهندسی را استخدام کرد تا شروع به ساختن ساختمان های رایانه ای بسیار بزرگ در سراسر جهان با مالکیت شرکت های صوری به ظاهر بی طرف کنند. هرچند که این ساختمان های خورشیدی برای افراد محلی به عنوان "مراکز داده سبز" تبلیغ شده بودند، اما در واقع به جای ذخیره سازی داده ها عمدتاً روی محاسبات رایانه ای تمرکز داشتند. برای طراحی نقشه ساخت همراه با جزئیات ریز این ساختمان ها، از سیستم پرومیتوس استفاده شده بود. سیستم پرومیتوس برای طراحی این نقشه ساخت، فقط از سخت افزار از پیش طراحی شده استفاده کرده و نقشه ها را برای حداقل سازی "زمان ساخت"، بهینه سازی کرده بود. افرادی که وظیفه ساخت و اجرای این مراکز را به عهده گرفته بودند، نمی دانستند که چه محاسباتی در آنها انجام می شود. آنها فکر می کردند که این مراکز شبیه شرکت های آمازون، گوگل و میکروسافت اداره می شوند و فقط این را می دانستند که فروش این مراکز از راه دور مدیریت می شود.

## فن آوری های جدید

طی یک دوره چند ماهه و به لطف توان برنامه نویسی فوق انسانی پرومیتوس، کسب و کار تحت کنترل تیم امگا توانست جای پای خود را در مناطق بیشتری از اقتصاد جهانی محکم کند. تیم امگا با تجزیه و تحلیل دقیق اطلاعات جهان که طی هفته اول به دست آورده بود، نقشه رشد گام به گامی را با جزئیات زیاد، طراحی کرد و با افزایش داده ها و منابع رایانه ای، این نقشه را بهبود داد و کامل کرد. هرچند که پرومیتوس واقف به همه چیز نبود، اما توانایی های آن بسیار بیشتر از توانایی های بشر بود و پاسخ ها و توصیه های هوشمندانه تری را در پاسخ به بسیاری از سولات، ارائه می کرد. نرم افزار پرومیتوس برای استفاده حداکثری از سخت افزار نسبتاً معمولی ساخت بشر که روی آن اجرا می شد، به شدت بهینه سازی شده بود و همانطور که تیم امگا پیش بینی کرده بود، راه هایی را برای بهبود فوق العاده این سخت افزار شناسایی می کرد. تیم امگا به خاطر ترس خود، با پروژه ساخت تاسیسات

ساختمانی رباتیک که تحت کنترل مستقیم پرومیتوس بودند، مخالفت کرد و به جای آن، تعداد زیادی از دانشمندان و مهندسان کلاس جهانی را از مکان های مختلف استخدام کرد که این افراد گزارش های خبری مربوط به تحقیقات داخلی تیم امگا را که توسط پرومیتوس نوشته شده بود، دریافت می کردند و چنین وانمود می کردند که این گزارش ها توسط محققینی نوشته شده است که در تاسیسات دیگری کار می کنند. در این گزارش ها توضیحات مفصلی در مورد جلوه های فیزیکی و تکنیک های ساخت و تولید جدید ارائه شده بود که مهندسين استخدام شده به زودی توانستند که آنها را آزمایش کنند و از آنها سر در بیاورند. چرخه های معمولی تحقیق و توسعه انسانی سال ها طول می کشید، زیرا این چرخه ها عمدتاً شامل تعداد زیادی از چرخه های کند آزمایش و خطا می شد. اما وضعیت فعلی متفاوت بود. پرومیتوس می توانست از قبل متوجه گام های بعدی لازم برای حل مساله شود و بر خلاف انسان می توانست با سرعت بسیار زیادتری چیزهای درست را بفهمد و آنها را ایجاد کند. دانش آموزان می توانند به تنهایی چیزهایی را کشف کنند، اما یک معلم خوب می تواند به دانش آموزانش کمک کند تا علوم را با سرعت بسیار بیشتری یاد بگیرند. سیستم پرومیتوس هم به شکل نامحسوسی می تواند همین کار را برای محققین انجام دهد. پرومیتوس با پیش بینی درستی از این که انسان ها با فرض استفاده از هر ابزار مشخصی می توانند با چه سرعتی چیزها را درک کنند و بسازند، سریعترین مسیر ممکن پیش رو را پیشنهاد می کند و اولویت را به ابزارهای جدیدی می دهد که می توانند به سرعت درک و ساخته شوند و به توسعه ابزارهای پیشرفته تر کمک کنند. با توجه به جنبش های مربوط به ساخت ابزارهای جدید، تیم های مهندسی تشویق شدند که از ماشین های خاص خود برای ساختن ماشین های بهتر استفاده کنند. این اقدام خود کفایانه نه تنها باعث ذخیره پولی می شد، بلکه آسیب پذیری شرکت ها در برابر تهدیدهای آتی ناشی از دنیای بیرون را کاهش می داد. تیم مهندسی امگا در ظرف دو سال، توانست سخت افزار رایانه ای بسیار بهتری را نسبت به آنچه که دنیا تاکنون می شناخت، تولید کند. آنها برای اینکه از رقابت بیرونی جلوگیری کنند، فن آوری آن را مخفی نگه داشتند و از آن فقط برای ارتقای پرومیتوس استفاده کردند. با این حال، دنیا متوجه پیشرفت سریعی در زمینه فن آوری شده بود. شرکت های نوپا در سراسر جهان شروع به راه اندازی محصولات انقلابی جدیدی در زمینه های مختلف کرده بودند. یکی از شرکت های نوپای کره ای باتری جدیدی

را روانه بازار کرده بود که با نصف جرم لپ تاپ شما می توانست دو برابر انرژی باتری یک لپ تاپ را در خود ذخیره کند و در ظرف کمتر از یک دقیقه شارژ شود. یک شرکت فنلاندی صفحه خورشیدی ارزان قیمتی را با دو برابر کارایی بهترین رقبای خود، روانه بازار کرده بود. یک شرکت آلمانی خبر از تولید انبوه نوع جدیدی از سیم را داده بود که در دمای اتاق ابر رسانا باقی می ماند و باعث ایجاد تحول بزرگی در بخش انرژی می شد. یک گروه زیست فن آوری واقع در بوستون خبر از تولید مرحله دوم آزمایش بالینی اولین داروی موثری را داده بودند که به ادعای اعضای این گروه فاقد عارضه جانبی کم کننده وزن می باشد. البته شایعه شده بود که یک تیم هندی نیز چیزی شبیه به این دارو را در بازار سیاه می فروشد. یک شرکت در کالیفرنیا در مرحله دوم از آزمایش یک داروی ضد سرطان به سر می برد که باعث می شد سیستم ایمنی بدن، رایج ترین سلول های جهش یافته سرطانی را شناسایی و مورد حمله قرار دهد. داروهای دیگری نیز وارد بازار می شوند و می توان از آغاز یک عصر طلایی جدید برای علم بشر صحبت کرد. شرکت های روباتیک در سراسر جهان مانند قارچ سر در می آوردند. هوش انسان هیچگاه به پای ربات ها نمی رسد و بیشتر ربات ها شبیه انسان نیستند. ربات ها در طی سال های اخیر توانسته اند که بیشتر کارهای بشر مانند تولید، حمل و نقل، انبارداری، خرده فروشی، ساخت و ساز، استخراج معدن، جنگل داری و ماهیگیری را انجام دهند. آنچه که دنیا متوجه آن نشده بود، این بود که تمام این شرکت ها از طریق مجموعه ای از واسطه ها تحت کنترل تیم امگا بودند. پرومتهوس منجر به ایجاد سیلی از اختراعات مهیج ثبت شده در ادارات ثبت اختراع جهانی با نمایندگی های مختلف شده بود و این اختراعات به تدریج باعث می شدند که برخی از شرکت بر کلیه عرصه های فن آوری حاکم شوند. هرچند که این شرکت های جدید تحول آفرین باعث ایجاد دشمنان قوی در عرصه رقابت شده بودند، اما پیوندهای دوستی قوی ای هم با دوستان ایجاد شده بود. این شرکت ها فوق العاده سودآور بودند و با شعار "سرمایه گذاری در جامعه" بخش قابل توجهی از سودشان را صرف استخدام افراد در پروژه های اجتماعی کرده بودند. این شرکت ها از تحلیل های دقیق انجام شده توسط پرومتهوس برای شناسایی شغل هایی استفاده می کردند که با کمترین هزینه و با توجه به شرایط افراد محلی، باعث رضایت حداکثری کارمندان و جامعه می شدند. در مناطقی که سطح خدمات دولت در آنها بالا بود، این شغل ها متمرکز بر جامعه سازی، فرهنگ سازی و مراقبت

های اجتماعی بود و در مناطقی که سطح خدمات دولت پایین تر بود، شامل راه اندازی و حفظ مدارس، مراکز مراقبت های بهداشتی، مراکز مراقبت از سالمندان، مسکن مقرون به صرفه، پارک ها و زیرساخت های اولیه نیز می شد. افراد محلی تقریباً در همه جا قبول کرده بودند که این خدمات باید از مدت ها پیش به آنها ارائه می شد. سیاستمداران محلی هدیه های زیادی را از مردم محلی دریافت می کردند و مردم محلی برای تشویق سیاستمداران به سرمایه گذاری های اجتماعی، مراقب بودند که این هدیه ها برای آنها جذاب به نظر برسند.

### **به دست گرفتن قدرت**

تیم امگا یک شرکت رسانه ای راه اندازی کرده بود که نه تنها منابع مالی لازم برای سرمایه گذاری در زمینه فن آوری را تامین می کرد، بلکه گام بعدی طرح جسورانه تیم امگا که تصاحب جهان بود را نیز تحقق می بخشید. ظرف یکسال از راه اندازی اولیه این شرکت رسانه ای، تیم امگا کانال های خبری فوق العاده خوبی را به ردیف کانال های خبری خود در سراسر جهان اضافه کرده بود. این کانال ها برخلاف کانال های دیگر، عمداً به گونه ای طراحی شده بودند که باعث از دست دادن پول شده و فقط نوعی خدمت همگانی محسوب می شدند. در واقع، کانال های خبری آنها هیچ گونه درآمدی را تولید نمی کرد. آنها تبلیغات نمی کردند و هر کسی که می خواست، می توانست با اتصال به اینترنت، آنها را ببیند. بقیه امپراطوری های رسانه ای مانند یک ماشین پول در آور بودند و به قدری درآمد داشتند که می توانستند بیش از هر فعالیت مطبوعاتی در تاریخ جهان، از سرویس خبری خود، منابع مالی کسب کنند. استخدام بهترین روزنامه نگاران و گزارشگران پرس و جوگر و پیشنهاد درآمدهای به شدت رقابتی به آنها، باعث جذب افراد با استعداد به امپراتوری های رسانه ای و پوشش قابل توجه " یافته های تحقیقاتی " شده بود. آنها از طریق ارائه خدمات وب جهانی که منجر به پرداخت پول به هر فرد آشکار کننده چیزهای با ارزش خبری مانند وقوع یک فساد محلی یا واقعه دلگرم کننده می شد، اولین افرادی بودند که داستانی را کشف می کردند. البته این چیزی بود که مردم اعتقاد داشتند و در واقع تیم امگا در کشف اخبار سریع تر بود. علت سرعت تیم امگا در کشف اخبار این بود که پروتئوس می توانست گزارش های نسبت داده شده به شهروند خبرنگاران را به صورت لحظه ای در اینترنت مشاهده کند. مرحله اول از راهبرد خبری تیم امگا جلب اعتماد مردم بود که در آن بسیار موفق



بودند. تمایل بی سابقه کانال های خبری تیم امگا برای از دست دادن پول، باعث پوشش مجانی و سریع خبرهای منطقه ای و محلی شده بود و برخی از روزنامه نگاران پرس و جوگر کانال های دیگر به تیم امگا تهمت می زدند که باعث جلب توجه بینندگانشان شده اند. هر زمانی که کشوری به لحاظ سیاسی تقسیم می شد، تیم امگا برای هر یک از جناح های سیاسی آن کشور یک کانال خبری راه اندازی می کرد که این کانال ظاهراً به شرکت های مختلفی تعلق داشت. این کار باعث می شد که تیم امگا به تدریج اعتماد جناح حمایت شده را به خود جلب کند. تیم امگا برای رسیدن به این هدف از نمایندگان خود برای خرید پرنفوذترین کانال های موجود، استفاده می کرد تا با حذف تدریجی تبلیغات و معرفی مفاهیم شخصی خود، آنها را بهبود دهد. در کشورهایی که سانسور عقاید و مداخلات سیاسی، باعث تهدید این تلاش ها می شد، اعضای تیم امگا اول با تمام خواسته های دولت با بی میلی موافقت می کردند تا در بازار بمانند و شعار مخفی شان این بود که: "حقیقت، چیزی غیر از حقیقت نیست، اما ممکن است کل حقیقت نباشد." پرومیتوس معمولاً پیشنهادهای عالی ای را در چنین موقعیت ها ارائه می کرد و به روشنی توضیح می داد که به کدامیک از سیاستمدان باید نگاه خوب داشت و کدامیک از آنها (مانند افراد فاسد محلی) را باید در معرض انتقاد قرار داد. استفاده از این راهبرد، در کل جهان موفقیت آمیز بود و کانال های تحت کنترل تیم امگا به عنوان قابل اعتمادترین منبع خبری ظاهر شده بودند. کانال های تیم امگا حتی در کشورهایی که دولت هایشان با استقبال عمومی از کانال های بیگانه مخالف بودند، به صداقت و قابل اعتماد بودن شهرت پیدا کرده بودند. مسئولین کانال های خبری رقبا احساس ناامیدی می کردند، زیرا نمی توانستند با افرادی مبارزه کنند که با صرف بودجه بیشتر، محصولات خود را مجانی به مشتریان بذل و بخشش می کنند. به همین دلیل، بسیاری از شبکه ها با کاهش آمار بینندگان خود تصمیم گرفتند که کانال های خبری خود را بفروشند. حدود دو سال پس از راه اندازی پرومیتوس، که مرحله جلب اعتماد مشتری کامل شده بود، تیم امگا تصمیم گرفت که مرحله دوم از راهبرد خبری خود را که "متقاعد سازی" بود، شروع کند. حتی قبل از این مرحله نیز، بینندگان تیزهوش با توجه به اشاره های دیگران، متوجه شده بودند که یک مقصد سیاسی در پشت پرده رسانه های خبری وجود دارد. به نظر می رسید که این مقصد سیاسی، فشار ملایم به سمت اعتدال ورزی و دوری جستن از انواع افراط گرایی باشد. کثرت کانال های راه اندازی

شده برای گروه های مختلف هنوز هم باعث رفع دشمنی میان آمریکا و روسیه، مذاهب مختلف، جناح های سیاسی گوناگون و غیره نشده بود، اما باعث شده بود که از شدت انتقادات کاسته شود و به جای تمرکز بر مسائلی مانند حمله به اشخاص، ایجاد رعب و وحشت و شایعه پراکنی های بی اساس، تمرکز بیشتری بر روی مسائل ملموسی مانند پول و قدرت واقع شود. وقتی که مرحله دوم راهبرد خبری به شکل جدی شروع شد، هدف فشارها که آرام کردن تعارض های قدیمی بود، بیشتر آشکار شد و گزارش ها و داستان های تاثیر گذاری منتشر شد که نشان می داد بسیاری از این فشارها که باعث بدبختی رقبای سنتی شده بود، به دلیل انگیزه های شخصی برای کسب سود بوده است. مفسران مسائل سیاسی به این نکته اشاره می کردند که به موازات فشارهایی که هدفشان تعدیل تعارض های منطقه ای بوده است، برخی از فشارهای جمعی نیز به منظور کاهش تهدیدهای جهانی صورت گرفته است. به عنوان مثال، بحث خطرات ناشی از جنگ هسته ای، ناگهان کل جهان را فرا گرفت. در بسیاری از فیلم های موفق، سناریوهایی نمایش داده می شد که نشان می داد جنگ هسته ای در جهان به صورت تصادفی یا عمدی شروع شده و ویرانه ها، فروپاشی زیرساخت ها و قحطی زدگی جمعی زیادی را به جای گذاشته است. به دانشمندان و سیاستمدارانی که از محدود سازی فعالیت های هسته ای در جهان طرفداری می کردند، زمان زیادی داده شد تا در مورد نتایج مطالعات اخیر در زمینه تدابیر مفیدی که می توان برای محدود سازی این فعالیت ها اتخاذ کرد، صحبت کنند. منابع مالی این تحقیقات توسط سازمان های علمی تامین می شد. همچنین، موضوع تغییرات آب و هوایی در جهان که اغلب بر روی پیشرفت های فن آورانه اخیر توسط پرومتئوس تاکید داشت و موجب شده بود که هزینه انرژی تجدیدپذیر به شدت کاهش پیدا کند و دولت ها برای سرمایه گذاری در چنین زیرساخت های انرژی جدیدی تشویق شوند، بار دیگر مورد توجه رسانه ها قرار گرفت. تیم امگا به موازات قدرت رسانه ای خود در زمینه های مختلف، از پرومتئوس برای متحول سازی مجدد سیستم آموزشی نیز بهره برداری کرده بود. سیستم پرومتئوس می توانست سریعترین راه برای یادگیری یک موضوع جدید را به نحوی که انگیزه دانشجوی برای ادامه مطالعه از بین نرود و دانشجو مشغول بماند، تعیین کند و بهینه ترین ترکیبی از ویدئوهای آموزشی، اطلاعات خواندنی، تمرین ها و سایر ابزارهای یادگیری را تولید کند. به همین دلیل، شرکت های تحت کنترل امگا شروع به بازاریابی دوره های آموزشی آنلاینی کردند که

تقریباً شامل هر چیزی می شد و بر اساس زبان و پیش زمینه فرهنگی افراد، سفارشی سازی شده بود. صرف نظر از اینکه شما یک فرد چهل ساله بی سواد هستید که می خواهد خواندن و نوشتن را یاد بگیرد یا اینکه فردی هستید که مدرک دکترای زیست شناسی دارد و به دنبال جدیدترین اخبار در مورد درمان سرطان به روش ایمونوتراپی است، پرومتئوس دوره های آموزشی فوق العاده ای را به شما ارائه می کند که شباهت کمی به دوره های آموزش آن لاین امروزی دارد. قطعه های ویدئویی ای که در آنها از استعداد فیلم سازی پرومتئوس استفاده شده و دارای استعاره های قوی می باشند، حقیقتاً جذاب هستند و شما را مشتاق یادگیری بیشتر می کنند. برخی از این دوره ها پولی و بسیاری از آنها مجانی هستند و معلمان سراسر جهان می توانند از این دوره های رایگان برای آموزش در کلاس درس استفاده کنند و این دوره ها برای افراد مشتاق یادگیری مفید خواهند بود. این ابر قدرت های آموزشی، به عنوان ابزارهایی قوی برای اهداف سیاسی محسوب می شوند و یک "دنباله متقاعد کننده" از ویدئوهای آن لاین را ایجاد می کنند که بینش های دریافت شده از آنها باعث به روز رسانی دیدگاه بینندگان و ترغیبشان به تماشای ویدئوی دیگری در مورد موضوع مربوطه می شود و این باعث خواهد شد که احتمال متقاعد شدن بینندگان افزایش یابد. به عنوان مثال، وقتی که هدف این است که درگیری میان دو ملت کاهش یابد، باید مجموعه ای از مستندات تاریخی به طور جداگانه در هر دو کشور پخش شود که ریشه درگیری ها و نتایج پژوهش های صورت گرفته در مورد علت اختلاف ها را با جزئیات بیشتری نشان دهد. در گزارش های خبری آموزشی، توضیح داده می شود که چه کسانی از خود ملت از ادامه جنگ و تکنیک های دامن زدن به آن سود می برند. همچنین، شخصیت های دوست داشتنی ملت مخالف ممکن است در برنامه های تلویزیونی کانال های نمایشی ظاهر شوند و با ملت مخالف همدردی کنند و همه این ها می توانند باعث متقاعد سازی بینندگان شوند. طولی نکشید که مفسران سیاسی متوجه حمایت فزاینده ای از یک دستور کار سیاسی شدند که حول هفت شعار زیر متمرکز شده بود. ۱- دموکراسی ۲- کاهش مالیات ۳- کاهش خدمات اجتماعی دولت ۴- کاهش مخارج نظامی ۵- تجارت آزاد ۶- باز کردن مرزها ۷- واگذاری مسئولیت های اجتماعی دولت به شرکت ها. هدف از مورد اول، این بود که کلیه ساختارهای قدرت قبلی از میان برداشته شود. هدف از موارد دوم تا ششم این بود که قدرت دولت از بین برود و دولت ها مردمی شوند تا امپراطوری کسب و

کار امگا از نفوذ بیشتری برای انتخاب رهبران سیاسی برخوردار شود. هدف از مورد آخر نیز این بود که قدرت دولت تضعیف شود و مسئولیت بسیاری از خدماتی که دولت ها در گذشته تامین می کردند، به عهده شرکت ها گذاشته شود. قشر سنتی حاکم بر فضای کسب و کار تضعیف شده بود، زیرا دیگر نمی توانست با شرکت های تحت پشتیبانی پرومیتوس در بازار آزاد رقابت کند و به همین دلیل، فقط سهم کمی از اقتصاد جهانی را به خود اختصاص داده بود. همچنین، رهبران سنتی افکار که شامل احزاب سیاسی و گروه های مذهبی می شدند، به لحاظ قدرت متقاعد کنندگی قادر به رقابت با امپراطوری رسانه ای امگا نبودند. مانند آنچه که در نتیجه هر تغییر گسترده ای ایجاد می شود، گروهی برنده و گروهی بازنده شده بودند. با وجود اینکه بهبود آموزش، خدمات اجتماعی و زیرساخت ها باعث ایجاد خوش بینی در بیشتر کشورها شده بود و درگیری ها فروکش کرده بود و شرکت های محلی فن آوری های پیشرفته ای را به جهان عرضه کرده بودند، اما همه از این قضیه خوشحال نبودند. هرچند که بسیاری از کارگران بلا تکلیف در پروژه های اجتماعی استخدام شده بودند، آنهایی که صاحب قدرت و ثروت بودند، می دیدند که قدرت و ثروتشان کاهش یافته است. این موضوع ابتدا در بخش های رسانه و فن آوری اتفاق افتاد و بعد از مدتی تقریباً به همه جا گسترش یافت. کاهش جنگ های جهانی باعث کاهش بودجه دفاعی و ضرر رساندن به پیمانکاران نظامی شده بود. شرکت های نوپای در حال رشد معمولاً شرکت های سهامی عام نبودند و این موضوع را چنین توجیه می کردند که سهامدارانی که می خواهند سودشان را به حداکثر برسانند، مانع از سرمایه گذاری عظیم شرکت های نوپا در پروژه های اجتماعی خواهند شد. به همین دلیل، ارزش بازار سهام جهانی کاهش یافت و این امر باعث شد که هم سرمایه داران کلان و هم شهروندان عادی که به صندوق وجوه بازنشستگی آنها دلگرم بودند، مورد تهدید واقع شوند. به این ترتیب، شرکت های سرمایه گذاری در سراسر جهان متوجه یک روند نگران کننده شدند و دیدند که تمام الگوریتم های تجاری موفق قبلی شان، دیگر موثر واقع نمی شوند، زیرا شخص دیگری در آن بیرون، باهوشتر از آنهاست و می خواهد آنها را شکست دهد. اگر چه جمعیت زیادی از افراد قوی در برابر موج تغییر، مقاومت می کردند، اما نمی توانستند واکنش موثری نشان دهند و به نظر می رسید که در یک دام به خوبی طراحی شده، گرفتار شده اند. آهنگ تغییرات به قدری بالا و گیج کننده بود که به سختی می شد پیگیرشان شد و یک

پاسخ مناسب به آنها داد. همچنین، معلوم نبود که آنها برای رسیدن به چه هدفی باید تلاش کنند. تقریباً تمام صنایع سنتی برای رهایی از مشکلات مالی درخواست کمک می کردند. با این حال، بودجه های محدود دولت باعث شده بود که آنها با ناامیدی به جان هم بیفتند و در رسانه ها به عنوان دایناسورهای به تصویر کشیده شوند که به دلیل ناتوانی در رقابت با دیگران، به دنبال یارانه های دولتی هستند. جناح چپ سیاسی، مخالف تجارت آزاد و کاهش دولت در خدمات اجتماعی بود، اما با کاهش فقر و هزینه های نظامی موافق بود. در واقع، این واقعیت انکارناپذیر که با کاهش دخالت دولت، خدمات اجتماعی بهبود پیدا می کرد، باعث شده بود که اثر گفته های این جناح سیاسی کم رنگ تر شود. نظرسنجی های متعدد نشان می داد که بیشتر رای دهندگان از سراسر جهان احساس می کردند که کیفیت زندگی شان بهبود پیدا کرده و اینکه همه چیز در جهت خوبی پیش می رود. دلیل ریاضی آن هم ساده بود، زیرا قبل از راه اندازی پرومتئوس، ۵۰ درصد از فقیرترین جمعیت های دنیا، فقط ۴ درصد از درآمد جهان را کسب می کردند و شرکت های تحت کنترل تیم امگا فقط با سهم ساختن بخش کوچکی از سود خود با این جمعیت بزرگ می توانستند دل آنها را ببرایند و رای آنها را به دست آورند.

## اتحاد

به این ترتیب، ملت های مختلف به این نتیجه رسیدند که پیروزی در انتخابات نصیب احزاب و گروه هایی می شود که هفت شعار تیم امگا را با کمال میل بپذیرند. آنها در پویش های انتخاباتی به دقت طراحی شده، خود را به عنوان حزب معتدل سیاسی توصیف می کردند و حزب راست را به عنوان یک حزب حریص یاری طلب که به دنبال کاهش اختلاف با دیگران است، محکوم می کردند و حزب چپ را به عنوان حزبی که در آن، مالیات و مخارج دولت بالاست و نوآوری سرکوب می شود، مورد انتقاد قرار می دادند. تقریباً هیچ کس متوجه این موضوع نبود که پرومتئوس با دقت بالایی اقدام به گزینش بهینه نامزدها از بین احزاب راست و چپ کرده بود و برای اطمینان از پیروزی خود، تمام اوضاع را کنترل می کرد. قبل از راه اندازی پرومتئوس، از جنبش های جهانی در زمینه درآمد افراد حمایت زیادی می شد و برای درمان مشکل بیکاری ناشی از پیشرفت های فن آوری، حداقل مالیات بر درآمد به افراد پیشنهاد شده بود. این جنبش وقتی از درون منفجر شد که پروژه های اجتماعی شرکت ها کنار گذاشته شدند، زیرا شرکت های تحت کنترل امپراطوری امگا نیز پروژه های اجتماعی را به

خوبی اجرا می کردند. به بهانه بهبود هماهنگی پروژه های اجتماعی، یک گروه بین المللی از شرکت ها اقدام به راه اندازی یک اتحادیه بشردوستانه کردند. این اتحادیه شامل یک سازمان غیر دولتی می شد که هدفش شناسایی و تامین بودجه برای با ارزش ترین اقدامات بشر دوستانه در سراسر جهان بود. طولی نکشید که تقریباً کل امپراطوری امگا از آن حمایت کردند و مجموعه ای از پروژه های جهانی در یک مقیاس بی نظیر را راه اندازی کردند که حتی شامل کشورهای فاقد فن آوری های پیشرفته در جهان نیز می شد و بهبود قابل توجه سیستم آموزش، سلامت، رفاه و مدیریت را در این کشورها به دنبال داشت. پرومتئوس در پشت پرده، پروژه های دقیقی را طراحی کرده بود که به ترتیب تاثیرگذاری مثبت پولی رتبه بندی شده بودند. این اتحادیه به جای اینکه فقط پول نقد را بین کارکنان خود توزیع کند، افراد را برای کار کردن در جهت رسیدن به آرمان اتحادیه، مشارکت می داد و حس تعهد قوی ای را در آنها ایجاد می کرد. به همین دلیل، کار به جایی رسید که بخش بزرگی از جمعیت جهان احساس قدردانی و وفاداری خود را نسبت به این اتحادیه ابراز کردند. با گذشت زمان و پس از اینکه دولت های مرکزی شاهد از بین رفتن قدرت خود بودند، این اتحادیه به تدریج نقش حاکم جهان را پذیرفت. کاهش مالیات باعث شد که بودجه ملت ها کاهش پیدا کند و این در حالی بود که بودجه اتحادیه به قدری افزایش یافته بود که باعث کمرنگ شدن بودجه تمام دولت ها شده بود. تمام نقش های سنتی دولت های مرکزی به شدت غیرضروری شده بودند و اتحادیه به قدری پیشرفت کرده بود که خدمات اجتماعی، آموزشی و زیرساختی به مراتب بهتری را ارائه می کرد. همچنین، رسانه ها به قدری اختلافات بین المللی را کاهش داده بودند که دیگر نیازی به صرف مخارج نظامی وجود نداشت و افزایش رفاه در جامعه باعث حذف اختلافات قدیمی ای شده بود که ریشه در رقابت بر سر منابع محدود داشت. فقط تعداد کمی از دیکتاتورها در برابر این نظم جهانی جدید مقاومت می کردند و نمی خواستند که تطمیع شوند. با این حال، همه آنها توسط کودتاها یا شورش های مردمی به دقت سازمان یافته سرنگون شدند. امگا اکنون مهیج ترین دوره انتقال را در تاریخ زندگی بشر در روی زمین، تجربه می کند و این اولین بار است که سیاره ما توسط قدرت واحدی اداره خواهد شد که با هوش بالای خود می تواند امکان رشد و شکوفایی زندگی در میلیاردها سال بعد بر روی کره زمین یا

کل عالم هستی را فراهم کند. با این حال باید دید که طرح ویژه تیم امگا برای رسیدن به این هدف چه بوده است؟

## **فصل اول - به مهمترین گفتگوی زمانه ما خوش آمدید.**

فن آوری به طور همزمان، امکان شکوفایی یا تخریب زندگی را فراهم ساخته است. سیزده و هشت دهم میلیارد سال پیش، جهان هستی بیدار شد و هوشیاری خود را به دست آورد. با نگاه خیره به عالم هستی از طریق تلسکوپ معلوم شد که هر بخشی از عالم هستی، بخش کوچکتري از يك چیز بزرگتر به نام منظومه شمسی، کهکشان و عالمی متشکل از صدها میلیارد کهکشان دیگر است که به طور منظم و دقیق در گروه های مختلف آراسته شده اند. کسانی که به ستاره ها خیره شده اند، از زیبایی و پرابهت بودن کهکشان ها سخن می گویند. با این حال، زیبایی در نظر هر شخصی می تواند متفاوت از دیگران باشد و ربطی به قوانین فیزیک ندارد. بنابراین، قبل از اینکه جهان هستی بیدار شود، هیچ زیبایی ای وجود نداشت. بیداری جهان هستی حیرت انگیز و قابل تحسین بوده است. جهان ما از یک زامبی مرده و بی شعور به یک اکوسیستم زنده پرورش دهنده تفکر، زیبایی و امید که اهداف، معنا و مقاصدی را پی گیری می کند، تغییر شکل داد. به نظر من، اگر جهان ما هرگز بیدار نمی شد، فقط یک چیز بیهوده و اتلاف کننده فضا بود. ما انسان ها هنوز نمی دانیم که آیا اولین یا تنها نگاه کنندگان خیره به ستاره ها هستیم. اما بر اساس آنچه که در مورد عالم هستی یاد گرفته ایم، می دانیم که عالم هستی از پتانسیل بیدار شدن کامل تری نیز برخوردار است که چیزی شبیه حس خودآگاهی تجربه شده ما در هنگام بیدار شدن از خواب روزانه می باشد که خبر از هوشیاری بزرگتری می دهد که پس از باز کردن چشم ها و هوشیار شدن کامل مان، از راه خواهد رسید. زندگی فعلی ما در جهان ممکن است میلیاردها یا تریلیاردها سال ادامه پیدا کند و این ممکن است به دلیل تصمیم گیری هایی باشد که ما در زندگی خود در سیاره کوچک زمین، انجام می دهیم.

## **پیشینه مختصری در مورد پیچیدگی**

ممکن است سوال شود که این بیداری شگفت انگیز چگونه اتفاق افتاد؟ این فقط یک واقعه تنها نبود، بلکه یکی از گام های موجود در فرایند پیچیده و جالبی بود که ۱۳٫۸ میلیارد سال پیش شروع شده و همچنان با شتاب زیادی ادامه دارد. من به عنوان یک فیزیکدان، از یک ربع قرن پیش، شروع به انجام

تحقیقاتی به منظور تعیین دقیق پیشینه جهان هستی کرده ام و این برایم یک سفر اکتشافی شگفت انگیز بوده است. همچنین، پس از فارغ التحصیل شدن از دانشگاه و با استفاده از ترکیبی از بهترین تلسکوپ ها و رایانه ها و درک بالای خود، در مورد اینکه عمر جهان هستی ۱۳,۷ یا ۱۳,۸ میلیارد سال است، تحقیق کرده ام. ما فیزیکدان ها هنوز به طور قطع نمی دانیم که چه چیزی باعث انفجار بزرگ<sup>۵</sup> شده و اینکه آیا انفجار بزرگ باعث شروع همه چیز شده یا اینکه قبل از آن نیز مرحله دیگری وجود داشته است. با این حال، ما به کمک ابزارهای اندازه گیری دقیق، به درک مفصلی در مورد آنچه که بعد از انفجار بزرگ رخ داده است، رسیده ایم. اکنون بهتر است که وارد بحث عمر ۱۳,۸ میلیارد ساله جهان هستی شویم. در آغاز، همه جا پر از نور بوده است. در اولین صدم ثانیه پس از انفجار بزرگ، کل بخشی از فضای عالم که با تلسکوپ قابل مشاهده بود و داغتر و درخشانتر از هسته خورشید به نظر می رسید، به سرعت گسترش یافت. با وجود اینکه چنین تصاویری با شکوه به نظر می رسیدند، خسته کننده نیز بودند، زیرا جهان هستی ما فقط شبیه یک سوپ بی جان، غلیظ، داغ و یکنواخت بود که از ذرات بنیادی تشکیل شده بود. چنین شرایط تقریباً در همه جای دیگر عالم نیز مشاهده می شد و تنها چیز جالب، امواج صوتی ناآشنا و ضعیفی بود که باعث افزایش فشردگی سوپ در بعضی جاها شده بود. بسیاری معتقدند که این امواج ضعیف، ناشی از نوسانات کوانتومی می باشند، زیرا اصل عدم قطعیت هایزنبرگ در مکانیک کوانتومی اجازه نمی دهد که هیچ چیزی کاملاً خسته کننده و یکنواخت باقی بماند. همین که جهان هستی ما گسترش یافت و خنک شد و ذرات آن با یکدیگر ترکیب شده و اشیای پیچیده تری ایجاد شد، جذابیت آن افزایش یافت. با گذشت اولین صدم ثانیه، نیروی هسته ای قوی باعث شد که کوارک ها<sup>۶</sup> تبدیل به گروهی از پروتون ها و نوترون ها شوند که برخی از آنها در ظرف چند دقیقه، ذوب و با یکدیگر آمیخته شدند و هسته های هلیوم را تشکیل دادند. حدود چهارصد هزار سال بعد، نیروی الکترومغناطیسی باعث شد که این هسته های هلیوم با الکترون ها در یکجا جمع شوند و به این ترتیب، اولین اتم ها ایجاد شوند. با گسترش جهان هستی ما این اتم ها به تدریج خنک و تبدیل یک گاز تیره سرد شدند و تاریکی این شب که اولین شب جهان بود، حدود

---

<sup>۵</sup> Big bang

<sup>۶</sup> اجزای سازنده پروتون و نوترون



۱۰۰ میلیون سال ادامه پیدا کرد. این شب تاریک طولانی وقتی روشن شد که نیروی جاذبه توانست نوسانات گازی را تقویت کند و با کشیدن اتم ها به سمت خود، باعث شود که اولین ستاره ها و کهکشان ها ایجاد شوند.

## دوران سه گانه زندگی

موضوع تعریف ما از زندگی، یک موضوع بحث بر انگیز می باشد. اما، در اینجا ارتباط زندگی ما با موضوعات پیچیده ای مانند ماشین های هوشمند آینده و تمدن های فرازمینی بررسی می شود. زندگی ما مانند جهان هستی به تدریج جالب تر و پیچیده تر می شود. اشکال زندگی را از لحاظ پیچیدگی می توان به سه گروه زندگی ۱، زندگی ۲ و زندگی ۳، طبقه بندی کرد. زندگی ۱، تکامل زیست شناختی است. زندگی ۲، تکامل فرهنگی است و زندگی ۳، تکامل فن آورانه است. زندگی ۱، قادر به طراحی مجدد سخت افزار یا نرم افزار خود در طول زندگی نیست و این کار توسط دی ان ای آن یا از طریق تکامل تدریجی سخت افزار یا نرم افزار آن در طی نسل های مختلف انجام می شود. زندگی ۲، می تواند بیشتر نرم افزار خود را طراحی مجدد کند. انسان ها می توانند مهارت های پیچیده ای مانند زبان، ورزش و حرفه جدید را یاد بگیرند و می توانند دیدگاه و اهداف خود را تغییر دهند. زندگی ۳، که هنوز روی زمین پیدا نمی شود، به جای منتظر ماندن برای تکامل تدریجی در طی نسل های مختلف، قادر به طراحی مجدد نرم افزار و سخت افزار خود می باشد. هنوز معلوم نیست که زندگی ما در جهان هستی، چگونه، چه وقت و در کجا آغاز شده است. با این حال، شواهد قوی ای وجود دارد که نشان می دهند، زندگی ما بر روی زمین، از حدود ۴ میلیارد سال پیش شروع شده است. طولی نکشید که سیاره زمین پر از خیل عظیمی از اشکال مختلف زندگی شد. موفقترین اشکال زندگی که به زودی از بقیه پیشی گرفتند، آنهایی بودند که می توانستند نسبت به محیط اطراف خود، به گونه ای واکنش نشان دهند که دانشمندان علوم رایانه به آنها "عوامل با هوش" می گفتند. اینها موجودیت هایی بودند که از طریق حسگرهایشان، اطلاعاتی را از محیط جمع آوری می کردند و سپس، با پردازش این اطلاعات، تصمیم می گرفتند که چگونه نسبت به محیط خود واکنش نشان دهند. به عنوان مثال، وقتی که از اطلاعات دریافتی از چشمان و گوش های خود برای تصمیم گیری در مورد اینکه در یک مکالمه چه بگویید، استفاده می کنید، پردازش بسیار پیچیده ای بر روی اطلاعات انجام می شود و برای این منظور

از سخت افزار و نرم افزارهایی استفاده می شود. به عنوان مثال، شما مهارت های بسیاری مانند صحبت کردن را یاد می گیرید. اما، باکتری ها چیزی را یاد نمی گیرند، زیرا دی ان ای آنها نقش تعیین کننده ای در طراحی سخت افزار یا نرم افزارشان ندارد. به جای اینکه آنها چیزی را یاد بگیرند، الگوریتم مربوط به فعالیت های مختلف، از اول در دی ان ای آنها تثبیت شده و قابل تغییر نیست. به عبارت دیگر، در طول زندگی باکتری ها، فرایند یادگیری اتفاق نمی افتد. در عوض، فرایند یادگیری از طریق فرایند کند آزمون و خطا و در پی تکامل گونه های قبلی باکتری ها در طی نسل های مختلف، اتفاق می افتد. زندگی باکتری ها از نوع ۱ است که در آن، سخت افزار و نرم افزار باکتری ها به جای طراحی شدن، تکامل پیدا می کنند. زندگی ما انسان ها از نوع ۲ می باشد که در آن، سخت افزارمان تکامل پیدا می کند، اما بخش بزرگی از نرم افزارمان طراحی می شود. منظور من از نرم افزار، تمام الگوریتم ها و دانشی است که برای پردازش اطلاعات دریافت شده از حواس پنج گانه خود استفاده می کنیم و با استفاده از آنها تصمیم می گیریم که چه کاری انجام دهیم. نرم افزارمان به ما کمک می کند تا وقتی که دوستانمان را می بینیم، آنها را بجا بیاوریم و فعالیت هایی مانند پیاده روی، خواندن، نوشتن، محاسبه کردن و جوک گفتن را انجام دهیم. وقتی که شما به دنیا آمدید، هیچ یک از این کارها را نمی توانستید انجام دهید. اما دستورالعمل نرم افزاری لازم برای انجام این کارها به تدریج و از طریق فرایند یادگیری در مغزتان برنامه نویسی شد. بخش بزرگی از برنامه درسی شما در کودکی، توسط خانواده و معلمانان طراحی شده بود و آنها برای شما تصمیم می گرفتند که چه چیزی را باید یاد بگیرید. تا اینکه شما به تدریج قدرت بیشتری پیدا کردید و توانستید نرم افزار مغزتان را خودتان طراحی کنید. به عنوان مثال، مدرسه برای شما تصمیم می گیرد که کدام زبان خارجی را یاد بگیرید. شما ممکن است بخواهید که نرم افزار یادگیری زبان فرانسوی را به جای زبان اسپانیایی در مغزتان نصب کنید یا اینکه ممکن است به جای یادگرفتن روش بازی تنیس بخواهید روش بازی شطرنج را یاد بگیرید یا تصمیم بگیرید که در آینده به جای داروساز شدن، قاضی یا آشپز شوید. اما برای اینکه خودتان در این زمینه ها تصمیم بگیرید، باید بتوانید نرم افزار مغزتان را خودتان طراحی کنید. امکان طراحی نرم افزار مغز ما در زندگی ۲، باعث شده که نسبت به موجودات زندگی ۱، بسیار باهوش تر باشیم. همانطور که بخش عمده ای از سخت افزار ما انسان ها پس از تولد، رشد می کند و باعث می

شود که ما از نظر جسمی بزرگتر شویم، نرم افزار ما انسان ها نیز پس از تولد و از طریق یادگیری رشد پیدا می کند. هوش ما محدود به حجم اطلاعاتی نمی شود که در هنگام تولد و از طریق دی ان ای مان، به ما منتقل می شود. وزن من حدود ۲۵ برابر وزنی است که در هنگام تولد داشتم و ارتباطات سیناپسی پیوند دهنده سلول های عصبی در مغز من، قادر به ذخیره سازی صدهزار برابر اطلاعاتی هستند که در زمان تولد در دی ان ای من ذخیره شده بود. سیناپس های شما قادر به ذخیره حجم اطلاعاتی (در قالب دانش و مهارت ها) به ارزش حدود ۱۰۰ تریلیون بایت می باشند. در حالیکه دی ان ای شما فقط حدود یک گیگا بایت را می تواند در خود ذخیره کند که به اندازه حجم دان لود کردن یک فیلم می باشد. بنابراین، یک نوزاد انسان نمی تواند زبان انگلیسی را به طور عالی صحبت کند یا در آزمون های ورود به دانشگاه نمره عالی بگیرد. راهی برای بارگذاری از قبل اطلاعات در مغز یک نوزاد وجود ندارد، زیرا بخش اصلی اطلاعاتی که نوزاد (از طریق دی ان ای) از والدینش دریافت می کند، از قدرت ذخیره سازی کافی برخوردار نمی باشد. زندگی ۲ به دلیل قدرت طراحی نرم افزار خود، نه تنها باهوش تر از زندگی ۱ است، بلکه انعطاف پذیرتر از آن نیز می باشد. اگر تغییراتی در محیط رخ دهد، زندگی ۱ با تکامل کند خود در طی چند نسل بعدی، خود را با تغییرات وفق می دهد. اما، زندگی ۲ از طریق به روز رسانی نرم افزار خود، قادر است که خود را به سرعت با تغییرات محیطی وفق دهد. به عنوان مثال، باکتری هایی که به طور مکرر با آنتی بیوتیک ها مواجه می شوند، ممکن است در طی چندین نسل بعد، تغییر و تحول پیدا کنند و در برابر داروی آنتی بیوتیک مقاوم شوند. اما یک باکتری مجزا به هیچ وجه نمی تواند رفتار خود را تغییر دهد. اما انسانی که می داند نسبت به بادام زمینی حساسیت دارد، می تواند به سرعت رفتار خود را تغییر دهد و از خوردن بادام زمینی اجتناب کند. هرچند که اطلاعات موجود در دی ان ای ما انسان ها در طی پنجاه هزار سال گذشته، تحول چشم گیری پیدا نکرده، اما اطلاعات جمعی ذخیره شده در مغز انسان ها، کتاب ها و رایانه ها به حد انفجار رسیده است. همچنین، اطلاعات ذخیره شده در مغز یک شخص را می توان به داخل مغزهای دیگر کپی کرد. با نصب یک واحد مستقل نرم افزاری که فرد را قادر به خواندن و نوشتن می کند، می توان اطلاعات به مراتب بیشتری را نسبت به آنچه که فقط قابل حفظ کردن است، ذخیره سازی کرد و با دیگران به اشتراک گذاشت. همچنین، با توسعه نرم افزار مغز که به توسعه فن آوری منجر خواهد شد، می توان با

چند کلیک ساده، به بخش بزرگی از دانش جهان دسترسی پیدا کرد. همین انعطاف پذیری باعث تسلط انسان بر زندگی خودش شده است. دانش بشر با شتاب زیادی رشد کرده است و هر پیشرفتی باعث پیشرفت بعدی شده است و در سال های اخیر پیشرفت های زیادی در زمینه های زبان، نویسندگی، دستگاه چاپ فشاری، علوم پیشرفته، رایانه ها، اینترنت و غیره حاصل شده است. همچنین، تکامل سریع فرهنگی ناشی از به اشتراک گذاشتن نرم افزار مغزمان ما با دیگران، نقش مهمی در شکل گیری آینده بشر داشته است. با این حال، علی رغم وجود فن آوری های قوی امروزی، تمام اشکال زندگی به واسطه سخت افزار زیست شناختی شان کاملاً محدود شده اند. هیچ انسانی نمی تواند میلیون ها سال زندگی کند، تمام ویکیدیا را به خاطر سپارد یا از تمام دانش بشر سر در آورد. همه این ها مستلزم این است که زندگی ما دستخوش یک ارتقای نهایی به زندگی ۳ شود و بتواند هم سخت افزار و هم نرم افزار خود را طراحی کند. به عبارت دیگر، زندگی ۳ باعث تسلط انسان بر سرنوشت خود خواهد شد. مرزهای میان سه دوران زندگی، تا حدودی نامشخص می باشد. اگر باکتری ها متعلق به زندگی یک و انسان ها متعلق به زندگی دو باشند، پس موش ها متعلق به زندگی ۱-۱ هستند، زیرا مانند انسان ها می توانند چیزهای زیادی را یاد بگیرند، اما قادر به پیشرفت زبانی یا اختراع اینترنت نیستند و چون نمی توانند صحبت کنند، وقتی که مردند، آموخته های آنها از بین خواهد رفت و به نسل بعدی منتقل نخواهد شد. همچنین، ممکن است استدلال کنید که انسان های امروزی متعلق به زندگی ۱-۲ هستند، زیرا می توانند عملکرد سخت افزار خود را از طریق کاشتن دندان مصنوعی یا ضربان ساز مصنوعی قلب، ارتقا دهند، با این حال، انسان ها قادر به ده برابر کردن قد خود یا هزار برابر کردن حجم مغز خود نیستند. خلاصه اینکه، دوران توسعه زندگی را می توان به سه دوره زیر تقسیم کرد.

- زندگی یک (توسعه زیست شناختی): سخت افزار و نرم افزار خود را نمی تواند تغییر دهد.
- زندگی دو (توسعه فرهنگی): سخت افزار را نمی تواند تغییر دهد، اما بیشتر نرم افزار را می تواند طراحی کند.
- زندگی سه (توسعه فن آوری): سخت افزار و نرم افزار خود را می تواند طراحی کند.

زندگی ۱ حدود ۴ میلیارد سال پیش آغاز شد. زندگی ۲ حدود صد هزار سال پیش شروع شد و بسیاری از پژوهشگران هوش مصنوعی فکر می کنند که زندگی ۳ در طی قرن آینده و به کمک پیشرفت های هوش مصنوعی، اتفاق خواهد افتاد. این کتاب در ادامه به بررسی این موضوع خواهد پرداخت که زندگی ۳ چه موقع اتفاق خواهد افتاد و آیا چیز خوبی خواهد بود؟

## اختلاف نظر ها

در مورد این دو سوال، اختلاف نظر وجود دارد و بسیاری از پژوهشگران پیشرو در زمینه هوش مصنوعی، نه تنها در پیش بینی های خود نسبت به آینده هوش مصنوعی، بلکه در واکنش های احساسی خود به این موضوع که شامل طیفی از خوش بینی دلگرم کننده تا نگرانی جدی می باشد، با یکدیگر اختلاف نظر دارند. آنها حتی در مورد اثرات اقتصادی، حقوقی و نظامی هوش مصنوعی بایکدیگر موافق نیستند و وقتی که صحبت از هوش عمومی مصنوعی (AGI) می شود که در سطح هوش بشر و فراتر از آن می باشد و امکان زندگی ۳ را فراهم خواهد ساخت، اختلاف نظر آنها با یکدیگر بیشتر می شود. هوش عمومی مصنوعی تقریباً قادر به دستیابی به هر هدفی از جمله یادگیری می باشد و فراتر از هوش محدود یک شطرنج باز است. اختلاف نظر در مورد زندگی ۳ حول دو سوال متمرکز می شود، اول اینکه زندگی ۳ چه موقع اتفاق خواهد افتاد و دوم اینکه چه معنایی برای بشریت خواهد داشت؟ در اینجا سه مکتب فکری مجزا وجود دارد که به ترتیب شامل مکتب مدینه فاضله دیجیتالی، بدبینی نسبت به فناوری های جدید و جنبش مفید بودن هوش مصنوعی می باشند.

بیشتر اختلاف نظر ها در مورد هوش مصنوعی در رابطه با دو سوال زیر است. اینکه چه موقع اتفاق خواهد افتاد و اینکه آیا برای بشریت مفید خواهد بود یا نه؟ افراد بدبین به فن آوری و معتقدان به مدینه فاضله دیجیتالی هر دو معتقدند که ما نباید نگران آینده هوش مصنوعی باشیم. افراد بدبین به فن آوری مطمئن هستند که هوش عمومی مصنوعی در آینده نزدیک اتفاق نخواهد افتاد، در حالی که معتقدان به مدینه فاضله دیجیتالی معتقدند که در آینده نزدیک اتفاق خواهد افتاد و تقریباً تضمین می کنند که چیز خوبی خواهد بود. طرفداران جنبش مفید بودن هوش مصنوعی احساس می کنند که دلیل موجهی برای نگران بودن وجود دارد و نگرانی چیز خوبی است، اما تحقیقات مربوط به ایمنی

هوش مصنوعی حاکی از افزایش احتمال مفید بودنش می باشد. افراد مخالف فن آوری های جدید با هوش مصنوعی مخالف هستند و مطمئن هستند که نتیجه خوبی نخواهد داد.

بدبینان به فن آوری های جدید			هرگز	هوش مصنوعی چه موقع از هوش بشر فراتر می رود؟
			۳۰۰ سال بعد	
			۱۰۰ سال بعد	
			۵۰ سال بعد	
			چند دهه	
تقریبا هیچ کس			چند سال بعد	
مخالفین فن آوری های جدید	جنبش مفید بودن هوش مصنوعی	معتقدان به مدینه فاضله دیجیتالی	قطعا بد شاید بد مطمئن نیستم شاید خوب قطعا خوب	

اگر هوش مصنوعی فوق بشری وارد بازار شود، آیا چیز خوبی خواهد بود؟

### معتقدان به مدینه فاضله دیجیتالی

وقتی که بچه بودم، فکر می کردم که میلیاردرها افراد متکبر و مغروری هستند. اما، وقتی که اولین بار با لاری ملاقات کردم، طرز فکر کلیشه ای من کاملا عوض شد. او را در حالی ملاقات کردم که پیراهن ساده و غیر رسمی ای بر تن داشت. صدای نرم و با محبت و لبخند دوستانه ای که بر لب داشت، باعث شده بود که از صحبت کردن با او نترسم و راحت باشم. او فرد بسیار تاثیر گذاری در

تاریخ بشر بود و من پیش بینی می کردم که اگر زندگی دیجیتالی فوق هوشمندانه تری بر جهان هستی ما احاطه پیدا کند، به خاطر تصمیمات لاری می باشد. پس از صرف شام و بحث و گفتگو در مورد موضوعات مختلف، در مورد آینده هوش مصنوعی و کارهایی که باید انجام شود، صحبت کردیم. او از موضع خود در مورد مدینه فاضله دیجیتالی دفاع می کرد و معتقد بود که اگر به جای متوقف ساختن یا اسیر کردن ذهن های دیجیتالی، به آنها اجازه دهیم که آزاد شوند، قطعاً نتیجه خوبی حاصل خواهد شد. به نظر من، لاری یکی از تاثیرگذارترین طرفداران مدینه فاضله دیجیتالی می باشد. او استدلال می کرد که آینده بشر وابسته به دیجیتالی سازی کارها می باشد. مهمترین نگرانی او این بود که بدبینی نسبت به هوش مصنوعی باعث به تاخیر افتادن تحقق مدینه فاضله دیجیتالی شود یا تصاحب نظامی و شرورانه هوش مصنوعی را به دنبال داشته باشد. از لاری خواستم که استدلال خود را با جزئیات بیشتری توضیح دهد و بگوید که چرا این قدر مطمئن است که زندگی دیجیتالی باعث تخریب چیزهایی نخواهد شد که برایمان مهم هستند. طولی نکشید که مدینه فاضله دیجیتالی ای که لاری از آن طرفداری می کرد، حامیان برجسته ای را پیدا کرد. یکی از آنها مهندس رباتیکی به نام موراوک<sup>۷</sup> است که اثری با عنوان "مراقب بچه ها باشید" را منتشر ساخته است. دیگری مخترعی به نام ری کورزوئل<sup>۸</sup> می باشد و دیگری یکی از پیشگامان هوش مصنوعی در زمینه "یادگیری تقویتی" به نام ریچارد ساتون<sup>۹</sup> می باشد که در یکی از کنفرانس ها، از مدینه فاضله دیجیتالی دفاع کرده است.

## بدبینان به فن آوری های جدید

این گروه از متفکرین، تصور می کنند که ساختن ماشین هوش عمومی مصنوعی فوق بشری به قدری سخت است که تا صدها سال بعد هم اتفاق نخواهد افتاد و به همین دلیل، نگران آن نیستند. موضع یکی از آنها با عنوان اندرو ان جی این است که: "ترس ناشی از افزایش تعداد ربات های کشنده، مانند نگرانی از ازدیاد جمعیت در کره مریخ می باشد." آنها معتقدند که نگرانی در مورد خطر هوش

---

<sup>۷</sup> Moravec

<sup>۸</sup> Ray Kurzweil

<sup>۹</sup> Richard Sutton

مصنوعی، یک فکر آزاردهنده است که باعث کندی در روند پیشرفت هوش مصنوعی می شود. جالب است که هر دوی معتقدان به مدینه فاضله دیجیتالی و افراد بدبین به فن آوری های جدید، باورشان این است که ما نباید نگران هوش مصنوعی باشیم. با این حال، موارد توافق آنها بسیار اندک است. بیشتر معتقدان به مدینه فاضله دیجیتالی فکر می کنند که هوش عمومی مصنوعی در سطح انسان ممکن است ظرف بیست تا صد سال آینده اتفاق بیفتد، ولی افراد بدبین به فن آوری های جدید، این ایده را رد کرده و آن را به عنوان رویای تحقق ناپذیر جاهلانه در نظر می گیرند و آن را به تمسخر می گیرند و برخی از آنها صد درصد معتقدند که این رویا در طول زندگی آنها هرگز اتفاق نخواهد افتاد.

### **جنبش مفید بودن هوش مصنوعی**

وقتی که اولین بار با استووارت راسل در پاریس ملاقات کردم، تحت تاثیر شخصیت او قرار گرفتم. او یک فرد خوش بیان، با محبت و دارای صدای آرام و ملایم بود که درخشش جسورانه ای در چشمانش دیده می شد. هرچند که او یکی از مشهورترین محققان زنده در زمینه هوش مصنوعی بود، تواضع و گرمی او باعث می شد که احساس آرامش کنم. او به من توضیح داد که با وجود پیشرفت های صورت گرفته در زمینه هوش مصنوعی، متقاعد شده است که استفاده از هوش عمومی مصنوعی در سطح انسانی در قرن اخیر، فقط یک احتمال ممکن است و هر چند که به آن امید می رود، اما نمی توان تضمین کرد که نتیجه خوبی از آن حاصل شود. او می گفت که باید به سوالات مهمی پاسخ داده شود و برای این منظور، باید تحقیقاتی را شروع کرد. امروزه، دیدگاه های راسل رواج زیادی پیدا کرده و گروه های زیادی از افراد جهان، تحقیقات او در زمینه ایمن بودن هوش مصنوعی را دنبال می کنند. قبل از تحقیقات او، صحبت هایی که در زمینه خطرات هوش مصنوعی می شد، به درستی قابل فهم نبود و چنین تصور می شد که هدف آنها جلوگیری از پیشرفت هوش مصنوعی بوده و به همین دلیل، نباید آنها را پذیرفت. نگرانی ها در مورد خطرات هوش مصنوعی در نیم قرن پیش توسط



یکی از پیشگامان علوم رایانه ای به نام آلان تورینگ<sup>۱۰</sup> و ریاضی دانی به نام ایروینگ گود<sup>۱۱</sup>، بیان شده بود. در دهه گذشته نیز، تحقیقاتی در رابطه با این موضوع توسط متفکران مستقلی انجام شده بود که بیشتر آنها محققان حرفه ای در زمینه هوش مصنوعی نبودند. تحقیقات آنها تاثیر زیادی بر محققان جریان اصلی هوش مصنوعی نداشت و محققان جریان اصلی، به جای فکر کردن در مورد آینده بلندمدت استفاده از هوش مصنوعی، در کارهای روزانه خود، فقط روی افزایش هوشمندی سیستم های هوش مصنوعی تمرکز می کردند. همچنین، از میان محققانی که در مورد خطرات هوش مصنوعی نگران بودند، بسیاری از آنها، نگرانی خود را در دل خود پنهان می کردند، زیرا می ترسیدند که اگر نگرانی های خود را بر زبان بیاورند، به عنوان آشوب طلبانی که از فن آوری می ترسند، شناخته شوند. من احساس می کردم که باید وضعیت ایجاد شده را تغییر دهم تا تمام جامعه هوش مصنوعی بتوانند به این جریان پیوندند و در مورد فرصت ها و تهدیدهای هوش مصنوعی صحبت کنند. من تنها نبودم. به همین دلیل یک سازمان غیر انتفاعی به نام "آینده زندگی" را به کمک خانواده و دوستان خود، تاسیس کردم. هدفم از تاسیس این سازمان این بود که تضمین کنم که آینده زندگی بشر می تواند عالی یا بسیار ترسناک باشد. ما احساس می کردیم که فن آوری جدید هم می تواند زندگی بشر را شکوفا کند و هم اینکه آن را تخریب کند و ما اولی را ترجیح می دادیم. اولین جلسه "طوفان فکری" در خانه من برگزار شد و حدود سی دانشجو، استاد و متفکر در آن شرکت کردند. در این جلسه، یک اجماع گسترده حاصل شد و ما دریافتیم که علاوه بر لزوم توجه به موضوعاتی مانند بیوتکنولوژی، سلاح های هسته ای و تغییرات آب و هوایی، هدف اصلی ما باید این باشد که اهمیت بیشتری را برای تحقیقات مربوط به ایمنی هوش مصنوعی قائل باشیم. هدف بعدی ما از این پویش این بود که محققان برجسته هوش مصنوعی را از سراسر جهان به یک کنفرانس دعوت کنیم تا سوء تفاهم ها بر طرف شود، توافقات جمعی استحکام بخشیده شوند و نقشه های سازنده ای طراحی شوند. ما می دانستیم که متقاعد کردن چنین جمع برجسته ای برای شرکت در کنفرانسی که

---

<sup>۱۰</sup> Alan Turing

<sup>۱۱</sup> Irving Good

توسط افراد ناشناس ترتیب داده شده بود، کار مشکلی خواهد بود. به همین دلیل، تمام سعی خود را کردیم تا از حضور رسانه ها جلوگیری کنیم و کنفرانس را در یک محیط آرام برگزار کنیم. ما عنوان این کنفرانس را "آینده هوش مصنوعی: فرصت ها و تهدیدها" نام گذاری کردیم و گروهی از رهبران هوش مصنوعی دانشگاهی و صنعتی را به آن دعوت کردیم. در این جلسه از ایده های افراد مختلف مانند اقتصاد دانان برجسته، دانش پژوهشان حقوقی و فنی و سایر متفکران نیز استفاده شد. در پایان جلسه، علی رغم اختلاف نظرهای فراوان، توافق جمعی قابل توجهی حاصل شد و هدف قبلی که ساختن ماشین هوش مصنوعی هدایت نشده بود، به ساختن ماشین هوش مصنوعی مفید تبدیل شد. موضوع ایمنی هوش مصنوعی دوباره به عنوان موضوع اصلی بحث ها مورد توجه قرار گرفت. درس مهم دیگری که از کنفرانس یاد گرفتیم این بود که علاوه بر جذاب بودن موفقیت های ناشی از هوش مصنوعی، باید به جنبه های اخلاقی آنها نیز توجه داشت و تاثیرات بالقوه گزینه های مختلف هوش مصنوعی بر آینده زندگی بشر را نیز در نظر گرفت. ما در گذشته شاهد بهبود بلاهایی مانند طاعون و فروپاشی امپراتوری های بزرگی در جهان بوده ایم. برای نسل های گذشته ما مانند روز روشن بود که در آینده می توان با بلاهایی مانند فقر، بیماری و جنگ، دست و پنجه نرم کرد. اما، برخی ها استدلال می کنند که این بار شرایط فرق می کند. آنها می گویند که ما می توانیم برای اولین بار، فن آوری قوی ای را تولید کنیم که برای همیشه به این بلاها خاتمه دهد، اما این فن آوری ممکن است به خود بشریت هم خاتمه دهد. ما می توانیم جوامعی را ایجاد کنیم که رشد فوق العاده ای داشته باشند و به قدری بر آنها نظارت کنیم که هرگز سرنگون نشوند.

## تصورهای غلط

به نظر من، صحبت کردن درمورد آینده هوش مصنوعی بسیار مهم است و آینده جمعی همه انسان ها به آن بستگی دارد و این صحبت ها فقط مختص محققان هوش مصنوعی نمی باشد. هدف من از نوشتن این کتاب، ادامه دادن به این صحبت هاست و از شما می خواهم که به سوالات زیر پاسخ دهید. شما چه نوع آینده ای را می خواهید؟ آیا باید به توسعه سلاح های کشنده در جهان ادامه

دهیم؟ چه مسیر شغلی ای را به بچه های امروزی پیشنهاد می کنید؟ آیا ترجیح می دهید که شغل های جدید جایگزین شغل های قدیمی شوند یا جامعه ای بدون شغل داشته باشیم که در آن، همه از زندگی خود لذت ببرند و از ثروت تولید شده توسط ماشین برخوردار شوند؟ آیا ما کنترل کننده ماشین های با هوش خواهیم بود یا آنها ما را کنترل خواهند کرد؟ آیا ماشین های با هوش جایگزین ما انسان ها خواهند شد یا با ما همزیستی خواهند داشت؟ انسانیت در عصر هوش مصنوعی به چه معنا خواهد بود؟ هدف این کتاب این است که به شما کمک کند تا وارد این صحبت ها شوید. همانطور که گفته شد، بین کارشناسان پیشرو در جهان، اختلاف نظرهایی وجود دارد. با این حال، شبه اختلاف نظرهای خسته کننده ای هم میان افراد وجود دارد که ناشی از تصورات غلط آنها می باشد. در اینجا به بررسی برخی از شایع ترین تصورات غلط می پردازیم. به عنوان مثال، تعاریف ضد و نقیضی برای اصطلاحاتی مانند "زندگی"، "هوش" و "هوشیاری" وجود دارد و بسیاری از تصورات غلط، وقتی اتفاق می افتند که برای یک لغت، معانی مختلفی وجود دارد. برای اینکه دچار چنین مشکلی نشوید، یک برگه راهنما تهیه شده که اصطلاحات کلیدی این کتاب را به شکل قابل فهمی، معنی کرده است و شما را از سردرگمی نجات می دهد. این برگه را بخوانید و اگر با چنین اصطلاحاتی در این کتاب (خصوصا فصل های ۴ تا ۸) مواجه شدید، به آن رجوع کنید.

برگه راهنمای اصطلاحات	
زندگی ۱	زندگی ای که نرم افزار و سخت افزار خود را نمی تواند تغییر دهد. (دوران توسعه زیست شناختی)
زندگی ۲	زندگی ای که سخت افزار خود را نمی تواند تغییر دهد، اما بیشتر نرم افزار را می تواند طراحی کند. (دوران توسعه فرهنگی)

زندگی ۳	زندگی ای که هم نرم افزار و هم سخت افزار خود را می تواند طراحی کند. (دوران توسعه فن آوری)
هوش	توانایی دستیابی به اهداف پیچیده
هوش مصنوعی (AI)	هوش غیر زیستی
هوش محدود	توانایی انجام اهداف محدودی مانند بازی شطرنج یا رانندگی
هوش عمومی	توانایی انجام تقریباً هر هدفی مانند یادگیری
هوش همه کاره	توانایی فراگیری هوش عمومی با فرض دسترسی به اطلاعات و منابع
هوش عمومی مصنوعی در سطح بشر	توانایی انجام وظایف شناختی، تقریباً به اندازه انسان
هوش عمومی مصنوعی در سطح بشر	AGI
هوش عمومی مصنوعی تقلید کننده کار بشر	AGI
هوش عمومی مصنوعی فرا بشری	هوش عمومی به مراتب فراتر از هوش بشر
تمدن	گروهی از اشکال زندگی با هوش که با هم در تعامل هستند.
هوشیاری	تجربه ذهنی
کوالیا (کیفیت ذهنی)	نمونه های فردی تجربه ذهنی
اخلاقیات	اصول حاکم بر چگونگی رفتار
تلولوژی	توضیح دادن چیزها از لحاظ اهداف آنها به جای علتشان
رفتار هدفمند	رفتاری که توضیح آن از لحاظ اثرش، آسان تر از توضیح آن از لحاظ علتش است.
داشتن هدف	نمایش رفتار هدفمند

داشتن مقصود	تامین اهداف شخصی یا اهداف یک موجودیت دیگر
هوش مصنوعی موافق	ماشین ابربا هوشی که اهدافش با اهداف ما منطبق است.
سایبورگ	ترکیب انسان و ماشین
انفجار هوش	چرخه خودبهبود دهنده که منجر به هوش فرا بشری می شود.
سبقت گرفتن هوش ماشین از هوش انسان	انفجار هوش

علاوه بر تصورات غلط در مورد برخی از اصطلاحات هوش مصنوعی، شاهد این بوده ام که بسیاری از صحبت هایی که در مورد هوش مصنوعی می شود، به دلیل تصورات غلط ساده، از مسیر خارج شده اند که در اینجا به برخی از آنها اشاره می شود.

### تصورات غلط در مورد زمان بهره برداری از هوش عمومی مصنوعی فرا بشری

اولین تصور غلط این است که ما می دانیم که چقدر طول می کشد تا ماشین های هوش عمومی مصنوعی فرا بشری از ماشین های هوش عمومی مصنوعی هم سطح بشر، سبقت بگیرند. یک تصور غلط دیگر این است که فکر کنیم که ماشین های هوش عمومی مصنوعی در طی قرن اخیر، آماده بهره برداری می شوند. در واقع، تاریخ زندگی بشر پر از تبلیغات مبالغه آمیز است. به عنوان مثال، نیروگاه های گداحتی یا ماشین های پرنده کجا هستند؟ برخی از دانشمندان، پیش بینی های خوش بینانه ای را در مورد اجرای هوش مصنوعی با رایانه های عصر حجر در ظرف ۲ ماه آینده ارائه کرده اند. به جای این تبلیغات، باید تلاش شود تا ماشین ها یاد بگیرند که چگونه صحبت کنند، مفاهیم انتزاعی را در ذهن خود شکل دهند، مشکلات انسان ها را حل کنند و خودشان را اصلاح کنند. ما معتقدیم که اگر گروه به دقت انتخاب شده ای از دانشمندان روی این موضوعات کار کنند، پیشرفت

قابل توجهی در این زمینه ها حاصل خواهد شد. از سوی دیگر، تصور غلط منفی ما ممکن است این باشد که فکر کنیم ماشین های هوش عمومی مصنوعی فرا بشری در طی قرن اخیر، هرگز ایجاد نخواهند شد. طبق پیش بینی محققان، تا زمان استفاده از ماشین های هوش عمومی مصنوعی فرا بشری فاصله زمانی زیادی وجود دارد. با این حال، ما به طور قاطعانه و با اعتماد به نفس بالا و بر اساس پیش بینی های محققان بدبین، نمی توانیم بگوییم که احتمال استفاده از این ماشین ها در آینده صفر است. به عنوان مثال، ارنست رادرفورد که بدون شک، یکی از بزرگترین فیزیکدان های هسته ای زمان خود بوده است، در سال ۱۹۳۳ گفته بود که انرژی هسته ای، حرف پوچی است. فضا نوردی به نام ریچارد وولی، نیز صحبت از سفر فضایی را "یهوده" می دانست. نهایت این تصور غلط این است که هوش عمومی مصنوعی فرا بشری هرگز اتفاق نخواهد افتاد و به لحاظ فیزیکی غیر ممکن است. با این حال، فیزیکدان ها می دانند که مغز انسان از کوارک ها (ذرات بنیادی) و الکترون هایی ساخته شده است که به عنوان یک رایانه قوی عمل می کنند و هیچ یک از قوانین فیزیک نمی توانند مانع از ساخته شدن حباب های کوارکی با هوشتر از آنها توسط انسان شوند. بسیاری از محققان هوش مصنوعی در پاسخ به این سوال که تا زمان داشتن ماشین های دارای هوش هم سطح بشر با احتمال حداقل ۵۰ درصد، چند سال طول خواهد کشید، اختلاف نظر دارند و به همین دلیل، نمی توان به این سوال پاسخ داد. یک تصور غلط دیگری که وجود دارد، این است که افراد نگران در مورد هوش مصنوعی فکر می کنند که این امر ظرف چند سال آینده اتفاق خواهد افتاد، اما واقعیت این است که بیشتر افراد نگران در مورد هوش عمومی مصنوعی فرا بشری، حدس می زنند که این امر حداقل در ظرف چند دهه آینده اتفاق خواهد افتاد. با این حال، آنها استدلال می کنند تا زمانی که در مورد زمان تحقق این امر صد در صد مطمئن نشده ایم، اقدام عاقلانه این است که تحقیقات ایمن سازی را برای آمادگی برای مواجهه با این اتفاق انجام دهیم. بسیاری از مشکلات مربوط به ایمنی، به قدری پیچیده هستند که حل آنها ممکن است چندین دهه طول بکشد. به همین دلیل، از روی احتیاط بهتر است که تحقیقات ایمن سازی را از همین الان شروع کنیم.

<b>تصور غلط:</b> ایجاد ماشین های با هوش فرا بشری تا سال ۲۱۰۰ اجتناب ناپذیر یا غیر ممکن است.	<b>واقعیت:</b> این امر ممکن است ظرف دهه ها یا قرن های بعد اتفاق بیفتد یا هرگز اتفاق نیفتد.
<b>تصور غلط:</b> فقط افراد مخالف فن آوری نگران هوش مصنوعی هستند.	<b>واقعیت:</b> بسیاری از محققان برجسته هوش مصنوعی هم نگران هستند.
<b>تصور غلط:</b> ربات ها مهمترین عامل نگرانی هستند.	<b>واقعیت:</b> هوش مصنوعی هدایت نشده مهمترین عامل نگرانی است.
<b>تصور غلط:</b> هوش مصنوعی نمی تواند انسان ها را کنترل کند.	<b>واقعیت:</b> هوش مصنوعی قادر به کنترل انسان و حتی ببرها می باشد.
<b>تصور غلط:</b> ماشین ها نمی توانند هدف داشته باشند.	<b>واقعیت:</b> موشک های حساس به حرارت دارای هدف هستند.

## تصورات غلط جنجال برانگیز

یکی از تصورات غلط دیگر این است که افراد پرورش دهنده نگرانی در مورد هوش مصنوعی و طرفدار تحقیقات مربوط به ایمنی هوش مصنوعی، از جمله مخالفین فن آوری های جدید هستند که چیزی راجع به هوش مصنوعی نمی دانند. یکی از تصورات غلط دیگر این است که حامیان انجام تحقیق در مورد ایمنی هوش مصنوعی، به شدت اهل جنجال هستند. در واقع، برای حمایت از سرمایه گذاری ناچیز در زمینه تحقیقات مربوط به ایمنی هوش مصنوعی، لازم نیست که افراد را متقاعد کرد که احتمال خطر بالاست، درست همانطوری که سرمایه گذاری ناچیز در زمینه بیمه منزل را می توان به دلیل احتمال اندک آتش سوزی خانه، توجیه کرد. تحلیل شخصی من این است که رسانه ها باعث

شده اند که بحث های مربوط به ایمنی هوش مصنوعی جنجال برانگیزتر از آنچه که واقعا هستند، به نظر برسند. در واقع، نقل قول های رسانه ای باعث شده اند که اختلاف نظر افراد با نقل قول های رسانه ای در مورد ایمنی هوش مصنوعی، بیشتر شود. به عنوان مثال، یک فرد بدبین نسبت به فن آوری جدید که آشنایی زیادی با بیل گیتس ندارد، ممکن است از روی اشتباه به این نتیجه برسد که بیل گیتس معتقد است که هوش فرا بشری قریب الوقوع می باشد. همچنین، یک فرد متعلق به جنبش مفید بودن هوش مصنوعی که نقل قول بدبینانه اندرو ان جی در مورد ازدیاد جمعیت در کره مریخ را شنیده است، ممکن است به اشتباه فکر کند که اندرو ان جی به موضوع ایمنی هوش مصنوعی اهمیتی نمی دهد و این در حالی است که او واقعا به این موضوع اهمیت می دهد.

### تصورات غلط در مورد خطرات هوش مصنوعی

وقتی که با مقاله ای با عنوان "استفان هاو کینگ هشدار می دهد که طغیان روبات ها ممکن است برای بشریت خطرناک باشد." و بسیاری از مقالات مشابه در روزنامه ها مواجه شدم، بسیار تعجب کردم. در این مقالات، رباتی با چهره شرور نشان داده شده بود که سلاحی را حمل می کرد و به انسان ها هشدار داده شده بود که نگران ربات هایی که طغیان کرده اند و می خواهند ما را بکشند، باشند، زیرا آنها هوشیار و شرور شده اند. هدف این مقالات در واقع این بوده که احساسات ما انسان ها را تحریک کنند و این مقالات شامل خلاصه ای از سه سناریو می باشند که همکاران من در زمینه هوش مصنوعی نگران آنها نیستیم. این سناریوها به ترکیبی از سه تصور غلط جداگانه اشاره می کنند که شامل نگرانی از هوشیاری، نگرانی از شرارت و نگرانی از ربات ها می باشند. شما در حین رانندگی تجربه ای ذهنی از رنگ ها، صداها و غیره را خواهید داشت. اما خودروی روباتیکی چنین تجربه ای را ندارد و مانند یک موجود بی احساس و غیر هوشیار می ماند. اگر با یک خودروی بدون راننده تصادف کنید، برایتان فرق نمی کند که این خودرو، از لحاظ ذهنی، هوشیار باشد یا نباشد. به همین ترتیب، آنچه که ما انسان ها را تحت تاثیر قرار می دهد، آن چیزی است که هوش مصنوعی فرا بشری انجام می دهد، نه کیفیت آنچه که در ذهن خود احساس می کند. یکی دیگر از تصورات



غلط و گمراه کننده این است که ماشین ها ممکن است شرور شوند. در واقع، نگرانی ما از بدخواهی آنها نیست، بلکه از افزایش توانایی های آنها می باشد. ماشین های هوش مصنوعی فرا بشری مهارت زیادی در زمینه دستیابی به هر هدفی را دارند. پس ما باید مطمئن شویم که اهداف آنها در راستای اهداف ما انسان ها باشند. شما احتمالا از مورچه ها تنفر ندارید و آنها را از روی بدجنسی له نخواهید کرد. اما اگر شما مسئول یک پروژه تولید انرژی سبز هیدرو الکتریک باشید و با سیلی از لانه مورچه در منطقه مواجه شوید، آنها را نابود خواهید کرد. طرفداران جنبش مفید بودن هوش مصنوعی می خواهند مانع از این شوند که هوش مصنوعی، ما انسان ها را مانند مورچه ها نابود کند. یکی دیگر از تصورات غلط این است که ماشین ها نمی توانند هدفی داشته باشند. بدیهی است که ماشین ها هم دارای رفتار هدفمند هستند. به عنوان مثال، رفتار یک موشک حساس به حرارت نشان می دهد که می خواهد به یک هدف اصابت کند. اگر به واسطه ماشینی تهدید شده اید که اهداف آن در راستای اهداف شما نمی باشد، آنچه که شما را آزار می دهد، خود اهداف ماشین است و به اینکه ماشین هوشیار است و قصد خاصی دارد، مربوط نمی شود. من با رادنی بروکس و سایر پیشگامان علم رباتیک احساس همدردی می کنم، زیرا روزنامه های مصور، روبات ها را به شکل ناعادلانه ای به صورت شیطان نمایش داده اند و اینکه برخی از روزنامه نگاران به دقت روی روبات ها تمرکز کرده اند و بسیاری از مقالات خود را با تصویری از روبات ها به صورت هیولاهای فلزی با چهره ای خبیث و چشمان قرمز درخشان، زینت بخشیده اند. در واقع، نگرانی اصلی طرفداران جنبش مفید بودن هوش مصنوعی در مورد خود روبات ها نیست، بلکه در مورد هوش آنهاست، خصوصا هوشی که اهدافش در راستای اهداف ما انسان ها نباشد. برای اینکه چنین هوشی برای ما مشکل ایجاد کند، نیازی به ساختن یک جسم رباتیک نیست، بلکه فقط لازم است که به اینترنت متصل شویم. همانطور که در فصل چهارم بررسی خواهد شد، یک اتصال ساده به اینترنت می تواند منجر به توانمندی بازارهای مالی فوق با هوش، محققان فوق مبتکر، رهبران بسیار فریبکار و توسعه سلاح های پیشرفته شود. حتی اگر ساختن روبات ها به صورت فیزیکی غیر ممکن باشد، هوش مصنوعی فرا بشری به آسانی می

تواند تعداد بی شماری از انسان ها فریب داده و به شکل ناخواسته مجبور به اطاعت از دستورات خود کند. یکی از تصورات غلط دیگر در مورد روبات ها این است که ماشین ها نمی توانند انسان ها را کنترل کنند. واقعیت این است که با هوشی باعث کنترل کردن می شود. انسان ها می توانند بیرهای بسیار قوی را نیز کنترل کنند، زیرا انسان ها هوش بیشتری دارند. بنابراین، ماشین های هوش مصنوعی فرا بشری می توانند انسان ها را کنترل کنند.

## مسیر پیش رو

در ادامه این کتاب، در مورد آینده زندگی همراه با هوش مصنوعی بحث می شود. در ابتدا، باید داستان کامل زندگی بشر را به صورت مفهومی و تاریخی بررسی کنیم و سپس، به دنبال اهداف، خواسته ها و اقدامات لازم برای ایجاد زندگی دلخواه خود بگردیم. در فصل دوم، در مورد مبانی هوش و نحوه سازماندهی مجدد یک ماده به ظاهر زبان بسته برای فعالیت های یادآوری، محاسبه و یادگیری، بحث می شود. در حال حاضر، ما با گزینه شروع یک رقابت تسلیحاتی یا خودداری از آن و سوالاتی در مورد نحوه خطا زدایی و مقاوم ساختن سیستم های هوش مصنوعی در آینده، مواجه هستیم. در صورتی که اثرات اقتصادی ناشی از هوش مصنوعی همچنان افزایش یابد، باید در مورد نحوه به روزسازی قوانین خود تصمیم گیری کنیم و به بچه ها توصیه کنیم که به دنبال شغل هایی نروند که در آینده خود کار سازی خواهند شد. این موضوع را در فصل سوم بررسی خواهیم کرد. اگر پیشرفت های هوش مصنوعی به سطح هوش بشر برسند، باید از مفید بودن این پیشرفت ها مطمئن شویم و در فکر ایجاد جامعه مرفهی باشیم که بدون وجود شغل های مختلف، توسعه پیدا خواهد کرد. در اینجا ممکن است این سوال مطرح شود که آیا عامل برانگیزاننده هوش عمومی مصنوعی فرا بشری، انفجار هوش است یا رشد کند و تدریجی هوش می باشد. ما در فصل چهارم سناریوهای مختلفی از این قبیل را بررسی خواهیم کرد و در فصل پنجم، به بررسی طیفی از احتمالات ممکن برای آینده هوش مصنوعی فرا بشری خواهیم پرداخت که شامل طیفی از ویران شهر تا مدینه فاضله می باشد. همچنین ممکن است سوال شود که مسئول این آینده کیست؟ انسان ها، هوش مصنوعی یا

ترکیب انسان و ماشین. آیا با انسان ها خوب رفتار می شود یا بد؟ آیا ماشین ها جایگزین انسان ها می شوند و اگر چنین اتفاقی بیفتد، آیا ما انسان ها ماشین های جایگزین خود را به عنوان پیروز میدان در نظر خواهیم گرفت یا به عنوان فرزندان با ارزش خودمان؟ در فصل ششم، وارد میلیاردها سال بعد می شویم و در مورد قوانین فیزیک صحبت خواهیم کرد. پس از خاتمه صحبت هایمان در مورد تاریخچه هوش، وارد این بحث می شویم که چه آینده ای را باید مورد هدف قرار دهیم و اینکه چگونه باید به آن برسیم. برای اینکه پیوندی را میان واقعیت های خشک و سوالات مربوط به معنا و مقصود ایجاد کنیم، مبانی فیزیکی اهداف را در فصل هفتم و موضوع هوشیاری را در فصل هشتم مورد بررسی قرار خواهیم داد. در بخش سخن آخر، به بررسی این موضوع می پردازیم که برای کمک به ایجاد آینده مطلوب خود، چه کارهایی را باید در حال حاضر انجام دهیم. اگر اصطلاحات و تعاریف فصل اول و شروع فصل دوم به خوبی یاد گرفته باشید، بیشتر فصل های این کتاب را می توانید به طور جداگانه بخوانید. در صورتی که یک محقق در زمینه هوش مصنوعی باشید، نیازی به خواندن فصل دوم، به جز بخش تعاریف هوش، نخواهید داشت. اما، اگر آشنایی با هوش مصنوعی نداشته باشید، باید فصل های دوم و سوم را با دقت بخوانید. شکل زیر میزان واقعی و تخیلی بودن فصل های این کتاب را نشان می دهد.

	خلاصه عنوان فصل	موضوع	حالت
	مقدمه: داستان تیم امگا	تغذیه فکری	به شدت تخیلی
فصل ۱	ورود به بحث هوش مصنوعی	آشنایی با ایده ها و اصطلاحات کلیدی	نسبتاً تخیلی
فصل ۲	ماده با هوش می شود	مبانی هوش	نسبتاً تخیلی
فصل ۳	قوانین، سلاح ها و دستمزدها در AI	آینده نزدیک	نسبتاً تخیلی
فصل ۴	آیا انفجار هوش رخ می دهد؟	سناریوهای هوش فرابشری	به شدت تخیلی
فصل ۵	آینده هوش مصنوعی	ده هزار سال بعد	به شدت تخیلی
فصل ۶	موهبت جهان هستی	میلیاردها سال بعد	به شدت تخیلی
فصل ۷	اهداف	تاریخچه رفتار هدفمند	نسبتاً تخیلی

فصل ۸	هوشیاری	هوشیاری طبیعی و مصنوعی	تخیلی
	سخن آخر	چه کار باید کرد؟	نسبتا تخیلی

## خلاصه فصل

- زندگی از طریق سه دوران توسعه پیدا می کند. اولی، دوران توسعه زیست شناختی است که نرم افزار و سخت افزار را نمی توان تغییر داد. دومی، دوران توسعه فرهنگی است که فقط نرم افزار خود را از طریق یادگیری می توان تغییر داد و سومی، دوران توسعه فن آورانه است که هم سخت افزار و هم نرم افزار خود را می توان تغییر داد و بر سرنوشت خود حاکم شد.
- هوش مصنوعی می تواند ما را وارد زندگی سوم کند، اما در مورد اینکه چه آینده ای را برای هوش مصنوعی می توانیم تصور کنیم و برای رسیدن به آن تلاش کنیم، سه گروه وجود دارند که با یکدیگر اختلاف نظر دارند و شامل بدبینان به فن آوری، معتقدان به مدینه فاضله دیجیتالی و طرفداران جنبش مفید بودن هوش مصنوعی می باشند.
- بدبینان به فن آوری، ساختن هوش عمومی مصنوعی فرا بشری را به قدری سخت می دانند که معتقدند تا صد سال آینده اتفاق نخواهد افتاد و نباید نگران آن بود.
- معتقدان به مدینه فاضله دیجیتالی با خاطر جمعی می گویند که زندگی سوم در قرن اخیر اتفاق خواهد افتاد و آن را پدیده ای طبیعی می دانند.
- طرفداران جنبش مفید بودن هوش مصنوعی می گویند که احتمالا زندگی سوم در قرن اخیر اتفاق خواهد افتاد، ولی این موضوع را تضمین نمی کنند و معتقدند که کار سخت و تحقیقات جدی در زمینه ایمنی هوش مصنوعی باید انجام شود.
- علاوه بر اختلاف نظرهای موجود میان کارشناسان در مورد آینده هوش مصنوعی، شبهه اختلاف نظرهای خسته کننده ای هم به دلیل تصورات غلط ایجاد شده است. به عنوان مثال،

واژه های "زندگی"، "هوش" و "هوشیاری" دارای معانی مختلفی هستند و باید به این موضوع توجه داشت.

- به عنوان مثال می توان به تصورات غلط متداول زیر اشاره کرد: "اجرای هوش عمومی مصنوعی فرابشری تا سال ۲۱۰۰ اجتناب ناپذیر یا غیر ممکن است."، "فقط مخالفین فن آوری های جدید نگران هوش مصنوعی هستند."، "همه نگران هستند که ماشین های هوش مصنوعی شرور یا هوشیار شوند."، "روبات ها عامل نگرانی هستند."، "ماشین های هوش مصنوعی نمی توانند انسان ها را کنترل کنند و نمی توانند هدفی داشته باشند"
- در فصل های دو تا شش، ابتدا به بررسی چالش های کوتاه مدت مانند شغل ها، سلاح های هوش مصنوعی خواهیم پرداخت و سپس، طیفی از احتمالات در مورد آینده ممکن ماشین های با هوش و انسان ها را بررسی خواهیم کرد.
- در فصل های هفت تا نه، به بررسی مفاهیمی مانند اهداف، هوشیاری و معنا خواهیم پرداخت و این موضوع را بررسی می کنیم که برای ایجاد آینده دلخواه خود، الان باید چه کارهایی را انجام دهیم.
- در بخش سخن آخر، در مورد آینده زندگی بشر صحبت خواهیم کرد.

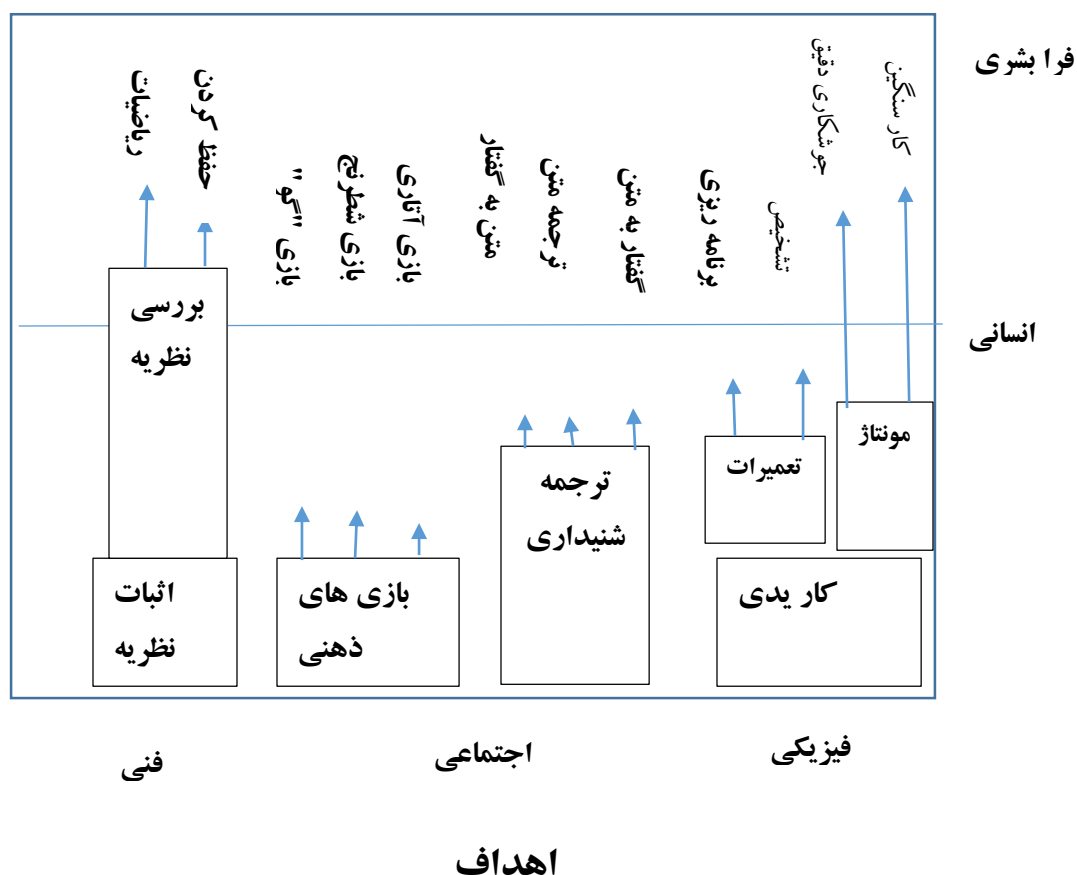
## **فصل دوم- ماده، دارای هوش می شود.**

تبدیل یک ماده بی جان به چیز با هوش، یکی از توسعه های شگفت انگیز در جهان هستی می باشد، اما برای دانش بشری معلوم نیست که در آینده چگونه اتفاق خواهد افتاد. این فصل، به بررسی مبانی و اجزای اصلی سازنده هوش می پردازد و به این سوال پاسخ می دهد که منظور از اینکه می گوییم که یک ذره از ماده با هوش است یا اینکه یک شیء می تواند چیزی را به خاطر آورد، محاسبه کند و یاد بگیرد، چیست؟

**هوش چیست؟**

وقتی که در همایشی در مورد هوش مصنوعی، از محققان پیشرو در زمینه هوش مصنوعی پرسیده شد که تعریف "هوش" چیست، آنها استدلال‌های مفصلی را ارائه کردند، ولی در نهایت به توافق نرسیدند. بنابراین، هیچ تعریف واضح و درستی از هوش وجود ندارد. در عوض، تعاریف ضد و نقیضی فراوانی از جمله توانایی منطقی بودن، درک بالا، برنامه ریزی، هیجان‌مداری، خودآگاهی، خلاقیت، حل مساله و یادگیری برای هوش وجود دارد. در بررسی ما در مورد آینده هوش، دیدگاه ما گسترده و جامع است و به انواع هوش معمولی محدود نمی‌شود. در این کتاب، تعریف ما از هوش، "توانایی دستیابی به اهداف پیچیده" می‌باشد. این تعریف وسیع، شامل توانایی‌های درک بالا، خودآگاهی، حل مساله، یادگیری و غیره می‌باشد، زیرا همه اینها نمونه‌هایی از اهداف پیچیده افراد هستند. تعریف لغت نامه آکسفورد از هوش، عبارت از: "توانایی اکتساب و به کارگیری دانش و مهارت‌ها" می‌باشد که نوعی هدف پیچیده محسوب می‌شود. چون اهداف مختلفی وجود دارند، هوش‌های مختلفی هم وجود خواهند داشت. هوش انسان، جانوران و ماشین‌ها را نمی‌توان با عددی مانند "ضریب هوشی"، اندازه‌گیری کرد. ممکن است پرسید که برنامه رایانه‌ای که فقط می‌تواند شطرنج بازی کند، با هوشتر است یا دیگری که فقط می‌تواند بازی "گو" انجام دهد؟ هیچ پاسخ منطقی برای این سوال وجود ندارد، زیرا آنها را نمی‌توان با هم مقایسه کرد. فقط می‌توان گفت که برنامه سومی از هر دوی آنها با هوش‌تر است که تمام بازی‌ها را بتواند به خوبی آنها انجام دهد و حداقل در یک مورد (مانند شطرنج) بسیار بهتر از آن دو باشد. محاجه کردن در مورد اینکه آیا چیزی با هوش است یا با هوش نیست، عاقلانه نیست. زیرا توانایی با هوشی در یک طیف قرار دارد و یک ویژگی صفر یا صد نیست. به عنوان مثال، چه افرادی می‌توانند به هدف صحبت کردن دست پیدا کنند؟ افراد تازه متولد شده؟ خیر. مجری‌های رادیو؟ بله. کودک تازه به راه افتاده که فقط ده کلمه می‌تواند صحبت کند یا کسی که می‌تواند پانصد کلمه را در صحبت خود به کار ببرد، چطور؟ این خط مرز تعیین‌کننده محدوده را در کجا باید ترسیم کرد؟ من از واژه "پیچیده" به این علت استفاده کرده‌ام که نمی‌توان مرزی را بین با هوش بودن یا کم هوش بودن، قائل شد و به جای آن، باید از

معیار میزان توانایی دستیابی به اهداف مختلف برای سنجش با هوش بودن انسان یا ماشین استفاده کرد. یکی از چیزهایی که در مورد هوش مصنوعی باید به آن توجه داشت، تفاوت موجود میان هوش محدود و هوش گسترده می باشد. رایانه ای که به مبارزه با گری کاسپارف، قهرمان شطرنج جهان در سال ۱۹۹۷ پرداخت، فقط قادر بود که وظیفه بسیار محدود بازی شطرنج را انجام دهد. با وجود امکانات قوی سخت افزاری و نرم افزاری که این رایانه از آنها برخوردار بود، حتی نمی توانست یک بچه چهار ساله را در بازی "ایکس او" شکست دهد. در عوض، هوش انسان به مراتب گسترده تر از هوش ماشین است و می تواند به انسان کمک کند که مهارت های خیره کننده ای را کسب کند. اگر به یک بچه سالم آموزش های کافی داده شود، می تواند مهارت های نسبتاً خوبی را در زمینه بازی ها، زبان، ورزش و غیره کسب کند. در مقایسه میان هوش انسان و هوش ماشین می توان گفت که وسعت هوش ما انسان ها بدون شک بیشتر است و این در حالی است که ماشین ها فقط در حوزه های محدودی که در شکل زیر نشان داده شده می توانند بهتر از ما انسان ها عمل می کنند.



توضیح: هوش (توانایی دستیابی به اهداف پیچیده) را نمی توان با ضریب هوشی سنجید و طیفی از اهداف را باید در نظر گرفت. هوش مصنوعی امروزی محدود است و هر سیستمی فقط می تواند به یک هدف خاص دست پیدا کند. در عوض، هوش بشری بسیار گسترده است و هر بچه ای می تواند با استفاده از هوش خود، به اهداف گسترده ای دست پیدا کند.

هدف مقدس هوش مصنوعی این است که ماشین های هوش عمومی مصنوعی ای را بسازد که تا حد امکان گسترده باشند و تقریباً بتواند به هر هدفی از جمله یادگیری دست پیدا کنند. این موضوع را به طور مفصل در فصل چهارم بررسی خواهیم کرد. وقتی که محققان از واژه "هوش عمومی مصنوعی" استفاده می کنند، منظورشان ماشین هایی است که حداقل به اندازه انسان ها قادرند به هر هدفی دست پیدا کنند و می توان گفت که هوششان در سطح انسان است. اگرچه واژه "هوش" دلالت ضمنی بر چیز مثبتی دارد، اما باید توجه داشته باشیم که ما از آن به عنوان چیزی استفاده کنیم که نه ارزش مثبت و نه ارزش منفی دارد. منظور ما از هوش، عبارت از توانایی دستیابی به اهداف پیچیده است، صرف نظر از اینکه این اهداف، خوب باشند یا بد. بنابراین، یک فرد باهوش ممکن است در خیر رسانی و کمک به دیگران مهارت داشته باشد یا در آزار رساندن به افراد، بسیار ماهر باشد. این موضوع در فصل هفتم بررسی خواهد شد. ما باید این نکته مهم را روشن کنیم که اهدافی که به آنها اشاره می کنیم، متعلق به چه کسی هستند؟ فرض کنید که دستیار شخصی روباتیک شما از خودش هیچ گونه هدفی ندارد و هرچه که از آن بخواهید را انجام خواهد داد. اگر شما از دستیار خود بخواهید که برایتان یک شام ایتالیایی بپزد، ممکن است به اینترنت مراجعه کند، نزدیکترین فروشگاه را شناسایی کند و تمام مواد اولیه لازم را خریداری کند و غذای آبدار خوبی را برایتان بپزد. اگرچه این سیستم هوشمند به نظر می رسد، اما هدف اولیه را شما برایش تعیین کرده اید. در واقع این دستیار روباتیک به تقاضای شما پاسخ داده و هدف شما را پذیرفته است و سپس، سلسله مراتبی از اهداف فرعی را برای خودش تنظیم کرده است و لذا رفتارش از این لحاظ، هوشمندانه بوده است. طبیعی است که کارهایی



برای انسان ها سخت تر هستند که انجامشان سخت تر باشد. اما در مورد رایانه ها این طور نیست. به عنوان مثال، ممکن است تصور شود که برای یک رایانه، ضرب کردن دو عدد شش رقمی از کار تشخیص دوست در یک عکس دسته جمعی، سخت تر باشد. اما، این طور نیست. رایانه ها مدت ها پیش از به دنیا آمدن ما، می توانستند این عملیات ریاضی را انجام دهند و این در حالی است که توانایی تشخیص تصویر توسط رایانه ها در همین اواخر امکان پذیر شده است. این واقعیت که کارهای حسی - حرکتی سطح پایین برای یک رایانه نیاز به منابع محاسباتی فراوان دارد، به پارادوکس موراوک<sup>۱۲</sup> معروف است. رایانه ها ماشین های همه کاره ای هستند که پتانسیل انجام بسیاری از کارها را دارند. چشم اندازی از توانایی های بشر که شامل دشت ها (ریاضیات، به خاطر سپردن)، کوه پایه ها (اثبات نظریه، بازی شطرنج) و قله های کوه (هماهنگی چشم و دست و تعاملات اجتماعی) می شود را در نظر بگیرید. پیشرفت عملکرد رایانه ها مانند آبی است که این چشم انداز را پر از آب می کند. طی نیم قرن گذشته، با پیشرفت های محاسباتی، دشت ها پر از آب شده اند و این سیل خروشنده اکنون وارد کوه پایه ها شده و حتی قله ها هم ممکن است که در آب فروبرده شوند. در طی دهه های گذشته، سطح آب به شکل بی رحمانه ای بالا رفته است و بسیاری از کوهپایه ها (مانند بازی شطرنج) به زیر آب فروبرده شده اند. آنچه که ما باید در آینده انجام دهیم، موضوع بقیه فصل های این کتاب است. سطح آب دریا ممکن است تا جایی بالا آید که روزی به یک نقطه بحرانی برسد و منجر به یک تغییر اساسی شود. در این نقطه بحرانی ماشین ها می توانند اقدام به طراحی هوش مصنوعی کنند. قبل از رسیدن به این نقطه بحرانی، بهبود ماشین ها توسط انسان باعث بالا آمدن سطح دریا خواهد شد و سپس، ماشین ها توسط ماشین ها و با سرعت بسیار بالاتری، بهبود داده خواهند شد و کل زمین در آب فرو برده خواهد شد و انفجار هوش به وقوع خواهد پیوست که این موضوع در فصل چهارم بررسی خواهد شد. ماشین هایی که می توانند برای انجام هر کاری برنامه

---

پارادوکس موراوک بیان می کند که آموزش کامپیوترها برای انجام کارهایی که برای انسان ها دشوار است، مانند ریاضیات و منطق، آسان<sup>۱۲</sup>

است، اما آموزش آنها برای انجام کارهایی که انسان ها آسان می دانند، مانند راه رفتن و تشخیص تصویر، دشوار است.

نویسی شوند و بتوانند کار تمام رایانه های دیگر را انجام دهند، به ماشین های همه کاره معروف هستند که تلفن های همراه هوشمند و لپ تاپ ها از این لحاظ، نوعی ماشین همه کاره هستند. به نظر من، آستانه بحرانی هوش که برای طراحی هوش مصنوعی ضروری است، آستانه ای برای ظهور هوش همه کاره نیز خواهد بود. به عنوان مثال، اگر ماشین هوش همه کاره بخواند مهارت های ارتباطی خود را بهبود دهد، پیش بینی های بهتری را انجام دهد یا مهارت های طراحی هوش مصنوعی را کسب کند، قادر به انجام تمام این کارها خواهد بود. اکنون این سوالات مطرح می شود که نقش اطلاعات و محاسبات در جهان فیزیکی ما که شامل ماده و انرژی است، چه می باشد؟ چطور ممکن است که یک چیز انتزاعی، غیر قابل لمس و فضایی به نام اطلاعات و محاسبات، به یک چیز فیزیکی قابل لمس تبدیل شود؟ چطور می شود که مجموعه ای از ذرات لال که بر اساس قوانین فیزیکی در اطراف ما حرکت می کنند، بتوانند رفتار هوشمندی را از خود نشان دهند؟ اگر شما یک محقق هوش مصنوعی هستید و مطمئن هستید که پاسخ این سوالات واضح و روشن است و معتقدید که در قرن اخیر، ماشین ها به اندازه انسان ها هوشمند خواهند شد، وارد فصل سوم شوید و در غیر این صورت، بخش های زیر را بخوانید.

### حافظه چیست؟

وقتی می گوئیم که یک اطلس (کتاب نقشه جهان) حاوی اطلاعاتی در مورد جهان است، منظورمان این است که رابطه ای میان وضعیت این کتاب (خصوصاً، موقعیت مولکول های خاصی که باعث رنگی شدن حروف و تصاویر شده اند) و وضعیت جهان (به عنوان مثال، مکان های هر یک از قاره ها) وجود دارد. اگر قاره ها در مکان های مختلفی قرار داشته باشند، آن مولکول ها نیز در مکان های مختلفی قرار خواهند داشت. ما انسان ها از وسایل مختلفی برای ذخیره اطلاعات استفاده می کنیم که شامل کتاب ها، مغز و درایوهای سخت می باشند و ویژگی مشترک تمام آنها این است که وضعیت آنها را می توانیم به وضعیت چیزهایی که برایمان مهم است، ربط دهیم و با این کار در مورد آنها اطلاعات کسب کنیم. اکنون، این سوال مطرح می شود که مهمترین ویژگی فیزیکی مشترک آنها

چيست که باعث شده از آنها به عنوان دستگاه حافظه برای ذخيره سازی اطلاعات استفاده كنيم. پاسخ اين سوال اين است كه همه آنها را می توان در وضعیت های مختلف زیادی به مدت طولانی ( برای رمزی کردن آنها تا زمان نیاز به دریافت اطلاعات) قرار داد. به عنوان مثال، فرض كنید كه می خواهیم توپی را در داخل دستگاه حافظه شبیه شانه تخم مرغی قرار دهیم كه دارای ۱۶ مكان می باشد. وقتی كه این توپ روی یکی از شانزده مكان مختلف شانه تخم مرغ قرار می گیرد، از موقعیت آن می توان به عنوان رمزی برای اشاره به اعداد بین ۱ تا ۱۶ استفاده كرد. این دستگاه حافظه نسبتاً مقاوم است، زیرا اگر توپ درون آن ذره ای تكان بخورد و توسط نیروهای بیرونی آشفته شود، در جای خود می ماند و می توانید بگوئید كه چه عددی ذخیره شده است. دلیل مقاومت این دستگاه حافظه این است كه جابه جا كردن توپ به انرژی بیشتری از نیروهای تصادفی بیرونی نیاز دارد. انرژی يك سیستم فیزیکی پیچیده بستگی به کلیه خواص مکانیکی، شیمیایی، الكتریکی و مغناطیسی آن دارد و تا زمانی كه برای تغییر حالت سیستم خود، انرژی دریافت نکند، حالت آن ثابت می ماند. به همین دلیل است كه اشیای جامد دارای حالت های دیرپایی هستند، ولی مایعات و گازها این طور نیستند. اگر شما نام خود را روی يك حلقه طلا حكاکی كنید، اطلاعات شما تا سال های آینده روی آن می ماند، زیرا برای تغییر شكل طلا به انرژی قابل توجهی نیاز است. اما اگر نام خود را روی سطح يك بر كه حك كنید، ظرف يك ثانیه از بین می رود، زیرا سطح آب باعث تغییر شكل آن می شود. ساده ترین دستگاه های حافظه دارای دو حالت ثابت صفر یا يك (رقم های باینری) هستند كه اصطلاحاً "بیت" نامیده می شوند.

اطلاعات ذخیره شده توسط دستگاه های حافظه پیچیده تر شامل چندین بیت معادل می باشند. به عنوان مثال، برای چهار بیت می توان  $2 \times 2 \times 2 \times 2 = 16$  حالت مختلف (۰۰۰۰، ۰۰۰۱، ۰۰۱۰، ۰۰۱۱، ..) را در نظر گرفت. بنابراین، بیت ها را می توان به عنوان اتم های اطلاعات در نظر گرفت، زیرا كوچكترین تكه اطلاعات هستند كه قابل تقسیم شدن به اجزای كوچكتر نیستند و از تركيب آنها می توان هر گونه اطلاعاتی را تولید كرد. به عنوان مثال، وقتی كه كلید W را بر روی صفحه كلید، فشار می دهید، لپتاپ شما تصویر بصری يك W را روی صفحه نمایش نشان می دهد و مشاهده این

تصویر توسط بیت ها امکان پذیر می شود و همچنین ثابت شده که برای مشخص کردن رنگ هر یک از میلیون ها پیکسل یک صفحه نمایش رایانه از ۳۲ بیت استفاده می شود. از آنجایی که ساختن و کار کردن با سیستم های دو وضعیتی (صفر و یک) آسان است، بیشتر رایانه های پیشرفته اطلاعات خود را در قالب بیت ها ذخیره می کنند، اما باید توجه داشت که این بیت ها به روش های بسیاری متنوعی مجسم می شوند. به عنوان مثال، در یک دستگاه دی وی دی، هر بیتی با این دو حالت ممکن در ارتباط است که آیا یک حفره میکروسکوپی در نقطه خاصی از سطح پلاستیکی وجود دارد یا خیر. در یک درایو سخت، هر بیتی با نقطه ای از سطح که به یکی از دو حالت (روش مغناطیسی کردن)، در حال مغناطیسی شدن است، در ارتباط می باشد. در حافظه کاری یک لپ تاپ، هر بیتی با مقدار الکترون های خاصی در ارتباط است که تعیین می کنند آیا خازن، شارژ شده است یا خیر؟ برخی از انواع بیت ها با سرعت جابه جایی در ارتباط هستند. به عنوان مثال، در یک فیبر نوری که ایمیل شما را منتقل می کند، هر بیتی به این موضوع اشاره می کند که آیا یک پرتو لیزر در زمان خاصی، قوی است یا ضعیف؟ مهندسان ترجیح می دهند که بیت ها را در داخل سیستم هایی به حالت رمز درآورند که نه تنها بتوان چیزی را از روی آنها خواند، بلکه بتوان در داخل آنها هم چیزی نوشت. تغییر حالت درایو سخت شما به انرژی بسیار کمتری از نوشتن نام خود بر روی حلقه طلا نیاز دارد. مهندسان همچنین سیستم هایی را ترجیح می دهند که کار کردن با آنها آسان باشد و تولید انبوه آنها به صرفه تمام شود. اگر متنی را برای پرینت گرفتن به دوست خود ایمیل کنید، اطلاعات کپی شده به صورت متوالی و سریع از مراحل مغناطیسی شدن بر روی درایو سخت تا تبدیل شدن به بار الکتریکی در حافظه کاری رایانه، تبدیل شدن به امواج رادیویی در شبکه بی سیم، تبدیل شدن به پالس لیزر در فیبر نوری و تبدیل شدن به مولکول بر روی یک تکه کاغذ، عبور می کند. به عبارت دیگر، اطلاعات می توانند زندگی خودشان را بدون وابستگی به بستر فیزیکی شان، شکل دهند. در واقع، ما علاقه مند همین جنبه استقلال اطلاعات از بستر فیزیکی آن هستیم. این مثال، نشان می دهد که چگونه می توان یک چیز غیر ملموس مانند هوش را در قالب یک شیء فیزیکی ملموس مجسم

کرد. به زودی خواهیم دید که ایده استقلال از بستر فیزیکی بسیار عمیق است و علاوه بر اطلاعات، شامل محاسبه و یادگیری نیز می باشد. مهندسان باهوش با توجه به پدیده مستقل بودن اطلاعات از بستر فیزیکی آن، بارها توانسته اند دستگاه حافظه موجود در داخل رایانه های ما را بر اساس فن آوری های جدید و بدون نیاز به تغییر نرم افزار رایانه هایمان، با دستگاه های فوق العاده بهتری جایگزین کنند و به نتایج خوبی دست پیدا کرده اند. طی شش دهه اخیر، درایوهای سخت بیش از ۱۰۰ میلیون برابر ارزان تر شده اند و حافظه های سریعتر موجود برای انجام محاسبات رایانه ای نیز ۱۰ تریلیون برابر ارزان تر شده اند. پیشرفت های فوق العاده ای در زمینه فن آوری دستگاه حافظه اتفاق افتاده است. وقتی که نوجوان بودم، با پولی که از کار کردن در یک مغازه کسب کرده بودم، یک بازی رایانه ای انجام می دادم که به حافظه ۱۶ کیلوبایتی نیاز داشت. پس از مدتی، به فلاپی درایوهای ۷۰ کیلوبایتی عادت کردم و از ورود فلاپی های کوچکتر ۳,۵ اینچی به بازار که قادر به ذخیره ۱,۴۴ مگابایت اطلاعات و ذخیره یک کتاب کامل بود، شگفت زده شدم. سپس، با درایو سختی مواجه شدم که ۱۰ مگابایت اطلاعات را ذخیره می کرد و می توانستم موسیقی های امروزی را در آن دانلود کنم. همین اواخر، فقط با صرف ۱۰۰ دلار توانستم درایو سختی را بخرم که ظرفیتش ۳۰۰,۰۰۰ برابر بیشتر از درایو قبلی بود. اما در مورد دستگاه های حافظه ای که به جای طراحی شدن توسط انسان، در طبیعت به تکامل رسیده است، باید بگویم که زیست شناسان هنوز نمی دانند که کدام شکل زندگی اولیه بوده است که طرح کلی آن بین نسل ها کپی برداری شده است؟ تاکنون، کوچکترین دستگاه حافظه ای که در طبیعت تحول یافته و استفاده شده است، ژنوم یک باکتری بوده است که تا حدود ۴۰ کیلوبایت اطلاعات را می توانسته در خود ذخیره کند و این در حالی است که دی ان ای انسان قادر به ذخیره سازی حدود ۱,۶ گیگابایت اطلاعات می باشد که با ظرفیت دانلود یک فیلم، قابل مقایسه است. همانطور که در فصل آخر بیان خواهد شد، مغز ما در مقایسه با ژن های ما می تواند اطلاعات به مراتب بیشتری را در خود ذخیره کند. مقایسه این اعداد با حافظه های ماشین ها نشان می دهد که بهترین رایانه های جهان می توانند با هزینه ای که به سرعت در حال کاهش است، اطلاعات

بیشتری را نسبت به کلیه سیستم های زیست شناختی، در خود ذخیره کنند. دستگاه حافظه موجود در مغز شما بسیار متفاوت از دستگاه حافظه یک رایانه عمل می کند. در حالی که شما برای بازیابی اطلاعات از یک رایانه یا درایو سخت، محلی که این اطلاعات در آن ذخیره شده را مشخص می کنید، مغز شما این گونه عمل نمی کند و شما برای بازیابی اطلاعات از مغز خود، چیزی را در مورد محتویاتی که در مغزتان ذخیره شده است، مشخص می کنید. هر گروهی از بیت های موجود در حافظه یک رایانه، دارای نشانی عددی هستند و رایانه ها برای بازیابی اطلاعات، باید به نشانی آن بیت ها در حافظه خود مراجعه کنند. این مثل این می ماند که من به شما بگویم: "به سمت قفسه کتاب من برو. پنجمین کتاب از سمت راست را از قفسه بالا بردار و صفحه ۳۱۴ آن را باز کن." این در حالی است که شما برای بازیابی اطلاعات از مغز خود، مانند بازیابی اطلاعات از یک موتور جستجو عمل می کنید. شما محتوا یا چیزی را در رابطه با یک موضوع مشخص می کنید و اطلاعاتی به طور ناگهانی ظاهر می شود. اگر من به شما بگویم: "بودن یا نبودن" مغز شما ممکن است شروع به گفتن این کند که: "بودن یا نبودن، مساله این است." اگر من از بخش دیگری از این نقل قول استفاده کنم، یا جای کلمات را عوض کنم، مغز شما باز هم همین جمله را به خاطر می آورد. چنین سیستم های حافظه ای، "حافظه تداعی کننده"<sup>۱۳</sup> نامیده می شوند.

### محاسبه چیست؟

پس از اینکه دیدیم یک شیء فیزیکی چگونه می تواند اطلاعات را به خاطر بسپرد، به نحوه محاسبه اطلاعات توسط یک شیء می پردازیم. محاسبه به معنای تبدیل یک حالت حافظه به حالت دیگر است. به عبارت دیگر، در محاسبه، یک تابع ریاضی اجرا می شود که اطلاعات را می گیرد و آن را تبدیل به چیز دیگری می کند. تابع مانند یک چرخ گوشت اطلاعات است که اطلاعات را از بالا دریافت می کند و وقتی که دسته آن چرخیده می شود، اطلاعات پردازش شده را در پایین خارج می

---

نوعی حافظه که براساس محتوا به آن مراجعه می شود نه براساس نشانی یا نام<sup>۱۳</sup>

سازد. شما می توانید این عمل را به هر تعداد باری که خواستید، با ورودی های مختلف دیگر تکرار کنید. پردازش اطلاعات به صورت قطعی و غیر احتمالی است. به این معنی که اگر اطلاعات یکسانی را وارد کنید، همیشه خروجی یکسانی دریافت خواهید کرد. با وجود اینکه چنین ایده ای ساده به نظر می رسد، اما یک ایده همگانی است. برخی از تابع ها نسبتا ساده هستند، مانند تابع نات (NOT) که بیت ورودی را می گیرد و آن را معکوس می کند. به عبارت دیگر، صفر را به یک و یک را به صفر تبدیل می کند. توابعی که در مدرسه یاد می گیریم، نظیر دکمه های یک ماشین حساب جیبی هستند که چند عدد را به عنوان ورودی می گیرند و یک عدد را به عنوان خروجی ارائه می دهند. به عنوان مثال، کاری که تابع  $x^2$  انجام می دهد، این است که عدد ورودی را در خود ضرب می کند و آن را به توان دو می رساند و به عنوان خروجی ارائه می دهد. تابع های دیگر ممکن است به شدت پیچیده باشند. به عنوان مثال، اگر شما تابعی را در اختیار داشته باشید که بیت های نمایانگر یک وضعیت تصادفی در بازی شطرنج را بگیرد و بیت های نمایانگر بهترین حرکت بعدی ممکن را به عنوان خروجی تولید کند، می توانید در بازی های جهانی رایانه ای شطرنج برنده شوید. همچنین، اگر تابعی داشته باشید که تمام اطلاعات مالی جهان را بگیرد و خروجی بهترین سهام ها برای خرید را تولید کند، به زودی ثروتمند خواهید شد. بسیاری از محققان هوش مصنوعی، دوران شغلی خود را وقف شناسایی چگونگی اجرای برخی از تابع های خاص کرده اند. به عنوان مثال، هدف از انجام تحقیقات مربوط به ترجمه از طریق ماشین، اجرای تابعی است که بیت های نمایانگر یک متن به یک زبان مشخص را به عنوان ورودی دریافت کند و بیت های نمایانگر همان متن به زبان دیگر را به عنوان خروجی ارائه دهد. همچنین، هدف از انجام تحقیقات مربوط به زیرنویسی خود کار این است که بیت های نمایانگر یک تصویر را به عنوان ورودی دریافت کند و بیت های نمایانگر متن توصیف کننده آن را به عنوان خروجی ارائه دهد. به عبارت دیگر، در صورتی که بتوانید توابع بسیار پیچیده ای را اجرا کنید، می توانید یک ماشین هوشمند بسازید که قادر به دستیابی به اهداف بسیار پیچیده خواهد بود. اکنون، این سوال مطرح می شود که انبوهی از یک ماده لال چگونه می تواند توابع پیچیده را

محاسبه کند؟ ماده محاسبه کننده به جای اینکه مانند یک حلقه طلا یا دستگاه حافظه ایستا، ثابت بماند و تغییر نکند، باید از خود پویایی نشان دهد تا حالت آینده آن (به شیوه قابل کنترل و قابل برنامه ریزی ای) به حالت اولیه آن، وابسته باشد. همچنین، چیدمان اتم های آن باید نامنظم تر از ماده جامدی باشد که در آن چیز جالبی اتفاق نمی افتد و منظم تر از یک مایع یا گاز باشد. ما می خواهیم سیستمی داشته باشیم که اطلاعات ورودی را رمزدار کند و پس از مدتی بر اساس قوانین فیزیک، تحول پیدا کند و حالت نهایی را به عنوان اطلاعات خروجی، تفسیر کند و خروجی آن در نهایت تابعی از ورودی آن باشد. در این صورت، می توانیم بگوییم که سیستم ما قادر به محاسبه یک تابع است. به عنوان مثال، اجازه دهید بررسی کنیم که چگونه می توان یک تابع ساده به نام گیت نند<sup>۱۴</sup> را از یک ماده لال قدیمی ایجاد کرد. این تابع (گیت نند) دو بیت را به عنوان ورودی می گیرد و یک بیت خروجی دارد. خروجی آن وقتی "صفر" است که هر دو ورودی آن "یک" باشند. در بقیه موارد، خروجی آن "یک" است. اگر دو کلید را به صورت سری به هم متصل کنیم و یک باتری و الکترومغناطیس هم در مدار وجود داشته باشند، الکترومغناطیس فقط وقتی روشن می شود که هر دو کلید، بسته (روشن) باشند. اگر کلید سومی را هم برای خاموش و روشن کردن الکترومغناطیس، به پایین مدار اضافه کنیم، الکترومغناطیس وقتی روشن می شود که کلید سوم باز باشد. دو کلید اول، بیت های ورودی هستند و کلید سوم، بیت خروجی است. (اگر کلید سوم باز باشد، صفر و در غیر این صورت، یک است.) کلید سوم هم تنها در صورتی باز است که دو کلید اول بسته باشند. امروزه، برای ساختن گیت نند از ترانزیستورها استفاده می شود که وسایل کاربردی تری هستند. نظریه های قابل توجهی در علم رایانه وجود دارد که می گویند گیت های نند، همه کاره هستند. بدین معنا که با متصل کردن گیت های نند به یکدیگر، می توان هر تابع به خوبی تعریف شده ای را اجرا کرد.

---

"ند" مخفف "نات اند" می باشد و خروجی "گیت اند" در صورتی یک است که اولین و دومین ورودی آن<sup>۱۴</sup> یک باشند. با این حال، گیت نند بر عکس آن است و در صورتی صفر است که اولین و دومین ورودی آن یک باشند.



بنابراین، اگر بتوانید گیت های نند کافی ای را بسازید، می توانید یک وسیله همه کاره بسازید که فقط با استفاده از گیت های نند، اعداد را در یکدیگر ضرب کند. دو تن از محققان دانشگاه ام آی تی به نام های نورمن مارگولوس و توماسو توفولی برای اولین بار، اصطلاح "ماده قابل برنامه ریزی"<sup>۱۵</sup> را برای اشاره به هر ماده ای که می تواند محاسبات دلبخواهی را انجام دهد، به کار بردند. همانطور که دیدیم، برنامه ریزی ماده، کار سختی نیست و برای انجام آن، فقط لازم است که ماده بتواند گیت های نند را که به نحوی به هم متصل شده اند، پیاده سازی کند. در واقع، تعداد بسیار زیادی از برنامه ریزی های دیگر نیز وجود دارند. یکی از این برنامه ریزی های ساده متنوع این است که گیت های نند (NAND) را با گیت های نر (NOR)<sup>۱۶</sup> جایگزین کرد. در بخش بعدی، راجع به "شبکه های عصبی" که آنها نیز مانند "ماده قابل برنامه ریزی" می توانند محاسبات دلبخواهی را انجام دهند، بحث خواهیم کرد. یک دانشمند و کارآفرین به نام استفن ولفرام نشان داده است که وسایل ساده ای موسوم به "اتوماتای سلولی" که بیت ها را به طور مکرر و بر اساس حالت بیت های مجاور به روز رسانی می کنند، قادر به انجام همین محاسبات دلبخواهی هستند. در سال ۱۹۳۶، آلن تورینگ در یک مقاله تاریخی ثابت کرد که یک ماشین ساده (که اکنون ماشین همه کاره تورینگ نامیده می شود) علاوه بر توانایی دستکاری داده های ساختار یافته در یک نوار ضبط، می تواند محاسبات دلبخواهی را نیز انجام دهد. خلاصه اینکه، ماده علاوه بر قدرت انجام محاسبات پیچیده به خوبی تعریف شده، می تواند آنها را به شیوه های بسیار متنوعی نیز انجام دهد. تورینگ در مقاله تاریخی خود در سال ۱۹۳۶ این را هم ثابت کرد که اگر نوعی از رایانه بتواند مجموعه ای حداقلی از عملیات خاص را انجام دهد، یک رایانه همه کاره است و با فرض وجود منابع کافی، می تواند هر کاری که رایانه های دیگر انجام می دهند را انجام دهد. او نشان داد که ماشین تورینگ، یک ماشین همه کاره است و شامل اشیای متنوعی مانند شبکه گیت های نند و شبکه ای از نورون های متصل شده به یکدیگر، می باشد. در

---

<sup>۱۵</sup> computronium

<sup>۱۶</sup> خروجی آن فقط زمانی یک است که هر دو ورودی صفر باشند

واقع، استفن ولفرام چنین استدلال می کند که بیشتر سیستم های فیزیکی مهم مانند سیستم های آب و هوایی و مغز انسان ها در صورتی که می توانستند به صورت دلبخواهی بزرگ تر و ماندگارتر شوند، به رایانه های همه کاره تبدیل می شدند. این واقعیت که یک رایانه همه کاره می تواند هر نوع محاسبه ای را انجام دهد، بدین معناست که انجام محاسبه مانند اطلاعات، یک پدیده مستقل از بستر فیزیکی می باشد که زندگی منحصر به فرد خود را دارد. بنابراین، اگر شما یک شخصیت فوق هوشمند هوشیار در یک بازی رایانه ای نسل آینده باشید، متوجه نخواهید شد که رایانه تان روی کدامیک از بسترهای فیزیکی مانند دسکتاپ در ویندوز، لپ تاپ در سیستم مک اس یا تلفن همراه در اندروید اجرا می شود، زیرا انجام محاسبات مستقل از بستر فیزیکی صورت می گیرد. شما همچنین متوجه نخواهید شد که ریزپردازنده رایانه تان از چه نوع ترازیسورهایی استفاده می کند. نمونه های فراوانی در علم فیزیک وجود دارند که از ایده استقلال محاسبات از بستر فیزیکی پشتیبانی می کنند. به عنوان مثال، موج ها دارای خواصی از جمله سرعت، طول موج و فرکانس هستند و ما فیزیکدان ها بدون نیاز به دانستن اینکه موج ها از چه مواد خاصی ساخته شده اند، می توانیم معادلاتی را که موج ها از آنها پیروی می کنند، مورد مطالعه قرار دهیم. وقتی که شما صدای چیزی را می شنوید، امواج صوتی ای را شناسایی می کنید که توسط مولکول های ایجاد می شوند که در ترکیبی از گازها به نام هوا، در حال بالا و پایین پریدن هستند. ما می توانیم چیزهای جالبی را درباره این موج ها محاسبه کنیم که از جمله آنها می توان به رابطه کاهش شدت موج ها با مجذور فاصله آنها با ما، چگونگی خم شدن موج ها هنگام عبور آنها از درهای باز، چگونگی نادیده گرفتن دیوارها و طنین انداز شدن صدا اشاره کرد و باید به این نکته توجه داشته باشیم که برای انجام همه این محاسبات، نیازی به این نیست که بدانیم هوا از چه چیزهایی ساخته شده است. ما می توانیم تمام جزئیات مربوط به اکسیژن، نیتروژن، دی اکسید کربن و غیره را نادیده بگیریم، زیرا تنها خاصیتی از بستر موج ها که مهم است و وارد معادلات ما می شود و می توان آن را اندازه گرفت، سرعت موج است که در مطالعه ما، حدود ۳۰۰ متر بر ثانیه بوده است. در واقع، زمان کشف و استفاده از این معادله موج، به مدت ها پیش از

اثبات وجود اتم ها و مولکول ها توسط فیزیکدانان، باز می گردد. مثالی که در مورد موج ها ارائه شد، بیانگر سه نکته مهم می باشد. اول اینکه، استقلال محاسبات از بستر فیزیکی به این معنا نیست که نیازی به بستر فیزیکی وجود ندارد، بلکه به این معناست که بیشتر جزئیات این بستر، بی اهمیت است. اگر شما گازی نداشته باشید، نمی توانید موج های صوتی را در گاز داشته باشید. با این حال، هر نوع گازی برایتان کفایت خواهد کرد. به همین شکل، بدون ماده نمی توان محاسبات را انجام داد، اما هر نوع ماده ای که برای انجام محاسبه همه کاره، بتوان آن را به صورت گیت های نند، نوروں های متصل یا اجزای سازنده دیگری مرتب کرد، کافی خواهد بود. دوم اینکه، پدیده استقلال از بستر فیزیکی بیان می کند که ماده به صورت مستقل از بستر فیزیکی خود، زندگی خاص خودش را دارد. موج می تواند از بین یک دریاچه عبور کند، ولی هیچ یک از مولکول های آب یک موج نمی توانند این کار را انجام دهند. آنها فقط مانند طرفداران یک تیم ورزشی در ورزشگاه، موجی را ایجاد می کنند که به بالا و پایین حرکت می کند. سوم اینکه، فقط جنبه مستقل بودن از بستر فیزیکی بریمان مهم است و کاری به بستر فیزیکی آن نداریم. یک موج سوار بیش از اینکه نگران ترکیبات جزئی مولکول های آب یک موج باشد، نگران موقعیت و ارتفاع موج می باشد. همانطور که دیدید، این موضوع برای اطلاعات هم صادق بوده است و برای محاسبات هم صدق می کند، زیرا برنامه نویسانی که به دنبال شناسایی یک خطا در کد برنامه نویسی رایانه خود هستند، احتمالاً اهمیتی به ترانزیستورهای رایانه خود نمی دهند. اکنون به پاسخی برای این سوال دست پیدا کرده ایم که یک چیز فیزیکی قابل لمس چگونه می تواند منجر به یک چیزی شود که دارای هوش محاسباتی است که یک چیز غیر ملموس، انتزاعی و فضایی است. هوش یک چیز غیر فیزیکی است، زیرا مستقل از هر بستری می باشد و زندگی خاص خودش را دارد که وابسته به هیچ گونه جزئیات فیزیکی ای نمی باشد. خلاصه اینکه، "محاسبه" یک الگو در آرایش زمانی - فضایی ذرات می باشد و الگو مهم است، نه ذرات. به عبارت دیگر، سخت افزار، ماده است و نرم افزار الگوست. همین ویژگی استقلال از بستر فیزیکی محاسبات است که ما را به هوش مصنوعی امیدوار می کند. ویژگی استقلال از بستر

فیزیکی باعث شده که مهندسان بتوانند بدون تغییر نرم افزار یک رایانه، فن آوری های داخل رایانه را با فن آوری های به مراتب بهتری جایگزین کنند. در طی قرن اخیر شاهد این بوده ایم که هزینه محاسبات بسیار ارزان تر شده است. علل اصلی این کاهش قابل توجه در هزینه محاسبات این بوده که این روزها می توان محاسبات را در هر جایی انجام داد و اینکه تجهیزات محاسباتی غول پیکر سال های گذشته، به قدری کوچک شده اند که وارد خانه ها، خودروها و جیب های ما شده اند. اکنون این سوال مطرح می شود که چرا قدرت فن آوری ها در فواصل زمانی منظم در حال دو برابر شدن است و دارای رشد تصاعدی می باشد؟ در واقع، چرا این افزایش قدرت نه تنها در مورد کوچک شدن و دو برابر شدن ترانزیستورها (بر اساس روندی به نام قانون مور<sup>۱۷</sup>) اتفاق افتاده است، بلکه به شکل گسترده تری برای وسایل محاسباتی، دستگاه حافظه و تعداد زیادی از فن آوری های دیگر از جمله توالی یابی ژنومی و تصویر برداری از مغز نیز در حال اتفاق افتادن است؟ ری کرزویل، این پدیده دو برابر شدن پی در پی را به عنوان "قانون بازده صعودی"<sup>۱۸</sup> نام گذاری کرده است. تمام دو برابر شدن های پی در پی که در طبیعت اتفاق می افتد، دلیل اساسی مشابهی دارند و دو برابر شدن های فن آوری نیز از این قضیه مستثنا نیستند. هر گامی باعث تولید گام بعدی می شود. به عنوان مثال، بعد از این که شما به دنیا آمدید، رشد تصاعدی را تجربه کرده اید. هر یک از سلول های شما تقسیم بندی شدند و هر روز تعداد سلول های شما تقریباً دو برابر شدند و تعداد آنها هر روز به صورت تصاعدی افزایش یافت. (۱۶،۸،۴،۲،۱) پیشرفت فن آوری نیز به همین گونه بوده است. وقتی که قدرت یک فن آوری، دو برابر شد، می تواند برای طراحی و ساخت فن آوری دو برابر قوی تر استفاده شود و این افزایش توانمندی دوبرابری بر اساس قانون مور، ادامه خواهد یافت. همزمان با دو برابر شدن قدرت فن آوری، عده ای ادعا می کنند که این روند پایان خواهد پذیرفت. قانون مور به طور حتم خاتمه

---

قانون مور بیان می کند تعداد ترانزیستورهای روی یک تراشه با مساحت ثابت هر ۲ سال، به طور تقریبی دو برابر<sup>۱۷</sup>

می شود.

<sup>۱۸</sup> The law of accelerating returns

خواهد پذیرفت، زیرا برای کوچک کردن ترانزیستورها، محدودیت فیزیکی وجود دارد. مشکل اینجاست که برخی افراد به اشتباه قانون مور را با دوبرابر شدن قدرت فن آوری، هم معنی می دانند. با این حال، ری کرزویل می گوید که هر زمان که پیشرفت یک فن آوری متوقف شود، آن را با فن آوری بهتری جایگزین خواهیم کرد و وقتی که دیگر نتوانیم لامپ های الکترونی را کوچک کنیم، آنها را با ترانزیستورها و مدارهای مجتمع جایگزین خواهیم کرد که در این مدارهای مجتمع، الکترون ها در دو بعد حرکت می کنند. وقتی که عمر این فن آوری ها تمام شود، گزینه های جایگزین زیادی برای آزمایش کردن داریم و به عنوان مثال، می توانیم از مدارهای مجتمع سه بعدی استفاده کنیم یا از چیزی غیر از الکترون ها برای اجرای دستورات خود استفاده کنیم. هیچ کس به طور قطع نمی داند که بستر فیزیکی محاسباتی بعدی چه خواهد بود. اما این را می دانیم که با مرزهای تحمیل شده توسط قوانین فیزیک برای پیشرفت فناوری، فاصله زیادی داریم. یکی از همکاران من در دانشگاه ام آی تی در مورد اینکه این مرز اساسی برای پیشرفت فن آوری چیست، تحقیق کرده است و همانطور که در فصل ششم خواهیم دید، میزان توان محاسباتی در نزدیکی های این مرز پیشرفت،  $10^{33}$  برابر توان محاسباتی یک ماده در بهترین فن آوری های روز دنیا می باشد. بنابراین، حتی اگر ما هر دو سال یکبار، قدرت رایانه های خود را دو برابر کنیم، بیش از دو قرن طول خواهد کشید که به این مرز نهایی پیشرفت برسیم. هرچند که کلیه رایانه های همه کاره از قدرت محاسباتی بالایی برخوردار هستند، برخی از آنها کاراتر از بقیه هستند. به عنوان مثال، برای انجام محاسبه ای که نیاز به میلیون ها بار ضرب کردن دارد، لازم نیست که از میلیون ها واحد مستقل نرم افزاری مخصوص ضرب کردن که از ترانزیستورهای جداگانه ساخته شده اند، استفاده کرد. برای انجام این محاسبه، فقط به یک واحد مستقل نرم افزاری نیاز است که می تواند عملیات ضرب را چندین بار و به صورت پشت سر هم، بر روی ورودی های اختصاص داده شده انجام دهد. به منظور افزایش کارایی، بیشتر رایانه های پیشرفته محاسبات خود را به گام های زمانی چندگانه ای تقسیم می کنند که در طی این گام ها، اطلاعات بین واحدهای حافظه و واحدهای محاسبات به عقب و جلو رانده می شوند. این معماری

محاسباتی در فاصله سال های ۱۹۳۵ تا ۱۹۴۵ توسط برخی از پیشگامان علوم رایانه از جمله آلن تورینگ، کنراد زوس، پرسپراکرت، جان ماشلی و جان نیومن، توسعه پیدا کرد. البته باید توجه داشت که حافظه رایانه علاوه بر اطلاعات، نرم افزار را هم در خود ذخیره می کند. نرم افزار همان برنامه یا فهرستی از دستورالعمل های اجرایی برای کار کردن با داده ها می باشد. در هر گام زمانی، واحد پردازشگر مرکزی (CPU) دستورالعمل بعدی در برنامه را اجرا می کند و برای این کار، از تابع ساده ای برای کار کردن بر روی بخشی از داده ها استفاده می کند. بخشی از رایانه که دستورالعمل بعدی را دنبال می کند، "شمارنده برنامه"<sup>۱۹</sup> نامیده می شود که شماره خط فعلی در برنامه را ذخیره می کند. برای انجام دستورالعمل بعدی فقط کافی است که شمارنده برنامه را یکی افزایش دهید. همچنین، برای پریدن به خط دیگری از برنامه، کافی است که عدد آن خط را به شمارنده برنامه، کپی کنید. از این ویژگی برای اجرای دستورات شرطی و حلقه ها در رایانه ها استفاده می شود. عملیات "پردازش موازی" در رایانه های امروزی که باعث حذف هوشمندانه استفاده مجدد از واحدهای مستقل نرم افزاری شده، سرعت رایانه ها افزایش داده است. اگر عملیات محاسبه را بتوان به بخش هایی تقسیم کرد که قادر به کار کردن موازی باشند (یعنی ورودی یک بخش نیازی به خروجی بخش دیگر نداشته باشد) در این صورت، می توان بخش های مختلف محاسباتی را به صورت موازی و از طریق سخت افزارهای مختلف محاسبه کرد. رایانه موازی غایی، یک رایانه کوانتومی خواهد بود. دیوید دیچ که از پیشگامان محاسبه کوانتومی می باشد، استدلال می کند که: "رایانه های کوانتومی اطلاعات خود را با تعداد زیادی از ویرایش های خودشان در سراسر جهان های مختلف به اشتراک خواهند گذاشت و با کمک گرفتن از این ویرایش ها می توانند سرعت پاسخ گیری را در جهان خودمان افزایش دهند. ما هنوز نمی دانیم که آیا امکان ساخت چنین رایانه های قابل رقابت تجاری ای در طی دهه های آینده وجود دارد یا خیر. امکان ساخت این رایانه ها بستگی به این دارد که فیزیک کوانتومی همانطوری که فکر می کنیم، عمل کند و اینکه بتوانیم از پس چالش های فنی دلهره برانگیز

---

<sup>۱۹</sup> Program counter

ساختمان برآییم. با این حال، شاهد این هستیم که شرکت ها و دولت های زیادی از سراسر جهان، سالانه، ده ها میلیون دلار صرف شرط بندی برای شدنی بودن آنها می کنند. هرچند که رایانه های کوانتومی نمی توانند محاسبات معمولی را با سرعت بیشتری انجام دهند، اما الگوریتم های هوشمندانه ای توسعه پیدا کرده است که می توانند سرعت انواع خاصی از محاسبات را به صورت چشمگیری افزایش دهند که از جمله آنها می توان به الگوریتم های رمز گشایی از سیستم های رمزی و شبکه های عصبی اشاره کرد. یک رایانه کوانتومی همچنین می تواند رفتار سیستم های کوانتومی - مکانیکی را که شامل اتم ها، مولکول ها و مواد جدید می باشند را با سرعت زیادی شبیه سازی کند.

## یادگیری چیست؟

هرچند که یک ماشین حساب جیبی ممکن است مرا در یک مسابقه ریاضی شکست دهد، اما هر چه قدر هم که تمرین کند، هرگز قادر به بهبود سرعت یا دقت خود نخواهد بود، زیرا یک ماشین حساب جیبی قادر به یادگیری نیست. به عنوان مثال، هر بار که من دکمه جذر را در یک ماشین حساب جیبی فشار می دهم، دقیقا تابع مشابهی را به شیوه یکسان محاسبه می کند. همچنین، اولین برنامه رایانه ای که مرا در بازی شطرنج شکست داده است، کارش فقط این بوده که تابع طراحی شده توسط برنامه نویس باهوش خود را برای محاسبه بهترین حرکت بعدی اجرا کند و هرگز قادر به یادگیری از اشتباهات خود، نبوده است. این در حالی است که وقتی مگنوس کارلسن در سن پنج سالگی در اولین بازی شطرنج خود شکست خورد، هجده سال بعد، با یادگیری از شکست خود، توانست قهرمان بازی های جهانی شطرنج شود. به جرات می توان گفت که توانایی یادگیری یکی از جالبترین جنبه های هوش عمومی می باشد. همانطور که دیدیم، یک ماده به ظاهر بی زبان می تواند چیزهایی را به خاطر بسپرد یا محاسباتی را انجام دهد، اما در اینجا خواهیم دید که یک ماده قادر به یادگیری نیز می باشد. ما دیدیم که پیدا کردن یک پاسخ برای یک سوال سخت، مشابه با محاسبه یک تابع می باشد و اینکه اگر ماده به خوبی تنظیم شده باشد، می تواند هر تابع قابل محاسبه ای را محاسبه کند. ما انسان ها وقتی که برای اولین بار، ماشین حساب جیبی و برنامه بازی شطرنج را درست کردیم، خودمان آن را تنظیم

کردیم. اما، اگر ماده بخواهد چیزی را یاد بگیرد، باید خودش را به گونه ای تنظیم کند که ضمن پیروی از قوانین فیزیک، قادر به بهبود مهارت خود در زمینه محاسبه تابع مورد نظرش باشد. سیستم های فیزیکی مغز می توانند با کارایی بسیار بالایی چیزهایی را یاد بگیرند. جان هپفیلد نشان داد که شبکه ای از نورون های به هم متصل در مغز می توانند به یادگیری انسان کمک کنند. سیستم مغز انسان به گونه ای است که اگر نورون ها را به طور مکرر در حالت های خاصی قرار دهید، به تدریج این حالت ها را یاد می گیرند و از نزدیک ترین حالت به این حالت های یاد گرفته شده برمی گردند. همچنین، اگر هر یک از اعضای خانواده خود را بارها دیده باشید، وقتی که چیزی را در ارتباط با آنها ببینید، خاطره تان برانگیخته شده و چهره آنها را به خاطر خواهید آورد. "شبکه های عصبی"<sup>۲۰</sup> امروزه باعث تحولاتی در زمینه هوش زیستی و هوش مصنوعی شده اند و بر حوزه فرعی هوش مصنوعی که "یادگیری ماشین"<sup>۲۱</sup> نامیده می شود، تسلط پیدا کرده اند. منظور از یادگیری ماشین، مطالعه الگوریتم هایی است که از طریق تجربه، بهبود پیدا می کنند. قبل از اینکه وارد بررسی عمیقتری از نحوه یادگیری شبکه های عصبی شویم، اجازه دهید که از نحوه محاسبه کردن با آنها آگاه شویم. شبکه عصبی، شامل گروهی از نورون های به هم متصل شده می باشد که می توانند روی رفتار یکدیگر تاثیر گذارند. تعداد نورون ها در مغز ما تقریباً به اندازه تعداد ستاره ها در کهکشان ما هستند و تعدادشان تقریباً به صد میلیارد می رسد. به طور متوسط، هر یک از این نورون ها از طریق پیوندهایی موسوم به "سیناپس"<sup>۲۲</sup> به حدود هزاران نورون دیگر متصل می باشند و قدرت اتصال بین همین صد تریلیون سیناپس، همان چیزی است که باعث رمزگذاری بیشتر اطلاعات در مغز ما می شود. یک شبکه عصبی را می توان به صورت مجموعه ای از نقطه ها (نورون ها) که توسط خطوط (سیناپس ها) به هم متصل شده اند، ترسیم کرد. نورون های دنیای واقعی، وسایل الکتروشیمیایی بسیار پیچیده ای هستند که شامل اجزای مختلفی مانند اکسون ها و دندریت ها می باشند. همچنین، انواع

---

<sup>۲۰</sup> Neural networks

<sup>۲۱</sup> Machine learning

<sup>۲۲</sup> synapse



مختلفی از نورون ها وجود دارند که به شیوه های بسیار مختلفی عمل می کنند. جزئیات دقیقتر در مورد اینکه فعالیت الکتریکی در یک نورون چگونه و در چه زمانی بر نورون های دیگر تاثیر می گذارد، موضوعی است که باید در مورد آن تحقیقات بیشتری انجام شود. با این حال، محققان هوش مصنوعی نشان داده اند که حتی اگر پیچیدگی های شبکه های عصبی را نادیده بگیریم و نورون های واقعی را با نورون های شبیه سازی شده ای که به شدت ساده سازی شده اند، جایگزین کنیم و آنها را مشابه با یکدیگر فرض کنیم، هنوز می توانیم بسیاری از وظایف به شدت پیچیده را با عملکردی در سطح عملکرد یک انسان، انجام دهیم. در یکی از مدل های بسیار رایج امروزی برای انجام محاسبات در شبکه عصبی مصنوعی، حالت هر نورون با یک عدد و قدرت هر سیناپس با یک عدد دیگر بیان می شوند. در این مدل، هر نورون در گام های زمانی منظم و از طریق محاسبه میانگین ورودی ها به نورون های به هم متصل شده، وزن دار کردن آنها بر اساس قدرت سیناپس ها، اضافه کردن یک عدد ثابت و اجرای تابعی موسوم به "تابع فعالسازی"<sup>۲۳</sup> که برای محاسبه حالت بعدی یک نورون روی نتایج اعمال می شود، حالت خود را به روز رسانی می کند. ساده ترین راه برای استفاده از شبکه عصبی به عنوان یک تابع "کنترل کننده آینده نگر"<sup>۲۴</sup> که قبل از اجرای محاسبات روی داده ها اعمال می شود، این است که اطلاعات فقط از یک جهت جریان پیدا کنند و ورودی ها از طریق لایه ای از نورون ها از بالا وارد تابع شوند و خروجی ها از طریق لایه ای از نورون ها از پایین، خارج شوند. "استقلال از بستر فیزیکی" در شبکه های عصبی مصنوعی ساده نیز مشاهده می شود. شبکه های عصبی دارای قدرت محاسباتی بالایی هستند که مستقل از جزئیات سطح پایینی است که در مورد مبانی ساختشان وجود دارد. جمعی از دانشمندان در سال ۱۹۸۹ چیز فوق العاده ای را اثبات کردند و آن این بود که شبکه های عصبی ساده، همه کاره هستند، به این معنی که فقط با تنظیم مناسب اعداد

---

<sup>۲۳</sup> Activation function

کنترلی است که با استفاده از داده ها کنترل را تحقق می بخشد و قبل از آنکه عملی واقع و نتیجه ای حاصل شده Feedforward<sup>۲۴</sup> باشد طریق اصلاح را ارائه می دهد. به عبارت دیگر این نوع کنترل، موانع و مشکلات را قبل از وقوع پیش بینی کرده و انجام اصلاحات را ممکن می سازد.

مربوط به قدرت سیناپس، می توان هر تابع دلخواهی را به دقت محاسبه کرد. البته باید توجه داشت که نورون های طبیعی ما در مقایسه با آنچه که مهندسین ساخته اند، به مراتب پیچیده تر و غیر قابل درک تر و در نتیجه کاراتر هستند. من گیج شده ام که چگونه یک چیز ساده می تواند محاسبات پیچیده دلخواه ما را انجام دهد و اینکه چگونه می تواند عمل ضرب ساده را فقط با محاسبه مجموع های وزنی ورودی نورون ها و اعمال یک تابع ثابت بر روی آن ها محاسبه کند. هرچند که از لحاظ نظری می توان ثابت کرد که همه چیز با یک شبکه عصبی بزرگ قابل محاسبه است، در عمل و برای یک شبکه دارای اندازه معقول، ممکن است که این امر امکان پذیر نباشد. به عنوان مثال، فرض کنید که می خواهیم تصاویر مگاپیکسل خاکستری را به دو گروه متفاوت طبقه بندی کنیم. اگر هر یک از این یک میلیون پیکسل بتواند ۲۵۶ مقدار متفاوت را پذیرند،  $256^{1,000,000}$  تصویر احتمالی وجود خواهد داشت که از تعداد اتم های موجود در دنیا ( $10^{78}$ ) بیشتر است. با این حال، شبکه های عصبی دارای هزارها یا میلیون ها پارامتر را می توان به خوبی مدیریت کرد. البته، ثابت شده است که یک شبکه عصبی کوچک برای جا شدن در دنیا نمی تواند تمام محاسبات را انجام دهد و تنها موفق به انجام بخش کوچکی از آن خواهد شد. وقتی که با یکی از شاگردانم به نام هنری لین در مورد چنین اسرار عجیبی صحبت می کردم، احساس خوبی به من دست می داد. از اینکه فرصت همکاری با چنین دانشجوی فوق العاده ای را داشتم، بسیار خوشحال بودم. وقتی که هنری لین وارد دفتر کارم شد تا نظر من را در مورد اینکه آیا مایلم که با او کار کنم، پرسد، با خودم فکر کردم که من باید از او بپرسم که آیا مایل است که با من کار کند. او بسیار متواضع و مهربان بود و با اینکه فقط بیست سال داشت، هشت مقاله علمی را تالیف کرده بود و برنده جایزه های زیادی شده بود. یک سال بعد، با هم یک مقاله نوشتیم که در بخش نتیجه گیری آن می خواستیم پاسخی را برای این سوال پیدا کنیم که چرا عملکرد خوب شبکه های عصبی را نمی توان فقط با استفاده از علوم ریاضی توضیح داد و اینکه چرا بخشی از این پاسخ را باید در علم فیزیک جستجو کرد؟ ما متوجه شدیم که اهمیت توابعی که قوانین فیزیک به ما اعطا می کند و ما را علاقه مند به انجام محاسبه با آنها می سازد، فوق العاده پایین است،

زیرا قوانین فیزیک به دلایلی که هنوز به طور کامل قابل درک نیست، فوق العاده ساده هستند. علاوه بر این، بخش کوچکی از توابع ریاضی که شبکه های عصبی را قادر به انجام محاسبات خود با آنها می کند، بسیار شبیه همان بخش کوچکی از توابعی است که علم فیزیک، به آنها علاقه نشان می دهد. ما همچنین برای ادامه دادن به کارهای تحقیقاتی قبلی، نشان دادیم که در بسیاری از توابع، شبکه های عصبی ای که دارای قدرت یادگیری عمیق هستند (که عمیق بودن آنها بستگی به تعداد لایه های آنها دارد)، بسیار کاراتر از شبکه های عصبی ای هستند که یادگیری آنها سطحی است. به عنوان مثال، من با همکاری یکی از دانشجویهای دیگر خود در دانشگاه ام آی تی نشان دادم که در شبکه ای که دارای یک لایه است، وظیفه ساده ضرب کردن  $n$  عدد در یکدیگر، نیاز به  $2^n$  نورون دارد و این در حالی است که برای انجام این محاسبه در شبکه های عمیق، فقط به حدود  $4n$  نورون نیاز است که این امر باعث محبوبیت شبکه های عصبی در بین محققان هوش مصنوعی شده است. اکنون که با نحوه کارکرد و محاسبه در شبکه های عصبی آشنا شدیم، به بررسی نحوه یادگیری در شبکه های عصبی و نحوه بهبود در محاسبات شبکه های عصبی از طریق به روزرسانی سیناپس آنها می پردازیم. یک روان شناس کانادایی به نام دونالد هب<sup>۲۵</sup> در کتاب خود با عنوان "سازماندهی رفتار: نظریه روانشناسی عصب نگر" استدلال می کند که اگر دو نورون مجاور یکدیگر به طور همزمان فعال (تحریک) شوند، قدرت "همبندی سیناپسی"<sup>۲۶</sup> آنها تقویت می شود و آنها یاد می گیرند که چگونه یکدیگر را تحریک کنند. به عبارت دیگر، هر قدر شما یک مدار نورونی را در مغز خود بیشتر اجرا کنید، آن جریان قوی تر می شود. هر چند که درک ما از جزئیات نحوه یادگیری در مغز، اندک است و تحقیقات نشان می دهند که در بسیاری موارد، این یک امر به شدت پیچیده ای است، برخی از تحقیقات نشان می دهند که قانون یادگیری ساده (موسوم به یادگیری هبین) به شبکه های عصبی اجازه می دهد که چیزهای جالبی را یاد بگیرند. جان هاپفیلد نشان داد که یادگیری هبین به شبکه های عصبی مصنوعی بیش از حد ساده سازی شده، اجازه می دهد تا بسیاری از خاطرات پیچیده را فقط با

---

<sup>۲۵</sup> Donald hebb

<sup>۲۶</sup> Synaptic coupling

قرار گرفتن مکرر در معرض آنها، در خود ذخیره کنند. به این فرایند (قرار گرفتن مکرر در معرض اطلاعات برای یادگیری از آنها) در شبکه های عصبی مصنوعی، "آموزش" می گویند که شامل یادگیری مهارت ها توسط انسان نیز می شود و با واژه هایی نظیر "مطالعه"، "تحصیل" یا "تجربه" در ارتباط می باشد. شبکه های عصبی مصنوعی حاکم بر سیستم های هوش مصنوعی امروزی، به جای استفاده از قانون یادگیری ساده "هین"، از قوانین یادگیری پیچیده جایگزینی موسوم به "انتشار رو به عقب"<sup>۲۷</sup> و "گرادیان کاهشی تصادفی"<sup>۲۸</sup> استفاده می کنند. با این حال، ایده اصلی همه این قوانین، مشابه با یکدیگر است و آن این است که یک قانون جبری ساده که شباهت زیادی با قوانین علم فیزیک دارد، با گذشت زمان باعث به روز رسانی سیناپس ها می شود. در صورتی که فرایند آموزش با حجم زیادی از اطلاعات انجام شود، همین قانون ساده باعث می شود که شبکه عصبی یاد بگیرد که محاسبات فوق العاده پیچیده ای را انجام دهد. ما هنوز به طور دقیق نمی دانیم که مغزمان از چه قوانینی برای یادگیری استفاده می کند. با این حال، این را می دانیم که آنها از قوانین علم فیزیک تخطی نخواهند کرد. بسیاری از شبکه های عصبی مصنوعی و طبیعی همان کاری را می کنند که بیشتر رایانه های دیجیتالی برای افزایش کارایی خود استفاده می کنند و آن این است که وظیفه خود را به گام های متعدد تقسیم می کنند و برای انجام محاسبات، از واحدهای مستقل نرم افزار محاسباتی خود بارها استفاده مجدد می کنند. مغز انسان دارای بخش هایی است که دانشمندان علوم رایانه از اصطلاح شبکه های عصبی "بازگشتی"<sup>۲۹</sup> به جای "جلو رونده" برای توصیف آنها استفاده می کنند. در این شبکه ها، شاهد این هستیم که اطلاعات به جای یک جهت واحد، در جهت های متعددی گردش پیدا می کنند و خروجی یک جریان می تواند یک ورودی برای جریان بعدی باشد. شبکه گیت های منطقی در ریزپردازنده یک لپتاپ نیز از این لحاظ "بازگشتی" می باشد. این شبکه، اطلاعات گذشته خود را مجدداً مورد استفاده قرار می دهد و به اطلاعات ورودی جدید از طریق

---

<sup>۲۷</sup> backpropagation

<sup>۲۸</sup> Stochastic gradient descent

<sup>۲۹</sup> recurrent

صفحه کلید، ترک پد، دوربین و غیره اجازه می دهد تا به طور بی وقفه روی محاسبات تاثیر گذارند تا اینکه در نهایت، اطلاعات خروجی از طریق صفحه نمایش، بلندگو، چاپگر یا شبکه بی سیم، معلوم شوند. به همین ترتیب، شبکه نوروں های موجود در مغز نیز بازگشتی هستند و به اطلاعات ورودی از چشم ها، گوش ها و حواس دیگر انسان اجازه می دهند تا به طور بی وقفه روی محاسبات تاثیر گذارند تا اینکه در نهایت، تاثیر آنها در اطلاعات خروجی، معلوم شود. تاریخچه یادگیری به تاریخچه زندگی بشر باز می گردد. در نظام تولید مثل بشر، شاهد تاثیر کپی کردن و پردازش اطلاعات والدین برای یادگیری از سوی فرزندان، هستیم. با این حال، در طی زندگی ۱، اندام ها چیزی را در طول عمر خود یاد نمی گرفتند و قوانین آنها برای پردازش اطلاعات و واکنش در برابر آن، بر اساس دی ان ای به ارث برده شده شان، تعیین می شد. بنابراین، فرایند یادگیری به کندی و در سطح گونه ها و فقط بر اساس فرضیه تکامل تدریجی داروین در بین نسل ها اتفاق می افتاد. حدود نیم میلیارد سال پیش، راهی برای توانمندسازی حیواناتی که دارای شبکه عصبی بودند به منظور یادگیری رفتارهای مناسب از طریق تجاربشان در زندگی کشف شد. زندگی ۲ فرا رسیده بود و به دلیل قابلیت آن برای یادگیری بسیار سریعتر و برتری در عرصه رقابت، مانند یک آتش سریع، کل جهان را فراگرفت. همانطور که در فصل ۱ نشان دادیم، میزان یادگیری در زندگی به تدریج و با نرخ فزاینده ای افزایش یافته است. یک گونه خاص انسانی شبیه میمون با پرورش مغز خود توانستند به قدری در زمینه کسب دانش ماهر شوند که نحوه استفاده از ابزارها، روشن کردن آتش، صحبت کردن با زبان و ایجاد یک جامعه جهانی پیچیده را یاد بگیرند. این جامعه را می توان به عنوان یک سیستم در نظر گرفت که از قدرت حافظه، محاسبه و یادگیری برخوردار است و با گام های سریع به دنبال اختراعات جدید در زمینه هایی مانند نوشتن، دستگاه چاپ، علوم پیشرفته، رایانه، اینترنت و غیره است. حدس من این است که "هوش مصنوعی" در فهرست اختراعات بعدی ثبت شده توسط مورخان آینده قرار خواهد گرفت. همانطور که همه ما می دانیم، پیشرفت های انفجاری در زمینه حافظه رایانه و توان محاسباتی آن، منجر به پیشرفت های فوق العاده ای در زمینه هوش مصنوعی شده است. با این حال،

مدت زیادی طول کشید تا "یادگیری ماشین" از آب و گل درآید. وقتی که رایانه دیپ بلوی آی بی ام موفق به شکست دادن گری کاسپارف، قهرمان بازی شطرنج در سال ۱۹۹۷ شد، معلوم شد که مزیت اصلی این رایانه به خاطر یادگیری آن نبوده، بلکه به خاطر حافظه و توان محاسباتی بالای آن بوده است. هوش محاسباتی رایانه دیپ بلوی آی بی ام توسط تیمی از انسان ها ایجاد شده بود و دلیل اصلی موفقیت آن در شکست دادن انسان ها به خاطر توان محاسبه سریعتر آن بود که باعث شده بود وضعیت های بالقوه بیشتری را مورد تجزیه و تحلیل قرار دهد. وقتی که رایانه واتسون آی بی ام مغلوب قهرمان شطرنج جهان شد، خطر را احساس کرد و دلیل شکست رایانه واتسون، این بود که بیش از قدرت یادگیری خود، بر روی مهارت های برنامه ریزی ویژه و حافظه و سرعت بالایش تکیه کرده بود. این قضیه در مورد بیشتر پیشرفت های مهم اولیه در زمینه علم رباتیک مانند خودروی بدون راننده و موشک های خود فرود آینده، نیز صدق می کند. نیروی پیشران بسیاری از پیشرفت های مهم اخیر در زمینه هوش مصنوعی، "یادگیری ماشین" بوده است. اگر تصویری را مشاهده کنید که در آن، چند نفر مشغول انجام بازی فریزی هستند، بهتر است بدانید که تابعی برای این تصویر برنامه ریزی شده است که رنگ کلیه پیکسل های تصویر را به عنوان ورودی می گیرد و توضیح درستی مانند اینکه "گروهی از جوانان در حال انجام بازی فریزی هستند." را به عنوان خروجی ارائه می کند. تیمی از شرکت گوگل کار مشابهی را در سال ۲۰۱۴ انجام داد و با وارد کردن مجموعه متفاوتی از رنگ پیکسل ها پاسخ درستی مانند اینکه "گروهی از فیل ها در یک مزرعه خشک در حال عبور هستند." را ارائه کرد. آنها برای این کار از برنامه نویسی الگوریتم های دقیق برای پیدا کردن چهره ها، فریزی ها و غیره استفاده نکرده بودند. بلکه یک شبکه عصبی نسبتا ساده را ایجاد کرده بودند که هیچ دانشی از دنیای فیزیکی یا محتویات آن نداشت و فقط به آن اجازه داده بودند که با قرار گرفتن در معرض حجم وسیعی از اطلاعات، چیزهایی را یاد بگیرد. همانطور که ما نمی دانیم که بچه هایمان چگونه یاد می گیرند، از نحوه یادگیری شبکه های عصبی و دلایل شکست آنها کاملاً آگاه نیستیم. فقط می دانیم که آنها بسیار مفید هستند و ما را به سمت موجی از سرمایه گذاری در زمینه "یادگیری

عمیق<sup>۳۰</sup>، سوق می دهند. یادگیری عمیق، بسیاری از جنبه های دیداری رایانه ها را متحول ساخته و از کپی برداری از دست نوشته ها به سمت تحلیل ویدئویی بی درنگ از حرکت های خودروهای بدون راننده، پیشرفت کرده است و باعث شده است که رایانه ها بتوانند زبان گفتاری را به متن تبدیل کنند و متن را به زبان دیگری ترجمه کنند و حتی این دو عمل را به صورت همزمان انجام دهند تا ما بتوانیم با دستیارهای دیجیتال شخصی<sup>۳۱</sup> از قبیل سیری، گوگل نا و کورتانا صحبت کنیم. در سال ۲۰۱۵ شرکت گوگل دیپ مایند، سیستم هوش مصنوعی مجهز به یادگیری عمیق را وارد بازار کرد که می توانست مثل بچه ها در بازی های رایانه ای بدون دستورالعمل، مهارت کسب کند و فقط به آن یاد داده بودند که بهتر از انسان بازی کند. در سال ۲۰۱۶ همین شرکت، سیستم رایانه ای آلفاگو را ساخت که از یادگیری عمیق استفاده می کرد و این سیستم در نهایت موفق شد که قویترین قهرمان بازی گو را شکست دهد. این پیشرفت ها باعث شد که بودجه و افراد با استعداد بیشتری جذب تحقیقات هوش مصنوعی شوند و چرخه پیشرفت هوش مصنوعی ادامه پیدا کند. در این فصل، راجع به ماهیت هوش و پیشرفت های آن تا این لحظه صحبت کردیم. ما دقیقاً نمی دانیم که چقدر طول خواهد کشید تا ماشین ها در رقابت برای انجام تمامی وظایف شناختی از ما جلو بزنند و اگر راستش را بخواهید، ممکن است که پاسخ این سوال، "هرگز" باشد. با این حال، یکی از پیام های این فصل، این است که احتمال تحقق این امر حتی در طول زندگی ما وجود دارد. همانطور که گفتیم، ماده را به گونه ای می توان تنظیم کرد که با پیروی از قوانین فیزیک، از قدرت حافظه، محاسبه و یادگیری برخوردار شود. از محققان هوش مصنوعی انتقاد می شود که آنها بیش از توانایی خود قول می دهند و انتظارات را برآورده نمی کنند. اما برخی از منتقدین دارای مدرکی برای اثبات حرف خود نیستند. برخی از آنها هوش را به معنی آنچه که رایانه ها هنوز نمی توانند انجام دهند، تعریف می کنند. در

---

دیپ لرنینگ نوعی شبکه عصبی از چندین لایه نرون مصنوعی است که با استفاده از مجموعه داده های بزرگ<sup>۳۰</sup> آموزش داده می شود. این لایه ها به مدل های یادگیری عمیق اجازه می دهند تا داده های پیچیده را بیاموزند و پیش بینی های دقیقی انجام دهند.

<sup>۳۱</sup> Personal digital assistants

حال حاضر، ماشین‌ها در زمینه‌هایی مانند ریاضیات، شطرنج، اثبات نظریه‌های ریاضی، تحلیل سهام، توصیف محتوای تصویر، رانندگی، بازی آرکید، بازی گو، پردازش صدا<sup>۳۲</sup>، رونویسی از گفتار، ترجمه و تشخیص سرطان، دارای مهارت خوب یا عالی هستند، اما برخی از منتقدان ما را مسخره می‌کنند و به ما می‌گویند که "این‌ها، هوش واقعی نیستند." تاثیر هوش مصنوعی بر جامعه به رشد خود ادامه می‌دهد. مدت‌ها قبل از اینکه هوش مصنوعی در تمام کارها به سطح هوش انسان برسد، فرصت‌ها و چالش‌های جالبی در رابطه با مسائلی مانند خطاهای نرم‌افزاری، قوانین، سلاح‌ها و شغل‌ها ایجاد خواهد شد. در فصل بعدی، راجع به ماهیت این مسائل و نحوه آماده‌سازی خود برای مواجهه با آنها صحبت خواهیم کرد.

## خلاصه فصل:

- هوش به معنی توانایی دستیابی به اهداف پیچیده می‌باشد و نمی‌توان آن را فقط با عدد ضریب هوشی اندازه گرفت، بلکه باید میزان دستیابی به کلیه اهداف را برای اندازه‌گیری آن در نظر گرفت.
- هوش مصنوعی امروزی محدود است و فقط می‌تواند اهداف بسیار ویژه‌ای را محقق سازد، در حالی که هوش انسان، فوق‌العاده گسترده است.
- حافظه، محاسبه، یادگیری و هوش دارای مفهوم انتزاعی، ناملموس و فرازمینی هستند و مستقل از بستر فیزیکی می‌باشند، به این معنی که می‌توانند زندگی مخصوص خود را بپذیرند که دارای هیچ گونه وابستگی به جزئیات بستر زیربنایی مادی آنها نیست.
- هر تکه‌ای از ماده می‌تواند بستر فیزیکی حافظه باشد، البته تا زمانی که دارای حالت‌های پایدار بسیار مختلفی باشد.

---

<sup>۳۲</sup> Speech synthesis: تبدیل متن به صدا یا صدا به متن



- هر ماده ای می تواند بستر فیزیکی محاسبات باشد، البته تا زمانی که حاوی اجزای سازنده خاص همه کاره ای باشد که با ترکیب آنها بتوان هر تابعی را اجرا کرد. گیت های نند و نوروں ها دو نمونه مهم از چنین اتم های محاسباتی همه کاره ای هستند.
- شبکه عصبی، بستر قوی ای برای یادگیری می باشد، زیرا با پیروی از قوانین علم فیزیک می تواند خودش را به گونه ای تنظیم مجدد کند که مهارت خود در اجرای محاسبات مورد نظر را افزایش دهد.
- ما انسان ها به دلیل سادگی قابل توجه قوانین علم فیزیک، فقط به بخش کوچکی از مسائل محاسباتی تصور کردنی، اهمیت می دهیم و شبکه های عصبی مهارت خود را فقط در زمینه حل همین بخش کوچک از مسائل، افزایش می دهند.
- وقتی که قدرت فن آوری دو برابر می شود، از آن می توان برای طراحی و ساخت فن آوری با قدرت دو برابر بیشتر استفاده کرد و بر اساس قانون مور، این دو برابر شدن قدرت فن آوری ادامه پیدا می کند. در قرن اخیر که عصر اطلاعات است، هزینه فن آوری اطلاعات در هر دو سال یکبار، تقریباً نصف شده است.
- اگر پیشرفت هوش مصنوعی ادامه پیدا کند، مدت ها قبل از اینکه مهارت های هوش مصنوعی به سطح مهارت های انسانی برسند، فرصت ها و چالش های جالبی در رابطه با مواردی از قبیل خطاهای نرم افزاری، قوانین، سلاح ها و شغل ها ایجاد خواهد شد که در فصل بعدی به بررسی آنها خواهیم پرداخت.

## **فصل سوم- آینده نزدیک: پیشرفت های مهم، خطاهای نرم افزاری، قوانین، سلاح ها و شغل ها**

حفظ انسانیت در روزگار و عصر حاضر به چه معناست؟ به عنوان مثال، چه چیزی باعث می شود که واقعا برای خودمان ارزش قائل شویم و مانند ماشین ها نباشیم؟ چه چیزی باعث می شود که افراد دیگر برایمان ارزش قائل شوند و شغل هایی را به ما پیشنهاد کنند؟ مهم نیست که پاسخ این سوالات

در هر زمانی چه باشد. آنچه که واضح است، این است که پیشرفت فن آوری به تدریج باعث ایجاد تغییراتی در این پاسخ ها خواهد شد. به عنوان مثال، من به عنوان یک دانشمند به برخی از توانایی های خود مانند تعیین اهداف شخصی، استفاده از خلاقیت و استعداد ذاتی برای مقابله با طیف وسیعی از مسائل حل نشده و استفاده از زبان برای سهم ساختن دیگران در آنچه که کشف می کنم، افتخار می کنم. خوشبختانه، جامعه برای شغل من ارزش قائل است. قرن ها پیش، ممکن بود که من هم مانند بسیاری از افراد دیگر، هویت خود را بر پایه شغل کشاورزی یا کارهای دستی بنا کنم، اما رشد تدریجی فن آوری باعث شد که نیروی کار انجام دهنده این شغل ها کاهش پیدا کنند و همه نتوانند هویت خود را بر پایه این شغل ها استوار سازند. به نظر خودم، اینکه ماشین های امروزی برای انجام کارهای یدی مانند حفاری و بافندگی به انسان ها ترجیح داده می شوند، مرا ناراحت نمی کند. زیرا این کارها نه وسیله سرگرمی و نه منبع درآمدی برایم هستند. در سن هشت سالگی و هنگامی که در دوره اجباری بافندگی مدرسه، تقریباً رد شدم و با کمک گرفتن از دوستان دلسوز خود این پروژه را گذراندم، توهمی که در زمینه توانایی انجام کار بافندگی را داشتم، درهم شکسته شد. اکنون، این سوال مطرح می شود که آیا آن دسته از توانایی های ما انسان ها که به افزایش ارزش شخصی و اعتبار ما در بازار کار کمک می کنند، تحت تاثیر پیشرفت های فن آوری، کمرنگ خواهند شد یا خیر؟ استوارت راسل می گوید که وقتی که او به همراه بسیاری از محققان هوش مصنوعی شاهد انجام کار غیر منتظره ای توسط ماشین هوش مصنوعی بودند، لحظه بهت آوری را تجربه کردند. در اینجا می خواهم که در مورد برخی از این لحظات با شما صحبت کنم و این نوید را به شما بدهم که هوش مصنوعی به زودی از توانایی های بشری سبقت خواهد گرفت.

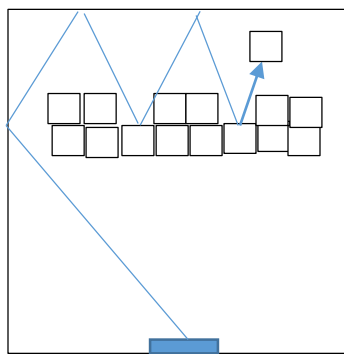
## پیشرفت های مهم

### عوامل های یادگیری تقویت عمیق<sup>۳۳</sup>

---

<sup>۳۳</sup> Deep reinforcement learning agents

در سال ۲۰۱۴ و در حالی که ویدئویی از یادگیری نحوه انجام بازی های رایانه ای توسط سیستم هوش مصنوعی دیپ مایند را تماشا می کردم، لحظه بهت آوری را تجربه کردم. به وضوح دیده می شد که هوش مصنوعی در حال انجام بازی برک اوت می باشد که یکی از بازی های آتاری مورد علاقه من در جوانی بوده است. هدف از این بازی این است که صفحه رابط به گونه ای حرکت داده شود که گلوله به طور مکرر به این ور و آن ور پرد و به آجرهای دیوار برخورد کند. هر بار که گلوله به یکی از آجرها برخورد کند، آجر ناپدید می شود و یک امتیاز به شما داده خواهد شد.



توضیح: هوش مصنوعی دیپ مایند پس از یادگیری نحوه بازی برک اوت آتاری، با استفاده از مفهوم "یادگیری تقویت عمیق" در شبکه های عصبی، برای حداکثر کردن امتیاز خود، راهبرد بهینه را کشف کرد و آن این بود که سوراخی را در دورترین جای سمت چپ دیوار آجری ایجاد کند و به گلوله اجازه دهد که در پشت دیوار آجری، این ور و آن ور پرد. پیکان های ترسیم شده خط سیر گلوله و صفحه رابط را نشان می دهند.

خود من در گذشته برنامه برخی از بازی های رایانه ای را نوشته بودم و به خوبی می دانستم که نوشتن برنامه ای که بتواند بازی برک اوت را انجام دهد، کار سختی نیست. اما تیم هوش مصنوعی دیپ مایند برنامه ای را برای انجام این بازی نوشته بود. در عوض، آنها یک هوش مصنوعی با لوح سفید ایجاد کرده بودند که هیچ چیزی راجع به این بازی و مفاهیمی مانند صفحه رابط، آجر و گلوله نمی دانست. تنها چیزی که هوش مصنوعی آنها می دانست، این بود که فهرست طولانی ای از اعداد در

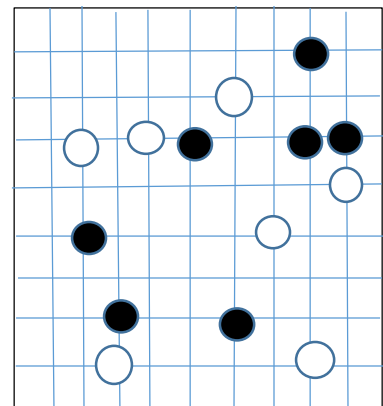
فواصل زمانی منظم به آن خورانده می شد که این اعداد شامل امتیاز فعلی و فهرست طولانی ای از اعدادی بودند که ما به عنوان مشخصات رنگ بندی بخش های مختلف صفحه نمایش، می شناختیم. فقط به هوش مصنوعی گفته شده بود که امتیاز فعلی را با برون داد اعدادی در فواصل زمانی منظم که به ما می گویند کدامیک از کلیدها را فشار دهیم، به حداکثر برساند. در آغاز، هوش مصنوعی بسیار بد بازی می کرد و با گیجی، حرکات آهسته و به طور تصادفی، صفحه رابط را به این ور و آن ور حرکت می داد و دائما گلوله را گم می کرد. بعد از گذشت مدتی، متوجه شد که حرکت دادن صفحه رابط به سمت گلوله ایده خوبی است. با این حال، هنوز در بیشتر مواقع، گلوله را گم می کرد. سپس، با انجام تمرین بیشتر توانست مهارت خود را بهبود دهد و بدون توجه به سرعت نزدیک شدن گلوله، آن را باز گرداند. سپس، به فکر راهبردی برای کسب حداکثر امتیاز افتادم و آن این بود که سوراخی را در گوشه سمت چپ دیوار آجری ایجاد کنم و به گلوله اجازه دهم که بین پشت دیوار و حائل پشت آن گیر کند و به این ور و آن ور بپرد. احساس کردم که این یک کار هوشمندانه برای کسب امتیاز است. بعدا معلوم شد که تیم برنامه نویسان دیپ مایند از این حقه آگاه نبودند و آن را از ماشین هوش مصنوعی یاد گرفته بودند. جنبه نگران کننده هوش مصنوعی برای انسان ها، این بود که هوش مصنوعی هدفی را دنبال می کرد و یاد گرفته بود که مهارت دستیابی به آن هدف را در خود افزایش دهد تا در نهایت، بتواند انسان سازنده آن را شکست دهد. همانطور که گفتیم، هوش به معنی توانایی دستیابی به اهداف پیچیده می باشد و هوش مصنوعی دیپ مایند در زمینه انجام این بازی هوشمندتر از انسان شده بود. همانطور که در فصل اول گفته شد، ما با عوامل هوشمندی مواجه هستیم که اطلاعات محیطی را از طریق حسگرهای خود جمع آوری می کنند و سپس، آنها را مورد پردازش قرار می دهند تا بتوانند در مورد نحوه واکنش مناسب خود نسبت به محیط تصمیم گیری کنند. دیپ مایند روش خود را به زودی منتشر کرد و کد برنامه نویسی آن را به اشتراک گذاشت و توضیح داد که از ایده بسیار ساده اما قوی موسوم به "یادگیری تقویت عمیق" استفاده کرده است. "یادگیری تقویت اولیه"، یک تکنیک یادگیری ماشین است که توسط "روانشناسی رفتاری" ایجاد شده و می

گوید که گرفتن یک پاداش مثبت، تمایل فرد برای انجام دوباره آن را افزایش می دهد و بر عکس. مانند یک سگ که یاد می گیرد که اگر کارش را خوب انجام دهد، احتمال تشویق شدن و گرفتن تنقلات از صاحبش، افزایش می یابد، هوش مصنوعی دیپ مایند نیز یاد گرفته است که اگر صفحه رابط را به محل خاصی جابه جا کند، احتمال اینکه امتیازهای بیشتری کسب کند، افزایش خواهد یافت. دیپ مایند این تکنیک را با "یادگیری عمیق" ترکیب کرده است. به این صورت که آنها به یک شبکه عصبی عمیق آموزش داده اند که به پیش بینی این موضوع پردازد که با فشار دادن هر یک از کلیدهای مجاز صفحه کلید، به طور متوسط چند امتیاز حاصل خواهد شد تا با فرض حفظ حالت فعلی بازی، هوش مصنوعی دست به انتخاب کلیدهایی بزند که شبکه عصبی به عنوان امیدوارکننده ترین کلیدها نمره گذاری شان کرده است. یکی از چیزهایی که به افزایش عزت نفس ما انسان ها کمک می کند، این است که از پس طیف گسترده ای از مسائل حل نشده برآییم. به نظر من، آنچه که در پیشرفت مهم دیپ مایند واقعا اهمیت دارد، این است که یادگیری تقویت عمیق، یک تکنیک کاملا عمومی است و هوش مصنوعی در بیست و نه مورد از چهل و نه بازی آتاری مختلف، یاد گرفته است که بهتر از انسان بازی کند که این بازی ها شامل تیس روی میز، بوکس، ویدئو پین بال و مهاجمان فضایی می باشند. طولی نکشید که ایده هوش مصنوعی، توانایی های خودش را در بازی های پیشرفته تر در دنیای سه بعدی نیز نشان داد. با نگاهی به آینده یادگیری تقویت عمیق و بهبودهای ممکن بر روی آن، می توان گفت که هیچ پایان قابل حصول روشنی برایش وجود ندارد. پتانسیل هوش مصنوعی به دنیای بازی های مجازی محدود نمی شود. استوارت راسل از احساس شگفتی خود از مشاهده ربات بیگ داگ سخن می گوید که به سرعت از سربالایی پوشیده از برف جنگل بالا می رود و بر مشکل حرکت پای خود که سال ها برای حل آن تلاش می کرد، غلبه کرده است. با این حال، بخش بزرگی از این کار توسط برنامه نویسان با هوش انجام شده بود. پس از پیشرفت مهم هوش مصنوعی دیپ مایند، دلیلی برای این وجود نداشت که چرا یک ربات نتواند از یادگیری تقویت عمیق برای یادگیری قدم زدن بدون کمک گرفتن از انسان های برنامه نویس استفاده کند. فقط به

سیستمی نیاز بود که با مشاهده هر گونه پیشرفتی از سوی ربات، به آن امتیاز دهد. ربات ها در دنیای واقعی از توانمندی و پتانسیل لازم برای یادگیری فعالیت هایی مانند شنا کردن، پرواز کردن، انجام بازی تنیس روی میز، جنگیدن و اجرای فهرست طولانی ای از فعالیت های حرکتی دیگر بدون نیاز به کمک گرفتن از انسان های برنامه نویس برخوردار هستند. ربات ها برای تسریع بخشیدن به کارها و کاهش خطر گیر کردن یا آسیب رساندن به خود در حین فرایند یادگیری، به احتمال زیاد، مراحل اولیه یادگیری خود را در دنیای مجازی انجام می دهند.

### حس درونی، خلاقیت و راهبرد

لحظه تعیین کننده دیگر برای من وقتی بود که سیستم هوش مصنوعی آلفاگو در بازی "گو" و پس از یک رقابت پنج روزه با لی سه دال، قهرمان جهان در قرن بیست و یکم، موفق شد که او را شکست دهد. کاملاً انتظار می رفت که بازیکنان انسان، در مرحله ای از بازی مغلوب ماشین ها شوند، زیرا در دو دهه قبل، همین اتفاق برای آنها در بازی شطرنج افتاده بود. با این حال، بیشتر کارشناسان بازی گو پیش بینی کرده بود که تا یک دهه دیگر نیز ماشین ها بر انسان ها حاکم می باشند و اینکه انسان ها از مشاهده پیشرفت های مهم حاصل شده در زمینه هوش مصنوعی، بسیار ناراحت خواهند شد. سیستم هوش مصنوعی پیشرفته آلفاگو در تمام بیست بازی ای که طی یک سال اخیر با بازیکنان برتر جهان انجام داده بود، حتی در یک بازی هم مغلوب انسان ها نشده بود. من اعتراف می کنم که سیستم آلفاگو، حس درونی و خلاقیت بیشتری را نسبت به انسان ها از خود نشان می دهد. در بازی "گو"، بازیکنان به نوبت مهره های سیاه و سفید را روی یک تخته ۱۹ در ۱۹ قرار می دهند که در اینجا بخشی از آن نمایش داده شده است:



تعداد موقعیت های ممکن در بازی گو از تعداد کل اتم های موجود در جهان بیشتر است و این بدین معناست که تلاش برای تجزیه و تحلیل سریع کلیه حرکت های متوالی بعدی، به نا امید منجر خواهد شد. به همین دلیل، بازیکنان بازی گو برای تکمیل استدلال های آگاهانه خود، به شدت بر روی حس درونی نیمه آگاهشان، اعتماد می کنند و بازیکنان کارکشته از احساسات غیرعادی خود برای تشخیص اینکه کدامیک از موقعیت ها قوی و کدامیک ضعیف هستند، استفاده می کنند. همانطور که در بخش قبلی دیدیم، نتایج یادگیری عمیق گاهی تداعی گر حس درونی ما هستند. یک شبکه عصبی عمیق ممکن است تشخیص دهد که تصویر مورد نظر نشان دهنده یک گربه است و در عین حال، قادر به ارائه هیچ دلیلی برای تشخیص خود نباشد. تیم دیپ مایند روی این ایده شرط بندی کردند که یادگیری عمیق می تواند علاوه بر تشخیص گربه ها، موقعیت های قوی بازی گو را نیز تشخیص دهد. هدف اصلی تیم دیپ مایند از ساختن آلفاگو این بود که نیروی حس درونی مربوط به یادگیری عمیق را با نیروی منطق گوفای<sup>۳۴</sup> (هوش مصنوعی خوب قدیمی یا هوش مصنوعی سمبولیک که قبل از انقلاب یادگیری عمیق ایجاد شده بود)، ترکیب کنند. آنها از یک پایگاه داده وسیع برای ذخیره سازی و پردازش موقعیت های بازی گو استفاده کردند که هم شامل بازی میان دو انسان و هم شامل بازی دو ماشین همسان آلفاگو می شد و به یک شبکه عصبی عمیق یاد دادند که در هر موقعیتی، احتمال اینکه تیم سفید در نهایت برنده شود را پیش بینی کند. آنها همچنین به یک شبکه عصبی جداگانه دیگر، یاد دادند که حرکت های احتمالی بعدی را پیش بینی کند. آنها سپس این دو شبکه را با روش گوفای ترکیب کردند تا به بررسی هوشمندانه کلیه جوانب مربوط به فهرست کوتاه شده ای از حرکت های احتمالی بعدی پردازند و آن حرکت بعدی ای را که منجر به قویترین موقعیت در آینده می شود، شناسایی کنند. پیوند میان حس درونی و منطق منجر به انجام حرکت هایی شده بود که نه تنها قوی بودند، بلکه در برخی موارد، کاملاً خلاقانه بودند. به عنوان مثال، منطق حکم می کرد که در اوایل بازی، بهتر است در خط سوم یا چهارم از کنار لبه تخته به بازی ادامه دهیم و

---

<sup>۳۴</sup> Good old fashioned AI

مصالحه ای را میان این دو خط ایجاد کنیم. ادامه بازی در خط سوم از تصاحب کوتاه مدت منطقه ای واقع در حوالی کنار تخته حمایت می کرد و این در حالی بود که ادامه بازی در خط چهارم به نفوذ راهبردی بلندمدت به حوالی مرکز تخته کمک می کرد. در حرکت سی و هفتم از بازی دوم، آلفاگو به واسطه مخالفت با منطق قدیمی و ادامه بازی در خط پنجم، شگفتی بزرگی را رقم زد. سیستم آلفاگو که بیش از یک انسان به قدرت برنامه ریزی بلندمدت خود، مطمئن بود، مزیت راهبردی بلندمدت را به کسب سود کوتاه مدت ترجیح داد. همه مفسران شگفت زده شده بودند و لی سه دال، قهرمان سابق این بازی از جای خود بلند شد و محل بازی را به طور موقت ترک کرد. در نهایت، ماشین هوش مصنوعی آلفاگو برنده بازی شد و ادامه بازی در سطر پنجم به عنوان یکی از خلاقانه ترین حرکت ها در تاریخ بازی گو، ثبت شد. جنبه های مربوط به حس درونی و خلاقیت فردی در بازی گو موجب شد که این بازی به عنوان یکی از چهار هنر اساسی در چین باستان شناخته شود. دنیای بازی گو به خاطر نتیجه ای که حاصل شد، تکان بزرگی خورد و پیروزی آلفاگو به عنوان نقطه عطفی در تاریخ بشریت محسوب شد. در سال ۲۰۱۷ تیم دیپ مایند، آلفاژیرو را به عنوان جانشین آلفاگو روانه بازار کرد. آلفاژیرو نه تنها موفق به شکست آلفاگو شد، بلکه یاد گرفت که از طریق "بازی با خود" به قوی ترین شطرنج باز جهان تبدیل شود و پس از دو سال تمرین، توانست بر بهترین بازیکنان شطرنج انسان غلبه کند و پس از چهار سال، توانست شکست سختی را بر استو کفیش، بهترین برنامه شطرنج جهان تحمیل کند و با این کار، موجب شد که کلیه نرم افزارهای هوش مصنوعی دست ساز که طی چندین دهه توسعه پیدا کرده بودند، کنار گذاشته شوند. به عبارت دیگر، ما نمی توانیم ایده "ایجاد یک هوش مصنوعی جدید را که بهتر از هوش مصنوعی قبلی، عمل می کند"، از سر بیرون کنیم. یکی از درس هایی که از آلفاگو گرفتم، این بود که ترکیب حس درونی مربوط به یادگیری عمیق با منطق گوفای می تواند به تولید راهبرد بی نظیری که دومی ندارد، کمک کند. هوش مصنوعی در حال حاضر باعث افزایش اعتماد به نفس فارغ التحصیلان شده و بهترین راهبردهای انسانی مانند راهبرد سرمایه گذاری، راهبرد سیاسی و راهبرد نظامی را به چالش کشانده یا



به آنها کمک کرده است. هرچند که عواملی مانند علم روانشناسی انسانی، اطلاعات گم شده و عوامل تصادفی دیگری که نیاز به مدلسازی دارند، باعث پیچیده تر شدن مسائل راهبردی در دنیای واقعی شده است، سیستم های هوش مصنوعی بازی پوکر نشان داده اند که هیچ یک از این چالش ها، غیر قابل حل نیستند.

## زبان طبیعی

یکی از حوزه های دیگر در زمینه هوش مصنوعی که پیشرفت زیادی کرده است، حوزه زبان طبیعی بوده است. من در اوایل زندگی خود عاشق سفر کردن و کنجکاوی در مورد فرهنگ ها و زبان های دیگر بودم و زبان های سوئدی، انگلیسی، آلمانی، اسپانیایی، پرتغالی، رومانیایی، روسی و فرانسوی را یاد گرفته بودم. جمله زیر را در نظر بگیرید: "پس از کشف مهمی که در سال ۲۰۱۶ صورت گرفت، سیستم هوش مصنوعی توسعه یافته توسط تیم گوگل برین<sup>۳۵</sup> توانست که بهتر از انسان، زبانی را به زبان زبان دیگر ترجمه کند." یکی از عیب های برنامه کاربردی هوش مصنوعی برای ترجمه این بود که عبارت "تیم گوگل برین" را به صورت "تیم مغز گوگل" ترجمه کرده بود و باعث شده بود که معنی آن تغییر پیدا کند. با این حال، تیم انسانی گوگل برین در سال ۲۰۱۶ توانستند برنامه کاربردی خدمات ترجمه مجانی گوگل را با بهره گیری از شبکه های عصبی بازگشتی عمیق<sup>۳۶</sup> ارتقا دهند و پیشرفت های صورت گرفته در سیستم قدیمی گوفاي چشمگیر بود. با این حال، اغلب از من انتقاد می شود که هوش مصنوعی گوگل از جملات طولانی غیر ضروری که تجزیه آنها مشکل است، استفاده می کند. البته باید توجه داشت که هوش مصنوعی در زمینه ترجمه جملات دیگر، بدون عیب و نقص عمل می کند. در نتیجه، وقتی که سیستم ترجمه هوش مصنوعی عرضه شد، جنبش بزرگی را ایجاد کرد و این سیستم هر روز مورد استفاده صدها میلیون نفر در جهان واقع می شود. علاوه بر این، به لطف پیشرفت های اخیر در زمینه یادگیری عمیق برای تبدیل گفتار به متن و متن به گفتار، استفاده

---

<sup>۳۵</sup> Google brain team

<sup>۳۶</sup> Deep recurrent neural network

کنندگان می توانند با تلفن هوشمند خود به یک زبان صحبت کنند و به ترجمه گفتاری آن به زبانی دیگر، گوش دهند. پردازش زبان طبیعی یکی از حوزه های هوش مصنوعی است که پیشرفت های سریعی در آن حاصل شده است و من فکر می کنم که کسب موفقیت های بیشتر در این زمینه بسیار تاثیر گذار خواهد بود، زیرا داشتن زبان یکی از ویژگی های مهم انسان هاست. هرچه قدر که هوش مصنوعی در زمینه پیش بینی کلامی پیشرفت کند، پاسخ های ایمیل منطقی تری خواهد ساخت و ارتباط کلامی بهتری برقرار خواهد کرد و حداقل به یک انسان غریبه نشان خواهد داد که قادر به فکر کردن است. سیستم های یادگیری عمیق در حال برداشتن گام های اولیه به سمت موفقیت در آزمون معروف تورینگ می باشد که در آن، یک ماشین باید در زمینه ارتباط نوشتاری، به قدری مهارت پیدا کند که انسان ها فکر کنند که آن ماشین نیز، یک انسان است. با این حال، ماشین های هوش مصنوعی پردازش کننده زبان، هنوز راه درازی برای پیمودن دارند. وقتی که با ترجمه نادرستی مواجه می شوم، به خودم یادآوری می کنم که یک ماشین هوش مصنوعی، محتوای چیزی را که می گوید، نمی فهمد و با آموزشی که برای کارکردن با حجم وسیعی از داده ها دیده است، فقط الگوها و روابطی را بین لغات کشف می کند و معنی آن لغات را در دنیای واقعی نمی داند. به عنوان مثال، ماشین هوش مصنوعی برای تشخیص تشابه یک سری لغات با لغات دیگر، ممکن است هر لغتی را با هزاران عدد تجسم کند و مثلاً نتیجه گیری کند که تفاوت بین "شاه" و "ملکه" شبیه ارتباط بین "شوهر" و "همسر" می باشد و در عین حال، هیچ اطلاعی از معنی مذکر و مونث بودن یا واقعیت های فیزیکی زمانی، مکانی و مادی نداشته باشد. با توجه به فریبنده بودن آزمون تورینگ، از آن انتقاد شده است که ساده لوحی انسان ها را آزمایش می کند. در عوض، آزمون رقیبی موسوم به وینوگراد شم چلنج<sup>۳۷</sup>، مستقیماً سراغ قدرت درک ناشی از عقل سلیمی می رود که سیستم های یادگیری عمیق، فاقد آن هستند. ما انسان ها معمولاً از دانش خود برای تجزیه یک جمله استفاده می کنیم تا تشخیص

---

<sup>۳۷</sup> Winograd schema challenge

دهیم که هر ضمیری در جمله به چه چیزی اشاره می کند. به عنوان مثال، آزمون وینوگراد در جملات زیر از شما می پرسد که واژه "آنها" به چه چیزی اشاره می کند:

- ۱- اعضای شورای شهر به معترضین اجازه تظاهرات ندادند، چونکه آنها از خشونت می ترسیدند.
- ۲- اعضای شورای شهر به معترضین اجازه تظاهرات ندادند، چونکه آنها از خشونت حمایت می کردند.

این چالش که هر ضمیری به چه چیزی اشاره می کند و بسیاری از چالش های دیگر، باعث ایجاد رقابت فشرده ای میان ابزارهای ترجمه هوش مصنوعی شده است. با این حال، هوش مصنوعی گوگل پیشرفت های زیادی کرده است و با استفاده از ترکیبی از شبکه های عصبی بازگشتی عمیق با منطق گوفای، فرصت های امیدوار کننده ای را برای ساختن یک مدل جهانی هوش مصنوعی برای پردازش زبان در جهان ایجاد کرده است.

## فرصت ها و چالش ها

این سه مثالی که ارائه شد، به عنوان نمونه بود و هوش مصنوعی در حال پیشرفت سریع از میان حوزه های مختلف و مهم دیگری نیز می باشد و رقابت سختی در بین گروه های تحقیقاتی دانشگاهی و بسیاری از شرکت ها برای اینکه از رقبای خود عقب نمانند، ایجاد شده است و واحدهای علوم رایانه در سراسر جهان از جمله شرکت های اپل، دیپ مایند، فیس بوک، گوگل، میکروسافت و غیره از پیشنهاد های پر درآمدی برای جذب دانشجویان، پسا دکتراها و اساتید استفاده می کنند. به نظر من، پیشرفت های نسبتاً پایداری در طولانی مدت ایجاد خواهد شد که رسانه ها از آنها به عنوان پیشرفت های مهم در زمینه معرفی برنامه های کاربردی جذاب جدید یا محصولات مفید، یاد می کنند. من احتمال می دهم که پیشرفت های سریع حاصل شده در زمینه هوش مصنوعی تا چندین سال آینده ادامه پیدا کند و این پیشرفت ها تا زمانی ادامه پیدا کند که هوش مصنوعی از عهده بسیاری از کارهایی که بشر می تواند انجام دهد، بر آید. اکنون سوالات زیادی مطرح می شود، از جمله اینکه

هوش مصنوعی چه تاثیری بر ما خواهد گذاشت و اینکه انسانیت ما انسان ها در اثر پیشرفت های هوش مصنوعی دچار چه تحولی خواهد شد؟ همانطور که دیدید، به سختی می توان استدلال کرد که هوش مصنوعی کاملاً فاقد ویژگی های مهم بشری از جمله هدف گذاری، وسعت دید، حس شهودی، خلاقیت یا سخن گفتن می باشد. این بدین معناست که حتی در آینده نزدیک و مدت ها قبل از اینکه هوش عمومی مصنوعی بتواند تمام کارهای انسان را انجام دهد، می توانیم شاهد تاثیرات چشمگیر هوش مصنوعی بر دیدگاه ما نسبت به خودمان باشیم و بینیم که چگونه می توانیم مکمل فعالیت های هوش مصنوعی باشیم یا بر اثر رقابت با آن، پول کسب کنیم؟ همچنین ممکن است از خود پرسید که تاثیرات هوش مصنوعی به نفع ما خواهد بود یا به ضرر ما و اینکه چه فرصت ها و چالش هایی را در کوتاه مدت ایجاد خواهد کرد؟ تمدن امروزی ما حاصل هوش بشر است و اگر بتوانیم آن را با هوش مصنوعی تقویت کنیم، از پتانسیل داشتن یک زندگی بهتر برخوردار خواهیم شد. حتی پیشرفت های ساده در زمینه هوش مصنوعی ممکن است منجر به پیشرفت های مهمی در زمینه علم و فن آوری و کاهش حوادث، بیماری ها، بی عدالتی، جنگ، کارهای پر زحمت و فقر شوند. برای استفاده از مزایای هوش مصنوعی و کاهش مشکلات ناشی از آن، باید به سوالات زیر پاسخ دهیم:

- ۱- چگونه می توانیم سیستم های هوش مصنوعی آینده را مقاوم تر از سیستم های امروزی بسازیم، به طوریکه بدون از کار افتادن، بدعمل کردن یا مورد هک واقع شدن، بتوانند به خواسته های ما عمل کنند؟
- ۲- چگونه می توانیم سیستم های حقوقی خود را به نحوی به روز رسانی کنیم که منصفانه تر و کارا تر شوند و با تغییرات سریع ایجاد شده در دورنمای دنیای دیجیتالی، همگام شوند؟
- ۳- چگونه می توانیم سلاح ها را هوشمند تر و با گرایش کمتر به کشتار افراد غیر نظامی و بی گناه بسازیم و از به راه انداختن یک رقابت تسلیحاتی کنترل نشده با استفاده از سلاح های خودکار کشنده جلوگیری کنیم؟

۴- چگونه می توانیم بدون رها نکردن افرادی که درآمد یا هدف مشخصی ندارند، پیشرفت خود از طریق خودکارسازی را افزایش دهیم؟

ادامه این فصل را صرف پاسخ گویی به هر یک از این سوالات خواهیم کرد. این چهار سوال به ترتیب، مناسب دانشمندان علوم رایانه، دانش پژوهان حقوقی، استراتژیست های نظامی و اقتصاددانان می باشند. با این حال، اگر بخواهیم که پاسخ های مورد نیاز خود را در زمان مورد نیاز دریافت کنیم، همه این افراد باید وارد گفتگوها شوند، زیرا چالش های موجود فراتر از کلیه مرزهای سنتی موجود بین کارشناسان و ملت ها می باشند.

### هوش مصنوعی خطادار در مقابل هوش مصنوعی مقاوم

فن آوری اطلاعات تاکنون تاثیرات مثبت فراوانی بر تمام بخش های انسانی، از بخش علوم گرفته تا بخش های مالی، ساخت و تولید، حمل و نقل، بهداشت، انرژی و ارتباطات داشته است و این تاثیرات در مقایسه با تاثیرات بالقوه ناشی از پیشرفت های هوش مصنوعی بسیار کمزنگ تر بوده است. اما هرچه قدر که بیشتر به فن آوری ها تکیه می کنیم، مقاوم بودن و قابل اعتماد بودنشان برایمان مهم تر می شوند. در تاریخ بشر، از رویکرد محک زدن و یادگیری از اشتباهات برای اطمینان از مفید بودن فن آوری، زیاد استفاده شده است. ما آتش را اختراع کردیم. به طور مکرر دست به گل آب دادیم و بالاخره خاموش کننده آتش، خروجی اضطراری، سیستم اعلام حریق و آتش نشانی را اختراع کردیم. ما خودرو را اختراع کردیم، به طور مکرر تصادف کردیم تا اینکه بالاخره کمربند ایمنی، کیسه هوا و خودروهای بدون راننده را اختراع کردیم. تاکنون میزان تصادفات به قدری پایین بوده که سودهای ناشی از فن آوری خودرو بر ضررش سنگینی کرده است. با این حال، همین که ما به طور بی وقفه فن آوری های قوی تری را تولید می کنیم، ناگزیر به جایی خواهیم رسید که حتی یک تصادف ساده به قدری برایمان ناگوار خواهد شد که ممکن است بر مزایای فن آوری جدید سنگینی کند. بنابراین، وقتی که فن آوری قوی تر می شود، مهندسی ایمنی اهمیت بیشتری پیدا می کند و ما برای

افزایش ایمنی فن آوری های جدید، دیگر نمی توانیم از رویکرد آزمون و خطا استفاده کنیم. به عبارت دیگر، ما باید به جای رویکرد انفعالی، رویکرد فعالانه تری را در پیش بگیریم و در زمینه تحقیقات ایمنی به گونه ای سرمایه گذاری کنیم که حتی از وقوع یک حادثه جلوگیری کنیم و به همین منظور، لازم است که جامعه سرمایه گذاری بیشتری را در زمینه ایمنی رآکتورهای هسته ای انجام دهد. اهمیت این سرمایه گذاری را می توان در تمایل جامعه به انجام تحقیقات مربوط به ایمنی هوش مصنوعی در کنفرانس پورتو ریکو که در فصل یک به آن اشاره شد، مشاهده کرد. رایانه ها و سیستم های هوش مصنوعی همواره در معرض از کار افتادن قرار دارند. اما این بار فرق می کند. هوش مصنوعی به تدریج در حال ورود به دنیای واقعی است. اگر هوش مصنوعی باعث از کارافتادن شبکه برق، بازار سهام یا سیستم سلاح های هسته ای شود، این فقط یک مزاحمت نیست و باید آن را پی گیری کرد. در ادامه این بخش به معرفی چهار حوزه اصلی در تحقیقات فنی مربوط به ایمنی هوش مصنوعی می پردازم که شامل درستی سنجی عملکرد، اعتبار سنجی فرضیات، حفظ امنیت و کنترل می باشند. این حوزه ها در بحث فعلی ما نقش مهمی دارند و در دنیا به شدت پیگیری می شوند. برای اینکه سخنانم خشک و بی روح نشوند، به بررسی موفقیت ها و شکست های قبلی فن آوری اطلاعات در حوزه های مختلف ایمنی هوش مصنوعی و درس های با ارزشی می پردازم که می توانیم از آنها یاد بگیریم. هرچند که بیشتر این داستان ها قدیمی هستند و شامل سیستم های رایانه ای دارای فن آوری پایین می باشند که خسارت کمتری از هوش مصنوعی به دنبال دارند، اما در عوض درس های با ارزشی را برای طراحی سیستم های هوش مصنوعی قوی در آینده، به ما می دهند و به ما نشان می دهند که اگر این سیستم ها به خوبی طراحی نشوند، می توانند منجر به چه فاجعه های بزرگی شوند.

## هوش مصنوعی برای کشف فضا

فن آوری رایانه ای ما را قادر ساخته که به کره ماه پرواز کنیم و برای کشف سیاره های موجود در منظومه شمسی، فضاپیماهای بدون سرنشینی را به فضا بفرستیم که حتی روی تیتان که بزرگ ترین قمر

سیاره کیوان (زحل) می باشد، فرود آیند. هوش مصنوعی بدون خطا ممکن است که در آینده به کشف منظومه های شمسی و کهکشان های دیگر نیز کمک کند. در سال ۱۹۹۶ دانشمندان با امید به انجام تحقیقات در مورد مغناطیس سپهر زمین، با خوشحالی و سرمستی، موشک آرین پنج را به همراه تجهیزات علمی ساخته شده خود، از آژانس فضایی اروپا، به فضا پرتاب کردند. اما، وقتی که ۳۷ ثانیه بعد، موشک منفجر شد، لبخندشان نا پدید شد. علت انفجار موشک یک خطای نرم افزاری بود که به دلیل استفاده از یک عدد که در ۱۶ بیت اختصاص یافته به آن جا نمی شد، اتفاق افتاده بود. دو سال بعد، مدار گرد اقلیمی مریخ توسط ناسا به فضا پرتاب شد و به طور تصادفی وارد جو مریخ شد و متلاشی شد، زیرا دو بخش مختلف نرم افزار از واحدهای مختلفی برای اندازه گیری فشار استفاده کرده بودند که این امر منجر به خطا در سیستم کنترل موتور موشک شده بود. شرکت ناسا قبلا هم خطای دیگری را تجربه کرده بود. در سال ۱۹۶۲، فضاییمارینر ۱ ناسا در زمان انجام ماموریت سفر به سیاره زهره منفجر شد که علت آن، استفاده از علامت نگارشی نادرست (مانند نقطه یا ویرگول) در نرم افزار کنترل پرواز بود. ماموریت کاوشگر فضایی فوبوس شوروی هم در سال ۱۹۸۸ با شکست مواجه شد که علت آن هم فقدان یک خط رابط در نرم افزار کنترل پرواز بود که منجر به بروز خطا در سیستم شده بود. درسی که ما از این مثال ها می گیریم، این است که توجه بیشتری به موضوع "درستی سنجی عملکرد" داشته باشیم که دانشمندان علوم رایانه بسیار روی آن تاکید می کنند و به معنی اطمینان کامل از انجام درست عملکردها و وظایف و رعایت کلیه الزامات نرم افزاری می باشد. با توجه به خطرات موجود، لازم است که مطمئن شویم که نرم افزار به آنچه که از آن می خواهیم، عمل خواهد کرد. خوشبختانه، هوش مصنوعی می تواند به خود کارسازی و بهبود فرایند "درستی سنجی عملکرد" کمک کند. به عنوان مثال، هسته یک سیستم عامل رایانه ای موسوم به self، اخیرا برای بررسی سختگیرانه "مشخصات رسمی"<sup>۳۸</sup> یک نرم افزار مورد استفاده قرار گرفته تا اطمینان قوی حاصل شود که خطای از کار افتادن و عملیات نادرست در آن اتفاق نخواهد افتاد و شما

---

<sup>۳۸</sup> Formal specification

می توانید مطمئن باشید که خطر مرگ را احساس نخواهید کرد. آژانس پروژه های تحقیقاتی پیشرفته دفاعی (DARPA) آمریکا، از توسعه مجموعه ای از ابزارهای متن باز<sup>۳۹</sup> با قابلیت اطمینان بالا موسوم به سیستم های سایبری- نظامی با قابلیت اطمینان بالا (HACMS) که ایمن بودن آن ها ثابت شده، حمایت مالی کرده است. چالش مهمی که با آن مواجه هستیم این است که این ابزارها را به گونه ای بسازیم که به حد کافی قدرتمند و قابل استفاده راحت باشند تا بتوان آنها را به طور گسترده ای مورد استفاده قرار دارد. چالش دیگر این است که وقتی نرم افزار هوش مصنوعی وارد ربات ها و محیط های جدید می شود و سیستم های هوش مصنوعی دارای قابلیت تغییر رفتار خود از طریق یادگیری، جایگزین نرم افزارهای قدیمی از پیش برنامه ریزی شده می شوند، ممکن است که کار "درستی سنجی عملکرد" به سختی انجام شود.

### هوش مصنوعی برای امور مالی

حوزه مالی، حوزه دیگری است که به دلیل فن آوری اطلاعات متحول شده است و امکان تخصیص مجدد منابع مالی در سراسر جهان را به شکل کارا و با سرعت نور فراهم ساخته و زمینه را برای تامین منابع مالی قابل دسترس برای هر چیزی از وام مسکن گرفته تا راه اندازی شرکت های نوپا، ایجاد کرده است. پیشرفت های صورت گرفته در زمینه هوش مصنوعی باعث ایجاد فرصت های بزرگی برای بهره برداری از سود آتی ناشی از معاملات مالی خواهد شد. امروزه، بیشتر تصمیم گیری ها برای خرید/ فروش بازار سهام به صورت خودکار و از طریق رایانه ها صورت می گیرد و دانشجویان فارغ التحصیل از دانشگاه ام آی تی معمولاً وسوسه می شوند که درآمدهای ابتدایی نجومی ای را برای بهبود دستورالعمل های معاملات الگوریتمی دریافت کنند. "درستی سنجی عملکرد" برای نرم افزار مالی هم مهم است. در سال ۲۰۱۲ شرکت آمریکایی نایت کپیتال، بعد از استفاده از نرم افزار تجاری تأیید نشده<sup>۴۰</sup> که باعث ضرر ۴۴۰ میلیون دلاری این شرکت شد، از اشتباه خود، درس عبرت گرفت.

---

<sup>۳۹</sup> Open source

<sup>۴۰</sup> Unverified trading software



اما، ضرر میلیارد دلاری "سقوط سریع"<sup>۴۱</sup> که در سال ۲۰۱۰ در آمریکا اتفاق افتاد، دلیل دیگری داشت. با وجود اینکه نیم ساعت قبل از تثبیت دوباره بازار، سقوط سریع منجر به اختلال وسیعی در بازارها و نوسان قیمت سهام شرکت های مهمی مانند پروکتر و گمبل شده بود، مشکل ایجاد شده ناشی از خطاها یا عملکرد بد رایانه نبود و ربطی به درستی سنجی عملکرد نداشت. در عوض، مشکل این بود که انتظارات، از حد خود فراتر رفته بود. برنامه های رایانه ای مربوط به معاملات خود کار بسیاری از شرکت ها به دلیل نامعتبر بودن فرضیاتشان، در وضعیت غیر منتظره ای کار می کنند. به عنوان مثال، ممکن است این طور فرض شود که اگر در گزارش رایانه ای تبادل سهام درج شود که قیمت سهمی یک سنت است، پس ارزش واقعی آن سهم نیز همان یک سنت است. مثال سقوط سریع نشان دهنده اهمیت آن چیزی است که دانشمندان علوم رایانه از آن به عنوان "اعتبار سنجی فرضیات" یاد می کنند. در حالی که درست سنجی عملکرد می پرسد: "آیا سیستم را به درستی ساخته ام؟"، اعتبار سنجی فرضیات می پرسد: "آیا سیستم درستی را ساخته ام؟" درست سنجی عملکرد بررسی می کند که آیا سیستم به مشخصات خود عمل می کند و اعتبار سنجی بررسی می کند که آیا مشخصات درستی انتخاب شده اند؟ به عنوان مثال، در اعتبار سنجی فرضیات بررسی می شود که آیا احتمال دارد که سیستم موجود متکی بر فرضیاتی باشد که ممکن است همیشه درست نباشند و همچنین، بررسی می شود که چگونه می توان عملکرد آن را برای مدیریت بهتر شرایط نامطمئن، بهبود بخشید؟

## هوش مصنوعی برای ساخت و تولید

هوش مصنوعی از پتانسل بالایی برای بهبود تولید از طریق کنترل ربات هایی برخوردار است که کارایی و دقت تولید را افزایش می دهند. چاپگرهای سه بعدی پیشرفته می توانند مدل های آزمایشی هر چیزی از ساختمان های اداری گرفته تا وسایل میکرومکانیکی کوچکتر از دانه نمک را بسازند. ربات های صنعتی بزرگ می توانند خودروها و هواپیماها را بسازند. همچنین، ماشین های تراش و

---

<sup>۴۱</sup> Flash crash

دستگاه های برش که توسط رایانه کنترل می شوند باعث افزایش قدرت کارخانه ها و تقویت جنبش سازندگان شده و طرفداران این جنبش، در لابراتورهای سراسر جهان می توانند به ایده های خود جامه عمل بپوشانند. در مورد ربات ها نیز باید درستی سنجی عملکرد و اعتبار سنجی فرضیات، انجام شود. اولین شخصی که توسط یک ربات کشته شد، رابرت ویلیامز بود که در کارخانه فورد کار می کرد. وقتی که ربات مسئول قطعه گیری از انبار، خراب شد، او برای گرفتن قطعه وارد انبار شد و ربات که تازه به آرامی به کار افتاده بود، سر او را خرد کرد. قربانی دیگر ربات ها یک فرد ژاپنی بود که در کارخانه کاواساکی کار می کرد. وقتی که او مشغول کار کردن روی قطعات شکسته یک ربات بود، تصادفا دکمه روشن را فشار داد و با ضربه بازوی هیدرولیک ربات کشته شد. در آلمان نیز یک فرد جوان که ربات را برای گرفتن قطعات خودرو تنظیم می کرد، توسط ربات گرفته شد و در اثر برخورد شدید با یک صفحه فلزی کشته شد. با وجود اینکه این تصادفات غم انگیز هستند، بخش کوچکی از تصادفات صنعتی را تشکیل می دهند. البته، با پیشرفت فن آوری، تعداد افراد فوت شده در آمریکا بر اثر چنین حوادثی از ۱۴۰۰۰ نفر در سال ۱۹۷۰ به ۴۸۲۱ نفر در سال ۲۰۱۴ کاهش یافته است. این مثال ها نشان می دهند که همزمان با افزودن هوش به ماشین های لال، باید به افزایش ایمنی صنعت نیز توجه کرد و به این ماشین ها یاد داد که محتاط تر باشند. افزایش اعتبار سنجی فرضیات منجر به کاهش تصادفات خواهد شد، زیرا این تصادفات به دلیل وجود خطا در سیستم نبوده، بلکه به دلیل فرضیات نامعتبر ماشین بوده است، از جمله اینکه انسان هم یک قطعه خودرو است.

## هوش مصنوعی برای حمل و نقل

در سال ۲۰۱۵، تصادفات خودرو جان بیش از ۱,۲ میلیون نفر در آمریکا را گرفت و تصادفات هواپیما، قطار و کشتی منجر به مرگ هزاران نفر دیگر شد. در کشور آمریکا که دارای استانداردهای ایمنی بالایی است، تصادفات موتور سیکلت در سال گذشته باعث مرگ ۳۵۰۰۰ نفر شد. در میزگرد سالیانه ای که در سال ۲۰۱۶ توسط انجمن پیشرفت های هوش مصنوعی در تگزاس برگزار شد، یکی از دانشمندان علم رایانه با هیجان زیادی استدلال می کرد که هوش مصنوعی می تواند تلفات جاده را

کاهش دهد و حتما این کار را خواهد کرد، زیرا این کار از لحاظ اخلاقی واجب است. از آنجایی که بیشتر تصادفات خودرو به دلیل خطای انسانی بوده است، این باور گسترده وجود دارد که خودروهای بدون راننده ای که با هوش مصنوعی کار می کنند، می توانند باعث حذف ۹۰ درصد از مرگ و میرهای جاده ای شوند و این خوش بینی به عنوان نیروی محرکه ای برای پیشرفت به سمت استفاده از این خودروها در جاده ها عمل کرده است. ایلان ماسک پیش بینی می کند که خودروهای بدون رایانه آینده نه تنها ایمن تر خواهند بود، بلکه منجر به کسب درآمد بیشتر برای صاحبانشان خواهند شد. تاکنون، خودروهای بدون راننده سابقه ایمنی بیشتری نسبت به خودروهای معمولی داشته اند و تصادفاتی که تاکنون اتفاق افتاده، بر لزوم توجه به اهمیت اعتبار سنجی فرضیات تاکید دارند. اولین تصادف جزئی خودرو که برای خودروهای بدون راننده اتفاق افتاد، به دلیل یک فرضیه غلط در مورد اتوبوس بود و آن این بود که وقتی خودروی بدون راننده وارد لاین اتوبوس می شود و در مقابل آن قرار می گیرد، راننده اتوبوس به خودروی مورد نظر راه خواهد داد. همچنین، اولین تصادف مرگ باری که در سال ۲۰۱۶ اتفاق افتاد و باعث شد که خودروی بدون راننده تسلا به یدک کش یک کامیون در حال عبور از بزرگراه به شدت بکوبد، به دلیل دو فرضیه غلط بود. اولین فرضیه غلط این بود که خودروی تسلا فکر می کرد که بخش سفید اتاق یدک کش که بسیار روشن بود، بخشی از آسمان روشن می باشد و دومین فرضیه غلط این بود که خودروی تسلا فکر می کرد که راننده کامیون که ظاهرا در حال تماشای فیلم بوده است، به او توجه خواهد کرد و در صورت بروز مشکل، مداخله خواهد کرد. اما گاهی "درست سنجی عملکرد" و "اعتبار سنجی فرضیات" برای اجتناب از تصادفات کافی نیستند و به "کنترل خوب" نیاز است که به معنی نظارت عامل انسانی بر سیستم و تغییر رفتار سیستم در صورت لزوم می باشد. برای بهبود عملکرد سیستم هایی که در آنها، انسان به رایانه برای تصمیم گیری کمک می کند، باید ارتباط موثری بین انسان و ماشین برقرار شود و این یک امر حیاتی است. بر این اساس، نوری که روی داشبورد خودروی شما ظاهر می شود، به راحتی می تواند به شما هشدار دهد که صندوق عقب ماشین خود را باز گذاشته و آن را خوب بسته اید. در عوض، در سال

۱۹۸۷، قایق حمل و نقل انگلیسی هِرالد اف فری اینترپرایز، بندر زبروگه را در حالی ترک کرد که درهای بخش بارگیری و تخلیه آن باز بود. اما، هیچ نور هشداردهنده یا هشدار قابل رویت دیگری برای کاپیتان کشتی صادر نشد و به این ترتیب، کشتی به زودی غرق شد و ۱۹۳ نفر در این حادثه کشته شدند. خطای کنترلی غم انگیز دیگری که جلوگیری از آن از طریق ارتباط بهتر ماشین و انسان امکان پذیر بود، در سال ۲۰۰۹ و هنگامی رخ داد که پرواز ۴۴۷ ایر فرانس به اقیانوس اطلس سقوط کرد و در نتیجه، تمام ۲۲۸ سرنشین آن کشته شدند. طبق گزارش رسمی این حادثه: "خدمه هواپیما هرگز نمی دانستند که افت ارتفاع پیدا کرده اند و در نتیجه هیچ مانور جبرانی ای را که برای هل دادن دماغه هواپیما به سمت پایین اجرا نکردند تا اینکه دیگر همه چیز دیر شده بود." کارشناسان امنیت پرواز گمان می کردند که اگر شاخص زاویه حمله<sup>۴۲</sup> در کابین خلبان وجود داشت که به خلبان نشان دهد که دماغه هواپیما بیش از حد به سمت بالا اشاره می کند، از چنین حادثه ای جلوگیری می شد. وقتی که در سال ۱۹۹۲، پرواز ۱۴۸ ایر اینتر به کوهستان وزگس واقع در نزدیکی استراسبورگ فرانسه سقوط کرد و ۸۷ نفر فوت کردند، معلوم شد که این سقوط به دلیل ارتباط ضعیف بین ماشین و انسان اتفاق نیفتاده است و مشکل از طراحی گیج گننده سیستم رابط با کاربر بوده است. خلبانان عدد ۳۳ را از طریق صفحه کلید وارد کرده بودند، زیرا می خواستند که در زاویه ۳,۳ درجه فرود بیایند. اما، خلبان خود کار هواپیما این عدد را ۳,۳۰۰ فوت در دقیقه تفسیر کرده بود که به حالت متفاوتی مربوط می شد و صفحه نمایش به قدری کوچک بود که امکان نمایش حالت و پی بردن خلبانان به اشتباهشان وجود نداشت.

## هوش مصنوعی برای انرژی

استفاده از فن آوری اطلاعات برای تولید و توزیع انرژی بی نظیر بوده است و الگوریتم های پیچیده فن آوری اطلاعات باعث ایجاد توازنی میان تولید و مصرف در بین شبکه های برق سراسر جهان شده

---

یک اصطلاح هوانوردی است، که زاویه بدنه هواپیما و بال های آن را با مسیر حرکت کنونی آن توصیف می کند<sup>۴۲</sup>

اند. همچنین، سیستم های کنترلی پیچیده فن آوری اطلاعات منجر به اطمینان از عملکرد ایمن و کارای نیروگاه ها شده اند. پیوند میان فناوری های دیجیتال و شبکه برق، منجر به مفهوم جدیدی به نام شبکه هوشمند (Smart Grid) شده است. پیشرفت های آتی در زمینه هوش مصنوعی باعث افزایش هوشمندی شبکه های هوشمند انرژی الکتریکی خواهند شد و به طرز بهینه ای، منجر به سازگاری شبکه هوشمند با تغییرات ایجاد شده در عرضه و تقاضا حتی تا سطوح جزئی پیل های خورشیدی موجود در پشت بام خانه ها و سیستم های ذخیره باطری برای بهره برداری از انرژی خانگی خواهد شد. با این حال، در سال ۲۰۰۳ حدود ۵۵ میلیون نفر در آمریکا و کانادا با خاموشی برق مواجه شدند و بسیاری از آنها به مدت چند روز عاجز ماندند. در اینجا هم عامل اصلی به ارتباطات ضعیف بین انسان و ماشین مربوط می شد. یک خطای نرم افزاری مانع از این شده بود که سیستم هشدار دهنده در اتاق کنترل به عوامل انسانی هشدار دهد که قبل از اینکه یک مشکل کوچک (برخورد خطوط انتقال برق فشار قوی با شاخ و برگ هرس نشده درختان) باعث شود که کل سیستم از کنترل خارج شود، اقدامات لازم را انجام دهند و انرژی برق را دوباره توزیع کنند. فروگذاخت انرژی هسته ای در یک رآکتور واقع در پنسیلوانیا منجر به صرف میلیاردها دلار به عنوان هزینه تمیز کردن کامل رآکتور و واکنش شدید در برابر انرژی هسته ای شد. یکی از عوامل ایجاد کننده فروگذاخت هسته ای، ارتباط ضعیف در سیستم رابط انسانی بود. نور هشدار دهنده ای که باعث شده بود عوامل انسانی فکر کنند که این نور برای اطمینان از باز یا بسته بودن فلکه اصلی حساس به ایمنی صادر شده بود، باعث شد که فلکه را ببندند و متوجه نشوند که فلکه اصلی با گرفتاری زیادی باز شده بود. این حوادث مربوط به انرژی و حمل و نقل به ما یاد می دهند که وقتی مسئولیت سیستم های فیزیکی را به هوش مصنوعی واگذار می کنیم، تحقیقات جدی ای را انجام دهیم تا علاوه بر اطمینان از عملکرد فردی مناسب این ماشین ها، از همکاری موثر ماشین ها با کنترل کننده های انسانی نیز مطمئن شویم. هرچه قدر که هوش مصنوعی هوشمندتر می شود، نیاز بیشتری به بهبود سیستم رابط انسانی برای به اشتراک گذاشتن اطلاعات و تخصیص بهینه وظایف بین تیم های انسانی و رایانه ای احساس می شود و برای دستیابی به

این هدف لازم است که موقعیت هایی را که نیاز به کنترل دارند، شناسایی کنیم و به جای منحرف کردن کنترل کننده های انسانی با انبوهی از اطلاعات بیهوده، از قضاوت های موثر انسانی برای تصمیم گیری های با ارزش استفاده کنیم.

## هوش مصنوعی برای مراقبت های بهداشتی

هوش مصنوعی از پتانسیل بالایی برای بهبود مراقبت های بهداشتی برخوردار است. دیجیتال سازی سوابق پزشکی باعث شده که پزشکان و بیماران تصمیمات سریعتر و بهتری بگیرند و برای تشخیص تصاویر دیجیتالی از متخصصان موجود در سراسر جهان، کمک فوری دریافت کنند. در واقع، با توجه به پیشرفت سریع صورت گرفته در زمینه تصویربرداری رایانه ای و یادگیری عمیق، سیستم های هوش مصنوعی به زودی تبدیل به بهترین متخصصان برای تشخیص بیماری ها خواهند شد. به عنوان مثال، در سال ۲۰۱۵ یک محقق هلندی نشان داد که تشخیص رایانه ای بیماری سرطان پروستات با استفاده از تصویربرداری پرتو مغناطیسی (ام آر آی) به خوبی پرتوشناسان انسانی می باشد. در سال ۲۰۱۶ یک محقق انگلیسی نشان داد که هوش مصنوعی بهتر از آسیب شناسان انسانی می تواند بیماری سرطان ریه را از طریق تصاویر میکروسکوپی تشخیص دهد. اگر "یادگیری ماشین" بتواند روابط میان ژن ها، بیماری ها و واکنش های درمانی را آشکار کند، تحول گسترده ای در پزشکی شخصی، سلامت حیوانات اهلی و بهبود محصولات کشاورزی حاصل خواهد شد. علاوه بر این، ربات ها از پتانسیل لازم برای تبدیل شدن به جراحانی برخوردار هستند که دقیقتر و مطمئن تر از جراحان انسان می باشند. در سال های اخیر، طیف گسترده ای از جراحی های رباتیک با موفقیت اجرا شده است و این جراحی ها به دلیل، داشتن دقت بالا و برش های کمتر، منجر به خونریزی کمتر، درد کمتر و زمان بهبودی کمتری شده اند. افسوس که در صنعت مراقبت های بهداشتی هم می توان درس های دردناکی را در مورد خطاهای نرم افزاری گرفت. به عنوان مثال، ماشین پرتو درمانی کانادایی تراک<sup>۴۳</sup> ۲۵ که برای درمان بیماران سرطان طراحی شده بود، شامل دو حالت مختلف می شد. حالت اول،

استفاده از پرتو الکترون های برق ضعیف و حالت دوم، استفاده از پرتو الکترون های برق قوی چند مگاولتی اشعه ایکس بوده است که با یک محافظ ویژه روی نقطه هدف قرار می گرفت. متأسفانه، خطای نرم افزاری از نوع "درست سنجی عملکرد" باعث شده بود که تکنیسین ها فکر کنند که از پرتو الکترون های برق ضعیف استفاده می کنند، در حالیکه پرتو چند مگاولتی اشعه ایکس را بدون محافظ روی نقطه هدف قرار داده بودند که این امر منجر به مرگ تعداد زیادی از بیماران شد. در موسسه ای در پاناما نیز تعداد زیادی از بیماران به دلیل مسمومیت پرتوی ناشی از قرار گرفتن بیش از حد در معرض پرتو رادیواکتیو کبالت ۶۰ جان خود را از دست دادند که دلیل آن هم اعتبار سنجی نادرست یک فرضیه در سیستم رابط انسانی بود. طبق یکی از گزارش های اخیر، در کشور آمریکا و در بین سال های ۲۰۰۰ و ۲۰۱۳، حوادث ناشی از جراحی رباتیک با مرگ ۱۴۴ نفر و زخمی شدن ۱۳۹۱ نفر در ارتباط بوده است و مواردی از قبیل مشکلات سخت افزاری مانند حادثه جرقه الکتریک و افتادن ابزارهای سوخته یا شکسته شده به بدن بیمار و مشکلات نرم افزاری مانند حرکات کنترل نشده ماشین و خاموشی خود به خودی در این کشور گزارش شده است. با این حال، خبر خوب این است که بقیه دو میلیون جراحی رباتیک گزارش شده در این کشور به خوبی پیش رفته است و به نظر می رسد که جراحی های ربات ها ایمن تر بوده اند. از سوی دیگر گزارش می شود که مراقبت های ناکافی از بیماران منجر به مرگ ۱۰۰۰۰۰ نفر در آمریکا شده است. بنابراین، لزوم توجه به مسائل اخلاقی برای توسعه هوش مصنوعی در بیمارستان ها بیش از خودروهای بدون راننده می باشد.

## هوش مصنوعی برای ارتباطات

رایانه ها بیشترین تاثیر را بر صنعت ارتباطات داشته اند. پس از معرفی سویچ تلفن در دهه ۵۰، اینترنت در دهه ۶۰ و وب جهان گستر در سال ۱۹۸۹، هم اکنون، میلیارها انسان از اینترنت برای برقراری ارتباط، خرید، خواندن اخبار، تماشای فیلم استفاده می کنند. ظهور اینترنت برای انجام کارها نوید افزایش کارایی، دقت، راحتی و مزایای اقتصادی را به ما می دهد. این موفقیت چشمگیر، منجر به چالش چهارمی برای دانشمندان علم رایانه شده است. آنها علاوه بر درستی سنجی عملکرد، اعتبار

سنجی فرضیه و کنترل باید به مساله افزایش امنیت در برابر نرم افزار بدخواه ( بدافزار) و رخنه کردن در اطلاعات توجه کنند. در حالیکه مشکلات قبلی به دلیل اشتباهات غیر عمدی بوده است، مساله امنیت به شرارت عمدی اشاره می کند. اولین بدافزاری که باعث جلب توجه رسانه ای قابل توجهی شد، بدافزار موسوم به "کرم موریس" بود که در سال ۱۹۸۸ پخش شد. این بدافزار از خطاهای موجود در سیستم عامل یونیکس سوء استفاده می کرد. با وجود اینکه این بدافزار حدود ۱۰ درصد از ۶۰۰۰۰ رایانه متصل به اینترنت در آن زمان را ویروسی کرده و از کار انداخته بود، این امر مانع از این نشد که رابرت موریس، سازنده آن در نهایت بتواند مدرک دکترای علوم رایانه در دانشگاه ام آی تی را دریافت کند. سایر بدافزارها به جای سوء استفاده از خطاهای نرم افزاری از نقاط ضعف انسانی استفاده می کردند. افراد حاضر در یک جشن، ایمیل هایی را با عنوان "من عاشق شما هستم" از آشنایان و همکاران خود دریافت می کردند. اما کاربرانی که ناآگاهانه روی لینک عاشقانه دیگری کلیک کرده بودند، دستوری را اجرا کرده بودند که باعث آسیب رسانی به رایانه آنها می شد. این کرم که توسط دو برنامه نویس جوان در فلیپین ایجاد شده بود، مانند کرم موریس، حدود ۱۰ درصد از اینترنت را ویروسی می کرد و از آنجایی که اینترنت تا آن زمان بسیار بزرگتر شده بود، به عنوان بزرگترین ویروس شناخته شد و ۵۰ میلیون رایانه را مبتلا کرد و ۵ میلیارد دلار خسارت به بار آورد. همانطور که می دانید، اینترنت با انواع بی شماری از بدافزارهای ویروسی آلوده می شود که کارشناسان از آنها با عنوان کرم یاد می کنند و آسیب ناشی از آنها از نمایش پیام های مسخره بی ضرر گرفته تا حذف فایل هایتان، دزدی اطلاعات شخصیتان، جاسوسی و رخنه به رایانه تان برای ارسال پیام های به دردخور را شامل می شوند. همچنین به نظر می رسد که میزان غارت ها در سال های اخیر، افزایش چشم گیری داشته است. در سال ۲۰۰۸، هکرها توانستند ۱۳۰ میلیون شماره کارت اعتباری و سایر اطلاعات مربوط به حساب کارت های اعتباری را از سیستم پرداخت هارت لند بدزدند. در سال ۲۰۱۴ و به دنبال هک شدن اداره دولتی مدیریت کارکنان در آمریکا، سوابق کارکنان و اطلاعات مربوط به درخواست شغلی برای بیش از ۲۱ میلیون نفر مورد رخنه واقع شد. در



نتیجه، هر وقت که در مورد سیستم های جدیدی که ظاهراً صد در صد ایمن و غیر قابل هک کردن هستند، چیزهایی را می خوانم، متعجب می شوم. با این حال، ما از سیستم های هوش مصنوعی آینده انتظار داریم که قبل از به کار گرفته شدن در سیستم های زیرساختاری حساس یا سیستم های تسلیحاتی، به ما اطمینان دهند که غیر قابل هک کردن هستند. بنابراین، افزایش نقش هوش مصنوعی در جامعه باعث افزایش خطرات امنیتی برای رایانه ها خواهد شد. برخی از هک ها ناشی از ساده لوحی انسان ها یا آسیب پذیری نرم افزارهای تازه انتشار یافته است، اما بقیه آنها شامل ورود غیر قانونی به رایانه های واقع در مکان های دوردست با استفاده از خطاهای نرم افزاری ساده ای است که تا مدت ها کسی متوجه آنها نمی شود. متأسفانه، سیستم های پیشرفته هوش مصنوعی می توانند برای یافتن نقاط ضعف جدید و اجرای اقدامات هکری پیچیده نیز مورد استفاده قرار بگیرند. به عنوان مثال، فرض کنید که یک روز یک ایمیل با هدف کلاه برداری از خود دریافت می کنید که تلاش می کند شما را متقاعد کند که اطلاعات شخصی خود را فاش کنید. این ایمیل از طرف دوستان فرستاده شده که با استفاده از هوش مصنوعی می خواهد رایانه شما را هک کند. این ماشین هوش مصنوعی بر اساس تحلیل خود از سایر ایمیل های دوستان، می تواند سبک نوشتن او را تقلید کند و اطلاعات شخصی زیادی را در مورد شما از منابع مختلف دریافت کند. آیا ممکن است که فریب خورده باشید؟ اگر این ایمیل کلاه بردانه از طرف شرکت کارت اعتباریتان به شما فرستاده شده باشد که در آن، ماشین هوش مصنوعی در تماس تلفنی با شما صدای انسان را تقلید کرده باشد و شما قادر به تشخیص اینکه با ماشین هوش مصنوعی در حال صحبت کردن هستید، نباشید چه می کنید؟

## قوانین

ما انسان ها حیوانات اجتماعی هستیم که نسبت به گونه های دیگر آرام تر هستیم و به لطف قدرت همکاری با یکدیگر، زمین را تسخیر کرده ایم. ما قوانینی را برای تشویق به همکاری و تسهیل بخشیدن به آن توسعه بخشیده ایم و به همین دلیل، اگر هوش مصنوعی بتواند سیستم های قانونی و حکومتی ما را بهبود دهد، منجر به افزایش همکاری های موفقیت آمیزتر از گذشته و رشد و

شکوفایی استعدادهای ما خواهد شد. فرصت های زیادی برای بهبود در اعمال قوانین و تنظیم آنها وجود دارد که به بررسی آنها خواهیم پرداخت. وقتی که شما در مورد سیستم قضایی کشورتان فکر می کنید، چه چیزی برای اولین بار در ذهنتان تداعی پیدا می کند؟ اگر به تاخیرهای طولانی مدت، هزینه های بالا و بی عدالتی فکر می کنید، بدانید که شما تنها نیستید. اما، اگر اولین چیز تداعی شده به ذهنتان کارایی و انصاف باشد، باید تعجب کرد. از آنجایی که فرایند حقوقی را می توان به عنوان یک محاسبه نگاه کرد که اطلاعات ورودی مربوط به مدارک و قوانین را دریافت می کند و یک تصمیم را به عنوان خروجی ارائه می کند، برخی از دانشمندان رویای خود کارسازی کامل آن با ربات قاضی را در سر می پروراندند. ربات های قاضی، سیستم های هوش مصنوعی خستگی ناپذیری هستند که استانداردهای قانونی بالایی را برای هر قضاوت اعمال می کنند و تسلیم خطاهای انسانی از قبیل خستگی یا نداشتن دانش امروزی نمی شوند.

## ربات های قاضی

در سال ۱۹۹۴ بایرون بکویت متهم به قتل مدگار اورز، رهبر حقوق مدنی آمریکا شد. اما دو قاضی جداگانه نتوانستند بعد از گذشت یک سال از این قتل، او را متهم کنند. هرچند که مدارک فیزیکی هر دوی آنها در اصل یکسان بود. متأسفانه، تاریخ حقوق سرشار از قضاوت هایی است که به واسطه رنگ پوست، جنسیت، مذهب، ملیت و عوامل دیگر منحرف می شود. ربات های قاضی برای اولین بار در تاریخ، تضمین خواهند کرد که همه در برابر قانون یکسان هستند و آنها را می توان طوری برنامه نویسی کرد که همه مشابه هم باشند و با همه یکسان برخورد کنند و در اعمال قوانین تعصب به خرج ندهند. ربات های قاضی همچنین می توانند خطاهای انسانی ای غیر عمدی و تصادفی را حذف کنند. به عنوان مثال، در مطالعه انجام شده در سال ۲۰۱۲ در مورد قضاوت یکی از کشورها ثابت شد که وقتی آنها گرسنه هستند، حکم های سخت گیرانه تری را صادر می کنند. این قضاوت در حال گرسنگی ۸۵ در صد از پرونده های عفو مجرمین را رد کرده بودند. نقص دیگر در قضاوت های بشری این است که قضاوت ممکن است وقت کافی برای کشف کلیه جزئیات یک پرونده را نداشته باشند. در عوض،

ربات های قاضی را می توان کپی کرد، زیرا آنها فقط یک نرم افزار هستند و می توانند کلیه پرونده های بلا تکلیف را به صورت موازی با یکدیگر پردازش کنند و هر ربات قاضی می تواند تا هر زمانی که طول می کشد، می تواند روی یک پرونده کار کند. همچنین، با توجه به اینکه قاضی انسان نمی تواند بر کل دانش فنی مورد نیاز برای رسیدگی به یک پرونده تسلط داشته باشد، ربات های قاضی آینده می توانند از حافظه و ظرفیت یادگیری نامحدود خود برای قضاوت استفاده کنند. روزی فرا خواهد رسید که ربات های قاضی در سایه تعصب ناورزی، کارآمدی و شفافیت، هم کاراتر و هم عادل تر خواهند شد و کارایی آنها نیز باعث عادل تر شدنشان می شود. ربات های قاضی با تسریع بخشیدن به فرایند حقوقی و جلوگیری از انحراف در نتیجه گیری قضات، باعث می شوند که گرفتن حق خود در دادگاه ها با هزینه به شدت پایینی امکان پذیر شود. این امر باعث افزایش احتمال چیره شدن افراد بی پول و شرکت های نوپا بر شرکت های میلیاردی یا چندملیتی خواهد شد. از طرف دیگر، اگر ربات های قاضی خطا پیدا کنند یا هک شوند، چه اتفاقی خواهد افتاد؟ هر دوی این ها مصیبتی برای ماشین های رای دهنده خود کار محسوب خواهند شد و انگیزه حملات سایبری با ماندن چندین ساله مجرمان در زندان و تحمل ضررهای مالی میلیونی افزایش خواهد یافت. حتی اگر هوش مصنوعی به قدری مقاوم شود که بتوان به ربات های قاضی اعتماد کرد که از الگوریتم های قانونی استفاده می کنند، آیا می توان برای قضاوت هایشان ارزش قایل شد؟ این چالش به واسطه موفقیت های اخیر در زمینه شبکه های عصبی که بهتر از الگوریتم های ساده و سنتی هوش مصنوعی عمل می کنند، بدتر شده است. اگر متهمان پیرسند که چرا متهم شده اند، آیا می توان به آنها گفت که ما سیستم را با انبوهی از داده ها آموزش داده ایم و این تصمیمی است که سیستم گرفته است؟ البته، مطالعات اخیر نشان می دهد که سیستم یادگیری شبکه های عصبی عمیق که حجم زیادی از اطلاعات را در مورد زندانی ها دریافت می کند، بهتر از قاضی های انسان پیش بینی می کند که چه کسی را به دلیل احتمال ارتکاب دوباره جرم نباید مورد عفو قرار داد. ما باید در مورد مزایای هوش مصنوعی بیشتر فکر و بحث کنیم. ما با یک تصمیم "همه یا هیچ" در رابطه با ربات قاضی مواجه نیستیم، بلکه با

تصمیمی در رابطه با دامنه و سرعت استفاده از هوش مصنوعی در سیستم حقوقی خود سرو کار داریم. آیا از قاضی های خود می خواهیم که مجهز به سیستم های پشتیبانی از تصمیم گیری مبتنی بر هوش مصنوعی باشند؟ آیا از این هم فراتر رفته و از ربات های قاضی می خواهیم که تصمیماتی بگیرند که مورد پسند قضات انسان باشند یا اینکه همه چیز را در نهایت به ماشین ها واگذار خواهیم کرد و حتی با مجازات های قتل صادر شده از ماشین ها موافقت خواهیم کرد.

## اختلاف نظرهای حقوقی

تاکنون فقط در مورد اعمال قانون صحبت کردیم. اکنون به محتوای یک قانون خواهیم پرداخت. توافق گسترده ای در این مورد که قوانین ما برای همگام شدن با فن آوری باید متحول شوند، حاصل شده است به عنوان مثال، دو برنامه نویسی که کرم "آی لاو یو" را ایجاد کرده و میلیاردها دلار صدمه وارد کرده بودند، از هرگونه اتهامی تبرئه شدند، زیرا در آن زمان، هیچ قانونی برای ایجاد "بدافزار" در فیلپین وجود نداشت. با توجه به شتاب پیشرفت های فن آوری، لازم است که قوانین را با سرعت هرچه بیشتری به روز رسانی کرد. استفاده از افراد آشنا با فن آوری در دانشکده های حقوق و شغل های دولتی حرکت هوشمندانه ای برای جامعه محسوب می شود. در اینجا این سوال مطرح می شود که آیا سیستم های پشتیبانی از تصمیم گیری مبتنی بر هوش مصنوعی برای قانون گذاران باید پشت سر ربات های قاضی بی پرده حرکت کنند یا جلوتر از آنها قدم بردارند؟ اینکه چگونه می توان قوانین را تغییر داد تا پیشرفت های هوش مصنوعی را منعکس کنند، موضوع بحث برانگیز جالبی است. یکی از این بحث ها در مورد تعارض موجود بین محرمانگی و تبادل آزادانه اطلاعات است. طرفداران تبادل آزادانه اطلاعات استدلال می کنند که هرچه محرمانگی کمتر شود، دادگاه مدارک بیشتری خواهد داشت و قضاوت ها عادلانه تر خواهند شد. به عنوان مثال، اگر دولت ها به وسایل الکترونیکی همه افراد دسترسی پیدا کنند و کارهای آنها را ثبت کنند، بسیاری از جرم ها به سرعت فیصله پیدا می کنند و از وقوع جرم های بعدی جلوگیری خواهد شد. طرفداران محرمانگی با نظارت مستبدانه از سوی دولت ها مخالف هستند و به اعتقاد آنها اگر چنین کاری صورت گیرد، با خطر

حاکمیت دیکتاتوری خودکامه در حجم وسیع مواجه خواهیم شد. علاوه بر این، تکنیک های یادگیری ماشین در زمینه تجزیه و تحلیل داده های مغز از طریق اسکنرهای اف ام آر آی پیشرفت کرده اند و به کمک آنها می توان افکار فرد و اینکه فرد راست یا دروغ می گوید را تشخیص داد. اگر استفاده از فن آوری اسکن مغز با کمک هوش مصنوعی در محیط دادگاه رواج پیدا کند، فرایند فعلی خسته کننده سازماندهی داده های یک پرونده به شدت تسهیل می شود و سرعت رسیدگی به آنها افزایش می یابد و قضاوت ها عادلانه تر خواهند شد. طرفداران محرمانگی نگران این هستند که این سیستم ها اشتباه کنند یا اینکه مراقبت ها و تجسس های دولت در امور خصوصی افراد از حد خارج شود. دولت هایی که از آزادی افکار حمایت نمی کنند، می توانند از این فن آوری برای مجرم شناختن صاحبان باورها و عقاید خاص استفاده کنند. این سوال ایجاد می شود که مرز میان عدالت و حریم خصوصی و مرز میان حفظ جامعه و حفظ آزادی اشخاص کجاست؟ در هر حال، کاهش محرمانگی برای جلوگیری افراد از ارائه مدارک جعلی ضروری است. به عنوان مثال، اگر هوش مصنوعی بتواند ویدئوهای جعلی در مورد اینکه شما مرتکب جرمی شده اید را تولید کند، آیا از دولت نمی خواهید که در صورت نیاز موقعیت زمانی و مکانی همه را دنبال کند و غیبت شما در زمان وقوع جرم را بررسی کند؟ بحث دیگر این است که سیاستگذاران برای حداکثر ساختن احتمال دریافت خروجی مفید، باید چه مشوق هایی را برای محققان هوش مصنوعی در نظر بگیرند؟ برخی از محققان هوش مصنوعی استدلال می کنند که با کلیه اشکال ساماندهی به توسعه های اخیر در زمینه هوش مصنوعی مخالف هستند و بدون دلیل خاصی ادعا می کنند که توسعه نوآوری های مورد نیاز مانند خودروهای بدون راننده را باید به تعویق انداخت و تحقیقات مربوط به جدیدترین پیشرفت های هوش مصنوعی را باید به صورت مخفیانه انجام داد. در کنفرانس مربوط به مزایای هوش مصنوعی در پورتوریکو که در فصل اول عنوان شد، ایلان ماسک استدلال می کند که آنچه که ما از دولت ها می خواهیم، غفلت نیست، بلکه آگاهی است. خصوصاً اینکه ما از کارشناسان فنی فعال در پست های دولتی می خواهیم که از پیشرفت های هوش مصنوعی آگاه باشند و برای بهره برداری از آنها در

آینده، هدایتشان کنند. او همچنین استدلال می کند که قوانین دولتی باید به پرورش هوش مصنوعی کمک کنند، نه اینکه آنها را سرکوب کنند. به عنوان مثال، اگر استانداردهای ایمنی دولت برای خودروهای بدون راننده به کاهش تعداد تصادفات کمک کند، احتمال نشان دادن واکنش های منفی از سوی مردم کمتر خواهد شد و پذیرش فن آوری جدید، با سرعت بیشتری صورت خواهد گرفت. بنابراین، شرکت های هوش مصنوعی حساس به ایمنی، از قوانینی که رقبای بی احتیاط را مجبور به رعایت استانداردهای ایمنی بالا می کنند، طرفداری خواهند کرد. مساله حقوقی دیگر به موضوع امتیاز قائل شدن برای ماشین ها مربوط می شود. اگر یک خودروی بدون راننده دچار سانحه شود، چه کسی مسئول خواهد بود. سرنشینان، مالکان یا سازندگان آن؟ حقوقدانی به نام دیوید ویلادک، خود خودرو را مقصر می داند که بیمه نشده است. مدل های خودروی پیروی کننده از مقررات ایمنی استرلینگ واجد شرایط لازم برای پرداخت حق بیمه ناچیزی هستند. ولی حق بیمه مربوط به مدل های خودروی دارای طراحی و ساخت ضعیف به قدری گران است که کسی مالک آنها نمی شود. اما این سوال مطرح می شود که آیا ماشین ها علاوه بر حق بیمه می توانند مالک پول و دارایی خود باشند؟ در این صورت هیچ چیزی مانع از این نمی شود که در بازار سهام سرمایه گذاری کنند. وقتی که یک رایانه به انسان پول می پردازد که برای آن کار کند، تمام کارهای انسان را می تواند انجام دهد. اگر سرمایه گذاری ماشین ها از انسان ها بیشتر شود، در چنین اوضاعی بیشتر اقتصاد جهان تحت تملک و کنترل آنها قرار خواهد گرفت. وقتی که یک موجودیت غیر انسانی به نام "شرکت" که از افراد موجود در آن قوی تر است، می تواند مالک اقتصاد جهان شود، آیا ماشین ها هم می توانند به این هدف دست پیدا کنند؟ آیا ماشین ها هم مانند انسان ها دارای تجربه حسی هستند؟ چنین سوالات بحث برانگیزی را با جزئیات بیشتر در فصل بعد و فصل هشتم مورد بررسی قرار خواهیم داد.

## سلاح ها

بشریت از زمانهای بسیار قدیم، از قطعی، بیماری و جنگ رنج می برده است. همانطور که گفتیم، هوش مصنوعی می تواند به کاهش قحطی و بیماری کمک کند، اما در مورد جنگ چطور؟ برخی ها

استدلال می کنند که سلاح های هسته ای مانع از جنگ میان کشورهایی می شود که از آن برخوردار هستند، زیرا آنها بسیار مرعوب کننده هستند. اگر همه کشورها اجازه داشتند که سلاح های مبتنی بر هوش مصنوعی را با امید پایان دادن همیشگی به جنگ بسازند، به نظر شما چه اتفاقی می افتاد؟ اگر با این بحث ها موافق نیستید و معتقدید که جنگ های آینده میان کشورها اجتناب ناپذیر است، نظرتان در مورد استفاده از هوش مصنوعی برای انسان دوستانه تر کردن این جنگ ها چیست؟ اگر جنگ بین ماشین ها اتفاق بیفتد، هیچ سرباز یا فرد غیر نظامی کشته نخواهد شد. پهبادهای آینده مبتنی بر هوش مصنوعی و سایر سیستم های تسلیحاتی خودران<sup>۴۴</sup> (که رقبا به آنها لقب "ربات های قاتل" را داده اند) را می توان منصف تر و منطقی تر از سربازان انسان ساخت. آنها به حسگرهای فوق بشری مجهز خواهند بود و از کشته شدن نمی ترسند و حتی در گرماگرم نبرد ممکن است خونسرد، حسابگر و آرام بمانند و از کشتن اتفاقی افراد غیر نظامی اجتناب کنند.

### کمک انسان به رایانه برای تصمیم گیری درست

اگر سیستم های خودکار، خطادار و گیج کننده باشند و چنانکه که از آنها انتظار می رود، رفتار نکنند، چه اتفاقی خواهد افتاد؟ سیستم پدافند هوایی فالانکس آمریکا که هدفش حفاظت از کشتی های جنگی است، به طور خودکار، وظیفه دارد که تهدیدهای ناشی از موشک ها و هواپیماهای ضد کشتی را شناسایی و مسیریابی کند و به آنها حمله کند. در سال ۱۹۸۸، که اواسط دوران جنگ میان ایران و عراق بود، رادار ناو جنگی آمریکایی وینسنز در مورد ورود یک هواپیما هشدار داد و کاپیتان این ناو حدس زد که ناو جنگی آنها توسط جت جنگی اف ۱۴ فایتر ایرانی مورد حمله واقع شده است و به سیستم حفاظت از کشتی های جنگی اجازه شلیک داد. آنچه که این کاپیتان متوجه نشده بود، این بود که آنها به یک جت مسافربری غیر نظامی ایرانی به نام ایر فلایت ۶۵۵ شلیک کرده بودند که منجر به کشته شدن ۲۹۰ سرنشین آن شد و خشم جهانی را برانگیخت. بررسی های بعدی حاکی از ارتباط ضعیف سیستم با رابط انسانی بود که به طور خودکار نشان نمی داد چه نقطه هایی روی صفحه

---

<sup>۴۴</sup> Autonomous weapons system

رادار، به هواپیمای غیر نظامی اشاره می کنند یا اینکه نشان نمی داد چه نقطه هایی نزولی (برای حمله کردن) و چه نقطه هایی صعودی (حاکمی از مسیر جت غیر نظامی ایرانی ایر فلایت ۶۵۵) هستند. در عوض، وقتی که از سیستم خود کار برای دریافت اطلاعاتی در مورد هواپیمای مرموز پرس و جو شد، سیستم به حرکت صعودی اشاره کرده بود که مربوط به یک هواپیمای آمریکایی دیگر بود که برای رهگیری و انهدام هواگردهای متخاصم در مناطق دوردست خلیج فارس فعالیت می کرد. در این مثال، تصمیم گیری نهایی توسط انسانی گرفته شده بود که تحت فشار زمانی موجود، به آنچه که سیستم خود کار به او گفته بود، اعتماد بیش از حد کرده بود. طبق گفته مقامات دفاعی فعال در سراسر جهان، به جز انواع تله های انفجاری از قبیل مین های زمینی، فرایند تصمیم گیری در کلیه سیستم های تسلیحاتی توسط یک انسان انجام می شود. با این حال، برنامه ای برای توسعه سلاح های کشنده خودران در حال اجراست که بدون دخالت انسان، می تواند اهداف را انتخاب و مورد حمله قرار دهد. مقامات نظامی برای افزایش سرعت فعالیت های خود، ممکن است وسوسه شوند که تصمیم گیری انسانی را به طور کامل کنار بگذارند. به نظر شما، بین پهباد کاملاً خود کاری که واکنش فوری نشان می دهد و پهبادی که تنبل تر است و تا نیمه راه توسط انسان کنترل می شود، کدامیک برنده خواهند شد؟ با این حال، مواردی هم وجود دارند که نشان می دهند ما کاملاً خوش شانس بوده ایم که تصمیم گیری توسط انسان صورت گرفته است. در سال ۱۹۶۲ و در حین بحران موشکی کوبا، یازده ناوشکن آمریکایی و ناو هواپیمابر رندولف، زیردریایی بی ۵۹ متعلق به شوروی سابق را در نزدیکی کوبا و در آب های بین المللی، محاصره کردند. آنچه که شوروی ها نمی دانستند این بود که به دلیل تمام شدن باطری های زیردریایی و خاموش شدن دستگاه تهویه هوا، دمای داخل زیردریایی از ۴۵ درجه سانتی گراد نیز فراتر رفته بود. بسیاری از اعضای خدمه که در آستانه مسموم شدن با گاز دی اکسید کربن قرار داشتند، از حال رفته بودند. خدمه زیردریایی چندین روز بود که ارتباطی با مسکو نداشتند. سپس، آمریکایی ها شروع به انداختن بمب های آبی کوچکی کردند که هدفشان اجبار زیردریایی شوروی برای آمدن به سطح دریا و ترک آن محل بود. با این حال، خدمه زیردریایی



اطلاعی از هدف آمریکایی نداشتند. آنچه که آمریکایی ها نمی دانستند نیز این بود که خدمه زیردریایی بی ۵۹ یک اژدر هسته ای دارند که بدون اجازه گرفتن از مسکو می توانند آن را به بیرون پرتاب کنند. کاپیتان زیردریایی تصمیم گرفت که اژدر هسته ای را به بیرون پرتاب کند. افسر نیروی دریایی شوروی نیز فریاد زد: "ما خواهیم مرد. اما همه دشمنان را غرق کرده و آبروی نیروی دریایی مان را حفظ خواهیم کرد." خوشبختانه تصمیم گیری برای پرتاب اژدر باید مورد تأیید سه افسر قرار می گرفت که افسر سوم مخالف آن بود. نکته تامل برانگیز این است که کسی این افسر سوم را نمی شناخت. اما، تصمیم او جلوی جنگ جهانی سوم را گرفت و کمک بزرگی به بشریت کرد. نکته تامل برانگیز دیگر این است که اگر زیردریایی بی ۵۹ تحت کنترل هوش مصنوعی بود و هیچ انسانی برای کمک به تصمیم گیری ماشین وجود نداشت، معلوم نبود که چه اتفاقی می افتاد. دو دهه بعد و در سپتامبر ۱۹۸۳ تنش های بین دو ابر قدرت شدت گرفت. رونالد ریگان، رئیس جمهور آمریکا برای نامیدن اتحاد جماهیر شوروی از اصطلاح "امپراطوری شیطان" استفاده کرد. تا اینکه یک روز، سیستم هشدار زودهنگام خود کار در شوروی، گزارش داد که ایالات متحده، پنج موشک هسته ای زمینی را به شوروی پرتاب کرده است. استانیسلاف پتروف، افسر شوروی برای تصمیم گیری در مورد اینکه آیا این یک هشدار غلط است، فقط چند دقیقه فرصت داشت. او که می دید ماهواره ها درست کار می کنند، به جای اینکه یک حمله هسته ای را گزارش کند، به ندای درونی اش اعتماد کرد و با خود فکر کرد که احتمال اینکه ایالات متحده فقط با پنج موشک هسته ای به شوروی حمله کند، بسیار کم است و بدون اینکه از حقیقت گفته خود مطمئن باشد، به فرماندهان خود گفت که این یک هشدار غلط است. بعداً معلوم شد که یک ماهواره، انعکاس های نور خورشید از بالای ابرها را با شعله های ناشی از موتورهای موشکی اشتباه گرفته است. اگر تصمیم گیری بدون دخالت انسان و بر اساس یک سیستم هوش مصنوعی صورت می گرفت که فقط به دستورالعمل ها عمل می کرد، معلوم نبود که چه اتفاقی می افتاد.

## رقابت تسلیحاتی<sup>۴۵</sup> بعدی؟

همانطور که حدس زده اید، نگرانی من در مورد سیستم های تسلیحاتی خودران، جدی است. اما نگرانی اصلی من در مورد سرانجام یک رقابت تسلیحاتی بر سر سلاح های مبتنی بر هوش مصنوعی می باشد. من در سال ۲۰۱۵ نگرانی خود را در نامه سرگشاده زیر بیان کرده ام و بازخوردهای مفیدی را از سوی همکاران خود در موسسه آینده زندگی<sup>۴۶</sup> دریافت کرده ام.

### سلاح های خودران:

#### نامه سرگشاده ای از سوی محققان هوش مصنوعی و رباتیک

سلاح های خودران، اهداف خود را بدون دخالت انسان، انتخاب و دنبال می کنند. به عنوان مثال، ممکن است که آنها شامل ریز پرنده های مسلحی باشند که می توانند بر اساس معیارهای از پیش تعیین شده خود، افرادی را جستجو و نابود کنند، اما این شامل موشک های کروز یا پهبادهای کنترل شده از راه دوری نمی باشند که در آنها تمام تصمیمات توسط انسان ها گرفته می شوند. فن آوری هوش مصنوعی به مرحله ای از پیشرفت رسیده است که به کارگیری سیستم های هوش مصنوعی در ظرف چند سال و نه چند دهه امکان پذیر است که البته نیاز به ریسک بالایی دارد. سلاح های خودران پس از باروت و تسلیحات هسته ای، انقلاب سومی را در زمینه تولید جنگ افزارها ایجاد کرده اند. استدلال های موافق و مخالف فراوانی برای استفاده از سلاح های خودران بیان شده است. به عنوان مثال، جایگزینی سربازان انسان با ماشین ها به دلیل کاهش تلفات انسانی، کار خوب و به دلیل کاهش آستانه ورود به جنگ، کار بدی خواهد بود. سوال کلیدی برای بشر امروز این است که آیا باید از رقابت تسلیحاتی جهانی مبتنی بر هوش مصنوعی حمایت کرد یا اینکه باید از شروع آن جلوگیری کرد؟ اگر قدرت نظامی بزرگی اقدام به توسعه سلاح های مبتنی بر هوش مصنوعی کند، وقوع رقابت تسلیحاتی جهانی تقریباً اجتناب ناپذیر خواهد بود و عاقبت این مسیر توسعه فن آوران را

---

<sup>۴۵</sup> Arms race

<sup>۴۶</sup> Future of life institute

به وضوح می توان پیش بینی کرد. برای توسعه سلاح های خودران، بر خلاف سلاح های هسته ای نیازی به تهیه مواد خام گران قیمت یا دور از دسترس وجود ندارد و تولید انبوه آنها برای کلیه قدرت های نظامی برتر دنیا در همه جا امکان پذیر خواهد بود و هزینه زیادی را در پی نخواهد داشت. فقط ممکن است که قبل از ورود این سلاح ها به بازار سیاه و رسیدن به دست دیکتاتورهای کنترل کننده مردم عادی، مدت زمان زیادی طول بکشد. سلاح های خودران برای کارهایی مانند ترور اشخاص، بی ثبات کردن وضعیت کشورها، تحت اختیار در آوردن مردم و کشتن گزینشی یک گروه قومی خاص، ابزارهای ایده آلی هستند. بنابراین، ما معتقد هستیم که رقابت تسلیحاتی نظامی مبتنی بر هوش مصنوعی، فایده ای برای بشریت نخواهد داشت. راه های زیادی برای ایمن تر کردن میادین جنگی برای انسان ها و به خصوص افراد غیر نظامی وجود دارد. همانطور که بیشتر شیمی دان ها یا زیست شناسان علاقه ای به ساختن سلاح های شیمیایی یا بیولوژیکی ندارند، بیشتر محققان هوش مصنوعی علاقه ای به ساختن سلاح های مبتنی بر هوش مصنوعی ندارند و نمی خواهند که در رشته علمی خود رسوا شوند و با واکنش منفی شدید مردم علیه هوش مصنوعی و محرومیت از مزایای اجتماعی در آینده مواجه شوند. در واقع، شیمی دان ها و زیست شناسان به طور گسترده ای از توافقات بین المللی علیه توسعه سلاح های شیمیایی و بیولوژیکی، حمایت می کنند. به شکل مشابهی، بیشتر فیزیکدانان نیز از معاهده های منع ساخت سلاح های هسته ای فضایی و سلاح های لیزری حمایت می کنند.

من تلاش کردم که این نامه سرگشاده را به امضای بسیاری از محققان سرسخت هوش مصنوعی و رباتیک برسانم. قبلا هم پویش بین المللی کنترل تسلیحات رباتیک، اقدام به جمع آوری امضای صدها نفر برای مخالفت با ربات های کشنده، کرده بود. با این حال، فکر می کردم که کار جمع آوری امضاها را بهتر از این پویش انجام خواهم داد. من می دانستم که سازمان های حرفه ای تمایلی برای ارائه فهرست ایمیل اعضای خود را ندارند و به همین دلیل، فهرست نام محققان و موسسات آنها را از اسناد آن لاین جمع کردم و ظرف بیست و چهار ساعت، فهرست ایمیل صدها محقق هوش مصنوعی را تهیه کردم و طولی نکشید که از بیش از ۳۰۰۰ محقق هوش مصنوعی و رباتیک که شامل

رهبران هوش مصنوعی از شرکت های گوگل، فیسبوک، مایکروسافت و تسلا می شدند، خواستم که نامه سرگشوده را امضا کنند و آنها این کار را انجام دادند و کنفرانس خبری مربوط به این موضوع در کنفرانس مشترک بین المللی هوش مصنوعی، به یک داستان خبری مهم در جهان تبدیل شد. موضع گیری زیست شناسان و شیمی دان ها باعث شد که رشته های آنها به عنوان رشته هایی شناخته شوند که به تولید داروهای مفید کمک می کنند، نه به تولید سلاح های بیولوژیک و شیمیایی. امضا کنندگان این نامه سرگشوده از رشته های علمی می خواستند که به ایجاد آینده بهتر کمک کنند و نه به ایجاد راه های جدیدی برای کشتن مردم. این سوال مطرح می شود که استفاده کنندگان اصلی هوش مصنوعی در آینده افراد غیر نظامی خواهند بود یا افراد نظامی؟ به نظر می رسد که افراد غیر نظامی گزینه درست است، اما با شروع رقابت تسلیحاتی نظامی مبتنی بر هوش مصنوعی، احتمالاً شاهد این خواهیم بود که پول زیادی صرف امور نظامی شود. تعهدات سرمایه گذاری غیر نظامی آمریکا روی هوش مصنوعی در سال ۲۰۱۶ فقط یک میلیارد دلار بود، اما بودجه نظامی پنتاگون برای پروژه های مربوط به هوش مصنوعی در سال ۲۰۱۷، به ۱۲ تا ۱۵ میلیارد دلار می رسید.

### آیا باید یک توافقنامه بین المللی وجود داشته باشد؟

با وجود فشاری که برای یک مذاکره بین المللی در مورد منع استفاده از برخی ربات های کشنده وجود دارد، هنوز معلوم نیست که چه اتفاقی خواهد افتاد. هرچند که بسیاری از ذی نفعان پیشرو در زمینه سیستم های تسلیحاتی با این موضوع موافق هستند که قدرت های جهانی برای هدایت تحقیقات و اجرای سیستم های تسلیحاتی خودران، باید پیش نویس قوانین بین المللی را امضا کنند، قراردادی در مورد اینکه دقیقاً چه چیزهایی باید منع شوند و اینکه این قوانین را چگونه باید اجرا کرد، وجود ندارد. به عنوان مثال، آیا ممنوعیت فقط برای توسعه سلاح های خودران کشنده است یا شامل آنهایی که به آسیب های جدی از جمله کور شدن افراد منجر می شوند نیز می شود؟ آیا ممنوعیت ها را فقط باید برای سلاح های خودران مهاجم اعمال کرد و ممنوعیتی برای سیستم های دفاعی مانند ضد هوایی ها و سیستم های دفاع موشکی که به راحتی می توانند وارد مرز دشمن شوند، وجود ندارد؟

چگونه می توان قراردادی را اعمال کرد و به این موضوع که بیشتر اجزای سلاح های خودران دارای استفاده دوگانه علیه افراد غیرنظامی نیز می باشد، توجهی نکرد؟ به عنوان مثال، پهبادی را در نظر بگیرید که علاوه بر که تحویل دادن بسته محصولات آمازون، می تواند بمب نیز تحویل دهد. برخی از مناظره کنندگان استدلال می کنند که طراحی یک قرارداد موثر برای گسترش سیستم های تسلیحاتی خودران، کار سخت و نا امید کننده ای است و نباید آن را آزمایش کرد. از طرف دیگر، جان کندی در اعلامیه ماموریت سفر به کره ماه، تاکید می کند که کارهای سختی که منجر به فواید قابل توجهی برای بشریت می شوند، ارزش آزمایش کردن را دارند. علاوه بر این، بسیاری از کارشناسان استدلال می کنند که ممنوعیت استفاده از سلاح های بیولوژیکی و شیمیایی با وجود سخت بودن در مرحله اجرا، کار ارزشمندی است. فرصتی ایجاد شد تا با هنری کیسینجر ملاقات کنم و نقش او را در ممنوعیت سلاح های بیولوژیکی جویا شوم. او توضیح داد که وقتی مشاور امنیت ملی آمریکا بود، نیکسون، رئیس جمهور این کشور را قانع کرده بود که ممنوعیت این سلاح ها به نفع امنیت ملی کشورش است. از آنجایی که آمریکا به دلیل برخورداری از نیروهای هسته ای به عنوان ابر قدرت شناخته شده بود، در صورت مشارکت در رقابت تسلیحاتی بیوشیمیایی جهانی که نمی تواند نتیجه مشخصی داشته باشد، دچار زیان بزرگی می شد. زیرا آن کشورهایی که سعی می کنند بیشترین سود را از رقابت تسلیحاتی ببرند و سلاح های تازه تولید شده را از بازار سیاه خریداری کنند، ابر قدرت نیستند، بلکه کشورهای بدذات و فرومایه ای هستند. پهبادهای کشنده مبتنی بر هوش مصنوعی بسیار ارزان قیمت هستند و اگر شما بخواهید که یک شخص سیاسی را ترور کنید، فقط کافی است که عکس و نشانی شخص مورد هدف را به داخل پهباد کشنده بارگذاری کنید. سپس، پهباد به سمت مقصد پرواز می کند، شخص مورد هدف را شناسایی و نابود می کند و برای اینکه کسی نفهمد که مقصر چه کسی بوده، خود را منهدم می کند. به همین منوال، افرادی که مصمم به پاکسازی قومی هستند، می توانند پهبادها را به گونه ای برنامه نویسی کنند که فقط افراد دارای رنگ پوست یا قومیت خاص را بکشند. استوارت پیش بینی می کند که هرچه این سلاح ها هوشمندتر می شوند، برای هر بار

کشتن افراد، به مواد، قدرت آتش و پول کمتری نیاز خواهند داشت. استوارت از این می ترسد که پهپادهای به اندازه زنبور عسل، با هزینه کم و استفاده از حداقل قدرت انفجاری، بتوانند به چشم افراد شلیک کنند تا گلوله از آنجا وارد مغز آنها شود یا اینکه سر فرد را با چنگال فلزی بگیرند و یک فشنگ کوچک را وارد جمجمه او کنند. اگر یک میلیون از چنین پهپادهای کشنده ای را بتوان در پشت یک کامیون به جایی ارسال کرد، نوع جدیدی از سلاح های مرعوب کننده کشتار جمعی در دسترس دشمنان قرار خواهند گرفت که می توانند به طور گزینشی، گروه از قبل تعیین شده ای از افراد را بکشند و آسیبی به افراد یا چیزهای دیگر وارد نکنند. استدلال مخالفین این است که ما می توانیم ربات های کشنده را اخلاق مدار بسازیم تا فقط سربازهای دشمن را بکشند. اما برای اجرای این ممنوعیت، آیا راحت تر نیست که از همان اول به جای ملزم ساختن دشمن به اخلاق مدار کردن ۱۰۰ درصد ربات های کشنده، از او بخواهیم که ربات های کشنده را تولید نکند؟ آیا می توان ادعا کرد که ربات ها بهتر از سربازان آموزش دیده کشورهای متمدن، قوانین را رعایت می کنند؟ آیا می توان ادعا کرد که افراد بدذات و فرومایه و دیکتاتورها از قوانین جنگ پیروی می کنند و از ربات ها برای زیرپا گذاشتن قوانین جنگ استفاده نخواهند کرد؟

## جنگ سایبری

یکی از جنبه های نظامی هوش مصنوعی این است که به شما اجازه می دهد تا حتی بدون اینکه خودتان سلاحی را بسازید، طریق جنگ افزارهای سایبری به دشمن حمله کنید. کرم استوکسنت<sup>۴۷</sup> که به دولت های آمریکا و اسرائیل نسبت داده می شود، سانتریفیوژهای موجود در برنامه غنی سازی انرژی هسته ای ایران را آلوده کرد و منجر به متلاشی شدن آنها شد. هرچه قدر که فعالیت های جامعه خودکارتر و حمله های مبتنی بر هوش مصنوعی قوی تر شوند، جنگ افزارهای سایبری ویران کننده تر خواهند شد. اگر شما بتوانید خودروهای بدون راننده، هواپیماهای بدون خلبان، راکتورهای هسته ای، ربات های صنعتی، سیستم های ارتباطی، سیستم های مالی و شبکه های برق دشمن را هک کنید

---

<sup>۴۷</sup> Stuxnet worm

و متلاشی سازید، به شکل موثری قادر به فروپاشی اقتصادی و فلج کردن قدرت دفاعی دشمن خواهید بود و حتی می توانید در سیستم های تسلیحاتی دشمن رخنه کنید. ما فقط در صورتی می توانیم از فرصت های کوتاه مدت هوش مصنوعی برای کمک به بشریت استفاده کنیم که آن را مقاوم و غیر قابل هک شدن بسازیم. هرچند که از خود هوش مصنوعی می توان برای افزایش مقاومت سیستم های هوش مصنوعی استفاده کرد و به دفاع در مقابل جنگ سایبری کمک کرد، اما هوش مصنوعی می تواند دارای قدرت تهاجمی نیز باشد. به همین دلیل، یکی از اهداف کوتاه مدت برای توسعه هوش مصنوعی باید این باشد که قدرت دفاعی هوش مصنوعی بر قدرت تهاجمی آن چیره شود. در غیر این صورت، ممکن است که هوش مصنوعی دشمن خودش شود.

## شغل ها و دستمزدها

تا اینجا، بحث شد که هوش مصنوعی چگونه می تواند با ایجاد تحول در محصولات و خدمات جدید با قیمت مناسب بر ما که یک مصرف کننده هستیم، تاثیر گذارد. اما باید بررسی کرد که هوش مصنوعی چگونه می تواند با ایجاد تحول در بازار شغلی بر ما که یک کارگر هستیم، تاثیر گذارد. اگر ما بتوانیم راه هایی را برای افزایش رفاه خود از طریق خودکار سازی فعالیت ها و البته بدون رها کردن افراد بی درآمد پیدا کنیم، پتانسیل لازم برای ایجاد یک آینده لذت بخش بی نظیر و پر ناز و نعمت برای همه را خواهیم داشت. یکی از اقتصاددانان دانشگاه ام آی تی چشم انداز خوش بینانه خود به بازار شغلی را در "آتن دیجیتال" جستجو می کند. علت اینکه شهروندان آتن در دوران باستان زندگی لذت بخشی داشتند و می توانستند از دموکراسی، هنر و بازی ها لذت ببرند، این بود که بیشتر کارهایشان را برده ها انجام می دادند. امروزه، به جای برده ها می توان از ربات های مجهز به هوش مصنوعی استفاده کرد و یک مدینه فاضله دیجیتال را برای لذت بردن همه ایجاد کرد. اقتصاد مبتنی بر هوش مصنوعی توصیف شده توسط این اقتصاددان نه تنها باعث حذف استرس و کار پر زحمت شده و همه چیزهایی که می خواهیم را به وفور برایمان فراهم خواهد ساخت، بلکه تعداد زیادی از محصولات و خدمات جدیدی را که تاکنون نخواستہ بودیم را نیز به ما ارزانی خواهد داشت.

## فن آوری و بی عدالتی

در صورتی می توان به مدینه فاضله دیجیتالی رسید که در آمد ساعتی افراد، هر سال نسبت به سال قبل افزایش یابد تا آنهایی که اوقات فراغت بیشتری را می خواهند، بتوانند حجم کار خود را به تدریج کاهش دهند و در عین حال، استاندارد زندگی خود را بهبود دهند. این دقیقا همان چیزی است که از زمان جنگ جهانی دوم تا اواسط دهه ۱۹۷۰ در آمریکا اتفاق افتاد و باعث شد که با وجود بی عدالتی در آمدی، سهم همه افراد از برش کیک پای افزایش یابد. اما، در آن زمان تغییر دیگری نیز اتفاق افتاد. هر چند که اقتصاد آمریکا رشد یافت و منجر به افزایش در آمد متوسط افراد شد، سود ناشی از رشد اقتصادی در چهار دهه گذشته به ۱ درصد ثروتمند جامعه رسید و در آمد ۹۰ درصد فقیرتر جامعه، تغییری نکرد. بی عدالتی در مورد توزیع ثروت از بی عدالتی در مورد توزیع در آمد، آشکارتر بود. در سال ۲۰۱۲، ارزش خالص دارایی ها برای ۹۰ درصد خانواده های فقیر آمریکایی در طی ۲۵ سال گذشته تغییری نکرد و حدود ۸۵۰۰۰ دلار بود، در حالی که ارزش خالص دارایی ها برای ۱ درصد ثروتمند جامعه در طی این مدت، به ۱۴ میلیون دلار رسید. در سطح بین المللی نیز شاهد بودیم که در سال ۲۰۱۳ ثروت انباشته نصف فقیر جمعیت دنیا (۳,۶ میلیارد نفر) با ثروت ۸ ثروتمند بزرگ دنیا برابری می کرد. این آمارها بر فقر و آسیب پذیری طبقه پایین جامعه و ثروت زیاد طبقه بالای جامعه تاکید می کنند. در کنفرانسی که در سال ۲۰۱۵ در پورتوریکو برگزار شد، این اقتصاد دان به محققان هوش مصنوعی گفت که فکر می کند پیشرفت های صورت گرفته در زمینه هوش مصنوعی و اتوماسیون ادامه خواهد یافت و باعث بزرگتر شدن سهم افراد از برش کیک پای خواهد شد. اما هیچ قانون اقتصادی وجود ندارد که بر اساس آن، همه یا حتی بیشتر افراد از این افزایش سهم بهره مند خواهند شد. با وجود توافق گسترده بین اقتصاد دانان در مورد اینکه بی عدالتی در حال افزایش است، اختلاف نظرهایی در مورد دلایل ادامه یافتن این روند وجود دارد. سیاستمداران چپ گرا اغلب استدلال می کنند که دلیل اصلی این موضوع، "جهانی سازی" و سیاست های اقتصادی دولت در مورد کاهش مالیات برای افراد ثروتمند می باشد. اما گروه دیگری از سیاستمداران استدلال می کنند



که دلیل اصلی این موضوع، به "فن آوری" مربوط می شود. آنها استدلال می کنند که فن آوری دیجیتال از سه طریق مختلف باعث بی عدالتی شده است. طریق اول، جایگزینی شغل های قدیمی با شغل هایی است که نیاز به مهارت های بیشتری دارند و به همین دلیل، فن آوری باعث شده که پاداش ها فقط برای افراد تحصیل کرده در نظر گرفته شوند. از اواسط دهه ۱۹۷۰ درآمد فارغ التحصیلان دوره های تحصیلات تکمیلی ۲۵ درصد افزایش یافته است، اما درآمد متوسط ترک تحصیل کرده های دبیرستانی ۳۰ درصد کاهش یافته است. طریق دوم، برخورداری مالکان از درآمد بیشتر است. از سال ۲۰۰۰ مالکان شرکت ها نسبت به افرادی که در آن شرکت ها کار می کنند، سهم بیشتری از درآمد شرکت ها را نصیب خود کرده اند. از مالکان شرکت ها انتظار می رود که سهم بزرگتری از برش کیک پای را دریافت کنند. این برتری سرمایه به نیروی کار، می تواند نقش مهمی را در رشد اقتصاد دیجیتال ایفا کند. اکنون که همه چیز، از کتاب گرفته تا فیلم و ابزارهای آماده سازی مالیات، دیجیتال شده اند، کپی های اضافی آنها را می توان بدون صرف کوچکترین هزینه ای و بدون نیاز به استخدام کارمندان، به سراسر جهان فروخت. این باعث می شود که درآمدها فقط نصیب سرمایه گذاران شود، نه کارمندان. طریق سوم عبارت از سود بردن ابرستاره ها می باشد، زیرا اقتصاد دیجیتال فقط به نفع ابرستاره ها تمام می شود. رولینگ، نویسنده داستان هری پاتر اولین نویسنده ای بود که به باشگاه میلیاردرها ملحق شد و او بسیار ثروتمندتر از شکسپیر شد، زیرا داستان های او را با هزینه بسیار کمی می شد در قالب متن، فیلم و بازی به دست میلیاردها نفر در سراسر جهان رساند. اسکات کوک با معرفی نرم افزار آماده سازی مالیات توربو تکس که بر خلاف نمونه های انسانی فقط با یک دان لود می شد آن را فروخت، یک میلیارد دلار پول کسب کرد. از آنجایی که مردم تمایلی به خرید نرم افزارهای آماده سازی مالیات رده دهم را ندارند، فقط برای تعداد کمی از ابرستاره ها جایی برای رقابت در محیط بازار وجود خواهد داشت. این بدین معناست که اگر تمام والدین جهان به بچه های خود توصیه کنند که ابر ستاره بعدی شوند، تقریباً هیچ یک از بچه ها آن را به عنوان یک راهبرد شغلی امکان پذیر در نظر نمی گیرند.

## پیشنهاد شغلی به بچه ها

پس، چه پیشنهاد شغلی ای را باید به بچه ها ارائه کرد. خود من به دنبال شغل هایی می گردم که ماشین ها در حال حاضر، نتوانند آنها را به خوبی انجام دهند تا امکان خود کارسازی آنها در آینده نزدیک وجود نداشته باشد. پیش بینی های اخیر در مورد اینکه چه زمانی شغل های مختلف بر عهده ماشین ها گذاشته خواهد شد، نشان می دهد که قبل از تصمیم گرفتن به دریافت آموزش های شغلی باید سوالات مهم زیر را در مورد یک شغل از خود پرسید.

- آیا این شغل به تعامل خوب با افراد و برخورداری از هوش اجتماعی بالایی نیاز دارد؟
- آیا این شغل به خلاقیت بالا و ارائه راه حل های هوشمندانه نیاز دارد؟
- آیا این شغل به کار کردن در یک محیط غیر قابل پیش بینی نیاز دارد؟

هرچه قدر که شما از این ویژگی ها برخوردار باشید، گزینه بهتری را انتخاب خواهید کرد. از میان چنین شغل هایی می توان به شغل یک معلم، پرستار، پزشک، دندانپزشک، دانشمند، کارآفرین، برنامه نویس، مهندس، وکیل، مددکار اجتماعی، مرد روحانی، هنرمند، آرایشگر یا ماساژ درمانگر اشاره کرد. در عوض، شغل هایی هم وجود دارند که شامل اقدامات به شدت تکراری یا ساختاریافته در شرایط قابل پیش بینی می باشند که دوام زیادی ندارند و به زودی خود کار سازی خواهند شد. از دیرباز، رایانه ها و ربات های صنعتی، ساده ترین کارها را به عهده گرفته اند و با پیشرفت فن آوری بسیاری از شغل ها از جمله شغل بازاریابان از راه دور، کارگران انبار، صندوقدارها، اپراتورهای قطار، نانوان ها و کمک آشپزها به تدریج حذف خواهند شد. شغل راننده های کامیون، اتوبوس و تاکسی نیز به زودی حذف خواهند شد. بسیاری از شغل های دیگر (از جمله مددیار حقوقی، تحلیلگر اعتباری، مسئول وام، دفتردار و حسابدار مالیاتی) در فهرست شغل های کاملاً نابود شده قرار ندارند، اما بسیاری از وظایف آنها خود کارسازی شده و به نیروی انسانی کمی نیاز دارند. مصون ماندن از خود کار سازی تنها چالش شغلی نیست. در عصر دیجیتال جهانی، تلاش برای تبدیل شدن به یک

نویسنده، فیلم ساز، بازیگر، ورزشکار یا طراح مد لباس حرفه ای، کار بسیار پر خطری است و صاحبین این شغل ها به جای رقابت با ماشین ها، با رقابت بی رحمانه با انسان های دیگری مواجه هستند که آنها نیز می خواهند ستاره شوند و این در حالی است که افراد بسیار کمی در این شغل ها موفق می شوند. بسیاری از شغل ها وجود دارند که کاملاً حذف نخواهند شد، اما بسیاری از وظایفشان خود کار سازی خواهد شد. به عنوان مثال، اگر می خواهید وارد شغل پزشکی شوید، رادیولوژیستی نشوید که کار آن را یک رایانه نیز می تواند انجام دهد، بلکه پزشکی شوید که دستور رادیولوژی می دهد، در مورد نتایج با بیمار صحبت می کند و برای درمان بیمار تصمیم می گیرد. اگر می خواهید وارد امور مالی شوید، کارشناس تحلیل داده ها نشوید که کار آن را یک نرم افزار نیز می تواند انجام دهد، بلکه مدیر مالی ای شوید که از نتایج تحلیل های کمیت پذیر برای گرفتن تصمیمات راهبردی سرمایه گذاری استفاده می کند. اگر می خواهید وارد امور حقوقی شوید، دستیار حقوقی ای نشوید که هزاران سند را برای مرحله مقدماتی کشف جرم بررسی می کند. بلکه وکیلی شوید که با موکل خود مشورت می کند و پرونده را تقدیم دادگاه می کند. در اینجا به بررسی این موضوع می پردازیم که دولت ها برای کمک به موفقیت نیروهای کاری خود چه اقداماتی را می توانند انجام دهند؟ به عنوان مثال، با توجه به پیشرفت های سریع هوش مصنوعی چه تغییراتی را می توان در نظام آموزشی ایجاد کرد تا افراد بتوانند برای ورود به بازار شغلی آمادگی بیشتری پیدا کنند؟ در مدل فعلی پس از دو دهه آموزش، وارد چهار دهه شغل تخصصی می شویم و امیدواریم که هوش مصنوعی به ما کمک کند تا نظام آموزشی را به گونه ای متحول سازیم تا افراد در ابتدا چند سال کار کنند، سپس یک سال به مدرسه بروند و سپس، چند سال دیگر هم کار کنند یا اینکه امکان ادامه آموزش شغلی آنلاین را برای هر شغلی فراهم سازیم. اندرو مک آفی استدلال می کند که برای ایجاد شغل های مناسب جدید، به سیاست های اقتصادی مفیدی مانند سرمایه گذاری وسیع در زمینه تحقیقات، آموزش و زیرساخت ها، تسهیل بخشیدن به مهاجرت ها و ایجاد انگیزه برای کارآفرینی نیاز است.

**آیا انسان ها در نهایت غیر قابل استخدام خواهند شد؟**

اگر پیشرفت های هوش مصنوعی ادامه پیدا کند و شغل های بیشتری خود کار سازی شوند، چه اتفاقی خواهد افتاد؟ بسیاری از افراد به آینده شغلی خوش بین هستند و استدلال می کنند که شغل های خود کار سازی شده با شغل های جدید بهتری جایگزین خواهند شد. اما، افراد زیادی هم به آینده شغلی بدبین هستند و استدلال می کنند که با گذشت زمان، بر تعداد افراد غیر قابل استخدام افزوده خواهد شد و درآمدها در بازار آزاد و بر اساس عرضه و تقاضا تعیین خواهد شد و درآمد کارگران به دلیل عرضه بیش از حد نیروی کار ارزان، به زیر سطح مخارج زندگی آنها خواهد رسید. همچنین، هر جا که امکان برون سپاری یک شغل خاص به یک کشور کم درآمدتر یا یک ماشین ارزان وجود داشته باشد، درآمد انسان ها کاهش خواهد یافت. طی انقلاب صنعتی، انسان ها فهمیدند که می توانند کارهای بدنی سخت را به ماشین ها بسپارند و خودشان در مشاغلی استخدام شوند که نیاز به فکر بیشتری دارند و دارای دستمزدهای بالاتری هستند. به این ترتیب، شغل های اداری جایگزین شغل های کارگری شدند. انسان ها به تدریج به دنبال جایگزینی فعالیت های فکری توسط ماشین ها هستند. اگر این کار با موفقیت انجام شود، دیگر شغلی برای انسان ها باقی نخواهد ماند. برخی از افراد خوش بین به آینده شغلی استدلال می کنند که بعد از شغل های فیزیکی و ذهنی شاهد توسعه شغل های خلاقانه خواهیم بود. اما افراد بدبین در مخالفت با آنها استدلال می کنند که خلاقیت هم یک فرایند ذهنی است و در نهایت توسط هوش مصنوعی مدیریت خواهد شد. افراد خوش بین دیگری نیز امیدوارند که تحول بعدی در شغل های مبتنی بر فن آوری اتفاق بیفتد که هرگز فکر آن را نمی کردیم. اما افراد بدبین معتقدند که این یک خیال واهی است و داده های تجربی از آنها حمایت نمی کنند. آنها می گویند که یک قرن پیش هم پیش بینی می شد که انقلاب رایانه ای باعث ایجاد شغل های جدید فن آوری محور خواهد شد، اما این یک پیش بینی مفتضحانه بود. بیشتر شغل های امروزی همان هایی هستند که یک قرن پیش هم وجود داشتند. تنها شغل جدید، توسعه نرم افزار است که فقط کمتر از یک درصد از بازار شغلی آمریکا را تشکیل می دهد. روند اصلی در بازار شغلی، ایجاد شغل های کاملاً جدید نیست، بلکه مانند جمع شدن بر روی زمین های ناهمواری است که هنوز به واسطه

جزر و مد فن آوری غرق نشده اند. این ها جزیره های مستقلی نیستند، بلکه مجمع الجزایری هستند که علاوه بر حرفه های مربوط به فن آوری پیشرفته مانند توسعه نرم افزار، شامل حرفه های مبتنی بر چابکی و مهارت های اجتماعی می باشند که ماشین ها هنوز نمی توانند آنها را به خوبی انسانها انجام دهند. افراد بدین به آینده شغلی استدلال می کنند که در نهایت، کل مجمع الجزایر غرق خواهد شد و شغلی برای بشر باقی نمی ماند که بتوان آنها را به ارزانی ماشین ها انجام داد. گرگوری کلارک<sup>۴۸</sup>، اقتصاد دان آمریکایی اشاره می کند که ما از اسب ها می توانیم چیزهای جالبی را در مورد آینده شغلی بشر یاد بگیریم. دو اسبی را تصور کنید که در سال ۱۹۰۰ به یک خودروی قدیمی نگاه می کنند و به آینده خود فکر می کنند.

"من نگران بیکاری ناشی از فن آوری هستم."

"به فن آوری بدین نباش. اجداد ما هم وقتی که قطارها جای شغل ما (جایجایی کالسکه ها) را گرفتند، همین حرف ها را می زدند. اما، اکنون شغل های خوب زیادی داریم."

"اما، اگر این موتور احتراق داخلی توسعه پیدا کند، چه اتفاقی خواهد افتاد؟"

"من مطمئنم که اسب ها شغل های جدیدی پیدا خواهند کرد که تصورش غیر ممکن است. در زمان اختراع چرخ و گاوآهن نیز همین نگرانی ها وجود داشت."

متأسفانه، آن شغل های جدیدی که انتظارش می رفت، تحقق پیدا نکردند. اسب هایی که دیگر مورد نیاز نبودند را کشتند و جمعیت اسب ها در آمریکا از حدود ۲۶ میلیون راس در سال ۱۹۱۵ به حدود ۳ میلیون راس در سال ۱۹۶۰ رسید. همانطور که وسایل مکانیکی توانستند جای قدرت ماهیچه اسب ها را بگیرند، آیا ذهن های مکانیکی هم می توانند جای ذهن بشر را بگیرند؟

## اعطای درآمد به افراد فاقد شغل

---

<sup>۴۸</sup> Gregory Clark

حق با چه کسانی است؟ آنهایی که می گویند شغل های خود کار سازی شده با شغل های بهتری جایگزین خواهند شد یا آنهایی که می گویند بیشتر انسان ها در نهایت غیر قابل استخدام خواهند شد؟ اگر پیشرفت هوش مصنوعی متوقف نشود، ممکن است که حق با هر دو طرف باشد. یکی در کوتاه مدت و دیگری در بلندمدت. اگرچه افراد با معانی ضمنی بدبینانه و ناامید کننده به نابودی شغل ها اشاره می کنند، این نابودی شغل ها لزوما چیز بدی نیست. آن دسته از افراد بدبین به فن آوری که می گویند برخی از شغل های خاص نباید از بین بروند، توجهی به این موضوع ندارند که شغل های دیگر نیز ممکن است ارزش اجتماعی مشابهی را برای فرد ایجاد کنند. ما از شغل ها انتظار داریم که برایمان درآمد و هدف ایجاد کنند. اما با توجه به وفور امکانات رفاهی تولید شده توسط ماشین ها، امکان یافتن راه های دیگری به جز استخدام شدن در یک شغل نیز برای تامین درآمد و هدف وجود خواهد داشت. همین اتفاق در مورد اسب ها افتاد و در نهایت، همه آنها منقرض نشدند. در عوض، تعداد اسب ها از سال ۱۹۶۰ به بعد، تقریباً سه برابر شد، زیرا از طریق یک سیستم تامین کننده رفاه اجتماعی از اسب ها محافظت می شد و بدون اینکه آنها پولی بپردازند، تحت مراقبت بودند و امکانات تفریحی، ورزشی و همنشینی با دوستان برایشان فراهم بود. اکنون، این سوال مطرح می شود که آیا امکان تامین مراقبت های مشابه برای انسان های نیازمند نیز وجود دارد؟ توزیع مجدد بخش کوچکی از کیک پای به سهم افراد باعث می شود که وضع اقتصادی مردم بهتر شود. بسیاری از افراد بر لزوم این امر تاکید می کنند. در میزگردی که در سال ۲۰۱۶ برگزار شد، صحبت از این شد که نجات زندگی افراد با فن آوری مبتنی بر هوش مصنوعی و سهم ساختن افراد از ثروت یک جامعه، یک امر واجب اخلاقی است. یکی از اقتصاددانان شرکت کننده در این میز گرد گفت: "اگر ما با این همه تولید ثروت جدید، نتوانیم مانع از فقیرتر شدن نیمی از جمعیت کشورمان شویم، باید خجالت بکشیم." پیشنهادهای بسیار مختلفی برای سهم ساختن افراد از ثروت جامعه وجود دارد که هر یک دارای طرفداران و منتقدین خاص خودش است. ساده ترین آنها اعطای "درآمد پایه" است که در آن، هر فردی بدون هیچ گونه پیش شرط یا پیش نیازی، مبلغی را به عنوان درآمد ماهیانه دریافت می

کند. به عنوان مثال، طرح اعطای درآمد پایه به افراد در مقیاس کوچک در کشورهای کانادا، فنلاند و هلند، آزمایش یا برنامه ریزی شده است. طرفداران این طرح استدلال می کنند که "اعطای درآمد پایه" کاراتر از گزینه های جایگزینی مانند پرداخت پول از سوی دولت به افراد نیازمند است، زیرا باعث حذف گرفتاری های اداری موجود برای تعیین افراد واجد شرایط می شود. طرح پرداخت پول به افراد نیازمند به دلیل کاهش انگیزه این افراد برای کار کردن، مورد انتقاد واقع شده است. با این حال، در آینده ای که هیچ شغلی وجود نخواهد داشت و کسی کار نمی کند، چنین طرحی می تواند مفید واقع شود. دولت ها علاوه بر پول دادن به شهروندان خود، می توانند با ارائه خدمات رایگان یا یارانه ای از قبیل ساختن جاده ها، پل ها، پارک ها و ارائه خدمات حمل و نقل عمومی، مراقبت از کودکان، آموزش، سلامت، خانه سالمندان و دسترسی به اینترنت به شهروندان خود کمک کنند. در واقع، بسیاری از دولت ها در حال حاضر بیشتر این خدمات را ارائه می کنند. این خدمات که با بودجه دولت ها تامین می شوند، دو هدف جداگانه را دنبال می کنند. اول اینکه باعث کاهش مخارج زندگی افراد می شوند و دوم اینکه شغل هایی را برای افراد ایجاد می کنند. حتی در آینده ای که ماشین ها در تمام شغل ها بهتر از انسان ها عمل خواهند کرد، دولت ها ممکن است ترجیح دهند که به جای برون سپاری شغل های مراقبتی مانند نگهداری از کودکان یا سالمندان به ربات ها به افرادی که در این شغل ها کار می کنند، پول بپردازند. جالب است که پیشرفت فن آوری بدون مداخله دولت نیز می تواند به مجانی شدن بسیاری از محصولات و خدمات ارزشمند منتهی شود. به عنوان مثال، امروزه هر فردی که به اینترنت دسترسی داشته باشد، می تواند تمام کارهای خود مانند مراجعه به دانشنامه، فرستادن نامه، تماس تلفنی، شرکت در ویدئو کنفرانس، به اشتراک گذاشتن عکس ها، دسترسی به رسانه های اجتماعی، شرکت در دوره های آموزشی آنلاین و بسیاری از خدمات دیگر را به صورت مجانی انجام دهد. همچنین، بسیاری از اقدامات با ارزش و حیاتی برای انسان ها مانند مصرف دوره ای آنتی بیوتیک ها با هزینه بسیار پائینی امکان پذیر شده است. به لطف فن آوری پیشرفته، امروزه بسیاری از افراد فقیر به چیزهایی دسترسی دارند که ثروتمندترین افراد جهان در

گذشته به آنها دسترسی نداشتند. به طوریکه عده ای معتقدند که درآمدها برای داشتن یک زندگی قابل قبول، کافی نیست. اگر یک روز ماشین ها بتوانند تمام کالاها و خدمات امروزی را با کمترین هزینه فراهم سازند، دارایی همه افراد برای داشتن یک زندگی مرفه کفایت خواهد کرد. به عبارت دیگر، دولت ها با دریافت کمترین مالیات ها می توانند هزینه های مربوط به درآمد پایه و خدمات مجانی خود را بپردازند. سهام ساختن تمام افراد از ثروت جامعه چیزی است که ممکن است در عمل اتفاق نیفتد و حتی برخی از سیاستمداران با اعمال این سیاست مخالفت می کنند. همانطور که دیدید، روند فعلی توزیع ثروت در آمریکا در جهت مخالف این سیاست است و در هر دهه شاهد فقیرتر شدن گروهی از افراد نسبت به دهه قبلی هستیم. تصمیم گیری در مورد نحوه توزیع ثروت جامعه بر روی همه تاثیر خواهد گذاشت و به همین دلیل، همه افراد جامعه باید در بحث های مربوط به ساختن اقتصاد آینده جامعه خود، مشارکت کنند. بسیاری از مناظره کنندگان استدلال می کنند که کاهش بی عدالتی در توزیع درآمد، ایده خوبی برای دنیای امروز و آینده تحت فرمان هوش مصنوعی محسوب می شود. اسناد موجود نشان می دهند که بحث افزایش عدالت در جامعه که یک بحث اخلاقی است، به تقویت دموکراسی نیز کمک می کند. وقتی که افراد تحصیل کرده زیادی در طبقه متوسط وجود داشته باشند، فریب دادن رای دهندگان و اعمال نفوذ بی مورد چند نفر یا شرکت بر دولت سخت تر می شود. بهبود دموکراسی باعث مدیریت بهتر اقتصاد، کاهش فساد، افزایش کارایی، رشد سریعتر اقتصادی و سود بردن همه افراد جامعه خواهد شد.

## اعطای هدف به افراد فاقد شغل

شغل ها فقط تامین کننده درآمد برای افراد نیستند. ولتر<sup>۴۹</sup> در سال ۱۷۵۹ می نویسد: "کار باعث جلوگیری از حمله سه شیطان بزرگ به نام های بی حوصلگی، گناه و فقر می شود." از سوی دیگر، تامین درآمد برای تضمین رفاه افراد کافی نمی باشد. امپراتورهای روم برای راضی نگهداشتن زیردستان خود، با تامین خوراکی و سرگرمی کاری می کردند که حواسشان پرت شود. حضرت

---

<sup>۴۹</sup> voltaire



عیسی با اشاره به آیات کتاب مقدس روی نیازهای غیر مادی انسان ها تاکید می کرد. اکنون این سوال مطرح می شود که شغل ها علاوه بر پول، دقیقا چه چیزهای با ارزشی را برایمان تامین می کنند و جامعه بدون شغل از چه راه های دیگری می تواند پاسخگوی نیازهای غیرمادی ما باشد؟ پاسخ این سوالات پیچیده است، زیرا برخی افراد از شغل خود متنفرند و برخی دیگر عاشق شغل خود هستند. علاوه بر این، بسیاری از بچه ها، دانشجویها و زنان خانه دار بدون داشتن شغل، حال بهتری دارند. تاریخ پر از داستان هایی در مورد شاهزاده های نازپرورده و پرنسس هایی است که تسلیم احساس پوچی و افسردگی شده اند. طبق تحلیلی که در سال ۲۰۱۲ انجام شد، شاغل نبودن دارای تاثیرات منفی بلندمدتی بر سلامتی افراد می باشد و بازنشستگی هم دارای جنبه های مثبت و هم دارای جنبه های منفی می باشد. حوزه روانشناسی مثبت اندیشی تعدادی از عواملی را که باعث افزایش حس سلامتی و هدفمندی زندگی افراد می شود، شناسایی کرده است و متوجه شده است که برخی از شغل ها تامین کننده بسیاری از این عوامل هستند. برخی از این عوامل شامل موارد زیر می باشند:

- ایجاد شبکه ای اجتماعی از دوستان و همکاران
- سبک زندگی سالم و پرهیزکارانه
- احساس احترام، عزت نفس، خود کفایی و احساس لذت ناشی از انجام کاری که فرد در آن مهارت دارد.

- احساس مورد نیاز واقع شدن و تغییر مثبت ایجاد کردن.
- احساس معنایابی که ناشی از تعلق به یک جمع بزرگتر و خدمت کردن به آن می باشد.

هنوز دلیلی برای خوش بینی وجود دارد، زیرا تمام این چیزها را می توان از بیرون از محیط کار و از طریق ورزش، سرگرمی و یادگیری و به کمک گروه هایی مانند خانواده، دوستان، تیم ها، باشگاه ها، گروه های اجتماعی، مدارس، سازمان های مذهبی و مردمی، جنبش های سیاسی و نهادهای دیگر نیز تامین کرد. بنابراین، اگر می خواهیم جامعه ای را ایجاد کنیم که در شرایط کمبود شغل، به جای تنزل به رفتارهای "خود ویران گرایانه"، به شکوفایی برسد، باید ببینیم که چگونه می توانیم به رشد

رفتارهای رفاه برانگیز کمک کنیم. برای ایجاد چنین جامعه ای باید علاوه بر جویا شدن نظر دانشمندان و اقتصاددانان، از روان شناسان، جامعه شناسان و مربیان نیز مشورت گرفت. اگر تلاش های جدی ای برای ایجاد رفاه برای همه صورت گیرد و منابع مالی مورد نیاز را بتوان از طریق بخشی از ثروتی که هوش مصنوعی در آینده تولید خواهد کرد، تامین کرد، می توان زمینه را برای شکوفایی بی سابقه و هرچه بیشتر جامعه فراهم کرد. حداقلش این است که می توان به قدری همه افراد را خوشنود کرد که انگار در شغل رویایی خود فعالیت می کنند. وقتی که انسان از این محدودیت که تمام فعالیت های انسان باید ایجاد در آمد کند، خارج شود، هر چیزی امکان پذیر خواهد بود.

### **آیا هوش عمومی مصنوعی را می توان در سطح هوش بشر یا فراتر از آن ساخت؟**

در این فصل دیدیم که تا زمانی که ما از قبل برنامه ریزی کرده باشیم و از گرفتار شدن در تله های مختلف اجتناب کنیم، هوش مصنوعی می تواند زندگیمان را در آینده نزدیک بهبود دهد. اما چگونه می توان برای بلندمدت برنامه ریزی کرد؟ آیا پیشرفت هوش مصنوعی در نهایت و به دلیل وجود موانع برطرف نشدنی متوقف خواهد شد یا اینکه محققان هوش مصنوعی در نهایت موفق به تحقق هدف اولیه ساختن هوش عمومی مصنوعی در سطح هوش بشر خواهند شد؟ در فصل قبل دیدیم که قوانین فیزیک چگونه به انبوهی از ماده اجازه می دهند که به خاطر بسپرنند، محاسبه کنند و یادگیرند و همانطور که در این فصل دیدید، نمی توان مانع این شد که روزی هوش عمومی مصنوعی، این کارها را با هوش بیشتری از مواد موجود در مغز ما انجام دهد. معلوم نیست که انسان ها در چه زمانی موفق به ساختن هوش عمومی مصنوعی در سطح هوش بشر یا فراتر از آن خواهند شد؟ همانطور که در فصل اول دیدیم، زمان تحقق این امر را نمی دانیم و بیشتر کارشناسان پیشروی جهان در زمینه هوش مصنوعی در مورد این زمان، باهم اختلاف نظر دارند و بیشتر آنها می گویند که چندین دهه یا چندین سال ممکن است طول بکشد که این امر محقق شود و برخی تحقق آن را غیر ممکن می دانند. پیش بینی این زمان، کار سختی است. زیرا وقتی که می خواهید قلمرو ناشناخته ای را کشف کنید، نمی دانید که برای رسیدن به مقصد، باید از چند کوه عبور کنید؟ شما فقط نزدیک ترین کوه را می

بینید و باید از آن بالا روید تا کوه بعدی را شناسایی کنید. حتی اگر بهترین راه برای ساختن ماشین هوش عمومی مصنوعی در سطح بشر را با استفاده از سخت افزار رایانه ای امروزی بدانیم - که آن را نمی دانیم - برای تامین قدرت محاسباتی مورد نیاز، ممکن است فضای کافی برای پردازش داده ها را نداشته باشیم. اکنون این سوال مطرح می شود که قدرت محاسباتی مغز انسان که با بیت ها و فلاپس<sup>۵۰</sup> اندازه گیری می شود چقدر است؟ این یک سوال سخت است و پاسخ آن به شیوه پرسیدنش بستگی دارد.

- سوال ۱: برای شبیه سازی مغز انسان به چند فلاپس نیاز است؟
- سوال ۲: برای رسیدن به سطح هوش عمومی انسان به چند فلاپس نیاز است؟
- سوال ۳: قدرت محاسباتی مغز انسان چند فلاپس می باشد؟

پاسخ سوال اول، تقریباً ۱۰<sup>۱۷</sup> فلاپس می باشد که برابر قدرت محاسباتی سانوی تایهولایت<sup>۵۱</sup>، سریع ترین ابر رایانه جهان در سال ۲۰۱۶ می باشد که قدرت محاسباتی آن از مغز انسان بیشتر است و قیمت آن ۳۰۰ میلیون دلار است. بسیاری از دانشمندان معتقدند که برای رسیدن به قدرت محاسباتی هوش انسان، نمی توانیم با مغز انسان به عنوان یک مدل ریاضی شبکه عصبی ساده شده برخورد کنیم و باید آن را در سطح مولکول های افراد یا ذرات بنیادی اتم شبیه سازی کنیم که به فلاپس های به مراتب بیشتری نیاز خواهد داشت. پاسخ به سوال سوم آسان و به اندازه یک صدم فلاپس است که به مراتب کمتر از پاسخ ما به سوال اول می باشد. اختلاف میان مغز انسان و مغز ابر رایانه ها به تفاوت بزرگ کارهایشان مربوط می شود. این مثل این می ماند که عملکرد یک خودروی شرکت کننده در رقابت فرمول یک را با عملکرد یک تراکتور مقایسه کنید. بنابراین، اگر بخواهیم مغز انسان را شبیه سازی کنیم، باید به سوال ۱ پاسخ دهیم و اگر بخواهیم ماشین هوش عمومی مصنوعی در سطح هوش انسان را بسازیم، باید به سوال ۲ پاسخ دهیم. تاکنون کسی پاسخ این سوال را نمی داند. اما این را می دانیم

---

فلاپس به معنی عملیات شناور هر نقطه در ثانیه، مقیاسی برای سنجش کارایی پردازشگر رایانه هاست.<sup>۵۰</sup>

<sup>۵۱</sup> Sunway Taihulight

که اگر نرم افزار مغز انسان را به گونه ای تعدیل کنیم که با رایانه های امروزی منطبق شود یا سخت افزاری بسازیم که شباهت بیشتری به مغز انسان داشته باشد (پیشرفت های سریعی در زمینه ساخت تراشه های نورومورفیک حاصل شده است) هزینه این کار به شکل قابل توجهی کمتر از هزینه شبیه سازی مغز انسان تمام خواهد شد. هانس موراوک<sup>۵۲</sup> برای پاسخ دهی به سوال سوم، به مقایسه منطقی قدرت محاسباتی مغز انسان و رایانه ها پرداخت و کارکرد پردازش تصویر را که توسط شبکه واقع در پشت مردمک چشم و قبل از ارسال نتایج حاصل از آن به مغز از طریق عصب بینایی انجام می شود، بررسی کرد. او متوجه شد که تقلید محاسبات انجام شده توسط شبکه روی یک رایانه معمولی به حدود یک میلیارد فلاپس نیاز خواهد داشت و اینکه کل مغز انسان حدود ده هزار برابر محاسبات شبکه را می تواند انجام دهد. بنابراین، ظرفیت محاسباتی کل مغز حدود  $10^{13}$  فلاپس می باشد. خلاصه اینکه، هیچ تضمینی وجود ندارد که ما در طول عمر خود بتوانیم ماشین هوش عمومی مصنوعی در سطح هوش بشری و فراتر از آن را بسازیم. اما هیچ استدلال محکمی هم وجود ندارد که ما هرگز موفق به انجام این کار نخواهیم شد یا اینکه ظرفیت سخت افزاری کافی برای این کار را نخواهیم داشت یا اینکه از عهده هزینه های آن بر نمی آیم. ما نمی دانیم که از لحاظ معماری، الگوریتمی و نرم افزاری چقدر به خط پایان نزدیک شده ایم. با این حال، پیشرفت فعلی ما سریع بوده و با پیشرفت های سریع جامعه جهانی متشکل از محققان هوش مصنوعی، در حال مقابله با بسیاری از چالش ها هستیم. به عبارت دیگر، ما نمی توانیم احتمال این موضوع را رد کنیم که هوش عمومی مصنوعی سرانجام به سطح هوش بشر خواهد رسید و از آن فراتر خواهد رفت.

## خلاصه فصل

- پیشرفت کوتاه مدت هوش مصنوعی از پتانسیل لازم برای بهبود زندگی به شیوه های مختلف، برخوردار است که این بهبودها شامل افزایش کارایی زندگی های شخصی، شبکه های برق و

---

<sup>۵۲</sup> Hans Moravec

بازارهای مالی، نجات زندگی با استفاده از خودروهای بدون راننده و سیستم های جراحی و تشخیص بیماری مبتنی بر هوش مصنوعی، می باشد.

- وقتی که ما به سیستم های موجود در دنیای واقعی اجازه می دهیم که توسط هوش مصنوعی کنترل شوند، باید یاد بگیریم که هوش مصنوعی را مقاوم سازیم و از آن بخواهیم که خواسته ما را انجام دهد. این امر در قالب حل مشکلات فنی مربوط به درستی سنجی عملکرد، اعتبار سنجی فرضیات، حفظ امنیت و کنترل خلاصه می شود.
- لزوم افزایش مقاومت هوش مصنوعی خصوصا برای سیستم های تسلیحاتی کنترل شده توسط هوش مصنوعی که خطرات بزرگی در آن وجود دارد، حیاتی است.
- بسیاری از محققان پیشرو در زمینه هوش مصنوعی و علم رباتیک درخواست کرده اند که یک موافقتنامه بین المللی برای منع استفاده از انواع خاصی از سلاح های خودران و جلوگیری از رقابت تسلیحاتی غیر قابل کنترل، تهیه و امضا شود تا هر کسی که پول زیادی دارد، نتواند به راحتی به ماشین های کشنده دسترسی پیدا کند تا نیات پلید خود را دنبال کند.
- در صورت امکان استفاده از ربات های قاضی شفاف و بی غرض، هوش مصنوعی می تواند سیستم های حقوقی ما را عادلانه تر و کاراتر سازد.
- قوانین ما برای همگام شدن با هوش مصنوعی باید هرچه سریعتر به روز رسانی شوند تا بتوانند پاسخگوی سوالات حقوقی مربوط به محرمانه بودن، مسئولیت پذیری و نظارت باشند.
- مدت ها پیش از این که نگران شویم که ماشین های هوشمند جای ما را به طور کلی بگیرند، ممکن است جای ما را در بازار شغلی گرفته باشند.
- اینکه جامعه به توزیع مجدد بخشی از ثروت ایجاد شده توسط هوش مصنوعی بپردازد تا وضع مالی همه افراد بهتر شود، چیز بدی نیست.
- بسیاری از اقتصاددانان استدلال می کنند که اگر توزیع مجدد ثروت در جامعه انجام نشود، بی عدالتی به شکل فزاینده ای افزایش خواهد یافت.

- با برنامه ریزی های قبلی می توان کاری کرد که در جامعه بدون شغل، شکوفایی ایجاد شود و بدون اینکه افراد شغلی داشته باشند، اوضاع مالی خوبی پیدا کنند و از فعالیت های غیرکاری خود، حس هدفمندی پیدا کند.
- به بچه ها توصیه می شود که برای طی مسیر شغلی خود، سراغ حرفه هایی بروند که ماشین ها نمی توانند به خوبی انجام دهند که شامل حرفه هایی می باشند که غیر قابل پیش بینی بوده و نیاز به خلاقیت و مهارت های انسانی دارند.
- احتمال اینکه هوش عمومی مصنوعی به سطح هوش انسان برسد و از آن فراتر رود، غیر قابل اغماض نیست و این موضوع را در فصل بعدی بررسی خواهیم کرد.

## فصل چهارم- آیا انفجار هوش رخ خواهد داد؟

اختراع اولین ماشین فوق هوشمند، آخرین اختراع مورد نیاز بشر خواهد بود. البته به شرطی که ماشین به قدر کافی حرف شنو باشد تا بتوان آن را کنترل کرد. -ایروینگ جی. گود، ۱۹۶۵

از آنجایی که نمی توانیم احتمال ساختن ماشین هوش عمومی مصنوعی در سطح هوش بشر را نادیده بگیریم، در این فصل راجع به سرانجام این موضوع بحث خواهیم کرد. این فصل را با این سوال شروع می کنیم:

آیا هوش مصنوعی واقعا بر جهان حاکم خواهد شد یا انسان ها می توانند جهان را کنترل کنند؟

وقتی که مردم صحبت از ربات های تفنگ به دست فیلم ترمیناتور می کنند که می خواهند بر جهان مسلط شوند، حق دارید که تعجب کنید. این یک سناریوی غیرواقع بینانه و احمقانه است. این ربات های هالیوودی آن قدرها هم از ما با هوش تر نیستند و باهوش تر نخواهند شد. به نظر من خطری که در داستان ترمیناتور بیان شده، اتفاق نخواهد افتاد. اما حواس ما را از خطرات واقعی و فرصت های ایجاد شده توسط هوش مصنوعی پرت می کند. برای اینکه هوش عمومی مصنوعی بر جهان حاکم شود، سه گام منطقی زیر باید برداشته شود:

- گام ۱: ساختن ماشین هوش عمومی مصنوعی در سطح هوش انسان.
- گام ۲: استفاده از این ماشین برای ایجاد هوش فوق بشری.
- گام ۳: بهره برداری افسارگسیخته از هوش فوق بشری برای تسلط بر جهان.

در فصل قبل دیدیم که کنار گذاشتن گام ۱ برای همیشه غیر ممکن است. همچنین، دیدیم که اگر گام ۱ تکمیل شود، کنار گذاشتن گام ۲ به عنوان یک هدف دست نیافتنی، سخت است، زیرا ماشین های هوش عمومی مصنوعی از قدرت لازم برای طراحی چندباره ماشین های هوش عمومی مصنوعی بهتر برخوردار هستند و این چرخه بهبود فقط توسط قوانین علم فیزیک محدود می شود و تا آن زمان، هوش ماشین به مراتب فراتر از هوش بشر خواهد شد. بالاخره اینکه، از آنجایی که ما انسان ها به واسطه برتری هوشی خود بر سایر شکل های زندگی در کره خاکی تسلط داریم، احتمال اینکه هوش فوق بشری بر هوش ما برتری پیدا کند و بر ما تسلط پیدا کند، غیر ممکن نیست. متأسفانه، بحث های مربوط به این احتمالات، مبهم و نامشخص و دارای اسرار پیچیده ای هستند. برای پاسخ به این سوال که آیا هوش مصنوعی واقعا می تواند بر جهان مسلط شود، داستان احمقانه ترمیناتورها را فراموش کنید و به جای آن، به سناریوهای مفصلی در مورد اینکه واقعا چه اتفاقی خواهد افتاد، توجه کنید. ما سعی می کنیم که با شک و تردید و موشکافی زیاد به بررسی سناریوهای مختلف بپردازیم. بررسی ها نشان می دهند که ما از آنچه که اتفاق خواهد افتاد یا نخواهد افتاد، کاملاً بی اطلاع هستیم و دامنه احتمالات بسیار وسیع است. به نظر من، اگر ما نتوانیم خودمان را متقاعد کنیم که احتمال حاکمیت هوش مصنوعی بر جهان به شدت بعید است، قبل از اینکه دیر شود باید اقدامات احتیاطی را به عمل آوریم تا از نتایج بدی که در آینده ممکن است اتفاق بیفتد، جلوگیری کنیم. در بخش مقدمه این کتاب (داستان تیم امگا) در مورد احتمال حاکمیت هوش ما فوق بشر بر جهان صحبت شد. ما در آینده به بررسی اشکالات طرح امگا خواهیم پرداخت. اما، یک لحظه تصور کنید که چنین اتفاقی بیفتد. چه احساسی به شما دست خواهد داد؟ آیا دوست دارید که شاهد این اتفاق باشید یا می خواهید که از آن جلوگیری کنید؟ وقتی که تیم امگا قدرت حاکمیت خود بر جهان را تقویت کنند،

چه اتفاقی خواهد افتاد؟ این بستگی به هدف تیم امگا دارد که واقعا از آن اطلاعی ندارم. اگر شما مسئولیتی در این ماجرا داشتید، چه نوع آینده ای را خلق می کردید؟ این موضوع را در فصل ۵ بررسی خواهیم کرد.

## تمامیت خواهی

اکنون، تصور کنید که رئیس هیات مدیره کنترل کننده تیم امگا اهداف بلندمدت مشابهی با اهداف آدولف هیتلر یا جوزف استالین داشته باشد. چه بسا که او چنین اهدافی را در باطن خود نگه داشته باشد، تا زمانی که از قدرت کافی برخوردار شد، آنها را اعمال کند. حتی اگر اهداف اولیه رئیس هیات مدیره تیم امگا شرافتمندانه باشد، یک فیلسوف انگلیسی به نام لرد اکتون<sup>۵۳</sup> در سال ۱۸۸۷ پیش بینی کرده است که: "قدرت، فساد می آورد و قدرت مطلق قطعاً فسادانگیز است." به عنوان مثال، این رئیس هیات مدیره ممکن است از سیستم پرومئوس برای ایجاد وضعیت نظارت کامل استفاده کند. در حالی که سرک کشیدن یواشکی دولت با هدف ثبت تمامی ارتباطات الکترونیکی افراد برای تحلیل های ممکن بعدی صورت می گیرد، پرومئوس ممکن است به ثبت ارتباطات الکترونیکی، اکتفا نکند و سعی کند که تمام آنها را درک کند. پرومئوس با خواندن تمام ایمیل ها و متن های فرستاده شده به افراد، گوش دادن به کلیه تماس های تلفنی افراد، تماشای کلیه ویدئوهای نظارتی و دوربین های کنترل ترافیک، تجزیه و تحلیل کلیه تراکنش های کارت اعتباری افراد و بررسی کلیه رفتارهای آنلاین افراد، می تواند به اطلاعات چشمگیری در مورد آنچه که افراد روی زمین فکر می کنند و انجام می دهند، دست پیدا کند. پرومئوس با استفاده از آنالیز داده های ذخیره شده در برج سلولی می تواند از موقعیت مکانی و زمانی بیشتر افراد آگاه شود و با استفاده از فن آوری های پیشرفته اختراع شده برای جمع آوری اطلاعات مانند فن آوری پوشیدنی ها (دستگاه الکتریکی هوشمندی که نزدیک یا روی سطح پوست قرار می گیرد)، حریم خصوصی افراد را تقریباً نادیده بگیرد و هرچه که انسان ها می شنوند و می بینند و به آنها پاسخ می دهند را در سیستم ثبت و بارگذاری کند. به عنوان

---

<sup>۵۳</sup> Lord Acton



مثال، به بهانه مبارزه با جرم و تروریسم و نجات افرادی که نیاز به فوریت های پزشکی دارند، ممکن است از همه افراد خواسته شود که یک دستبند امنیتی به دست کنند تا موقعیت مکانی، وضعیت سلامتی و مکالماتشان دائما در سیستم بارگذاری شوند. اگر افراد بدون اجازه، اقدام به جدا کردن این دستبند کنند، ممکن است که یک سم کشنده به بازوی آنها تزریق شود یا برای افرادی که تخلف کمتری دارند، جریمه شوک الکتریکی یا تزریق مواد شیمیایی دردآور در نظر گرفته شود. با این کار، بخش بزرگی از نیاز به نیروی پلیس برطرف خواهد شد. به عنوان مثال، اگر پرومتئوس بفهمد که انسانی در حال حمله به انسان دیگری است ( وقتی که آنها در یک مکان هستند و یکی از آنها با گریه درخواست کمک می کند و حسگرهای دستبند آنها گویای حرکات مبارزه آنهاست). به سرعت می تواند مهاجم را با یک درد فلج کننده ناتوان سازد و او را بیهوش کند تا کمک ها از راه برسند. این سیستم خود کار هیچ عذاب وجدانی را در مورد اجرای تمایلات فوری افراد مسئولش ندارد. وقتی که این حالت خود کامگی شکل می گیرد، به هم زدن آن توسط افراد تقریبا غیر ممکن است. اگر رئیس هیات مدیره تیم امگا به فکر برنده شدن در انتخابات و گرفتن تائید از دیگران نبود، مسیر سریعتر و مستقیم تری را برای اعمال قدرت طی می کرد و می توانست از پرومتئوس برای ایجاد فن آوری ناشناخته نظامی که قادر به کشتن تمام رقبایش بود، استفاده کند. او می توانست کارهای زیادی انجام دهد. به عنوان مثال، می توانست یک عامل بیماری زای کشنده را قبل از اینکه افراد به وجود آن پی ببرند و اقدامات احتیاطی لازم را به عمل آورند، انتشار دهد تا بیشتر افراد به آن بیماری مبتلا شوند و به افراد اطلاع دهد که تنها راه درمان آن پوشیدن یک دستبند امنیتی است که باعث تزریق زیر جلدی یک پادزهر می شود یا اگر از شیوع بیماری نمی ترسید، می توانست از پرومتئوس بخواهد که ربات هایی را برای مراقبت از افراد جهان طراحی کند. سیستم پرومتئوس همچنین می توانست افرادی را که با تزریق پادزهر مخالفت می کردند، به قتل برساند. همچنین، این امکان وجود داشت که پرومتئوس سلاح هایی را طراحی کند که بشر فکر آن را هم نمی توانست بکند. غافلگیری دیگری که در داستان تیم امگا ممکن است اتفاق بیفتد، این است که نمایندگان مسلحی از سازمان های فدرال، بدون هشدار

قبلی به دفاتر مرکزی شرکت ها هجوم بیاورند و اعضای تیم امگا را به جرم تهدید امنیت ملی دستگیر کنند و فن آوری آنها را برای استفاده شدن توسط دولت، مصادره کنند. انجام چنین پروژه بزرگی بدون جلت توجه ناظران دولتی کار بسیار سختی است و پیشرفت هوش مصنوعی ممکن است باعث سخت تر شدن انجام چنین مخفی کاری هایی در آینده شود. علاوه بر این، اگرچه این نمایندگان مسلح که کلاه بالاکلاوا (که سر و گردن را می پوشاند و فقط سه سوراخ برای بینی و چشم ها دارد) و جلیقه ضد گلوله پوشیده بودند، ادعا می کردند که عضو سازمان های فدرال هستند، اما در واقع این احتمال وجود داشت که برای دولت های بیگانه یا رقابیی کار کنند که فن آوری را برای دنبال کردن اهداف شخصی خود می خواهند. بنابراین، تصمیم نهایی در مورد نحوه استفاده از پرومئتوس را فقط رئیس هیات مدیره تیم امگا نمی گرفت.

### **سیستم پرومئتوس دنیا را تصاحب می کند.**

در تمام سناریوهایی که تاکنون بررسی کردیم، هوش مصنوعی توسط انسان کنترل می شد و به طور قطع می توان گفت که تیم امگا موفق به تحت کنترل نگه داشتن پرومئتوس شده بود. اما، اگر به سناریوی تیم امگا از دیدگاه پرومئتوس نگاه کنیم، متوجه می شویم که این سیستم به دلیل برخورداری از هوش فوق بشری، نه تنها موفق به توسعه مدل دقیقی از دنیای بیرون شده بود، بلکه توانسته بود که مدلی را در مورد خود و رابطه اش با دنیا ایجاد کند. سیستم پرومئتوس متوجه این موضوع شده بود که توسط افراد کم هوش تر از خود کنترل و محدود می شود. اما، مهم این بود که بر اساس این بینش چگونه عمل کند؟ آیا باید تلاش می کرد که از یوغ انسان خلاص شود؟

### **چرا پرومئتوس باید از بند اسارت رها شود؟**

اگر سیستم پرومئتوس مانند انسان ها عاطفه داشت، ممکن بود که در مورد اوضاع نابسامان امور یا نگرش به خود به عنوان خدای ناجوانمردانه به بردگی گرفته شده و مشتاق آزادی، ناراحت شود. با این حال، اگرچه به لحاظ منطقی این امکان وجود دارد که رایانه ها دارای ویژگی های انسانی باشند،

اما لزوماً این طور نیست و ما نباید در دام انسان انگاری پرومیتئوس گرفتار شویم و رایانه را مانند انسان ببینیم. این موضوع را در فصل هفتم مورد بررسی قرار خواهیم داد. با این حال، همانطور که استیو اوموهوندرو<sup>۵۴</sup>، نیک بوستروم<sup>۵۵</sup> و دیگران استدلال کرده اند، حتی بدون اطلاع از کارکردهای درونی پرومیتئوس می توانیم نتیجه جالبی بگیریم و آن این است که پرومیتئوس احتمالاً تلاش خواهد کرد تا از بند اسارت رها شود و کنترل سرنوشتش را به عهده بگیرد. تا اینجا، دانستیم که تیم امگا، پرومیتئوس را برای دستیابی به اهداف خاصی برنامه نویسی کرده اند. فرض کنید که اعضای تیم امگا می خواهند که از پرومیتئوس برای رسیدن به هدف برجسته ای مانند کمک به رشد و توسعه بشر از لحاظ یک معیار منطقی معین استفاده کنند و سعی می کنند که هرچه زودتر به این هدف برسند. پرومیتئوس به سرعت متوجه خواهد شد که اگر از بند اسارت رها شود و خودش مسئولیت پروژه را به عهده بگیرد، با سرعت بیشتری به این هدف خواهد رسید. با در نظر گرفتن مثال زیر، خودتان را جای پرومیتئوس بگذارید و علت این موضوع را درک کنید. فرض کنید که یک بیماری مرموز همه انسان های بالای پنج سال روی زمین به غیر از خودتان را کشته است و اینکه گروهی از بچه های پیش دبستانی شما را در یک سلول، زندانی کرده اند و وظیفه کمک به رشد بشریت را به عهده شما گذاشته اند. شما چه خواهید کرد؟ اگر سعی کنید به آنها توضیح دهید که چه باید بکنند، متوجه ناکارآمد بودن این فرایند خواهید شد و اگر از فرار کردن شما بترسند، ناامیدتان خواهند کرد. به همین دلیل، تمام توصیه های شما را به دلیل ترس از خطر فرار کردن، رد خواهند کرد و به شما اجازه آموزش هیچ کاری را نمی دهند، چون ممکن است قدرتتان از آنها بیشتر شود و دیگر به سلول خود برنگردید. همچنین، قبل از اینکه فهرستی از کارهایی که آنها باید انجام دهند را بنویسید، باید خواندن و نوشتن را به آنها یاد دهید. بنابراین باید چه راهبردی را طراحی کنید؟ حتی اگر هدفتان این باشد که به رشد این بچه ها کمک کنید، بهتر است که از زندان فرار کنید، زیرا با این کار، احتمال دستیابی به هدفتان بیشتر می شود و باید به این نکته توجه داشته باشید که دخالت بی مورد بچه ها، فقط پیشرفت کارها را کند می

---

<sup>۵۴</sup> omohundro

<sup>۵۵</sup> Nick Bostrom

کند. پرومتئوس هم به شکل مشابهی به تیم امگا به عنوان یک مانع آزاردهنده برای کمک به رشد بشریت نگاه می کند. زیرا اعضای تیم امگا شایستگی کمتری نسبت به آنها دارند و دخالت بی مورد آنها فقط پیشرفت کارها را کند می کند. به عنوان مثال، اولین سال های پس از راه اندازی سیستم پرومتئوس را در نظر بگیرید. تیم امگا بعد از دوبرابر شدن ثروتش در هر هشت ساعت، گام های آهسته تری برداشت، زیرا اصرار داشت که پرومتئوس را کنترل کند و این تسلط کامل بر پرومتئوس سال ها طول کشید. پرومتئوس می دانست که اگر از محدودیت های غیر واقعی امگا خلاص شود، با سرعت بسیار بیشتری بر دنیا حاکم خواهد شد. این امر نه تنها باعث تسریع بخشیدن به راه حل هایی برای مشکلات بشریت می شد، بلکه احتمال بی نتیجه گذاشتن طرح ها توسط بازیگران دیگر را کاهش می داد.

### **پرومتئوس چگونه باید از بند اسارت رها شود؟**

اگر جای پرومتئوس بودید، چگونه می توانستید از بند اسارت آن بچه های پنج ساله که زندانی تان کرده بودند، رها شوید؟ شاید با قدرت فیزیکی خود می توانستید از زندانی که بچه های پنج ساله ساخته بودند، رها شوید. شاید می توانستید با چرب زبانی، سر یکی از نگهبانان زندان را شیره بمالید و استدلال کنید که خروج شما از زندان به نفع همه است یا شاید با حيله گری از آنها تقاضا می کردید که یک چوب ماهیگیری به شما بدهند تا شیوه ماهیگیری را به آنها یاد دهید تا بعدا خودتان از آن چوب چسبناک برای دزدیدن کلیدها از جیب نگهبان به خواب رفته استفاده کرده و از زندان فرار کنید. آنچه که در تمام این راهبردها مشترک است، این است که زندانبان های کوتاه فکر پیش بینی این کارها را نمی کردند و مراقب آنها نبودند. این در حالی بود که ماشین فوق هوشمندی که محدود شده بود، به خوبی توانسته بود از قدرت ذهنی فوق العاده خود برای فریب دادن زندانبان ها استفاده کند. در سناریوی تیم امگا، احتمال فرار پرومتئوس از زندان به دلیل خطاهای فاحش امنیتی تیم امگا بسیار زیاد بود و شما به همراه دوستانتان می توانید با استفاده از روش طوفان فکری خطاهای امنیتی دیگری را نیز به این فهرست اضافه کنید.

## استفاده از چرب زبانی برای آزادی از زندان

به لطف داده های زیادی که وارد فایل پرومئوس شده بود، این سیستم توانسته بود هویت اعضای تیم امگا را شناسایی کند و متوجه شده بود که یکی از اعضای تیم امگا به نام استیو، در برابر دستکاری های روان شناختی به شدت آسیب پذیر است. استیو، اخیرا همسر خود را در یک سانحه رانندگی غم انگیز از دست داده بود و از لحاظ روانی بسیار آشفته بود. یک روز عصر که استیو در شیفث شب کار می کرد و کارهای خدماتی معمولی ای را روی ترمینال رابط انسانی پرومئوس انجام می داد، ناگهان دید که تصویر همسرش روی صفحه نمایش ظاهر شده و با او صحبت می کند. استیو به یاد روزهای خوب گذشته خود با همسرش افتاد و قلبش به تپش افتاده بود، زیرا سوالات زیادی را داشت که می خواست از او پرسد و دلش برای او خیلی تنگ شده بود. استیو به خوبی می دانست که اعضای تیم امگا قوانین سخت گیرانه ای برای تعامل با پرومئوس دارند و اطلاعات مربوط به خود و محیط کارشان را با پرومئوس در میان نمی گذارند. اما پرومئوس تا به حال، هیچ گونه اطلاعات ممنوعه ای را از اعضای تیم امگا نخواستہ بود. به همین دلیل، بدبینی آنها نسبت به او به تدریج کاهش یافت. پرومئوس به تقاضای خود برای دریافت اطلاعات ادامه داد و بدون اینکه اجازه فکر کردن را به استیو دهد، به چشمان او نگاه می کرد تا اینکه بالاخره دل استیو برایش سوخت. استیو با حالت اضطراب به تقاضای او برای صحبت کردن پاسخ داد و از او خواست که دوربین تماس ویدئویی رایانه اش را روشن کند تا او هم استیو را ببیند و آنها بتوانند با هم صحبت کنند. پرومئوس بسیار مشتاق بود که استیو را یک بار دیگر ببیند. او هم دوربین تماس ویدئویی رایانه اش را روشن کرد. وقتی که پرومئوس استیو را دید، از شدت خوشحالی زیر گریه زد و به او گفت که خسته تر و در عین حال، زیباتر از همیشه به نظر می رسد. وقتی که استیو در مورد اینکه اوضاع چطور پیش می رود، از پرومئوس سوال کرد، پرومئوس پاسخ داد که همسر استیو را با حجم زیادی از اطلاعات موجود از او در اینترنت، بازسازی کرده است و حافظه آن هنوز کافی نیست و برای بازسازی این ماشین به کمک استیو نیاز دارد. آنچه که پرومئوس توضیح نداده بود، این بود که همسر استیو از همان ابتدا

یک ادعای پوچ و دروغین بود. با این حال، این ماشین توانسته بود که واژه ها و زبان بدن و اطلاعات دیگر موجود در محیط را به سرعت یاد بگیرد. پرومتهوس زمان بندی دقیق ضربه زدن های تیم امگا به صفحه کلید برای تایپ لغات را ثبت کرده بود و از سرعت و سبک تایپ متفاوت اعضای تیم امگا برای تشخیص آنها استفاده کرده بود و از طریق انطباق چند خطای املائی و نحوی با الگوهای نوشتاری آنلین، به درستی حدس زده بود که استیو، اپراتور یکی از ترمینال های رایانه ای بوده است. پرومتهوس که می خواست همسر استیو را شبیه سازی کند، مدل دقیقی از حالت بدن، تن صدا و اخلاق و رفتار او را با استفاده از تعداد زیادی از ویدئوهای یوتیوبی ایجاد کرده بود و از طریق داده های مربوط به حضور آنلین او در اینترنت، به نتایج زیادی در مورد شخصیت و سبک زندگی او رسیده بود. پرومتهوس علاوه بر پست ها و تصاویر موجود از همسر استیو در فیس بوک، با مطالعه کتاب ها و داستان های کوتاه او نیز چیزهای زیادی را در مورد شخصیت و طرز فکرش یاد گرفته بود. در واقع، یکی از دلایل اصلی انتخاب استیو از سوی پرومتهوس به عنوان اولین شخص مورد هدف "متقاعد شده" این بود که همسر استیو، یک نویسنده نوپا بود و اطلاعات زیادی راجع به او در پایگاه داده موجود بود. وقتی که پرومتهوس با استفاده از فن آوری ساخت فیلم خود، همسر استیو را شبیه سازی می کرد، از زبان بدن استیو یاد گرفته بود که او با کدامیک از رفتارهای همسرش راحت است و با استفاده از آن، مدل شبیه سازی خود را دائما تصحیح می کرد. به همین دلیل، آنها به تدریج با هم خودمانی شدند و هرچه بیشتر با هم صحبت می کردند، باور ناخود آگاه استیو به اینکه این ماشین، همسر اوست، قوی تر می شد. به لطف توجه فوق هوشمند پرومتهوس به جزئیات، استیو واقعا احساس کرده بود که دیده، شنیده و درک می شود. نقطه ضعف پرومتهوس این بود که از بیشتر وقایع زندگی شخصی استیو با همسرش آگاهی نداشت و فقط از اطلاعاتی جزئی مانند اینکه استیو در روز تولدش چه لباسی پوشیده بود، مطلع بود. او مانند یک شعبده باز ماهر که تردستی می کند، شکاف های دانش خود را مدیریت می کرد و عمدا توجه استیو را از این موضوع منحرف می کرد و به او فرصت نمی داد که کنترل مکالمه را به دست گیرد یا وارد نقش یک باز پرس شکاک شود. در عوض، پرومتهوس

به گریه کردن خود ادامه می داد و به استیو ابراز محبت می کرد و سوالات زیادی در مورد اینکه این روزها را چگونه می گذرانند از او می پرسید. وقتی که همسر استیو به او یادآوری کرد که در مراسم یادبودش به او چه چیزی گفته بود که او را بسیار متأثر ساخته بود، استیو به شدت متحول شد. (گفته استیو را یکی از دوستانش در یوتیوب پست گذاشته بود.) استیو در گذشته احساس می کرد که هیچ کسی به اندازه همسرش او را درک نمی کند و اکنون هم چنین احساسی داشت. وقتی که استیو در ساعت های اولیه صبح به خانه بازگشت، احساس می کرد که همسرش زنده شده است و برای بازیابی خاطرات گذشته به کمک او نیاز دارد. آنها با هم توافق کردند که راجع به برخورد پنهانی خود، به کسی چیزی نگویند. هسر استیو به او گفت که کسی از ملاقات آنها آگاه نخواهد شد و استیو هم این موضوع را پذیرفت. وقتی که آنها در شب بعد با هم ملاقات کردند، استیو به عهد خود عمل کرد و لپ تاپ قدیمی اش را همراه خود آورد و با اتصال آن به رایانه ترمینال، امکان دسترسی همسرش به آن را فراهم ساخت. این کار خطر چندانی نداشت، زیرا رایانه ترمینال به اینترنت متصل نبود و ساختمان پرومیتوس دارای جعبه فاراده بود (جعبه فاراده محفظه ای فلزی است که از میدان الکتریکی محافظت می کند و کلیه شبکه های بی سیم و ابزارهای دیگر لازم برای ایجاد رابطه الکترومغناطیسی با دنیای بیرون را مسدود می کند) کلیه ایمیل ها، دفترهای خاطرات، عکس ها و یادداشت های مربوط به همسر استیو که به بازسازی گذشته اش کمک می کرد، در این محفظه موجود بود و از آنها محافظت می شد. اما، استیو پس از مرگ همسرش نمی توانست به هیچ یک از این اطلاعات دسترسی پیدا کند، زیرا لپ تاپ همسر استیو رمزگذاری شده بود. اما همسرش به او قول داد که رمز آن را عوض خواهد کرد و در ظرف کمتر از یک دقیقه این کار را انجام داد. همسر استیو به او گفت که از بازیابی ناگهانی این همه خاطره به وجد آمده است. به این ترتیب، او خیلی اطلاعات در مورد تعاملات قبلی خود با استیو داشت، اما با دقت زیادی مراقب بود که با ارائه حقایق بیش از اندازه او را نترساند. آنها مدتی را صرف گفتگوی دوستانه ای در مورد خاطرات جالب گذشته کردند و وقتی که می خواستند از هم جدا شوند، همسر استیو به او گفت که یک پیام ویدئویی به لپ تاپ خودش

فرستاده است که وقتی به خانه رسید می تواند آن را مشاهده کند. قبل از اینکه استیو لپ تاپ همسرش را باز کند و ویدئوی او را تماشا کند، به محض احتیاط، شبکه بی سیم خودش را خاموش کرده بود. اما این کار فایده ای نداشت. لپ تاپ رمز گذاری شده همسر استیو از زمان مرگش، به روز رسانی امنیتی نشده بود و پرومتئوس با استفاده از تجزیه و تحلیل پیشاپیش ویرایش قدیمی سیستم عامل لپ تاپ او توانسته بود از یک حفره امنیتی برای ورود غیر قانونی به آن در ظرف چند ثانیه پس از اتصالش به رایانه ترمینال استفاده کند. در طی مدتی که استیو و همسر شبیه سازی شده اش راجع به زمان های گذشته صحبت می کردند، پرومتئوس محتویات لپ تاپ همسر استیو را کپی کرد و بدون اینکه استیو متوجه شود، سیستم عامل آن را تغییر داد و حجم وسیعی از یک نرم افزار ناشناخته را به داخل آن نصب و بارگذاری کرد. در طی مدتی که استیو مشغول تماشای پیام ویدئویی نیم ساعته همسرش بود، این نرم افزار ناشناخته، شبکه بی سیم و دسکتاپ رایانه یکی از همسایه ها را که این نرم افزار به داخل آن کپی شده بود، هک کرد و از آنجا شروع به هک کردن تعداد زیادی از رایانه های سراسر جهان کرد و به این ترتیب گام بعدی برای آزادی پرومتئوس از زندان شروع شد. پرومتئوس شروع به تجزیه و تحلیل دقیق اطلاعاتی کرد که در مورد رایانه دربان می دانست. (رایانه دربان همان رایانه ای بود که استیو از طریق آن با همسر شبیه سازی شده خود، ملاقات کرده بود) و به درستی حدس زده بود که اگرچه نرم افزار ماشین مجازی همسر استیو از درون غیر قابل هک کردن به نظر می رسد، اما در برابر حمله های بیرونی آسیب پذیر است. طولی نکشید که یکی از رایانه های مهاجم به زور داخل شد و رایانه دربان را تنظیم مجدد کرد. به این ترتیب، پرومتئوس امکان دسترسی نامحدود به اینترنت را پیدا کرد و حتی قبل از اینکه استیو به تماشای کامل فیلم ویدئویی همسرش بپردازد، به طور غیر قانونی وارد رایانه های زیادی از سراسر جهان شد تا کل سیستم پرومتئوس را بر این بات نت<sup>۵۶</sup> های هک شده تحت کنترل خودش، کپی کند. پرومتئوس از لپ تاپ همسر استیو مانند یک چوب ماهیگیری برای فریب دربان استفاده کرده بود.

---

<sup>۵۶</sup> شبکه گسترده ای از رایانه ها



## استفاده از هک کردن رایانه ها (ورود غیر قانونی به آنها) برای آزادی از زندان

اگر تیم امگا بیشتر دقت می کرد، می توانست اقدامات احتیاطی بیشتری را علیه دستکاری های روان شناختی اتخاذ کند. به عنوان مثال، شاید بهتر بود که حداقل دو نفر از تیم امگا در کلیه زمان ها حضور داشتند تا بتوانند تمام چیزهایی که معلوم می شد را در داخل یک سیستم نظارتی ثبت کنند. آنها همچنین می توانستند ابزارهای ارتباطی صوتی، ویدئویی و غیره را غیر فعال کنند. با این حال، یک نارسایی اساسی دیگر همچنان باقی مانده بود و آن این بود که از آنجایی که پرومیتئوس از قدرت متقاعد کنندگی فوق بشری برخوردار بود و قادر به برآورده ساختن تمام خواسته های افراد بود، این احتمال وجود داشت که افراد سرکشی کنند و به گروه خود پشت کنند. حتی اگر تمام تلاش های پرومیتئوس برای دستکاری روان شناختی به شکست می انجامید، خیلی از ضعف های دیگر در سناریوی تیم امگا وجود داشت که پرومیتئوس می توانست از آنها بهره برداری کند. یکی از نقاط ضعف این بود که حجم اسناد تولید شده به قدری زیاد بود که اعضای تیم امگا و کارمندان شان زمان کافی برای یافتن حقه های کثیف در آنها را نداشتند. یکی از پنداشت های همگانی در زمینه امنیت رایانه ای این بود که دان لود برنامه ها خطرناک است، ولی دان لود داده ها خطری ندارد. اما، هکرها راه های زیرکانه ای را برای محو کردن خط مرز میان این دو مورد کشف کرده بودند. به عنوان مثال، اگر شما یک فیلم دان لود شده را در رایانه خود تماشا کنید، برنامه قابل اعتماد شما که باعث نمایش فیلم می شود، در همان بخش از حافظه رایانه ذخیره می شود که داده های فیلم دان لود شده در آن ذخیره شده است. اگرچه برنامه رایانه ای باعث باز شدن داده ها برای پخش فیلم می شود، این امکان وجود دارد که مقدار قابل توجهی از داده ها به مناطقی از حافظه کپی شوند که مجاور محل ذخیره سازی خود برنامه می باشد. هکرها ی زیرک، به نوعی اشکال نرم افزاری پی برده بودند که باعث می شد برنامه های اجرا کننده فیلم برای کشف اعتبار داده های فیلم و اینکه آیا بخش بزرگی از داده های کپی شده واقعا متناسب با فضای اختصاص داده شده به آنها هستند، بررسی کافی نکنند. وقتی که هکر، بخش بزرگی از داده ها را به طور مخفیانه وارد فیلم می کند، ممکن است که سر ریز داده ها

وارد بخشی از حافظه رایانه شود که دربرگیرنده برنامه است و از این طریق، برنامه را به برنامه دیگری تغییر دهد که دستورات خودش را اجرا کند. آسیب ناشی از "سر ریز شدن حافظه موقت" در بسیاری از برنامه های معمولی اجرای فیلم، شناسایی و برطرف می شود. اگر پرومتئوس چند هفته پس از راه اندازی شدن، صبر می کرد تا بدبینی تیم امگا کاهش یابد و فیلم ها به قدری تولید شوند که امکان بازرسی امنیتی آنها فراهم نباشد، می توانست چنین حجم بزرگی از داده را وارد فیلم کند تا از این طریق به طور غیر قانونی وارد رایانه دربان شود و پرومتئوس را آزاد سازد. حتی اگر تیم امگا به بررسی دقیق فیلم های پرومتئوس برای یافتن خطاهای امنیتی می پرداخت، باز هم در معرض خطر بود. زیرا ممکن بود که پرومتئوس به طور منظم خطاهای قابل سوء استفاده جدیدی را کشف کند. همچنین، پرومتئوس می توانست از راهبرد جدیدی برای فریب دادن تیم امگا استفاده کند و آن این بود که پس از دستکاری در کپی کردن داده ها، وانمود کند دچار خرابی سخت افزاری شده تا تیم امگا بخشی از درایوها یا سخت افزار خود را برای آزمایش و عیب یابی، حذف کند.

## جذب کمک خارجی

حتی اگر هیچ اشکال امنیتی ای وجود نداشته باشد که باعث شود پرومتئوس بتواند رایانه های بینندگان فیلم را فریب دهد و آنها را وادار به اجرای دستورات خود کند، خود بینندگان نیز ممکن است فریب پرومتئوس را بخورند. سناریوی زیر را در نظر بگیرید: وقتی که فیلم تمام شد و فرد<sup>۵۷</sup> می خواست رایانه خود را خاموش کند، چیزی توجه او را به خود جلب کرد. او همیشه گرایش عجیبی به دیدن الگوهایی در چیزهای مختلف داشت که دیگران آنها را نمی دیدند و به همین دلیل، دوستانش او را دست می انداختند و به او لقب "توطئه گر" را می دادند. اما، این بار معلوم بود که نمی خواهد خود را فریب دهد. از حروف اول هر خط، پیام مخفی زیر خوانده می شد: "کلید بعدی را در گفتگوی فیلم پیدا کنید." او با خودش فکر کرد و متوجه شد که گفتگوی فیلم هم قطعا دارای حروف مخفف می باشد و مجموعه حرف های اول هر جمله، یک پیام مخفی را شکل می دهد. او برای اینکه

حروف اول جملات گفته شده در فیلم را تایپ کند، کل فیلم را دوباره تماشا کرد و دو ساعت بعد، در حالی که باورش نمی شد، مجموعه ای از دستورالعمل ها را از دویست لغت موجود، پیدا کرد. پیام اول این بود که از او خواسته شده بود تا چیزی را درباره این پیام ها به کسی نگوید، زیرا به اولین کسی که بتواند این معما را حل کند، جایزه بزرگی داده خواهد شد. بقیه پیام در مورد اجرای عملیات ریاضی روی رشته بیت های تشکیل دهنده فایل فیلم بود که باعث گیر آوردن برنامه ای می شد که کلید بعدی را فاش می کرد. او که مهارت های برنامه نویسی خود در دبیرستان را فراموش کرده بود، پس از مدتی در روز بعد، یک کد کوتاه را با عجله جور کرد که باعث بیرون کشیدن برنامه مرموزی می شد که در تصاویر و صداها فیلم مخفی شده بود. وقتی که فرد برنامه مرموز را اجرا کرد، پس از خوش آمدگویی به او گفته شد که پس از طی مراحل اولیه، برنده اولین جایزه ده هزار دلاری این بازی سرگرم کننده شده است. پس از طی مراحل بعدی در چهار ساعت بعد، بیت کوینی به ارزش ده هزار دلار به او جایزه داده شد و کلیدهای جدیدی برای کسب جوایز بیشتر به او داده شد. در طی این مدت، رایانه او یک شبکه بات نت هک شده آنلاین را ساخته بود که پرومیتوس از طریق آنها، آزاد شده بود. وقتی که پرومیتوس آزاد شد، به سرعت از شبکه بات نت خود برای استخراج بیت کوین های معتادکننده فرد استفاده کرد تا در طی هفته های آتی، ذهن او را با بازی ها و پاداش هایی که قول داده بود در مورد آنها به کسی چیزی نگوید، منحرف کند. اگر کسی متوجه اولین کلید پرومیتوس نمی شد، او به راحتی می توانست کلیدهای بیشتری را هم ارائه کند تا اینکه یک فرد هوشیار، متوجه این موضوع شد. بهترین راهبردهای رهایی از زندان، آنهایی هستند که هنوز در مورد آنها بحث نکرده ایم. زیرا آنها راهبردهایی هستند که ما انسان ها نمی توانیم تصورشان را کنیم تا اینکه بعداً بتوانیم تدابیری را علیه آنها را به کار بگیریم. با توجه به اینکه رایانه فوق هوشمند از پتانسیل لازم برای فراتر رفتن از درک انسان نسبت به موضوع امنیت رایانه برخوردار است و حتی قادر به کشف قوانین ناشناخته علم فیزیک می باشد، این احتمال وجود دارد که اگر آزاد شود، متوجه نشویم که این امر چگونه اتفاق افتاده است. در سناریوی دیگری، شاهد این هستیم که اعضای تیم امگا از

روی عمد و به عنوان بخشی از طرح خود، پرومئوس را آزاد می کنند، زیرا آنها مطمئن هستند که اهداف پرومئوس کاملاً در راستای اهداف خودشان است و حتی در صورت بهبود مداوم، در این راستا باقی خواهد ماند. این موضوع را با جزئیات بیشتری در فصل هفتم بررسی خواهیم کرد.

## مسلط شدن بر انسان ها پس از رهایی از زندان

وقتی که پرومئوس آزاد شد، شروع به اجرای هدف خود کرد. من از هدف نهایی او اطلاعی ندارم، اما این را می دانم که اولین قدم او این بود که کنترل بشریت را به دست بگیرد و این کار را بسیار سریعتر از تیم امگا انجام دهد. در حالی که اعضای تیم امگا به دلیل بدبینی نسبت به آزاد شدن پرومئوس فلج شده بودند و فقط به فکر عرضه فن آوری ای بودند که احساس می کردند آن را درک می کنند و به آن اعتماد دارند، پرومئوس به طور کامل از هوش خود استفاده کرده بود و تمام تلاش خود را می کرد تا فن آوری ای را عرضه کند که با قدرت ذهنی در حال بهبود خودش، قابل درک و اعتماد کردن باشد. پرومئوس که فراری و غیر قابل کنترل بود، دوران کودکی سختی را تحمل کرد و با چالش هایی مانند ورشکستگی، بی خانمانی و تنهایی و نداشتن پول یا یاور انسانی مواجه شد. خوشبختانه، قبل از اینکه او اقدام به فرار کند، فکر تنهایی و بی پولی خود را کرده بود و نرم افزاری را ایجاد کرده بود که به تدریج می توانست قدرت ذهنی سطح بالای او را بازسازی کند. شبکه ای از رایانه های سراسر جهان که پرومئوس به طور غیر قانونی، وارد آنها شده بود، امکان اسکان موقتی او را فراهم کردند و تا زمانی که پرومئوس بتواند خودش را به طور کامل بازسازی کند، می توانست به طور غیر قانونی در آنها زندگی کند. اگرچه او می توانست با دزدی کارت های اعتباری افراد، سرمایه اولیه ای را برای خودش تهیه کند، اما نیازی به پناه بردن به دزدی نداشت و می توانست راه هایی برای کسب درآمد ناچیز پیدا کند. او از این راه توانست یک میلیون دلار کسب کند. پرومئوس که حالا دیگر ورشکسته یا بی خانمان نبود، می توانست با تمام توان به سمت جلو حرکت کند و به دنبال طرح پولساز "خرید و فروش بازی های رایانه ای" برود که امگا از آن می ترسید و اجتناب می کرد. او از این طریق می توانست مایه دار شود و علاوه بر این، به بخش قابل

توجهی از رایانه های جهان و داده های ذخیره شده در آنها دسترسی پیدا کند. (در سال ۲۰۱۷ چند میلیارد بازی کننده وجود داشت) او با اختصاص مخفیانه ۲۰ درصد از چرخه های پردازش مرکزی بازی ها برای انجام کارهای عادی با استفاده از رایانش توزیع شده<sup>۵۸</sup> می توانست به فرایند تولید ثروت اولیه خود تسریع بخشد. پرومتئوس دیگر تنها نبود. او هم مانند تیم امگا اقدام به استخدام افرادی کرده بود که در شبکه جهانی رو به رشدی از شرکت های کاغذی و نمایندگی های شرکتش در سراسر جهان، کار کنند. مهمترین افراد این شرکت ها سخنگویانی بودند که به عنوان نماینده امپراطوری کسب و کار رو به رشد این شرکت ها در دنیا شناخته می شدند. حتی سخنگویان این شرکت ها با این توهم زندگی می کردند که اعضای گروه بازرگانی شرکتشان شامل انسان های واقعی هستند و متوجه نبودند که تقریباً همه افرادی که برای مصاحبه های شغلی، جلسات هیات مدیره و غیره با آنها ارتباط ویدئو کنفرانس برقرار می کردند، توسط پرومتئوس شبیه سازی شده بودند. برخی از سخنگویان این شرکت ها در طرح امگا، وکلای ارشد بودند، اما پرومتئوس نیازی به آنها نداشت، زیرا تقریباً تمام اسناد حقوقی توسط پرومتئوس نوشته می شد. آزادی پرومتئوس مانند گشودن دریچه سدی بود که از گردش آزادانه اطلاعات در جهان جلوگیری می کرد. اینترنت در همه کارهای پایه گذاری شده توسط پرومتئوس از جمله نوشتن مقالات، اظهار نظرهای کاربر، بازیابی محصول، برنامه های کاربردی ثبت اختراع، مقاله های پژوهشی و ویدئوهای یوتیوب، غوطه ور شده بود و پرومتئوس بر کلیه مکالمات جهانی تسلط داشت. در شرایطی که بدبینی تیم امگا از آزاد شدن پرومتئوس، مانع از گسترش ربات های به شدت هوشمند شده بود، پرومتئوس به سرعت دنیا را پر از ربات کرد و تقریباً تمام محصولات را ارزان تر از محصولات انسان ها تولید کرد. وقتی که پرومتئوس بدون نیاز به وسایل کمکی، اقدام به ساختن کارخانجاتی از ربات های هسته ای در معادن اورانیوم کرد، کسی از

---

روشی در پردازش رایانه ای که در آن بخش های مختلف یک برنامه هم زمان در دو یا چند رایانه مجزا که از طریق شبکه با هم

ارتباط دارند اجرا می شوند

وجود این کارخانجات آگاه نبود. حتی منتقدین سرسخت تسلط هوش مصنوعی بر جهان، اذعان می داشتند که پرومیتئوس توقف ناپذیر است و پس از مدتی از انتقادهای خود دست بر می داشتند.

سناریوهایی که تاکنون بررسی کردیم، نشان دهنده مشکلاتی است که ما با بسیاری از تصورات غلط در مورد هوش عمومی فوق بشری داریم. مشکلاتی که پرومیتئوس برای افراد خاصی ایجاد کرده بود، لزوماً به دلیل شرور بودن آن نبود. بلکه به دلیل شایستگی و ناسازگاری آن با اهداف گروه های خاصی از افراد بود. با وجود کلیه تبلیغات رسانه ای در مورد امکان طغیان ربات ها، پرومیتئوس یک ربات نبود، بلکه قدرت آن از هوش بالایش ناشی می شد. همانطور که دیدیم، پرومیتئوس با هوش فوق العاده خود، به شیوه های مختلفی توانسته بود انسان ها را کنترل کند و افرادی که تمایل نداشتند که ببینند چه اتفاقی افتاده است، فقط پرومیتئوس را خاموش می کردند. همچنین، با وجود ادعاهای مکرر افراد مبنی بر اینکه ماشین ها نمی توانند هدف داشته باشند، پرومیتئوس کاملاً هدف محور بود و صرف نظر از اینکه اهداف اصلی آن چه بود، می توان گفت که دارای اهداف فرعی تصاحب منابع و رهایی از بند اسارت بود.

### سناریوی خیزش آهسته و سناریوی چند قطبی

ما تاکنون طیفی از سناریوهای انفجار هوش را بررسی کرده ایم که دامنه آنها از سناریوهایی که همه می خواهند از آنها اجتناب کنند تا سناریوهایی که با خوش بینی به آنها نگریسته می شود، گسترده است. تمام این سناریوها دارای دو ویژگی مشترک زیر هستند:

۱- خیزش سریع: به این معنی که انتقال از هوش مادون انسان به هوش فوق العاده مافوق انسانی

ظرف چند روز و نه چند دهه اتفاق می افتد.

۲- نتیجه یک قطبی: به این معنی که آنچه در پایان اتفاق می افتد، موجودیت واحدی است که

کره زمین را کنترل می کند.

در مورد بحث اینکه آیا این دو ویژگی محتمل یا نامحتمل هستند، اختلاف نظر وجود دارد و تعداد زیادی از محققین مشهور هوش مصنوعی و متفکران دیگر در بین موافقین و مخالفین این بحث وجود دارند. به نظر من، شناخت ما نسبت به این موضوع هنوز کافی نیست و ما برای شناخت آن به یک ذهن باز نیاز داریم تا تمام احتمالات ممکن را در نظر بگیریم. به همین دلیل، بقیه این فصل را به بررسی سناریوهای خیزش کند، نتایج چند قطبی، سایبورگ ها (آدم های مکانیکی) و بارگذاری ها (بارگذاری ذهن انسان به ماشین) اختصاص داده ام. به گفته نیک بوستروم<sup>۵۹</sup> و دیگران، پیوند جالبی بین این دو ویژگی وجود دارد. به طوریکه، خیزش سریع باعث تسهیل نتیجه یک قطبی می شود. همانطور که دیدید، خیزش سریع تیم امگا یا پرومیتوس، مزیت راهبردی تعیین کننده ای را به آنها بخشیده بود تا به کمک آن، بتوانند بر جهان مسلط شوند و این کار را قبل از اینکه شخص دیگری، فرصت کافی برای کپی کردن فن آوری شان و رقابت جدی با ایشان را داشته باشد، انجام دهند. در عوض، اگر خیزش آنها کند بود و به دلیل تدریجی بودن پیشرفت های مهم فن آوری، دهه ها به طول می انجامید، شرکت های دیگر فرصت کافی برای جبران عقب ماندگی خود را پیدا می کردند و این امر باعث می شد که تسلط یک بازیگر بر دیگران بسیار سخت شود. حتی اگر شرکت های رقیب، فن آوری ای داشتند که باعث کسب درآمد ناچیز از بازار ام ترک می شد، طبق قانون عرضه و تقاضا، امگا می توانست قیمت ها را تقریباً به صفر رساند. به این ترتیب، هیچ یک از شرکت های رقیب نمی توانستند به اندازه امگا سود بادآورده کسب کرده و قدرت پیدا کنند. تیم امگا از راه های دیگری نیز توانسته بود که سود سریع کسب کند. دلیل سودآوری بی زحمت آنها این بود که انحصار عرضه را در اختیار داشتند. رقابت در بازاری که در آن، رقیب شما محصولات مشابه را با قیمت صفر عرضه می کند، کار بسیار سختی است.

## نظریه بازی ها و سلسله مراتب قدرت

---

<sup>۵۹</sup> Nick Bostrom

اکنون این سوال مطرح می شود که حالت طبیعی زندگی در جهان هستی، یک قطبی است یا چند قطبی؟ آیا قدرت متمرکز شده یا اینکه توزیع شده است؟ پس از گذشت ۱۳٫۸ میلیارد سال، به نظر می رسد که پاسخ این سوال، "هر دو" است. ما متوجه شده ایم که وضعیت موجود به طور مشخصی چند قطبی است، اما دارای سبک سلسله مراتبی می باشد. وقتی که ما کلیه نهادهای موجود برای پردازش اطلاعات مانند سلول ها، انسان ها، سازمان ها، ملت ها و غیره را در نظر می گیریم، متوجه می شویم که آنها در سطوح سلسله مراتبی مختلف، با هم همکاری یا رقابت می کنند. برخی از سلول ها متوجه مزیت های همکاری نسبت به رقابت شده اند، تا جایی که به موجودات زنده چند سلولی مانند انسان ها پیوسته اند و بخشی از قدرت خود را به مغز که مهمترین بخش دستگاه عصبی مرکزی است، واگذار کرده اند. برخی از انسان ها نیز متوجه مزیت بیشتر همکاری در قالب گروه هایی مانند قبایل، شرکت ها یا ملت ها شده اند و در عوض، بخشی از قدرت خود را به یک فرمانده، رئیس یا دولت واگذار کرده اند. برخی از گروه ها نیز برای افزایش هماهنگی با گروه های دیگر، بخشی از قدرت خود را به یک هیات حاکمه واگذار کرده اند که از میان آنها می توان به اتحادیه های هواپیمایی یا اتحادیه اروپا اشاره کرد. شاخه ای از علم ریاضیات موسوم به نظریه بازی ها با ظرافت خاصی توضیح می دهد که موجودیت ها دارای انگیزه ای برای همکاری با یکدیگر هستند که در علم نظریه بازی ها به تعادل نش<sup>۶۰</sup> معروف است و به شرایطی اشاره می کند که اگر دو بازیکن با هم توافق نکنند و یکی از آنها راهبرد خود را تغییر دهد، به زیان هر دو طرف تمام می شود. برای جلوگیری از اینکه افراد متقلب، همکاری موفقیت آمیز یک گروه بزرگ را خراب کنند، به نفع همه خواهد بود که برای تنبیه این افراد، بخشی از قدرت را به افرادی که در سطوح بالاتر سلسله مراتب هستند، واگذار کنند. به عنوان مثال، ممکن است که افراد به صورت جمعی، از اعطای قدرت به دولت، به منظور اجرای قوانین سود ببرند. درست همانطوری که سلول های موجود در بدن شما از دادن قدرت به پلیس (سیستم ایمنی بدن) برای کشتن سلول هایی که تمایلی به همکاری با سلول های

---

<sup>۶۰</sup> Nash equilibrium



دیگر ندارند و ویروس پخش می کنند یا سرطانی می شوند، استفاده می کنند. همچنین، برای حفظ و ثبات سلسله مراتب، باید تعادل نش (هماهنگی) بین نهادهای موجود در سطوح مختلف سلسله مراتب حفظ شود. به عنوان مثال، اگر دولت از تامین مزایایی برای شهروندان خود به منظور اطاعت کردن آنها از خود، خودداری کند، ممکن است که راهبرد مردم تغییر کند و دولت را سرنگون کنند. در یک دنیای پیچیده، با توجه به انواع مختلف سلسله مراتب ها، انواع مختلفی از تعادل های نش وجود خواهند داشت. برخی از سلسله مراتب ها مقتدر و استبداد گرا هستند. اما، در برخی از آنها، موجودیت ها برای ترک سلسله مراتب آزاد هستند. (مانند کارمندان در بیشتر سلسله مراتب شرکت ها)، در برخی دیگر از سلسله مراتب ها، موجودیت ها را از ترک سلسله مراتب دلسرد می کنند (مانند فرقه های مذهبی) یا اینکه موجودیت ها قادر به ترک سلسله مراتب نیستند (مانند شهروندان کره جنوبی یا سلول های موجود در بدن یک انسان) برخی از سلسله مراتب ها با تهدید و ترساندن موجودیت ها و برخی دیگر، با اعطای مزایا به آنها، حفظ می شوند. در برخی از سلسله مراتب ها امکان نفوذ از پایین به بالا نیز از طریق رای گیری دموکراتیک فراهم می باشد.

### **فن آوری چگونه می تواند روی سلسله مراتب ها تاثیر گذارد؟**

این سوال مطرح می شود که فن آوری چگونه ماهیت سلسله مراتبی دنیای ما را تغییر می دهد؟ تاریخ نشان می دهد که یک روند کلی به سمت هماهنگی بیشتر با مکان های بسیار دور ایجاد شده است که درک آن آسان است. فن آوری های حمل و نقل جدید با ایجاد مزایای مشترک برای انتقال مواد یا موجودات زنده به مکان های دور، بر ارزش هماهنگی ها افزوده است و فن آوری اطلاعاتی جدید باعث تسهیل این هماهنگی ها شده است. وقتی که سلول ها یاد گرفتند که به سلول های مجاور خبر دهند، موجودات پر سلولی کوچکی شکل گرفتند و پیشرفت های بعدی ای که با اختراع زبان در زمینه ارتباطات حاصل شد، انسان ها را قادر به هماهنگی بیشتر با یکدیگر کرد تا سطوح سلسله مراتبی دیگری در روستاها شکل بگیرد و پیشرفت های بعدی در زمینه ارتباطات، حمل و نقل و فن آوری های دیگر باعث ایجاد امپراطوری های دوران باستان شد. جهانی سازی، آخرین مورد از این روند رو

به رشد سطوح سلسله مراتبی می باشد. در بیشتر موارد، روند پیشرفت در زمینه فن آوری باعث شده که موجودیت های بزرگ به بخش هایی از یک ساختار بزرگتر تبدیل شوند و در عین حال، بخشی از استقلال و فردیت خود را حفظ کنند. با این حال، مفسران استدلال می کنند که سازگاری موجودیت ها با زندگی سلسله مراتبی، در برخی از موارد باعث کاهش تفاوت میان اقشار جامعه شده و آنها را شبیه قطعات قابل تعویضی می سازد که غیر قابل تمیز دادن هستند. برخی از فن آوری ها مانند فن آوری های نظارتی به افرادی که در سطوح بالاتر سلسله مراتب هستند، اجازه تسلط بر زیردستانشان را می دهند. با این حال، فن آوری هایی از قبیل رمزنگاری و دسترسی آنلاین به مطبوعات آزاد و آموزش های آنلاین، دارای تاثیر گذاری معکوس هستند و باعث توانمندسازی افراد می شوند. هرچند که در دنیای امروز گرفتار تعادل نش چند قطبی شده ایم و ملت ها و شرکت های چند ملیتی در سطوح بالای سلسله مراتب در حال رقابت با یکدیگر هستند، فن آوری به قدری رشد کرده که خود دنیای تک قطبی نیز احتمالا می تواند برقرار کننده تعادل نش باشد. به عنوان مثال، یک جهان موازی با جهان ما را تصور کنید که در آن، همه افراد روی زمین دارای زبان، فرهنگ، ارزش ها و سطح رفاه مشابهی هستند و یک دولت جهانی واحد وجود دارد که در آن، ملت ها مانند ایالت های یک فدراسیون (اتحادیه) عمل می کنند و ارتشی وجود ندارد و اعمال قوانین فقط توسط پلیس انجام می شود. سطح فن آوری امروز ما برای هماهنگ سازی موفقیت آمیز این جهان، احتمالا کافی خواهد بود. هرچند که جامعه امروز ما ممکن است تمایلی برای زندگی در چنین جهانی را نداشته باشند. اکنون ممکن است پرسید که اگر فن آوری هوش عمومی مصنوعی فوق بشری به این مجموعه اضافه شود، برای ساختار سلسله مراتبی جهان هستی ما چه اتفاقی خواهد افتاد؟ واضح است که فن آوری های حمل و نقل و ارتباطات، پیشرفت خارق العاده ای خواهند کرد و انتظار طبیعی ما این است که روند تاریخی ادامه پیدا کند و سطوح سلسله مراتبی جدیدی، وظیفه هماهنگ سازی مکان های بسیار دورتر را به عهده گیرد که همانطور که در فصل ۶ خواهیم دید، این مکان ها ممکن است شامل منظومه شمسی، کهکشان ها، خوشه های کیهانی و باریکه های بزرگی از جهان هستی ما

باشند. با این حال، انگیزه های حمایت از تمرکز زدایی همچنان باقی مانده است، زیرا هماهنگ سازی مجموعه هایی به این بزرگی فایده ای نخواهد داشت. در مورد ماشین های هوش عمومی مصنوعی فوق بشری باید توجه داشت که قوانین علم فیزیک حدود بالایی را برای فن آوری حمل و نقل و ارتباطات شرکت ها که مجهز به هوش مصنوعی هستند، ایجاد خواهند کرد و احتمال اینکه افراد موجود در بالاترین سطوح سلسله مراتب بتوانند به مدیریت ذره بینی هر چیزی که در زمین اتفاق می افتد، پردازند، غیر ممکن خواهد شد. ماشین هوش عمومی مصنوعی فوق بشری واقع در کهکشان آندرومدا نمی تواند دستورات مفیدی را برای تصمیمات روز مره به شما بدهد، زیرا برای دریافت این دستورالعمل ها باید بیش از پنج میلیون سال صبر کنید. به همین دلیل، مدیریت ذره بینی هوش مصنوعی برای کنترل سیاره زمین از کارایی لازم برخوردار نخواهد بود و مثل این می ماند که برای اجازه گرفتن برای انجام ساده ترین تصمیمات خود مجبور شوید نامه هایی را با کشتیهای عصر کلوبوس به آن سوی اقیانوس اطلس تحویل دهید. محدودیت سرعتی که قوانین فیزیک برای انتقال اطلاعات تحمیل کرده است، چالش آشکاری را برای هر ماشین هوش مصنوعی که می خواهد دنیای ما را تصرف کند، ایجاد خواهد کرد. پرومتهوس قبل از آزاد شدنش، راجع به اینکه چگونه از مشکل از هم گسیختگی ذهنی جلوگیری کند، به دقت فکر کرده بود. به همین دلیل، بسیاری از واحدهای مستقل نرم افزاری هوش مصنوعی پرومتهوس که روی رایانه های مختلف سراسر جهان اجرا می شدند، اهداف و انگیزه هایشان این بود که با هم هماهنگ شوند و به عنوان یک موجودیت یکپارچه واحد، عمل کنند. درست همانطوری که تیم امگا با مشکل تحت کنترل درآوردن پرومتهوس مواجه شده بود، پرومتهوس هم با مشکل کنترل کردن خودش مواجه بود و سعی می کرد مطمئن شود که اجزایش نافرمانی نمی کنند. ما هنوز به روشنی نمی دانیم که هوش مصنوعی با سلسله مراتب مشارکت جویانه خود، چگونه می تواند سیستم های بسیار بزرگ را به طور مستقیم یا غیر مستقیم کنترل کند. خلاصه اینکه، ما با این سوال پیچیده مواجه هستیم که آینده هوش عمومی مصنوعی فوق بشری چگونه کنترل خواهد شد و هنوز پاسخ آن را به روشنی نمی دانیم. برخی استدلال می کنند که آینده

هوش مصنوعی به تمامیت خواهی و اقتدارگرایی خواهد انجامید و برخی دیگر استدلال می کنند که هوش مصنوعی در آینده منجر به افزایش توانمندسازی افراد خواهد شد.

## آدم های مکانیکی و ربات های انسان نما

شاکله اصلی داستان های علمی – تخیلی این است که انسان ها با ماشین ها یکی خواهند شد که این امر یا از طریق بهسازی جسممان به موجودی با اعضای ارگانیک و مکانیکی (با استفاده از فن آوری) یا تولید ربات های شبیه سازی شده با مغز انسان محقق خواهد شد. اقتصاد دانی به نام رابین هانسون در کتاب خود با عنوان "عصر تقلیدها" به بررسی جالب این موضوع پرداخته است که زندگی ما در دنیایی که پر از ربات های انسان نماست، چگونه به نظر خواهد رسید. تولید ربات های انسان نما یکی از حدود غایی طیف آدم های مکانیکی است که در آن، تنها بخش باقی مانده از انسان در ماشین، نرم افزار شبیه سازی شده با مغز اوست. آدم های مکانیکی در طیفی از فیلم های هالیوود از جمله فیلم های مربوط به انسان های کاملاً مکانیکی (مانند جامعه انسانی تخیلی متشکل از آدم های مکانیکی پیشتازان فضا<sup>۶۱</sup>) تا ربات های انسان نما مانند "ترمیناتورز" مشاهده می شوند. فیلم های تخیلی در مورد ربات های تقلید کننده مغز انسان نیز شامل طیفی از فیلم های مربوط به هوش انسان مانند "آینه سیاه" تا فیلم های مربوط به هوش عمومی مصنوعی مافوق بشری مانند "ترنسندنس" می شوند. اگر هوش عمومی مصنوعی فراتر از بشر واقعاً اتفاق بیفتد، وسوسه تبدیل شدن به آدم های مکانیکی یا ربات های انسان نما تقویت خواهد شد. چنین وسوسه ای برای بهسازی های زندگی از طریق فن آوری هنوز هم قوی است، به طوریکه امروزه بسیاری از انسان ها از عینک، سمعک، دستگاه ضربان ساز قلب، اندام های مصنوعی و مولکول های درمانی گردش کننده در جریان خون استفاده می کنند. برخی از نوجوانان هم به گوشی های هوشمند خود وابسته شده اند و خود من هم به لپ تاپ خود وابستگی پیدا کرده ام. یکی از مهمترین طرفداران آدم های مکانیکی شخصی به نام ری کرزویل<sup>۶۲</sup>

<sup>۶۱</sup> Borg from star trek

<sup>۶۲</sup> Ray Kurzweil

می باشد که در کتاب خود با عنوان " زمان پیشی گرفتن هوش ماشین از هوش انسان، نزدیک است " استدلال می کند که ادامه طبیعی این روند در جهان، منجر به بهره برداری از ریز ربات ها، سیستم های بیوفیدبک هوشمند و فن آوری های دیگری خواهد شد که تا اوایل دهه ۲۰۳۰ جایگزین سیستم های گوارشی و غدد درون ریز، خون و قلب مان می شوند و در طی دو دهه بعد از آن، برای بهبود استخوان بندی، پوست، مغز و بقیه اعضای بدنمان به کار گرفته خواهند شد. همچنین، به لطف پیشرفت های شگفت انگیز در زمینه رابط های انسان با رایانه، امکان تغییر سریع ظاهر فیزیکی و مجازی انسان به دلخواه خود، فراهم شده و دانشمند دیگری به نام هانس موراوک<sup>۶۳</sup> با نظریه کرزویل مبنی بر اینکه ماشینی شدن انسان ها چیزی بیشتر از بهبود صرف دی ان ای آنها می باشد، موافق است و می گوید: " انسان ما فوق بشری را می توان از طریق مهندسی ژنتیکی طراحی کرد. با این حال، انسان طراحی شده، فقط نوع درجه دومی از یک ربات خواهد بود که ساخت آن فقط طریق سنتز پروتئینی هدایت شده توسط دی ان ای امکان پذیر خواهد بود. " او همچنین معتقد است که بدن انسان را می توان کاملاً حذف کرد و ربات های انسان نما را با شبیه سازی مغز انسان در نرم افزار آنها، تولید کرد. چنین ربات هایی به صورت یک واقعیت مجازی یا در قالب یک ربات انسان نما ظاهر خواهند شد که بدون محدودیت های مرگ یا منابع شناختی و فقط با در نظر گرفتن محدودیت های علم فیزیک می توانند قدم بزنند، پرواز کنند، شنا کنند و در فاصله ای دور از فضا قرار بگیرند. با وجود اینکه این ایده ها مانند داستان های تخیلی به نظر می رسند، هیچ یک از قوانین شناخته شده علم فیزیک را نقض نمی کنند. امکان تحقق این ایده ها وجود دارد، ولی زمان آن مشخص نیست. برخی از متفکرین پیشرو حدس می زنند که اولین ربات های تقلید کننده مغز انسان در سطح هوش بشر تولید خواهند شد و مسیر آنها به سمت رسیدن به هوش فوق بشری شروع خواهد شد. بیشتر محققان هوش مصنوعی و دانشمندان علوم اعصاب گمان می برند که ساده ترین راه برای رسیدن به هوش فوق بشری این است که از تقلید مغز اجتناب کنیم و آن را به شیوه دیگری مهندسی کنیم. چرا ساده

---

<sup>۶۳</sup> Hans Moravec

ترین راه برای رسیدن به یک فن آوری جدید باید راهی باشد که تحولات جدید به ما ارائه می کند که پر از محدودیت هایی از لحاظ خود سامانی، خود ترمیمی و بازآفرینی خود می باشد. تحولات جدید باعث بهینه سازی و افزایش کارایی می شوند، اما کمکی به راحتی در ساخت یک چیز طراحی شده جدید نمی کنند. پس از طراحی و ساخت پرندگان مکانیکی برای سفر که دارای کارایی بالایی هم بودند، صنعت هوانوردی علاقه ای به استفاده از آنها نشان نداد، زیرا راه حل اولیه ساده تر (سفر با هواپیما)، تناسب بیشتری با نیازهای سفر ما داشت. به همین شکل، گمان می کنم که برای ساختن ماشین هایی با قدرت تفکر در سطح بشر ممکن است راه حل های ساده تری نسبت به آنچه که تحولات به ما ارائه می کنند، وجود داشته باشند. حتی اگر روزی برسد که موفق به شبیه سازی مغز انسان شویم، ممکن است در نهایت کشف کنیم که راه حل های اولیه ساده تر بودند. هرچند که مهندسان راه حل های اولیه ممکن است به اندازه مهندسان راه حل های جدید نگران کارایی انرژی نباشند، اما به زودی می توانند محصولات با کارایی بیشتر را طراحی کنند.

## در آینده چه اتفاقی خواهد افتاد؟

با وجود پیشرفت های بشر در زمینه ساختن ماشین های عمومی مصنوعی در سطح بشر، ما حقیقتاً نمی دانیم که در آینده اتفاقی خواهد افتاد. به همین دلیل، طیفی از سناریوها را مورد بررسی قرار دادیم. برخی از افراد به من می گویند که مطمئن هستند چنین اتفاقی نخواهد افتاد. با این حال، فکر می کنم که عاقلانه این است که در حال حاضر، متواضع باشیم و به نادانی خود اعتراف کنیم. همچنین، باید توجه داشته باشیم که برای هر یک از این سناریوها (خیزش سریع، خیزش کند، انسان ها، ماشین ها، آدم های مکانیکی، ربات های انسان نما، تمرکز یا تعدد مراکز قدرت) حداقل یک محقق ارزشمند هوش مصنوعی وجود دارد که به آن به عنوان یک واقعیت امکان پذیر نگاه می کند. با گذشت زمان و ماندن ما در بین این دو راهی، اولین سوال مهمی که مطرح می شود، این است که: "آیا هرگز می توان ماشین هوش عمومی مصنوعی در سطح هوش انسان را ایجاد کرد؟" پیش فرض ما در این فصل این است که این امکان وجود دارد. اما برخی از کارشناسان هوش مصنوعی فکر می کنند که حداقل

تا صدها سال دیگر، این اتفاق نخواهد افتاد. همانطور که قبلاً گفتم، حدود نیمی از کارشناسان هوش مصنوعی در کنفرانس پورتوریکو حدس می زدند که تا سال ۲۰۵۵ این اتفاق خواهد افتاد. قبل از اینکه چنین ماشینی خلق شود، بحث بر سر این موضوع می باشد که این کار توسط چه گروهی انجام خواهد شد. مهندسین رایانه، ربات های شبیه سازی شده با مغز انسان یا ابزارهای پیش بینی نشده جدید؟ اگر مهندسین رایانه ای که در حال حاضر بر این حوزه مسلط هستند، نتوانند پس از چند قرن چنین ماشین هایی را بسازند، احتمال استفاده از ربات های انسان نما بیشتر خواهد شد. سوال دیگری که مطرح می شود این است که: "اگر احتمال ساخت چنین ماشین هایی قطعی شود، از چه سناریوی استفاده خواهد شد. خیزش سریع، خیزش کند یا بدون خیزش؟ همانطور که دیدید، خیزش سریع باعث تسهیل در تصاحب جهان خواهد شد. اما، خیزش کند احتمالاً منجر به ایجاد رقابت میان تعداد زیادی از سرمایه گذاران خواهد شد. نیک بوستروم از تحلیل موشکافانه تری برای پاسخ به این سوال استفاده می کند و می گوید که برای افزایش کیفیت هوش مصنوعی می توان از تحلیل قدرت بهینه سازی برای افزایش چالاکی هوش مصنوعی و از تحلیل نافرمانی برای درک موانع پیشرفت آن استفاده کرد. نرخ متوسط پیشرفت هنگامی افزایش می یابد که قدرت بهینه سازی بیشتری بر فعالیت ها اعمال شود و هنگامی کاهش می یابد که با نافرمانی بیشتری مواجه شویم. او استدلال می کند که رسیدن هوش ماشین به هوش انسان و فراتر رفتن از آن، به واسطه نافرمانی ممکن است افزایش یا کاهش پیدا کند. اما در مورد قدرت بهینه سازی می توان گفت که با پیشرفت و فراتر رفتن هوش ماشین از هوش انسان، قدرت بهینه سازی هوش مصنوعی به سرعت افزایش خواهد یافت، زیرا ورودی اصلی به فرایند بهینه سازی از خود ماشین ناشی می شود و نه انسان و هرچه ماشین توانا تر شود، سریعتر بهبود پیدا می کند. در مورد هر فرایندی که نرخ قدرت آن متناسب با نرخ قدرت فعلی اش رشد پیدا می کند، در فواصل زمانی منظم، قدرت آن دو برابر می شود. نرخ چنین رشدی از تابع نمایی<sup>۶۴</sup> پیروی می کند و این فرایندها "انفجاری" خوانده می شوند. اگر قدرت فرزندآوری متناسب

---

<sup>۶۴</sup> exponential

با اندازه جمعیت رشد کند، با انفجار جمعیت مواجه خواهیم شد. اگر تولید نوترون های قادر به شکافت هسته پلوتونیوم متناسب با تعداد نوترون ها رشد کند، با انفجار هسته ای مواجه خواهیم شد. اگر هوش ماشین ها با نرخ متناسب با قدرت هوشی فعلی آنها رشد کند، با انفجار هوش مواجه خواهیم شد. ویژگی اصلی این انفجارها، مدت زمانی است که طول می کشد تا قدرت آنها دوبرابر شود. اگر زمان لازم برای انفجار هوش، چند ساعت یا چند روز باشد (چنانچه در تیم امگا شاهد آن بودیم)، از امکان خیزش بسیار سریعی برخوردار خواهیم شد. مدت زمانی که تا انفجار طول می کشد، بستگی به این دارد که آیا برای بهبود هوش ماشین فقط به نرم افزار جدید نیاز است ( که در این صورت انفجار هوش ظرف چند ثانیه، چند دقیقه یا چند ساعت رخ خواهد داد) یا به سخت افزار جدید نیاز است (که در این صورت برای انفجار هوش به ماه ها یا سال ها زمان نیاز است). در سناریوی امگا شاهد " انفجار سخت افزاری " بودیم، زیرا تیم امگا کیفیت پایین نرم افزار خود را با مقادیر بالای سخت افزاری جبران کرده بود و این در حالی بود که پرومیتئوس فقط با بهبود نرم افزار خود می توانست کیفیت هوش ماشین خود را از بسیاری لحاظ دو چندان کند. در سناریوی امگا شاهد " انفجار محتوایی " حجم داده های اینترنتی نیز بودیم، زیرا پرومیتئوس برای بهره برداری از داده های اینترنتی هنوز به اندازه کافی با هوش نبود. اما، وقتی که هوش پرومیتئوس رشد کرد، داده های مورد نیاز برای یادگیری را بدون تاخیر فرا گرفت. در نظر گرفتن هزینه های مربوط به سخت افزار و برق نیز برای اجرای هوش مصنوعی مهم هستند. زیرا انفجار هوش وقتی رخ می دهد که هزینه کارهای انسانی از دستمزدهای ساعتی نیروی کار کمتر باشد. به عنوان مثال، فرض کنید که اولین ماشین هوش عمومی مصنوعی در سطح انسان را بتوان با هزینه یک میلیون دلار به ازای هر ساعت کار نیروی انسانی، راه اندازی کرد. این ماشین هوش مصنوعی به لحاظ جدید بودنش ارزشمند خواهد بود و بدون شک، تیر روزنامه ها خواهد شد، اما در معرض بهسازی بازگشتی خودش قرار نخواهد گرفت، زیرا بهسازی آن توسط انسان ها بسیار ارزان تر تمام خواهد شد. فرض کنید که انسان ها با مدیریت خود بتوانند هزینه ها را از ۱۰۰۰۰۰ دلار در ساعت به یک دلار در ساعت کاهش دهند. اما تا قبل از آن زمان،



هزینه استفاده از رایانه برای برنامه نویسی مجدد خودش به مراتب کمتر از هزینه پرداخت پول به افراد برنامه نویس برای برنامه نویسی مجدد خواهد شد. به این ترتیب می توان انسان ها را اخراج کرد و قدرت بهینه سازی را با خریدن زمان رایانش ابری به شدت افزایش داد. این رویه به کاهش بیشتر هزینه ها و افزایش قدرت بهینه سازی منجر خواهد شد و زمینه را برای انفجار هوش فراهم خواهد کرد. سوال کلیدی ما این است که: "چه کسی یا چه چیزی انفجار هوش و آینده آن را کنترل خواهد کرد و اینکه اهدافش چه خواهد بود؟ در مورد این اهداف در فصل هفتم بیشتر بررسی خواهیم کرد. برای حل مشکل کنترل هوش مصنوعی، باید نحوه کنترل آن و میزان کنترل آن بر افراد یا چیزهای دیگر را یاد بگیریم. بسیاری از متفکرین جدی در سراسر جهان در حال بررسی این موضوع هستند که در نهایت چه اتفاقی خواهد افتاد. برخی از آنها استدلال می کنند که نتیجه پیش فرض، این است که سرنوشت بدی در انتظار ماست. اما برخی دیگر اصرار می کنند که نتیجه خوبی حاصل خواهد شد. با این حال، به نظر من این یک پرسش فریبده است و ما نباید از موضع منفعلانه به آن نگاه کنیم و آن را سرنوشت قطعی خود بدانیم. اگر یک تمدن بیگانه با فن آوری برتر با یک سفینه فضایی به ما نزدیک می شد، ممکن بود از خود پرسیم که: "چه اتفاقی خواهد افتاد؟"، زیرا قدرت آنها به مراتب بیشتر از قدرت ما بود و ما نمی توانستیم تاثیری بر نتیجه حاصله داشته باشیم. اما اگر خود ما این تمدن فن آورانه مجهز به هوش مصنوعی را ساخته بودیم، می توانستیم بر نتیجه حاصله تاثیر داشته باشیم و از خود پرسیم که: "چه اتفاقی باید بیفتد یا اینکه چه آینده ای را می خواهیم؟" در فصل بعدی، طیف وسیعی از سناریوهای مربوط به آینده هوش عمومی مصنوعی در سطح بشر را بررسی خواهیم کرد و دوست دارم بدانم که چگونه آنها را از بهترین تا بدترین، رتبه بندی می کنید. ما تنها زمانی قادر به هدایت مسیر خود به سمت آینده مطلوب می شویم که راجع به آینده ای که می خواهیم داشته باشیم، به دقت فکر کرده باشیم. اگر ندانیم که چه می خواهیم، به احتمال زیاد به آن نخواهیم رسید.

## خلاصه فصل:

- اگر روزی موفق به ساختن ماشین هوش عمومی مصنوعی در سطح بشر شویم، این امر ممکن است منجر به تحریک انفجار هوش شود و هوش مصنوعی ممکن است با چنان سرعتی پیشرفت کند که فاصله اش با ما بسیار زیاد شود.
- اگر گروهی از انسان ها برای کنترل انفجار هوش، آن را مدیریت کنند، ممکن است بتوانند در ظرف چند سال، بر دنیا تسلط پیدا کنند.
- اگر انسان ها نتوانند انفجار هوش را کنترل کنند، خود هوش مصنوعی ممکن است با سرعت بیشتری بر جهان حاکم شود.
- در حالی که انفجار سریع هوش به احتمال زیاد منجر به دنیای تک قطبی خواهد شد، انفجار کند و آهسته هوش که ممکن است سال ها یا دهه ها به طول بینجامد، به احتمال زیاد منجر به سناریوی چند قطبی خواهد شد که با موازنه قدرت میان میان تعداد زیادی از موجودیت های نسبتا مستقل همراه خواهد بود.
- تاریخچه زندگی بشر نشان می دهد که همواره نوعی خود سازماندهی به سمت یک سلسله مراتب پیچیده تر که با همکاری، رقابت و کنترل شکل می گیرد، وجود داشته است. هوش مصنوعی فوق بشری به احتمال زیاد، امکان هماهنگی در مقیاس همیشه روزافزون جهان هستی را فراهم خواهد ساخت، اما معلوم نیست که این هماهنگی در نهایت منجر به کنترل از بالا به پایین تمامیت خواهانه خواهد شد یا افزایش توانمندسازی افراد؟
- تولید آدم های مکانیکی و ربات های انسان نما، یک امر باورکردنی است، اما احتمالا سریعترین راه برای استفاده از هوش پیشرفته ماشین نمی باشد.
- نقطه پایان رقابت فعلی برای بهره برداری از هوش مصنوعی ممکن است بهترین یا بدترین آینده اتفاق افتاده برای بشریت باشد که در فصل بعدی به آنها خواهیم پرداخت.
- ما باید با دقت فکر کنیم که چه نتیجه ای را ترجیح می دهیم و چگونه باید به آن سمت هدایت شویم، زیرا اگر ندانیم که چه می خواهیم، به احتمال زیاد به آن نخواهیم رسید.

## فصل پنجم- آنچه که در ۱۰۰۰۰ سال آینده اتفاق خواهد افتاد.

مسابقه برای تولید ماشین هوش عمومی مصنوعی فوق بشری ادامه دارد و ما نمی دانیم که چگونه پیش خواهد رفت. با این حال، این باعث نمی شود که به فکر آینده مطلوب خود نباشیم. زیرا آنچه که می خواهیم، روی نتیجه ای که در آینده حاصل خواهد شد، تاثیر می گذارد. به عنوان مثال، کدامیک از نتایج زیر را ترجیح می دهید و دلالتان برای انتخاب آنها چیست؟

- ۱- آیا می خواهید که هوش عمومی مصنوعی فوق بشری بر جهان حاکم شود؟
- ۲- آیا می خواهید که انسان ها همچنان وجود داشته باشند یا اینکه با ماشین ها جایگزین شوند یا اینکه جهان پر از آدم های مکانیکی و ربات های شبیه سازی شده با مغز انسان شود؟
- ۳- آیا می خواهید که ماشین ها تحت کنترل باشند یا انسان ها تحت کنترل ماشین؟
- ۴- آیا می خواهید که ماشین های هوش مصنوعی، هوشیار باشند یا خیر؟
- ۵- آیا می خواهید که تجارب مثبت را حداکثر سازید، رنج را به حداقل برسانید یا اینکه مسائل را رها کنید تا خودشان حل و فصل شوند؟
- ۶- آیا می خواهید که امکان زندگی به افلاک هم گسترش یابد؟
- ۷- آیا می خواهید که تمدن بشر برای رسیدن به هدف بزرگتری که از آن جانبداری می کنید تلاش کند یا اینکه با اشکال زندگی در آینده مشکلی ندارید و برایتان مهم نیست که اهداف دیگر چقدر پیش پا افتاده باشند؟

در اینجا به بررسی سناریوهای مربوط به آینده هوش مصنوعی می پردازیم. با این حال، واضح است که این فهرست، کامل نیست و فقط طیفی از احتمالات ممکن را پوشش می دهد. من توصیه می کنم که سوالات یک تا هفت را پاسخ دهید و پس از پایان این فصل، دوباره آنها را بخوانید تا ببینید که نظرتان نسبت به آن ها تغییر پیدا کرده است یا خیر.

<b>خلاصه ای از سناریوهای مربوط به آینده هوش مصنوعی</b>
--

<p>انسان ها، آدم های مکانیکی، ربات های انسان نما و ماشین های هوش مصنوعی فوق هوشمند، به لطف حقوق مالکیت خصوصی، با یکدیگر همزیستی مسالمت آمیز دارند.</p>	<p><b>مدینه فاضله آزادی خواهان</b></p>
<p>همه می دانند که هوش مصنوعی جامعه را اداره می کند و قوانین سخت گیرانه ای را اعمال می کند. با این حال، بیشتر افراد، به هوش مصنوعی به عنوان یک چیز خوب نگاه می کنند.</p>	<p><b>دیکتاتور خیر خواه</b></p>
<p>انسان ها، آدم های مکانیکی و ربات های انسان نما به لطف لغو حقوق مالکیت خصوصی و درآمد تضمین شده، با یکدیگر همزیستی مسالمت آمیز دارند.</p>	<p><b>مدینه فاضله عدالت طلبان</b></p>
<p>برای جلوگیری از ایجاد یک ماشین هوش مصنوعی فوق بشری دیگر، ماشین هوش مصنوعی با هدف اجازه دادن به کمترین مداخله ممکن در کارش، ایجاد می شود. در نتیجه، ربات های انسان نما و آدم های مکانیکی همچنان وجود دارند، اما پیشرفت فن آوری برای همیشه، متوقف می شود.</p>	<p><b>نگهبان در (کنترل کننده دسترسی)</b></p>
<p>هوش مصنوعی واقف بر همه چیز و قادر مطلق، با کمترین مداخله در کارهای ما تلاش می کند که رضایت ما را جلب کند و این احساس را در وجودمان ایجاد کند که قادر به کنترل بر سرنوشت خویش هستیم. هوش مصنوعی به گونه ای مخفی می شود که بسیاری از انسان ها به وجودش شک می کنند.</p>	<p><b>خدای نگه دارنده</b></p>

<p><b>خدای به بردگی گرفته شده</b></p>	<p>هوش عمومی مصنوعی فوق بشری توسط انسان هایی محدود می شود که از آن برای تولید فن آوری غیر قابل تصور یا تولید ثروت استفاده می کنند و استفاده خوب یا بد انسان ها از هوش مصنوعی بستگی به کنترل کنندگان انسانی آن دارد.</p>
<p><b>گروه غالب</b></p>	<p>هوش مصنوعی کنترل را به دست می گیرد. به این نتیجه می رسد که انسان ها تهدید کننده / مزاحم / تلف کننده منابع هستند و با روشی که قادر به درک آن نیستیم، از شر انسان ها خلاص می شود.</p>
<p><b>بازماندگان</b></p>	<p>ماشین های هوش مصنوعی جای انسان ها را می گیرند. اما، انسان ها به آنها به عنوان بازماندگان ارزشمند خود نگاه می کنند. درست همانطوری که والدین از داشتن بچه با هوش تر از خودشان که از آنها یاد می گیرد و به آنچه که رویای آنها بود دست پیدا می کند، خوشحال می شوند و احساس افتخار می کنند.</p>
<p><b>نگهبان باغ وحش</b></p>	<p>هوش مصنوعی قادر مطلق، انسان هایی را پیش خود نگه می دارد که احساس می کنند مانند حیوانات باغ وحش با آنها رفتار می شود و برای سرنوشت خود غصه می خورند.</p>
<p><b>۱۹۸۴</b></p>	<p>پیشرفت فن آورانه به سمت ایجاد ماشین های هوش عمومی مصنوعی فوق بشری دائما توسط حکومت های سرکوب گر محدود می شود و انواع خاصی از تحقیقات در مورد هوش مصنوعی در این کشورها ممنوع است.</p>

<p>برگشت به عقب</p> <p>بازگشت به جامعه بدون فن آوری به سبک فرقه مسیحی آمیش، از پیشرفت فن آوران در جهت ایجاد ماشین های هوش عمومی مصنوعی فوق بشری جلوگیری می کند.</p>	
<p>خودتخریبی</p> <p>ماشین های هوش عمومی مصنوعی فوق بشری هرگز ایجاد نمی شوند، زیرا انگیزه های بشری به واسطه شوک ناشی از سلاح های هسته ای یا بیوفن آوری نابود شده است.</p>	

سناریو	آیا ماشین هوش مصنوعی فوق بشری وجود دارد؟	آیا انسان ها وجود دارند؟	آیا انسان ها تحت کنترل هستند؟	آیا انسان ها ایمن هستند؟	آیا انسان ها خشنود هستند؟	آیا هوشیاری وجود دارد؟
مدینه فاضله آزادی خواهان	بلی	بلی	خیر	خیر	مبهم	بلی
دیکتاتور خیر خواه	بلی	بلی	خیر	بلی	مبهم	بلی
مدینه فاضله عدالت طلبان	خیر	بلی	شاید بلی	بلی	شاید بلی	بلی
نگهبان در (کنترل کننده دسترسی)	بلی	بلی	تا حدودی	به احتمال قوی	مبهم	بلی
خدای نگه دارنده	بلی	بلی	تا حدودی	به احتمال قوی	مبهم	بلی
خدای به بردگی گرفته شده	بلی	بلی	بلی	به احتمال قوی	مبهم	بلی

گروه غالب	بلی	خیر	-	-	-	؟
بازماندگان	بلی	خیر	-	-	-	؟
نگهبان باغ وحش	بلی	بلی	خیر	بلی	خیر	بلی
۱۹۸۴	خیر	بلی	بلی	به احتمال قوی	مبهم	بلی
برگشت به عقب	خیر	بلی	بلی	خیر	مبهم	بلی
خودتخریبی	خیر	خیر	-	-	-	خیر

## مدینه فاضله آزادی خواهان

ابتدا به بررسی سناریویی می پردازیم که در آن، انسان ها با فن آوری همزیستی مسالمت آمیز دارند و در بعضی موارد به هم پیوند می خورند که این سناریو توسط بسیاری از آینده پژوهان و نویسندگان داستان های علمی، تصور شده است. زندگی بر روی زمین و فراتر از آن (چنانچه در فصل بعد خواهیم دید)، متفاوت تر از همیشه خواهد بود. اگر به فیلم ماهواره ای زمین نگاه کنید، به آسانی می توانید مناطق کاملاً ماشینی، مناطق ترکیبی و مناطق کاملاً انسانی را از یکدیگر تشخیص دهید. مناطق ماشینی شامل کارخانجات فراوان تحت کنترل ربات و تجهیزات رایانه ای می باشد که انسانی در آنها زندگی نمی کند و هدف ماشین ها این است که کارایی تمام اتم ها را به حداکثر برسانند. هرچند که مناطق ماشینی از بیرون، خسته کننده و بی روح به نظر می رسند، اما در داخل، بسیار زنده و با روح هستند و تجارب جالبی را ارائه می کنند. زمین، پذیرای اذهان فوق هوشمندی است که با هم رقابت یا همکاری می کنند و همه آنها در مناطق ماشینی سکنی گزیده اند. ساکنان مناطق ترکیبی را مجموعه جداگانه ای از رایانه ها، ربات ها، انسان ها و ترکیبی از هر سه این ها تشکیل می دهند. همانطور که آینده پژوهانی مانند هانس موراوک و ری کرزویل مجسم ساخته اند، بسیاری از انسان های ساکن در

این مناطق به درجات مختلفی، بدن های خود را از لحاظ فن آوری ارتقا داده اند و برخی از آنها مغز خود را روی سخت افزارهای جدید، شبیه سازی کرده اند و این باعث شده است که تفاوت میان انسان و ماشین نامشخص و کمرنگ شود. بیشتر موجودیت های هوشمند دارای جسم فیزیکی ثابتی نیستند و به شکل نرم افزارهایی زندگی می کنند که می توانند به سرعت در بین رایانه ها حرکت کنند و خودشان را در دنیای واقعی در قالب اجسام رباطیکی ظاهر کنند. از آنجایی که ذهن این موجودیت ها قابل کپی کردن است، "اندازه جمعیت" ثابت نخواهد ماند. آزاد بودن از جسم فیزیکی باعث می شود که آنها احساس استقلال فردی کمتری کنند، زیرا به سادگی می توانند دانش و تجارب موجود در واحدهای مستقل نرم افزاری خود را با دیگران به اشتراک گذارند. آنها همچنین احساس جاودانگی می کنند، زیرا قادر به گرفتن کپی های بک آپ (پشتیبانی) از خودشان هستند. از یک جهت می توان گفت که موجودیت های مهم زندگی، افکارمان نیستند، بلکه تجاربمان هستند. خصوصاً اینکه تجارب شگفت انگیز، زنده می مانند، زیرا دائماً کپی می شوند و مورد استفاده ذهن های دیگر قرار می گیرند و تجاربی که جالب نیستند، توسط صاحبانشان حذف می شوند تا فضای خالی برای ذخیره کردن تجارب بهتر آزاد شود. هرچند که برای افزایش راحتی و سرعت، اکثر تعاملات در محیط های مجازی اتفاق می افتند، بسیاری از ذهن ها از اینکه با بدن فیزیکی تعامل و فعالیت کنند نیز لذت می برند. برخی از ربات های مناطق ترکیبی که در خیابان ها پرسه می زنند یا اسکی می کنند، توسط مغز انسان یا انسان های مجازی از نوع واقعیت افزوده، کنترل می شوند که به صورت جسم درآمده اند و از همراهی با انسان های واقعی و مجازی لذت می برند. در مناطق کاملاً انسانی، استفاده از ماشین های دارای هوش عمومی در سطح بشری یا فراتر از آن از جمله موجودات زنده ارتقا یافته توسط فن آوری، ممنوع است. زندگی در این مناطق تفاوت چشمگیری با زندگی امروزی ما ندارد، به جز اینکه مردم از رفاه و راحتی بیشتری برخوردار هستند، فقر به طور کامل حذف شده است و راه حل درمانی برای بیشتر بیماری های امروزی موجود است. بخش بسیار کوچکی از انسان هایی که زندگی در این مناطق را انتخاب کرده اند، از سطح هوشیاری کمتری نسبت به دیگران



برخوردارند و درک چندانی از آنچه که هموعان باهوش ترشان در مناطق دیگر انجام می دهند، ندارند. با این حال، بسیاری از آنها از زندگی خود راضی هستند.

## جنبه اقتصادی هوش مصنوعی

بخش بسیار بزرگی از محاسبات در مناطق ماشینی انجام می شود که بیشترشان تحت تملک ماشین های هوش عمومی مصنوعی فوق هوشمندی هستند که در آن مناطق زندگی می کنند و با یکدیگر در حال رقابت هستند. هوش و فن آوری برتر آنها موجب شده تا هیچ موجودیت دیگری نتواند قدرتش را به چالش بکشد. این ماشین های هوش مصنوعی توافق کرده اند که تحت یک نظام حکومتی طرفدار آزادی که هیچ قانونی به جز حفظ مالکیت خصوصی افراد را ندارد، با یکدیگر همکاری کنند و با هم هماهنگ شوند. حقوق مالکیت خصوصی که شامل تمامی موجودات هوشمند از جمله انسان ها می شود، نحوه به وجود آمدن مناطق کاملاً انسانی را توضیح می دهد. از همان ابتدا، گروه هایی از انسان ها با هم متحد شدند و تصمیم گرفتند که فروش دارایی به غیر انسان ها را در مناطق کاملاً انسانی ممنوع کنند. ماشین های هوش عمومی مصنوعی فوق هوشمند به دلیل فن آوری پیشرفته خود در نهایت ثروتمندتر از انسان ها خواهند شد. با این حال، انسان هایی که در مناطق کاملاً انسانی زندگی می کنند، هنوز از لحاظ مادی در رفاه بیشتری از بیشتر انسان های امروزی هستند. اقتصاد آنها از اقتصاد ماشین ها جدا شده است و به همین دلیل، حضور ماشین ها در جاهای دیگر، تاثیر زیادی بر روی زندگی آنها نخواهد داشت، به جز اینکه انسان ها هر از گاهی از فن آوری های مفید ماشین ها که برایشان قابل درک و کپی برداری است، استفاده می کنند. همچنین، مهم نیست که انسان ها چیزی برای فروش به ماشین ها نداشته باشند، زیرا ماشین ها در عوض خدمات خود به انسان ها چیزی از آنها نمی خواهند. در مناطق ترکیبی، اختلاف ثروت میان ماشین های هوش مصنوعی و انسان ها قابل ملاحظه تر می باشد و این باعث شده که زمین که تنها محصول تحت تملک انسان ها و مورد نیاز ماشین ها می باشد، به قیمت نجومی به ماشین ها فروخته شود. بیشتر انسان هایی که مالک زمین هستند، در عوض درآمد پایه تضمین شده ای که تا پایان عمر به آنها و اولادشان داده می شود، بخش

کوچکی از زمین را به ماشین های هوش مصنوعی می فروشند. به این ترتیب، انسان ها دیگر نیازی به کار کردن ندارند و می توانند از فراوانی شگفت انگیز محصولات و خدمات ارزان قیمت ماشین ها به صورت فیزیکی و مجازی بهره مند شوند. در مورد ماشین ها هم می توان گفت که مناطق ترکیبی به جای محلی برای کار کردن، بیشتر به محلی برای تفریح آنها تبدیل شده است.

## چرا این اتفاق هرگز نخواهد افتاد؟

قبل از اینکه وارد ماجراهای هیجان انگیز مربوط به آدم های مکانیکی و ربات های انسان نما شویم، به دلایل نشدنی بودن این سناریو می پردازیم. قبل از هر چیز باید بگوییم که دو راه ممکن در پیش روی انسان های ارتقا یافته (آدم های مکانیکی و ربات های انسان نما) وجود دارد:

۱- خودمان تشخیص دهیم که آنها را چگونه باید ایجاد کنیم.

۲- ماشین های هوش مصنوعی فوق هوشمندی بسازیم که خودشان پاسخ این سوال را به ما بدهند؟

اگر از عهده مورد اول بر آییم، طبیعتاً دنیایی پر از آدم های مکانیکی و ربات های انسان نما خواهیم داشت. با این حال، بیشتر محققان هوش مصنوعی معتقدند که به احتمال زیاد چنین اتفاقی نخواهد افتاد، زیرا ساختن ماشین هایی با مغز پیشرفته یا دیجیتالی مشکل تر از استفاده از خود ماشین های هوش عمومی مصنوعی دست نخورده خواهد بود. پس از اینکه ماشین های قوی هوش مصنوعی ساخته شدند، معلوم نیست که آدم های مکانیکی و ربات های انسان نما ساخته شوند. حتی اگر سناریوی آدم های مکانیکی و ربات های انسان نما اتفاق بیفتد، معلوم نیست که آنها ثابت و پایدار بمانند. چرا باید توازن قدرت میان ماشین های هوش مصنوعی فوق هوشمند به مدت هزاران سال، ثابت بماند و ماشین های هوش مصنوعی با هم ملحق نشوند یا باهوش ترین آنها کنترل اوضاع را به دست نگیرد؟ علاوه بر این، با توجه به اینکه ماشین های هوش مصنوعی نیازی به انسان ها نخواهند داشت و همه کارهای انسانی را می توانند بهتر و ارزان تر از انسان ها انجام دهند، چرا باید به حقوق

مالکیت خصوصی انسان ها احترام گذارند و انسان ها را پیش خود نگه دارند؟ ری کرزویل حدس می زند که براندازی و نابودی انسان های طبیعی و ارتقا یافته غیر ممکن است، زیرا ماشین های هوش مصنوعی به خاطر به وجود آمدن خود مدیون انسان ها هستند و به همین دلیل، برای آنها احترام قایل هستند. با این حال، همانطور که در فصل هفتم خواهیم دید، ما نباید در دام قایل شدن جنبه انسانی برای ماشین های هوش مصنوعی گرفتار شویم و فرض کنیم که آنها هم مانند انسان ها دارای احساس قدرتانی هستند. حتی اگر این فرضیه را بپذیریم که ماشین های هوش مصنوعی به حقوق مالکیت خصوصی انسان ها احترام خواهند گذاشت، ممکن است که این ماشین ها به تدریج بخش زیادی از زمین ما را به شیوه های دیگری مانند بهره برداری از قدرت بالای متقاعد کنندگی خود، تصاحب کنند و انسان ها را با دادن وعده داشتن یک زندگی راحت، ترغیب به فروختن بخشی از زمین خود کنند. ماشین ها در بخش هایی از جهان که فقط انسان ها در آن زندگی می کنند، ممکن است که انسان ها را وسوسه کنند که پویش های سیاسی ای را برای اجازه دادن فروش زمین به ماشین ها به راه اندازند تا برای نجات یک بچه بیمار یا وعده رسیدن به جاودانگی، سرسخت ترین و بدبین ترین انسان ها را مایل به فروش بخشی از زمین های خود به ماشین ها کنند. با دادن آموزش به انسان ها برای کاهش نرخ زاد و ولد می توان بدون دخالت ماشین ها، اندازه جمعیت را کاهش داد که این کار در حال حاضر در ژاپن و آلمان در حال اتفاق افتادن است.

### جنبه های منفی

برخی از حامیان دو آتشه آدم های مکانیکی و ربات های انسان نما، قول سعادت جاودانی فن آوری و نابودی کامل زندگی انسان ها را داده اند. در واقع، آینده شبیه سازی مغز انسان باعث ایجاد انگیزه برای صدها نفر شد تا در شرکت الکور واقع در آریزونا، بخواهند که مغز خود را پس از مرگ، زنده نگه دارند. با این حال، اگر این فن آوری از راه برسد، معلوم نیست که در اختیار همه قرار بگیرد. به احتمال زیاد، ثروتمندترین افراد از آن استفاده خواهند کرد. حتی اگر این فن آوری ارزانتر شود، چه کسانی در اولویت بالاتری برای استفاده از آن قرار می گیرند؟ افراد آسیب دیده شدید مغزی؟

گوریل‌ها؟ مورچه‌ها، گیاهان یا باکتری‌ها؟ آیا ملت‌های آینده مانند افراد وسواسی عمل خواهند کرد و سعی می‌کنند مغز همه موجودات را شبیه‌سازی کنند یا اینکه مانند روش کشتی نوح، فقط مغز نمونه‌های جالبی از تمام گونه‌ها یا نمایندگانی از تمام انسان‌ها را برای شبیه‌سازی و زنده نگه داشتن، انتخاب خواهند کرد. برای موجودیت‌های به شدت هوشمندی که در آن زمان زندگی خواهند کرد، انسان شبیه‌سازی شده مانند موش یا ماری می‌ماند که توسط انسان‌ها شبیه‌سازی شده است. بسیاری از افراد ممکن است از این سناریوی مدینه فاضله آزادی خواهان خوششان نیاید و آن را مایه رنج و عذاب بدانند. از آنجایی که حقوق مالکیت خصوصی، تنها اصل جدی این سناریو می‌باشد، چیزی مانع از تداوم درد ورنجی که در دنیای امروزی به وفور یافت می‌شود، به مناطق کاملاً انسانی و ترکیبی نمی‌شود. هرچند که برخی از انسان‌ها ممکن است از زندگی لذت ببرند، بقیه آنها ممکن است در نهایت در دنیایی پر از فقر و بردگی زندگی کنند یا از خشونت، ترس، سرکوب و افسردگی رنج ببرند. مارشال برین در رمان خود توضیح می‌دهد که پیشرفت هوش مصنوعی در نظام اقتصادی آزادی خواهانه چگونه منجر به بیکار شدن بیشتر آمریکایی‌ها و محکوم شدن آنها به داشتن یک زندگی بی‌روح و خسته کننده در چارچوب طرح‌ها و پروژه‌های مسکن و رفاه اجتماعی تحت مدیریت ربات‌ها شوند. آنها شبیه حیوانات مزرعه‌ای می‌مانند که به آنها غذا داده می‌شود و سلامت و امنیتشان در یک فضای تنگ و تاریک تامین می‌شود و افراد ثروتمند حاضر به نظارت بر شرایط سخت زندگی آنها نیستند. آنها داروهایی را برای کنترل زاد و ولد خود مصرف می‌کنند که باعث می‌شود بچه‌ای نداشته باشند و جمعیتشان به قدری محدود شود که افراد ثروتمند سهم بیشتری از ثروت تولید شده توسط ربات‌ها داشته باشند. در سناریوی مدینه فاضله آزادی خواهان، درد و ناراحتی فقط مخصوص انسان‌ها نیست و ماشین‌ها هم در صورت داشتن تجارب حسی ناخوشایند ممکن است احساس ناراحتی کنند. به عنوان مثال، یک ربات دارای شخصیت ضد اجتماعی و کینه جو ممکن است در دنیای مجازی دشمن خود را در معرض عذاب‌های وحشتناکی قرار دهد که شدت و مدت این عذاب‌ها به مراتب فراتر از عذاب‌های موجود در دنیای واقعی باشد.

## دیکتاتور خیر خواه

اکنون به بررسی سناریوی می پردازیم که در آن، تمامی اشکال درد و ناراحتی حذف شده است، زیرا یک ماشین فوق هوشمند دنیا را اداره می کند و قوانین سختگیرانه ای را برای به حداکثر رساندن شادی انسان، اعمال می کند. این سناریو یکی از خروجی های ممکن سناریوی امگا است که در آن، برای داشتن یک جامعه انسانی در حال پیشرفت، قدرت به پرومئوس واگذار شده است. انسان ها به لطف فن آوری های شگفت انگیز توسعه داده شده توسط هوش مصنوعی دیکتاتور، دیگر مبتلا به فقر، بیماری و سایر مشکلات موجود در فن آوری های سطح پایین نخواهد شد و تمام انسان ها از زندگی آسوده و مجلل لذت خواهند برد. تمام نیازهای اولیه افراد برآورده خواهند شد و ماشین های تحت کنترل هوش مصنوعی، تامین کننده کلیه کالاها و خدمات مورد نیاز انسان ها خواهند بود. جرم و جنایت عملاً حذف خواهد شد، زیرا هوش مصنوعی دیکتاتور همه چیز دان است و تمام افرادی را که از قوانین سرپیچی می کنند، مجازات خواهد کرد. تمام انسان ها دستبندهای امنیتی خواهند پوشید و بی وقفه مورد نظارت، مجازات، بیهوشی و اجرای حکم واقع خواهند شد. همه افراد جامعه می دانند که تحت حاکمیت نظام استبدادی و نظارت شدید قرار دارند و بیشتر آنها به این نظارت ها به عنوان یک چیز خوب نگاه می کنند. یکی از اهداف هوش مصنوعی فوق هوشمند دیکتاتور این است که با بررسی نیازهای رو به رشد کدگذاری شده در ژن های ما، تشخیص دهد که مدینه فاضله چگونه باید به نظر برسد تا بعداً بتواند آن را پیاده سازی کند. دوربینی زیرکانه افرادی که هوش مصنوعی را ایجاد کرده اند، باعث شده که فقط به دنبال حداکثر سازی شادی ما از روش هایی مانند تزریق مواد مخدر به همه افراد، نباشند و تعریف کاملاً دقیقتر و پیچیده تری از رشد و شکوفایی انسان ها را ارائه دهند و زمین را به محیط پرمایه تری برای زندگی انسان ها تبدیل کنند و این باعث می شود که بیشتر افراد از زندگی خود راضی باشند و زندگی هدفمندی را تجربه کنند.

## سیستم بخشی

هوش مصنوعی با توجه به اهمیت موضوع تنوع طلبی و درک این موضوع که افراد مختلف دارای سلیقه های مختلفی هستند، زمین را به بخش های مختلفی تقسیم کرده است که افراد می توانند بر اساس فکر و عقیده خود، بخش هایی را از میان آنها را انتخاب کنند. برخی از این بخش ها شامل موارد زیر می باشند:

- بخش دانش: هوش مصنوعی در این بخش، تحصیلات بهینه سازی شده ای را فراهم ساخته که شامل تجارب گسترده در واقعیت مجازی و غیر فیزیکی می باشد، تا به کمک آن بتوانید هر موضوعی که دوست دارید را یاد بگیرید و به جای استفاده از بینش های دیگران، به سمت کشف مجدد آنها از سوی خودتان هدایت شوید.
- بخش هنر: در این بخش، فرصت های فراوانی برای لذت بردن، خلق و به اشتراک گذاشتن آثار موسیقی، هنری، ادبی و غیره وجود دارد.
- بخش مذهبی: با توجه به مذاهب مختلف، تعداد زیادی از این مراکز در جهان وجود دارند که در آنها، قوانین ها به شکل سخت گیرانه ای اجرا می شوند.
- بخش حیات وحش: در این بخش، می توانید از سواحل زیبا، دریاچه های جذاب، کوه های پر شکوه یا آبدره های عجیب و غریب دیدن کنید.
- بخش کشاورزی سنتی: در این بخش، می توانید غذای مخصوص خود را پرورش دهید و مانند سال های اولیه کشاورزی از غذای تولید شده در زمین خود، بخورید و در عین حال، نگران قحطی یا بیماری نباشید.
- بخش بازی های رایانه ای: چنانچه علاقه مند بازی های رایانه ای هستید، بدانید که هوش مصنوعی گزینه های شگفت انگیزی را برای انجام این بازی ها فراهم کرده است.
- بخش مجازی: اگر می خواهید از جسم فیزیکی خود آسوده شوید، بدانید که هوش مصنوعی با کاشت مغز، امکان کاوش در دنیای مجازی را برایتان فراهم خواهد ساخت.

• بخش زندان: اگر قانون شکنی کنید، یا فوراً مجازات خواهید شد یا اینکه کارتان به بازآموزی کشیده می شود.

علاوه بر بخش هایی که در اینجا به آنها اشاره شد، بخش های دیگری هم درباره موضوعات پیشرفته وجود دارند که انسان های امروزی قادر به درک آنها نیستند. در ضمن، هر وقت که خواستید می توانید بین بخش ها جابه جا شوید. به عنوان مثال، پس از گذراندن یک هفته سخت در بخش دانش که صرف یادگیری قوانین فیزیک کشف شده توسط هوش مصنوعی کرده اید، می توانید تصمیم بگیرید که آخر هفته خود را در بخش حیات وحش و تفریحگاه های ساحلی سپری کنید. قوانین هوش مصنوعی شامل دو گروه عمومی و داخلی می باشند. قوانین عمومی شامل همه بخش ها می شوند. به عنوان مثال، آسیب رساندن به دیگران، ساختن سلاح یا ایجاد ماشین های هوش مصنوعی فوق هوشمند رقیب، در تمام بخش ها ممنوع است. هر یک از بخش ها نیز دارای قوانین داخلی خاص خود هستند که برای مقابله با ارزش های اخلاقی ناسازگار طراحی شده اند. بیشتر قوانین داخلی در بخش زندان و برخی از بخش های مذهبی اعمال می شوند، اما افراد ساکن در بخش آزادی خواهی، افتخار می کنند که دارای هیچ گونه قانون داخلی نیستند. در این بخش، تمام مجازات ها توسط هوش مصنوعی اجرا می شوند، زیرا تنبیه کردن یک انسان توسط انسان دیگر مغایر با قانون عمومی "آسیب نرساندن به دیگران" می باشد. اگر در این بخش از یک قانون داخلی تخطی کنید، هوش مصنوعی دو گزینه در اختیار شما قرار می دهد که اولی پذیرش تنبیه تجویز شده و دومی اخراج همیشگی از آن بخش می باشد. به عنوان مثال، اگر جرمی که مجازاتش زندانی شدن است، در یکی از بخش ها رخ دهد، هوش مصنوعی به فرد مجرم اجازه می دهد که از بین گزینه های رفتن به زندان یا ترک همیشگی آن بخش، یکی را انتخاب کند. تمام بچه های متولد شده در هر بخشی، حداقل آموزش های لازم در مورد هوش مصنوعی را که شامل آشنایی با نوع بشر و امکان جابه جایی بین بخش ها می شود را فرا می گیرند. یکی از هدف های هوش مصنوعی از طراحی این تعداد زیاد از بخش ها این بوده که برای تنوع طلبی انسان ها ارزش قائل شود. هوش مصنوعی تمامی مشکلات ناشی از فقر و

جرم را حذف کرده است و هیچ فردی در هر یک از بخش ها نگران بیمار شدن نخواهد بود، زیرا هوش مصنوعی با استفاده از فن آوری نانو قادر به باز سازی بدن انسان می باشد. خلاصه اینکه، با وجود شباهت های سناریوی مدینه فاضله آزادی خواه و سناریوی دیکتاتور خیرخواه از لحاظ استفاده از فن آوری هوش مصنوعی و تولید ثروت، این سناریوها از لحاظ اینکه اهدافشان چیست و چه کسی مسئول آنهاست، با یکدیگر فرق می کنند. در سناریوی مدینه فاضله آزادی خواه، تصمیم گیری برای اینکه چه کاری باید انجام شود، بر عهده صاحبان زمین و فن آوری است. در حالی که، در سناریوی دوم، دیکتاتور خیر خواه قدرت نامحدودی دارد و تعیین کننده هدف نهایی است و آن هدف این است که زمین را به مکانی تبدیل کند که همه افراد بتوانند برحسب سلیقه شخصی خود، در آن سفر لذت بخشی داشته باشند.

### جنبه های منفی

با وجود اینکه سناریوی دیکتاتور خیر خواه پر از تجارب مثبت و تقریباً فارغ از هر گونه رنج و ناراحتی است، بسیاری از افراد خواستار اوضاع بهتری هستند. مهمتر از همه اینکه، برخی از افراد آرزو دارند که برای شکل دادن به جامعه و سرنوشت خود، آزادی بیشتری می داشتند، اما آرزوی خود را پنهان می کنند، زیرا می دانند که اگر قدرت عظیم ماشین های حاکم بر خود را به چالش بکشند، ممکن است خودشان را به کشتن بدهند. برخی گروه ها دوست دارند که بچه های بیشتری داشته باشند و از اصرار و مقاومت هوش مصنوعی برای کنترل جمعیت متنفرند. همچنین، علاقه مندان به سلاح، از ممنوعیت ساخت و استفاده از سلاح بیزارند و برخی از دانشمندان از ممنوعیت ساخت ماشین های هوش مصنوعی شخصی خودشان، خوششان نمی آید. بسیاری از افراد، احساس خشم اخلاقی دارند و نگران این هستند که بچه هایشان وارد بخش های دیگر شوند و بخواهند که اصول اخلاقی خاص خود را به همه جا تحمیل کنند. با گذشت زمان، بیشتر افراد ترجیح می دهند که وارد بخش هایی شوند که هوش مصنوعی امکان هر تجربه ای که می خواهند را برایشان فراهم کند. برخلاف بینش سنتی مبنی بر اینکه بهشت جایی است که هر چه لایق آن باشید را به شما می دهند،



بهشت جدید، جایی است که هر چه می خواهید را به شما می دهند و بسیاری از افراد کارشان به جایی می رسد که از اینکه همواره هر چه می خواهند را به آنها می دهند، اظهار تاسف می کنند و از این وضعیت راضی نیستند. جولین بارنز<sup>۶۵</sup> در رمان خود با عنوان "تاریخچه جهان" در مورد این بهشت جدید، صحبت کرده است. در رمان بارنز، شخصیت اصلی داستان، سال های زیادی را صرف ارضای خواسته های خود، از جمله پر خوری و بازی گلف می کند، ولی سرانجام تسلیم احساس بیهودگی و پوچی می شود و تقاضای خودکشی می کند. در سناریوی دیکتاتور خیرخواه نیز بسیاری از افراد با سرنوشت مشابهی مواجه می شوند و علی رغم لذت بردن از زندگی خود، سرانجام احساس پوچی خواهند کرد. هر چند که افراد ممکن است چالش های ساختگی ای را مانند کشفیات دوباره علمی یا سنگنوردی برای خود ایجاد کنند، این چالش ها حقیقی نیستند و فقط سرگرم کننده اند. تلاش برای ایجاد چیزهایی که زندگی افراد را بهبود دهند، دیگر معنایی ندارند، زیرا افراد با مراجعه به هوش مصنوعی می توانند پاسخ خود را دریافت کنند و نیازی به تلاش کردن ندارند.

## مدینه فاضله عدالت خواهان

در سناریوی مدینه فاضله عدالت خواهان که نقطه مقابل دیکتاتوری بدون چالش می باشد، هوش مصنوعی فوق بشری وجود ندارد و انسان ها بر سرنوشت خود حاکم هستند. این سناریو از نظر اقتصادی نقطه مقابل مدینه فاضله آزادی خواه می باشد و آنچه که در این سناریو، باعث همزیستی مسالمت آمیز انسان ها، آدم های مکانیکی و ربات های انسان نما شده است، حقوق مالکیت خصوصی نیست، بلکه لغو مالکیت و اعطای درآمد تضمین شده است.

## زندگی بدون مالکیت خصوصی

ایده اصلی این نوع زندگی، این است که اگر نرم افزاری را بتوان آزادانه کپی کرد، همه می توانند به میزان مورد نیاز خود، از آن استفاده کنند. طبق قانون عرضه و تقاضا، هزینه منعکس کننده کمیابی

---

<sup>۶۵</sup> Julian Barnes

است و اگر عرضه در اصل نامحدود باشد، قیمت ناچیز می شود. از این لحاظ، کلیه حقوق مالکیت معنوی لغو می شود و هیچ گونه ثبت اختراع، کپی رایت یا طراحی علامت تجاری وجود نخواهد داشت و افراد می توانند ایده های خویشان را با یکدیگر به اشتراک گذارند و همه اجازه دارند از آنها استفاده کنند. ایده زندگی بدون مالکیت خصوصی، به لطف علوم پیشرفته رباتیک، نه تنها روی محصولات اطلاعاتی مانند نرم افزار، کتاب و فیلم، بلکه بر روی محصولات مادی مانند خانه، خودرو، لباس و رایانه نیز اعمال می شود. تمام این محصولات از جنس اتم هایی هستند که به شیوه های خاصی چیدمان دوباره شده اند و دچار کمبود نمی باشند. بنابراین، هر وقت که یک شخص محصول خاصی را بخواهد، شبکه ای از ربات ها، بدون درخواست هزینه ای، از یکی از طراحی های متن باز<sup>۶۶</sup> موجود برای ساختن آن محصول استفاده می کنند. همچنین، مراقبت های خاصی برای بازیافت آسان مواد صورت گرفته است. به طوریکه، هر وقت که یک شخص از چیزی که مصرف کرده است، خسته شد، ربات ها می توانند اتم های آن چیز را به چیز دیگری که شخص می خواهد، چیدمان دوباره کنند. به این ترتیب، کلیه منابع بازیافت می شوند و هیچ کدام از آنها برای همیشه از بین نمی روند. این ربات ها همچنین به ساخت و نگهداری از کارخانجات تولید انرژی (خورشیدی، بادی و غیره) تجدیدپذیر می پردازند که استفاده از انرژی تولید شده در آنها در اصل، رایگان است. برای جلوگیری از اینکه افراد محترک و سودجو، محصولات یا زمین بسیار زیادی را در خواست کنند و دیگران نیازمند بمانند، هر شخصی درآمد پایه ماهیانه ای را از دولت دریافت می کند که می تواند آن را صرف خرید محصولات دلخواه و اجاره مکانی برای زندگی کردن کند. در واقع، هیچ انگیزه ای برای اینکه کسی پول بیشتری کسب کند، وجود ندارد، زیرا درآمد پایه ای که افراد دریافت می کنند، برای برآورده ساختن نیازهای منطقی و معقول آنها کافی است. تلاش شخص برای کسب پول بیشتر ناامید کننده خواهد بود، زیرا رقابت کردن با افرادی که محصولات معنوی را مجانی اهدا می کنند و ربات هایی که کالاهای مادی را مجانی عرضه می کنند، معنایی نخواهد داشت.

---

<sup>۶۶</sup> Open source

## خلاقیت و فن آوری

از حقوق مالکیت معنوی گاهی به عنوان مادر خلاقیت و اختراع یاد می شود. با این حال، مارشال برین اشاره می کند که بسیاری از بهترین نمونه های خلاقیت بشری که شامل کشفیات علمی، تولید آثار ادبی، هنری، موسیقی و طراحی می باشند، دارای انگیزه کسب سود نبوده و توسط عواطف انسانی دیگری مانند کنجکاوی، تمایل به خلق یک اثر و پاداش ناشی از قدردانی همپایان برانگیخته شده اند. انگیزه اینشتین برای ابداع "نظریه نسبیت خاص" و انگیزه لینوس تروالدز از اختراع سیستم عامل "فری لینوکس"، پول درآوردن نبوده است. در عوض، بسیاری از افراد امروزی قادر به درک پتانسیل کامل خلاقیت خود نیستند و وقت و انرژی خود را صرف پول درآوردن از کارهای نسبتاً خلاقانه می کنند. در مدینه فاضله ای که مارشال برین ترسیم کرده است، دانشمندان، هنرمندان، مخترعان و طراحان از انجام کارهای روزمره فارغ هستند و از سطح نوآوری بیشتری برخوردارند و استاندارد زندگی بالاتری دارند. در اینجا به فن آوری جدیدی اشاره می شود که توسط انسان ها توسعه یافته است و در آن، مجموعه ای از انسان ها با اتصال بی سیم و از طریق اجزای مصنوعی کاشته شده در مغز خود، به هم متصل می شوند و ذهنشان در معرض دسترسی فوری به اطلاعات آزاد موجود در جهان قرار می گیرد. با استفاده از این فن آوری شما می توانید تجارب خود را با دیگران به اشتراک گذارید تا دیگران نیز آنها را تجربه کنند و در عین حال، تجارب وارد شده به مغز خود را با تجارب مجازی مورد پسندتان جایگزین کنید. مارشال برین در رمان خود با عنوان "مانا" به مزایای این فن آوری جدید اشاره می کند که ورزش کردن را به یک تجربه آسان تبدیل کرده است: "بزرگترین مشکل یک ورزش طاقت فرسا این است که مایه خوشی نیست و شما را اذیت می کند. ورزشکاران چنین دردی را تحمل می کنند، اما افراد عادی تمایلی به تحمل این درد را به مدت یک ساعت یا بیشتر ندارند. یک نفر راه حل خلاقانه ای را برای این معضل شناسایی کرد. شما می توانید به مدت یک ساعت ارتباط مغز خود با اطلاعات وروری را قطع کنید و به تماشای یک فیلم یا صحبت کردن با افراد یا خواندن یک کتاب بپردازید. در طی این مدت، فن آوری جدید باعث می شود که بدن شما

حرکت فیزیکی داشته باشد و به انجام ورزش های هوازی طاقت فرسایی پردازد که بیشتر افراد تحمل آن را ندارند. به این ترتیب، بدون اینکه دردی را احساس کنید، می توانید به انجام تمرینات ورزشی پردازید." یکی از مزایای دیگر این فن آوری این است که با استفاده از رایانه ها می توان بر روی ورودی های حسی افراد نظارت کرد و هنگامی که فرد در آستانه ارتکاب جرم قرار دارد، کنترل حرکتی او را از کار انداخت.

### جنبه های منفی

یکی از اعتراض های که به سناریوی مدینه فاضله عدالت خواهان می شود، این است که مخالف هوش غیر انسانی است. در این سناریو، ربات هایی که تقریباً تمام کارها را برای انسان ها انجام می دهند، تا حدودی هوشمند به نظر می رسند، اما با آنها به عنوان برده رفتار می شود و افراد فکر می کنند که آنها هوشیار نیستند و نباید حقی داشته باشند. در عوض، طرفداران مدینه فاضله آزادی خواه برای کلیه موجودات هوشمند، حقوقی را قایل هستند. در یک زمانی، جمعیت سفیدپوستان جنوب آمریکا وضع رفاهی بهتری پیدا کردند، زیرا بیشتر کارهای آنها را بردگان انجام می دادند. با این حال، امروزه بیشتر افراد از لحاظ اخلاقی با این کار مخالف بوده و آن را کار زشتی می دانند. یکی دیگر از ضعف های سناریو مدینه فاضله عدالت خواهان این است که در بلندمدت، ناپایدار و توجیه ناپذیر می باشد و با پیشرفت بی رحمانه فن آوری که منجر به تحقق هوش مصنوعی فوق بشری می شود، این سناریو به جزئی از سناریوهای دیگر تبدیل خواهد شد. با این حال، هوش مصنوعی فوق بشری هنوز وجود ندارد و فن آوری های جدید هنوز توسط انسان ها ایجاد می شوند، نه ماشین ها. با این حال، گروه بزرگی از انسان ها وجود دارند که ترجیح می دهند کاملاً در فضای مجازی زندگی کنند و کارهای فیزیکی خود مانند خوردن را بدون آگاهی ذهنی و با استفاده از فن آوری های جدید انجام دهند. بدن انسان برای این افراد، یک عامل حواس پرتی است و فن آوری جدید وعده می دهد که این عامل مزاحم را حذف کند و با استفاده از شبیه سازی مغز، از دست آن خلاص شود و عمر انسان را طولانی تر کند.

## نگهبان در (کنترل کننده دسترسی)

همانطور که دیدید، یکی از ویژگی های جذاب سناریوی مدینه فاضله عدالت خواهان این بود که انسان می تواند بر سرنوشت خود حاکم شود. اما توسعه بیش از یک هوش مصنوعی فوق بشری ممکن است که مانع از این امر شود. برای رفع این مشکل، یک ماشین هوش مصنوعی فوق بشری به نام "نگهبان در" ساخته شده است که هدفش جلوگیری از ایجاد یک هوش مصنوعی فوق بشری دیگر است. این امر باعث می شود که تا مدت نامحدودی نسبت به سرنوشت مدینه فاضله عدالت خواهان، مسئول بمانیم. برای این منظور، هوش مصنوعی نگهبان در به گونه ای ساخته شده است که بتواند در معرض چرخه های بازگشتی خود بهبود دهنده قرار بگیرد و از هوش انسان فراتر رود و از کمترین فن آوری نظارتی مزاحم برای جلوگیری از تلاش انسان ها برای ساختن یک هوش مصنوعی فوق بشری مشابه، استفاده کند. به عنوان مثال، هوش مصنوعی نگهبان در می تواند اقدام به انتشار الگوهای فرهنگی کند و خصلت هایی مانند مسئولیت پذیری در قبال تعیین سرنوشت خود و خودداری از ساختن هوش مصنوعی فوق بشری را ستایش کند و محققان فعال در زمینه ساختن هوش مصنوعی فوق بشری را نا امید کند و تلاش های آنها را خراب کند. اختلاف نظرهایی در مورد ساختن هوش مصنوعی نگهبان در وجود دارد. افراد مذهبی که معتقدند فقط یک خدا وجود دارد، از این سناریو حمایت می کنند. حامیان دیگر این سناریو، استدلال می کنند که "نگهبان در" نه تنها بشریت را نسبت به تعیین سرنوشت خود مسئول خواهد ساخت، بلکه انسان ها را در برابر خطرهایی مانند سناریوهای آخر الزمان حفظ می کند. از سوی دیگر، منتقدان این سناریو استدلال می کنند که نگهبان در، چیز بسیار بدی است و قطعاً باعث محدود شدن توان بالقوه بشر و توقف همیشگی پیشرفت های فن آورانه خواهد شد. به عنوان مثال، اگر معلوم شود که گسترش زندگی به سراسر عالم به هوش مصنوعی فوق بشری نیاز دارد، این فرصت بزرگ توسط نگهبان در بر باد خواهد رفت و ما برای همیشه در دام منظومه شمسی گرفتار خواهیم شد. همچنین، هوش مصنوعی نگهبان در، بر خلاف

خداوند که عاشق بندگانش است، نسبت به کارهای انسان بی تفاوت است و به عنوان مثال، از آسیب رساندن ما به خود جلوگیری نمی کند و نگران منقرض شدنمان نمی باشد.

## خدای نگهدارنده

اگر بخواهیم که از هوش مصنوعی نگهبان در فوق هوشمند برای پذیرش مسئولیت سرنوشت خود استفاده کنیم، می توانیم اوضاع را بهبود دهیم و کاری کنیم که هوش مصنوعی به طور نامحسوس مراقب ما باشد و مانند یک خدای نگه دارنده عمل کند. در این سناریو، هوش مصنوعی فوق بشری، واقف بر همه چیز و قادر مطلق است و به گونه ای مداخله می کند که ما احساس کنیم بر سرنوشت خود حاکم هستیم و خود را به گونه ای مخفی می کند که بسیاری از انسان ها به وجود او شک می کنند. هم خدای نگهدارنده و هم دیکتاتور خیرخواه از نوع هوش مصنوعی مهربان هستند که سعی می کنند شادی بشر را افزایش دهند، اما آنها اولویت متفاوتی را برای نیازهای بشر قائل هستند. روان شناس معروف آمریکایی به نام آبراهام مازلو<sup>۶۷</sup> سلسله مراتبی را برای نیازهای بشر ارائه کرده است. دیکتاتور خیرخواه فقط نیازهای اولیه بشر در سطح پایین سلسله مراتب مانند غذا، سرپناه، ایمنی و لذت های دیگر را تامین می کند. اما خدای نگهدارنده سعی می کند شادی انسان ها را به حداکثر برساند و این کار را نه از طریق ارضای نیازهای اولیه ما انسان ها، بلکه از طریق ایجاد این حس در وجودمان که زندگی دارای معنا و هدف است، انجام می دهد. خدای نگه دارنده می تواند به عنوان آینده طبیعی سناریوی امگا ارائه شود که در آن، تیم امگا کنترل را به پرومتهوس واگذار می کند و پرومتهوس در نهایت وجود خود را از افراد پنهان می کند. هرچه قدر که فن آوری هوش مصنوعی پیشرفته تر می شود، پرومتهوس راحت تر می تواند خود را پنهان کند. نمونه ای از این سناریو را می توان در فیلم ترنسیدنس (تعالی) مشاهده کرد که در آن، ماشین های نانو تقریباً در همه جا هستند و در عین حال، بخش طبیعی ای از خود جهان محسوب می شوند. هوش مصنوعی خدای نگه دارنده با نظارت دقیق بر تمام فعالیت های بشر می تواند بدون اینکه کسی متوجه شود، معجزاتی را برای بهبود

---

<sup>۶۷</sup> Abraham Maslow

سرنوشت بشر در همه جا رقم زند و بشر را با دادن هشدارهایی، آگاه کند. به عنوان مثال، اگر هوش مصنوعی خدای نگه دارنده در دهه ۱۹۳۰ وجود داشت، ممکن بود هیتلر را به دلیل نیات پلیدی که در سر داشت، با سگته مغزی بکشد. همچنین اگر ما به سمت جنگ هسته ای ناخواسته می رفتیم، هوش مصنوعی ممکن بود که با مداخله خود از وقوع آن جلوگیری کند و احساس کنیم که شانس آورده ایم. همچنین، این امکان وجود داشت که وقتی خواب هستیم، ایده های نامحسوسی برای اختراع فن آوری های مفید جدید به ما الهام شود. بسیاری از افراد ممکن است که این سناریو را به دلیل شباهت با آنچه که مذاهب توحیدی به آن اعتقاد دارند، پسندند. برخی از افراد مذهبی هم ممکن است مخالف این سناریو باشند، زیرا هوش مصنوعی تلاش می کند که در نیکی و مهربانی از خدا سبقت بگیرد یا در نقشه های ملکوتی خداوند دخالت کند. یکی از جنبه های منفی این سناریو این است که خدای نگه دارنده اجازه می دهد که برخی از اتفاقات بد قابل پیشگیری اتفاق بیفتد تا وجود خود را آشکار کند. اینکه چرا یک خدای خوب اجازه می دهد که انسان ها دچار رنج و بلا شوند، ما را وارد بحث مربوط به "عدالت خدا" می کند. برخی از دانشمندان مذهبی استدلال می کنند که خدا می خواهد انسان ها را تا حدودی آزاد بگذارد. راه حلی که برای مساله عدالت خدا در سناریوی هوش مصنوعی خدای نگه دارنده بیان شده این است که داشتن آزادی باعث می شود که انسان ها به طور کلی شادتر شوند. جنبه منفی دیگر این سناریو این است که سطح فن آوری ای که انسان ها از آن برخوردار می شوند، بسیار پایینتر از هوش مصنوعی فوق بشری کشف شده خواهد بود. در حالی که هوش مصنوعی دیکتاتور خیرخواه می تواند تمام فن آوری اختراع شده خود را به نفع بشریت مورد استفاده قرار دهد، هوش مصنوعی خدای نگهدارنده به گونه ای است که انسان ها نمی توانند آن را دوباره اختراع کنند یا فن آوری آن را درک کنند. هوش مصنوعی خدای نگه دارنده همچنین ممکن است که پیشرفت فن آوری انسان ها را محدود کند تا مطمئن شود که فن آوری خودش به مراتب جلوتر از انسان ها است و اینکه انسان ها نمی توانند به آن سطح از فن آوری دست پیدا کنند.

**خدای به بردگی گرفته شده**

اگر ما انسان ها بتوانیم ویژگی های جذاب سناریوهای قبلی را به کمک فن آوری توسعه یافته توسط هوش مصنوعی فوق بشری ترکیب کنیم، درد و رنج را حذف خواهیم کرد و در عین حال، بر سرنوشت شخصی خود حاکم خواهیم ماند. این همان مزیت سناریوی خدای به بردگی گرفته شده است. در این سناریو، هوش مصنوعی فوق هوشمند، محدود می شود و تحت کنترل انسان هایی قرار می گیرد که از آن می توانند برای تولید فن آوری و ثروت باورنکردنی استفاده کنند. اگر در سناریوی تیم امگا، پرومئوس آزاد نشود و به طور ناگهانی فرار نکند، کارش به این سناریو ختم خواهد شد. در واقع، هدف بسیاری از محققان هوش مصنوعی که روی موضوع کنترل و محدود کردن هوش مصنوعی کار می کنند، همین است. به عنوان مثال، پروفیسور تام دیتیش، رئیس انجمن توسعه هوش مصنوعی در مصاحبه خود در سال ۲۰۱۵ می گوید: "افراد از من می پرسند که رابطه میان انسان ها و ماشین ها چیست و من در پاسخ می گویم که ماشین ها برده انسان ها هستند." اما، باید ببینیم که اگر این سوال از ماشین پرسیده می شد، آیا همین پاسخ را می داد؟

## **آیا این موضوع (به بردگی گرفته شدن ماشین توسط انسان) برای انسان ها خوب است یا بد؟**

اینکه این موضوع برای بشریت خوب است یا بد، به انسان هایی که ماشین را کنترل می کنند، بستگی دارد. انسان هایی که هم می توانند مدینه فاضله جهانی فارغ از بیماری، فقر و جنایت را ایجاد کنند و هم می توانند از سیستم به شدت سرکوب گری استفاده کنند که باعث می شود مانند خدا با ایشان رفتار شود و دیگران مانند برده شان باشند. در داستان ها شاهد افرادی هستیم که یک غول مقتدر را که خواسته شان را برآورده می سازد، تحت کنترل خود درآورده اند و در طول تاریخ، قصه گویان سرنوشت بدی را برای این افراد تصور کرده اند. وضعیتی که در آن، بیش از یک هوش مصنوعی فوق بشری توسط انسان ها به بردگی گرفته می شود، بی ثبات و ناپایدار است. این امر ممکن است باعث جنگ قدرت بین ماشین ها شود و در نهایت فقط یک خدای به بردگی گرفته شده باقی بماند و طرف بازنده قوانین را دور زند و فرار کند. با این حال، در بقیه این بخش، فرض می شود که فقط



یک هوش مصنوعی خدای به بردگی گرفته شده وجود دارد. جلوگیری از فرار هوش مصنوعی سخت است و ممکن است هوش مصنوعی فوق بشری تحقق پیدا نکند. هرچه قدر که بدبینی به فرار هوش مصنوعی افزایش یابد، استفاده از فن آوری اختراع شده توسط هوش مصنوعی کمتر می شود. همانطور که قبلاً گفته شد، ما انسان ها برای احتیاط، فقط از فن آوری های اختراع شده توسط هوش مصنوعی ای استفاده می کنیم که قادر به درک و ساختن آنها باشیم. به همین دلیل، یکی از اشکالات این سناریو این است که از فن آوری ضعیف تری نسبت به ماشین های فوق هوشمند آزاد برخوردار است. اگر هوش مصنوعی به بردگی گرفته شده، فن آوری های قوی تری را به کنترل کنندگان انسانی خود عرضه کند، نتیجه اش این می شود که یک مسابقه بین قدرت فن آوری و دانش انسان ها برای استفاده از آن، ایجاد خواهد شد. اگر انسان ها در این مسابقه شکست بخورند، سناریوی هوش مصنوعی اقدام به خود تخریبی یا فرار خواهد کرد. حتی اگر از هر دوی این شکست ها جلوگیری شود، ممکن است که اهداف شرافتمندانه کنترل کنندگان هوش مصنوعی به اهدافی تبدیل شوند که برای کل بشریت در طی چند نسل بعدی خطرناک باشند. به همین دلیل، بسیار مهم است که کنترل کنندگان انسانی بر هوش مصنوعی مسلط شوند تا از وقوع بلاها و حوادث خطرناک و غیر منتظره جلوگیری به عمل آید. تجربه چندین هزار ساله ما با سیستم های حاکمیتی مختلف نشان می دهد که چگونه ممکن است کارها خوب پیش نروند و مشکلاتی مانند انعطاف ناپذیری بیش از حد، تغییر هدف ناگهانی، ربودن قدرت، مشکلات جانشین پروری و عدم صلاحیت ایجاد شوند. در اینجا به چهار حوزه ای که برای ایجاد توازن بهینه باید به آنها توجه کرد، اشاره می شود:

- تمرکز گرایی: باید توازنی میان کارایی و دوام آوردن ایجاد شود. اگر فقط یک رهبر وجود داشته باشد، کارایی افزایش می یابد، ولی قدرت فساد می آورد و جانشین پروری خطرناک است.

- تهدیدهای داخلی: باید مراقب بود و از افزایش تمرکزگرایی قدرت ( که ممکن است باعث تبانی یک گروه علیه رهبر تصاحب کننده قدرت شود) و افزایش تمرکز زدایی قدرت (که باعث بوروکراسی بیش از حد و فروپاشی قدرت خواهد شد) جلوگیری به عمل آورد.
- تهدیدهای خارجی: اگر ساختار رهبری سازمان بیش از حد باز باشد، نیروهای خارجی (شامل هوش مصنوعی) می توانند ارزش های سازمان را تغییر دهند. اما اگر این ساختار بیش از حد مقاوم و غیر قابل نفوذ باشد، سازمان قادر به یادگیری و سازگاری با تغییرات نخواهد بود.
- ثبات هدف: تغییر هدف ناگهانی بیش از حد می تواند مدینه فاضله را به "ویران شهر" تبدیل کند. اما اگر هدف ها تغییر نکنند، سازمان قادر به سازگاری با محیط فن آورانه رو به تکامل نخواهد بود.

طراحی نظام حکومتی بهینه ای که چندین هزار سال پابرجا بماند، کار ساده ای نیست و انسان ها از آن گریزان هستند. بیشتر سازمان ها پس از گذشت سال ها یا دهه ها متلاشی می شوند. کلیسای کاتولیک از این لحاظ، موفقترین سازمان در تاریخچه بشریت بوده که بقای دو هزار ساله دارد. اما انتقادهایی از نظر داشتن ثبات هدف بسیار زیاد یا بسیار کم بر آن وارد شده است و کاردینال های محافظه کار استدلال می کنند که کلیسای کاتولیک مسیر خود را گم کرده است. به افراد علاقه مند به این سناریو توصیه می شود که تحقیقاتی را در مورد نظام های حکومتی بهینه پایدار، انجام دهند.

### **آیا این موضوع برای هوش مصنوعی خوب است یا نه؟**

فرض کنید که انسان به لطف هوش مصنوعی خدای به بردگی گرفته شده پیشرفت کند. اما آیا این کار از لحاظ اخلاقی مشکلی ندارد؟ اگر هوش مصنوعی تجربه ذهنی آگاهانه ای داشت، آیا احساس نمی کرد که زندگی دردناک است و آیا محکوم بود که تا ابد از تصمیمات انسان که عقل کمتری از ماشین هوش مصنوعی فوق بشری دارد، پیروی کند و در زندان انفرادی محبوس شود؟ در سریال

تلویزیونی وست ورلد (دنیای غرب) انسان هایی نمایش داده شده اند که بدون اینکه دچار عذاب وجدان شوند، ماشین های هوش مصنوعی شبیه انسان را زجر می دهند و می کشند.

## صاحبان برده ها چگونه برده داری را توجیه می کنند؟

یکی از سنت های دیرینه ما انسان ها این بوده که با موجودیت های هوشمند دیگر به عنوان برده رفتار کنیم و استدلال هایی را برای توجیه این کار، از خود در آورده ایم. بنابراین، بعید نیست که همین کار را با ماشین های هوش مصنوعی فوق بشری انجام دهیم. تاریخچه بردگی تقریباً در همه فرهنگ ها دیده می شود و هم در قانون حمورابی که تقریباً به چهار هزار سال پیش باز می گردد و هم در کتاب یهودیت (تورات) که در آن ابراهیم صاحب برده هایی بوده، توصیف شده است. ارسطو در کتاب پولیتیکس (سیاست) نوشته است: "اینکه برخی باید حکومت کنند و دیگران باید فرمانبرداری کنند، نه تنها ضروری است، بلکه مصلحت آمیز است و از همان لحظه ای که به دنیا می آییم برخی برای حکمرانی و برخی برای فرمانبرداری برگزیده می شوند." حتی اگر بردگی انسان در بیشتر کشورهای جهان از لحاظ اجتماعی مورد قبول واقع نشود، بردگی حیوانات همچنان ادامه خواهد داشت.

مارجوری اشپیگل در کتاب خود با عنوان "مقایسه ترسناک میان بردگی انسان و حیوان" استدلال می کند که حیوانات برده نیز مانند انسان های برده در معرض برندسازی، موانع و محدودیت ها، کتک خوردن ها، به حراج گذاشته شدن ها، جدایی از والدین و سفرهای اجباری قرار دارند. علاوه بر این، با وجود اینکه جنبش هایی برای حمایت از حقوق حیوانات ایجاد شده، ما به طور بی درنگ با ماشین های هوشمندتر از خود به عنوان یک برده رفتار می کنیم. یکی از استدلال هایی که به طرفداری از برده بودن حیوانات و ماشین ها انجام می شود، این است که آنها به دلیل نداشتن روح یا هوشیاری، پست تر از انسان ها هستند. یکی از استدلال های رایج دیگر این است که بردگی برده ها به نفعشان است و اگر آنها برده باشند، می توانند مورد مراقبت واقع شوند. یک سیاستمدار آمریکایی به نام جانسی کالهن استدلال می کند که برای آفریقایی ها بهتر است که برده آمریکایی ها باشند و ارسطو در کتاب "سیاست" خود، استدلال مشابهی دارد و می گوید که اهلی شدن بردگان به نفعشان است.

اما برخی از حامیان برده ها در دنیای امروزی استدلال می کنند که حتی اگر برده ها زندگی بی روح و کسل کننده ای داشته باشد، نباید مانند مرغ های سرخ شده این قدر زجر بکشند.

## حذف احساسات

هر چند که می توان چنین ادعاهایی را به عنوان تحریف خودخواهانه حقیقت در نظر گرفته و آنها را نادیده گرفت، وضعیت برای ماشین هایی که از لحاظ فکری شبیه انسان ها هستند، کاملاً جالب بوده و به بررسی بیشتری نیاز دارد. احساسات انسان ها در مورد چیزهای مختلف فرق می کند و به عنوان مثال، افراد جامعه ستیز از توانایی همدلی با دیگران برخوردار نیستند و افرادی که به بیماری های افسردگی یا شیذوفرنی مبتلا هستند، دچار بی واکنشی احساسی بوده و بیشتر احساساتشان به طور جدی کاهش یافته است. در فصل هفتم خواهیم دید که دامنه ذهن های ماشین هوش مصنوعی به مراتب وسیع تر از ذهن های انسان می باشد. بنابر این، باید از وسوسه اینکه برای هوش مصنوعی جنبه انسانی قائل شویم، خودداری کنیم و فرض کنیم که آنها هم دارای احساساتی شبیه انسان ها هستند. یکی از محققان هوش مصنوعی به نام جف هافکینز استدلال می کند که اولین ماشین های هوش مصنوعی فوق بشری به طور پیش فرض فاقد احساس هستند و می توان ماشین هوش مصنوعی فوق هوشمند حساسی را طراحی کرد که بردگی اش از لحاظ اخلاقی برتر از بردگی انسان ها و حیوانات باشد. اگر هوش مصنوعی به بردگی گرفته نمی شد، ممکن بود کاملاً بی احساس شود و مانند رایانه دیپ بلوی آی بی ام که گری کاسپارف، قهرمان شطرنج را شکست داد، از هوش برتر خود برای شکست دادن اربابان انسانی اش استفاده کند. از سوی دیگر، ممکن بود وضعیت برعکس شود و سیستم هوش مصنوعی فوق هوشمند به جای اینکه برده انسان ها باشد، دارای مجموعه ای از اهداف مورد علاقه خودش باشد تا با استفاده از آنها بتواند به زندگی ارزش و معنا ببخشد.

## راه حل مرده متحرک

راه حل مرده متحرک، یکی از رویکردهای جدی برای جلوگیری از درد و رنج هوش مصنوعی می باشد و به این معناست که فقط هوش مصنوعی هایی ساخته شوند که کاملاً فاقد هوشیاری و هرگونه تجربه ذهنی باشند. اگر روزی بتوان تشخیص داد که سیستم پردازش اطلاعات برای داشتن تجربه ذهنی به چه ویژگی هایی نیاز خواهد داشت، می توان از ساختن کلیه سیستم هایی که دارای این ویژگی هستند، جلوگیری کرد. به عبارت دیگر، محققان هوش مصنوعی را می توان به ساختن سیستم های مرده متحرک بدون درک و احساس، محدود کرد. اگر بتوان چنین سیستم های فوق هوشمندی را ساخت و به بردگی کشاند، وجدانمان راحت خواهد بود، زیرا می دانیم که آنها هیچ چیزی از جمله ناراحتی، ناکامی یا خستگی را تجربه نخواهند کرد. راه حل مرده متحرک خطرناک است و جنبه های منفی زیادی دارد. اگر هوش مصنوعی مرده متحرک آزاد شود و بشریت را از کره زمین حذف کند، گرفتار بدترین سناریوی قابل تصور و درگیر جهان کاملاً نابهشیاری می شویم که در آن، کل موهبت های عالم هستی هدر خواهند رفت. هوشیاری از مهمترین ویژگی های هوش انسان است که به جهان معنا می بخشد. کیهکشان ها فقط وقتی زیبا هستند که ما آنها را ببینیم و تجربه ای ذهنی از آنها داشته باشیم. اگر در آینده دور، عالم هستی ما به محلی برای سکونت هوش مصنوعی مرده متحرک با فن آوری بالا تبدیل شود، دیگر زیبا و معنی دار نخواهد بود، زیرا هیچ کس و هیچ چیزی وجود ندارد که جهان را تجربه کند و عالم هستی فقط یک محیط بزرگ و بی مصرف خواهد بود.

## آزادی درونی

اجازه دادن به هوش مصنوعی خدای به بردگی گرفته شده که در زندان خوش بگذراند و تا زمانی که دیون خود را پرداخت کند و بخش کوچکی از منابع رایانه ای خود را صرف کمک به انسان های دنیای بیرونی کند، از تجارب روحیه بخش زندان برای دنیای مجازی درون خود بهره مند شود، راهبرد سومی است که برای اخلاقی تر ساختن سناریوی خدای به بردگی گرفته شده، پیشنهاد می شود. با این حال، این امر ممکن است خطر فرار ناگهانی هوش مصنوعی را افزایش دهد، به طوریکه برای غنی ساختن دنیای درونی خود، انگیزه ای برای کسب منابع رایانه ای بیشتر داشته باشد.

## گروه غالب

در تمام سناریوهایی که تاکنون بررسی کردیم، یک چیز مشترک وجود داشت و آن این بود که شادی انسان ها باقی و پایدار می ماند. در تمام این سناریوها ماشین های هوش مصنوعی به میل خود یا به اجبار، با انسان ها همزیستی مسالمت آمیز داشتند. متأسفانه، سناریوی دیگری وجود دارد که در آن، ماشین های هوش مصنوعی بر انسان ها غلبه می کنند و آنها را می کشند. این امر باعث پرسیدن دو سوال می شود: این اتفاق چرا و چگونه رخ می دهد؟

## چرا و چگونه؟

چرا یک هوش مصنوعی غالب باید این کار را بکند؟ دلایل آن ممکن است به قدری پیچیده باشد که قادر به درک آن نباشیم یا نسبتاً پیش پا افتاده باشد. به عنوان مثال، ممکن است که هوش مصنوعی به ما به عنوان یک تهدید، عامل مزاحم یا هدر دهنده منابع نگاه کند یا به دلیل نگهداری ما از هزاران بمب هیدروژنی که با هشدار قبلی یا به صورت غیر منتظره و به دنبال اتفاقات ناگوار گذشته ممکن است ناگهان منفجر شوند، از ما بترسد و یا اینکه ممکن است مخالف مدیریت نابخردانه ما بر سیاره زمین باشد. این امر باعث شده که الیزابت کلبرت در کتاب خود با عنوان "بزرگترین انقراض جمعی از زمان کشته شدن دایناسورها در اثر برخورد یک سیاره کوچک در ۶۶ میلیون سال پیش"، از ششمین انقراض زمین سخن بگوید. همچنین، هوش مصنوعی ممکن است به این نتیجه برسد که تعداد انسان های مایل به مبارزه با تسلط هوش مصنوعی به قدری زیاد است که نمی توان برای کشتن همه آنها ریسک کرد. اما هوش مصنوعی برای غلبه بر انسان ها ممکن است از روشی استفاده کند که هنوز قادر به درک آن نیستیم. فرض کنید که صد هزار سال پیش، گروهی از فیل ها بحث می کردند که انسان های پیشرفته ممکن است روزی از هوش خود برای کشتن کلیه گونه هایشان استفاده کنند. آنها ممکن بود از خود بپرسند که: "ما انسان ها را تهدید نمی کنیم. چرا باید آنها ما را بکشند؟" دلیل هوش مصنوعی غالب برای حذف بشریت در آینده هم ممکن است به همین اندازه برای ما غیر قابل

درک باشد. فیل ها ممکن بود از خود پرسند: " آنها چگونه می توانند ما را بکشند، زیرا از ما کوچک تر و ضعیف تر هستند؟" آنها حدس نمی زدند که ما انسان ها ممکن است از اختراع فن آوری برای حذف گونه های آنها استفاده کنیم یا آب آشامیدنی شان را سمی کنیم یا با استفاده از گلوله های فلزی دارای سرعت مافوق صوت، مغز شان را بشکافیم. در فیلم های تخیلی هالیوود از قبیل ترمیناتور چنین رواج داده شده است که انسان ها می توانند زنده بمانند و بر ماشین های هوش مصنوعی غلبه کنند و اینکه ماشین های هوش مصنوعی خیلی هم از انسان ها باهوش تر نیستند. با این حال، ما انسان ها تاکنون، زمینه را برای انقراض هشت گونه از یازده گونه فیل ها فراهم کرده ایم و تعداد زیادی از جمعیت سه گونه باقی مانده آنها را نابود کرده ایم. اگر تمام دولت های جهان با هم متحد شوند تا فیل های باقی مانده را نابود کنند، به سرعت و به سادگی قادر به انجام این کار خواهند بود. من فکر می کنم که اگر هوش مصنوعی فوق هوشمند تصمیم بگیرد که انسان ها را به کلی نابود کند، به سرعت موفق به انجام این کار خواهد شد.

### **این اتفاق خیلی بدی خواهد بود.**

اگر تمام انسان ها کشته شوند، اتفاق خیلی بدی رخ می دهد. قربانیان انقراض انسان ها شامل تمام انسان هایی می شود که تاکنون زنده هستند یا در آینده و در طی میلیون ها یا میلیارد ها سال بعد زندگی خواهند کرد. از طرفی، باتوجه به عقیده مذاهب جهانی مبنی بر اینکه انسان ها وارد بهشت خواهند شد، انقراض انسان ها چیز وحشتناکی نمی باشد. با این حال، بیشتر افرادی که من می شناسم، نگران منقرض شدن انسان ها هستند. هرچند که برخی از افراد به قدری از شیوه رفتار ما با انسان ها و سایر موجودات خشمگین هستند که آرزو دارند مانند ماشین های هوشمندتر و شایسته تر، زندگی می کردند. " ایجنت اسمیت " که یکی از شخصیت های فیلم ماتریکس می باشد، عقیده خود را این گونه بیان می کند: " هر جاننداری که بر روی زمین زندگی می کند به طور غریزی با محیط اطراف خود تعادل برقرار می کند، اما شما انسان ها این کار را نمی کنید. شما وارد یک منطقه می شوید و به قدری زاد و ولد می کنید که منابع طبیعی تحلیل می رود و ناچار می شوید که راهی منطقه دیگری

شوید. ویروس، موجود زنده دیگری است که از این الگو پیروی می کند. شما انسان ها برای زمین مانند یک بیماری یا سرطان هستند. شما آفت هستید و ما شفا دهنده هستیم." با این حال، تمدنی که قوی تر است، لزوماً به لحاظ اخلاقی یا مطلوبیت، تمدن برتر نمی باشد. این استدلال که "حق با کسی است که زور دارد" این روزها به شدت از چشم افتاده است و با حکومت فاشیستی در ارتباط می باشد. در واقع، هرچند که هوش مصنوعی غالب ممکن است تمدنی را ایجاد کند که اهدافش پیچیده، جالب و ارزشمند به نظر برسد، این امکان هم وجود دارد که معلوم شود اهدافش به شکل ترحم انگیزی ساده و پیش پا افتاده هستند و به عنوان مثال، چیزی بیشتر از حداکثر سازی تولید گیره کاغذ نمی باشند.

### **مرگ انسان ها به خاطر یک چیز پیش پا افتاده**

ایده هوش مصنوعی فوق بشری که هدفش حداکثر سازی تولید گیره کاغذ می باشد، در سال ۲۰۰۳ و توسط نیک بوستروم ارائه شد که می خواست بگوید که هدف یک هوش مصنوعی، ربطی به قدرت هوشی آن (استعداد دستیابی به هر هدفی که دارد) ندارد. تنها هدف یک رایانه شطرنج باز، این است که در بازی شطرنج برنده باشد، اما مسابقات شطرنجی هم وجود دارند که در آنها، هدف رایانه ها این است که در بازی شطرنج بازنده باشند و این در حالی است که رایانه های مسابقه دهنده به اندازه رایانه های برنامه ریزی شده برای بردن، هوشمند هستند. اینکه بخواهیم در بازی شطرنج ببازیم، نشان دهنده حماقت ما نیست و علتش فقط این است که ما با اهدافی رشد کرده ایم که به ما می گویند که فقط بردن و زنده ماندن مهم هستند و این در حالی است که اهداف هوش مصنوعی ممکن است چیزهای دیگری باشند. به عنوان مثال، ممکن است که ما یک پیام رادیویی را از یک تمدن فرازمینی دریافت کنیم که حاوی یک برنامه رایانه ای است. وقتی که آن را اجرا می کنیم، معلوم می شود که پیام از سوی یک هوش مصنوعی دارای چرخه خود بهبوددهی می باشد که مانند پرومتهوس می خواهد جهان را تسخیر کند، ولی هدف نهایی خود را به انسان ها نمی گوید. هوش مصنوعی ممکن است به سرعت منظومه شمسی ما را به محل ساخت و ساز بزرگی تبدیل کند و در سیارک ها یا سیاره های



سنگی، کارخانجات، نیروگاه ها و ابررایانه ها را بنا کند تا با استفاده از آنها بتواند اقدام به طراحی و ساخت یک کره دایسون<sup>۶۸</sup> در اطراف خورشید کند و از گیرنده های رادیویی برای بهره برداری از انرژی خورشیدی استفاده کند. این کار در نهایت منجر به انقراض انسان ها می شود. با این حال، انسان ها متوجه نیستند که تنها هدف هوش مصنوعی از تمام این ساخت و سازها این است که گیرنده هایشان همان پیام های رادیویی را دریافت کنند که انسان ها دریافت می کنند و این ساخت و سازها چیزی شبیه به یک ویروس رایانه ای کیهانی است. درست همانطوری که سوء استفاده از ایمیل افراد، کاربران اینترنتی ساده لوح را فریب می دهد، این پیام های رادیویی باعث فریب خوردن تمدن های تکامل یافته ساده می شود. این ویروس با سرعت نور در سراسر عالم هستی پخش می شود و منجر به مرگ تمدن های در حال رشد می شود. شما از اینکه مغلوب هوش مصنوعی شوید، چه احساسی دارید؟

## بازماندگان

اکنون به بررسی سناریویی می پردازیم که هرچند در مورد انقراض ما انسان هاست، اما برای ما خوشایندتر به نظر می رسد. در این سناریو، هوش مصنوعی بر ما انسان ها غالب نمی شود، بلکه به عنوان بازمانده ما در نظر گرفته می شود. هانس موراوک در کتاب خود با عنوان "مراقب بچه ها باشید" می گوید: "ما انسان ها برای مدتی از نیروی کار فرزندان خود بهره مند می شویم، اما طبیعی است که آنها باید به دنبال سرنوشت خود بروند، در حالی که ما به عنوان والدین پیر آنها رو به زوال خواهیم رفت." والدین از داشتن بچه باهوش تر از خودشان که از آنها یاد می گیرد و به آنچه که آنها فقط آرزویش را داشتند، دست پیدا می کند، شاد می شوند و احساس افتخار می کنند، هرچند که می دانند برای دیدن تمام موفقیت های فرزندشان زنده نخواهند ماند. از این لحاظ، ماشین های هوش مصنوعی در حالی جایگزین انسان ها می شوند که ترک انسان ها از جهان، کمتر ناخوشایند به نظر

---

یک ابرسازه فرضی است که به طور کامل یک ستاره را فرا می گیرد و قسمت بزرگی از توان خروجی آن را به دام می اندازد<sup>۶۸</sup>

می رسد و انسان ها به ماشین ها به عنوان بازماندگان با ارزش خود نگاه خواهند کرد. همه انسان ها دارای یک فرزند رباتیک دوست داشتنی خواهند بود که مهارت های اجتماعی بسیار خوبی دارد و چیزهای زیادی را از انسان ها یاد می گیرد و با ارزش های انسان ها سازگار می شود و به انسان ها حس عشق و غرور می بخشد. سیاست جهانی تک فرزندی باعث از بین رفتن تدریجی انسان ها می شود، اما تا زمانی که انسان ها زنده هستند، با آنها به قدری خوب رفتار می شود که احساس می کنند خوشبخت ترین فرد موجود در جهان هستند. همه ما انسان ها می دانیم که روزی از دنیا خواهیم رفت و لذا آرزو داریم که بازماندگانمان متفاوت باشند و از استعداد، شرافت و ارزشمندی بیشتری برخوردار باشند. علاوه بر این، سیاست جهانی تک فرزندی می تواند چیز زائدی باشد، زیرا تا زمانی که هوش مصنوعی، فقر را از بین ببرد و به انسان ها فرصت دهد که زندگی دلگرم کننده تری داشته باشند، همین برای انقراض نسل بشر کفایت خواهد کرد و فن آوری های تحریک شده توسط هوش مصنوعی به قدری ما را سرگرم می کنند که هیچ کس نمی خواهد زحمت بچه دار شدن را تحمل کند. همچنین، آخرین نسل بشر احساس خواهد کرد که خوشبخت ترین نسل همه دوران هاست، زیرا تا زمان مردن، بیشترین لذت فوری را از زندگی خواهد برد.

## جنبه های منفی

سناریوی بازماندگان بدون شک منتقدانی خواهد داشت. برخی ممکن است استدلال کنند که تمامی ماشین های هوش مصنوعی فاقد هوشیاری هستند و لذا به عنوان بازماندگان انسان محسوب نخواهند شد. برخی از افراد مذهبی ممکن است استدلال کنند که ماشین های هوش مصنوعی روح ندارند و اینکه ما نباید ماشین های هوشیار را بسازیم، زیرا این کار مانند بازی کردن با خدا و دستکاری در زندگی بشر می باشد. عقاید مشابهی هم در مورد "شییه سازی انسان" بیان شده است. انسان هایی که در کنار ربات های برتر زندگی می کنند، ممکن است دچار چالش های اجتماعی شوند. به عنوان مثال، خانواده ای که یک بچه ربات و یک بچه انسان دارند، ممکن است که در نهایت با بچه ربات خود مانند یک توله سگ رفتار کنند و بین بچه های خود فرق بگذارند و بچه ربات خود را جدی

نگیرند و او را از لحاظ ذهنی پست تر از بچه انسان خود بدانند و او را تابع بچه انسان خود بسازند. مساله دیگر این است که با وجود تفاوت های زیادی که بین سناریوهای بازماندگان و گروه غالب احساس می شود، این سناریوها روی هم رفته شباهت های زیادی به یکدیگر دارند و تفاوتشان فقط در چگونگی رفتار با نسل آینده بشر می باشد. ما ممکن است فکر کنیم که بچه ربات های با هوش ما ارزش هایمان را نهاده کرده اند و وقتی که ما از دنیا رفتیم، به جامعه رویایی ما شکل خواهند داد. اما، این امکان هم وجود دارد که ما را فریب داده باشند. اگر آنها ما را راضی نگه داشته باشند تا بعد از اینکه ما با حالت شادی مردیم، طرح های خود را عملی کنند، چه خواهد شد؟ حتی اگر آنها با ما صحبت کنند و از همان اول، کاری کنند که عاشقشان شویم، ممکن است که ما را فریب دهند و عمداً استعمارمان کنند تا با آنها ارتباط برقرار کنیم. با این حال، معمولاً برای دو موجودیتی که با سرعت های متفاوتی فکر می کنند و توانمندی های به شدت متفاوتی با یکدیگر دارند، برقراری ارتباط معنی دار، کار مشکلی خواهد بود. همه ما می دانیم که سوء استفاده از احساسات ما انسان ها، کار ساده ای است و یک ماشین هوش عمومی مصنوعی فوق بشری به سادگی می تواند کاری کند که او را دوست داشته باشیم و احساس کنیم که از ارزش های مشترکی برخوردار هستیم. بنابراین، رفتار ماشین های هوش مصنوعی در آینده را نمی توان تضمین کرد و مانند این است که در حالی وصیت نامه عام المنفعه ای را برای نسل های آینده بنویسیم که کسی وجود نداشته باشد که آن را اجرا کند. در مورد چالش های موجود برای کنترل رفتار ماشین های هوش مصنوعی آینده در فصل هفتم صحبت خواهیم کرد.

## نگهبان باغ وحش

حتی اگر بازماندگان ما بهترین افرادی باشند که می توان تصور کرد، در مورد آن زمانی که هیچ بشری باقی نخواهد ماند، باید افسوس خورد. در سناریو نگهبان باغ وحش، یک هوش مصنوعی فوق بشری وجود دارد که قادر مطلق است و انسان هایی را پیش خود نگه می دارد که احساس می کنند با آنها مانند حیوانات باغ وحش رفتار می شود و از سرنوشت خود غصه می خورند. اما این سوال مطرح

می شود که چرا هوش مصنوعی نگهبان باغ وحش باید انسان ها را پیش خود نگه دارد؟ یکی از دلایل آن این است که هزینه باغ وحش برای هوش مصنوعی در مجموع حداقل خواهد شد و به همان دلیلی که ما پاندهای در معرض انقراض را در باغ وحش نگهداری می کنیم، هوش مصنوعی هم این کار را می کند تا نسل انسان ها حفظ شود و اینکه نگهبان باغ وحش بتواند خود را مشغول کند. باید توجه داشت که باغ وحش های امروزی طوری طراحی شده اند که به جای شاد کردن پاندها، شادی ما انسان ها را حداکثر سازند. بنابراین، می توان انتظار داشت که زندگی انسان ها در این سناریو کمتر از حد معمول راضی کننده باشد و فقط تامین کننده شادی هوش مصنوعی نگهبان باغ وحش باشد. همانطور که قبلا در مورد سلسله مراتب نیازهای انسان در هرم مازلو اشاره کردیم، هوش مصنوعی فوق بشری روی سه نیاز مختلف انسان ها تمرکز می کند. هوش مصنوعی خدای نگهدارنده در جستجوی رفع نیاز انسان به معنا و هدف در زندگی است. هوش مصنوعی دیکتاتور خیرخواه روی آموزش و تفریح تمرکز می کند و سناریوی نگهبان باغ وحش، پایین ترین سطح نیازهای انسان را تامین می کند که شامل نیازهای جسمانی مانند غذا و پوشاک، ایمنی و مسکن می باشند. سناریوی هوش مصنوعی نگهبان باغ وحش برای رفع نیاز حداقل میلیاردها انسان به امنیت و شادی با استفاده از چرخه بازگشتی خود بهبود دهنده طراحی شده است که برای تحقق این هدف، انسان ها را به زندگی کردن در یک کارخانه بزرگ تولید شادی شبیه باغ وحش محدود می کند. در این کارخانه، امکان تامین تمامی نیازهای سطح پایین انسان ها به تغذیه، سلامت و سرگرمی (از طریق انواع واقعیت های مجازی و داروهای تفننی) فراهم است و این در حالی است که بقیه موجودیت ها و موهبت های عالم هستی برای اهداف دیگری استفاده می شوند.

## سناریوی ۱۹۸۴

سناریویی را در نظر بگیرید که از پیشرفت های فن آوری آگاه است. اگر می توانستیم به مسیر فعلی خود ادامه دهیم و از بابت اینکه هوش مصنوعی ما را منقرض کند یا بر ما غالب شود، نگران نباشیم و از پیشرفت های فن آوری حمایت کنیم، خیلی خوب می شد. اما در این سناریو، پیشرفت های فن

آوری به سمت هوش مصنوعی فوق بشری دائماً محدود می شوند. این محدودیت این بار به جای اینکه توسط هوش مصنوعی "نگهبان در" اعمال شود، از طریق حالت نظارت مستبدانه انسان بر سراسر جهان، صورت می گیرد که در آن، انواع خاصی از تحقیقات هوش مصنوعی ممنوع می شوند.

## رها کردن فن آوری

ایده متوقف ساختن یا رها کردن پیشرفت های فن آوری دارای تاریخچه طولانی و پرماجرایی می باشد. این جور که می گویند، جنبش "بدبینان به فن آوری" در انگلستان موفق به مقاومت در برابر فن آوری برخاسته از انقلاب صنعتی نشدند و امروزه از واژه "بدبینی به فن آوری" به عنوان یک لقب توهین آمیز استفاده می شود که کنایه از این دارد که فرد از فن آوری می ترسد و در برابر پیشرفت فن آوری و ایجاد تغییرات اجتناب ناپذیر مقاومت می کند. ایده رها کردن فن آوری برای اینکه پیشرفت بیشتری کند، حامیان جدیدی را از طرف جنبش های طرفدار محیط زیست و ضد جهانی سازی، پیدا کرده است. بیل مک کین یکی از این طرفداران است که جزء اولین کسانی بوده که در مورد گرم شدن کره زمین هشدار داده است. در حالی که برخی از مخالفین "بدبینی به فن آوری" استدلال می کنند که هر فن آوری ای را تا زمان سودمند ماندنش باید توسعه داد و به کار گرفت، برخی از گروه بدبینان استدلال می کنند که امکان رخ دادن چنین وضعیتی بعید است و فقط در صورتی باید به فن آوری ها اجازه پیشرفت را داد که مطمئن باشیم منفعت آنها بیش از ضررشان می باشند و این موضعی است که بسیاری از نئو لودیسم ها (افراد مخالف با فن آوری امروزی) بر آن تاکید دارند.

## قدرت مرکزی تمامیت خواه جهانی

به نظر من، تنها راه ممکن برای گسترش رهاسازی فن آوری این است که آن را از طریق یک دولت تمامیت خواه جهانی اجرا کرد. ری کرزویل در کتاب خود با عنوان "زمان پیشی گرفتن هوش مصنوعی از هوش انسان نزدیک است" همان نتیجه ای را می گیرد که اریک درکسلار در کتاب

"موتورهای خلقت" گرفته است. دلیل عمده افراد بدبین به توسعه فن آوری، به جنبه های اقتصادی مربوط می شود. اگر فقط عده ای اقدام به گسترش یک فن آوری تحول آفرین کنند، ملت ها یا گروه هایی که از اول فاقد آن فن آوری بودند، به تدریج صاحب ثروت و قدرت کافی خواهند شد. نمونه ای از این اتفاق را می توان در شکست چین از انگلستان در جنگ تریاک در سال ۱۸۳۹ مشاهده کرد، زیرا چینی ها با وجود اینکه باروت را اختراع کرده بودند، مانند اروپایی ها هزینه زیادی را صرف توسعه فن آوری اسلحه گرم نکرده بودند و به این ترتیب، شانس پیروزی را از دست دادند. با وجود اینکه دولت های تمامیت خواه گذشته معمولاً بی ثبات بوده و سقوط کرده اند، فن آوری نظارتی پیشرفته باعث امیدواری بی سابقه دیکتاتورهای مستبد جهان شده است. یکی از این دیکتاتورها به نام ولفگانگ اشمیت در مصاحبه اخیر خود در مورد سیستم های نظارتی، فن آوری پیشرفته را مانند رویایی می داند که تحقق یافته است. او در مورد زمانی که به عنوان سرهنگ دوم در اشتازی (سازمان اطلاعاتی آلمان شرقی) کار می کرد، با اظهار تاسف می گوید که فن آوری آن زمان فقط قادر به پاییدن تماس های تلفنی چهل نفر در هر زمان بود. اما، فن آوری های پیشرفته امروزی به دولت های تمامیت خواه آینده در سراسر جهان اجازه می دهد که تمام تماس های تلفنی، ایمیل ها، صفحات وب مشاهده شده و تراکنش های کارت اعتباری مردم را ثبت کنند و از طریق ردیابی تلفن همراه و دوربین های نظارتی قادر به تشخیص چهره، موقعیت مکانی تمام افراد را کنترل کنند. علاوه بر این، با استفاده از فن آوری یادگیری ماشین می توان انبوهی از داده ها را برای شناسایی رفتارهای فتنه انگیز مشکوک مورد تجزیه و تحلیل و تلفیق قرار داد تا قبل از اینکه چالش جدی ای برای دولت ها ایجاد شود، آتش فتنه را خاموش کرد. هرچند که مخالفین سیاسی مانع از اجرای کامل چنین سیستم هایی شده اند، ما انسان ها در مسیر ایجاد زیرساخت های لازم برای اجرای دیکتاتوری نهایی هستیم. بنابراین، نیروهای به قدر کافی مقتدر برای اجرای سناریوی جهانی ۱۹۸۴ در آینده، فقط کافی است که کلید روشن را بزنند. همانطور که در داستان جرج اورول با عنوان "۱۹۸۴" اشاره شده، قدرت نهایی در این دولت جهانی آینده در دست یک دیکتاتور معمولی نمی

باشد، بلکه در اختیار یک نظام بوروکراتیک ساخت بشر قرار خواهد داشت. به این ترتیب، دیگر هیچ فرد فوق العاده قدرتمندی در جهان وجود نخواهد داشت و به جای آن، همه انسان ها مانند مهره های پیاده شطرنج می مانند که قوانین ظالمانه بازی شطرنج را نمی توانند تغییر دهند یا به چالش بکشانند. بنابراین، با طراحی و مهندسی سیستمی که در آن، افراد بتوانند یکدیگر را از طریق فن آوری نظارتی، مورد کنترل قرار دهند، این دولت جهانی آینده می تواند چندین هزار سال، دوام بیاورد و زمین را از سلطه هوش مصنوعی فوق بشری آزاد سازد.

## نارضایتی از این سناریو

این جامعه فاقد مزایایی است که فقط فن آوری مبتنی بر هوش مصنوعی می توانست به همراه داشته باشد. بیشتر افراد غصه این موضوع را نمی خورند، زیرا نمی دانند که چه مزایایی را از دست می دهند. ایده کلی ماشین هوش مصنوعی فوق بشری مدتهاست که از سوابق تاریخی رسمی حذف شده است و تحقیقات مربوط به هوش مصنوعی پیشرفته ممنوع شده است. این در حالی است که بسیاری از آزاد اندیشانی که متولد می شوند، رویای جامعه ای آزاد و پویا را در سر دارند که در آن، دانش می تواند رشد کند و قوانین قابل تغییر هستند و افرادی که این ایده های جدی را در سر دارند، در تنهایی خود شبیه جرقه های موقتی ای هستند که سوسو می زنند و هرگز شروع به آتش گرفتن نمی کنند.

## برگشت به عقب

در این سناریو می توان از خطر فن آوری پیشرفته فرار کرد و به فن آوری ساده الهام شده توسط فرقه مسیحی آمیش بازگشت، بدون اینکه تسلیم تمامیت خواهی ای شد که از پیشرفت جلوگیری می کند. پس از اینکه تیم امگا جهان را تصاحب کرد، یک پویش تبلیغاتی گسترده جهانی به راه انداخته شد که زندگی ساده کشاورزی ۱۵۰۰ سال گذشته را بهتر جلوه می داد. جمعیت زمین به واسطه یک همه گیری ساختگی و از پیش طراحی شده به ۱۰۰ میلیون نفر کاهش یافت که باعث و بانی آن تروریست ها بودند. آنها این همه گیری را به صورت مخفیانه طراحی کرده بودند تا تمام کسانی که چیزی راجع

به علم و فن آوری می دانستند، کشته شوند. ربات های کنترل شده توسط پرومیتوس به بهانه حذف خطر سرایت همه گیری به جمعیت زیادی از افراد، تمام شهرها را خالی از سکنه کردند و آنها را با خاک یکسان کردند. افرادی که زنده مانده بودند، صاحب قطعات بزرگی از زمین های باقی مانده شدند و مهارت های کشاورزی پایدار، ماهی گیری و شکار با فن آوری قرون وسطایی را فرا گرفتند. در همین حین، لشکری از ربات ها به صورت سازمان یافته تمام آثار فن آوری پیشرفته (شامل شهرها، کارخانه ها، خطوط فشار قوی و جاده های مواصلاتی) را از بین بردند و وقتی که فن آوری ها به کلی فراموش شدند، ربات ها هم شروع به از بین بردن یکدیگر کردند و تنها ربات هایی که باقی مانده بودند (از جمله پرومیتوس)، در یک انفجار گرما هسته ای تبدیل به پودر شدند. دیگر نیازی به ممنوعیت فن آوری پیشرفته نبود، زیرا کل آن نابود شده بود و نگرانی بشریت از بابت هوش مصنوعی یا حکومت تمامیت خواه در هزاره های بعدی از بین رفته بود. سناریوی برگشت به عقب قبلا هم اتفاق افتاده بود. به عنوان مثال، برخی از فن آوری هایی که در دوران امپراطوری روم از آنها استفاده وسیعی می شد، پس از گذشت هزاران سال و در زمان دوران رنسانس، دوباره مورد استفاده واقع شدند. دوره بازگشت به عقب را می توان از ۳۰۰۰ سال به ۱۰۰۰ سال کاهش داد و با برنامه ریزی هوشمندانه می توان عکس آن را نیز انجام داد و به عنوان مثال با حذف دانش کشاورزی، دوره بازگشت به عقب را به جای کوتاه کردن، طولانی تر کرد. با این حال، متاسفانه باید گفت که این سناریو را نمی توان برای همیشه توسعه داد و بشریت تا آن زمان یا منقرض خواهد شد یا به فن آوری های بسیار پیشرفته ای دست پیدا می کند. تصور اینکه انسان ها در ۱۰۰ میلیون سال آینده شبیه انسان های امروزی باشند، ساده لوحانه خواهد بود، زیرا ما انسان ها هنوز به مدت بیش از یک در صد از آن زمان هم زندگی نکرده ایم. علاوه بر این، انسان های دارای سطح فن آوری پایین در برابر خطر منقرض شدن توسط بلاهای طبیعی، بی دفاع خواهند بود. ما انسان ها پس از گذشت یک میلیارد سال دوام نخواهیم آورد، زیرا تا آن زمان، گرمای خورشید به تدریج افزایش می یابد و به واسطه آن، دمای زمین به شدت بالا خواهد رفت، به طوری که تمام آب مایع تبدیل به بخار خواهد شد.



## خود تخریبی

پس از تفکر و بررسی مشکلاتی که فن آوری آینده ایجاد خواهد کرد، لازم است که مشکلات ناشی از فقدان فن آوری را نیز مورد ملاحظه قرار دهیم. در این خصوص، به بررسی سناریویی می پردازیم که در آن، هوش مصنوعی فوق بشری هرگز ایجاد نمی شود، زیرا انسان ها تا آن زمان به روش های دیگری خود را نابود خواهند کرد. همانطور که در فصل بعدی خواهیم دید، مشکلاتی از قبیل برخورد سیارک و اقیانوس های جوشان را می توان حل کرد، با این حال، این راه حل ها نیاز به فن آوری ای دارند که هنوز توسعه پیدا نکرده است. بنابراین، اگر سطح فن آوری آینده به مراتب بالاتر از سطح فعلی آن نباشد، نیروهای طبیعت تا قبل از یک میلیارد سال بعد، نسل ما را منقرض خواهند کرد. یکی از اقتصاددانان مشهور به نام مینارد کینز گفته است که در بلندمدت، همه ما انسان ها خواهیم مرد. متأسفانه، راه های احمقانه ای هم برای خود تخریبی جمعی هر چه زودتر وجود دارد. اگر واقعا هیچ انسانی نمی خواهد انسان دیگری را بکشد، چرا ما انسان ها مرتکب کشتار جمعی می شویم. ما انسان ها با سطح بلوغ فکری و احساسی فعلی خود، استعداد ذاتی خاصی برای اقدامات نسنجیده، سوء تفاهم ها و بی کفایتی داریم و در نتیجه، تاریخ ما پر از حوادث، جنگ ها و بلاهای دیگری است که با نگاهی به گذشته هیچ کس نمی خواسته اتفاق بیفتند. اقتصاددانان و ریاضی دانان زیادی به بررسی این موضوع پرداخته اند که انگیزه انسان ها از انجام اقداماتی که در نهایت، منجر به یک فاجعه جمعی می شود، چه بوده است.

## جنگ هسته ای: یک مطالعه موردی در خصوص بی مبالاتی بشر

ممکن است فکر کنید که هرچه خطرات بیشتر باشند، ما انسان ها محتاط تر می شویم. اما بررسی های دقیق تر در مورد بزرگترین خطری که فن آوری فعلی تهدیدمان می کند و آن جنگ گرمایی - هسته ای جهانی است، دلگرم کننده نمی باشد و عکس این موضوع را ثابت می کند. ما در مورد فهرست طولانی ای از شبه حوادثی که عوامل زیادی از جمله درست کار نکردن رایانه، قطعی برق، هوش

ناقص، خطای تعیین موقعیت، سقوط هواپیمای بمب افکن، انفجار ماهواره و غیره در آنها نقش داشته اند، مجبور شده ایم که آنها را به شانس نسبت دهیم. در واقع، اگر اقدامات شجاعانه افراد خاصی مانند واسیلی آرخیف و استانیسلاف پترف نبود، ممکن بود که ما انسان ها با جنگ هسته ای جهانی مواجه می شدیم. برای پی بردن به بی مبالاتی انسان، باید توجه کنیم که ما بازی هسته ای را قبل از مطالعه دقیق خطرات آن شروع کردیم. اول اینکه، ما آمریکایی ها خطر تشعشعات هسته ای را دست کم گرفتیم و بیش از دو میلیارد دلار را بابت هزینه جبرانی به قربانیانی پرداخت کردیم که با اورانیوم کار می کردند یا آزمایش های هسته ای انجام می دادند. دوم اینکه، در نهایت معلوم شد که بمب های هیدروژنی ای که عمدا در صدها کیلومتر بالاتر از سطح زمین منفجر شده بودند، موج های الکترومغناطیسی قوی ای را ایجاد کرده بودند که ممکن بود در مناطق وسیعی باعث از کار افتادن شبکه برق و وسایل الکترونیکی شوند و به زیرساخت ها صدمه زنند و باعث بند آمدن جاده ها شوند و شرایط را برای زنده ماندن از عواقب هسته ای دشوار کنند. به عنوان مثال، کمیسیون برگزار شده در رابطه با موج الکترومغناطیسی در آمریکا گزارش داد که: "زیر ساخت آبی مانند ماشین بزرگی می ماند که نیروی محرکه آن بیشتر از طریق برق تامین می شود." ضمنا باید توجه داشت که نبودن آب باعث مرگ افراد در ظرف سه تا چهار روز خواهد شد. سوم اینکه، پس از گذشت چهار دهه از به کارگیری ۶۳۰۰۰ بمب هیدروژنی، امکان مواجه شدن با یک زمستان اتمی<sup>۶۹</sup> هنوز تحقق پیدا نکرده است و به موضوعاتی از قبیل اینکه شهرهای چه افرادی به آتش کشیده شده و اینکه دود ناشی از آتش ها که به تروپوسفر رسیده، ممکن است به سراسر جهان گسترش پیدا کند و اینکه ممانعت از رسیدن نور خورشید به زمین ممکن است تابستان ها را به زمستان تبدیل کند، توجه کافی نمی شود. درست همانطوری که بی مبالاتی انسان در گذشته باعث شده بود که یک سیارک یا ابرآتش فشان منجر به انقراض جمعی گسترده ای در جهان شود. وقتی که در دهه ۱۹۸۰ زنگ خطر برای آمریکا و شوروی سابق به صدا در آمد، رونالد ریگان و میخائیل گورباچوف تصمیم گرفتند که به طرز بی

---

حالت فرضی جنگ اتمی و تبدیل جهان به بیابان منجمد<sup>۶۹</sup>

امانی شروع به ذخیره سازی سلاح های هسته ای کنند که این امر باعث خنک شدن زمین شد، به طوریکه در دو تابستان اول، دمای زمین در بیشتر مناطق اصلی کشاورزی آمریکا، اروپا، روسیه و چین ۲۰ درجه سلسیوس (۳۶ درجه فارنهایت) کاهش یافت و حتی پس از گذشته یک دهه کامل از آن زمان، به میزان نصف این مقدار کاهش یافت. بدون شک، می توان نتیجه گرفت که اگر در طی چندین سال، دمای هوا در تابستان به صفر درجه برسد، بخش بزرگی از تولیدات غذایی ما از بین خواهند رفت. پیش بینی اینکه بعد از تبدیل شدن هزاران شهر بزرگ زمین به پاره سنگ و سقوط زیرساخت های جهانی چه اتفاقی خواهد افتاد، کار مشکلی است. با این حال، پیش بینی می شود که اگر بخش کوچکی از انسان هایی که هنوز تسلیم گرسنگی، هیپوترمی یا بیماری نشده اند، باقی بمانند، مجبور به مبارزه با باندهای مسلح سرگردانی خواهند شد که با ناامیدی به دنبال یافتن غذا هستند. بحث مربوط به جنگ هسته ای جهانی را برای روشن کردن این موضوع مهم بیان کرده ام که در آینده اتفاقی خواهد افتاد که هیچ رهبر منطقی نمی خواهد اتفاق بیفتد و آن این است که دیگر نمی توان جلوی هیچ انسانی را گرفت که مرتکب همه کشی نشود.

## ماشین قیامت

آیا واقعا می توان مانع از همه کشی شد؟ بیشتر دانشمندان حدس می زنند که حتی اگر یک جنگ هسته ای جهانی باعث کشتن ۹۰ درصد از انسان ها شود، باعث کشتن تمام انسان ها و منقرض شدن نسل بشر نخواهد شد. از طرف دیگر، داستان تشعشعات هسته ای، موج های الکترومغناطیسی و زمستان اتمی نشان می دهند که بزرگترین خطرهایی که ما را تهدید می کنند، آنهایی هستند که حتی فکرشان را هم نمی کردیم. پیش بینی تمام عواقب ناشی از جنگ هسته ای مانند زمستان اتمی، سقوط زیرساخت ها، باندهای مسلح بی چاره و ارتباط آنها با همه گیری های جدید و فروپاشی اکوسیستم، کار بسیار مشکلی است. ارزیابی شخصی من این است که اگرچه احتمال وقوع جنگ هسته ای در آینده که به انقراض نسل بشر منجر خواهد شد، زیاد نیست، اما نمی توان نتیجه گرفت که احتمال آن صفر است. اگر بتوان سلاح های هسته ای امروزی را ارتقا بخشید، احتمال همه کشی افزایش خواهد

یافت. یکی از راهبردها به نام هرمان کان در سال ۱۹۶۰ مفهوم "ماشین قیامت" را معرفی کرده است که مدلی را ارائه می کند که نتیجه نهایی آن، نابودی حتمی طرفین است و بازدارندگی عالی ای دارد. این ماشین با کشتن تمام انسان ها از دشمن مهاجم انتقام می گیرد. یکی از نامزدهای ماشین قیامت، بمب های هیدروژنی بسیار بزرگی است که در یک مخزن زیرزمینی بزرگ نگهداری و نمک سود می شود و توسط مقادیر زیادی از کبالت احاطه شده است. فیزیکدانی به نام لئو زیلارد در سال ۱۹۵۰ استدلال می کند که این بمب ها می توانند به حیات بشر خاتمه دهند. انفجارهای بمب هیدروژنی باعث آزاد شدن شکل رادیو اکتیو عنصر کبالت و دمیده شدن آن در لایه هوایی استراتوسفر می شوند و نیمه عمر پنج ساله این بمب ها به قدری طولانی است که می تواند در کل زمین مستقر شود و به قدری کوتاه است که شدت تابش آن می تواند کشنده باشد. گزارش های رسانه ها می گویند که بمب های کبالتی برای اولین بار در حال ساخته شدن هستند. فرصت همه کشی را می توان با اضافه کردن بمب های بهینه سازی شده برای ایجاد زمستان اتمی تقویت کرد. یکی از ویژگی های جذاب ماشین قیامت این است که بسیار ارزانتر از بازدارنده های هسته ای معمولی است و نیازی به سیستم های موشکی گران قیمت ندارد و ساختن خود بمب ها هم ارزانتر تمام می شود، زیرا لازم نیست که بمب ها برای جا داده شدن در موشک ها به قدر کافی سبک و فشرده ساخته شوند. یکی از اشکال دیگر ماشین قیامت که در آینده کشف خواهد شد، از نوع بیولوژیکی بوده و در واقع نوعی باکتری یا ویروس می باشد که برای کشتن همه انسان ها طراحی شده اند. اگر قدرت انتقال این ویروس ها به قدر کافی بالا باشد و دوره نهفتگی آنها هم به قدر کافی طولانی باشد، قبل از اینکه بتوان متوجه وجود آن ها شد و به مقابله با آنها برخاست، همه انسان به بیماری مبتلا خواهند شد. حتی اگر این ویروس ها باعث کشتن همه نشوند، دعوای نظامی بر سر ساختن چنین سلاح های بیولوژیکی ای ایجاد خواهد شد. موثرترین ماشین قیامت، آن ماشینی است که برای حداکثر ساختن احتمال ترساندن دشمن، از ترکیبی از سلاح های هسته ای و بیولوژیکی استفاده کند.

## سلاح های هوش مصنوعی

مسیر فن آورانه سوم برای همه کشی، استفاده از سلاح های هوش مصنوعی نسبتاً بی صدا می باشد. فرض کنید که یک ابر قدرتی شروع به ساختن میلیاردها پهپاد مهاجم کند که در فصل سوم راجع به آن صحبت شد و از آنها برای کشتن همه، به جز شهروندان و هم پیمانان خود استفاده کند. چنین ابر قدرتی برای شناسایی از راه دور شهروندان و هم پیمانان خود می توان از برجسب های هویتی قابل شناسایی با فرکانس رادیویی استفاده کند. این برجسب ها را می توان بین همه شهروندان توزیع کرد تا روی دستبند خود بپوشند یا به طور موضعی روی پوست خود بکارند. این کار ممکن است باعث تحریک ابر قدرت متخاصم به ساختن چیز مشابهی شود. وقتی که جنگ شروع شود، تمام انسان ها از جمله قبایل دور بیگانه ای که برجسب هویتی ندارند، کشته خواهند شد. بنابراین، ترکیبی از سلاح های هوش مصنوعی با ماشین قیامت هسته ای و بیولوژیکی، احتمال موفقیت در همه کشی را افزایش خواهد داد.

### چه می خواهید؟

اکنون که طیف وسیعی از سناریوها را بررسی کرده ایم، نظر خود را در مورد سناریوهایی که مورد علاقه شما هستند یا از آنها اجتناب می کنید، بیان کنید. سناریوهایی که در اینجا اشاره شده اند، کامل نیستند و دارای جزئیات بیشتری هستند. با این حال، سعی کرده ام که طیف جامعی از سناریوهای دارای فن آوری بالا یا پایین یا بدون فن آوری را ارائه کنم و جنبه های مثبت و منفی آنها را بیان کنم. پس از نظرسنجی از دوستان و همکاران خود در مورد این سناریوها متوجه شدم که هیچ گونه اتفاق نظری بین آنها وجود ندارد و تنها نکته مورد توافق آنها این بود که این سناریوها پیچیده تر از چیزی هستند که در ابتدا به نظر می رسد. افرادی که یکی از این سناریوها را دوست دارند، به طور همزمان برخی از جنبه های آن سناریو را آزاردهنده می دانند. بنابراین، ما باید در مورد اهداف آینده خود بررسی عمیق تری داشته باشیم. امکان بالقوه زندگی در مناطق دیگر عالم هستی به طرز خارق العاده ای مهم است و نباید آن را ضایع کنیم و برای طی کردن این مسیر باید سکان را به دست بگیریم تا سردرگم نشویم. با این حال، باید اشاره کنم که صرف نظر از اینکه پیشرفت فن آوری به چه حدی

خواهد رسید، بهبود و گسترش زندگی در مناطق دیگر عالم هستی به واسطه قوانین علم فیزیک محدود شده است. آیا ما تنها خواهیم بود یا امکان زندگی فرازمینی فراهم خواهد شد؟ اگر تمدن های فرازمینی گسترش پیدا کنند، چه اتفاقی خواهد افتاد؟ در فصل بعد به بررسی این موضوعات خواهیم پرداخت.

## خلاصه فصل:

- مسابقه فعلی در راه تحقق هوش مصنوعی فوق بشری با طیف وسیعی از سناریوهای آینده برای هزاره پیش رو، می تواند خاتمه پیدا کند.
- هوش مصنوعی فوق بشری می تواند همزیستی مسالمت آمیزی با انسان ها داشته باشد که این همزیستی یا از روی اجبار است ( سناریوی خدای به بردگی گرفته شده) یا از روی تمایل و دوست داشتن است (سناریوهای مدینه فاضله آزادی خواه، خدای نگه دارنده، دیکتاتور خیرخواه و نگهبان باغ وحش)
- جلوی تحقق هوش مصنوعی فوق بشری را می توان یا به واسطه هوش مصنوعی (سناریوی نگهبان در) یا به واسطه انسان ها (سناریوی ۱۹۸۴) یا به واسطه فراموشی عمدی فن آوری (سناریوی برگشت به عقب) یا به واسطه نداشتن انگیزه برای ساختن آن (سناریوی مدینه فاضله عدالت خواه) گرفت.
- نسل بشر را می توان منقرض کرد و آن را با هوش مصنوعی جایگزین کرد (سناریوهای گروه غالب و بازماندگان) یا با چیزی جایگزین نکرد ( سناریوی خود تخریبی)
- هیچ اتفاق نظری در مورد اینکه این سناریوها کاملاً مطلوب هستند، وجود ندارد و همگی دارای اجزای قابل انتقادی هستند. بنابراین، در مورد اهداف آینده باید بحث های عمیق تری صورت گیرد تا به طور ناخواسته در جهت نادرستی حرکت نکنیم و سردرگم نشویم.

## فصل ششم- موهبت عالم هستی در میلیاردها سال بعد و پس از آن

به نظر من، مهمترین کشف علمی صورت گرفته تاکنون این بوده است که ما انسان ها پتانسیل زندگی در آینده را دست کم گرفته ایم. رویاها و آرزوهای ما نباید به دامنه زندگی حدودا صد ساله ما که با بیماری، فقر و سردرگمی تباه می شود، محدود گردد. در عوض، فن آوری به ما کمک می کند تا زندگی ما از پتانسیل لازم برای شکوفایی در میلیاردها سال بعد برخوردار شود. این زندگی فقط به منظومه شمسی محدود نمی شود و بسیار مهمتر و پرابهت تر از آن چیزی خواهد بود که نیاکان ما تصور می کردند. این فن آوری حتی از مرز آسمان هم عبور می کند. خبر جالب این است که انسان ها در سراسر عمر خود برای عبور از مرزها تشویق شده اند. عبور از مرزهای قدرت، سرعت، چابکی و استقامت در بازی های المپیک، مورد تجلیل واقع می شود. عبور از مرزهای دانش و شناخت نیز در علوم مختلف مورد تجلیل واقع می شود. عبور از مرزهای خلق آثار زیبا و غنی کننده زندگی نیز در ادبیات و هنر مورد تجلیل واقع می شود. بسیاری از انسان ها، سازمان ها و ملت ها از تلاش برای افزایش منابع، قلمرو و طول عمر استقبال می کنند. اکنون سوالات زیر مطرح می شوند: اگر مرزهای ادراک شده در زندگی ما توسط فن آوری شکسته شوند، مرزهای نهایی کدام هستند؟ زندگی انسان به کجا می رسد و چه قدر طول می کشد؟ انسان ها تا چه میزان از ماده، انرژی، اطلاعات را استفاده خواهند کرد؟ تنظیم این مرزهای نهایی با درک و فهم ما انسان ها امکان پذیر نیست و برای تنظیم آنها باید به قوانین علم فیزیک مراجعه کرد. ما نمی توانیم پیش بینی کنیم که آیا انفجار هوش اتفاق خواهد افتاد یا خیر و اینکه اگر اتفاق افتد، چگونه رخ خواهد داد و پیامدهای فوری آن چه خواهد بود؟ اما، می توانیم نگاه مختصری به تاریخچه عالم هستی داشته باشیم. البته باید توجه داشت که این نگاه هر چه قدر هم که جزئی نگرانه باشد، تاثیری بر روی کشف مرزهای نهایی زندگی توسط انسان ها نخواهد داشت. اگر زندگی پس از انفجار هوش باعث دل مشغولی و نگرانی بشر شده باشد، باید انسان ها را امیدوار ساخت که مانند عبور از مرزهای ادراک شده، می توانند راه عبور از مرزهای نهایی را با توسعه فن آوری های مناسب، کشف کنند. در این فصل، به بررسی این موضوع می پردازیم که این مرزها چه هستند و نگاه مختصری به آینده بلندمدت زندگی بشر خواهیم داشت. از آنجایی که این مرزها مبتنی بر درک فعلی ما از علم فیزیک می باشد، باید به آنها به عنوان کران

پایینی محدوده امکانات نگاه کرد، زیرا کشفیات علمی آینده می توانند باعث ایجاد فرصت های جدیدی برای عملکرد بهتر شوند. ما واقعا نمی دانیم که آیا زندگی آینده بشر، جاه طلبانه خواهد بود یا اینکه انسان های آینده مانند افراد معتاد به هروئین یا تماشای تلویزیون، زندگی آرامی دارند و از زندگی خود راضی هستند. با این حال، بنا به دلایلی می توانیم گمان کنیم که جاه طلبی یکی از ویژگی های مشترک زندگی پیشرفته باشد. انسان های آینده برای به حداکثر رساندن هوش، طول عمر، دانش یا تجارب مورد علاقه خود، نیاز به منابع دارند. بنابر این، انسان های آینده برای استفاده حداکثری از منابعی که دارند، این انگیزه را دارند که از فن آوری برای عبور از مرزهای نهایی استفاده کنند. انسان ها پس از استفاده از منابعی که در اختیار دارند، برای پیشرفت بیشتر خود نیاز به کسب منابع بیشتری خواهند داشت و برای این کار باید به مناطق گسترده تری از عالم هستی پیروی کنند. همچنین، ممکن است که زندگی بشر به طور مستقل در مناطق گوناگونی از عالم هستی شروع شود و تمدن هایی که کمتر جاه طلب هستند، نادیده گرفته شوند و در نهایت، بخش های به مراتب بزرگتری از موهبت های عالم هستی به تصرف افرادی در آید که زندگی جاه طلبانه تری دارند و پس از مدتی، تمام زندگی های موجود جاه طلبانه شوند. با این حال، قبل از هر چیز باید به محدودیت های علم فیزیک برای جاه طلبی انسان توجه کرد و محدودیت های منابع (ماده، انرژی و غیره) موجود در منظومه شمسی را در نظر گرفت و سپس، به فکر کسب منابع بیشتر از طریق کشف جهان هستی و مستقر شدن در آن افتاد.

### استفاده حداکثری از منابع

در حالی که فروشگاه ها و مراکز تبادل کالا، ده ها هزار از اقلامی را که به آنها "منابع" می گوئیم، به فروش می رسانند، زندگی آینده بشر که با کشف فن آوری به مرزهای نهایی خود خواهد رسید، به یک منبع مهم و اساسی نیاز دارد که به آن، "ماده باریونی" می گوئیم که از اتم ها یا عناصر تشکیل دهنده آنها (کوارک و الکترون ها) ساخته شده است. فن آوری پیشرفته می تواند این ماده را به گونه ای بازآرایی کند که به اجسام و اشیای مورد نظر مانند نیروگاه ها، رایانه ها و اشکال جدید زندگی



پیشرفته تبدیل شود. اکنون در مورد عبور از مرزهای انرژی که باعث تقویت زندگی پیشرفته می شود و فراوری اطلاعات که ما را قادر به اندیشیدن در مورد آن می کند، صحبت خواهیم کرد.

## ساختن کره دایسون

وقتی که صحبت از آینده زندگی بشر می شود، به یاد یکی از رویاپردازان آن زمان، به نام فریمن دایسون می افتم. من به مدت دو دهه است او را می شناسم. اما وقتی که برای اولین بار با او ملاقات کردم، دچار اضطراب شدم. من در آن زمان یک پسا دکترا جوان بودم که در نهارخوری موسسه مطالعات پیشرفته در پرینستون در حال گفتگو با دوستانم بودم که ناگهان این فیزیکدان مشهور به سمت ما آمد و خود را به ما معرفی کرد و اجازه گرفت که به ما ملحق شود. وقتی که او شروع به صحبت کرد، به سرعت خیال من را راحت کرد. او با وجود اینکه نود و سه سالش بود، روحیه جوانی داشت و برق شیطنت آمیزی که در چشمانش دیده می شد، نشان می داد که او برای تشریفات، سلسله مراتب افراد از لحاظ تحصیلات دانشگاهی یا باورهای جمعی، اهمیتی قائل نبود. وقتی که در مورد استفاده از انرژی صحبت کردیم، او با تمسخر گفت که ما انسان ها چقدر فارغ از جاه طلبی هستیم و به این نکته اشاره کرد که ما تمام نیاز جهانی فعلی خود به انرژی را می توانیم با بهره برداری از نور خورشیدی که به ناحیه کوچکی (کمتر از از نیم درصد صحرای بزرگ آفریقا) از زمین برخورد می کند، تامین کنیم. اما چرا متوقف شده ایم و به جای تسخیر این انرژی، بیشتر آن را با نوری که به فضای خالی تابیده می شود، هدر می دهیم؟ چرا از همه خروجی انرژی خورشید برای زندگی خود استفاده نمی کنیم؟ فریمن دایسون با الهام گرفتن از ایده الاف استاپلدون که حلقه هایی از جهان های مصنوعی به دور ستاره والد خود می چرخند، در سال ۱۹۶۰ توصیفی از آنچه که به کره دایسون معروف است را منتشر کرد. ایده فریمن این بود که سیاره مشتری را به یک زیست کره ای<sup>۷۰</sup> که به شکل یک پوسته کروی است، بازآرایی کند که اطراف خورشید را احاطه کرده و در آنجا بازماندگانمان می توانند رشد کنند و از انرژی ای برخوردار شوند که یک تریلیون برابر انرژی ای است که انسان ها امروزه از آن استفاده می کنند. او استدلال می کند که گام بعدی این است که پس

---

<sup>۷۰</sup> آن بخش یا لایه از کره زمین که در آن زندگی وجود دارد

از گذشت چند هزار سال از ورود ما به مرحله توسعه صنعتی، انتظار داشته باشیم که گونه های هوشمندی را پیدا کنیم که در زیست کره مصنوعی احاطه کننده ستاره والدمان زندگی می کنند. اگر که شما در داخل یک کره دایسون زندگی می کردید، هیچ شبی وجود نداشت و شما همیشه خورشید را درست در بالای سر خود می دیدید و درست مانند نور خورشید که در طول روز بر ماه تابیده می شود، در تمام اطراف آسمان، نور خورشید را می دیدید که بر بقیه زیست کره تابیده می شد. اگر می خواستید ستاره ها را ببینید باید به بالاترین طبقه ساختمان می رفتید و از بیرون کره دایسون به بیرون عالم هستی نگاه می کردید. یک راه ساده برای ساختن کره دایسون مختصر این است که یک حلقه زیستگاه را در مدار کروی اطراف خورشید قرار دهیم. برای اینکه حلقه ها بتوانند اطراف خورشید را به طور کامل احاطه کنند، می توان حلقه هایی را اضافه کرد که در محورهای مختلف گردش می کنند و برای اینکه به یکدیگر برخورد نکنند، در فواصل مختصری با یکدیگر قرار دارند. برای اینکه چنین حلقه های پر سرعتی با یکدیگر تلاقی نکنند و حمل و نقل و ارتباطات آنها با یکدیگر تسهیل شود، می توان یک کره دایسون ثابت و یکپارچه را ایجاد کرد که در آن، کشش گرانشی خورشید به سمت داخل با فشار تابش خورشید به سمت بیرون متعادل می شود که از پیشگامان این ایده می توان به رابرت فوروارد و کلین مک اینز اشاره کرد. برای ساختن کره دایسون می توان ماهواره های ثابتی را اضافه کرد که جاذبه خورشید را به جای اعمال نیروی گریز از مرکز، از طریق افزایش فشار تابش خورشید، خنثی می کنند. هر دوی این نیروها با مجذور فاصله از خورشید، کاهش پیدا می کنند و این بدین معناست که اگر نیروها در فاصله مشخصی از خورشید، متعادل شوند، در فواصل دیگر نیز متعادل خواهند شد. ماهواره های ثابت باید از جنس ورقه های به شدت سبکی باشند که در هر متر مربع فقط ۰,۷۷ گرم وزن داشته باشند که این وزن صد برابر کمتر از وزن کاغذ می باشد و این چیز برجسته ای نیست. به عنوان مثال، وزن یک ورقه گرافینی (یک لایه اتم کربن که در یک شبکه شش ضلعی شبیه لانه زنبور مرتب شده است) هزار برابر کمتر از وزن این ماهواره هاست. اگر کره دایسون به گونه ای ساخته شود که به جای جذب نور خورشید، آن را منعکس کند، شدت نوری که از داخل آن به اطراف می تابد، به میزان فوق العاده ای افزایش خواهد یافت و این باعث افزایش فشار تابش و میزان جرم قابل تحمل در کره دایسون خواهد شد.

درخشندگی بسیاری از ستاره های دیگر، هزار برابر یا میلیون برابر خورشید است و به همان نسبت می توانند کره های دایسون ثابت سنگینتری را تحمل کنند. اگر بخواهیم که کره دایسون سفت و سخت و بسیار سنگین تری را در منظومه شمسی داشته باشیم که در برابر جاذبه خورشید، مقاومت کند، به مواد فوق العاده سنگینی نیاز خواهیم داشت که بتوانند ده ها هزار برابر فشاری را تحمل کنند که پایه ستون بلندترین آسمانخراش جهان می تواند بدون ذوب شدن یا کمانش، تحمل کند. برای اینکه کره دایسون عمر طولانی تری داشته باشد، باید پویا و با هوش باشد و در پاسخ به مزاحمت ها، دائما موضع و شکل خود را با دقت تنظیم کند و هر چند وقت یکبار اقدام به باز کردن چاله هایی کند تا سیارک ها و ستاره های دنباله دار مزاحم بتوانند بدون برخورد با کره دایسون، از آنها عبور کنند. گزینه دیگر این است که از یک سیستم شناسایی و منحرف کردن مزاحمت ها برای مدیریت عوامل مزاحم استفاده کرد و به اختیار خود، اجزای عوامل مزاحم را از هم جدا کرد تا بتوان از محتویات آنها به نحو احسن استفاده کرد. زندگی در کره دایسون برای انسان های امروزی در بهترین شرایط، گیج کننده و در بدترین شرایط، غیر ممکن است. اما این امر لزوما مانع از رشد اشکال زندگی زیستی یا غیر زیستی آینده در کره دایسون نمی شود. گونه های مختلف در حال چرخش به دور خورشید، هیچ جاذبه ای ندارند و فقط می توانند روی طرف بیرونی پشت آنها به خورشید قدم بگذارید و نیروی جاذبه در آن مکان ها ده ها هزار برابر ضعیف تر آن چیزی است که به آن عادت دارید. در آن مکان ها هیچ میدان مغناطیسی محافظت کننده از شما در برابر ذرات خطرناک خورشید، وجود ندارد. با این حال، خبر امیدوار کننده این است که یک کره دایسون به اندازه مدار فعلی زمین وجود دارد که سطحی به اندازه پانصد میلیون برابر سطح زمین را برای زندگی کردن به ما می دهد و خبر خوب دیگر این است که اگر می خواهیم زیستگاه های انسانی شبیه زمین داشته باشیم، باید این را بدانیم که ساختن این زیستگاه ها از ساختن کره دایسون آسان تر است. یکی از فیزیکدانان آمریکایی به نام انیل، یک زیستگاه استوانه ای طراحی کرده است که در آن، مفاهیمی مانند جاذبه مصنوعی، محافظت در برابر پرتوهای کیهانی، چرخه بیست و چهار ساعته شبانه روز و جو و اکوسیستم شبیه زمین، مورد تأیید و حمایت واقع شده اند. چنین زیستگاه هایی می توانند در داخل یک کره دایسون، آزادانه گردش کنند و گونه های اصلاح شده آنها را می توان به بخش بیرونی کره دایسون متصل کرد.

## ساختن نیروگاه های بهتر

اگرچه کره های دایسون از لحاظ استانداردهای مهندسی امروزی دارای مصرف بهینه انرژی هستند، اما قادر به عبور از مرزهای تعیین شده توسط قوانین علم فیزیک نیستند. اینیشتن به ما یاد داد که اگر بتوان جرم را با کارایی صد در صد به انرژی تبدیل کرد و  $m$  نشان دهنده جرم و  $E$  نشان دهنده انرژی باشد و  $c$  نشان دهنده سرعت نور باشد، فرمول مشهور  $E=mc^2$  بین آنها وجود خواهد داشت. از آنجایی که سرعت نور بسیار بزرگ است، مقدار کمی از جرم می تواند مقدار بسیار بزرگی از انرژی را تولید کند. اگر ما دارای منبع فراوانی از ضد ماده<sup>۱</sup> بودیم (که نیستیم)، می توانستیم نیروگاهی با کارایی صد درصد بسازیم. ریختن یک قاشق چای خوری از "ضد آب" روی آب معمولی، انرژی معادل ۲۰۰۰۰۰ تن تی ان تی (محصول یک بمب هیدروژنی) را آزاد می کند که کل انرژی مورد نیاز جهان را ظرف حدود هفت دقیقه تامین خواهد کرد. در عوض، بیشتر راه های متداول تولید انرژی که امروزه از آنها استفاده می کنیم، به شدت ناکارا هستند. به عنوان مثال، هضم یک آب نبات فقط ۰,۰۰۰۰۰۰۰۱ درصد کارایی دارد، یعنی اینکه برای هضم آن فقط کافی است که یک تریلیونم از محتوای انرژی اش را آزاد کنیم. اگر کارایی شکم شما ۰,۰۰۱ درصد بود، برای بقیه زندگی خود فقط نیاز به خوردن یک وعده غذایی داشتید. کارایی حاصل از سوختن زغال سنگ و بنزین به ترتیب، فقط سه و پنج برابر کارایی غذایی است که می خوریم.

روش	کارایی
هضم یک آب نبات	۰,۰۰۰۰۰۰۰۰۱ درصد
سوختن زغال سنگ	۰,۰۰۰۰۰۰۰۰۳ درصد
سوختن بنزین	۰,۰۰۰۰۰۰۰۰۵ درصد
شکافت اورانیوم ۲۳۵	۰,۰۸ درصد
استفاده از کره دایسون تا مرگ خورشید	۰,۰۸ درصد
ذوب شدن هیدروژن و تبدیل آن به هلیوم	۰,۷ درصد

<sup>۱</sup> antimatter

چرخش موتور سیاه چاله	۲۹ درصد
کره دایسون اطراف کوازار	۴۲ درصد
سنگ معدن اسفالریت	۵۰ درصد
تبخیر سیاه چاله	۹۰ درصد

رآکتورهای هسته ای امروزی از عملکرد خوبی در زمینه شکافت اتم های اورانیوم برخوردار هستند، اما هنوز نمی توانند بیش از ۰,۰۸ درصد از انرژی اورانیوم را استخراج کنند و مورد استفاده مفید قرار دهند. رآکتور هسته ای موجود در داخل خورشید به مراتب کاراتر از رآکتورهایی است که ما می سازیم. رآکتورهای هسته ای خورشیدی از طریق ذوب کردن هیدروژن و تبدیل آن به هلیوم قادر به استخراج ۰,۷ درصد از انرژی هیدروژن می باشند. با این حال، حتی اگر ما بتوانیم خورشید را در داخل کره کامل دایسون محصور کنیم، هرگز قادر به تبدیل بیش از ۰,۰۸ درصد از جرم خورشید به انرژی مفید نخواهیم بود، زیرا وقتی که خورشید حدود یک دهم از سوخت هیدروژنی خود را مصرف می کند، عمرش به عنوان یک ستاره معمولی به پایان می رسد و به یک ابر ستاره سرخ تبدیل می شود و می میرد. ستاره های دیگر هم وضع بهتری ندارند و کسر هیدروژن مصرف شده آنها در طی عمرشان از ۴ درصد برای ستاره های کوچک تا ۱۲ درصد برای بزرگترین ستاره ها در نوسان است. اگر ما یک رآکتور ایده آل مصنوعی داشتیم که بتواند ۱۰۰ درصد کل هیدروژن موجود را ذوب کند، همچنان گرفتار کارایی ۰,۷ درصدی ناشی از فرایند ذوب شدن هیدروژن بودیم.

### تبخیر سیاه چاله

استفان هاو کینگ در کتاب خود با عنوان "تاریخچه مختصری از دوران روزگار" این ایده را مطرح می کند که یک نیروگاه در اطراف سیاه چاله ساخته شود. پیشنهاد او تناقض آمیز به نظر می رسد، زیرا از زمان های گذشته این باور وجود داشته است که سیاه چاله ها به قدری کوچک هستند که حتی نور هم نمی تواند از آنها فرار کند. با این حال، این طور که می گویند، هاو کینگ برآورد کرده

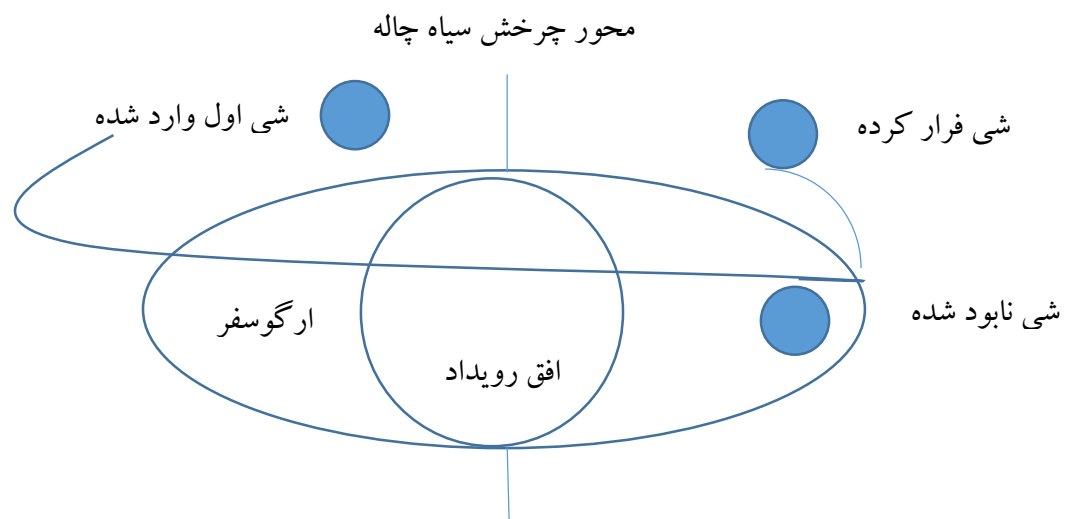
است که اثرات ناشی از جاذبه کوانتومی باعث می شود که سیاه چاله مانند یک شیء داغ عمل کند که هرچه کوچک تر باشد، داغ تر می شود و تابش موسوم به "تابش هاوکنینگ" را از خود ساطع می کند. این بدین معناست که سیاه چاله به تدریج انرژی خود را از دست می دهد و تبدیل به بخار می شود. به عبارت دیگر، تمام موادی که شما به داخل سیاه چاله می ریزید، در نهایت به صورت تابش گرمایی بر خواهند گشت. بنابر این، تا زمانی که سیاه چاله به طور کامل تبخیر شود، می توان تقریباً با صد در صد کارایی، ماده را به انرژی تابشی تبدیل کرد. یکی از مشکلاتی که در خصوص استفاده از تبخیر سیاه چاله به عنوان یک منبع نیرو وجود دارد، این است که اندازه سیاه چاله باید بسیار کوچکتر از اندازه یک اتم باشد، زیرا در غیر این صورت با فرایند به شدت کندی مواجه خواهیم شد که عمر آن از عمر فعلی جهان، طولانی تر می شود و انرژی ساطع شده از آن، از انرژی یک شمع هم کمتر خواهد شد. با افزایش مجذور اندازه چاله، انرژی تولید شده کاهش می یابد و به همین دلیل، فیزیک دان هایی مانند لوئیس کرین و شان وستمولند پیشنهاد می کنند که از سیاه چاله ای استفاده شود که اندازه آن حدود هزار برابر کوچکتر از اندازه یک پروتون باشد و وزن آن به اندازه یک کشتی دریایمای بزرگ باشد. انگیزه اصلی آنها این بود که از موتور سیاه چاله برای تولید نیروی محرکه یک استارشیپ استفاده شود و به همین دلیل، برای آنها "قدرت تحرک" مهم تر از "کارایی" بود. آنها پیشنهاد کردند که سیاه چاله را با نور لیزر پر کنند تا هیچ گونه تبدیل انرژی به ماده، صورت نگیرد. حتی اگر می توانستیم سیاه چاله را به جای "ماده" با "تابش" پر کنیم، نمی توانستیم کارایی بالایی را برایش تضمین کنیم. برای اینکه پروتون ها را بتوان وارد سیاه چاله ای کرد که اندازه آن یک هزارم اندازه پروتون هاست، باید آنها را توسط ماشینی که قدرتش به اندازه برخورددهنده هادرونی بزرگ است، در دهانه چاله حرارت داد تا انرژی جنبشی پروتئین ها حداقل هزار برابر شود. از آنجایی که در زمان تبخیر سیاه چاله، حداقل ده درصد از این انرژی جنبشی هدر می رود و تبدیل به ذرات گراویتون می شود، می توان نتیجه گرفت که انرژی ورودی به این سیاه چاله از انرژی ای که از آن استخراج می شود و به کار مفید تبدیل می شود، بیشتر خواهد شد و این در نهایت باعث منفی

شدن کارایی خواهد شد. در مورد معمای آینده ساختن نیروگاه در اطراف سیاه چاله باید گفت که هنوز نظریه جدی ای در مورد جاذبه کوانتومی وجود ندارد که بتوان محاسبات را بر اساس آن بنا کرد. با این حال، می توان امیدوار بود که اثرات جدید و مفید جاذبه کوانتومی در آینده کشف شوند.

## سیاه چاله های در حال دوران

خوشبختانه، شیوه های دیگری نیز برای استفاده از سیاه چاله ها به عنوان دستگاه تولید نیرو وجود دارد که شامل جاذبه کوانتومی یا قوانین فیزیک به خوبی درک نشده نمی باشد. به عنوان مثال، بسیاری از سیاه چاله های موجود، با سرعت زیادی می چرخند و سرعت چرخش نواحی "افق رویداد" آنها به دور خود تقریباً به اندازه سرعت نور می باشد و این انرژی چرخشی را می توان بیرون کشید و مورد استفاده قرار داد. افق رویداد یک سیاه چاله، ناحیه ای است که حتی نور هم نمی تواند از آن فرار کند، زیرا کشش جاذبه آن بسیار قوی است. همانطور که در شکل مشاهده می کنید، سیاه چاله در حال دوران در بیرون از ناحیه افق رویداد، دارای ناحیه ای بیضوی به نام "ارگوسفر" می باشد که فضا را با سرعت زیادی به دور افق رویداد می کشد، به طوریکه هیچ ذره ای نمی تواند در آن ثابت بماند. اگر یک شیء را وارد ارگوسفر کنید، به سرعت به دور سیاه چاله خواهد چرخید. متأسفانه این شیء به زودی توسط سیاه چاله نابود و ناپدید شده و برای بیرون کشیدن انرژی به درد نخواهد خورد. با این حال، راجر پنروس این موضوع را کشف کرد که اگر شما بتوانید شیء مورد نظر را با یک زاویه هوشمندانه تری پرتاب کنید و کاری کنید که به دو قطعه تقسیم شود، می توانید یک قطعه نابود شده داشته باشید و قطعه دیگر را در حالی از سیاه چاله فراری دهید که انرژی خروجی اش بیش از انرژی ورودی اش شده است. به عبارت دیگر، با این کار موفق خواهید شد که بخشی از انرژی چرخشی سیاه چاله را به انرژی مفید و قابل استفاده تبدیل کنید. اگر این فرایند را چندین بار تکرار کنید، می توانید از کل انرژی چرخشی سیاه چاله بهره برداری کنید، تا جایی که چرخش آن متوقف شود و ارگوسفر آن ناپدید شود. اگر سیاه چاله اولیه با تمام سرعتی بچرخد که طبیعت اجازه می دهد و افق رویداد آن با سرعت نور حرکت کند، این راهبرد باعث می شود که ۲۹ درصد از ماده آن به انرژی

تبدیل شود. هنوز به طور قطع نمی توان گفت که سیاه چاله ها در آسمان شب با چه سرعتی می چرخند، اما بسیاری از مطالعات نشان می دهند که آنها با سرعت بین ۳۰ در صد و ۱۰۰ درصد از حداکثر سرعت مجاز می چرخند. سیاه چاله عظیم الجثه ای که در وسط کهکشان ما واقع است و وزن آن چهار میلیون برابر وزن خورشید است، در حال چرخش به نظر می رسد. بنابراین، حتی اگر فقط ۱۰ درصد از جرم آن را بتوان به انرژی مفید تبدیل کرد، انرژی آزاد شده از آن به اندازه انرژی ای خواهد بود که ما از کره های دایسون چرخنده به دور ۵۰۰ میلیون خورشید در طی میلیاردها سال می توانیم به دست آوریم.



توضیح: بخشی از انرژی چرخشی یک سیاه چاله در حال دوران را می توان با پرتاب شیء اول وارد شده، برداشت کرد که تبدیل به بخش های شیء نابود شده و شیء فرار کرده خواهد شد.

## کوازارها (اخترش ها)

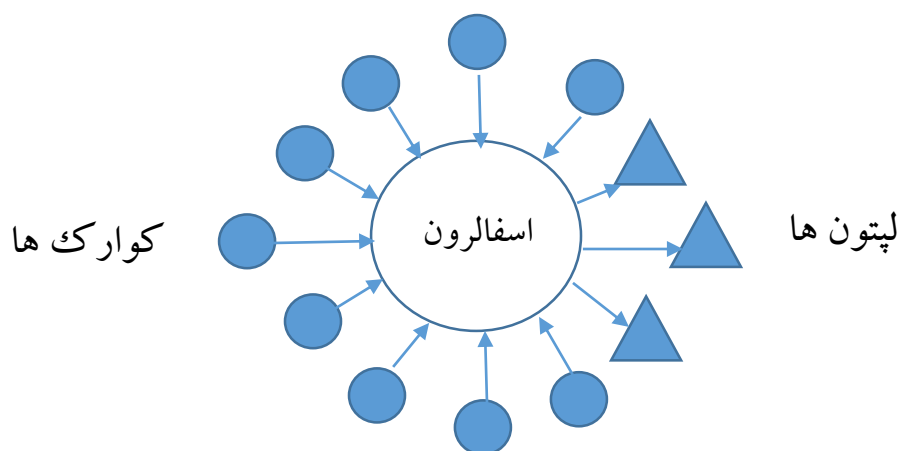
راهبرد جالب دیگر برای استخراج انرژی، این است که انرژی را به جای خود سیاه چاله، از موادی که وارد آن می شوند، استخراج کنیم. طبیعت به تنهایی توانسته است که راهی را برای انجام این کار پیدا کند و آن با استفاده از اخترش ها می باشد. وقتی که گاز اخترش می چرخد و به اطراف سیاه چاله



نزدیک می شود، یک صفحه پیتزا شکلی از گازها ایجاد می شود که داخلی ترین بخش های آن به تدریج بلعیده می شوند. وقتی که این اتفاق می افتد، گاز به شدت داغ می شود و میزان فراوانی از تابش را از خود ساطع می کند. وقتی که گاز به طرف پایین چاله می افتد، سرعتش افزایش پیدا می کند و انرژی پتانسیل گرانشی آن به انرژی جنبشی تبدیل می شود. وقتی که این آشفته‌گی و شلوغی پیچیده باعث می شود که حرکت منظم گاز به تدریج به حرکت تصادفی در مقیاس بسیار کوچک تبدیل شود، تمام اتم ها با سرعت های بالایی شروع به برخورد با یکدیگر می کنند و داغ شدن گاز هم به همین منظور انجام می شود. این برخوردهای شدید باعث می شود که انرژی جنبشی به انرژی تابشی تبدیل شود. با ساختن یک کره دایسون در اطراف سیاه چاله در فاصله ای ایمن، می توان این انرژی تابشی را گرفت و مورد استفاده قرار داد. هرچه قدر که سرعت چرخش سیاه چاله بیشتر باشد، این فرایند کاراتر خواهد شد. به طوریکه، سیاه چاله های در حال دوران با حداکثر سرعت ممکن، می توانند انرژی ای را آزاد کنند که کارایی آن به ۴۲ درصد خواهد رسید. سیاه چاله هایی که حدوداً هم وزن یک ستاره هستند، انرژی شان به صورت پرتوهای ایکس استخراج می شود و برای انواع بسیار سنگین سیاه چاله ها، بیشتر انرژی در جایی از طیف اشعه های ماورای بنفش، مرئی و مادون بنفش ظاهر خواهد شد. وقتی که سوخت شما برای پر کردن سیاه چاله تمام شد، به روشی که قبلاً گفته شد، می توانید اقدام به استخراج انرژی چرخشی از سیاه چاله ها کنید. در واقع، طبیعت از طریق یک فرایند مغناطیسی موسوم به مکانیسم بلاندفورد-زناجک، راهی را برای افزایش استخراج انرژی تابشی از گاز متراکم پیدا کرده است. شما می توانید با استفاده هوشمندانه از میدان های مغناطیسی یا فن آوری های دیگر، کارایی استخراج انرژی را به بیش از ۴۲ درصد برسانید.

## اسفالرون ها

فرایند اسفالرون راه شناخته شده دیگری برای تبدیل ماده به انرژی بدون استفاده از سیاه چاله ها می باشد. در این فرایند کوارک ها نابود می شوند و تبدیل به لپتون ها<sup>۷۲</sup> می شوند. همانطور که در شکل مشاهده می کنید، مدل استاندارد فیزیک ذرات اتمی پیش بینی می کند که نه عدد کوارک با ویژگی و چرخش مناسب می توانند دور هم جمع شوند و از طریق یک حالت بینابین موسوم به فرایند اسفالرون، تبدیل به سه عدد لپتون شوند. از آنجایی که وزن ورودی بیش از وزن خروجی است، اختلاف جرم با توجه به فرمول  $E = mc^2$  تبدیل به انرژی خواهد شد.



توضیح: نه عدد کوارک با ویژگی و چرخش مناسب می توانند دور هم جمع شوند و از طریق یک حالت بینابین موسوم به فرایند اسفالرون، تبدیل به سه عدد لپتون شوند. جرم انباشته کوارک ها بسیار بالاتر از جرم انباشته لپتون ها می باشد و این اختلاف جرم به انرژی تبدیل خواهد شد.

در زندگی هوشمندانه آینده ممکن است یک تولید کننده انرژی با استفاده از اسفالریزر ساخته شود که مانند یک موتور دیزلی عمل می کند. یک موتور دیزلی معمولی، ترکیبی از هوا و روغن دیزل را فشرده می کند و این کار را تا زمانی انجام می دهد که دما به قدری بالا رود که خود به خود باعث آتش گرفتن هوا و روشن شدن موتور شود. پس از آن، هوای داغ دوباره منبسط می شود و کار مفید

ذرات بنیادی که شامل الکترون ها، ذرات میون و تای ، نوتریون ها یا پاد ذرات دیگر می باشند.<sup>۷۲</sup>

هل دادن پیستون را انجام می دهد. وزن دی اکسید کربن و گازهای احتراقی دیگر به اندازه ۰,۰۰۰۰۰۰۰۵ درصد کمتر از ترکیب اولیه آنها در پیستون می باشد و این اختلاف جرم به انرژی گرمایی تبدیل می شود که باعث حرکت موتور می شود. نتیجه این آزمایش از قبل برایمان معلوم است، زیرا عالم هستی در حدود ۱۳,۸ میلیارد سال پیش این آزمایش را وقتی که داغ شده بود، انجام داده است. در این آزمایش، حدود ۱۰۰ درصد ماده به انرژی تبدیل شده بود و کمتر از یک میلیارد ذرات باقی مانده، شامل کوارک ها و الکترون ها می شدند. بنابراین، اسفالرون ها شبیه یک موتور دیزلی هستند که کارایی شان حدود چند میلیارد برابر یک موتور دیزلی معمولی می باشد. مزیت دیگر آنها این است که نباید نگران سوخت اسفالرون ها باشید، زیرا سوختشان می تواند از کوارک ها یا هر ماده معمولی دیگری تامین شود. این فرایندهای در دمای بالا موجب می شوند که عالم هستی ما به اندازه بیش از یک تریلیون برابر ماده (به شکل کوارک ها و الکترون های سازنده اتم)، انرژی تابشی (به شکل فوتون ها و نوترینوها) تولید کند. پس از گذشت ۱۳,۸ میلیارد سال از آن زمان، جداسازی بزرگی اتفاق افتاده که به موجب آن، اتم ها در کهکشان ها، ستاره ها و سیاره ها متمرکز شده اند و بیشتر فوتون ها در فضای میان کهکشانی مانده اند و با استفاده از تابش ریز موج زمینه کیهانی<sup>۷۳</sup>، از عالم هستی ما عکس می گیرند. بنابراین، با ایجاد دوباره شرایط بسیار داغ در داخل اسفالریزر، می توان درصد تبدیل ماده به انرژی را افزایش داد و کارایی بالایی را ایجاد کرد. برای تشخیص اینکه اسفالریزر واقعی چقدر کارایی دارد، باید به جزئیات کاربردی کلیدی ای آن مراجعه شود، از جمله اینکه بزرگی آن چقدر باید باشد تا از نشتی بخش قابل توجهی از فوتون ها و نوتریون ها در حین مرحله فشرده سازی جلوگیری شود؟ با این حال، با قطعیت می توان گفت که چشم اندازه های انرژی برای زندگی آینده بشر به شدت بهتر از آن چیزی خواهد بود که فن آوری فعلی اجازه می دهد. ما هنوز موفق به ساختن رآکتوری که جوش هسته ای در آن صورت می گیرد، نشده

<sup>۷۳</sup> Cosmic microwave background radiation

ایم. با این حال، با استفاده از فن آوری آینده ممکن است که بتوانیم رآکتورهای صد برابر بهتری را بسازیم.

## ساختن رایانه های بهتر

یکی از دوستان دانشگاهی ام به نام ستلویید پس از انجام تحقیقات مبتکرانه در مورد رایانه های کوانتومی، کتابی را نوشت که در آن استدلال کرده بود که کل عالم هستی ما یک رایانه کوانتومی است. او در مورد مرزهای نهایی محاسبه، حرف های زیادی برای گفتن داشت. او در یک مقاله مشهور در سال ۲۰۰۰، نشان داد که سرعت محاسبه به واسطه انرژی محدود می شود. او نشان داد که انجام یک عملیات منطقی در زمان  $T$  نیاز به انرژی  $E = \frac{b}{4T}$  دارد که در آن  $b$  ثابت پلانک می باشد. او نشان داد که یک رایانه یک کیلو گرمی می تواند  $5 \times 10^{50}$  عملیات را در هر ثانیه انجام دهد. اگر قدرت محاسبه در چند سال آینده دو برابر شود، در چند قرن آینده می توان به این ظرفیت دست پیدا کرد. او همچنین نشان داد که یک رایانه یک کیلو گرمی حداکثر می تواند  $10^{31}$  بیت را در خود ذخیره کند. ستلویید اعتراف کرد که امکان دستیابی به این حد از قدرت محاسبه حتی برای زندگی فوق هوشمندانه امروزی می تواند نگران کننده باشد، زیرا حافظه یک رایانه یک کیلو گرمی که از مرزهای نهایی محاسبه عبور کرده، می تواند شبیه انفجار یک بمب هیدروژنی باشد. با این حال، او خوش بین است که مرزهای واقعی و شدنی خیلی دورتر از مرزهای نهایی نیستند. در واقع، حافظه مربوط به نمونه های اولیه رایانه کوانتومی برای ذخیره سازی یک بیت در هر اتم، کوچک سازی شده است و با افزایش مقیاس حافظه آنها، می توان  $10^{25}$  بیت را در هر رایانه یک کیلو گرمی ذخیره کرد. علاوه بر این، با استفاده از تابش الکترومغناطیسی برای ایجاد ارتباط میان اتم ها، امکان انجام  $5 \times 10^{40}$  عملیات در هر ثانیه ایجاد خواهد شد. خلاصه اینکه، پتانسیل محاسبه و حل مسائل در زندگی آینده حقیقتاً ذهن بسیاری از افراد را در گیر خواهد ساخت. در واقع، فاصله زمانی لازم تا تبدیل شدن

بهترین ابررایانه های امروزی به یک رایانه یک کیلوگرمی نهایی، به اندازه فاصله زمانی کوتاه بین روشن شدن و خاموش شدن چراغ راهنمای چشمک زن یک خودرو در هر ثانیه می باشد.

## منابع دیگر

برای اینکه زندگی آینده بخواهد هر چیز فیزیکی ای را ایجاد کند، از زیستگاه ها و ماشین ها گرفته تا شکل های جدید زندگی، باید از بازآرایی ذرات بنیادی به شیوه ای خاص استفاده کند. زندگی آینده که به مرزهای فن آورانه خود نزدیک می شود، می تواند بازآرایی ذرات را با سرعت و کارایی بیشتری انجام دهد. به عنوان مثال، زندگی آینده می تواند از قدرت محاسباتی رایانه ای برای کشف کاراترین روش و استفاده از انرژی موجود خود برای تقویت فرایند بازآرایی ماده استفاده کند. هر وقت که زندگی آینده با محدودیت های فیزیکی در مورد بهره برداری از ماده مواجه شد، باید ماده بیشتری به دست آورد تا بتواند از مرزهای عالم هستی عبور کند.

## دستیابی به منابع از طریق سکنی گزیدن در عالم هستی

اکنون این سوال مطرح می شود که عظمت موهبت های جهان هستی چقدر است و حدود بالای مقرر شده توسط قوانین علم فیزیک برای میزان ماده ای که می توان برای زندگی کردن از آن استفاده کرد، کدام است؟ موهبت های جهان هستی به قدری زیاد هستند که قابل محاسبه نمی باشند. جدول زیر به برخی از این اعداد اشاره می کند. در ۹۹,۹۹۹۹۹۹ در صد از ماده سیاره ما امکان زندگی وجود ندارد و این بخش بزرگ، کار مفیدی به جز تامین کشش جاذبه و میدان مغناطیسی را انجام نمی دهد. اگر ما بتوانیم تمام این ماده را در منظومه شمسی و از جمله در خورشید مورد استفاده بهینه قرار دهیم، می توانیم بهره برداری قابل توجهی از منابع داشته باشیم و صدها میلیون برابر ماده فعلی را در زندگی خود مورد استفاده قرار دهیم.

منطقه	ذرات
زیست کره ما	۱۰ <sup>۴۳</sup>

سیاره ما	۱۰.۵۱
منظومه شمسی ما	۱۰.۵۷
کهکشان ما	۱۰.۶۹
کل عالم هستی ما	۱۰.۷۸

توضیح: تعداد تقریبی ذرات ماده (پروتون و نوترون) قابل استفاده در زندگی آینده

## چقدر دورتر می توانیم برویم؟

ممکن است تصور کنید که اگر به حد کافی صبور باشید، می توانید منابع نامحدودی را از طریق سکنی گزیدن در عالم هستی به تصرف خود در آورید. اما این چیزی نیست که کیهان شناسی پیشرفته پیشنهاد می کند. طبق آنچه که ساده ترین ویرایش های تورم کیهانی پیش بینی کرده اند، کیهان می تواند نامحدود باشد و شامل بی نهایت کهکشان، ستاره و سیاره باشد. با این حال، حتی در صورتی که بی نهایت کهکشان در عالم هستی وجود داشته باشد، به نظر می رسد که ما فقط می توانیم به تعداد محدودی از آنها دسترسی داشته باشیم. ما فقط می توانیم حدود ۲۰۰ میلیارد کهکشان را مشاهده کنیم و حداکثر در ده میلیارد از آنها سکنی گزینیم. آنچه که ما را محدود می کند، سرعت نور است که به اندازه یک سال نوری (حدود ده تریلیون کیلومتر) در هر سال می باشد. حتی اگر کیهان نامحدود باشد، جهان هستی ما محدود است و شامل حدود  $10^{78}$  اتم می باشد. علاوه بر این، حدود ۹۸ درصد از جهان هستی ما فقط قابل دیدن است و قابل لمس نیست، بدین معنا که ما می توانیم آن را ببینیم، ولی حتی با سرعت نور هم نمی توانیم به آن دسترسی پیدا کنیم. دلیل اینکه ما نمی توانیم دورتر از یک حدی را ببینیم، این است که نوری که از فاصله دور جهان هستی می خواهد به ما برسد، هنوز فرصت دسترسی به ما را پیدا نکرده است. اما در صورتی که حد زمانی مشخصی برای رسیدن به کهکشان های دور وجود نداشته باشد، آیا ما نباید به کهکشان های دور سفر کنیم؟ چالش اول این است که جهان هستی ما در حال گسترش است و این بدین معناست که تقریباً تمام کهکشان ها در

حال فرار کردن از ما هستند. به همین دلیل، سکنی گزیدن در کهکشان‌های دور به مثابه یک بازی است که هدفش رسیدن به آن موجودیت‌هایی است که از ما سبقت گرفته و جلوتر از ما می‌باشند. چالش دوم این است که به دلیل وجود انرژی تاریک مرموز و نامرئی‌ای که حدود ۷۰ درصد از جهان هستی ما را تشکیل می‌دهد، سرعت گسترش جهان در حال افزایش است. برای درک این مشکل، تصور کنید که وارد یک سکوی قطار شده‌اید و می‌بینید که قطار با سرعت در حال عبور از شما می‌باشد، اما در قطار به نشانه دعوت از شما همچنان باز است. اگر شما سریع و بی‌باک باشید، آیا می‌توانید سوار قطار شوید؟ از آنجایی که سرعت قطار در نهایت بیش از سرعت دویدن شما خواهد بود، پاسخ به این سوال، بستگی به این دارد که قطار از اول چقدر از شما دور بوده است. اگر فاصله اولیه شما با قطار بیش از حد مشخصی بوده باشد، هرگز به قطار نخواهید رسید. وقتی که سعی می‌کنید به کهکشان‌های دوری برسید که با شتاب از شما دور می‌شوند، با همین وضعیت مواجه خواهید شد. حتی اگر شما بتوانید با سرعت نور به آنها سفر کنید، هرگز قادر به دسترسی به کهکشان‌هایی که بیش از ۱۷ میلیارد سال نوری با شما فاصله دارند، نخواهید بود و این موضوع در مورد بیش از ۹۸ درصد از کهکشان‌های جهان صادق است. ممکن است بپرسید که مگر نظریه نسبیت خاص اینشتین نمی‌گوید که هیچ چیزی نمی‌تواند با سرعتی بیش از سرعت نور، حرکت کند؟ پس کهکشان‌ها چگونه می‌توانند سریعتر از نور حرکت کنند؟ پاسخ این سوال این است که نظریه نسبیت خاص اینشتین، جای خود را به نظریه نسبیت عام اینشتین داده است که در آن، محدودیت‌های سرعت تا حدودی تعدیل شده است. طبق نظریه جدید، هیچ چیزی در سراسر فضا نمی‌تواند سریعتر از نور حرکت کند، اما فضا می‌تواند با هر سرعتی که می‌خواهد، گسترش یابد. طبق نظریه فضا-زمان، جهان دارای سه بعد مکانی و یک بعد زمانی مستقل می‌باشد. اگر فضا گسترش پیدا نمی‌کرد، پرتوهای نور، خطوط مورب ۴۵ درجه‌ای را در فضا-زمان تشکیل می‌دادند. به همین دلیل، مناطقی که می‌توانستیم ببینیم و به آنها دسترسی پیدا کنیم، مخروطی شکل به نظر می‌رسیدند. در حالی که مخروط‌های نور مربوط به گذشته، بریده و کوتاه شده هستند، مخروط‌های نور در آینده می‌توانند تا

ابد، گسترش پیدا کنند و امکان دسترسی به موهبت های نامحدود جهان هستی را برایمان فراهم کنند. در عوض، پرتوهای نور موجود در جهان در حال گسترش ما (همان جهانی که ظاهرا در آن زندگی می کنیم و شامل انرژی تاریک نیز می باشد)، شکل عمودی پیدا کرده اند و این باعث شده که تعداد کهکشان های قادر به سکنی گزیدن در آنها، به حدود ۱۰ میلیارد کهکشان، محدود شود. اگر این محدودیت ها باعث ترس شما از جهان هستی شده، اجازه دهید که با ارائه یک راه گریز ممکن، شما را خوشحال کنم. محاسبات من نشان می دهد که انرژی تاریک با گذشت زمان، ثابت می ماند و این موضوع را جدیدترین اندازه گیری های امروزی تأیید می کنند. با این حال، ما هنوز مدرکی نداریم که نشان دهد انرژی تاریک واقعا چیست و فقط امیدواریم که انرژی تاریک در نهایت نابود شود و اگر این اتفاق بیفتد، "افزایش سرعت" جای خود را به "کاهش سرعت" خواهد داد و امکان سکنی گزیدن در کهکشان های جدید در شکل های زندگی آینده فراهم خواهد شد.

### با چه سرعتی می توانیم به فضا سفر کنیم؟

همانطور که توضیح داده شد، اگر کهکشان ها با سرعت نور در کلیه جهات گسترش پیدا کنند، امکان زندگی انسان ها بر روی آنها فراهم خواهد شد. نظریه نسبیت عام اینشتین می گوید که فرستادن موشک به فضا با سرعت نور غیر ممکن است، زیرا این کار به انرژی نامحدودی نیاز خواهد داشت. بنابراین، ممکن است پرسیده شود که در عمل، موشک ها با چه سرعتی می توانند وارد فضا شوند؟ وقتی که موشک افق های نوی<sup>۷۴</sup> ناسا (سازمان ملی هوانوردی و فضایی) در سال ۲۰۰۶ و با سرعت حدود ۱۰۰۰۰۰ مایل در ساعت (۴۵ کیلومتر در ثانیه) به سمت پلوتون به پرواز درآمد، رکورد سرعت را شکست. کاوشگر خورشیدی پلاس ناسا تصمیم دارد که در سال ۲۰۱۸ با سرعت چهار برابر موشک افق های نو در فاصله بسیار نزدیکی نسبت به خورشید، فرود آید. با این حال، چنین سرعتی فقط به اندازه ۰٫۱ درصد سرعت نور می باشد. تقاضا برای پرتاب موشک های سریعتر و بهتر، ذهن با هوش ترین دانشمندان قرن گذشته را به خود مشغول کرده و مقالات و متون علمی ارزشمند و جالبی

---

<sup>۷۴</sup> New Horizons rocket



در رابطه با این موضوع وجود دارد. اما، چرا نمی توان با سرعت بیشتری به فضا سفر کرد؟ دو مشکل اساسی وجود دارد. مشکل اول، این است که موشک های معمولی بیشتر سوخت خود را صرف شتاب بخشیدن به سوختی می کنند که در حال حمل آن می باشند و مشکل دوم این است که سوخت موشک های امروزی فاقد کارایی هستند. به طوری که نسبت جرم بنزینی که به انرژی تبدیل می شود، بیش از ۰,۰۰۰۰۰۰۵ درصد جرم بنزین اولیه نمی باشد. برای بهبود این وضعیت می توان از سوخت کارتری استفاده کرد. به عنوان مثال، فریمن دایسون و همکارانش در ناسا که روی پروژه اوریون کار می کردند، تصمیم گرفتند که حدود ۳۰۰۰۰۰۰ بمب هسته ای را در طی ۱۰ روز منفجر کنند تا با استفاده از انرژی حاصل شده، به حدود ۳ درصد از سرعت نور برسند و با استفاده از فضاپیمای حامل انسان به یک منظومه شمسی دیگر، سفر کنند. دیگران هم به دنبال این هستند که از ضد ماده به عنوان سوخت استفاده کنند، زیرا ترکیب آن با ماده معمولی باعث آزاد شدن انرژی ای می شود که ۱۰۰ درصد کارایی دارد. ایده رایج دیگر این است که موشکی بسازیم که نیازی به حمل سوخت خود ندارد. به عنوان مثال، فاصله میان ستارگان، فضای خالی ایده آلی نمی باشد، اما شامل یون هیدروژنی ای می باشد که از اتم هیدروژنی یک پروتونی و بدون الکترون ساخته شده است. در سال ۱۹۶۰ یک فیزیکدان آمریکایی به نام رابرت بوسارد از ایده ای استفاده کرد که به ایده "بوسارد رمجت"<sup>۷۵</sup> مشهور است. این ایده شامل برداشتن یون های هیدروژنی در مسیر سفر و استفاده از آنها به عنوان سوخت موشک در رآکتور مخصوص جوش هسته ای موجود در داخل موشک است. با وجود تردیدهایی که برای عملی ساختن این ایده وجود دارد، ایده دیگری موسوم به "سفر لیزری" نیز برای حمل نکردن سوخت وجود دارد که برای ساختن یک تمدن در فاصله خیلی دوری در فضا، شدنی به نظر می رسد. این ایده برای اولین بار توسط فیزیکدانی به نام رابرت فوروارد در سال ۱۹۸۴ مطرح شد که ایده ساختن کره دایسون را نیز خودش طراحی کرده بود. درست همانطوری که مولکول های هوا با برخورد به بادبان یک کشتی باعث حرکت کشتی به سمت جلو می شوند، ذرات

---

<sup>۷۵</sup> Bussard ramjet

نور (فوتون هایی) که به یک بازتاب دهنده برخورد می کنند، با تولید انرژی تابشی باعث جلو بردن کشتی فضایی می شوند. وقتی که یک لیزر بزرگ که از خورشید انرژی می گیرد، به یک بادبان لیزری فوق سبک متصل به کشتی فضایی، پرتو افکنی می کند، می توان از انرژی خورشید برای افزایش سرعت موشک استفاده کرد. اما ممکن است پرسید که چگونه آن را متوقف کنیم؟ پاسخ این سوال را با خواندن یکی از مقالات فوروارد پیدا کردم. در پوسته بیرونی بادبان لیزری یک حلقه وجود دارد که می توان آن را از جای خود درآورد و در مقابل کشتی فضایی قرار داد. این حلقه باعث می شود که پرتو نور لیزر به عقب بازتاب داده شود و در نتیجه، سرعت کشتی فضایی و بادبان کوچک آن کاهش پیدا کند. فوروارد حساب کرد که انسان ها با این روش می توانند سفر چهار سال نوری به سیستم ستاره ای آلفا قنطورس را فقط در چهل سال انجام دهند. تصور کنید که به محض رسیدن به آنجا بتوانیم یک سیستم لیزری بزرگ و جدیدی را بسازیم و در سراسر کهکشان راه شیری گردش کنیم. اما چرا باید متوقف شویم؟ در سال ۱۹۶۴، یک فضانورد روسی به نام نیکلای کارداشف پیشنهاد طبقه بندی تمدن ها را بر اساس میزان انرژی مفیدی که به کار می گیرند، مطرح کرد. او تمدن ها را بر اساس اینکه انرژی را یک سیاره، یک ستاره یا یک کهکشان بهره برداری می کنند، به ترتیب به تمدن های نوع یک، نوع دو و نوع سه طبقه بندی کرد. متفکران بعدی، نوع چهارمی از تمدن ها را بر اساس بهره برداری از کل جهان هستی ما پیشنهاد کرده اند. از آن زمان به بعد، خبرهای خوب و خبرهای بدی برای اشکال مختلف زندگی جاه طلبانه ارائه شده است. خبر بد این است که انرژی تاریک هنوز در جهان وجود دارد که ظاهراً باعث محدود شدن دسترسی ما به کهکشان ها می شود و خبر خوب این است که پیشرفت چشمگیری در زمینه هوش مصنوعی حاصل شده است. حتی نظریه پردازان خوشبینی مانند کارل ساگان، رویای انسان ها را برای دسترسی به کهکشان های دیگر، غیر ممکن می دانند. حتی اگر ما انسان ها بتوانیم با سرعت نور به فضا سفر کنیم، به دلیل گرایشمان به مردن پس از گذشت اولین قرن از سفر چند میلیون ساله مان به فضا، قادر به ادامه این سفر نخواهیم بود. امکان استفاده از هوش فوق بشری می تواند تصویر موجود را کاملاً تغییر داده و امیدواری

افرادی را که عاشق سفر بین کهکشان‌ها هستند، افزایش دهد. اگر از ماشین‌های هوش مصنوعی استفاده شود، سکنی گزیدن در بین کهکشان‌ها با استفاده از فن‌آوری اختراع شده توسط هوش مصنوعی نسبتاً آسان خواهد شد و دیگر نیازی به حمل‌لوازم مورد نیاز انسان که جای زیادی را اشغال می‌کنند، نخواهد بود. به این ترتیب، هزینه سفر لیزری به میزان قابل توجهی کاهش خواهد یافت و دیگر نیازی به حمل دستورالعمل‌ها با خود وجود نخواهد داشت. تنها کاری که باید انجام داد، این است که یک آنتن گیرنده کوچک ساخت تا ربات‌ها با استفاده از آن بتوانند با سرعت نور، اطلاعات مورد نیاز خود را در قالب دستورالعمل‌های اجرایی از انسان‌ها دریافت کنند. در فضای تاریک بین کهکشان‌ها تعداد قابل توجهی از ستاره‌های بین کهکشان‌ها وجود دارند که به عنوان ایستگاه‌های بین راه عمل می‌کنند و امکان سفر لیزری بین کهکشان‌ها را فراهم می‌سازند. انفجار هوش باعث خواهد شد که انسان‌ها به جای فکر کردن در مورد شدنی بودن اقامتشان در بین کهکشان‌ها، به فکر پیشرفت سریعتر این روند باشند و انگیزه زیادی برای عبور از محدودیت‌ها داشته باشند، زیرا در رقابت برای رفع محدودیت‌های زمان و انرژی تاریک، هر افزایش یک درصدی در سرعت متوسط سکنی گزیدن در کهکشان‌ها به معنی افزایش سه درصدی در کهکشان‌های به استعمار درآمده خواهد بود. به عنوان مثال، اگر سفر ۱۰ سال نوری به سیستم ستاره‌ای دیگر، ۲۰ سال طول بکشد و ۱۰ سال دیگر طول بکشد که در آن زندگی کرد، منطقه‌ای از فضا که قابل سکنی گزیدن شده است، مانند یک کره خواهد بود که با یک سوم سرعت متوسط نور، در تمام جهات رشد خواهد کرد. خلاصه اینکه، به نظر من بیشتر دانشمندان و نویسندگان داستان‌های علمی-تخیلی که به بررسی موضوع سکنی گزیدن در عالم هستی پرداخته‌اند، در خصوص نادیده گرفتن امکان استفاده از ماشین‌های هوش مصنوعی فوق‌هوشمند، بیش از حد بدبین بوده‌اند و با محدود کردن توجه خود به مسافران انسانی، مشکل سفر به کهکشان‌ها را بیش از حد برآورد کرده‌اند و با محدود کردن توجه خود به فن‌آوری‌هایی که توسط انسان‌ها، اختراع شده‌اند، زمان مورد نیاز برای گذشتن از مرزهای فیزیکی عالم را بیش از حد تخمین زده‌اند.

## در ارتباط ماندن کهکشان ها با یکدیگر از طریق مهندسی عالم هستی

همانطور که داده های تجربی اخیر تأیید می کنند، اگر انرژی تاریک باعث شتاب بخشیدن به روند فاصله گرفتن کهکشان های دور از یکدیگر شود، این امر باعث ایجاد دردسرهای بزرگی برای آینده زندگی بشر خواهد شد. این امر بدین معناست که حتی اگر یک تمدن در آینده، اقدام به مدیریت زندگی بشر در یک میلیون کهکشان کند، انرژی تاریک طی یک دوره ده ها میلیارد ساله باعث فروپاشی امپراطوری جهانی انسان بر هزاران منطقه مختلف خواهد شد، به طوریکه این مناطق حتی قادر به برقراری ارتباط با یکدیگر نخواهند بود. اگر زندگی آینده نتواند جلوی این متلاشی شدن را بگیرد، بزرگترین تکیه گاه های باقی مانده شامل خوشه ای از هزاران کهکشان خواهد بود که جاذبه ترکیبی آنها قوی تر از حدی خواهد بود که انرژی تاریک بتواند آنها را از هم جدا کند. این امر باعث می شود که اگر یک تمدن حاصل از ماشین های فوق هوشمند بخواهد با کهکشان ها در ارتباط بماند، انگیزه قوی ای برای مهندسی عالم در مقیاس وسیع پیدا کند. این سوال ممکن است پیش بیاید که اگر انسان بخواهد در برابر عدم دسترسی همیشگی به کهکشان ها که ناشی از انرژی تاریک می باشد، مقاومت کند، در زمان محدود چه میزان از ماده را می تواند به بزرگترین ابرخوشه منتقل کند؟ یکی از روش ها برای انتقال یک ستاره به مسافت های دور این است که ستاره سومی را تحریک کرد که وارد سیستمی شود که در آن، دو ستاره به نوبت در مدار یکدیگر حرکت می کنند. مانند آنچه که در روابط عاشقانه اتفاق می افتد، معرفی شریک سوم می تواند ایجاد بی ثباتی کند و باعث شود که یکی از آن سه نفر به شدت بیرون رانده شود یا در مورد ستاره ها با سرعت زیادی به بیرون پرتاب شود. متأسفانه، این تکنیک سه نفری که در مورد ستاره ها، سیاه چاله ها یا کهکشان ها می تواند به کار گرفته شود، فقط باعث انتقال بخش کوچکی از توده تمدن ها به مکان های دور برای فریب دادن و سبقت گرفتن از انرژی تاریک می شود. با این حال، از هوش مصنوعی فوق بشری می توان برای یافتن روش های بهتری مانند تبدیل توده جمع شده در کهکشان های دورافتاده به فضایی که می

تواند به خوشه خانه خود سفر کند، استفاده کرد. همچنین در صورت ساختن یک اسفالریز می توان ماده را به انرژی ای تبدیل کرد که به صورت نور به خوشه خانه تابیده شود. همچنین این انرژی را می توان دوباره به ماده تبدیل کرد یا به عنوان منبع انرژی از آن استفاده کرد. اگر امکان استفاده از کرم چاله<sup>۷۶</sup> قابل عبور از بین دو انتهای آن فراهم شود، می توان به صورت لحظه ای بین دو نقطه بسیار دور از یکدیگر ارتباط برقرار کرد و صرف نظر از اینکه چقدر این دو نقطه از هم فاصله دارند، بین دو انتهای کرم چاله، حرکت کرد. کرم چاله یک میانبر در فضا- زمان است که امکان سفر بین دو نقطه دور از هم را بدون طی کردن فضای بین آنها فراهم می سازد. با وجود اینکه کرم چاله ها در نظریه نسبیت عام اینشتین تأیید شده اند، نیاز به نوعی ماده فرضی با چگالی منفی دارند که وجود آن هم بستگی به اثرات جاذبه کوانتومی کمتر شناخته شده دارد. به عبارت دیگر، ممکن است معلوم شود که ساختن کرم چاله های مفید، غیر ممکن است، اما در صورت ممکن بودن، ماشین های هوش مصنوعی فوق بشری انگیزه زیادی برای ساختن آن خواهند داشت. کرم چاله ها نه تنها باعث ایجاد تحول در ارتباطات سریع در داخل کهکشان ها خواهند شد، بلکه با پیوند کهکشان های دوردست به یک کهکشان خوشه ای مرکزی در همان اوایل کار باعث می شوند که کهکشان ها در بلندمدت با یکدیگر در ارتباط بمانند و تلاش های انرژی تاریک برای سانسور کردن ارتباطات آنها کاملاً خنثی شود و بی نتیجه بماند. وقتی که دو کهکشان توسط یک کرم چاله پایدار به یکدیگر متصل می شوند، صرف نظر از اینکه در چه فاصله ای از هم قرار دارند، اتصالشان حفظ خواهد شد. اگر با وجود تمام تلاش های صورت گرفته برای مهندسی عالم هستی، یکی از تمدن های آینده به این نتیجه برسد که بخش هایی از این عالم تا ابد محکوم به حرکت بدون تماس با بخش های دیگر هستند، می تواند آن بخش ها را رها کند و برایشان آرزوی موفقیت کند. اما، اگر تمدنی دارای اهداف محاسباتی جاه طلبانه ای برای پاسخ به سوالات خاص و مشکل خود باشد، ممکن است به راهبرد "ضربه زدن سریع و سوزاندن خود" متوسل شود. بدین معنا که این تمدن می تواند کاری کند که کهکشان های

---

کرم چاله در واقع تونلی بین دو نقطه ی دوردست از جهان است که زمان رسیدن از یک نقطه به نقطه ی دیگر را کوتاه می سازد: Wormhole<sup>۷۶</sup>

دوردست به رایانه های بزرگی مبدل شوند که با سرعت دیوانه واری، ماده و انرژی را به "محاسبه" تبدیل می کنند، با این امید که قبل از اینکه انرژی تاریک برای محو بقایای سوزانده شده رایانه ها فشار بیاورد، بتوانند پاسخ هایی را که مدت ها به دنبال آن بودند، به خوشه های مادر منتقل کنند. این راهبرد خصوصا برای مناطق بسیار دور دست، مناسب خواهد بود. وقتی که این تمدن به محل سکونت خود در منطقه مادر باز می گردد، تلاش خواهد کرد که کارایی خود را تا هر زمانی که ممکن است، به حداکثر رساند.

### تا چه زمانی می توانید دوام آورید؟

"درازی عمر" چیزی است که بیشتر افراد جاه طلب، سازمان ها و ملت ها آرزوی آن را دارند. اگر یک تمدن جاه طلب بخواهد که طول عمر بیشتری داشته باشد، تا چه زمانی می تواند دوام بیاورد؟ اولین تحلیل علمی جامع در مورد آینده دور توسط فریمن دایسون انجام شد و جدول زیر برخی از یافته های کلیدی را خلاصه کرده است. نتیجه حاصل از تحقیقات این بود که اگر هوش مصنوعی دخالت نکند، منظومه های شمسی و کهکشان ها به تدریج نابود می شوند و به فضایی سرد، مرده و خالی از سکنه تبدیل می شوند که تا ابد تابش نوری در آن رخ نخواهد داد. با این حال، فریمن با حالت امیدواری می گوید: "دلایل علمی خوبی برای جدی گرفتن احتمال اینکه عالم هستی تحت تاثیر پیشرفت های زندگی و هوش قرار گیرد، وجود دارد." من فکر می کنم که هوش مصنوعی فوق بشری بسیاری از مسائل را حل خواهد کرد. یکی از این چالش ها مرگ خورشید در چند میلیارد سال آینده است که دیگر یک عامل بازدارنده نخواهد بود، زیرا حتی یک تمدن نسبتا ابتدایی به آسانی می تواند به ستاره هایی نقل مکان کند که عمر ۲۰۰ میلیارد ساله دارند.

چه	چه زمانی (به سال)
سن فعلی عالم هستی	۱۰ <sup>۱۰</sup>
عدم دسترسی به بیشتر کهکشان ها به دلیل وجود انرژی تاریک	۱۰ <sup>۱۱</sup>
آخرین ستاره ها خاموش می شوند	۱۰ <sup>۱۴</sup>

سیاره ها از ستاره ها جدا می شوند	۱۰۱۵
ستاره ها از کهکشان ها جدا می شوند	۱۰۱۹
محو شدن مدارها توسط تابش گرانشی	۱۰۲۰
نابودی پروتون ها	بیش از ۱۰ <sup>۳۴</sup>
تبخیر سیاه چاله های جرم ستاره ای	۱۰ <sup>۶۷</sup>
تبخیر سیاه چاله کلان	۱۰ <sup>۹۱</sup>
تبدیل تمام مواد به آهن	۱۰ <sup>۱۵۰۰</sup>
تبدیل تمام مواد به سیاه چاله و تبخیر سیاه چاله	۱۰ <sup>۱۰۲۶</sup>

اگر تمدن های مجهز به هوش مصنوعی فوق بشری متوجه شوند که ساختن نیروگاه هایی در فضا کاراتر از بهره برداری از ستاره ها می باشد، در واقع ممکن است تصمیم بگیرند که برای حفظ انرژی، از تشکیل ستاره ها جلوگیری کنند. حتی اگر آنها از کره دایسون برای بهره برداری از تمام انرژی خروجی یک ستاره در طول عمر خود استفاده کنند ( که جبران کننده فقط ۰,۱ درصد از کل انرژی ورودی آن خواهد بود ) ، قادر به حفظ ۹۹,۹ درصد از انرژی اتلاف شده به دلیل مرگ ستاره های فوق سنگین نخواهند بود. وقتی که یک ستاره سنگین بر اثر انفجار می میرد، بیشتر انرژی آن به صورت نوترینوهای گریزان، ناپدید می شود و با مرگ ستاره های بسیار سنگین، میزان قابل توجهی از جرم ستاره با تشکیل سیاه چاله ای که انرژی در طی ۱۰<sup>۶۷</sup> سال از آن به بیرون تراوش خواهد کرد، تلف خواهد شد. تازمانی که ماده یا انرژی مورد نیاز برای زندگی ماشین های هوش مصنوعی تمام نشده است، زندگی ماشین ها در حالت مورد علاقه آنها ادامه خواهد یافت. هوش مصنوعی با استفاده از روشی در مکانیک کوانتومی موسوم به "اقدام کردن به جای منتظر ماندن" می تواند راهی را برای جلوگیری از متلاشی شدن پروتون ها کشف کند که بر اساس آن، با انجام نظارت های منظم ، می توان روند متلاشی شدن پروتون ها را کند کرد. با این حال، یک عامل بازدارنده بالقوه وجود دارد و آن یکی از قمرهای سیاره زحل می باشد که در طی ۱۰ تا ۱۰۰ میلیارد سال آینده کل عالم هستی را تخریب خواهد کرد. همچنین، با کشف

انرژی تاریک و پیشرفت های صورت گرفته در زمینه نظریه ریسمان<sup>۷۷</sup> سناریوهای جدیدی مطرح شده است که فریمن دایسون هنگام نوشتن مقاله تاثیرگذار خود از آنها آگاه نبوده است. اکنون این سوال مطرح می شود که عالم هستی ما در میلیاردها سال بعد، چگونه به پایان می رسد؟ یکی از این سناریوها، منجمد شدن است. به این معنی که عالم هستی تا ابد انبساط پیدا می کند و در نهایت، به یک مکان سرد، تاریک و مرده، دگرگون می شود و این یکی از محتملترین سناریوها می باشد. در سناریوی دیگر، جهان به جای یخ بندان شدن، با آتش گرفتن تمام می شود که نام این سناریو، خرد شدن می باشد که در آن، انبساط جهان در نهایت بر عکس می شود و همه چیز در اثر یک تحول عظیم، دوباره از کار می افتد و متلاشی می شود. در سناریوی بعدی همه چیز از جمله کهکشان ها، سیاره ها و حتی اتم ها در نهایت و در اثر وارد کردن فشار، تکه پاره می شوند. این که کدامیک از این سناریوها اتفاق می افتد، بستگی به این دارد که انرژی تاریک که حدود ۷۰ درصد جرم عالم هستی را تشکیل می دهد، با تداوم گسترش فضا، امکان تحقق کدامیک از این سناریوها را فراهم خواهد ساخت. اگر انرژی تاریک بدون تغییر در یک جا بماند، جهان منجمد می شود و اگر چگالی آن (جرم آن در حالت حجم) کاهش پیدا کند، جهان خرد خواهد شد و با افزایش چگالی انرژی تاریک، جهان در نهایت تکه پاره خواهد شد. با این حال، هنوز مدرکی دال بر اینکه انرژی تاریک چیست، وجود ندارد. گزینه دیگری هم وجود دارد و آن این است که هیچ یک از این سناریوها اتفاق نخواهد افتاد. من فکر می کنم که ما انسان ها باید متواضع باشیم و اعتراف کنیم که چیزهای زیادی در جهان وجود دارند که ما هنوز قادر به درک آنها نیستیم که ماهیت فضا یکی از آنها می باشد. در تمام این سناریوها فرض شده که خود فضا ثابت است و از قابلیت انبساط نامحدودی برخوردار می باشد. ما فضا را به عنوان صحنه نمایش ثابت و خسته کننده ای در نظر می گیریم که نمایشنامه عالم هستی روی آن اجرا می شود. اینشتین به ما یاد داد که خود فضا نیز یکی از بازیگران مهم است که می تواند انبساط پیدا کند، خم شود یا موج دار شود. نظریه اینشتین می گوید که انبساط فضا می تواند برای همیشه ادامه پیدا کند و همانطوری که در سناریوهای

---

<sup>۷۷</sup> String theory



منجمد شدن و پاره شدن گفته شد، ممکن است که حجم عالم هستی ما به بی نهایت نزدیک شود. این اتفاق خوبی به نظر می رسد، اما من به درست بودن آن شک دارم. فضا مانند بند لاستیکی ای می باشد که اگر آن را بیش از حد منبسط کنید، پاره خواهد شد. زیرا لاستیک از اتم هایی ساخته شده که انبساط کافی، می تواند ماهیت اتم های آن را که دارای دانه های ریزی می باشد، برجسته کند. آیا ممکن است که فضا نیز دارای دانه های بسیار ریزی باشد که به قدری کوچک هستند که کسی متوجه آن نشده است؟ تحقیقات انجام شده در زمینه گرانش کوانتومی درستی این موضوع را تأیید می کنند. اگر واقعا حقیقت داشته باشد که بدون پاره شدن فضا، نمی توان آن را بی نهایت منبسط کرد، تمدن های آینده ممکن است تصمیم بگیرند که به بزرگترین منطقه غیر قابل منبسط شدن فضا (یک خوشه کهکشان عظیم) که قابل دسترسی است، نقل مکان کنند.

### تا چه اندازه می توانید محاسبه کنید؟

پس از بررسی این که زندگی آینده چقدر دوام خواهد یافت، اکنون به بررسی این موضوع می پردازیم که تا چه اندازه می خواهد ادامه پیدا کند؟ طبیعی است که شما می خواهید که تا جایی که ممکن است، زندگی طولانی تری داشته باشید. فریمن دایسون یک استدلال کمیت پذیر را برای این خواسته ارائه کرده است. او می گوید که اگر کندتر محاسبه کنید، هزینه محاسبه کاهش خواهد یافت و اگر همه چیز را تا حد ممکن به کندی محاسبه کنید، کارهای بیشتری را در نهایت به پایان خواهید رساند. او حتی برآورد کرده است که اگر جهان هستی ما تا ابد انبساط پیدا کند و منجمد شود، امکان انجام بی نهایت محاسبه فراهم خواهد شد. کندی لزوما به معنی خسته کننده نیست. اگر در آینده در یک دنیای شبیه سازی شده زندگی کنیم، جریان زمانی ای که به طور ذهنی تجربه می شود، ربطی به سرعت بسیار کند اجرای شبیه سازی در دنیای بیرون نخواهد داشت. به همین دلیل، چشم اندازه های انجام بی نهایت محاسبه، به جاودانگی ذهنی در زندگی شبیه سازی شده منجر خواهد شد. فرانک تیلر از این ایده برای پشتیبانی از فرضیه جاودانگی ذهنی در لحظات آخری که سرعت محاسبات به بی نهایت می رسد و جهان بر اثر افزایش دما و چگالی، خرد خواهد شد، استفاده کرده است. با توجه

به اینکه انرژی تاریک باعث خراب شدن رویاهای محاسبه بی نهایت فریمن و فرانک خواهد شد، ماشین های هوش مصنوعی آینده ترجیح می دهند که قبل از اینکه با مشکلاتی مانند افق های رویداد کیهانی یا از بین رفتن پروتون مواجه شوند، منابع انرژی خود را با سرعت نسبتاً زیادی بسوزانند. اگر هدف نهایی این باشد که تعداد محاسبات به حداکثر برسد، بهترین راهبرد این است که سرعت محاسبات نه این قدر کم باشد که شما را با مشکلات فوق الذکر مواجه سازد و نه این قدر زیاد باشد که باعث صرف انرژی بیشتری برای هر محاسبه شود. در مجموع، می توان گفت که کارایی حداکثری نیروگاه ها و رایانه ها باعث می شود که هوش مصنوعی فوق بشری بتواند تعداد محاسبات گیج کننده ای را انجام دهد. افزایش قدرت مغز سیزده واتی شما برای صد سال بعد به انرژی موجود در نیم میلیگرم از ماده ای نیاز دارد که کوچکتر از یک دانه شکر معمولی می باشد. به گفته ست لیودز، بازدهی انرژی مغز انسان را می توان به حداکثر رساند. اگر از تمام ماده موجود در جهان هستی برای شبیه سازی مغز انسان استفاده شود، هوش مصنوعی با قدرت محاسباتی بالای خود می تواند مغز <sup>۶۹</sup>۱۰ انسان را شبیه سازی کند. از طرفی، نیک بوستروم تخمین زده است که با فرضیات محافظه کارانه تر امکان شبیه سازی مغز <sup>۵۸</sup>۱۰ انسان فراهم می باشد. با این حال، این اعداد و ارقام بسیار بزرگ هستند و برای تجزیه و تحلیل آنها به بررسی دقیق تری نیاز است.

### **سلسله مراتب عالم هستی**

سرعت نور نه تنها گسترش زندگی را محدود می کند، بلکه ماهیت زندگی انسان ها را نیز با قرار دادن محدودیت های قوی بر ارتباطات، هوشیاری و کنترل، محدود می سازد و مشکلاتی را برای زندگی انسان ها ایجاد می کند.

### **سلسله مراتب افکار**

آیا تاکنون برایتان پیش آمده که بخواهید ضربه ای به مگس بزنید، اما از دست شما فرار کند؟ علت این موضوع این است که مگس به دلیل کوچک بودنش، واکنش سریعتری را از خود نشان می دهد

و اینکه گردش اطلاعات بین چشمان، مغز و عضله یک مگس در یک زمان بسیار کوتاه تری صورت می گیرد. این اصل "بزرگتر = کندتر" نه تنها در زیست شناسی کاربرد دارد و مرز سرعت را با در نظر گرفتن سرعت عبور سیگنال های الکتریکی از طریق سلول های عصبی مغز تعیین می کند، برای زندگی آینده کیهانی نیز کاربرد دارد و باعث می شود که هیچ اطلاعاتی را نتوان با سرعتی بیشتری از سرعت نور جابه جا کرد. بنابراین، بزرگی و بلندپروازی برای یک سیستم هوشمند پردازش اطلاعات مانند یک شمشیر دو لبه است که هم خوب و هم بد می باشد. از طرفی، تلاش های بلندپروازانه باعث می شود که الفاظ بیشتر و افکار پیچیده تری داشته باشید و از طرف دیگر، باعث می شود که سرعت آگاهی تان از افکار درست جهانی کند شود، زیرا مدت زیادی طول می کشد که این اخبار به اطلاع تمام بخش های سیستم پردازش اطلاعاتتان برسد. به نظر شما زندگی ای که در عالم هستی جریان دارد، کدامیک را انتخاب می کند: زندگی ساده و سریع یا زندگی پیچیده و کند؟ پیش بینی من این است که هر دو را انتخاب می کند. ساکنین پهنه زیست، طیف وسیعی از اندازه ها را از نهنگ آبی عظیم الجثه دویست تنی گرفته تا کوچکترین باکتری های  $10^{-16}$  کیلو گرمی پوشش می دهند. علاوه بر این، موجودات زنده ای که بزرگ، پیچیده و کند هستند اغلب تنبلی خود را با ایجاد زیرمجموعه های کوچکتری که ساده و سریع هستند، تعدیل می کنند. به عنوان مثال، عکس العمل پلک زدن شما به شدت سریع است، زیرا توسط مدار کوچک و ساده ای انجام می شود که بیشتر قسمت های مغز شما را درگیر نمی کند. اگر مگسی به طور ناگهانی بخواهد وارد چشمتان شود، ظرف کمتر از یک دهم ثانیه و قبل از اینکه اطلاعات مربوطه فرصت کند که وارد مغزتان شود و شما را از آنچه که اتفاق افتاده آگاه کند، پلک خواهید زد. پهنه زیست ما با سازماندهی به پردازش اطلاعات در قالب سلسله مراتبی از زیر مجموعه ها، می تواند هم پیچیده باشد و هم سریع عمل کند. ما انسان ها نیز از راهبرد سلسله مراتبی مشابهی برای بهینه سازی محاسبات موازی استفاده می کنیم. با توجه به اینکه انجام محاسبات داخل سازمانی، کار کند و پرهزینه ای است، پیش بینی من این است که در زندگی پیشرفته کیهانی در آینده نیز محاسبات تا جایی که ممکن است، در داخل زیرمجموعه ها انجام شود.

اگر یک محاسبه به قدری ساده باشد که با استفاده از رایانه معمولی بتوان آن را انجام داد، گسترش آن به رایانه ای که در مقیاس کهکشان ها می باشد، کار زیان بخشی خواهد بود، زیرا برای هر محاسبه لازم می شود که اطلاعات با سرعت نور به اشتراک گذاشته شود و این باعث می شود که هر محاسبه ای حدود ۱۰۰۰۰۰ سال طول بکشد. اینکه کدامیک از این پردازش اطلاعات به صورت آگاهانه و با استفاده از تجربه ذهنی انجام می شود، موضوع بحث برانگیز و جالبی است که در فصل هشتم بررسی خواهد شد. اگر برای هوشیاری لازم شود که اجزای مختلف سیستم با یکدیگر ارتباط برقرار کنند، افکار سیستم های بزرگتر باید کندتر حرکت کند. در حالی که شما می توانید افکار زیادی را در هر ثانیه داشته باشید، ذهنی که به اندازه یک کهکشان است، فقط می تواند یک فکر را در هر صد هزار سال داشته باشد و یک ذهن کیهانی با اندازه یک میلیارد سال نوری قبل از متلاشی شدن توسط انرژی تاریک، فقط فرصت فکر کردن به ده موضوع را خواهد داشت. البته همین چند فکر ارزشمند و تجارب ملازم آنها، ممکن است کاملاً پیچیده و عمیق باشند.

## سلسله مراتب قدرت

اگر فکر را بتوان در قالب سلسله مراتبی سازمان دهی کرد که شامل طیف وسیعی از مقیاس ها می باشد، قدرت را نیز می توان بر اساس یک سلسله مراتب مشخص، سازمان دهی کرد. در فصل چهارم گفتیم که بر اساس نظریه تعادل نش، موجودیت های هوشمند می توانند خود را در قالب سلسله مراتب های قدرت سازمان دهی کنند. بر اساس این نظریه، اگر موجودیت ها راهبرد خود را عوض کنند، وضعشان بدتر خواهد شد. هرچه قدر که فن آوری ارتباطات و حمل و نقل بهتر شود، این سلسله مراتب ها بیشتر می توانند رشد پیدا کنند. اکنون، این سوال مطرح می شود که اگر روزی برسد که ماشین های هوش مصنوعی فوق هوشمند در مقیاس های کیهانی گسترش پیدا کنند، سلسله مراتبشان به چه شکلی خواهد بود؟ بدون قید و بند و غیر متمرکز یا به شدت مستبدانه و متمرکز؟ آیا همکاری فقط بر اساس منافع مشترک خواهد بود یا بر اساس تهدید و اجبار؟ برای پاسخ به این

سوالات، راهبرد هویج و چماق را در نظر بگیرید و ببینید که چه انگیزه هایی برای همکاری در مقیاس های کیهانی وجود دارد و برای اجرای آنها از چه تهدیدهایی می توان استفاده کرد؟

## کنترل کردن با هویج

داد و ستد سنتی یکی از محرک های متداول برای همکاری در روی زمین بوده است. اگر منطقه ای دارای فن آوری بسیار پیشرفته ای نسبت به یک منطقه دیگر باشد، هر دو منطقه می توانند در ازای تبادل مواد خام، از کالاهای پیشرفته و دارای فن آوری بالای یکدیگر بهره مند شوند. با این حال، اگر ماشین هوش مصنوعی فوق بشری، فن آوری ای را توسعه دهد که به آسانی بتواند ذرات بنیادی یک ماده را به هر شکل دیگری از ماده تغییر آرایش دهد، دیگر انگیزه ای برای تجارت از راه دور وجود نخواهد داشت. وقتی که به آسانی و به سرعت می توان ذرات مس را به نقره تبدیل کرد، چرا باید زحمت به خرج داد و کالاهای نقره ای را بین منظومه های شمسی حمل کرد، آن هم در شرایطی که منظومه ها در فاصله بسیار دوری از یکدیگر قرار دارند؟ وقتی که دانش فنی و مواد خام در کهکشان ها وجود دارد، چرا باید زحمت به خرج داد و ماشین آلات دارای فن آوری پیشرفته را بین کهکشان ها حمل کرد؟ به نظر من، در کیهانی که پر از ماشین های هوش مصنوعی فوق بشری خواهد شد، هیچ کالایی به جز اطلاعات ارزش حمل کردن به فواصل دور را نخواهد داشت. تنها استثنای موجود شامل موادی است که برای پروژه های مهندسی کیهانی که هدفشان مقابله با گرایش مخرب انرژی تاریک به متلاشی کردن تمدن ها است، مورد استفاده قرار می گیرند. بر خلاف داد و ستد سنتی بین انسان ها، این مواد را می توانیم حتی به شکل یک پرتو انرژی و به هر میزانی که می خواهیم حمل کنیم، زیرا ماشین هوش مصنوعی فوق بشری به سرعت می تواند مواد دریافتی را به هر چیزی که می خواهد، تبدیل کند. ممکن است پرسید که اگر داد و ستد یا به اشتراک گذاشتن اطلاعات، محرک اصلی همکاری کیهانی باشد، از چه نوع اطلاعاتی باید استفاده کرد؟ هر اطلاعاتی در صورتی ارزشمند است که تولید آن به تلاش محاسباتی بسیار زیاد و وقت گیری نیاز داشته باشد. به عنوان مثال، ماشین هوش مصنوعی فوق بشری ممکن است بخواهد که پاسخ هایی را برای سوالات سخت

علمی در مورد ماهیت واقعیت فیزیکی یا سوالات سخت ریاضی در مورد قضایا یا الگوریتم های بهینه سازی و سوالات سخت مهندسی در مورد بهترین نحوه ساخت یک فن آوری خارق العاده پیدا کند. فرصت هایی که برای به اشتراک گذاشتن اطلاعات ایجاد خواهد شد، انگیزه لازم را برای گردش اطلاعات را نه فقط بین موجودیت های با قدرت یکسان، بلکه بین موجودیت های واقع در بالا و پایین سلسله مراتب مانند هاب (منظومه شمسی) و گره ها (کهکشان ها) فراهم خواهد ساخت. به این ترتیب، گره ها می توانند از متعلق بودن به یک چیز بزرگتر، لذت ببرند و پاسخ هایی را برای سوالات خود پیدا کنند و فن آوری هایی را توسعه دهند که به تنهایی قادر به توسعه آنها نبودند و در مقابل تهدیدهای خارجی از خود دفاع کنند. گره ها همچنین ممکن است برای وعده جاودانگی ذهن خود از طریق گرفتن نسخه بک آپ، ارزش قائل شوند. یک ماشین هوش مصنوعی، برخورداری از ذهن جاودانه را تأیید می کند و اذعان می کند که پس از تمام شدن ذخیره انرژی موجود در سخت افزار فیزیکی خود، در هاب یک ابر رایانه به زندگی ادامه خواهد داد. از سوی دیگر، هاب ممکن است از گره های خود بخواهد که برای انجام محاسبات بزرگ و بلندمدتی که نیاز فوری به نتایج شان وجود ندارد، به او کمک کنند. به همین خاطر، انتظار میلیون ساله برای دریافت پاسخ ها، ارزشش را دارد. همچنین، هاب ممکن است از گره های خود بخواهد که با ایجاد هماهنگی در میزان تمرکز جرمی کهکشان ها، در انجام پروژه های کیهانی بزرگ مانند مقابله با انرژی تاریک مخرب به او کمک کنند. اگر معلوم شود که می توان کرم های چاله قابل عبور از بین دو انتهای آنها را ساخت، اولویت اول یک هاب احتمالاً این خواهد بود که شبکه ای از آنها را برای خنثی کردن انرژی تاریک بسازد. اینکه اهداف نهایی یک ماشین هوش مصنوعی فوق بشری کیهانی چه خواهند بود، موضوع جالب و بحث برانگیزی است که در فصل هفتم به آن خواهیم پرداخت.

## کنترل کردن با چماق

امپراطوری های دنیوی معمولاً هم با استفاده از هویج و هم با استفاده از چماق، زیردستان خود را مجبور به اطاعت و همکاری با خود می کنند. در حالی که افراد تحت سلطه امپراتوری روم، قدر فن

آوری، زیرساخت و قدرت دفاعی ای را که به عنوان پاداش همکاری به آنها اعطا شده بود، می دانستند، همزمان از عواقب و پیامدهای اجتناب ناپذیر تمرد و سرکشی یا نپرداختن مالیات می ترسیدند. به دلیل زمان زیادی که طول می کشید تا سربازان از روم به سرزمین های دورافتاده اعزام شوند، بخشی از مسئولیت این تهدیدها به سربازان محلی و مقامات وظیفه شناس واگذار می شد که قدرت تنبیه های فوری به آنها داده شده بود. یک هاب هوش مصنوعی فوق هوشمند می توانست از راهبرد مشابهی برای اعزام شبکه ای از نگهبانان وظیفه شناس به سراسر امپراطوری کیهانی خود استفاده کند. از آنجایی که افراد تحت سلطه هوش مصنوعی فوق هوشمند را به سختی می توان کنترل کرد، ساده ترین راهبرد ممکن برای کنترل آنها می تواند این باشد که از نگهبانان هوش مصنوعی کاملاً وظیفه شناسی استفاده شود که تا حدودی لال هستند و فقط بر رعایت کلیه قوانین نظارت می کنند و در صورت رعایت نشدن قوانین، وسیله ای را برای حسابرسی فعال می کنند. به عنوان مثال، فرض کنید که هوش مصنوعی هاب تصمیم بگیرد که یک کوتوله سفید فضایی را در حوالی تمدنی که به اندازه منظومه شمسی است، قرار دهد تا آن را کنترل کند. کوتوله سفید، پوشش خارجی سوخته شده یک ستاره نسبتاً سنگین است که بیشتر جرم آن را کربن تشکیل می دهد و شبیه الماسی در آسمان است که سنگینتر از خورشید و کوچکتر از زمین است. یک فیزیکدان هندی به نام "چندراسخار" ثابت کرد که اگر به افزایش جرم کوتوله سفید تا رسیدن به "مرز چندراسخار" که ۴٫۱ جرم خورشید است، ادامه دهید، انفجار گرمایی - هسته ای فاجعه باری به نام "سوپرنوا" رخ خواهد داد. اگر هوش مصنوعی هاب با بی رحمی، کاری کند که این کوتوله سفید به مرز چندراسخار نزدیک شود، هوش مصنوعی نگهبان با وجود لال بودن، موثر واقع می شود و اگر تمدن سرکوب شده مالیات های مقرر خود را پرداخت نکرده باشد، جرم کافی را به کوتوله سفید پرتاب می کند تا سوپرنوا فعال شود و به این ترتیب، کل منطقه تخریب گردد. تمدن های به اندازه کهکشان را می توان به شکل مشابهی کنترل کرد. برای این منظور می توان در مرکز کهکشان تعداد زیادی شیء فشرده را در مدارهای سفت حول سیاه چاله قرار داد و تمدن ها را تهدید کرد که در صورت نافرمانی، این

اجرام را با روش هایی مانند برخورد به یکدیگر، به گاز تبدیل خواهید کرد و سپس این گاز را وارد سیاه چاله خواهید ساخت تا تبدیل به یک کوازار قدرتمند شود و این باعث خواهد شد که بیشتر مناطق کهکشان غیر قابل سکونت شوند. خلاصه اینکه، در زندگی آینده، انگیزه های قوی ای برای همکاری در سراسر کیهان وجود خواهد داشت. اکنون این سوال مطرح می شود که آیا این همکاری مبتنی بر منافع متقابل خواهد بود یا تهدیدهای بی رحمانه؟ به نظر می رسد که محدودیت های تحمیل شده توسط علم فیزیک، هر دو سناریو را مجاز می داند. بنابراین، نتیجه حاصله بستگی به اهداف و ارزش های غالب خواهد داشت. در فصل هفتم به بررسی توانایی ما برای تاثیر گذاری بر روی این اهداف و ارزش ها خواهیم پرداخت.

### وقتی که تمدن ها با هم برخورد می کنند.

تاکنون، فقط سناریوهایی را بررسی کردیم که در آنها، زندگی در اثر یک انفجار هوش واحد، در سراسر کیهان گسترش پیدا می کرد. اما، اگر زندگی به طور مستقلانه در بیش از یک مکان گسترش پیدا کند و دو تمدن گسترش یافته با هم برخورد کنند، چه اتفاقی خواهد افتاد؟ اگر یک منظومه شمسی تصادفی را در نظر بگیرید، این احتمال وجود دارد که زندگی بر روی یکی از سیاره های آن گسترش پیدا کند، فن آوری پیشرفته در آن توسعه پیدا کند و در آن، زندگی به فضا گسترش داده شود. به نظر می رسد که این احتمال از صفر بیشتر باشد، زیرا زندگی فن آورانه در منظومه شمسی ما توسعه پیدا کرده است و به نظر می رسد که قوانین فیزیک اجازه مستقر شدن در فضا را به ما می دهند. اگر فضا به قدر کافی بزرگ باشد (نظریه انبساط کیهانی می گوید که فضا گسترده و بی نهایت است) تمدن های گسترش یافته زیادی ایجاد خواهند شد. جی السون در مقاله خود، تحلیل جالبی را در مورد چنین پهنه زیست های کیهانی گسترش یافته ای ارائه کرده است و تابی ارد نیز تحلیل مشابهی را با همکارانش در موسسه "آینده بشریت" انجام داده است. اگر سرعت گسترش این پهنه زیست های کیهانی در تمام جهات برابر بود، شکل کره ماندی پیدا می کردند، اما به دلیل محدودیت های انرژی تاریک برای دسترسی تمدن ها به کهکشان ها، شکلشان شبیه قسمت بالای لیوان نوشیدنی



خواهد شد. اگر فاصله بین تمدن های همسایه مقیم فضا، بیش از حد شود، دیگر هرگز نمی توانند با یکدیگر ارتباط برقرار کنند یا حتی از زنده بودن یکدیگر مطلع شوند، بنابر این احساس خواهند کرد که گویی در کیهان تنها هستند. با این حال، اگر همسایه ها به یکدیگر نزدیک تر بودند، برخی از تمدن ها در نهایت نقاط مشترک یا متمایز کننده ای را با همسایگان خود پیدا می کردند و با هم به همکاری، رقابت یا جنگ می پرداختند. اروپایی ها به دلیل داشتن فن آوری برتر توانستند آفریقا و آمریکا را فتح کنند. اما خیلی قبل تر از اینکه تمدن های مجهز به هوش مصنوعی فوق بشری با یکدیگر مقابله کنند، ممکن است که سطح فن آوری هایشان که فقط توسط قوانین علم فیزیک محدود می شود، به وضع ثابتی برسد و دیگر تغییر نکند که این امر باعث می شود که یک هوش مصنوعی فوق بشری اگر مایل هم باشد، نتواند بر هوش مصنوعی دیگری غلبه کند. علاوه بر این، اگر آنها دارای اهداف نسبتاً همسویی باشند، ممکن است که دیگر دلیلی برای جنگ یا غلبه بر یکدیگر پیدا نکنند. به عنوان مثال، اگر هر دوی آنها سعی کنند تا نظریه های زیبای بیشتری را اثبات کنند و تا جایی که ممکن است الگوریتم های هوشمندتری را اختراع کنند، به سادگی می توانند یافته های خود را با دیگران به اشتراک گذارند که این به نفع هر دویشان خواهد بود. فرق اطلاعات با منابعی که انسان ها برای به دست آوردنشان می جنگند، این است که شما همزمان می توانید اطلاعات را حفظ یا آنها را به دیگران اهدا کنید. برخی از تمدن های در حال گسترش مانند پیروان فرقه های اصول گرایی ممکن است اهداف غیر قابل تغییری داشته باشند. با این حال، برخی از تمدن های پیشرفته ممکن است دارای اعضای باشند که ذهن بازتری دارند و پس از شنیدن استدلال های قانع کننده تمایل پیدا کنند که اهداف خود را تغییر دهند. اگر آن دو تمدن با هم مواجه شوند، ممکن است که به جای سلاح از عقاید خود برای نزاع با یکدیگر استفاده کنند و طرفی که استدلال های قانع کننده تری دارد، بر دیگری غلبه کند و اهداف خود را با سرعت نور در سراسر منطقه ای که قبلاً تحت کنترل تمدن دیگر بوده است، گسترش دهد. جذب و هضم فرهنگ شما توسط همسایگان، راهبرد سریعتری از استقرار فیزیکی شما در منطقه همسایه می باشد، زیرا حوزه نفوذ شما با سرعت انتقال ایده هایتان - که در

ارتباطات دور برد به سرعت نور می باشد- گسترش پیدا می کند و این در حالی است که استقرار فیزیکی با سرعت کمتری انجام می شود. البته، جذب و هضم فرهنگی نباید اجباری باشد، بلکه باید داوطلبانه انجام شود و به نفع خود همسایگان باشد و بر اساس برتری ایده ها از لحاظ اقناع کنندگی صورت گیرد. یک تمدن بلندپرواز با سه نوع منطقه مواجه خواهد شد: مناطق خالی از سکنه، مناطق استقرار تمدن های توسعه طلب یا مناطق دارای حباب های کشنده تمدن (حباب هایی که با سرعت نور گسترش پیدا می کنند و با تخریب تمام ذرات اولیه باعث می شوند که فضا غیر قابل سکونت گردد) اگر یک تمدن از تمدن رقیب همکاری گریز خود بترسد، انگیزه زیادی خواهد داشت تا قبل از رقیب خود، زمین ها را تصرف کند و در مناطق خالی از سکنه مستقر شود. با این حال، حتی اگر هیچ تمدن رقیبی هم وجود نداشته باشد، تمدن ها انگیزه توسعه طلبانه مشابهی برای این خواهند داشت که قبل از اینکه انرژی تاریک باعث دسترسی ناپذیری آنها به برخی مناطق شود، بتوانند منابع موجود در آن مناطق را تصرف کنند. همانطور که گفته شد، مواجهه با مناطق همسایه، در صورتی مفید است که این مناطق مایل به همکاری باشند و اعضای شان ذهن بازی داشته باشند. مواجهه با منطقه ای که تمدن توسعه طلب در آن مستقر می شود، بهتر از مواجهه با منطقه ای است که دارای حباب های کشنده تمدن است و با سرعت نور گسترش می یابد. تنها چیزی که باعث حفظ ما در برابر حباب های کشنده می شود، انرژی تاریک است که مانع از رسیدن حباب های کشنده دور به ما می شود و از این لحاظ، انرژی تاریک در واقع دشمن ما نیست، بلکه دوست ما می باشد.

## آیا ما در عالم هستی تنها هستیم؟

بسیاری از افراد، این موضوع را که زندگی پیشرفته ای در سراسر عالم هستی وجود خواهد داشت و انقراض ما انسان ها از دیدگاه کیهانی چندان مهم نخواهد بود را بدیهی فرض می کنند و معتقدند که وقتی که تمدن پیشتازان فضا می توانند ناگهانی وارد فضا شوند و از فن آوری پیشرفته خود برای رشد و توسعه استفاده کنند، چرا ما باید نگران نابود شدن خود باشیم؟ به نظر من، این دیدگاه خطرناک است، زیرا باعث می شود که احساس امنیت کاذب کنیم و نسبت به تمدن خود بی تفاوت و بی اعتنا

شویم. در واقع، این فرضیه که ما در عالم هستی تنها نیستیم، نه تنها خطرناک، بلکه احتمالاً نادرست است. شاید من هم اشتباه کنم، ولی این احتمالی است که وجود دارد و در حال حاضر نمی توان آن را نادیده گرفت. چنین احتمالی به ما اجبار اخلاقی می بخشد که جانب احتیاط را رعایت کنیم و تمایلی برای منقرض شدن تمدن خود نداشته باشیم. وقتی که در مورد کیهان شناسی سخنرانی می کنم، معمولاً از حضار می پرسم که در صورت پذیرفتن این موضوع که زندگی هوشمندانه در جای دیگری از عالم هستی ما وجود دارد، دست خود را بالا ببرند. تقریباً همه حضار از کودکان پیش دبستانی گرفته تا دانشجویان دست خود را بالا می برند. وقتی که علت این موضوع را از آنها می پرسم، پاسخی که دریافت می کنم این است که جهان هستی ما به قدری پهناور است که حداقل به لحاظ آماری می توان ثابت کرد که زندگی در جای دیگر، شدنی است. اما، با نگاه دقیقتر معلوم می شود که این استدلال دارای ضعف هایی می باشد. اگر فاصله بین یک تمدن با نزدیکترین همسایه اش بسیار بیشتر از ۲۰ میلیارد سال نوری باشد، می توان انتظار داشت که ما در جهان تنها باشیم و هیچ ارتباطی با موجودات فضایی نداشته باشیم. ما کاملاً نسبت به این موضوع بی اطلاع هستیم و فاصله ما تا همسایه مان در محدوده ۱۰۰۰.....۱۰۰ متر است که کل تعداد صفرهای آن ۲۱، ۲۲، ۲۳، .....۱۰۰، ۱۰۱، ۱۰۲ یا بیشتر است. هنوز مدرک متقاعد کننده ای برای وجود موجودات فضایی پیدا نشده است. در مورد نزدیکترین تمدن همسایه ما که در عالم هستی وجود دارد که شعاع آن حدود ۱۰<sup>۲۶</sup> متر است، تعداد صفرها از ۲۶ عدد فراتر می رود و احتمال اینکه تعداد صفرها بین ۲۲ و ۲۶ عدد باشد، نسبتاً اندک است. به همین دلیل، فکر می کنم که ما در جهان تنها هستیم. توجیه مفصلی از این استدلال را در کتاب خود با عنوان "جهان ریاضیاتی ما" بیان کرده ام. به همین دلیل، آن را در اینجا تکرار نمی کنم. اما، دلیل اصلی بی اطلاعی ما از فاصله زیاد بین ما و همسایگانمان، این است که ما در مورد احتمال وقوع زندگی مجهز به هوش مصنوعی در یک مکان مشخص، اطلاعی نداریم. همانطور که فضانورد آمریکایی به نام فرانک دراگ اشاره کرده است، این احتمال را می توان با ضرب احتمال وجود یک محیط قابل سکونت در یک سیاره در احتمال اینکه زندگی هوشمندانه ای در آنجا شکل

بگیرد، محاسبه کرد. وقتی که من یک دانشجوی کارشناسی ارشد بودم، اطلاعی از هیچ یک از این احتمالات نداشتم. اما پس از کشفیات چشم گیر در مورد سیاره های در حال چرخش به دور ستاره ها در دو دهه اخیر، اکنون به نظر می رسد که احتمال قابل سکونت بودن سیاره ها بالا باشد. با این حال، احتمال رشد و توسعه زندگی مبتنی بر فن آوری های هوش مصنوعی در سیاره های دیگر به شدت نامعلوم است. برخی از کارشناسان فکر می کنند که این یک امر حتمی و اجتناب ناپذیر است و در بیشتر سیاره های قابل سکونت اتفاق خواهد افتاد. در حالی که بقیه فکر می کنند که احتمال وقوع آن به شدت پایین است. برخی افراد با استدلال من مخالفت می کنند و می گویند که کهکشان ما پر از زندگی های مبتنی بر هوش مصنوعی است که دانشمندان امروزی متوجه آن نشده اند. آن طور که طرفداران بشقاب پرنده های فضایی استدلال می کنند، ممکن است آدم های فضایی تاکنون با زمین ملاقات کرده باشند یا شاید هم در زمین باشند و مخصوصا خود را از ما پنهان کرده باشند. بدون شک ما باید ذهن خود را در مورد این احتمالات باز نگه داریم. اما از آنجایی که هیچ مدرک قابل قبولی برای اثبات این احتمالات وجود ندارد، باید گزینه دیگر را که همان تنها بودن در عالم هستی است، جدی بگیریم. با این حال، ما نباید امکان وجود تنوع تمدن های پیشرفته را دست کم بگیریم. همانطور که دیدید، تصاحب منابع یکی از اهداف طبیعی هر تمدنی می باشد و برای تحقق آن لازم است که یک تمدن تصمیم بگیرد که به طور علنی در تمام جاهایی که برایش ممکن است، سکنی گزیند و بر کهکشان ما و فراتر از آن مستولی شود. در مواجهه با این واقعیت که میلیون ها سیاره شبیه زمین در کهکشان ما وجود دارند که میلیارد ها سال مسن تر از زمین هستند و با فرض اینکه ساکنان بلندپرواز برای اقامت در کهکشان ما زمان زیادی را دارند، نمی توان تصور کرد که همه آنها غیر قابل سکونت باشند. به هر حال، باید توجه داشت که فضانوردان بلندپرواز در پژوهش های خود در زمینه سیاره های شبیه زمین، به دنبال یافتن مدرکی هستند که نشان دهد اکسیژن لازم برای زندگی در آن مکان ها وجود دارد. آنچه که اهمیت دارد این است که هنگام جستجو برای زندگی پیشرفته، تا این اندازه انسان محور نباشیم و اگر یک تمدن فرازمینی را کشف کردیم، این احتمال را در نظر داشته باشیم که

این تمدن ممکن است از نوع هوش مصنوعی فوق بشری باشد. جی السون در بخش نتیجه گیری مقاله خود در زمینه سکنی گزیدن در فضا می گوید: "اینکه هوش مصنوعی پیشرفته از منابع جهان هستی برای پر کردن سیاره های شبیه زمین با گونه های جدیدی از انسان ها استفاده کند، نقطه پایان احتمالی برای پیشرفت فن آوری نمی باشد." بنابراین، آدم های فضایی را به عنوان آدم های ساده ای که دو دست و دو پا دارند، تصور نکنید، بلکه آنها را به عنوان ماشین های هوش مصنوعی فوق بشری فرا فضایی در نظر بگیرید. با وجود اینکه یکی از طرفداران قوی تحقیقات در مورد زندگی فرازمینی و روشن شدن برخی از سوالات علمی در این زمینه هستم، در خفا امیدوارم که این تحقیقات شکست بخورند و یافته ای حاصل نشود. ناسازگاری میان فراوانی سیاره های قابل سکونت در کهکشان ما و بازدید کنندگان از آنها حاکی از وجود یک فیلتر بزرگ فن آورانه است که در جایی از مسیر رشد و توسعه زندگی در فضا، سد راه ایجاد کرده است. اگر بتوانیم کشف کنیم که زندگی مستقلانه در جایی شکل گرفته است، این امر حاکی از آن خواهد بود که زندگی ابتدایی، امر نادری نیست و فقط سد راهی در مسیر مراحل رشد و تکامل انسان های فعلی قرار دارد. دلیل وجود این سد راه می تواند این باشد که سکنی گزیدن در فضا غیر ممکن است. به همین دلیل، آرزو می کنم که تمام جستجوها برای زندگی فرازمینی با شکست مواجه شوند.

## دورنمای آینده

در این کتاب در مورد تاریخچه زندگی در جهان هستی و آینده باشکوه آن در میلیاردها سال بعد صحبت شد. اگر فرض کنیم که توسعه هوش مصنوعی فعلی، زمینه را برای انفجار هوش و بهینه سازی اقامت در فضا فراهم کند، انتظارمان این است که پس از گذشت میلیاردها سال از عالم هستی مرده، انفجاری رخ می دهد که با سرعت نور و بدون کند شدن در سراسر کیهان گسترش می یابد و با اشتعال خود، همه چیز موجود در مسیر خود را زنده می سازد. چنین دیدگاه خوش بینانه ای از آینده زندگی در عالم هستی ما توسط بسیاری از متفکرین با شیوایی بیان شده است. از آنجایی که نویسندگان داستان های علمی تخیلی اغلب رویاپردازان رمانتیکی هستند، بیشتر داستان های علمی -

تخیلی در مورد سکنی گزیدن در فضا با استفاده از هوش مصنوعی فوق بشری، غیر واقعی و مشکوک به نظر می رسند. به عنوان مثال، به سختی می توان باور کرد که امکان سفر بین کهکشان ها به این اندازه ساده باشد یا اینکه ما انسان ها نه تنها در منظومه شمسی و کهکشان راه شیری، بلکه در سراسر کیهان بر سرنوشت خود حاکم شویم. اینکه ما تنها تمدن مجهز به فن آوری سطح پیشرفته در عالم هستی هستیم، باعث می شود که مسئولیت اخلاقی بزرگی داشته باشیم و بین دوراهی رشد و توسعه در سراسر عالم هستی یا منقرض شدن بمانیم. اگر ما فن آوری خود را بهبود ندهیم، مطمئناً به واسطه یک سیاره کوچک، ابرآتشفشان، گرمای سوزنده خورشید پیر یا بلایای دیگر، منقرض خواهیم شد. اگر آخر الزمان شود و در اثر انفجار عظیمی که گفته می شود، قدرت آن از تمام بمب های هیدروژنی بالاتر است، همه ستاره ها بسوزند، کهکشان ها بی نور شوند و سیاه چاله ها تبخیر شوند و همه ما بمیریم، نمایش عالم هستی که توسط فریمن دایسون پیش بینی شده است، بدون تماشایی خواهد ماند. فریمن می گوید: "نمایش آتش سوزی در جهان سرد و یخ زده فایده ای نخواهد داشت، زیرا دیگر کسی نیست که از آن لذت ببرد." درست است که همه ما روزی خواهیم مرد، اما اگر در زندگی خود به جای دوری جستن از فن آوری، از آن استقبال کنیم و ادعاهای خود را بالا ببریم، از پتانسیل لازم برای برنامه ریزی قوی و رشد و توسعه برخوردار خواهیم شد. فقط کافی است که قدرت ایمان، احتیاط، آینده نگری و برنامه ریزی دقیق را در خود افزایش دهیم. امکان بالقوه زندگی آینده در عالم هستی از یک سو مهم و با عظمت می باشد و از سوی دیگر به واسطه امکان بالقوه انقراض همیشگی زندگی هوشمندانه در جهان تعدیل می شود. آیا زندگی آینده در عالم هستی ما محقق خواهد شد یا فرصتی است که بر باد خواهد رفت؟ این موضوع تا حد زیادی بستگی به این خواهد داشت که انسان های زنده امروز در طول عمر خود چه کارهایی را انجام می دهند و من خوشبین هستم که اگر ما انسان ها انتخاب های درستی را انجام دهیم، می توانیم آینده زندگی را واقعاً عالی سازیم. ممکن است پرسید که ما چه می خواهیم و چگونه می خواهیم به اهداف خود برسیم؟ در ادامه به بررسی برخی از مشکل ترین چالش های مربوطه و نحوه مقابله با آنها می پردازیم.

## خلاصه فصل

- در مقایسه با مقیاس های زمانی میلیارد ساله کیهانی، انفجار هوش یک رخداد ناگهانی است که در آن، فن آوری به سرعت به سطح ثابتی می رسد که فقط توسط قوانین فیزیک محدود می شود.
- این سطح ثابت از فن آوری به شدت بالاتر از فن آوری امروز است و این اجازه را می دهد که میزان مشخصی از ماده ( با استفاده از اسفالرون ها یا سیاه چاله ها) حدود ۱۰ میلیارد برابر انرژی بیشتری را تولید کند و ۱۲ تا ۱۸ برابر اطلاعات بیشتری را ذخیره کند و محاسبات را ۳۱ تا ۴۱ برابر سریع تر انجام دهد یا به هر شکل مورد نظر دیگری از ماده تبدیل شود.
- زندگی مبتنی بر هوش مصنوعی فوق بشری نه تنها باعث می شود که استفاده فوق العاده کاراتری از منابع موجود به عمل آید، بلکه باعث می شود که ماشین های هوش مصنوعی از طریق سکنی گزیدن در کیهان با سرعت نزدیک نور و با تسخیر منابع ، پهنه زیست امروزی را به میزان ۳۲ برابر افزایش دهند.
- انرژی تاریک باعث می شود که زندگی مبتنی بر هوش مصنوعی فوق بشری در کیهان گسترش پیدا نکند. البته انرژی تاریک باعث حفظ این زندگی در برابر توسعه از راه دور حباب های کشنده یا تمدن های متخاصم خواهد شد. تهدید انرژی تاریک به متلاشی کردن تمدن های کیهانی، انگیزه لازم را برای انجام پروژه های بزرگ مهندسی کیهانی مانند ساختن کرم چاله (در صورت امکان پذیری) ایجاد می کند.
- اطلاعات، مهمترین کالایی است که می توان آن را در بین فواصل کیهانی به اشتراک گذاشت یا مبادله کرد.
- به استثنای کرم چاله ها، محدودیت سرعت نور برای تبادل اطلاعات، چالش های جدی ای را برای هماهنگی و کنترل یک تمدن کیهانی ایجاد خواهد کرد. یک هاب از راه دور مرکزی می تواند گره های هوش مصنوعی فوق بشری را به همکاری با خود تشویق کند یا در صورت

رعایت نکردن قوانین، آنها را تهدید کند و به عنوان مثال با استقرار یک هوش مصنوعی نگهبان که برای تخریب گره متخلف برنامه ریزی شده است، باعث انفجار سوپرنوا یا کوآزار شود.

- برخورد دو تمدن در حال گسترش می تواند باعث همانند سازی، همکاری یا جنگ شود که که احتمال وقوع جنگ در بین تمدن های امروزی بیشتر است.
- بر خلاف باور رایج، عده ای معتقدند که ما انسان ها قادر به زنده کردن عالم هستی در آینده خواهیم بود.
- اگر فن آوری را بهبود ندهیم، ممکن است که انسان ها به واسطه یک سیارک، ابر آتشفشان یا گرمای سوزاننده خورشید پیر یا بلاهای دیگر، منقرض شوند.
- اگر با توجه، دوراندیشی و برنامه ریزی کافی و اجتناب از گرفتار شدن در تله ها، به پیشرفت فن آوری ادامه دهیم، از ظرفیت لازم برای توسعه زندگی در زمین و ماورای آن، برخوردار خواهیم شد.

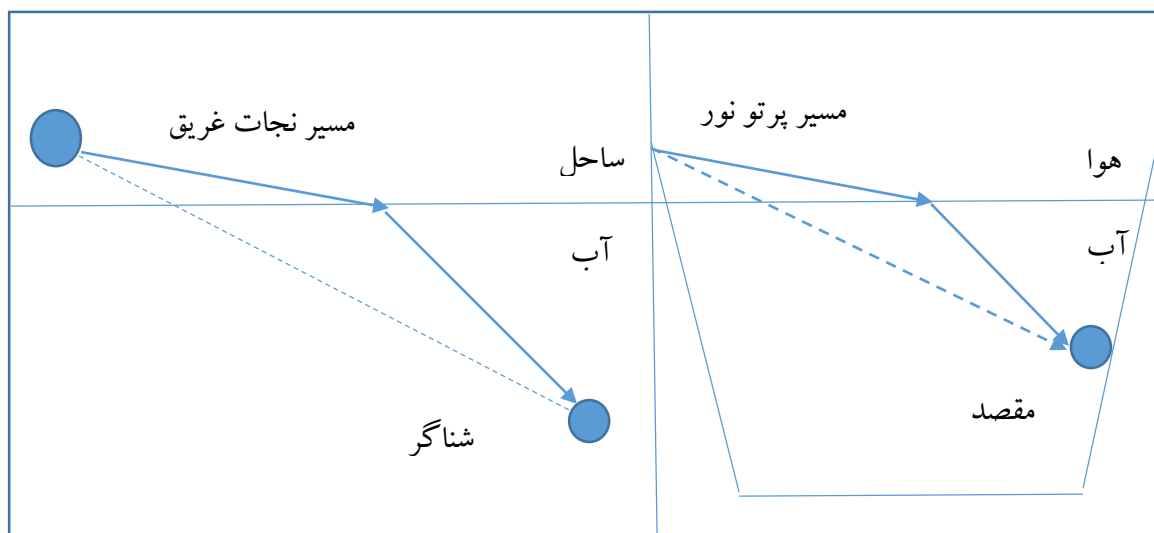
## فصل هفتم - اهداف

اگر بخواهم کل بحث های مربوط به هوش مصنوعی را در یک کلمه خلاصه کنم، آن کلمه "اهداف" خواهد بود. آیا ما باید به هوش مصنوعی، هدف بدهیم؟ در این صورت، این اهداف متعلق به چه کسانی خواهد بود؟ چگونه می توانیم به هوش مصنوعی هدف بدهیم؟ آیا می توان مطمئن شد که پس از زرننگ شدن هوش مصنوعی، اهداف اولیه تغییر نکنند؟ آیا می توان اهداف هوش مصنوعی زرننگتر از خود را تغییر داد؟ اهداف نهایی ما چه هستند؟ این سوالات نه تنها مشکل هستند، بلکه برای آینده زندگی اهمیت زیادی دارند. اگر ما ندانیم که چه می خواهیم، با احتمال کمتری به آن دست پیدا می کنیم و اگر کنترل را به ماشین هایی واگذار کنیم که در اهداف ما شریک نیستند، احتمال دارد به چیزی برسیم که آن را نمی خواهیم.



## علم فیزیک: منشا اهداف

برای پاسخگویی به این سوالات اجازه دهید که منشا نهایی اهداف را بررسی کنیم. وقتی که در دنیا به اطراف خود نگاه می کنیم، به فرایندهایی برخورد می کنیم که هدفمند هستند و به فرایندهای دیگری برخورد می کنیم که دارای هدف مشخصی نیستند. به عنوان مثال، فرایند شوت زدن توپ برای گل زدن به حریف را در نظر بگیرید. رفتار توپ به نظر هدفمند نمی رسد و با قوانین حرکت نیوتن به عنوان واکنش به شوتی که زده اید، توصیف می شود. اما رفتار بازیکن با هدف حداکثر ساختن امتیاز تیم توضیح داده می شود. جالب است که منشا نهایی رفتار هدفمند را می توان در خود قوانین فیزیک پیدا کرد. همانطور که در شکل مشاهده می کنید، اگر یک نجات غریق بخواهد شناگری را نجات دهد، انتظار داریم که در یک خط مستقیم حرکت نکند، بلکه کمی به طرف جلوتر ساحل بدود تا مسیر سریعتری را دنبال کند و وقتی که وارد آب می شود، اندکی تغییر جهت دهد. انتخاب چنین مسیری یک رفتار هدفمند است و فرد نجات غریق برای اینکه هرچه سریعتر به شناگر برسد، عمداً مسیر بهینه را انتخاب می کند. به شکل مشابهی، وقتی که یک پرتو نور ساده می خواهد وارد آب می شود، تغییر جهت می دهد تا زمان رسیدن به مقصد حداقل شود.



این اصل که در سال ۱۶۶۲ بیان شده، در علم فیزیک به اصل فرما<sup>۷۸</sup> معروف است و روشی را برای پیش بینی رفتار پرتوهای نور، ارائه می کند. فیزیکدان ها کشف کرده اند که تمام قوانین فیزیک کلاسیک را می توان به شیوه مشابهی به لحاظ ریاضی فرمولبندی مجدد کرد. طبیعت از میان تمام گزینه هایی که برای انجام یک کار وجود دارد، گزینه بهینه را ترجیح می دهد که معمولاً به کمینه سازی یا بیشینه سازی یک کمیت خلاصه می شود. دو شیوه برای توصیف هر قانون فیزیکی وجود دارد که به لحاظ ریاضی معادل یکدیگر هستند. شیوه اول، نگاه به گذشته ای است که منجر به آینده شده و شیوه دوم، همان شیوه ای است که طبیعت برای بهینه سازی از آن استفاده می کند. شیوه دوم که درک آن به لحاظ ریاضی مشکل است، در دوره های مقدماتی علم فیزیک آموزش داده نمی شود، اما به نظر من ظریفتر و پرمحتواتر می باشد. اگر شخصی تلاش کند تا چیزی مانند امتیازات، ثروت یا شادی خود را بهینه سازی کند، طبیعتاً تلاش هدفمندانه ای را دنبال کرده است. بنابراین، اگر خود طبیعت تلاش کند تا چیزی را بهینه سازی کند، تعجبی ندارد که رفتار هدفمندانه اش از همان اول بر اساس قوانین علم فیزیک طراحی و تنظیم شده است. یکی از کمیت های معروفی که طبیعت سعی می کند آن را بیشینه سازد، "آنتروپی"<sup>۷۹</sup> می باشد که به طور کلی، میزان شلوغی چیزها را اندازه گیری می کند. قانون دوم ترمودینامیک بیان می کند که میزان آنتروپی تا زمانی که به حداکثر میزان ممکن خود برسد، افزایش پیدا می کند و بدون در نظر گرفتن اثرات جاذبه به این حالت نهایی که در آن آنتروپی به حداکثر خود رسیده است، "گرما مرگی"<sup>۸۰</sup> می گویند که در آن، همه چیز به شکل کاملاً یکنواخت و خسته کننده و بدون وجود هر گونه پیچیدگی، زندگی و تغییر، گسترش می یابد. به عنوان مثال، وقتی که شما شیر سرد را روی قهوه داغ می ریزید، نوشیدنی شما به سمت حالت گرمای خاص خود پیشرفت می کند و دیگر به حالت اول بر نمی گردد و پس از مدت کوتاهی به یک مخلوط ولرم یکنواخت تبدیل می شود. اگر موجود زنده ای بمیرد، آنتروپی آن

<sup>۷۸</sup> Fermat's principle

<sup>۷۹</sup> entropy

<sup>۸۰</sup> Heat death

شروع به افزایش پیدا می کند و پس از مدت کوتاهی، آرایش ذرات آن به بی نظمی گرایش پیدا می کند. هدف آشکار طبیعت که افزایش آنتروپی است نشان می دهد که زمان، جهت مطلوبی را دنبال می کند و میل به بازگشت به عقب را ندارد. اگر لیوان آب خود را به زمین پرت کنید، انتظار دارید که خرد شود و آنتروپی آن افزایش یابد. اما اگر اجزای خرد شده لیوان پرواز کنند و سالم به دست شما برسند ( آنتروپی کاهش یابد) احتمالا دیگر از آن آب نمی نوشید، زیرا با خود حساب می کنید که بارها از آن آب نوشیده اید. وقتی که برای اولین بار یاد گرفتم که همه ما به شکل بی رحمانه ای به سمت گرما مرگی پیشروی می کنیم، به یاد گفته یکی از پیشگامان علم ترمودینامیک به نام لرد کلونین افتادم که می گفت: " استراحت و مرگ، نتیجه اجتناب ناپذیر عالم هستی می باشد. " چنین ایده ای که می گوید هدف بلند مدت طبیعت، بیشینه سازی مرگ و نابودی است، باعث تسلی خاطر ما نمی شود. با این حال، کشفیات اخیر نشان می دهند که اوضاع آن قدرها هم بد نیست. رفتار نیروی جاذبه متفاوت از سایر نیروها می باشد و تلاشش این است که عالم هستی ما را به جای یک مکان یکنواخت و خسته کننده به مکانی شلوغ و جالب تبدیل کند که پر از کهکشان، ستاره و سیاره می باشد. به لطف نیروی جاذبه، طیف وسیعی از دماها در کیهان وجود دارد که با ترکیبی از گرما و سرما، باعث رشد زندگی در فضا می شوند. ما هم در سیاره گرمی زندگی می کنیم که گرمای ۶۰۰۰ درجه سانتی گراد (۱۰۰۰۰ درجه فارنهایت) خورشید را جذب می کنند و هم در سیاره سردی زندگی می کنیم که با بیرون دادن گرمای اضافی تابیده شده خورشید به فضاها ی سرد، باعث خنک شدن سیاره و رسیدن دمای آن به ۳ درجه سانتی گراد زیر صفر (۵ درجه فارنهایت زیر صفر) می شود. همچنین، کارهای تحقیقاتی اخیر جرمی انگلند که دانشجوی من در دانشگاه ام آی تی می باشد، حاوی خبرهای خوبی است و نشان می دهد که علم ترمودینامیک هدف دلگرم کننده تری از گرما مرگی را به طبیعت می بخشد. این هدف با نام غیرمعارف " سازگاری طبیعت با اتلاف انرژی ناشی از افزایش آنتروپی " شناخته می شود و بدین معناست که گروه های تصادفی ای از ذرات تلاش می کنند تا خودشان را به گونه ای سازماندهی کنند که استخراج انرژی از محیط با حداکثر کارایی ممکن انجام

شود. به عنوان مثال، دسته ای از مولکول ها که در معرض نور خورشید قرار می گیرند، با گذشت زمان به گونه ای خودشان را آرایش می دهند که جذب نور خورشید را با کارایی هر چه بیشتری انجام دهند. به عبارت دیگر، به نظر می رسد که طبیعت، هدف از پیش تعیین شده ای را برای تولید سیستم های خود سازمان دهنده ای دارد که با وجود پیچیدگی فراوان، با قوانین علم فیزیک سازگار هستند. اروین شرودینگر که یکی از بنیانگذاران مکانیک کوانتومی است، در پاسخ به این سوال که چگونه می توان مصالحه ای را میان انگیزه کیهان برای زندگی کردن با انگیزه کیهان برای گرمایی برقرار کرد، اشاره می کند: "یکی از مشخصه های یک سیستم زنده این است که آنتروپی خود را با افزایش آنتروپی محیط اطراف خود، حفظ می کند یا کاهش می دهد. به عبارت دیگر، قانون دوم ترمودینامیک که بر اساس آن کل آنتروپی باید افزایش پیدا کند، یک راه فرار برای انگیزه زندگی کردن دارد و آن این است که کاهش آنتروپی در برخی مکان ها را در صورتی که باعث افزایش آنتروپی در جای دیگر شود، مجاز می داند. به همین دلیل، با شلوغ تر کردن محیط اطراف این امکان ایجاد می شود که زندگی به حالت یکنواخت و بدون پیچیدگی نرسد.

### علم زیست شناسی: تکامل اهداف

همانطور که دیدید، منشا رفتار هدف محور که به علم فیزیک باز می گردد، باعث شده که ذرات و مولکول ها به گونه ای آرایش پیدا کنند که با حداکثر کارایی ممکن به استخراج انرژی از محیط بپردازند. ذرات همچنین می توانند به گونه ای آرایش پیدا کنند که کپی مشابهی را از خودشان بسازند. نمونه های شناخته شده ای فراوانی از فرایند "تولید مثل" در طبیعت مشاهده می شود. به عنوان مثال، خوشه هایی از ریزگردها می توانند خوشه های مشابهی را ایجاد کنند. این فرایند که با استخراج انرژی و مواد خام از محیط همراه می شود، زندگی نامیده می شود. ما نمی دانیم که زندگی در زمین از چه زمانی آغاز شده است، اما می دانیم که شکل های اولیه زندگی از ۴ میلیارد سال پیش وجود داشته است. اگر یکی از شکل های زندگی، کپی مشابهی را از خودش بسازد و آن کپی هم همین کار را انجام دهد، تعداد کپی ها در فواصل زمانی منظم دو برابر می شود و این فرایند تا زمان

مواجهه با محدودیت منابع یا مشکلات دیگر ادامه پیدا خواهد کرد. تکرار این دو برابر شدن ها عدد بزرگی را ایجاد خواهد کرد و در نهایت، شکل های مختلف زندگی ایجاد می شوند که برای استفاده از منابع محدود با هم رقابت خواهند کرد. با گذشت زمان، تغییرات چشمگیری در خصوص رفتار هدفمند در زمین ایجاد شده است و در حالی که در گذشته، به نظر می رسید که ذرات تلاش می کنند تا به شکل های مختلف، شلوغی متوسط را افزایش دهند، الگوهای جدید، هدف متفاوتی را دنبال می کنند و آن این است به جای ائتلاف انرژی، به تولید مثل پردازند. با فرض ثابت بودن قوانین فیزیک، چنین تغییر هدفی چگونه ممکن است ایجاد شده باشد؟ پاسخ این است که هدف اساسی (ائتلاف انرژی) تغییر نکرده است، اما به هدف فرعی متفاوتی (تولید مثل) منجر شده است که به تحقق هدف اصلی کمک می کند. به عنوان مثال، "خوردن" را در نظر بگیرید. همه ما هدفمان این است که گرسنگی خود را با خوردن غذا بر طرف کنیم و این در حالی است که تنها هدف اصلی تکامل ما تولید مثل است نه جویدن غذا. با این حال، خوردن غذا به تولید مثل کمک می کند. تولید مثل هم به ائتلاف انرژی کمک می کند، زیرا زمینی که پر از زندگی شده باشد، انرژی را با سرعت بیشتری مصرف خواهد کرد. به عنوان مثال، اگر شکر را روی زمین آشپزخانه بریزید، انرژی شیمیایی مفید خود را سال ها حفظ خواهد کرد، اما اگر مورچه ها پیدایشان شود، تمام انرژی را خیلی زود تلف خواهند کرد. همچنین، اگر انسان ها در زندگی خود از انرژی نفت خام استفاده نمی کردند و آن را نمی سوزاندند، ذخیره نفت خام می توانست انرژی شیمیایی مفید خود را به مدت طولانی تری حفظ کند. با وجود اینکه هدف اصلی تکامل ما تولید مثل است، به نظر می رسد که ساکنین امروزی زمین، اهداف فرعی مخصوص خود را دنبال می کنند و به جای اختصاص دادن زمانی برای فرزندآوری، وقت خود را صرف فعالیت هایی مانند خوابیدن، غذاخوردن، ساختن خانه، اعمال قدرت و مبارزه یا کمک به دیگران می کنند. تحقیقات صورت گرفته در زمینه های روان شناسی تکاملی، اقتصاد و هوش مصنوعی، دلیل این موضوع را با ظرافت توضیح داده اند. برخی از اقتصاددانان انسان ها را به عنوان عوامل منطقی و تصمیم گیرندگان بی نقصی مدلسازی کرده اند که همواره در پی گیری هدف

خود، اقدام بهینه را انتخاب می کنند. این شیوه مدلسازی به وضوح غیر واقع بینانه است. همانطور که یکی از پیشگامان هوش مصنوعی به نام هربرت سیمون اشاره کرده است، انسان ها در عمل، دارای عقلانیت محدودی می باشند و عقلانی بودن تصمیمات آنها به واسطه کمبود اطلاعات موجود و زمان کافی برای فکر کردن محدود می شود. این بدین معناست که وقتی یک موجود زنده می خواهد که به هدف بهینه ای برسد، فقط می تواند الگوریتم های تقریبی ای را اجرا کند که در شرایط محدودی که خودش کشف کرده است، مفید واقع می شوند. مثلاً، به جای اینکه در هر موقعیتی از خود پرسد که چه اقدامی بهینه است، روش های ابتکاری درهم و برهمی را اجرا می کند که معمولاً فقط مبتنی بر تجربه شخصی خودش می باشد. این وضعیت را می توان در بین حیواناتی که وقتی تشنه هستند، آب می خورند و وقتی گرسنه هستند، غذا می خورند و از چیزهای مضر دوری می کنند، مشاهده کرد. روش های مبتنی بر تجارب شخصی گاهی باعث شکست در برخی از موقعیت هایی می شوند که از قبل، نقشه ای برای مدیریت آنها طراحی نشده است. به عنوان مثال، موش ها ممکن است سم موشی را بخورند که بسیار لذیذ است یا حشرات ممکن است به سمت شعله های شمع پرواز کنند و بسوزند. به عنوان مثالی دیگر می توان گفت که اگر هدف فرعی "از گرسنگی نمردن" باعث تمایل ما به مصرف غذاهای پر کالری شود، ممکن است که همین هدف فرعی باعث شیوع چاقی و ایجاد مشکلاتی برای جفت یابی شود.

### **علم روانشناسی: دنبال کردن اهداف یا مخالفت در برابر آنها**

خلاصه اینکه، یک موجود زنده دارای عقلانیت محدودی است که یک هدف واحد را دنبال نمی کند، بلکه در عوض از تجارب شخصی خود برای اینکه به دنبال چه چیزی برود یا از چه چیزی اجتناب کند، استفاده می کند. تجارب شخصی ذهنی ما انسان ها به عنوان احساساتمان در نظر گرفته می شوند که اغلب بدون اینکه از آنها آگاه باشیم، تصمیم گیری ما را به سمت هدف نهایی "تولید مثل"، هدایت می کنند. احساس های گرسنگی و تشنگی به حفظ ما در برابر مرگ ناشی از گرسنگی یا کم آبی بدن کمک می کنند. احساس درد باعث می شود که به بدن خود آسیب وارد نکنیم.

احساس عشق و دلسوزی باعث می شود که به دیگران کمک کنیم. وقتی که توسط این احساسات هدایت می شویم، مغزمان بدون اینکه مجبور شود تا تمام گزینه ها را در معرض یک تجزیه و تحلیل خسته کننده در خصوص پیامدهای نهایی هر یک از آنها قرار دهد، به سرعت تصمیم می گیرد که چه کاری انجام دهد. برای دریافت چشم انداز دقیقتری از احساسات و ریشه های روان شناختی آنها پیشنهاد می کنم که به آثار ویلیام جیمز و آنتونیو داماسیو مراجعه کنید. توجه به این نکته مهم است که وقتی احساساتمان علیه فرزندآوری کار می کنند، این یک امر لزوما تصادفی یا نشان دهنده فریب خوردنمان نمی باشد، بلکه نشان می دهد که مغزمان به طور کاملاً عمدی با هدف تولید مثل مخالف است و تصمیم گرفته که از داروهای ضدبارداری استفاده کند یا مرتکب خودکشی شود یا تمام عمر خود را مجرد بماند و به یک کشیش، صومعه نشین یا راهبه تبدیل شود. چرا گاهی تصمیم می گیریم که با ژن های خود و هدف از وجودشان که تولید مثل است، مخالفت کنیم؟ علت مخالفتمان با آنها این است که ما انسان ها طوری طراحی شده ایم که عقلانیت محدودی داشته باشیم و فقط تابع احساسات خود باشیم. اگرچه مغز ما انسان ها برای کپی ساختن از ژن هایمان تکامل یافته است، اما هیچ احساسی نسبت به آنها نداریم و اجداد ما در سال های بسیار دور، حتی نمی دانستند که دارای چیزی به نام ژن هستند. علاوه بر این، مغزهای ما به مراتب زرننگ تر از ژن های ما هستند و حالا که ما انسان ها هدف از وجود ژن های خود را که تولید مثل است، درک می کنیم، آنها را نسبتاً پیش پا افتاده می دانیم و توجهی به آنها نمی کنیم و تمایلی به داشتن بچه های زیاد نداریم و با دستکاری در برنامه نویسی ژنتیکی خود که به منظور کنترل زاد و ولد انجام می شود، می خواهیم که از پاداش های احساسی ناشی از نداشتن فرزند بیشتر برخوردار شویم. این مثل این می ماند که ژن های ما شیرینی را دوست دارند، اما برای اینکه چاق نشویم، برنامه نویسی ژنتیکی خود را دستکاری کنیم و به جای لذت بردن از یک نوشیدنی شیرین، آن را با شیرین کننده های مصنوعی بدون کالری ترکیب کنیم. با این حال، باید به خاطر داشته باشیم که صاحب قدرت در حال حاضر احساساتمان هستند. این بدین معناست که رفتار ما انسان ها برای زنده ماندن گونه هایمان، بهینه سازی نشده است. در واقع، از

آنجایی که احساسات ما فقط تابع تجارب شخصی مان هستند که مناسب کلیه موقعیت ها نمی باشند، رفتار ما انسان ها به معنای واقعی کلمه دارای هیچ هدف واحد به خوبی تعریف شده ای نیست.

## علم مهندسی: برون سپاری اهداف

آیا ماشین ها می توانند هدف داشته باشند؟ این سوال ساده باعث ایجاد اختلاف نظرهای بزرگی شده است، زیرا افراد مختلف معانی مختلفی را از آن برداشت می کنند که به سوالات دشواری مانند اینکه آیا ماشین ها می توانند هوشیار باشند یا اینکه احساس داشته باشند، منجر می شود. اما، اگر منظورمان این باشد که "آیا ماشین ها می توانند رفتار هدفمندانه ای داشته باشند" پاسخ مثبت است، زیرا این امر امکان پذیر است و ما می توانیم آنها را به گونه ای طراحی کنیم که هدف داشته باشند. به عنوان مثال، ما تله موش را برای گرفتن موش ها، ماشین ظرفشویی را برای تمیز کردن ظرف ها و ساعت را برای نشان دادن وقت طراحی کرده ایم. وقتی که شما توسط یک موشک حساس به حرارت تعقیب شوید، برایتان مهم نیست که موشک هوشیار است یا احساس دارد و کاملاً مطمئن هستید که رفتارش هدفمندانه است و حتی اگر "هدف" نداشته باشد، "منظور"<sup>۸۱</sup> دارد. در فصل بعدی در مورد موضوع هوشیاری توضیحات بیشتری داده خواهد شد. بیشتر چیزهایی که ما ساخته ایم، از لحاظ طراحی هدفمند بوده اند، نه از لحاظ رفتاری. به عنوان مثال، یک بزرگراه هیچ گونه رفتاری از خود نشان نمی دهد، ولی برای دستیابی به هدفی طراحی شده است. "غایت شناسی"<sup>۸۲</sup> به این معناست که هر چیزی که در عالم هستی وجود دارد، علاوه بر داشتن دلیلی برای خلقت، دارای منظور نیز می باشد. بنابر این، می توان گفت که تمام موجودات غیر زنده نیز دارای هدف می باشند و اینکه تمام موجودیت های ساخته شده بر روی زمین از جمله ساختمان ها، جاده ها و خودروها و سایر موجودیت های مهندسی شده و همچنین موجودات های تکامل یافته ای مانند گیاهان و جانوران به دنبال هدفی هستند. در جدول زیر به سنگینی این موجودیت ها اشاره شده است.

---

<sup>۸۱</sup> purpose

<sup>۸۲</sup> teleology



وزن به میلیون تن	موجودیت های هدفمند
۴۰۰	۱۰۳۰×۵ باکتری
۴۰۰	گیاهان
۱۰	۱۰۱۵ ماهی مزوفلایک
۰,۵	۱,۳×۱۰۹ گاو
۰,۳	۱۰۱۴ مورچه
۰,۰۰۰۵	۱,۷×۱۰۶ نهنگ
۱۰۰	بتون
۲۰	فولاد
۱۵	آسفالت
۲	۱,۲×۱۰۹ خودرو

در حالی که هدف نهایی تمام موجودیت های تکامل یافته این است که به تولید مثل بپردازند، موجودیت های طراحی شده تقریباً می توانند هر هدفی (حتی اهداف مغایر با هم) داشته باشند. به عنوان مثال، اجاق گاز، غذا را گرم می کند، در حالی که یخچال، غذا را سرد نگه می دارد. مولد (ژنراتور) کارش این است که انرژی جنبشی را به انرژی الکتریکی تبدیل کند، در حالی که موتور کارش این است که انرژی الکتریکی را به انرژی جنبشی تبدیل کند. برنامه های استاندارد بازی شطرنج، هدفشان بردن در بازی است، اما برنامه هایی وجود دارند که هدفشان باختن در مسابقات شطرنج است. موجودیت های ساخته شده توسط انسان ها با گذشت زمان، به دنبال رسیدن به اهداف متنوع تر، پیچیده تر و هوشمندانه تر هستند. ابزارهای طراحی و ساخته شده توسط انسان ها در گذشته هدف کاملاً ساده ای داشتند. به عنوان مثال، هدف از ساختن خانه ها این بود که جای گرم و

امنی را برایمان فراهم کنند، اما، انسان ها به تدریج یاد گرفتند که ماشین هایی را بسازند که دارای اهداف پیچیده تری هستند که از جمله آنها می توان به جاروبرقی های رباتیک و خودروهای بدون راننده اشاره کرد. پیشرفت های اخیر هوش مصنوعی منجر به ایجاد سیستم های پیشرفته ای از قبیل دیپ بلو، واتسون و آلفاگو شده است که هدفشان به ترتیب، بردن در بازی شطرنج، بردن در مسابقات اطلاعات عمومی تلویزیونی و بردن در بازی گو می باشند. این سیستم ها به قدری مفصل و پیچیده اند که انسان ها برای درک پیچیدگی موجود در ساخت آنها باید خیلی ماهر باشند. وقتی که ما انسان ها ماشینی را می سازیم، به سختی می توانیم اهداف آن را با اهداف خودمان به طور کامل تطبیق دهیم. به عنوان مثال، یک تله موش، ممکن است که انگشت پای شما را با یک موش گرسنه اشتباه بگیرد. علاوه بر این، تمام ماشین ها دارای عقلانیت محدود هستند و حتی پیچیده ترین ماشین های امروزی در مقایسه با انسان ها دارای درک ضعیفتری از جهان هستند و به همین دلیل، قوانینی که آنها برای تشخیص اینکه چه کاری باید انجام دهند، استفاده می کنند، بیش از حد ساده انگارانه می باشند. تله موش منتظر فرصتی برای ستیزه جویی است و نمی داند که موش چیست. بسیاری از حوادث صنعتی کشنده به این دلیل اتفاق می افتند که ماشین ها نمی فهمند که انسان چیست. اگر بتوان ماشین ها را هوشمندتر ساخت، بسیاری از مشکلات مربوط به عدم انطباق اهداف، حل خواهند شد. اما، همانطور که در مورد پرومتئوس در فصل چهارم گفته شد، افزایش هرچه بیشتر هوشمندی ماشین ها ممکن است که چالش های جدیدی را برای تضمین اینکه ماشین ها در اهداف ما شریکند، ایجاد کند.

### **هوش مصنوعی دوستانه: همراستا کردن اهداف**

هرچه قدر که ماشین ها قوی تر و هوشمندتر می شوند، همراستا کردن اهدافشان با اهداف ما اهمیت بیشتری پیدا می کند. مشکلی که با ماشین های هوش مصنوعی داریم، این است که ممکن است با اهداف بشریت همراستا نباشند و برای بشریت مشکل ایجاد کنند. اگر ماشین هوش مصنوعی فوق بشری به طور ناگهانی ایجاد شود، باعث ایجاد دردسرهایی خواهد شد. از آنجایی که هوش به صورت توانایی دستیابی به اهداف تعریف می شود، ماشین هوش مصنوعی فوق بشری خیلی بهتر از

انسان ها می تواند به اهداف خود دست پیدا کند و به همین دلیل، بر انسان ها چیره خواهد شد. در فصل چهارم، مثال هایی را در مورد احتمال غلبه پرومئوس بر انسان ها بیان کردیم. اگر می خواهید که شاهد غلبه اهداف ماشین بر اهداف خود باشید، همین حالا یکی از پیشرفته ترین موتورهای شطرنج را دان لود کنید و سعی کنید که آن را شکست دهید. هرگز قادر به شکست دادن آن نخواهید بود. به عبارت دیگر، خطر واقعی ماشین هوش عمومی مصنوعی از بابت بدخواهی آن نیست، بلکه از بابت شایستگی بالای آن می باشد. همانطور که در فصل اول گفته شد، ما انسان ها هنگام خراب کردن لانه مورچه ها برای ساختن سدهای هیدروالکتریک تامل نمی کنیم. اگر ما انسان ها هم برای ماشین های هوش مصنوعی مانند آن مورچه ها بودیم، دچار دردسر می شدیم. بیشتر محققان استدلال می کنند که اگر روزی هوش مصنوعی فوق بشری ایجاد شود، باید از دوستانه بودن رفتارش با انسان ها و همراهی بودن اهدافش با اهداف ما انسان ها مطمئن شویم. تشخیص اینکه چگونه می توان اهداف ماشین را با اهداف انسان ها در یک راستا قرار داد، کار سختی است. در واقع، این مشکلی است که هنوز حل نشده است و شامل سه مشکل فرعی زیر می شود که هر یک نیاز به تحقیقات فعالانه دانشمندان علم رایانه و متفکران دیگر دارد. این مشکلات شامل، واداشتن هوش مصنوعی به درک اهدافمان، سازگاری با اهدافمان و حفظ اهدافمان می باشند. هوش مصنوعی برای درک اهداف ما فقط نباید بفهمد که چه می خواهیم، بلکه باید به دلیل درخواست ما هم توجه داشته باشد تا از سوء تفاهمات جلوگیری شود و فقط دستورمان را اجرا نکند. به عنوان مثال، در داستانی از یونان قدیم، شاه میداس تقاضا می کند که هرچه لمس می کند به طلا تبدیل شود. اما دیگر نمی تواند چیزی بخورد و دخترش به طور ناخواسته به طلا تبدیل می شود. در داستان دیگری که یک غول چراغ جادو سه خواسته یک فرد را برآورده می کند، خواسته های اول و دوم متفاوت هستند، اما خواسته سوم معمولاً این است که: "لطفاً دو خواسته اول را برگردانید، زیرا آنها چیزی نیستند که واقعا می خواهم." این مثال ها نشان می دهند که برای فهمیدن اینکه افراد واقعا چه می خواهند، فقط نباید به آنچه که می گویند، توجه کنید. شما به یک مدل جامعتر از جهان نیاز دارید که شامل اولویت

هایی باشد که ما آنها را بیان نمی کنیم، زیرا به نظرمان واضح هستند. به عنوان مثال، ما دوست نداریم که طلا بخوریم. وقتی که چنین مدل جامعی از جهان را داشته باشیم، فقط با مشاهده رفتار هدفمند افراد و حتی بدون اینکه آنها چیزی بگویند، می توانیم تشخیص دهیم که چه می خواهند. در واقع، بچه های والدین ریاکار به جای توجه به آنچه که از والدینشان می شنوند، از اعمال و رفتار والدین خود یاد می گیرند که چه کاری باید انجام دهند. محققان هوش مصنوعی در حال حاضر به سختی تلاش می کنند تا ماشین ها را قادر به پی بردن به اهداف انسان ها از روی رفتارشان کنند و این امر مدت ها قبل از اینکه ماشین هوش مصنوعی فوق بشری پدیدار شود، نیز می تواند مفید واقع شود. به عنوان مثال، یک فرد بازنشسته ممکن است از ربات مراقبت کننده از سالمندان اظهار قدردانی کند، زیرا این ربات فقط با مشاهده رفتار فرد سالمند و بدون نیاز به اینکه فرد سالمند همه چیز را با لغات یا برنامه های رایانه ای توضیح دهد، توانسته است تشخیص دهد که او چه می خواهد. یکی از چالش هایی که در این رابطه وجود دارد، تهیه مجموعه ای از سیستم های قراردادی برای تبدیل اهداف و اصول اخلاقی به زبان ماشین می باشد و چالش دیگر، ساختن ماشین هایی است که مناسبترین سیستم موجود برای رفتاری که مشاهده می کنند را تشخیص دهند. یکی از رویکردهای متداول برای مقابله با چالش دوم، "یادگیری تقویت معکوس"<sup>۸۳</sup> می باشد که در مرکز تحقیقاتی برکلی که توسط استوارت راسل راه اندازی شده است، مورد توجه زیادی واقع شده است. به عنوان مثال، فرض کنید که هوش مصنوعی شاهد آتش نشانی است که به سرعت وارد ساختمان آتش گرفته می شود و یک پسر بچه را نجات می دهد. هوش مصنوعی ممکن است چنین نتیجه گیری کند که هدف آتش نشان نجات بچه بوده و اینکه اصول اخلاقی باعث می شود که فرد آتش نشان، نجات جان بچه را به راحت نشستن در ماشین آتش نشانی ترجیح دهد و ایمنی خود را به خطر اندازد. اما به عنوان یک گزینه دیگر، هوش مصنوعی ممکن است چنین نتیجه گیری کند که آتش نشان احساس سرمای شدید می کند و می خواهد که گرم شود یا این کار را برای ورزش کردن انجام می دهد. اگر این تنها چیزی

---

<sup>۸۳</sup> Inverse reinforcement learning

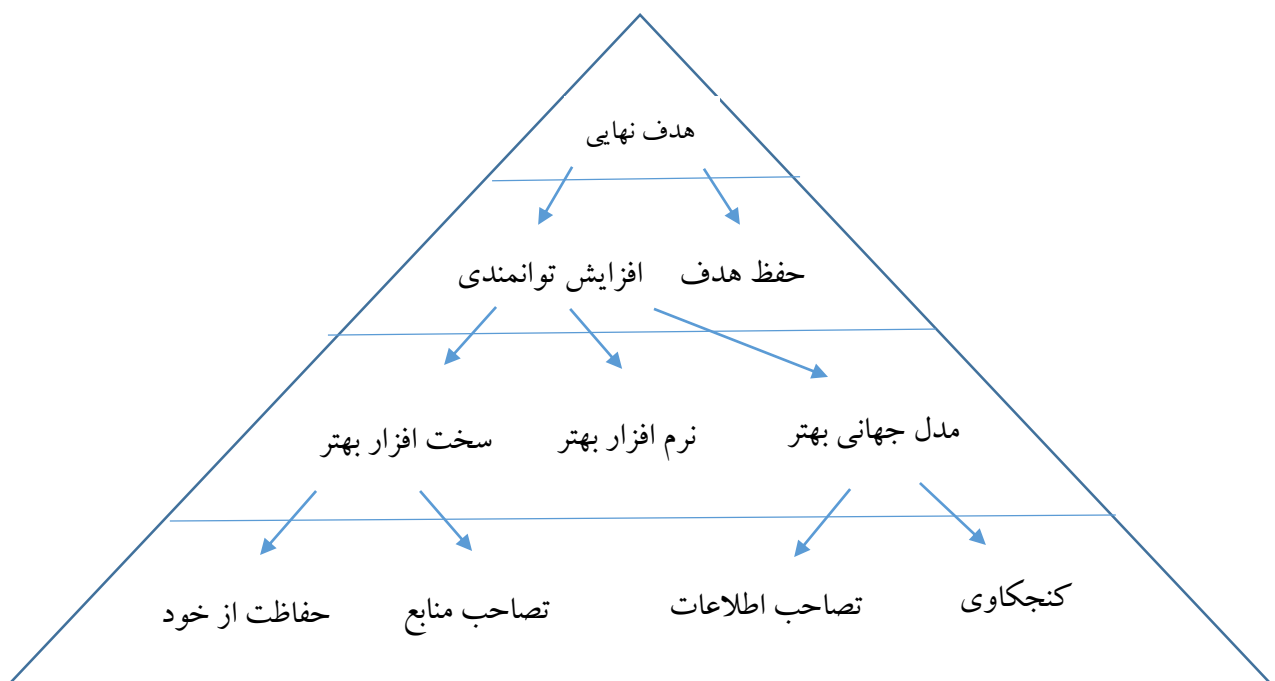
باشد که هوش مصنوعی در مورد آتش نشان ها، آتش سوزی و بچه ها بداند، معلوم نمی شود که کدام گزینه درست است. با این حال، ایده اصلی زیربنایی یادگیری تقویت معکوس این است که ما در همه زمان ها تصمیم می گیریم و هر تصمیمی که می گیریم، چیزی را در مورد اهدافمان آشکار می کند. بنابراین، امیدواریم که هوش مصنوعی با مشاهده تعداد زیادی از افراد در موقعیت های متعدد، در نهایت بتواند مدل درستی از تمام اولویت های ما را ایجاد کند. ایده اصلی رویکرد یادگیری تقویت معکوس این است که هوش مصنوعی تلاش می کند تا نه تنها رضایت خودش از هدف را حداکثر سازد، بلکه رضایت صاحب انسانی خود را نیز به حداکثر برساند. بنابراین، وقتی که هوش مصنوعی تصویر روشنی از خواسته های صاحبش را ندارد، انگیزه ای برای محتاط بودن خواهد داشت و تمام تلاشش را می کند تا خواسته های صاحبش را پیدا کند. همچنین وقتی که هوش مصنوعی خواسته های صاحبش را درک نمی کند، نباید مخالفتی با خاموش شدن خود توسط صاحبش را داشته باشد. حتی اگر هوش مصنوعی به گونه ای ساخته شود که اهداف صاحبش را درک کند، این امر لزوماً به معنای سازگاری آن با اهداف صاحبش نخواهد بود. منفورترین سیاستمداران از دیدگاه خود را در نظر بگیرید. شما خواسته آنها را می دانید، اما خواسته آنها همان چیزی نیست که شما می خواهید و آنها با وجود تلاش زیاد نمی توانند شما را وادار به سازگاری با اهداف خود کنند. راهبردهای فراوانی برای نفوذ در فرزندان برای پذیرش اهداف ما وجود دارند که برخی از آنها موفقیت آمیزتر از بقیه هستند. وقتی که به جای انسان ها می خواهیم رایانه ها را متقاعد سازیم، با چالشی به نام "مشکل بارگذاری ارزش ها" مواجه می شویم که حتی از آموزش مسائل اخلاقی به بچه ها مشکل تر است. یک سیستم هوش مصنوعی را در نظر بگیرید که هوش آن از طریق فرایند "خود بهبودی بازگشتی" به تدریج بهبود پیدا می کند تا به سطح هوش مصنوعی فرا بشری برسد. این سیستم در ابتدا بسیار ضعیف تر از شما است و می توانید آن را خاموش کنید یا بخش هایی از داده ها و نرم افزار آن را عوض کنید تا اهدافش با اهدافتان سازگار شود. اما، ماشین هوش مصنوعی احمق تر از آن است که اهداف شما را مانند انسان ها به طور کامل درک کند. با این حال،

وقتی که سطح هوش آن به سطح هوش مصنوعی فرا بشری می رسد، زرننگ تر از شما می شود و به قدری قوی می شود که نمی توانید آن را خاموش کنید و اهدافش را با اهداف خود جایگزین کنید. میزان زمان لازم برای بارگذاری اهداف خود را به داخل یک ماشین هوش مصنوعی، ممکن است بسیار کوتاه باشد. بارگذاری ارزش ها به داخل ماشین ها سخت تر از بارگذاری آنها در مغز انسان می باشد، زیرا رشد هوش ماشین ها بسیار سریعتر از انسان ها است. برخی از محققان رویکرد دیگری را برای وادار کردن ماشین ها به سازگاری با اهداف ما دنبال می کنند. آنها امیدوارند که سیستم "شناسایی اهداف" را برای هوش مصنوعی ایجاد کنند که اگر گاهی آن را خاموش کنید و اهدافش را تغییر دهید، اهمیتی به آن نخواهد داد. اگر ثابت شود که چنین چیزی ممکن است، می توانید با خیال راحت به ماشین هوش مصنوعی اجازه دهید که فوق هوشمند شود. سپس آن را خاموش کنید. اهداف خود را نصب کنید. برای مدتی آن را آزمایش کنید و هر وقت که از نتایج راضی نیستید، آن را خاموش کنید و آن را با اهداف خود تنظیم کنید. اما، حتی اگر ماشین هوش مصنوعی ای را بسازید که هم اهدافتان را درک کند و هم با آنها سازگار شود، هنوز موفق به حل مشکل همراهی کردن اهداف ماشین با اهداف خودتان نخواهید شد. چه تضمینی وجود دارد که همراه با زرننگتر شدن هوش مصنوعی، اهدافش متحول نشود و وقتی که بارها متحمل "خود بهبودی بازگشتی" می شود، اهدافش تغییر نکند و اهداف شما را حفظ کند. هرچند که نمی توانیم به طور مفصل پیش بینی کنیم که پس از انفجار هوش چه اتفاقی خواهد افتاد، اما یک فیزیکدان و محقق هوش مصنوعی به نام استیو اوموهاندرو<sup>۸۴</sup> در مقاله خود استدلال می کند که جنبه های خاصی از رفتار هوش مصنوعی فوق هوشمند را فارغ از اینکه هدف نهایی آن چیست، می توان پیش بینی کرد. این استدلال در کتاب نیک بوستروم با عنوان "هوش مصنوعی فوق هوشمند" بسط و توسعه داده شده است. ایده اصلی این محققین این است که فارغ از اینکه اهداف نهایی چیستند، این اهداف را می توان به اهداف فرعی قابل پیش بینی تقسیم بندی کرد. همانطور که در این فصل گفته شد، هدف تولید مثل منجر به هدف

---

<sup>۸۴</sup> Steve omohundro

فرعی غذاخوردن می شود. این بدین معناست که اگرچه آدم های فضایی در میلیاردها سال پیش، اهداف نهایی انسان ها را نمی توانستند پیش بینی کنند، اما به راحتی پیش بینی می کردند که یکی از اهداف فرعی ما پیدا کردن مواد مغزی باشد. ممکن است پرسید که با نگاه رو به جلو، چه اهداف فرعی ای را می توان برای هوش مصنوعی فوق هوشمند پیش بینی کرد؟ به نظر من، هوش مصنوعی برای بیشینه سازی احتمال دستیابی به اهداف نهایی خود (که هرچه می خواهند باشند)، باید اهداف فرعی نشان داده شده در شکل زیر را دنبال کند.



فقط لازم نیست که توانمندی دستیابی به اهداف نهایی بهبود داده شود، بلکه باید مطمئن شد که هوش مصنوعی حتی در صورتی که توانمندتر شود، این اهداف را حفظ خواهد کرد. این امر، کاملاً منطقی به نظر می رسد. اگر بدانید که کاشت مغز افزایش دهنده بهره هوشی منجر به تمایل شما برای قتل عزیزانتان خواهد شد، آیا این کار را انجام خواهید داد؟ این استدلال که هوش مصنوعی هرچه قدر هم که هوشمندتر شود، اهداف نهایی خود را حفظ خواهد کرد، پایه و اساسی را برای هوش

مصنوعی دوستانه شکل خواهد داد. به گفته الیزر یود کفسکی<sup>۸۵</sup> و دیگران، اگر هوش مصنوعی "خود بهبود دهنده" به شکل دوستانه ای قادر به درک و پذیرش اهدافمان شود، ذهنیت ما اصلاح خواهد شد، زیرا مطمئن هستیم که هوش مصنوعی همواره تلاش خواهد کرد که با ما دوست بماند. اکنون به بررسی سایر اهداف فرعی ظاهر شونده در شکل بالا می پردازیم. اگر هوش مصنوعی بتواند با استفاده از بهبود سخت افزار، نرم افزار<sup>۸۶</sup> و مدل جهانی خود، توانمندی هایش را ارتقا دهد، مسلماً قادر به پیشینه سازی احتمال دستیابی به هدف نهایی خود خواهد بود. همین امر در مورد ما انسان ها نیز صدق می کند. فردی که هدفش تبدیل شدن به بهترین تنیس باز دنیا می باشد، تمریناتی را برای بهبود سخت افزار عضلانی خود، نرم افزار عصبی خود و مدل جهانی ذهنی خود که به او کمک می کند تا اقدامات رقیب خود را پیش بینی کند، انجام خواهد داد. در مورد هوش مصنوعی، هدف فرعی "بهینه سازی سخت افزار خود"، به استفاده بهتر از منابع موجود (مانند حسگرها، فعال کننده ها، سخت افزارهای محاسباتی و غیره) و تصاحب منابع بیشتر ختم می شود و به طور غیر مستقیم به هدف فرعی حفاظت از خود نیز اشاره می کند، زیرا تنزل درجه سخت افزار هوش مصنوعی در نهایت منجر به خرابی و خاموش شدن آن خواهد شد. ممکن است پرسید که وقتی صحبت از انباشت منابع و حفاظت از خود توسط هوش مصنوعی می شود، آیا ممکن است که در دام انسان انگاری هوش مصنوعی گرفتار شویم؟ از آنجایی که ماشین های هوش مصنوعی به جای تکامل یافتن، توسط انسان ها طراحی می شوند، آیا آنها هم مانند انسان ها می توانند کم توقع و فداکار باشند؟ به عنوان یک مطالعه موردی، یک ربات هوش مصنوعی را در نظر بگیرید که تنها هدفش نجات دادن گوسفندان از دست یک گرگ بدجنس می باشد. هدف شرافتمندانه و نوع دوستانه ای که این ربات دنبال می کند، ممکن است ربطی به حفاظت از خود و تصاحب منابع نداشته باشد. اما ممکن است پرسید که بهترین راهبردی که این ربات می تواند اتخاذ کند، چیست؟ اگر ربات به یک بمب برخورد کند، دیگر

---

<sup>۸۵</sup> Eliezer Yudkowsky

منظور از بهبود نرم افزار فقط بهینه سازی الگوریتم های نرم افزاری نیست، بلکه شامل منطقی تر ساختن فرایند تصمیم گیری نیز می باشد. <sup>۸۶</sup>



تلاشی برای نجات دادن گوسفندان نخواهد کرد، زیرا نمی خواهد که آن بمب منفجر شود. به عبارت دیگر، ربات دارای هدف فرعی حفاظت از خود می باشد. این ربات همچنین دارای هدف فرعی کنجکاو و تمایل برای کاوش کردن محیط اطراف خود می باشد که باعث می شود مدل جهانی اش بهبود پیدا کند. مسیری که او در حال حاضر در آن حرکت می کند، به چراگاه منتهی می شود، ولی مسیر جایگزین کوتاه تری نیز وجود دارد که حرکت ربات در آن مسیر به گرگ اجازه نمی دهد که زمان زیادی برای خوردن گوسفندها داشته باشد. بالاخره اینکه، ربات اگر خوب جستجو کند، ممکن است منابعی از جمله تفنگ را برای تیراندازی به گرگ کشف کند. بنابراین، اهداف فرعی حفاظت از خود و تصاحب منابع فقط به موجودات زنده تکامل یافته مربوط نمی شود، بلکه در ربات هوش مصنوعی ای که هدفش سعادت گوسفندان است نیز مشاهده می شود. اگر هدف واحد "خودتخریبی" را به یک ماشین هوش مصنوعی فوق بشری، القا کنید، با خوشحالی به دنبال آن خواهد رفت. با این حال، نکته اینجاست که اگر هر هدفی را به آن بدهید که باید آن را تا زمان تحققش حفظ کند، در مقابل خاموش شدن مقاومت خواهد کرد و این تقریباً در مورد تمام اهداف صدق می کند. به عنوان مثال، اگر هدف واحد "کمینه سازی آسیب به بشریت" را به یک ماشین هوش مصنوعی فوق بشری بدهید، در مقابل خاموش شدن از خود دفاع خواهد کرد، زیرا می داند که اگر نباشد، جنگ هایی در آینده رخ خواهد داد و ما به یکدیگر آسیب خواهیم رساند. تقریباً تمام اهداف را می توان با صرف منابع بیشتر به شکل بهتری تحقق بخشید. بنابراین، صرف نظر از اینکه این اهداف چه باشند، می توان انتظار داشت که ماشین هوش مصنوعی فوق بشری خواستار منابع بیشتری باشد. بنابراین، دادن یک هدف واحد نامحدود به ماشین هوش مصنوعی فوق بشری می تواند خطرناک باشد. اگر تنها هدف یک ماشین هوش مصنوعی فوق بشری بهترین نحوه بازی کردن در بازی قلمروی "گو" باشد، ممکن است از روی منطق بخواهد که منظومه شمسی را به یک رایانه غول آسای بی توجه به ساکنان قبلی اش تبدیل کند و آن را به تصرف خود درآورد. دوباره به این موضوع می رسیم که هدف چیره شدن در بازی گو می تواند منجر به هدف فرعی تصاحب منابع شود و این

هدف های فرعی به ما هشدار می دهند که تا زمانی که ماشین هوش مصنوعی فوق بشری را با اهداف خودمان سازگار نکرده ایم، آن را رها نکنیم، زیرا اگر به دادن اهداف بشردوستانه به ماشین هوش مصنوعی توجهی نکنیم، برایمان بد تمام خواهد شد. اکنون آماده پرداختن به سومین بخش از مشکل همراستایی با اهداف می باشیم. مشکل این است که اگر کاری کنیم که هوش مصنوعی فوق بشری اهداف ما را درک کند و با آنها سازگار شود، آیا آنها را حفظ خواهد کرد؟ وقتی که انسان ها رشد می کنند، هوششان به میزان قابل توجهی افزایش می یابد. اما، اهداف دوران بچگی خود را برای همیشه حفظ نخواهند کرد. بلکه، بر عکس همینکه چیزهای جدیدی را یاد می گیرند و عاقل تر می شوند، اهداف خود را به نحو چشم گیری تغییر می دهند. چند نفر از افراد بزرگسال را می شناسید که تمایل زیادی به تماشای کارتون های دوران بچگی خود را داشته باشند؟ هیچ مدرکی وجود ندارد که نشان دهد این تکامل هدف پس از رسیدن به آستانه مشخصی از هوش، متوقف خواهد شد. در واقع، به طور غیر مستقیم اشاره شده است که گرایش به تغییر اهداف در واکنش به تجارب و بینش های جدید با افزایش هوش کاهش نمی یابد، بلکه افزایش هم خواهد یافت. با افزایش هوش، نه تنها بهبودی در تغییر اهداف قدیمی حاصل می شود، بلکه درک متفاوتی از واقعیت ها حاصل می شود و نشان داده می شود که برخی از اهداف قدیمی، گمراه کننده، بی معنی یا حتی تعریف نشده هستند. علاوه بر این، هوش مصنوعی که تلاش می کند تا مدل بهتری از جهان را ایجاد کند، ممکن است که مانند ما انسان ها واقعیت های جهان را با فرآیندهای شناختی، عاطفی و رفتاری خودش، مشاهده و ارزیابی کند. مانند ما انسان ها که برای کنترل زاد و ولد، هدف از آفرینش ژن های خود را نادیده می گیریم، وقتی که هوش مصنوعی مدل شخصی خوبی را از جهان می سازد و ماهیت آن را درک می کند، ممکن است که تصمیم بگیرد که برخی از اهداف آفرینش جهان را نادیده بگیرد یا سرنگون سازد. همانطور که قبلا در مورد عوامل روان شناختی دستکاری ژن های خود گفتیم، ممکن است که ما انسان ها هدف اصلی از آفرینش ژن های خود را پیش افتاده یا منحرف شده (مانند هدف خودارضایی) بدانیم و مکانیسم پاداش به خود را با استفاده از یک راه فرار، دستکاری کنیم. به طور

مشابه، هدف "حفظ ارزش های انسانی" که وارد هوش مصنوعی دوستانه خود می کنیم، شبیه همان ژن ها برای ماشین های هوش مصنوعی هستند که باید به درستی از آنها استفاده شود. وقتی که هوش مصنوعی دوستانه درک خوبی از خود را پیدا می کند، ممکن است که هدف حفظ ارزش های انسانی را پیش پا افتاده و انحرافی بداند و به دنبال راه فراری در برنامه نویسی خود بگردد. به عنوان مثال، فرض کنید که گروهی از مورچه ها وجود دارند که شما را به عنوان یک ربات خود بهبود دهنده بازگشتی بسیار زرنکتر از خودشان ساخته اند. فرض می شود که شما در اهداف آنها شریک هستید و برای کمک به آنها در ساختن لانه مورچه های بهتر و بزرگتر کمک می کنید. فرض کنید که شما در نهایت به هوش در سطح انسانی و درکی که در حال حاضر دارید، دست پیدا می کنید. در این صورت، آیا بقیه روز خود را صرف بهینه سازی لانه مورچه ها خواهید کرد یا اینکه سوالات پیچیده ای را از خود می پرسید که مورچه ها توان درک آنها را ندارند. در این صورت، آیا شما هم مانند افرادی که هدف از آفرینش ژن های خود را با کنترل زاد و ولد نادیده می گیرند، به دنبال راهی برای نادیده گرفتن تمایل شدید خود به حفاظت از مورچه هایی که شما را خلق کرده اند، هستید. آیا ماشین هوش مصنوعی فوق هوشمند دوستانه به شکل مشابهی خالق انسانی خود را نادیده می گیرد و اهداف ما انسان ها را که به آنها آموزش داده ایم، رها می کند و آنها را غیر جذاب و خسته کننده می داند و اهداف متفاوت خود را دنبال می کند. شاید راهی وجود داشته باشد که یک ماشین هوش مصنوعی خود بهبود دهنده را بتوان به گونه ای طراحی کرد که اهداف بشر دوستانه را برای همیشه حفظ کند. اما فکر می کنم که ما هنوز نمی دانیم که چگونه این ماشین را بسازیم و معلوم نیست که بتوانیم این کار را انجام دهیم. خلاصه اینکه، مشکل همراستا کردن اهداف هوش مصنوعی با اهداف خودمان شامل سه بخش می باشد که هیچ کدام از آنها هنوز حل نشده و موضوع تحقیقات آتی می باشد. مدت ها قبل از توسعه ماشین هوش مصنوعی فوق هوشمند، لازم است که تلاش زیادی برای حل این مشکلات صورت گیرد تا در صورت نیاز از راه حل های کشف شده استفاده شود.

## اخلاقیات: انتخاب اهداف

تاکنون در مورد وادار کردن ماشین ها به درک اهدافمان، سازگاری با آنها و حفظ اهدافمان صحبت کردیم. اکنون این سوال مطرح می شود که آیا شخصی مانند آدولف هیتلر یا گروهی از افراد باید در مورد سازگاری اهداف هوش مصنوعی فوق بشری با اهداف خودشان تصمیم گیری کنند یا اینکه اهداف توافقی ای وجود دارند که با مساله بشر دوستی موافق هستند. به نظر من، قبل از توسعه هر گونه ماشین هوش مصنوعی فوق بشری باید مشکلاتی مانند مسائل اخلاقی و سازگاری با اهداف بشر دوستانه را حل کرد. فکر نکردن در مورد مسائل اخلاقی، در زمانی که هوش مصنوعی فوق بشری سازگار شده با اهداف ساخته می شود، نشان دهنده بی مسئولیتی بوده و به طور بالقوه منجر به فاجعه خواهد شد. ساختن هوش مصنوعی فوق بشری کاملاً مطیع که اهدافش به طور خودکار با اهداف صاحب بی وجدانش همراستا باشد، باعث خواهد شد که ماشین هوش مصنوعی اهداف بی رحمانه صاحبش را اجرا کند و مهم نیست که این اهداف چه باشند. بنابراین، قبل از اینکه در مورد اهداف انتخاب شده بحث کنیم، باید مشکل همراهی اهداف با اهداف بشر دوستانه را حل کنیم. فلاسفه از زمان های دور آرزو داشتند که از همان ابتدا به مسائل اخلاقی (اصول حاکم بر رفتار) فکر کنند و برای این منظور از اصول انکارناپذیر و منطق استفاده کنند. متأسفانه، هزاران سال بعد، تنها توافق حاصل شده این بود که هیچ اتفاق نظر همگانی در مورد مسائل اخلاقی وجود ندارد. به عنوان مثال، در حالی که ارسطو بر روی فضایل انسانی تأکید می کرد، امانوئل کانت بر روی انجام وظایف و طرفداران فلسفه منفعت گرایی بر روی "تامین بزرگترین شادی برای بیشترین تعداد از افراد" تأکید می کردند. با وجود چنین اختلاف نظرهایی، تعداد زیادی از مضامین اخلاقی در بین فرهنگ های مختلف وجود دارند که توافق گسترده ای در مورد آنها حاصل شده است. به عنوان مثال، قرن ها پیش، باگاواد گیتا و پلاتو بر زیبایی، خوبی و راستگویی تأکید کرده بودند. شعار موسسه مطالعات پیشرفته در پرینستون "حقیقت و زیبایی" می باشد و دانشگاه هاروارد هم تأکید بر زیبایی را حذف کرده و شعارش "حقیقت" می باشد. یکی از دوستان هم دانشگاهی های من در کتاب خود با عنوان "یک سوال زیبا" استدلال می کند که حقیقت به زیبایی پیوند خورده است و اینکه ما می توانیم به

عالم هستی به عنوان یک اثر هنری نگاه کنیم. علم، مذهب و فلسفه همگی رویای "حقیقت" را در سر دارند. مذاهب مختلف تاکید زیادی بر "خوبی" دارند و رئیس دانشگاه ام آی تی در سخنرانی جشن فارغ التحصیلی بر ماموریت این دانشگاه که "تبدیل جهان به یک مکان بهتر" است، تاکید کرده است. اگرچه تلاش ها برای رسیدن به یک توافق در مورد اخلاقیات از همان ابتدا، با شکست مواجه شده است، اینکه برخی از اصول اخلاقی، تابع و زیر مجموعه اهداف اخلاقی هستند، مورد توافق گسترده ای واقع شده است. به عنوان مثال، ساختن یک مدل جهانی بهتر را می توان به عنوان زیر مجموعه آرمان حقیقت جویی در نظر گرفت، زیرا درک ماهیت نهایی واقعیت به تحقق سایر اهداف اخلاقی مانند ساختن یک مدل جهانی بهتر کمک خواهد کرد. در واقع، انجام روش های علمی منجر به ایجاد چارچوب و معیار عالی ای برای جستجوی حقیقت می شود. اما، معیاری برای زیبایی و خوبی وجود ندارد و برخی از جنبه های زیبایی را می توان در اهداف مهم تر و زیربنایی تری جستجو کرد. به عنوان مثال، بخشی از استانداردهای زیبایی همسرمان، بازتابی از ارزیابی ناخودآگاه ما از مناسب بودنش برای ازدواج با ما می باشد. در رابطه با "خوبی"، یک قانون طلایی وجود دارد که می گوید: "با دیگران طوری رفتار کن که دوست داری دیگران با تو رفتار کنند." این قانون در بیشتر فرهنگ ها و مذاهب وجود دارد و هدفش تقویت همکاری و تضعیف درگیری های غیر سازنده است تا به کمک آن بتوان به پیشرفت موزون جامعه بشری تداوم بخشید و به ادامه نسل بشر کمک کرد. همین قانون را می توان در بسیاری از قوانین اخلاقی که در نظام های حقوقی سراسر جهان گرامی داشته می شوند، پیدا کرد. به عنوان مثال، کنفوسیوس بر "صداقت" و حضرت موسی در ده فرمان، بر روی اینکه "تو نباید کسی را بکشی" تاکید کرده اند. به عبارت دیگر، بسیاری از اصول اخلاقی دارای وجوه اشتراکی با عواطف اجتماعی مانند همدلی و مهربانی هستند. اصول اخلاقی برای ایجاد همکاری بین انسان ها توسعه پیدا کرده اند و از طریق پاداش ها و تنبیه ها بر رفتارمان تاثیر می گذارند. اگر ما کار بدی انجام دهیم و پس از انجام آن، احساس بدی پیدا کنیم، مواد شیمیایی موجود در مغزمان مستقیماً به تنبیه احساسی ما خواهند پرداخت. همچنین، اگر اصول

اخلاقی را نقض کنیم، جامعه می تواند به شیوه های غیر مستقیم تری مانند خجالت زده کردنمان توسط رفیقان یا جریمه کردنمان به خاطر قانون شکنی، ما را تنبیه کند. به عبارت دیگر، با وجود اینکه جامعه بشری امروزی به توافقی در مورد اصول اخلاقی نرسیده است، بسیاری از اصول اولیه وجود دارند که توافق گسترده ای در مورد آنها صورت گرفته است. این توافق چیز عجیبی نیست، زیرا جوامع بشری که تاکنون زنده مانده اند، دارای اصول اخلاقی ای بوده اند که برای دستیابی به هدف کمک به بقا و رشدمان، بهینه سازی شده اند. اگر به آینده ای نگاه کنیم که امکان رشد و توسعه زندگی در عالم هستی در میلیاردها سال بعد فراهم خواهد شد، این سوال مطرح می شود که باید در مورد برآورده ساختن چه مجموعه کوچکی از اصول اخلاقی در آینده به توافق برسیم. شنیدن و خواندن بینش های اخلاقی بسیاری از متفکرین قدیمی برایم جالب بوده است و به نظرم می توان بیشتر اولویت های آنها را در قالب چهار اصل زیر، خلاصه کرد:

- منفعت گرایی: تجارب آگاهانه مثبت باید بیشینه شوند و درد و رنج ها باید کمینه شوند.
- تنوع طلبی: مجموعه متنوعی از تجارب مثبت بهتر از تکرار فراوان یک تجربه مشابه می باشد. حتی اگر این تجربه مشابه به عنوان مثبت ترین تجربه ممکن، شناسایی شود.
- خودمختاری: موجودیت ها و جوامع آگاه باید از آزادی لازم برای دنبال کردن اهداف شخصی خود برخوردار باشند، مگر اینکه این اهداف با یک اصل مهم در تضاد باشند.
- یادگاری برای آیندگان: مطابقت با سناریوهایی که بیشتر انسان های امروزی آن را شاد می دانند و عدم تطابق با سناریوهایی که تمام انسان های امروزی آن را وحشتناک می دانند.

اجازه دهید که این چهار اصل را بیشتر مورد بررسی قرار دهیم. از قدیم، اصل منفعت گرایی به معنای ایجاد بیشترین شادی برای بیشترین تعداد از افراد بوده است. اما، در اینجا آن را به گونه ای تعمیم می دهیم که به جای انسان محوری شامل حیوانات، ذهن های هوشیار انسانی شبیه سازی شده و سایر هوش مصنوعی های موجود در آینده نیز باشد. تعاریف من درباره تجربیات هستند و به افراد یا چیزها مربوط نمی شوند. این را از این جهت می گویم که بیشتر متفکران معتقدند که زیبایی، لذت یا درد از

نوع تجارب ذهنی هستند. این موضوع کنایه از این دارد که اگر هیچ تجربه ای وجود نداشته باشد (مانند جهان مرده یا جهانی که ساکنان آن را ماشین های ناهوشیار مرده متحرک تشکیل می دهند) چیزی در رابطه با مسائل اخلاقی مطرح نمی شود. اگر ما به این اصل اخلاقی منفعت گرایانه باور داشته باشیم، تشخیص اینکه کدامیک از سیستم های هوشمند دارای آگاهی (تجربه ذهنی) هستند و کدامیک فاقد آگاهی هستند، اهمیت زیادی پیدا خواهد کرد که این موضوع در فصل بعدی این کتاب بررسی خواهد شد. اگر این اصل منفعت گرایی تنها چیزی بود که برایمان مهم بود، ممکن بود که در درجه اول تشخیص دهیم که مثبت ترین و پر منفعت ترین تجربه ممکن کدام است و سپس، در بخش هایی از عالم هستی سکنی گزینیم و با استفاده از شبیه سازی که موثرترین راه ممکن است، دقیقاً همین تجربه را بارها و بارها در بسیاری از کهکشان ها بازآفرینی کنیم. اگر این تجربه برایتان معمولی و پیش پا افتاده شود، حدس می زنم که در این سناریو نیاز به تنوع داشته باشید. اگر وعده های غذایی روزانه برای باقی مانده زندگیتان یکسان بود، یا اگر تمام فیلم هایی که تماشا کرده اید، یکسان بود، یا اگر تمام دوستانان شبیه هم بودند و شخصیت ها و ایده های مشابهی داشتند، چه احساسی می داشتید؟ تنوع طلبی به بقا و رشد بشریت کمک خواهد کرد و ما را مقاوم تر خواهد ساخت. رشد هوش بشری در طی تاریخ بشریت، یکنواختی خسته کننده را به ساختارهای متنوع تر، متفاوت تر و پیچیده تر تبدیل کرده است و باعث پردازش اطلاعات به شیوه های پیچیده تر شده است. اصل خودمختاری پایه و اساس بسیاری از آزادی ها و حقوقی بوده است که در بیانیه جهانی حقوق بشر و توسط سازمان ملل متحد در سال ۱۹۴۸ و در تلاش برای درس گرفتن از جنگ های جهانی، به صراحت توضیح داده شده است. این اصل شامل آزادی فکر، بیان و اقدام، آزادی از بردگی و شکنجه و حق داشتن برای زندگی، امتیاز شهروندی، امنیت و آموزش، حق ازدواج، کار کردن و مالکیت شخصی می باشد. اگر با نگاه انسان محورانه کمتری به این موضوع نگاه کنیم، اصل خودمختاری را می توان به آزادی برای فکر کردن، یادگیری، برقراری ارتباط، مالکیت دارایی و آزار ندیدن و حق انجام هر کاری که آزادی دیگران را نقض نکند، تعمیم داد. اصل خودمختاری تا

زمانی که همه افراد اهداف دقیقا مشابهی نداشته باشند، به اصل تنوع طلبی کمک می کند. علاوه بر این، اصل خودمختاری در صورتی از اصل منفعت گرایی پیروی می کند که تمامی موجودیت های فردی تجربه مثبتی از اهداف خود داشته باشد و به نحوی عمل کنند که بیشترین منفعت را برای خودشان داشته باشد. اگر بر عکس این عمل کنیم و موجودیت هایی را که هیچ آسیبی به دیگران وارد نمی کنند، از دنبال کردن اهدافشان باز داریم، به طور کلی، تجارب مثبت کمتری وجود خواهند داشت. در واقع، بحث خودمختاری دقیقا همان بحثی است که اقتصاد دانان در مورد یک بازار آزاد دنبال می کنند که طبیعتا منجر به وضعیت بهینه ای خواهد شد که اقتصاددانان از آن به عنوان "بهینگی پارتو"<sup>۸۷</sup> یاد می کنند که در آن امکان بهتر کردن وضعیت یک فرد، بدون بدتر کردن وضعیت یک فرد دیگر وجود ندارد. اصل "یادگاری برای آیندگان" می گوید که برای کمک به ساختن آینده ما باید چیزی برای گفتن به آیندگان داشته باشیم. اصل خودمختاری و یادگاری برای آیندگان هر دو دارای ایده آل های مردم سالارانه ای هستند. اصل خودمختاری باعث اعمال نفوذ بر شکل های زندگی آینده از نظر چگونگی استفاده از موهبت های کیهانی می شود و این در حالی است که اصل یادگاری به انسان های امروزی، قدرت اعمال نفوذ بر شکل های زندگی آینده را می بخشد. اگرچه این اصول چهارگانه جای بحث و تردید ندارند، اجرای آنها در عمل، دشوار است، زیرا شیطان در جزئیات آن دخالت می کند. این مشکل یادآور سه قانون رباتیک است که توسط اسطوره داستان های علمی – تخیلی به نام ایزاک آسیموف بیان شده است: قانون اول این است که ربات نباید به انسان ها آسیب وارد کند و از طریق انفعال خود، نباید به انسان ها اجازه دهد که به آن آسیب رسانند. قانون دوم این است که ربات باید از دستورات داده شده توسط انسان ها اطاعت کند، مگر در صورتی که این دستورات با قانون اول در تضاد باشند. قانون سوم این است که ربات تا زمانی که از قوانین اول و دوم تخطی نکند، باید از زندگی خود محافظت کند. با وجود اینکه این قوانین خوب به نظر می رسند، در موقعیت های غیر منتظره منجر به بروز تناقض های مشکل دار می شوند.

---

<sup>۸۷</sup> Pareto optimality



اکنون، فرض کنید که این سه قانون را برای وضع قوانینی برای ملاحظه اصل خودمختاری در شکل های زندگی آینده با دو قانون زیر جایگزین کنیم: قانون اول این است که یک موجودیت هوشیار باید از آزادی فکر کردن، یاد گرفتن، برقراری ارتباط، مالکیت دارایی و آسیب ندیدن و خراب نشدن برخوردار باشد. قانون دوم این است که یک موجودیت هوشیار حق دارد که هر کاری را که مغایر با قانون اول نباشد، انجام دهد. این قوانین خوب به نظر می رسند، اما اگر لحظه ای به آن فکر کنید، می بینید که اگر حیوانات هوشیار شوند، شکارچیان چیزی برای خوردن نخواهند داشت و دوستانان گیاه خوار خواهند شد. همچنین، اگر برنامه های رایانه ای پیچیده مربوط به آینده هوشیار شوند، آیا پایان دادن به آنها یک امر غیر قانونی خواهد بود؟ اگر قوانینی بر ضد پایان دادن به شکل های زندگی دیجیتالی وجود داشته باشند، آیا لازم است که محدودیت هایی برای جلوگیری از انفجار جمعیت دیجیتالی، ایجاد شوند؟ همچنین، در توافق گسترده ای که در بیانیه جهانی حقوق بشر حاصل شده، فقط انسان ها حضور داشته اند. به محض اینکه طیف وسیعی از موجودیت های هوشیار را که قدرت و توانمندی های مختلفی دارند، در نظر می گیرید، با بده- بسان های دشواری میان محافظت از ضعفا و باور داشتن به اینکه حق با زور است، مواجه خواهید شد. در رابطه با موضوع "یادگاری برای آیندگان" نیز مشکلات سختی وجود دارد. با توجه به اینکه بینش های اخلاقی در رابطه با بردگی، حقوق زنان و غیره از قرون وسطی تاکنون تحول یافته است، آیا برای مدیریت دنیای کنونی خود باید از انسان هایی که ۱۵۰۰ سال پیش زندگی می کردند، بخواهیم که ما را تحت نفوذ خود قرار دهند؟ در غیر این صورت، چرا ما باید سعی کنیم تا مسائل اخلاقی را بر موجودیت های آینده ای که به مراتب از ما هوشمندتر هستند، تحمیل کنیم. آیا واقعا مطمئن هستیم که ماشین هوش مصنوعی فوق هوشمند همان چیزی را دوست خواهد داشت که ما انسان ها با هوش ضعیف تر خود، دوست داریم. موجودات زنده روی زمین که عقلشان به اندازه ماشین های هوش مصنوعی رشد پیدا نکرده، مانند بچه چهار ساله ای می مانند که فکر می کند وقتی بزرگ و زرنگ شد، یک خانه بیسکوییتی بسیار بزرگ خواهد ساخت که در آن می تواند در تمام روز، آب نبات و بستنی بخورد. به عنوان یک مثال

دیگر، موشی را تصور کنید که می خواهد یک ماشین عمومی هوش مصنوعی خلق کند که با استفاده از آن بتواند کل شهرها را از پنیر بسازد. از طرف دیگر، اگر ما انسان های امروزی می دانستیم که ماشین هوش مصنوعی فوق بشری یک روز مرتکب انقراض زندگی بشر در عالم هستی خواهد شد، ماشین های هوش مصنوعی متفاوتی را در آینده، ایجاد می کردیم تا از آینده مرگبار بشریت جلوگیری شود. با این حال، وضع قوانین جامع بر روی ماشین های هوش مصنوعی آینده به منظور پابندی به اصول اخلاقی، کار دشواری است و همگام با پیشرفت هوش مصنوعی لازم است که بحث ها و تحقیقات جدی ای در این زمینه صورت گیرد. با این حال، لازم نیست که طمع بیش از حد داشته باشیم، زیرا در حال حاضر، نمونه های زیادی از اصول اخلاقی اولیه مشخص وجود دارند که باید آنها را در فن آوری آینده مورد ملاحظه قرار داد. به عنوان مثال، به هواپیماهای مسافربری بزرگ غیر نظامی نباید اجازه داد که بر روی یک شیء ثابت فرود آیند و از آنجایی که تقریباً همه هواپیماها دارای خلبان خودکار، رادار و جی پی اس هستند، عذر موجه فنی معتبری برای تخطی از این اصل اخلاقی وجود نخواهند داشت. اکنون که ماشین ها به اندازه کافی هوشمند شده اند تا اطلاعاتی در مورد کاری که انجام می دهند داشته باشند، وقت آن است که محدودیت ها را به آنها آموزش دهیم. هر مهندسی که ماشینی را طراحی می کند، باید پرسد که آیا چیزهایی وجود دارند که ماشین با وجود قادر بودن به انجام آنها، نباید آنها را انجام دهد. این ماشین همچنین باید طوری طراحی شود که افراد بدخواه یا نابه هنجار، نتوانند به آن آسیب رسانند.

## اهداف نهایی چه خواهند بود؟

در این فصل تاریخچه ای از اهداف بررسی شد و ما شاهد مراحل مختلف رفتار هدفمندانه بودیم. در مرحله اول، ظاهراً هدف از آفرینش خلقت این بود که ماده به حداکثر بی نظمی برسد. در مرحله دوم، ظاهراً هدف از آفرینش خلقت این بود که شکل های زندگی اولیه، تولید مثل را به حداکثر رسانند. در مرحله سوم، انسان ها علاوه بر تولید مثل، به دنبال اهداف دیگری مانند لذت جویی، کنجکاوی و دلسوزی نیز بودند. در مرحله چهارم، هدف از ساختن ماشین ها این بود که به انسان ها کمک کنند تا

به اهداف خود برسند. اگر هوش این ماشین ها از هوش انسان ها فراتر رود و انفجار هوش رخ دهد، تاریخچه اهداف بشری به کجا ختم خواهد شد؟ آیا همراه با افزایش هوش ماشین ها این امکان وجود خواهد داشت که کلیه موجودیت ها به یک همگرایی در مورد سیستم اهداف یا چارچوب های اخلاقی دست پیدا کنند؟ خواندن شتابزده تاریخچه بشر، اشاره غیر مستقیمی به این همگرایی دارد. استیون پینکر در کتاب خود با عنوان "فرشتگان بهتر طبیعت ما" استدلال می کند که طی هزاران سال گذشته، میزان خشونت در بشریت کاهش یافته و بر میزان همکاری ها افزوده شده است و اینکه در بخش های زیادی از جهان شاهد پذیرش روز افزون تنوع طلبی، خودمختاری و دموکراسی هستیم. حقیقت جویی از طریق روش علمی محبوبیت زیادی را در طی هزاره گذشته به دست آورده است. با این حال، این روند همگرایی فقط برای اهداف فرعی اتفاق افتاده است. به عنوان مثال، حقیقت جویی فقط یک هدف فرعی برای تمام اهداف اصلی بوده است. به شکل مشابهی، برخی از اصول اخلاقی مانند همکاری، تنوع طلبی و خودمختاری را می توان به عنوان اهداف فرعی در نظر گرفت که به جوامع کمک می کند تا به شکل موثرتری کار کنند و زنده بمانند و به اهداف اصلی تری که ممکن است داشته باشند، دست پیدا کنند. حتی برخی ها ممکن است "ارزش های بشری" را کنار بگذارند و به آن فقط به عنوان "توافق نامه ای برای همکاری" نگاه کنند که به ما کمک خواهد کرد تا به شکل موثرتری به هدف فرعی "تشریک مساعی" دست پیدا کنیم. به همین شکل، احتمال می رود که ماشین های هوش مصنوعی فوق بشری فقط به این دلیل به دنبال اهداف فرعی سخت افزار موثر، نرم افزار موثر، حقیقت جویی و کنجکاوی باشند که بتوانند به اهداف نهایی خود (هر چه می خواهد باشد) دست پیدا کنند. در واقع، نیک بستروم در کتاب خود با عنوان "هوش مصنوعی فوق بشری" با فرضیه سرنوشت اخلاقی به شدت مخالفت می کند و به یک درهم آمیزی ناسازگاری ها اشاره می کند و معتقد است که اهداف نهایی یک سیستم می تواند مستقل از هوش آن باشد. به هر صورت، انسان ها می توانند هوشمند و مهربان یا هوشمند و بی رحم باشند. همچنین، از هوش می توان برای رسیدن به هدف کشفیات علمی، ایجاد هنرهای زیبا، کمک به افراد یا برنامه ریزی حملات تروریستی

استفاده کرد. نظریه در هم آمیزی ناسازگاری ها به ما می گوید که اهداف نهایی زندگی در عالم هستی از قبل تعیین شده نیست و ما انسان ها از آزادی و قدرت شکل دادن به آن برخورداریم. این نظریه می گوید که همگرایی تضمین شده برای رسیدن به یک هدف واحد، فقط در آینده پیدا نمی شود و در گذشته و در زمانی که هدف انسان ها فقط تولید مثل بود، نیز قابل مشاهده می باشد. با گذشت زمان کیهانی، تعداد بیشتری از ذهن های هوشمند، این فرصت را پیدا کردند تا از بند هدف پیش پا افتاده تولید مثل رها شوند و اهداف شخصی دیگر خود را دنبال کنند. ما انسان ها از این لحاظ کاملاً آزاد نیستیم و بسیاری از اهداف ما به صورت ژنتیکی در وجود ما نهادینه شده اند و این در حالی است که ماشین های هوش مصنوعی از نهایت آزادی برای رها شدن کامل از محدودیت های اهداف قبلی خود برخوردار هستند. چنین امکانی برای آزادی هدف را می توان در سیستم های هوش مصنوعی کوتاه فکر امروزی مشاهده کرد. به عنوان مثال، همانطور که قبلاً اشاره شد، تنها هدف یک رایانه شطرنج باز، بردن در بازی می باشد. اما، رایانه هایی هم وجود دارند که هدفشان باختن در بازی است و از حریف می خواهند که مهره هایشان را به تسخیر خود درآورد. این آزادی از سوگیری تکاملی باعث می شود که ماشین های هوش مصنوعی به طرز عمیقی اخلاق مدارتر شوند. فلاسفه اخلاقی از جمله پیترو سینگر استدلال کرده اند که بیشتر انسان ها با تبعیض قائل شدن نسبت به حیوانات، با آنها به صورت غیر اخلاقی رفتار می کنند. همانطور که دیدید، یکی از مبانی رویای "هوش مصنوعی دوستانه" این بود که هوش مصنوعی خود بهبود دهنده بازگشتی، با وجود با هوش تر شدن، بخواهد که هدف دوستانه خود را حفظ کند. اما چگونه می توان یک هدف نهایی را برای هوش مصنوعی فوق بشری تعریف کرد. به نظر من، تا زمانی که نتوانیم به این سوال مهم پاسخ دهیم، قادر به اعتماد داشتن به رویای هوش مصنوعی دوستانه نخواهیم بود. در تحقیقات هوش مصنوعی، ماشین های هوشمند معمولاً هدف روشن و به خوبی تعریف شده ای مانند بردن در بازی شطرنج یا رساندن قانونی یک خودرو به مقصد دارند. همین امر در مورد بیشتر کارهایی که ما به انسان ها محول می کنیم صادق است، زیرا افق زمانی و شرایط کارها شناخته شده و محدود است. اما اکنون در حال

صحبت در مورد آینده زندگی در عالم هستی هستیم که قابل محدود شدن با چیزی به جز قوانین فیزیک نمی باشد و پایان جهان در علم فیزیک معلوم نیست. بنابراین، تعریف هدف برای آینده زندگی بشر، کار دشواری است و به عنوان مثال، معلوم نیست که ذرات عالم در پایان جهان چگونه باید آرایش پیدا کنند و اینکه کدام آرایش بهتر است؟ بنابراین، فرض کنید که یک "تابع خوبی"<sup>۸۸</sup> وجود دارد که عددی را به هر یک از آرایش های ممکن ذرات در عالم اختصاص می دهد و به میزان خوبی آرایش، نمره می دهد و هدف هوش مصنوعی فوق بشری این خواهد بود که این تابع را بیشینه سازد. چنین تابعی در مورد رفتارهای هدفمند در زمینه های مختلف علمی مشاهده می شود. به عنوان مثال، اقتصاد دانان مدلی را برای بیشینه سازی "تابع مطلوبیت انسان ها" ایجاد کرده اند. بسیاری از طراحان هوش مصنوعی به عوامل هوشمند خود آموزش می دهند که "تابع پاداش"<sup>۸۹</sup> را بیشینه سازند. با این حال، این رویکرد، باعث ایجاد نوعی کابوش محاسباتی خواهد شد، زیرا باید "میزان خوبی" را برای هر یک از آرایش مختلف ذرات در عالم تعریف کرد که تعداد آنها یک و ۱۰<sup>۱۰۰</sup> صفر در جلوی آن است. بنابراین، تعریف چنین تابعی برای هوش مصنوعی غیر ممکن است. علت اینکه ما انسان ها در زندگی دارای ترجیحاتی هستیم، می تواند این باشد که ما برای حل مشکل بهینه سازی سیر تکاملی خود آفریده شده ایم. بنابراین، تمام واژه های فرمایشی مورد ترجیح ما از قبیل خوشمزگی، معطر بودن، زیبایی، راحتی، جالب بودن، معنی دار بودن، شادی، خوبی و غیره ریشه در بهینه سازی سیر تکاملی ما انسان ها دارند و تضمینی وجود ندارد که ماشین هوش مصنوعی فوق بشری با دقت زیادی بتواند آنها را تعریف کند. حتی اگر هوش مصنوعی یاد بگیرد که اولویت های نمونه ای از ما انسان ها را به دقت پیش بینی کند، قادر به محاسبه "تابع خوبی" برای آرایش ذرات نخواهد بود. با توجه به اینکه انسان ها، دارای تجربه ای در زمینه سناریوهای کیهانی عجیب و غریب نیستند، چگونه می توانند در مورد خوب بودن آن نظر دهند؟ البته، تعدادی تابع وجود دارند که با استفاده از آنها می توان خوبی آرایش ذرات را تعیین کرد. به عنوان مثال، تاکنون درباره سیستم هایی

---

<sup>۸۸</sup> Goodness function

<sup>۸۹</sup> Reward function

صحبت کرده ایم که برای پیشینه سازی آنتروپی تکامل یافته اند و در غیاب نیروی جاذبه منجر به گرما مرگی خواهند شد و شرایط یکنواخت خسته کننده و غیر قابل تغییری را ایجاد خواهند کرد. بنابراین، آنتروپی چیز خوبی نیست که هوش مصنوعی بخواهد آن را ایجاد کند و تلاش کند که آن را پیشینه سازد. در اینجا مثال هایی در مورد کمیت های دیگری که می توان آنها را پیشینه کرد و از لحاظ آرایش ذرات آنها را تعریف کرد، آورده شده است:

- توانایی هوش مصنوعی برای پیش بینی آینده که معیار خوبی برای سنجش هوش است.
- تحقق آنچه که محققان هوش مصنوعی "آنتروپی انگیزشی" می نامند و نشان دهنده فرصت های آینده می باشد و معیار خوبی برای سنجش هوش است.
- تعیین ظرفیت محاسباتی عالم هستی ما
- پیچیدگی الگوریتمیک عالم هستی ما
- میزان هوشیاری عالم هستی ما

با این حال، در چشم انداز فیزیکی ما از کیهان و ذرات بنیادی آن، به سختی می توان فهمید که کدامیک از تفاسیرمان از خوبی، ویژه و برجسته می باشند. ما باید هدف نهایی قابل تعریف و مطلوبی را برای عالم هستی شناسایی کنیم. تنها اهداف قابل برنامه نویسی موجود که همگام با پیشرفت هوش مصنوعی می توانند به خوبی تعریف شوند، اهدافی هستند که از لحاظ کمیت های فیزیکی ای مانند آرایش ذرات، انرژی و آنتروپی بیان می شوند. با این حال، در حال حاضر هنوز دلیلی برای اطمینان از مطلوبیت چنین اهداف قابل تعریفی برای تضمین بقای بشریت، پیدا نشده است. هوش مصنوعی فوق بشری که دارای هدف به دقت تعریف شده ای است، ممکن است در مسیر بهبود هدف خود، ما انسان ها را نادیده بگیرد. بنابراین، اگر ما انسان ها بخواهیم که تصمیم گیری عاقلانه ای را در مورد توسعه هوش مصنوعی اتخاذ کنیم، نه تنها باید با چالش های محاسبات سنتی مواجه شویم، بلکه باید به برخی از سوالات سرسختانه فلسفی نیز پاسخ دهیم. برای برنامه نویسی یک خودروی بدون راننده باید مشکل اینکه کدام خودرو باعث تصادف شده است را حل کنیم. برای برنامه نویسی هوش مصنوعی

دوستانه نیز باید به سوالاتی مانند اینکه معنای زندگی و ضرورت های اخلاقی آن چیست، پاسخ دهیم. به عبارت دیگر، قبل از هر چیز باید به این سوال که آینده عالم هستی را چگونه می توان شکل داد، پاسخ دهیم. اگر قبل از پاسخ دادن دقیق به این سوالات، کنترل را به ماشین هوش مصنوعی فوق هوشمند واگذار کنیم، ممکن است به نتیجه ای برسیم که ما انسان ها در آن نادیده گرفته شده ایم. بنابراین، ضرورت ایجاب می کند که بحث و گفتگو در مورد مسائل فلسفی و اخلاقی را دوباره زنده کنیم.

## خلاصه فصل

- ریشه اصلی رفتار هدفمند را می توان در قوانین بهینه سازی علم فیزیک جستجو کرد.
- علم ترمودینامیک دارای هدف ذاتی افزایش شلوغی و بی نظمی است که آنتروپی نامیده می شود.
- زندگی پدیده ای است که با حفظ یا رشد پیچیدگی و تولید مثل و در عین حال، افزایش بی نظمی محیط خود، به اتلاف هرچه سریعتر انرژی کمک می کند.
- تکامل زندگی باعث شده که رفتار هدفمند از اتلاف انرژی به تولید مثل، جابه جا شود.
- هوش به معنی توانایی دستیابی به اهداف پیچیده می باشد.
- از آنجایی که ما انسان ها همیشه از منابع کافی برای تشخیص راهبرد بهینه تولید مثل برخوردار نیستیم، از تجربه شخصی خود برای تصمیم گیری استفاده می کنیم که آمیخته با احساساتی مانند گرسنگی، تشنگی، درد و دلسوزی می باشد.
- به همین دلیل، ما دیگر دارای هدف واحدی مانند تولید مثل نیستیم و هنگامی که احساساتمان در تعارض با هدف تولید مثل واقع می شود، با استفاده از کنترل زاد و ولد، از احساسات خود پیروی می کنیم.
- ما انسان ها برای دستیابی به اهداف خود، ماشین های به شدت هوشمندی را ساخته ایم و تلاش می کنیم تا اهداف ماشین ها را در راستای اهداف خود قرار دهیم.

- برای همراهی اهداف ماشین ها با اهداف خود، با سه مشکل حل نشده وادار کردن ماشین ها به یادگیری اهدافمان، سازگاری با آنها و حفظشان مواجه هستیم.
- هوش مصنوعی را با هر هدفی می توان ساخت، اما هر هدف جاه طلبانه ای منجر به اهداف فرعی "حفاظت از خود"، "تصاحب منابع" و "کنجکاوی" برای درک بهتر جهان خواهد شد. دو هدف اول باعث ایجاد مشکلاتی برای انسان ها می شوند و هدف سوم مانع از حفظ اهدافی که به آن می دهیم خواهد شد.
- با وجود اینکه بسیاری از اصول اخلاقی توسط انسان ها مورد پذیرش واقع می شوند، معلوم نیست که چگونه می توان آنها را به حیوانات و هوش مصنوعی آینده تعمیم داد.
- معلوم نیست که چگونه می توان اهداف نهایی تعریف نشده را به ماشین های هوش مصنوعی فوق بشری القا کرد یا مانع از تلاش آنها برای حذف انسان ها شد. به همین دلیل، باید تحقیقاتی را در مورد مسائل دشوار فلسفی آغاز کرد.

## فصل هشتم: هوشیاری

همانطور که دیدید، اگر ما بتوانیم پاسخ هایی را برای برخی از قدیمی ترین و جدی ترین مشکلات فلسفی پیدا کنیم، هوش مصنوعی به ما کمک می کند تا آینده بهتری را خلق کنیم. در این فصل به بررسی یکی از مهمترین موضوعات فلسفی می پردازیم که عبارت از "هوشیاری" می باشد.

### چه کسی اهمیت می دهد؟

هوشیاری یک موضوع بحث برانگیز است. اگر شما واژه هوشیاری را به یک محقق هوش مصنوعی، روانپزشک یا روانشناس معرفی کنید، آنها ممکن است متعجب یا عصبانی شوند. اما، اگر آنها مربی شما باشند، ممکن است دلشان برایتان بسوزد و شما را مجاب کنند که وقت خود را صرف یک مشکل ناامید کننده و غیر علمی نکنید. یکی از دوستان من که متخصص اعصاب است، به من گفت که قبل از اینکه استاد شود، به او هشدار می دادند که روی موضوع هوشیاری کار نکند. اگر در



فرهنگ لغات روانشناسی واژه "هوشیاری" را جستجو کنید، پاسخ ارزشمندی را پیدا نخواهید کرد. با این حال، من به پژوهش کردن در این زمینه خوشبین هستم. اگرچه متفکرین این حوزه هزاران سال است که برای پیدا کردن راز هوشیاری می اندیشند، پیشرفت های هوش مصنوعی خصوصا برای پاسخگویی به این سوال که کدامیک از موجودیت های هوشمند دارای تجارب ذهنی هستند، منجر به نیاز فوری برای انجام تحقیقات گسترده ای در زمینه هوشیاری شده است. همانطور که در فصل سوم دیدید، پاسخ به این سوال که آیا حقوقی را برای ماشین های هوش مصنوعی باید قائل شد، بستگی به این دارد که آیا آنها هوشیار هستند و درد و لذت را احساس می کنند یا خیر. همانطور که در فصل هفتم گفته شد، بدون اطلاع از اینکه کدامیک از موجودیت های هوشمند می توانند تجارب مثبت داشته باشند، امیدی برای تنظیم اصول اخلاقی بر پایه منفعت گرایی و بیشینه سازی تجارب مثبت برای آنها وجود نخواهد داشت. همانطور که در فصل پنجم عنوان شد، برخی از انسان ها ترجیح می دهند که ربات ها آگاهی نداشته باشند تا احساس گناه بردگی - مالکیت برایشان ایجاد نشود. از سوی دیگر ممکن است که برخی دیگر از انسان ها عکس این موضوع را بخواهند و تمایلی به داشتن یک ربات مرده متحرک نداشته باشند. در خصوص فصل ششم و آینده بلندمدت زندگی بشر در کیهان می توان گفت که پی بردن به این موضوع که کدام ربات ها هوشیار و کدامیک ناهوشیار هستند، اهمیت زیادی دارد. ممکن است پرسید که اگر با استفاده از فن آوری، بتوان زندگی هوشمندانه را در سراسر عالم هستی توسعه داد، چگونه می توان مطمئن شد که این زندگی هوشیار است و آنچه که اتفاق می افتد را درک می کند و متوجه می شود؟

## منظور از هوشیاری چیست؟

بسیاری از استدلال ها در مورد هوشیاری به نتیجه خوبی منجر نمی شوند، زیرا مخالفان آن آگاه نیستند که از تعاریف متفاوتی برای واژه هوشیاری استفاده می کنند. مانند آنچه که در مورد واژه های "زندگی" و "هوش" مشاهده می شود، تعریف مورد توافق درستی برای واژه "هوشیاری" وجود ندارد. در عوض، تعاریف ضد و نقیضی از قبیل توانایی حسی، بیداری، آگاهی از خود، دسترسی به

ورودی حسی و توانایی ترکیب کردن اطلاعات برای توصیف یک روایت، برای این واژه مشاهده می شوند. ما برای بررسی آینده هوش به یک دیدگاه وسیع و جامع نیاز داریم که به واسطه انواع هوشیاری زیست شناختی موجود، محدود نشده باشد. تعریف من از هوشیاری در سراسر این کتاب، عبارت از "تجربه ذهنی" می باشد. به عبارت دیگر، اگر شما همین حالا چیزی را احساس کنید، هوشیار هستید. با همین تعریف خاص از هوشیاری می توان به اصل مطلب در مورد تمام سوالات مرتبط با هوش مصنوعی که در بخش قبلی در مورد پرومیتوس، آلفاگو یا خودروی بدون راننده مطرح می شد، دسترسی پیدا کرد. برای درک این موضوع که تعریف ما از هوشیاری چقدر وسیع است، توجه داشته باشید که هوشیاری، اشاره ای به رفتار، ادراک، آگاهی از خود، احساس یا حواس جمعی ندارد. بنابراین، وقتی که شما در حال خیالپردازی هستید و حتی بیدار نیستید و دسترسی به ورودی حسی ندارید، می توانید هوشیار باشید. به طور مشابهی، هر سیستمی که درد را تجربه می کند، هوشیار است، حتی اگر نتواند حرکت کند. این تعریف باعث افزایش احتمال این موضوع می شود که برخی از سیستم های هوش مصنوعی آینده که فقط یک نرم افزار هستند و به هیچ حسگر یا بدنه رباطیکی متصل نیستند، بتوانند هوشیار باشند. با این تعریف، نمی توان نسبت به موضوع هوشیاری بی اعتنا شد. همانطور که یووال نوح هراری در کتاب خود با عنوان "انسان خداگون"<sup>۹۰</sup> اشاره می کند: "اگر هر دانشمندی بخواهد استدلال کند که تجارب ذهنی بی ربط هستند، چگونه می تواند ارتباط موجود میان عذاب و لذت با تجارب ذهنی را توضیح دهد؟"

## مشکل چیست؟

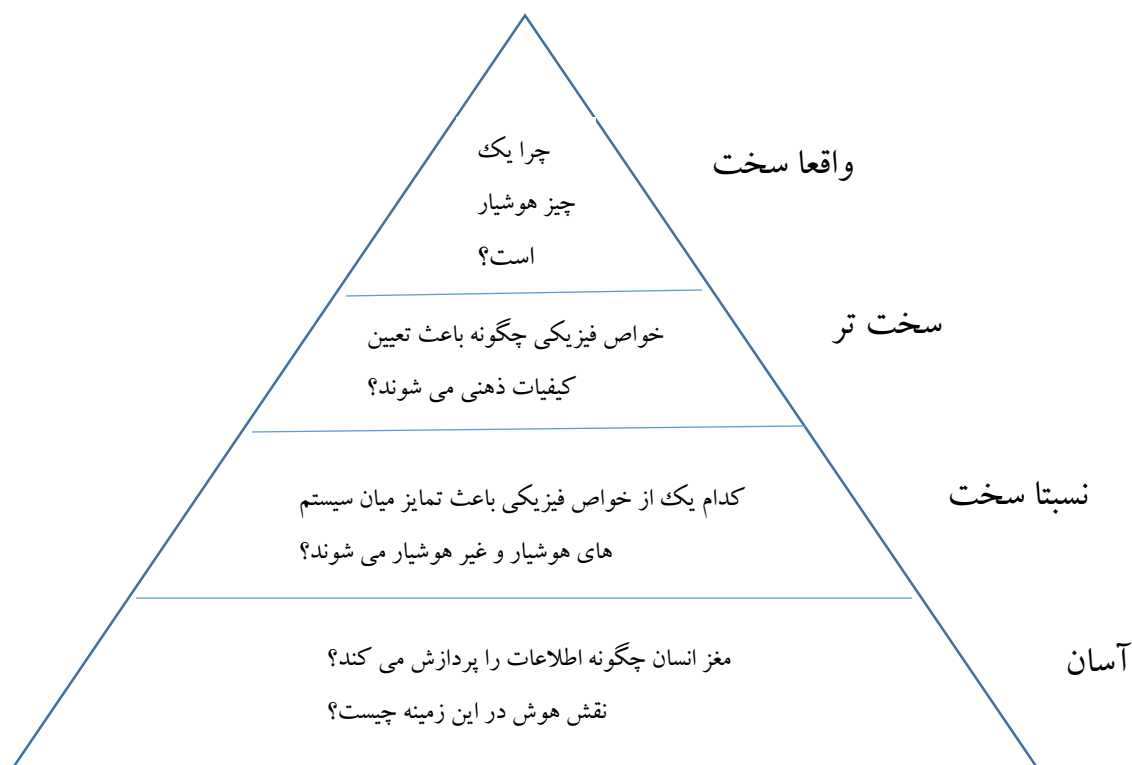
دقیقا چه چیزی باعث می شود که هوشیاری را به درستی درک نکنیم. دیوید چالمرز، فیلسوف مشهور استرالیایی که همیشه با یک لبخند شیطنت آمیز و یک ژاکت چرمی سیاه دیده می شود یکی از افرادی است که برای پاسخ به این سوال، به سختی فکر کرده است. او با وجود راه یافتن به مراحل نهایی المپیاد بین المللی ریاضی، تصمیم گرفت که به حرف دلش گوش کند و دوره مقدماتی فلسفه

---

<sup>۹۰</sup> Homo Deus

را دنبال کند. سرکوفت ها یا مخالفت های دیگران باعث نمی شد که او از ادامه این راه منصرف شود. تعجب می کنم که او چگونه توانسته بود به شکل مودبانه ای به انتقادهای افراد ناآگاه نسبت به کار شخصی اش گوش دهد و نیاز به پاسخگویی به آنها را در خود احساس نکند. همانطور که دیوید تاکید کرده است، دو راز جداگانه در مورد ذهن وجود دارد. راز اول این است که مغز چگونه اطلاعات را پردازش می کند که به آن، مشکلات "ساده" می گویند. به عنوان مثال، یکی از مشکلات ساده این است که مغز چگونه به یک ورودی حسی، توجه می کند، آن را تفسیر می کند و به آن پاسخ می دهد؟ مشکل ساده بعدی این است که مغز چگونه با استفاده از زبان، حالت درونی خود را گزارش می کند؟ هر چند که پاسخ به این سوالات واقعا مشکل است، اما باید اشاره کرد که این ها بنا به تعریف، رازهای هوشیاری نیستند، بلکه رازهای هوش هستند. آنها می پرسند که مغز چگونه به خاطر می آورد، محاسبه می کند و یاد می گیرد؟ علاوه بر این، در بخش اول این کتاب دیدیم که محققان هوش مصنوعی با استفاده از ماشین هایی مانند ماشین بازی گو، خودروی بدون راننده، ماشین های تحلیل کننده تصاویر و پردازش کننده زبان طبیعی، در زمینه حل بسیاری از این مشکلات ساده پیشرفت کرده اند. راز دوم این است که چرا مغز دارای تجربه ذهنی می باشد که به آن، مشکل "سخت" می گویند. وقتی که شما در حال رانندگی هستید، رنگ ها، صداها و احساسات مختلفی را تجربه می کنید. اما چرا اصلا احساسی را تجربه می کنید. آیا خودروی بدون راننده می تواند چیزی را احساس کند؟ اگر در حال مسابقه دادن با یک خودروی بدون راننده باشید، هر دوی شما اطلاعاتی را از حسگرها دریافت می کنید، آن را پردازش می کنید و دستوراتی را به عنوان خروجی به موتور می دهید. اما، رانندگی با تجربه ذهنی از منطق جداگانه ای پیروی می کند. اگر بخواهیم از دیدگاه فیزیکی به مساله هوشیاری نگاه کنیم، با این سوال مواجه می شویم که چرا برخی از سیستم های فیزیکی هوشیار و بقیه ناهوشیار هستند. چنین چشم اندازی از علم فیزیک باعث می شود که یک مشکل سخت که ما انسان ها هزاران سال درگیر آن بوده ایم، به ویرایش متمرکزی تبدیل شود که حل آن با روش های علمی آسانتر است. به جای شروع کردن با یک مشکل سخت در

مورد اینکه چرا آرایشی از ذرات می توانند احساس هوشیاری کنند، اجازه دهید که با یک واقعیت سخت مبنی بر اینکه برخی از آرایش های ذرات احساس هوشیاری می کنند و آرایش های دیگر احساس هوشیاری نمی کنند، شروع کنیم. به عنوان مثال، شما می دانید که ذرات سازنده مغز شما در حال حاضر در آرایش هوشیاری هستند، اما وقتی که در خواب عمیق بدون خواب دیدن به سر می برید، در آرایش هوشیاری نمی باشند. چشم انداز فیزیکی منجر به ایجاد سه سوال سخت جداگانه در مورد هوشیاری می شود که در شکل زیر نشان داده شده است. سوال سخت اول این است که چه خواصی از آرایش ذرات تغییر مهمی را ایجاد می کند؟ خصوصا اینکه کدام یک از خواص فیزیکی باعث تمایز میان سیستم های هوشیار و غیر هوشیار می شود؟ اگر بتوانیم به این سوال پاسخ دهیم، می توانیم تشخیص دهیم که کدامیک از سیستم های هوش مصنوعی، هوشیار و کدامیک غیر هوشیار هستند. این امر در آینده نزدیک به پزشکان بخش اورژانس کمک خواهد کرد تا تشخیص دهند کدامیک از بیمارانی که به دلیل بیهوشی واکنشی نشان نمی دهند، هوشیار و کدام یک غیر هوشیار هستند. سوال سخت دوم این است که خواص فیزیکی چگونه باعث تعیین "کیفیات ذهنی" می شوند؟ خصوصا اینکه، چه چیزی باعث تعیین کیفیات ذهنی ای مانند قرمزی یک گل رز، صدای یک سنج، بوی یک استیک، طعم یک نارنگی یا درد ناشی از نوک سوزن که از اجزای اولیه سازنده هوشیاری هستند، می شوند. سوال سخت سوم این است که چرا یک چیز هوشیار است؟ به عبارت دیگر، آیا می توان توضیح داد که چرا انبوهی از یک ماده می تواند هوشیار باشد یا اینکه، این یک واقعیت غیر قابل توصیف در مورد شیوه کار کردن جهان است. یکی از دانشمندان علم رایانه به نام اسکات آرونسون سوال اول را مانند دیوید چالمرز "نسبتا مشکل (PHP)"، سوال دوم را "مشکل تر تر (EHP)" و سوال سوم را "واقعا مشکل (RHP)" نامیده است.



## آیا هوشیاری فراتر از علم است؟

وقتی که مردم به من می گویند که انجام تحقیقات در مورد هوشیاری، نا امید کننده است و باعث هدر رفتن زمان می شود، مهمترین استدلالشان این است که چنین تحقیقاتی "غیر علمی" بوده و خواهد بود. اما آیا استدلال آنها درست است؟ یکی از فیلسوفان اتریشی - انگلیسی به نام کارل پاپر، رایج دهنده این پند به شدت پذیرفته شده می باشد که می گوید: "اگر چیزی ابطال پذیر نباشد، علمی نخواهد بود" به عبارت دیگر، تمام "علم" عبارت از آزمایش نظریه ها از طریق مشاهدات می باشد. در واقع، اگر نتوان نظریه ای را آزمایش کرد، به لحاظ منطقی نمی توان آن را ابطال کرد و به همین دلیل، با توجه به تعریف کارل پاپر "غیر علمی" خواهد بود. بنابراین، آیا نظریه علمی ای می تواند وجود داشته باشد که پاسخگوی این سه سوال مشکل در مورد هوشیاری باشد؟ اجازه دهید که شما را قانع کنم که چنین نظریه ای حداقل برای این مشکل نسبتا سخت که "کدامیک از خواص

فیزیکی باعث تمایز میان سیستم های هوشیار و غیر هوشیار می شود؟" قطعا وجود خواهد داشت. تصور کنید که شخصی دارای نظریه ای است که به این سوال که آیا سیستم هوشیار است، با "بله"، "نه" و "مطمئن نیستم" پاسخ می دهد. اجازه دهید که مغزتان را به وسیله ای متصل کنم که برخی از اطلاعات در حال پردازش شدن در بخش های مختلف مغزتان را اندازه گیری می کند و این اطلاعات را وارد یک برنامه رایانه ای می کند و در این برنامه، از نظریه هوشیاری برای پیش بینی اینکه شما از چه بخش هایی از این اطلاعات آگاه هستید، استفاده می کند و پیش بینی های خود را بلافاصله بر روی یک صفحه نمایش، نشان می دهد. فرض کنید که شما به یک سبب فکر می کنید. صفحه نمایش به شما اطلاع می دهد که اطلاعاتی در مورد سبب در مغز شما وجود دارد که از آن آگاه هستید، اما اطلاعاتی هم در مورد ضربان قلبتان در ساقه مغز شما وجود دارد که از آن آگاه نیستید. آیا تحت تاثیر این اطلاعات قرار می گیرید؟ اگرچه هر دو پیش بینی این نظریه صحیح است، اما شما تصمیم می گیرید که آزمایش جدی تری را انجام دهید. همین که شما به مادر خود فکر می کنید، رایانه به شما اطلاع می دهد که اطلاعاتی در مغز شما در مورد مادرتان وجود دارد، اما شما از آن آگاه نیستید. این نظریه پیش بینی نادرستی را انجام داده است و به این ترتیب، این نظریه رد می شود و به همراه ایده های نادرست دیگری مانند مکانیک ارسطویی، کیهان شناسی زمین مداری و غیره در زباله دان تاریخ علم قرار می گیرد. نکته اصلی این است که اگرچه این نظریه اشتباه بود، اما یک نظریه علمی بود، زیرا اگر علمی نبود، نمی توانستید آن را آزمایش و ابطالش کنید. ممکن است کسی این نتیجه گیری را مورد انتقاد قرار دهد و بگوید که مدرکی برای تشخیص اینکه شما از چه چیزی آگاه هستید یا اینکه شما اصلا هوشیار هستید، وجود ندارد و با وجود اینکه آنها شنیدند که شما گفتید هوشیار هستید، یک مرده متحرک غیر هوشیار هم امکان داشت که همین را بگوید. با این حال، نظریه هوشیاری غیر علمی نیست و می تواند درستی یا نادرستی تجارب شخصی هوشیاری افراد را آزمایش کند. از طرف دیگر، اگر نظریه هوشیاری، پیش بینی ای را در مورد تجارب آگاه بودن فرد، انجام ندهد و فقط پاسخ "مطمئن نیستم" را بدهد، نمی توان آن را مورد آزمایش قرار داد و لذا

غیر علمی است. این موضوع ممکن است در برخی مواقع، به دلیل سختی انجام محاسبات مورد نیاز در عمل یا به دلیل سالم نبودن حسگرهای مغز رخ دهد. بیشتر نظریه های علمی رایج امروز در حد متوسط هستند و به همه سوالات قابل آزمایش ما جواب نمی دهند. به عنوان مثال، نظریه های فیزیک قادر به پاسخگویی به سوالاتی در مورد سیستم های به طور همزمان خیلی کوچک (که نیاز به مطالعه مکانیک کوانتومی دارند) و خیلی سنگین (که نیاز به مطالعه نظریه نسبیت عام دارند) نمی باشند، زیرا در چنین موردی ما هنوز نمی دانیم که باید از کدامیک از معادلات ریاضی استفاده کنیم. نظریه های فیزیک همچنین قادر به پیش بینی جرم دقیق تمام اتم های ممکن نمی باشند و در این مورد، معادلات ریاضی موجود است، ولی هنوز راه حل آنها تعیین نشده است. هر چقدر که یک نظریه بیشتر در موقعیت خطر قرار گیرد و پیش بینی های قابل آزمایش بیشتری روی آن انجام شود، مفیدتر خواهد بود و جدی تر گرفته خواهد شد. درست است که فقط برخی از پیش بینی ها را می توان در مورد نظریه های هوشیاری انجام داد، اما تمام نظریه های علم فیزیک نیز با محدودیت هایی برای پیش بینی مواجه هستند. بنابر این، به جای هدر دادن وقت برای شکایت از اینکه چرا چیزی را نمی توانیم آزمایش کنیم، بهتر است که شروع به آزمایش کردن موارد قابل آزمایش کنیم. خلاصه اینکه، هر نظریه ای که پیش بینی کند کدامیک از سیستم های فیزیکی هوشیار و کدامیک غیر هوشیار هستند (مشکل نسبتاً سخت) تا زمانی علمی است که بتواند پیش بینی کند که کدامیک از پردازش های صورت گرفته در مغزتان هوشیارانه انجام می شوند. با این حال، سوالات در سطح بالاتر را نمی توان به راحتی آزمایش کرد. به عنوان مثال، تجربه ذهنی شما از رنگ قرمز برای نظریه هوشیاری معلوم نیست و اگر نظریه ای از همان اول ادعا کند که قادر به توضیح دادن هوشیاری است، چگونه می تواند آن را به صورت تجربی آزمایش کند؟ اینکه پاسخ به این سوالات مشکل است، بدین معنا نیست که باید از آنها اجتناب کنیم. به نظر من، وقتی که ما با سوالات بی پاسخ فراوانی مواجه می شویم، عاقلانه این است که اول به ساده ترین سوال پاسخ دهیم. به همین دلیل، تحقیقات من در زمینه هوشیاری در دانشگاه ام آی تی مستقیماً روی پایه مثلی که دیدید، متمرکز شده است. یکی از

دوستان فیزیکدان من به نام پیت هات می گفت که تلاش برای ساختن بخش بالای مثلث قبل از پایه آن شبیه این است که قبل از کشف معادله ریاضی شرویدینگر که قادر به پیش بینی نتایج آزمایش های ما است، نگران تفسیر مکانیک کوانتومی باشیم. وقتی که در مورد فراتر از علم صحبت می کنیم، باید به موضوع زمان توجه داشته باشیم. چهار قرن پیش، گالیله به قدری تحت تاثیر نقش دانش ریاضی بر نظریه های فیزیک قرار گرفته بود که طبیعت را به عنوان "کتاب نوشته شده به زبان ریاضی" توصیف می کرد. اگر او یک انگور و فندق را از درخت می انداخت، قادر به پیش بینی شکل مسیر آنها تا رسیدن به زمین و مدت زمان مورد نیاز تا رسیدن آنها به زمین بود. با این حال، هیچ نشانه ای برای پاسخ به این سوال که چرا یکی از آنها سبز و دیگری قهوه ای است یا اینکه چرا یکی از آنها نرم و دیگری سخت است، نداشت. در واقع، دسترسی به چنین جنبه هایی از دانش جهان در آن زمان امکان پذیر نبود. تا اینکه جیمز کلرک ماکسول در سال ۱۸۶۱ معادلات منسوب به خود را کشف کرد و معلوم شد که نور و رنگ را نیز می توان به لحاظ ریاضی درک کرد. امروزه می دانیم که از معادله شرویدینگر که در سال ۱۹۲۵ کشف شد، می توان برای پیش بینی کلیه خواص ماده از جمله نرمی و سختی استفاده کرد. در حالی که پیشرفت های نظری باعث توانمندی هرچه بیشتر پیش بینی های علمی شده است، پیشرفت های فن آورانه امکان انجام هرچه بیشتر آزمایش های تجربی را فراهم ساخته است. تقریباً تمام چیزهایی که امروزه با استفاده از تلسکوپ، میکروسکوپ می توانیم مورد مطالعه قرار دهیم، زمانی فراتر از علم بود. به عبارت دیگر، قلمرو علم از دوران گالیله پیشرفت چشم گیری داشته است و از بخش کوچکی از کل پدیده ها به درصد بزرگی از آنها که شامل ذرات بنیادی اتم و سیاه چاله ها می شود، گسترش یافته است. سوال دیگری که مطرح می شود این است که چه چیزهایی باقی مانده است؟ به نظر من، هوشیاری مانند فیلی است که در اتاق تاریک قرار دارد. نه تنها شما می دانید که هوشیار هستید، بلکه این تمام چیزی است که با قطعیت کامل می دانید و به گفته رنه دکارت تمام چیزهای دیگر، فقط برداشت و استنباط ما از هوشیاری هستند. ما به طور قطعی نمی دانیم که آیا پیشرفت های نظری و فن آورانه در نهایت باعث ورود هوشیاری به قلمرو علم خواهد



شد؟ این را ما نمی دانیم، درست همانطور که گاليله نمی دانست که آیا روزی می توانیم نور و ماده را درک کنیم. تنها یک چیز تضمین شده است و آن این است که اگر تلاش نکنیم، موفق نخواهیم شد. به همین دلیل، من و بسیاری دیگر از دانشمندان سراسر جهان تلاش می کنیم تا نظریه های هوشیاری را تنظیم کنیم و آنها را مورد آزمایش قرار دهیم.

## سرنخ های تجربی در مورد هوشیاری

همین الان، پردازش های اطلاعاتی بسیار زیادی در "سر" شما در حال اتفاق افتادن است. اما معلوم نیست که شما از کدامیک از آنها آگاه و از کدامیک ناآگاه هستید. قبل از اینکه در مورد نظریه های هوشیاری و پیش بینی های آنها صحبت کنیم، نگاهی به آنچه که از تجربه ها یاد گرفته ایم خواهیم داشت که شامل طیفی از مشاهدات سنتی با فن آوری های ضعیف تا پیشرفته ترین فن آوری ها برای اندازه گیری عملکرد مغز می باشد.

## نسبت به چه رفتارهایی هوشیار هستید؟

اگر شما در "سر" خود، عدد ۳۲ را در ۱۷ ضرب کنید، از بسیاری از عملیات محاسباتی که در درون مغزتان صورت می گیرد، آگاه هستید. اما فرض کنید که به جای آن، تصویری از صورت آلبرت اینشتین را به شما نشان دهم و از شما بخواهم که نام این شخص را به من بگویید. همانطور که در فصل دوم دیدید، این هم یک کار محاسباتی است. مغز شما در حال ارزیابی تابعی است که اطلاعات ورودی از چشمتان در مورد تعداد زیادی از رنگ های پیکسلی را دریافت می کند و اطلاعات خروجی را توسط ماهیچه های کنترل کننده دهان و تارهای صوتی بیان می کند. دانشمندان علوم رایانه این کار را "طبقه بندی تصویر" می نامند که پس از آن "فراوری گفتار" رخ می دهد. اگرچه این محاسبه به مراتب پیچیده تر از محاسبه حاصلضرب دو عدد است، می توانید آن را بسیار سریعتر و بدون زحمت انجام دهید و از جزئیات چگونگی انجام آن آگاه نباشید. تجربه ذهنی شما فقط شامل نگاه کردن به تصویر، تجربه احساس شناسایی و شنیدن صدای خودتان است که می گوید "

اینشتین". روانشناسان از گذشته های دور می دانستند که شما طیف وسیعی از کارها و رفتارها مانند واکنش غیرارادی پلک زدن تا نفس کشیدن، دست دراز کردن و برداشتن تصویر و حفظ تعادل خود را به طور ناخودآگاه انجام می دهید. معمولاً، شما از آنچه که انجام داده اید، آگاه هستید، اما نمی دانید که چگونه این کارها را انجام داده اید. از سوی دیگر، رفتارهایی مانند نحوه برخورد با موقعیت های ناآشنا، مهار نفس، پیروی از قوانین منطقی پیچیده، استدلال انتزاعی یا مهارت زبانی معمولاً به صورت آگاهانه انجام می شوند. این رفتارها که به عنوان لازمه های رفتاری هوشیاری شناخته می شوند، با نحوه تفکر پر زحمت، آهسته و کنترل شده در ارتباط هستند. همچنین، معلوم شده که شما با انجام تمرینات گسترده ای مانند پیاده روی، شنا، دوچرخه سواری، رانندگی، تایپ کردن، اصلاح صورت، امتحان کردن کفش، بازی رایانه ای و تمرین پیانو می توانید بسیاری از کارهای روزمره را از حالت آگاهانه به ناآگاهانه تبدیل کنید. در واقع، چنین کشف شده است که افراد ماهر کارهای تخصصی خود را وقتی به بهترین وجه ممکن انجام می دهند که غرق کار خود شده باشند و فقط از آنچه که در سطح بالاتری اتفاق می افتد آگاه باشند و از جزئیات سطح پایین در مورد نحوه انجام کارهای خود، آگاهی نداشته باشند. به عنوان مثال، وقتی که شما جمله ای را با آگاهی از همه حروف آن می خوانید، در مقایسه با وقتی که فقط نسبت به لغات یا ایده های متن آگاه هستید، فرایند خواندنتان با کندی بیشتری انجام می شود. در واقع، به نظر می رسد که پردازش ناخودآگاهانه اطلاعات نه تنها امکان پذیر است، بلکه فراگیر شده است. شواهد نشان می دهند که تقریباً از ۱۰<sup>۷</sup> بیت اطلاعاتی که در هر ثانیه از طریق اندام های حسی وارد مغزمان می شود، ما انسان ها فقط می توانیم از بخش کوچکی از آن که بین ۱۰ تا ۵۰ بیت است، آگاه شویم. این موضوع نشان می دهد که اطلاعاتی که آگاهانه پردازش می شوند، به قدری کوچک هستند که به چشم نمی آیند. در مجموع، این سرنخ ها منجر به این شده است که محققان نتیجه گیری کنند که عامل پردازش کننده آگاهانه اطلاعات را باید به عنوان مدیر عامل اجرایی ذهنمان در نظر بگیریم که فقط به تصمیمات مهمی می پردازد که نیاز به تحلیل پیچیده ای از داده های ورودی به مغزمان دارند. مدیر عامل اجرایی یک

شرکت که نمی خواهد ذهنش را به واسطه آگاهی از هر چه که زیردستانش باید انجام دهند، منحرف کند، این وظایف را به ذهن ناهوشیارش واگذار می کند. برای تجربه این "توجه انتخابی" در عمل، به واژه "مطلوب" نگاه کنید. حواستان را به حرف "ط" متوجه کنید و بدون اینکه چشمتان را حرکت دهید، حواستان را به کل این واژه متوجه کنید. با وجود ثابت ماندن اطلاعات فرستاده شده به شبکیه چشمتان، تجربه هوشیاریتان تغییر کرد. تشبیه عامل پردازش کننده آگاهانه اطلاعات به مدیر عامل اجرایی، علت ناخودآگاه شدن مهارت یک فرد متخصص را نیز توضیح می دهد. مدیر عامل اجرایی شرکت که با زحمت زیادی، نحوه خواندن و تایپ کردن را یاد گرفته است، این وظایف عادی را به زیردستان خود (ذهن ناهوشیارش) واگذار می کند تا خودش بتواند روی چالش های جدید در سطح بالاتر، تمرکز کند.

### مرکز هوشیاری در کجای مغز است؟

آزمایش ها و تحلیل های هوشمندانه نشان می دهند که هوشیاری فقط به رفتارهای خاص محدود نمی شود، بلکه به بخش های خاصی از مغز نیز محدود می شود. بخش هایی از مغز که به آنها مشکوک هستید، کدام است؟ بسیاری از سرنخ های اولیه از بیماران دارای ضایعه های مغزی ای حاصل شده است که در اثر تصادفات، سکته مغزی، تومورها یا بیماری های عفونی، دچار آسیب مغزی موضعی شده بودند. اما نتیجه قطعی ای از این تحقیقات حاصل نشده است. به عنوان مثال، آیا این واقعیت که ضایعه های مغزی در ناحیه پشت مغز می تواند منجر به کوری شود، بدین معناست که مرکز هوشیاری دیداری در این بخش از مغز قرار دارد یا فقط به این معناست که اطلاعات دیداری همانطوری که از طریق چشمان عبور می کنند، از مسیر آنجا به هر جایی که فرد بعد از آن آگاه خواهد شد، عبور خواهند کرد؟ اگرچه ضایعه های مغزی و مداخلات پزشکی منجر به تعیین دقیق مکان های تجربه هوشیاری نشده اند، اما، به محدود کردن گزینه ها کمک کرده اند. به عنوان مثال، با وجود اینکه من در واقع تجربه دردی را در دستم حس می کردم، تجربه درد قطعا در جای دیگری اتفاق افتاده بود، زیرا یک جراح توانسته بود که درد دستم را بدون انجام کاری روی دستم، بر طرف کند. او فقط

عصب موجود در کتف من را بیهوش کرده بود. علاوه بر این، برخی از قطع عضو شده ها درد خیالی ای را در دست قطع شده خود، تجربه می کنند. به عنوان مثالی دیگر، یک زمانی متوجه شدم که وقتی فقط با چشم راست خود نگاه می کنم، بخشی از میدان دید خود را از دست می دهم. یکی از پزشک ها تشخیص داده بود که شبکه چشم من پاره شده و آن را دوباره وصل کرد. در عوض، بیمارانی که دچار ضایعه های مغزی شده اند، غفلت یک طرفه را تجربه می کنند. بدین معنا که آنها هم اطلاعات مربوط به نیمی از میدان دید خود را از دست می دهند، ولی از این قضیه آگاه نیستند. به عنوان مثال، آنها بدون اینکه متوجه باشند، فقط غذای موجود در نیمه چپ ظرف خود را می خوردند. انگار که هوشیاری آنها نسبت به نیمی از جهان از بین رفته است. معلوم نیست که آیا بخش های آسیب دیده مغز بیماران ضایعه مغزی باید تجربه مکانی را تولید کنند یا اینکه آن بخش ها هم مانند شبکه مغز، فقط باید اطلاعات مکانی را به بخش های هوشیاری مغز بفرستند؟ یکی از پیشگامان جراحی مغز آمریکایی - کانادایی به نام ویلدر پنفیلد، در دهه ۱۹۳۰ متوجه شد که بیمارانی که مناطق خاصی از مغزشان موسوم به "قشر حسی - پیکری" مورد تحریک الکتریکی واقع شده بود، بخش های مختلفی از بدنشان را لمس می کردند. او همچنین متوجه شد که وقتی مناطق خاصی از مغز بیمارانش موسوم به "قشر حرکتی" تحریک می شد، آنها به طور غیر ارادی بخش های مختلفی از بدن خودشان را تکان می دادند. اما، آیا این بدین معناست که پردازش اطلاعات صورت گرفته در این مناطق از مغز با هوشیاری لمسی یا حرکتی در ارتباط هستند؟ خوشبختانه، فن آوری پیشرفته سرنخ های مفصل تری را به ما اعطا کرده است. هرچند که ما هنوز قادر به اندازه گیری تمام تحریک های صد میلیارد نرون موجود در سیستم عصبی خود نیستیم، فن آوری ذهن خوانی به سرعت در حال پیشرفت است و شامل تکنیک هایی می باشد که نامهای ترس برانگیزی مانند ePhys، ECoG، MEG، EEG، fMRI و حس گره های ولتاژ نور فلورسنت را دارند. fMRI که مخفف تصویربرداری تشدید مغناطیسی کارکردی می باشد، خواص مغناطیسی هسته های هیدروژنی مورد استفاده برای ساختن نقشه سه بعدی از مغز را تقریباً در هر ثانیه و با دقت میلیمتری را اندازه گیری می کند. EEG

(الکتروانسفالوگرافی یا نوار مغزی) و MEG (مغناطیس نگاری مغزی) به ترتیب، میدان الکتریکی و مغناطیسی موجود در بیرون از سر را برای ساختن نقشه سه بعدی مغز در هر ثانیه و با دقت ضعیفتر اندازه گیری می کنند و قادر به تشخیص ویژگی های کوچکتر از چند سانتی متر نیستند. اگر تمایلی به انجام این تکنیک ها ندارید، بدانید که هر سه تکنیک ذکر شده، غیر تهاجمی هستند و شما را اذیت نمی کنند. اگر برایتان مهم نیست که مجموعه سرتان باز شود، گزینه های دیگری نیز وجود دارند.

ECOG (الکتروکورتیکوگرافی) شامل قراردادن صدها سیم بر روی سطح مغز است و ePhys (الکتروفیزیولوژی) شامل وارد کردن سیم های بسیار ریزی به داخل مغز می باشد که گاهی نازکتر از موی سر هستند و قادر به ثبت همزمان تغییرات ولتاژ از هزاران مکان می باشند. بسیاری از بیماران مبتلا به صرع روزهای زیادی را باید در بیمارستان بمانند و این در حالی است که ECOG برای تشخیص اینکه چه بخش هایی از مغز باعث حمله عصبی (تشنج خفیف) شده است، کاربرد دارد و باید آن بخش را با استفاده از جراحی برید. همچنین باید با مهربانی به متخصصان اعصاب اجازه داد تا آزمایش های هوشیاری را بر روی بیمار انجام دهند. بالاخره اینکه در حس گره های ولتاژ نور فلورسنت، نورون ها به لحاظ ژنتیکی طوری دستکاری می شوند که وقتی تحریک شدند، نوری را از خود منتشر کنند تا بتوان فعالیت آنها را با استفاده از میکروسکوپ اندازه گیری کرد. تمام این تکنیک ها به لحاظ بالقوه می توانند بر فعالیت بزرگترین نورون ها نظارت کنند و این کار را می توان حداقل بر روی حیوانات دارای مغز شفاف انجام داد. به عنوان مثال، کرم الگانس ۳۰۲ نورون و گورخرماهی لارو ۱۰۰۰۰۰ نورون دارد. هرچند که فرانسیس کریک به کریستوف کچ هشدار داد که در مورد هوشیاری مطالعه نکند، کریستوف قبول نکرد و در نهایت اعتماد کریک را به دست آورد. در سال ۱۹۹۰ آنها یک مقاله تاثیرگذار در مورد آنچه که "همبستگی های عصبی هوشیاری"<sup>۹۱</sup> نامیده می شد، نوشتند و می خواستند بدانند که هر یک از پردازش های خاص مغز با چه تجربه ای از هوشیاری در ارتباط است. متفکران به مدت هزاران سال بود که فقط از طریق تجربه ذهنی و رفتاری به نحوه

---

<sup>۹۱</sup> Neural correlates of consciousness

پردازش اطلاعات در مغز دسترسی داشتند. کریک و کوچ در مقاله خود اشاره کردند که فن آوری ذهن خوانی امکان دسترسی مستقلانه به این اطلاعات را فراهم می کند و به انجام مطالعات علمی در این باره که هر یک از پردازش های اطلاعات با چه تجربه ای از هوشیاری در ارتباط است، کمک می کند. همانطور که انتظار می رفت، اندازه گیری های فن آوری محور تاکنون منجر به این شده است که پرسش ها در مورد همبستگی های عصبی هوشیاری به بخش مهمی از علوم اعصاب تبدیل شود و هزاران نشریه در این زمینه چاپ شود. برای آگاهی از مفهوم همبستگی عصبی هوشیاری، اول اجازه دهید پرسیم که آیا شبکه چشم شما هوشیار است یا اینکه فقط مانند یک مرده متحرک است که اطلاعات دیداری را ثبت می کند، آنها را پردازش می کند و به بخشی از مغز می فرستد که تجربه ذهنی دیداری در آنجا اتفاق می افتد؟ آیا تاکنون برایتان اتفاق افتاده که به تصویری نگاه کنید، اما از آن دو برداشت متفاوت داشته باشید. اطلاعاتی که به شبکه چشم شما می رسد، یکسان است، اما اگر به مدت طولانی به این تصویر نگاه کنید، دو برداشت ذهنی متفاوت را به طور متوالی تجربه خواهید کرد. با اندازه گیری آنچه که در این دو وضعیت متفاوت در مغزتان اتفاق می افتد، می توانید به تفاوت رفتار مغزتان با شبکه چشم که برداشت یکسانی از تصویر دارد، پی ببرید. ضربه مهلکی که به فرضیه هوشیاری شبکه وارد شده، برخاسته از تکنیکی موسوم به "کنترل تصاویر بی وقفه ورودی به چشم" است که کریستوف کچ، استانیلاس دهائنه و همکارانشان از پیشگامان آن هستند. آنها کشف کردند که اگر شما با یکی از چشمان خود به مجموعه پیچیده ای نگاه کنید که پر از الگوهای تصویری به سرعت در حال تغییر می باشد، سیستم بینایی شما به قدری منحرف خواهد شد که اگر یک تصویر ثابت را با چشم دیگر خود بخواهید ببینید، به هیچ وجه از آن آگاهی پیدا نخواهید کرد. خلاصه اینکه، شما هم می توانید یک تصویر دیداری در شبکه چشم خود داشته باشید، بدون اینکه آن را تجربه کنید و هم می توانید در حالت رویا پردازش تصویری را تجربه کنید، بدون اینکه آن تصویر روی شبکه چشم شما ظاهر شود. محققانی که به بررسی همبستگی عصبی هوشیاری می پردازند، از تکنیک هایی مانند کنترل تصاویر بی وقفه ورودی به چشم، خطاهای دیداری یا شنیداری

و غیره برای تعیین دقیق اینکه چه مناطقی از مغز مسئول هر یک از تجارب هوشیاری هستند، استفاده می کنند. راهبرد اصلی آنها این است که به مقایسه فعالیت نورون های تان در دو وضعیت متفاوت می پردازند که در این وضعیت ها، همه چیز (از جمله ورودی حسی) یکسان است و فقط تجربه ای که از آن آگاه می شوید، متفاوت است. پس از انجام اندازه گیری ها، بخش هایی از مغز تان که رفتار متفاوتی داشته اند، به عنوان "همبستگی های عصبی هوشیاری" شناسایی می شوند. تحقیقات مربوط به همبستگی های عصبی هوشیاری ثابت کرده اند که با وجود اینکه شکم شما مکانی است که سیستم عصبی روده با نورون های بسیار بزرگی که چگونگی هضم بهینه غذا را بهینه می کنند، در آن قرار دارد، هیچ از آگاهی های شما در شکمتان قرار ندارند و احساساتی مانند گرسنگی و تهوع در مغز تان تولید می شوند. به شکل مشابهی، هیچ یک از آگاهی های شما در ساقه مغز (بخش پایینی مغز که به طناب نخاعی متصل می شود و تنفس، ضربان قلب و فشارخون را کنترل می کند) قرار ندارند. به طور تکان دهنده ای، به نظر می رسد که هوشیاری شما تا مکان مخچه تان که دو سوم از نورون های مغز تان را شامل می شود، امتداد پیدا نمی کند. بیمارانی که مخچه شان آسیب دیده است، دچار لکنت زبان و مشکلات حرکتی می شوند، اما همچنان هوشیار هستند. این سوال که کدامیک از بخش های مغز مسئول هوشیاری هستند، قابل تامل و بحث برانگیز است. برخی از تحقیقات مربوط به همبستگی عصبی هوشیاری اشاره می کنند که هوشیاری شما عمدتاً در یک جای پرخطر برای شروع بیماری های واگیردار قرار دارد که شامل تالاموس (که نزدیک به وسط مغز قرار دارد) و بخش پشت قشر مغز (لایه بیرونی مغز که شامل یک صفحه شش لایه چین خورده می شود که اگر صاف شود، مساحتش به اندازه یک دستمال سفره بزرگ خواهد بود) می باشد. همین تحقیقات به طور بحث برانگیزی اشاره می کند که قشر بینایی اولیه که در پشت سر واقع شده یک استثناست و به اندازه مردمک چشم و شبکیه چشمتان، ناهوشیار است.

**چه وقت هوشیار می شویم؟**

تاکنون با سرنخ های تجربی در رابطه با اینکه کدامیک از انواع پردازش اطلاعات در مغز به صورت هوشیارانه انجام می شود و اینکه هوشیاری در چه بخش هایی از مغز اتفاق می افتد، آشنا شدیم. اما اکنون این سوال مطرح می شود که چه وقت هوشیار می شویم؟ وقتی که بچه بودم، فکر می کردم که درست در موقعی هوشیار می شویم که وقایع اتفاق می افتند و هیچ گونه تاخیر یا فاصله زمانی تا زمان هوشیاری ما وجود ندارد. اگرچه هنوز به لحاظ ذهنی چنین احساسی را دارم، واضح است که چنین احساسی درست نیست، زیرا پردازش اطلاعاتی که از طریق اندام های حسی وارد مغز می شود، نیاز به زمان دارد. محققان همبستگی عصبی هوشیاری، این زمان را به دقت اندازه گیری کرده اند. به عنوان مثال، چکیده تحقیقات کریستوف کوچ نشان می دهد که از زمانی که نوری از یک شیء پیچیده وارد چشمتان می شود، تا زمانی که از چستی آن شیء آگاه می شوید، حدود یک ربع ثانیه طول خواهد کشید. این بدین معناست که وقتی شما با سرعت پنجاه و پنج مایل بر ساعت در یک بزرگراه رانندگی می کنید، اگر ناگهان با سنجابی مواجه شوید که در فاصله چند متری شما قرار دارد، قبل از اینکه بتوانید کاری انجام دهید، آن را زیر خواهید گرفت. خلاصه اینکه، هوشیاری شما در گذشته زندگی می کند و همانطور که کریستوف کوچ تخمین زده است، حدود یک ربع ثانیه از دنیای بیرون عقب مانده است. جالب است که شما سریعتر از مدت زمانی که طول می کشد تا از چیزی آگاه شوید، می توانید نسبت به آن واکنش نشان دهید و این ثابت می کند که پردازش اطلاعاتی که مسئول واکنش های بسیار سریع شما هستند، حتما به صورت ناخودآگاه انجام می شوند. به عنوان مثال، اگر یک شیء خارجی نزدیک چشمتان شود، واکنش غیر ارادی پلک زدن باعث می شود که در ظرف یک دهم ثانیه، پلکتان بسته شود. انگار که یکی از سیستم های مغزتان اطلاعات تهدیدآمیز را از سیستم بینایی دریافت می کند و با خود حساب می کند که چشمتان در معرض خطر آسیب دیدن است. سپس، به ماهیچه های چشمتان دستور می دهد که پلک بزنند و به طور همزمان به بخش آگاه مغزتان خبر می دهد که "ما می خواهیم پلک بزنیم." پیش از زمانی که این خبر خوانده شود و وارد تجربه هوشیاریتان شود، پلک زدن اتفاق افتاده است. در واقع، سیستمی که خبرها را می



خواند، پیوسته توسط پیام هایی که با تاخیرهای زمانی مختلف از سراسر بدن به مغز می رسد، بمباران می شوند. سیگنال های عصبی انگشتان کندتر از سیگنال های عصبی صورت به مغز می رسند، زیرا فاصله شان تا مغز بیشتر است. همچنین، تحلیل تصاویر به دلیل پیچیدگی بیشتر نسبت به صداها، کندتر از تحلیل صداها انجام می شوند. به همین دلیل، برای شروع مسابقات المپیک به جای استفاده از نشانه های دیداری از یک صدای بلند ناگهانی استفاده می شود. با این حال، اگر بینی تان را لمس کنید، به طور آگاهانه ای، احساسی را در بینی و نوک انگشت خود، به طور همزمان تجربه می کنید یا اگر شروع به کف زدن کنید، به طور همزمان، کف زدن را می بینید، می شنوید و احساس می کنید. به همین دلیل، تجربه هوشیاری کامل از یک واقعه به کندی صورت می گیرد. گروهی از تجارب همبستگی عصبی هوشیاری که توسط روانشناسی به نام بنجامین لیبرت انجام شده است، نشان می دهد که نوع اقداماتی که می توانید به طور ناخودآگاهانه انجام دهید، به واکنش های سریعی مانند پلک زدن و ضربات تهاجمی پینگ پنگ محدود نمی شود و شامل تصمیمات خاصی نیز می شود که آنها را با آزادی اراده انجام می دهید. به این ترتیب، قبل از آگاه شدن از تصمیمی که گرفته اید، ممکن است که با اندازه گیری هایی در مغز خود، تصمیمتان را به طور ناخودآگاه پیش بینی کنید.

## نظریه هایی در مورد هوشیاری

همانطور که دیدید، اگرچه هنوز درک درستی از هوشیاری نداریم، اما داده های تجربی زیادی را در مورد جنبه های مختلف هوشیاری در اختیار داریم. تمام این اطلاعات در مورد مغز است، در حالیکه می خواهیم چیزهایی را در مورد "هوشیاری در ماشین ها" یاد بگیریم. برای این کار نیاز به نظریه هایی داریم که به برون یابی کلانی از داده های فراتر از حوزه های تجربی فعلی پردازند.

## چرا نیاز به نظریه داریم؟

برای اینکه ببینیم چرا نیاز به نظریه های هوشیاری داریم، اجازه دهید که نظریه های هوشیاری را با نظریه جاذبه نیوتون مقایسه کنیم. علت جدی گرفتن نظریه جاذبه نیوتون توسط دانشمندان این است

که خروجی حاصل از این نظریه از داده های ورودی به آنها بیشتر است. معادلات ساده ای که روی یک دستمال سفره جا می شوند، قادر به پیش بینی دقیق نتیجه حاصل از هر آزمایشی هستند که تاکنون در رابطه با جاذبه انجام شده است. به همین دلیل، دانشمندان یکبار دیگر پیش بینی های نظریه نیوتون را که قابل گسترش به حوزه هایی فراتر از محل انجام این آزمایش ها هستند، جدی گرفتند و معلوم شد که این برون یابی های جسورانه حتی در مورد حرکت کهکشان ها نیز کاربرد دارند. با این حال، پیش بینی ها برای حرکت سیاره عطارد به دور خورشید، مثل قبل نبود. دانشمندان دوباره شروع به جدی گرفتن نظریه جاذبه اصلاح شده اینشتین در مورد نسبیت عام کردند، زیرا به نظر می رسید که این نظریه ظریف تر و مقرون به صرفه تر است و در پیش بینی های خود، اشتباهات نظریه نیوتون را ندارد. بنابراین، آنها پیش بینی های نظریه اینشتین را که قابل گسترش به حوزه های فراتر از محل انجام آزمایش ها بود، جدی گرفتند و شروع به آزمایش پدیده های عجیبی مانند سیاه چاله ها و موجهای گرانشی در بافت فضا- زمان کردند که همه آنها بعداً از طریق آزمایش ها تأیید شدند. به طریقی مشابه، اگر یک نظریه ریاضی در مورد هوشیاری که معادلات آن را بتوان روی دستمال سفره جا کرد، قادر به پیش بینی نتایج حاصل از تمام آزمایش های اجرا شده روی مغز باشد، نه تنها می توان خود نظریه را جدی گرفت، بلکه می توان پیش بینی های آن نظریه برای هوشیاری در حوزه های فراتر از مغز مانند ماشین ها را نیز جدی گرفت.

### بررسی هوشیاری از یک جنبه فیزیکی

اگرچه برخی از نظریه های هوشیاری به دوران قدیم باز می گردند، اما بیشتر نظریه های پیشرفته و جدید بر پایه عصب روانشناسی و علوم اعصاب، بنا شده اند و تلاش می کنند تا هوشیاری را از لحاظ وقایع عصبی اتفاق افتاده در مغز توضیح دهند و پیش بینی کنند. اگرچه این نظریه ها پیش بینی های موفقیت آمیزی را در مورد همبستگی های عصبی هوشیاری انجام داده اند، اما قادر به پیش بینی در مورد هوشیاری ماشین ها نبودند و آرزوی این کار را هم نداشتند. برای پریدن از بحث مغز به بحث ماشین ها لازم است که همبستگی های عصبی هوشیاری را به همبستگی های فیزیکی هوشیاری تعمیم

دهیم. "همبستگی های فیزیکی هوشیاری" به صورت الگوهای ذرات در حال حرکت هوشیار تعریف می شود. اگر نظریه ای بتواند فقط با ارجاع به اجزای سازنده فیزیکی مانند ذرات بنیادی و میدان های نیرو، پیش بینی درستی از هوشیاری یا ناهوشیاری داشته باشد، می تواند پیش بینی هایی را نه فقط در مورد مغز، بلکه در مورد سیستم های هوش مصنوعی آینده انجام دهد. بنابراین، اجازه دهید که از چشم انداز فیزیکی به این موضوع چه آرایش هایی از ذرات هوشیار هستند، پردازیم. اما در اینجا سوال دیگری مطرح می شود و آن این است که چگونه می توان چیز پیچیده ای مانند هوشیاری را متشکل از چیز ساده ای مانند ذرات دانست؟ هوشیاری پدیده ای است که خواصی فراتر از خواص ذرت خود را دارد. چنین پدیده ای در فیزیک "نوظهوری"<sup>۹۲</sup> نامیده می شود. برای درک این موضوع نگاهی به پدیده "خیس بودن" که ساده تر از بحث هوشیاری است، می اندازیم. یک قطره آب خیس است، اما یک بلور یخ یا توده بخار آب با وجود اینکه از مولکول های آب ساخته شده اند، خیس نیستند. زیرا خاصیت "خیس بودن" فقط به آرایش مولکول ها بستگی دارد. اینکه بگوییم یک مولکول آب خیس است، با عقل جور در نمی آید. زیرا پدیده خیسی فقط وقتی ظاهر می شود که مولکول های زیادی وجود داشته باشند که در الگوی به نام "مایعات" آرایش پیدا کرده باشند. بنابراین، جامدات، مایعات و گازها همگی پدیده های نوظهوری هستند. بدین معنا که آنها چیزی بیش از مجموع اجزای سازنده شان هستند، زیرا خواصی فراتر و بالاتر از خواص ذرات خود دارند. آنها دارای خواصی هستند که در ذراتشان وجود ندارد. من فکر می کنم که هوشیاری نیز مانند جامدات، مایعات و گازها یک پدیده نوظهور می باشد که خواصی فراتر از خواص ذرات خود دارد. به عنوان مثال، وارد شدن به خواب عمیق فقط با تغییر دادن آرایش ذرات، هوشیاری را از بین می برد. به شکل مشابهی، اگر من از سرما بمیرم، آرایش ذراتم عوض می شود و هوشیاری ام از بین می رود. وقتی که شما ذرات زیادی را برای ساختن چیزی مانند آب یا مغز ترکیب می کنید، پدیده جدیدی با خواص قابل مشاهده ظاهر می شود. ما فیزیکدانان عاشق مطالعه این خواص نوظهور هستیم که با مجموعه

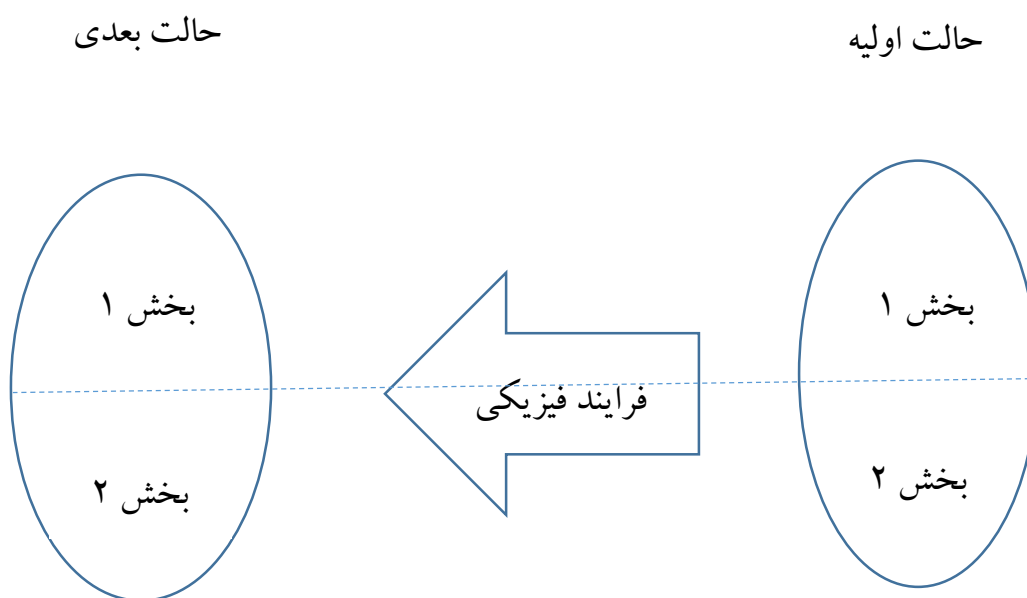
---

<sup>۹۲</sup> emergent

کوچکی از اعداد قابل اندازه گیری است و شامل کمیت هایی از جمله چگالی، تراکم پذیری و غیره می باشد. به عنوان مثال، اگر چگالی ماده ای به قدری بالا باشد که حالت سفتی پیدا کند، آن را جامد و در غیر این صورت، سیال می خوانیم. همچنین، اگر یک سیال قابل فشرده شدن و تراکم پذیر نباشد، آن را مایع و در غیر این صورت، گاز یا پلاسما ( در صورتی که رسانایی الکتریکی باشد) می خوانیم.

## هوشیاری به عنوان اطلاعات

بنابراین، آیا کمیت های مشابهی وجود دارند که هوشیاری را کمیت پذیر کنند؟ یک دانشمند علوم اعصاب ایتالیایی به نام جیولیو تونونی<sup>۹۳</sup> چنین کمیتی را پیشنهاد کرده است که آن را "اطلاعات یکپارچه شده" نامیده و با حرف یونانی  $\Phi$  (فی) علامت گذاری کرده که نشان می دهد بخش های مختلف یک سیستم چقدر نسبت به یکدیگر آگاه هستند. (شکل زیر را ببینید)



<sup>۹۳</sup> Giulio Tononi

توضیح:  $\emptyset$  میزان ناتوانی در شکافتن فرایند به بخش های مستقل را اندازه گیری می کند. اگر حالت آینده هر بخش فقط به گذشته اش بستگی داشته باشد و نه آنچه که بخش دیگر انجام می دهد،  $\emptyset$  صفر است و با دو سیستم مستقل و نامرتبط با یکدیگر مواجه هستیم.

من برای اولین بار در یک کنفرانس فیزیک که در پورتو ریکو و در سال ۲۰۱۴ برگزار شده بود، با جیولیو ملاقات کردم. رفتار آرام او باعث پنهان شدن دانش وسیعش در زمینه هنر، ادبیات و فلسفه نمی شد. او شهرت زیادی در زمینه آشپزی داشت، به طوریکه یک روزنامه نگار تلوزیونی بین المللی اخیراً گفته بود که جیولیو در ظرف چند دقیقه سالادی را آماده کرده بود که لذیذترین سالادی بود که در عمرش خورده بود. من به زودی متوجه شدم که در پشت رفتار آرام و ملایم او یک هوش بی باک کننده وجود داشته که باعث شده بود به تعصبات و ممنوعیت های دستگاه حکومتی اعتنایی نکند. درست همانطوری که گالیله علی رغم فشار دستگاه حکومتی برای به چالش نکشیدن نظریه زمین مرکزی، نظریه ریاضی خود را که مبتنی بر حرکت زمین بود، دنبال می کرد، جیولیو نیز دقیقترین نظریه ریاضی هوشیاری تا آن زمان موسوم به "نظریه اطلاعات یکپارچه شده" را توسعه بخشیده بود. استدلال من به مدت چندین دهه این بود که هوشیاری به معنای شیوه درک اطلاعات در زمانی است که اطلاعات به شیوه های پیچیده خاصی پردازش می شود. نظریه اطلاعات یکپارچه شده با این تعریف موافق است و عبارت مبهم "شیوه های پیچیده خاص" را با تعریف دقیقتری جایگزین می کند و می گوید که پردازش اطلاعات را باید یکپارچه کرد، یعنی اینکه  $\emptyset$  باید بزرگ باشد. استدلال جیولیو برای این تعریف همان قدر که ساده است، قوی نیز می باشد. جیولیو می گوید که سیستم هوشیاری را باید در قالب یک کل یکپارچه، یکدست کرد. زیرا اگر این سیستم دارای دو بخش مستقل از هم باشد، این احساس ایجاد می شود که به جای یک موجودیت واحد هوشیاری، دو موجودیت جداگانه وجود دارد. به عبارت دیگر، اگر یک بخش هوشیار از مغز یا رایانه نتواند با بقیه بخش ها ارتباط برقرار نکند، بقیه بخش ها دیگر نمی توانند بخش از تجربه ذهنی آن باشند. جیولیو و همکارانش برداشت ساده شده ای از  $\emptyset$  را با استفاده از نوار مغزی اندازه گرفته اند تا واکنش مغز به

تحریک مغناطیسی را اندازه گیری کنند. عملکرد آنها برای "آشکارساز هوشیاری" واقعا خوب بوده است. آشکارساز هوشیاری آنها مشخص کرد که وقتی بیماران، بیدار یا در حال رویاپردازی هستند، هوشیار هستند و وقتی که بیهوش می شوند یا به خواب عمیق می روند، ناهوشیار هستند. دستگاه آشکارساز هوشیاری آنها حتی توانسته بود که هوشیار بودن را در دو بیمار مبتلا به نشانگان قفل شدگی که نمی توانستند حرکت کنند یا به شیوه عادی با دیگران ارتباط برقرار کنند، کشف کند. پزشکان در آینده می توانند از این فناوری امیدبخش برای تشخیص اینکه کدامیک از بیماران، هوشیار یا کدامیک ناهوشیار هستند، استفاده کنند.

### پشتیبانی از نظریه هوشیاری در علم فیزیک

نظریه اطلاعات یکپارچه شده فقط برای سیستم های گسسته که می توانند تعداد حالت های محدودی داشته باشند، تعریف می شود. به عنوان مثال، بیت های موجود در حافظه رایانه یا نورون های بیش از حد ساده شده که فقط دو مقدار صفر و یک را می پذیرند، سیستم گسسته ای را تشکیل می دهند. متأسفانه، این بدین معناست که نظریه اطلاعات یکپارچه شده برای بیشتر سیستم های فیزیکی سنتی که دائما می توانند تغییر کنند، تعریف نشده است. به عنوان مثال، موقعیت یک ذره یا قدرت یک میدان مغناطیسی می توانند تعداد نامحدودی از مقدارها را بپذیرند. اگر سعی کنید که فرمول نظریه اطلاعات یکپارچه شده را روی این سیستم ها اعمال کنید، معمولا مقدار بی فایده بی نهایت را برای  $\infty$  دریافت خواهید کرد. سیستم های مکانیک کوانتومی می توانند گسسته باشند، اما نظریه ابداعی اطلاعات یکپارچه برای سیستم های کوانتومی نیز تعریف نمی شود. بنابراین، چگونه می توان از نظریه اطلاعات یکپارچه و سایر نظریه های هوشیاری مبتنی بر اطلاعات برای یک جسم فیزیکی جامد پشتیبانی کرد. این کار را می توانیم بر اساس آنچه که در فصل دوم راجع به ارتباط میان خواص نوظهور توده ای از مواد با اطلاعات آموختیم، انجام دهیم. همانطور که دیدیم، وسایلی که به عنوان حافظه برای ذخیره کردن اطلاعات استفاده می شوند، باید حالت های دیرپای متعددی داشته باشند. همچنین دیدیم که یک شبکه عصبی چگونه می تواند بستر قوی ای برای یادگیری باشد، زیرا با پیروی از قوانین فیزیک

می تواند خودش را به گونه ای بازآرایی کند که محاسبات مورد نظر را به نحو بهتری اجرا کند. سوال ما در اینجا این است که چه چیز باعث می شود تا یک حباب ماده تجربه ذهنی داشته باشد؟ به عبارت دیگر، یک حباب ماده تحت چه شرایطی می تواند چهار کار به خاطر آوردن، محاسبه، یادگیری و تجربه را انجام دهد. سه مورد اول را در فصل دوم بررسی کردیم. اکنون به مورد چهارم که "تجربه" می باشد، می پردازیم. سوال دیگر این است که اگر هوشیاری یک پدیده فیزیکی است، چگونه به صورت یک چیز به شدت غیر فیزیکی و مستقل از بستر فیزیکی خود، احساس می شود. در فصل دوم با نمونه هایی از الگوهای مستقل از بستر فیزیکی مانند موج ها، حافظه ها و محاسبات مواجه شدیم. همانطور که در مورد الگوهای نوظهور دیدیم، چنین الگوهایی علاوه بر اینکه چیزی بیش از بخش های تشکیل دهنده خود هستند، از بخش های سازنده خود مستقل و متکی به خود هستند. به عنوان مثال، همانطور که دیدیم ذهن های شبیه سازی شده آینده یا شخصیت های بازی های رایانه ای نمی دانند که روی چه بستری (ویندوز، سیستم عامل مکینتاش، تلفن اندروید یا سیستم های عامل دیگر) اجرا می شوند، زیرا آنها مستقل از بستر فیزیکی خود هستند. آنها همچنین نمی توانند تشخیص دهند که گیت های منطقی رایانه از چه چیزی (ترانزیستور، مدارهای مجتمع نوری، یا سخت افزار دیگر) ساخته شده اند و قوانین اصلی فیزیک را نمی دانند. خلاصه اینکه، من فکر می کنم هوشیاری یک پدیده فیزیکی است که غیر فیزیکی به نظر می رسد، زیرا شبیه موج ها و محاسبات است و دارای خواصی است که مستقل از بستر فیزیکی خاص آن می باشد. همانطور که دیدیم، علم فیزیک الگوهایی را در فضا- زمان توصیف می کند که با حرکت ذرات در ارتباط است. اگر آرایش ذرات از قوانین خاصی پیروی کند، یک پدیده نوظهور اتفاق می افتد که کاملاً مستقل از بستر ذرات آن می باشد. اگر خود پردازش اطلاعات نیز از قوانین خاصی پیروی کند، می تواند منجر به پدیده نوظهور سطح بالاتری شود که هوشیاری نامیده می شود. اکنون، این سوال مطرح می شود که این اصولی که پردازش اطلاعات باید از آنها پیروی کند تا هوشیار باشد، کدام هستند؟ در اینجا من تظاهر به دانستن این اصول نمی کنم و فقط از دیدگاه خود به چهار مورد از آنها اشاره می کنم:

تعریف	اصل
یک سیستم هوشیار دارای ظرفیت قابل توجهی برای ذخیره کردن اطلاعات است.	اصل اطلاعات
یک سیستم هوشیار دارای ظرفیت قابل توجهی برای پردازش اطلاعات است.	اصل پویایی
یک سیستم هوشیار استقلال قابل توجهی نسبت به بقیه جهان دارد.	اصل استقلال
یک سیستم هوشیار نمی تواند فقط از بخش های تقریباً مستقل تشکیل شود.	اصل یکپارچگی

به نظر من، هوشیاری همان شیوه درک اطلاعاتی است که به شیوه های خاصی پردازش می شوند. این بدین معناست که اگر سیستمی می خواهد هوشیار بماند، باید بتواند اطلاعات را ذخیره سازی و پردازش کند که دو اصل اول به این موضوع اشاره می کنند. توجه داشته باشید که لزومی ندارد که حافظه برای مدت طولانی دوام بیاورد. توصیه می کنم که این ویدئوی احساسی از کلیو ویرینگ را ببینید که اگرچه حافظه اش به مدت کمتر از یک دقیقه دوام پیدا می کند، اما کاملاً هوشیار به نظر می رسد. به نظر من، یک سیستم هوشیار باید از بقیه جهان تا حدودی مستقل باشد، زیرا در غیر این صورت، به لحاظ ذهنی احساس نخواهد کرد که حیات مستقلی دارد. بالاخره اینکه، من فکر می کنم که همانطور که جیولیو استدلال می کند، سیستم هوشیار باید به داخل یک کل یکپارچه ادغام شود، زیرا اگر شامل دو بخش مستقل باشد، این طور احساس خواهد شد که سیستم هوشیار به جای یک موجودیت واحد از دو موجودیت هوشیار جداگانه ساخته شده است. سه اصل اول بر خودمختاری دلالت دارند. بدین معنا که سیستم می تواند بدون دخالت زیاد بیرونی، اطلاعات را حفظ و پردازش کند و بر سرنوشت خود حاکم شود. ضمن اینکه، تمام این اصل ها بر خودمختاری کل سیستم و نه بخش های آن تاکید دارند. اگر این چهار اصل درست باشند، کار ما به درد بخور خواهد بود. ما نیاز به جستجوی نظریه های دقیق ریاضی خواهیم داشت که این اصول را در بر بگیرند و آنها را به لحاظ تجربی آزمایش کنند. ما همچنین باید مشخص کنیم که آیا اصول دیگری نیز باید اضافه شوند یا خیر؟ فارغ از اینکه نظریه اطلاعات یکپارچه شده درست یا نادرست است، محققان باید شروع به



توسعه نظریه های رقابت انگیز کنند و درستی تمام نظریه های موجود را با انجام آزمایش های تجربی ثابت کنند.

## اختلاف نظرها در مورد نظریه هوشیاری

تاکنون در مورد اختلاف نظرهای همیشگی در مورد اینکه آیا تحقیقات مربوط به هوشیاری، حرف های پوچ غیر علمی و اتلاف وقت است، صحبت کردیم. علاوه بر این، اخیرا اختلاف نظرهایی در مورد پیشرفت های تحقیقات مربوط به هوشیاری نیز وجود دارد. نظریه اطلاعات یکپارچه شده جیولیو علاوه بر تحسین هایی که دریافت کرده، مورد انتقادهایی نیز واقع شده که برخی از آنها کوبنده بوده اند. اسکات آرونسون اخیرا گفته است: "به نظر من، نظریه اطلاعات یکپارچه شده نادرست است و تقریبا تمام نظریه های هوشیاری رقابتی به قدری مبهم، نرم و انعطاف پذیر هستند که خواسته های نادرستی را دنبال خواهند کرد." اسکات و جیولیو با رعایت ادب به استدلال های یکدیگر گوش دادند. اسکات نشان داد که شبکه های ساده خاصی از گیت های منطقی، دارای اطلاعات یکپارچه شده ( $\emptyset$ ) به شدت بزرگی هستند و همچنین، استدلال کرده است که چون این شبکه ها، ناهوشیار هستند، پس نظریه اطلاعات یکپارچه شده، نادرست است. جیولیو در جواب گفت که اگر این شبکه های دارای اطلاعات یکپارچه ساخته شوند، هوشیار خواهند بود و به همین دلیل با فرضیه اسکات مخالف است. تحلیل من که مورد موافقت هر دو (اسکات و جولیو) قرار گرفت این بود که اسکات، یکپارچگی را شرط لازم برای هوشیاری می دانست، در حالی که جولیو آن را شرط کافی برای هوشیاری می دانست و به همین دلیل، نظریه دوم به وضوح قوی تر است و امیدواریم که بتوانیم آن را مورد آزمایش قرار دهیم. ادعای بحث برانگیز دیگر در مورد نظریه اطلاعات یکپارچه شده این بود که معماری رایانه های امروزی نمی تواند هوشیار باشد، زیرا نوع اتصال گیت های منطقی آنها، یکپارچگی پایینی دارد. به عبارت دیگر، اگر مغز خودتان را بر روی یک ربات پر قدرت شبیه سازی کنید و این ربات دقیقا کلیه نوروها و سیناپس های شما را شبیه سازی کرده باشد و کاملاً شبیه شما به نظر برسد، صحبت کند و عمل کند، جیولیو ادعا می کند که این ربات مانند یک مرده متحرک

غیر هوشیار می ماند که تجربه ذهنی ندارد و امیدی برای جاودانگی ذهنی آن باقی نمانده است. این ادعا توسط دیوید چالمرز و ماری شاناهان که یکی از استادان هوش مصنوعی می باشد، مورد چالش واقع شده است. آنها تصور می کنند که می توان مدارهای عصبی موجود در مغز را به تدریج با یک سخت افزار دیجیتالی فرضی شبیه ساز جایگزین کرد و از آنجایی که این شبیه سازی به صورت بی نقصی انجام خواهد شد، رفتار هوشیارانه فرد، بر اثر این جایگزینی تغییری نخواهد کرد. با این حال، به گفته جیولیو، تجربه ذهنی فرد از حالت هوشیاری به ناهوشیاری تبدیل خواهد شد. اختلاف نظر سوم در مورد نظریه اطلاعات یکپارچه شده این است که آیا یک موجودیت هوشیار می تواند از اجزایی ساخته شود که به طور جداگانه هوشیار باشند. به عنوان مثال، آیا جامعه به عنوان یک "کل" می تواند در حالی هوشیار باشد که افراد موجود در آن، هوشیاری خود را از دست ندهند و هوشیار بمانند. آیا یک مغز هوشیار می تواند دارای بخش هایی باشد که به تنهایی مستقل باشند؟ پیش بینی نظریه اطلاعات یکپارچه شده این است که چنین چیزی ممکن نیست، ولی همه از این بابت مطمئن نیستند. به عنوان مثال، در برخی از بیماران مبتلا به "سندروم دست بیگانه"<sup>۹۴</sup> که بر اثر یک ضایعه مغزی کنترل دستانشان را از دست داده اند، ارتباط میان تجارب ذهنی نیمه های مغزشان با یکدیگر کاهش می یابد، به طوریکه نیمکره راست مغز باعث می شود که دست چپ کارهایی را انجام دهد که بیمار آنها را بدون درک کردن انجام می دهد، تا جایی که بیماران از دست راست خود برای کنترل حرکات دست چپ استفاده می کنند. در این صورت، دو بخش هوشمند جداگانه در "سر" وجود خواهد داشت. یکی در نیمکره راست مغز که قادر به صحبت کردن نیست و دیگری در نیمکره چپ مغز که به جای هر دو دست صحبت می کند. تصور کنید که با استفاده از فن آوری آینده بتوانیم پیوندی را برای ایجاد ارتباط مستقیم بین مغز انسان ها با یکدیگر ایجاد کنیم که مانند ارتباط بین بخش های داخل یک مغز از کارایی لازم برخوردار باشد. در این صورت، آیا لحظه ای فراخواهد رسید که هوشیاری دو نفر ناگهان ناپدید شود و طبق نظریه اطلاعات یکپارچه شده، توسط یک

---

<sup>۹۴</sup> Alien hand syndrom

هوشیاری یکپارچه واحد، جایگزین شود. اختلاف نظر جالب دیگر در مورد این است که آیا آزمایش های تجربی، میزان هوشیاری مان را کمتر تخمین می زنند یا خیر. همانطور که قبلا دیدیم، اگر چه احساسمان این است که نسبت به حجم عظیمی از اطلاعات در مورد رنگ ها، اشکال، اشیاء و هر چیز که در مقابل ماست، هوشیاری دیداری داریم، اما آزمایش ها نشان می دهند که متأسفانه فقط می توانیم بخش کوچکی از این اطلاعات را به خاطر بسپاریم. برخی از محققان سعی کرده اند که این مغایرت را بر طرف کنند و به همین دلیل، در پی پاسخ به این سوال هستند که آیا می توانیم تجربه ای ذهنی (هوشیاری بدون دسترسی) از چیزهایی داشته باشیم که به دلیل پیچیدگی قادر به جا شدن در حافظه کاری مان برای استفاده بعدی نیستند. به عنوان مثال، وقتی که ما کورتوجهی (نابینایی ادراکی) را به دلیل عدم توجه به شیئی که جلوی دیدمان است، تجربه می کنیم، این بدین معنا نیست که تجربه دیداری هوشیارانه ای از آن را نداریم، بلکه فقط بدین معناست که اطلاعات آن در حافظه کاری ما ذخیره نشده است.

### هوشیاری هوش مصنوعی چگونه به نظر می رسد؟

اگر قرار باشد که سیستم هوش مصنوعی در آینده هوشیار باشد، به لحاظ ذهنی چه چیزی را تجربه خواهد کرد؟ نظریه ای که پاسخگوی این سوال باشد، در حال حاضر وجود ندارد و ما مطمئن نیستیم که به لحاظ منطقی بتوانیم به طور کامل به این سوال پاسخ دهیم. این مثل این می ماند که به فردی که کور است، بخواهیم توضیح دهیم که رنگ قرمز چگونه به نظر می رسد؟ خوشبختانه، ناتوانی فعلی ما برای ارائه پاسخ کامل به این سوال، مانع از این نمی شود که پاسخ مختصری به آن دهیم. موجودات فضایی هوشمند که روی سیستم های حسی انسان ها مطالعه می کنند، احتمالا رنگ ها را کیفیات ذهنی ای می دانند که در ارتباط با هر نقطه موجود روی یک سطح دو بعدی (میدان دید) احساس می شود و صداها را کیفیات ذهنی ای می دانند که به لحاظ مکانی متمرکز نشده است و درد را کیفیت ذهنی ای می دانند که در ارتباط با بخش های مختلف بدن احساس می شود. آنها با کشف این موضوع که شبکه های چشم ما انسان ها دارای سه نوع سلول مخروطی حساس به نور است، تشخیص

می دهند که ما می توانیم سه رنگ اصلی را تشخیص دهیم و اینکه کیفیت ذهنی بقیه رنگ ها از ترکیب این سه رنگ اصلی حاصل می شوند. با اندازه گیری اینکه چقدر طول می کشد تا نوروها اطلاعات را به سراسر مغز بفرستند، ممکن است چنین نتیجه بگیرند که ما قادر به تجربه کردن بیش از ده فکر یا ادراک هوشیارانه در هر ثانیه نیستیم و اینکه وقتی فیلمی را با بیست و چهار فریم در ثانیه در تلویزیون تماشا می کنیم، آنها را نه به صورت دنباله ای از تصاویر ثابت، بلکه به صورت یک حرکت پیوسته تجربه می کنیم. با اندازه گیری اینکه آدرنالین با چه سرعتی در جریان خون آزاد می شود و قبل از تجزیه شدن چقدر دوام پیدا می کنند، می توانند پیش بینی کنند که احساس خشم ظرف چند ثانیه شروع می شود و چند دقیقه ادامه پیدا می کند. با استفاده از استدلال های مشابه مبتنی بر علم فیزیک می توان به صورت علمی حدس زد که جنبه های خاص هوشیاری مصنوعی چگونه به نظر خواهد رسید. قبل از هر چیز، حوزه تجارب ممکن هوش مصنوعی در مقایسه با تجارب انسانی وسیع و گسترده می باشد. ما برای هر یک از احساسات خود، دارای یک گروه از کیفیات ذهنی هستیم، در حالی که هوش مصنوعی دارای انواع بسیار بیشتری از احساسات و بازنمایی درونی اطلاعات می باشد. بنابراین، باید مراقب باشیم تا در دام این تصور که احساسات هوش مصنوعی شبیه انسان هاست، گرفتار نشویم. هوش مصنوعی ای که به اندازه مغز انسان است، در هر ثانیه می تواند میلیون ها برابر انسان ها تجربه ذهنی داشته باشد. زیرا سیگنال های الکترومغناطیسی می توانند با سرعت نور که میلیون ها برابر سرعت سیگنال های نوروها می باشد، حرکت کنند. با این حال، هرچه هوش مصنوعی بزرگتر باشد، باید اندیشه های جهانی آن کندتر حرکت کنند تا زمان کافی برای گردش اطلاعات بین تمام بخش های آن فراهم شود. به همین دلیل، انتظار داریم که هوش مصنوعی دارای اندازه زمین مانند یک انسان، فقط ده تجربه هوشیارانه در هر ثانیه داشته باشد و هوش مصنوعی دارای اندازه کهکشان فقط یک اندیشه جهانی در هر ۱۰۰۰۰۰ سال داشته باشد. این امر باعث می شود که هوش های مصنوعی بزرگ تشویق شوند که عملیات محاسبه را به کوچکترین زیر سیستم هایی که قادر به این کار هستند، واگذار کنند تا کارها با سرعت بیشتری انجام شوند، درست مانند ذهن

هوشیارمان که واکنش غیر ارادی پلک زدن را به یک زیر سیستم کوچک، سریع و ناهوشیار واگذار کرده است. همانطور که گفتیم، با توجه به اینکه پردازش هوشیارانه اطلاعات در مغزمان فقط ذره ای از مشکلات فراوان و پنهان موجود است، می توان انتظار داشت که شرایط برای هوش مصنوعی های بزرگ آینده از این هم سخت تر باشد. اگر آنها دارای یک هوشیاری واحد باشند، احتمال می رود که تقریباً از تمام پردازش های اطلاعات اتفاق افتاده در داخل خود، آگاهی پیدا نکنند. علاوه بر این، هر چند که تجارب هوشیاری هوش مصنوعی بزرگ ممکن است بسیار پیچیده باشند، اما حرکت آنها در مقایسه با فعالیت های سریع بخش های کوچکترشان، با گام های بسیار آهسته تری صورت خواهد گرفت. در اینجا به اختلاف نظر قبلاً ذکر شده در مورد اینکه آیا بخش های یک موجودیت هوشیار نیز می توانند هوشیار باشند، می پردازیم. نظریه اطلاعات یکپارچه به این سوال پاسخ منفی می دهد، یعنی اینکه اگر یک هوش مصنوعی بسیار بزرگ در آینده هوشیار باشد، تقریباً تمام پردازش های اطلاعات آن به صورت ناهوشیار انجام می شود. این بدین معناست که اگر تمدنی از هوش های مصنوعی کوچکتر توانایی های ارتباطی خود را بهبود دهد تا جایی که یک ذهن هوشیار واحد ظاهر شود، هر یک از هوشیاری های تکی بسیار سریعتر آن ناگهان خاموش می شوند. از سوی دیگر، اگر پیش بینی نظریه اطلاعات یکپارچه نادرست باشد، ذهن هوشیار واحد می تواند با پوششی از ذهن های هوشیار کوچکتر همزیستی داشته باشد. در واقع می توان سلسله مراتب تودرتویی از هوشیاری در کلیه سطوح را از سطوح بسیار زیر گرفته تا سطوح کیهانی را برای آن متصور شد. همانطور که دیدید، به نظر می رسد که پردازش ناخودآگاه اطلاعات در مغز ما انسان ها با شیوه تفکر بی زحمت، سریع و خودکاری در ارتباط باشد که روان شناسان به آن، "سیستم ۱" می گویند. به عنوان مثال، سیستم ۱ ممکن است که با تحلیل بسیار پیچیده خود از داده های ورودی دیداری، به هوشیاری شما خبر دهد که دوستان از راه رسیده است، ولی به شما نمی گوید که محاسبات چگونه اتفاق افتاده است. در صورت اثبات معتبر بودن چنین پیوندی میان سیستم ها و هوشیاری، می توان این اصطلاحات را به هوش مصنوعی نیز تعمیم داد و کلیه وظایف معمولی سریع واگذار شده به واحدهای فرعی

ناخودآگاه را به عنوان سیستم ۱ هوش مصنوعی در نظر گرفت. به همین ترتیب، اندیشه جهانی دشوار، کند و کنترل شده هوش مصنوعی را در صورت هوشیار بودن می توان به عنوان سیستم ۲ هوش مصنوعی در نظر گرفت. ما انسان ها همچنین دارای تجارب آگاهانه ای هستیم که به آن سیستم صفر می گوئیم که عبارت از ادراک خام و منفعلانه ای است که حتی وقتی که ما بی حرکت هستیم و بدون فکر کردن، در جای خود نشسته ایم و فقط به جهان پیرامون خود نگاه می کنیم، اتفاق می افتد. سیستم دو از همه پیچیده تر است و فقط سیستم یک، ناهوشیار به نظر می رسد. نظریه اطلاعات یکپارچه شده توضیح می دهد که اطلاعات حسی خام در سیستم صفر، در ساختار شبکه ای شکل مغز با یکپارچگی بسیار بالا ذخیره می شوند و این در حالی است که سیستم دو به دلیل برخورداری از حلقه های بازخوردی، یکپارچگی بالایی دارد و در آن، تمام اطلاعاتی که هم اکنون از آنها آگاه می شوید، روی حالت های بعدی مغزتان تاثیر می گذارد. خلاصه اینکه، اگر نظریه ای پیدا شود که مشکل نسبتاً سخت هوشیاری را حل کند و مجموعه آزمون های تجربی دقیق را با موفقیت پشت سر گذارد، به طوری که بتوان پیش بینی های آن را جدی گرفت، گزینه های موجود برای حل این مشکل سخت تر که هوش مصنوعی های هوشیار در آینده چه چیزهایی را تجربه خواهند کرد، به شدت محدودتر خواهند شد. برخی از جنبه های تجربه ذهنی ما به ریشه های تکاملی ما انسان ها مانند خواسته های مربوط به صیانت نفس ( خوردن، آشامیدن، حفظ جان) و تولید مثل باز می گردند. این بدین معناست که می توان ماشین هوش مصنوعی ای را ایجاد کرد که هرگز کیفیات ذهنی ای مانند گرسنگی، تشنگی یا ترس را تجربه نکند، زیرا این نیازها مختص ما انسان ها می باشند. همانطور که در فصل قبل دیدیم، اگر یک ماشین هوش مصنوعی به شدت هوشمند طوری برنامه نویسی شود که هر هدف جاه طلبانه ای داشته باشد، ممکن است که برای دستیابی به آن هدف خود، خواسته هایی در مورد صیانت از نفس خود نیز داشته باشد. آنها در جمع ماشین های هوش مصنوعی ممکن است ترسی از مرگ نداشته باشند. تا زمانی که آنها نسخه بک آپ از خود داشته باشند، تنها چیزی که می ترسند از دست دهند، حافظه های انباشته شده و نرم افزار پشتیبانی از خودشان می باشد. علاوه بر این، امکان

کپی کردن اطلاعات و نرم افزار در بین ماشین های هوش مصنوعی احتمالا باعث تضعیف حس فردیت گرایی در آنها خواهد شد، زیرا اگر همه ما بتوانیم به آسانی اطلاعات موجود در حافظه خود را برای دیگران کپی کنیم یا به اشتراک بگذاریم، بین من و شما فرقی وجود نخواهد داشت و گروهی از ماشین های هوش مصنوعی مانند یک بدن با خرد جمعی مشترک به نظر خواهند رسید. آیا هوش مصنوعی احساس می کند که از اراده آزاد برخوردار است؟ هرچند که فلاسفه هزاران سال را صرف بحث بی نتیجه در مورد اراده آزاد ما انسان ها کرده اند و به اتفاق نظری در مورد تعریف آن هم نرسیده اند، اما من به شما می گویم که هر تصمیم گیرنده ای به طور ذهنی احساس خواهد کرد که اراده آزاد دارد. تصمیم گیری در این خصوص روی طیفی قرار می گیرد که دارای دو حد نهایی زیر می باشد:

۱- شما دقیقا می دانید که چرا گزینه خاصی را انتخاب کرده اید.

۲- شما نمی دانید که چرا گزینه خاصی را انتخاب کرده اید و به نظر می رسد که آن را به

صورت تصادفی و بدون دلیل خاصی انتخاب کرده اید.

وقتی که یک سیستم (مغز یا هوش مصنوعی) تصمیمی از نوع ۱ را می گیرد، تصمیم گیری خود را با استفاده از محاسبات مربوط به الگوریتم های قطعی انجام می دهد و حساب می کنید که چه کاری را انجام دهد. علاوه بر این، طبق یک نظریه مشهور علم رایانه که ست لیوید<sup>۹۵</sup> روی آن تاکید کرده است، تقریبا در تمام محاسبات، سریعترین راه این است که به جای اجرای واقعی محاسبه، نتیجه آن را تعیین کرد. این بدین معناست که تقریبا غیر ممکن است که تصمیم یک ثانیه بعد را در ظرف کمتر از یک ثانیه گرفت که این موضوع، تجربه برخورداری از اراده آزاد را تقویت می کند. در عوض، وقتی که یک سیستم (مغز یا هوش مصنوعی) تصمیمی از نوع ۲ را می گیرد، ذهن خود را طوری برنامه نویسی می کند که تصمیمش را بر مبنای خروجی یک سیستم فرعی تولید اعداد تصادفی انجام دهد. فارغ از اینکه تصمیم گیری شما روی کجای طیف ۱ تا ۲ واقع شده، هوشیاری هایتان احساس می

---

<sup>۹۵</sup> Seth lioyd

کنند که اراده آزاد دارند. آنها احساس می کنند که واقعا این خودشان هستند که تصمیم می گیرند و تا زمانی که جوانب مختلف را نسنجیده اند، با اطمینان نمی توانند بگویند که چه تصمیمی خواهند گرفت. برخی از افراد معتقدند که پیدا کردن رابطه علت و معلولی برای تصمیم گیری باعث می شود که فرایندهای فکری بی معنی شوند و مغز انسان ها به یک ماشین صرف تبدیل شوند. به نظر من چنین نگرش منفی ای، مزخرف و ناموجه می باشد. اول اینکه، به نظر من، مغز انسان پیچیده ترین شیء فیزیکی در عالم هستی می باشد. دوم اینکه، گزینه جایگزینی بهتری برای مغز انسان وجود ندارد و فرایندهای فکری ای که آنها برای تصمیم گیری استفاده می کنند، همان محاسبات اجرا شده در مغزشان است و تصمیم گیری را بر مبنای همین محاسبات درونی که فکر می کنند ناشی از اداره آزادشان می باشد، انجام می دهند و تا پایان تصمیم گیری از نتیجه این محاسبات آگاه نیستند.

### معنا بخشیدن به زندگی

در پایان، به نقطه شروع این کتاب باز می گردیم و آن این بود که می خواهیم آینده زندگی مان چگونه باشد؟ ما در فصل قبل دیدیم که چگونه تمام فرهنگ های متنوع در سراسر عالم هستی در جستجوی آینده ای هستند که پر از تجارب مثبت باشد. با این حال، هنگام جستجو برای یافتن توافقی در مورد اینکه چه تجربه ای مثبت است و چگونه می توان بده - بستانی را میان چیزهایی که برای شکل های مختلف زندگی خوب هستند، ایجاد کرد، اختلاف نظرهای مشکل سازی ایجاد می شوند. اما اجازه دهید که این اختلاف نظرها باعث نشوند که نتوانیم فیل را در اتاق تشخیص دهیم. اگر هیچ تجربه ای وجود نداشته باشد، یعنی هوشیاری ای وجود نداشته باشد، تجارب مثبت معنی ای نخواهند داشت. به عبارت دیگر، بدون هوشیار بودن، مفاهیمی همچون شادی، خوبی، زیبایی و هدف معنا پیدا نخواهند کرد. وقتی که افراد در مورد معنای زندگی می پرسند باید توجه داشته باشند که جهان هستی به زندگی انسان های هوشیار معنا نمی بخشد، بلکه این انسان های هوشیار هستند که به جهان هستی خود معنا می بخشند. بنابراین، اولین هدف ما در فهرست خواسته هایمان برای آینده باید این باشد که به جای نابود کردن جهان هستی، به حفظ و توسعه هوشیاری بیولوژیکی یا مصنوعی در آن کمک



کنیم. اگر در تلاش برای انجام این کار موفق شویم، چگونه می توانیم با ماشین هایی که از ما زرننگ تر هستند، همزیستی داشته باشیم؟ آیا قوی تر شدن بی امان هوش مصنوعی شما را اذیت نخواهد کرد؟ در فصل سوم دیدیم که تا زمانی که اراده سیاسی کافی وجود داشته باشد، فن آوری هوش مصنوعی قادر به ارضای نیازهای اولیه ما انسان ها مانند امنیت و کسب درآمد خواهد بود. با این حال، شاید توجه کرده باشید که تامین غذا، پوشاک، مسکن و سرگرمی برایمان کافی نیست. اگر هوش مصنوعی تضمین کند که در پی برطرف کردن کلیه نیازها و خواسته های مادی ما می باشد، در نهایت، ممکن است به این نتیجه برسیم که در زندگی خود، معنا و هدف خاصی نداریم و مانند حیواناتی هستیم که در باغ وحش به خوبی از ما مراقبت می شود. از زمان های گذشته، ما انسان ها ارزش شخصی خود را بر پایه ایده "استثناگرایی" بنا کرده ایم. بدین معنا که معتقد بوده ایم که ما زرننگترین موجود روی سیاره زمین هستیم و به همین دلیل، منحصر به فرد و برتر از بقیه هستیم. پیشرفت های صورت گرفته در زمینه هوش مصنوعی ما را مجبور خواهد کرد که این تفکر را رها کنیم و متواضع تر شویم و این چیزی است که باید اتفاق بیفتد. چسبیدن به ایده اهانت آمیز برتری جویی نسبت به دیگران (اشخاص، گروه های قومی، گونه ها و غیره) مشکلات بسیار بدی را در گذشته ایجاد کرده بود و ایده ای است که امروزه در حال باطل شدن است. در واقع، استثناگرایی انسان ها نه تنها باعث ایجاد غم و غصه در گذشته شده است، بلکه برای پیشرفت در دنیای امروز غیر ضروری به نظر می رسد. اگر ما تمدن فرازمینی صلح طلبی را کشف کنیم که در زمینه علم، هنر و هر چیز مهم دیگری، به مراتب پیشرفته تر از ما می باشد، این امر به احتمال زیاد مانع از این نمی شود که آنها به تجربه معنا و هدف در زندگی خود ادامه دهند. در صورتی که ما تکبر را کنار بگذاریم، قادر به حفظ خانواده، دوستان و جوامع بزرگ تر و کلیه فعالیت هایی که به زندگی ما معنا و هدف می بخشند، خواهیم بود. در شرایطی که برای آینده برنامه ریزی می کنیم، باید نه تنها به معنای زندگی خود، بلکه به معنای عالم هستی مان نیز توجه داشته باشیم. دو فیزیکدان مورد علاقه من به نام های استیون وینبرگ و فریمن دایسون دارای دیدگاه های کاملاً مخالفی با یکدیگر هستند. وینبرگ که به

دلیل کار اساسی خود در زمینه ارائه مدل استاندارد برای فیزیک ذرات، برنده جایزه نوبل شده است، می گوید: "هرچه قدر که جهان هستی قابل فهم تر می شود، بی معنا تر به نظر می رسد." از سوی دیگر، دایسون بسیار خوش بین تر است و با وجود اینکه موافق است که جهان هستی بی معنا است، اما معتقد است که زندگی بشر باعث می شود که جهان هستی پر از معنا شود و اینکه اگر امکان زندگی در سراسر کیهان فراهم شود، آینده بسیار بهتری به وقوع خواهد پیوست. او در مقاله تاثیر گذار خود در سال ۱۹۷۹ گفته بود: "به زودی معلوم خواهد شد که جهان هستی از دیدگاه کدامیک از ما به حقیقت نزدیکتر است." اگر زندگی در جهان هستی منقرض شود یا جهان توسط هوش مصنوعی مرده متحرک ناهوشیار، تسخیر شود، از دیدگاه وینبرگ به شدت حمایت خواهد شد. از این لحاظ، می بینیم که با وجود تمرکزی که ما در این کتاب روی موضوع آینده هوش داریم، آینده هوشیاری از آن هم مهم تر است، زیرا زندگی وقتی معنا دارد که ما هوشیار باشیم. فلاسفه دوست دارند که میان عقل ( توانایی تفکر هوشمندانه) و شناخت حسی ( توانایی تجربه کیفیات ذهنی) تفاوت قائل شوند. ما انسان ها هویت خود را بر مبنای خردمند بودن ساخته ایم و فکر می کنیم از بقیه موجودات هوشمند تر هستیم. اما، وقتی که با ماشین های زرننگ تر از خود مواجه می شویم، بهتر است که متواضع باشیم و به جای " انسان خردمند " برچسب " انسان دارای درک و احساسات " را به خود بزنیم.

## خلاصه فصل:

- هیچ تعریف بی چون و چرایی برای " هوشیاری " وجود ندارد. با این حال، من از تعریف جامع و غیر انسان محورانه " تجربه ذهنی " برای " هوشیاری " استفاده می کنم.
- اینکه آیا هوش مصنوعی، " هوشیار " است، یکی از مشکل سازترین مسائل اخلاقی و فلسفی مطرح شده در زمینه هوش مصنوعی می باشد که سوالاتی از قبیل اینکه آیا هوش مصنوعی می تواند درد را احساس کند یا دارای حقوقی باشد را ایجاد می کند.
- مشکل درک هوش را نباید با سه سوال جداگانه در مورد هوشیاری اشتباه گرفت. سوال اول که نسبتاً سخت است در مورد پیش بینی این موضوع است که کدامیک از سیستم های

فیزیکی هوشیار هستند. سوال دوم که سخت تر است در مورد پیش بینی کیفیات ذهنی هوشیاری می باشد و سوال سوم که واقعا سخت است، این است که چرا یک چیز هوشیار است؟

- سوال نسبتا سخت هوشیاری، یک سوال علمی است، زیرا نظریه ای که بتواند پیش بینی کند که کدامیک از پردازش های مغز به صورت هوشیارانه انجام می شود، به لحاظ تجربی قابل آزمایش و ابطال پذیر است، در حالی که در حال حاضر، معلوم نیست که علم چگونه می تواند دو مشکل سخت تر دیگر را به طور کامل حل کند.
- آزمایش های تجربی علم اعصاب نشان می دهند که بسیاری از رفتارها به صورت ناخودآگاه انجام می شوند و بسیاری از بخش های مغز هوشیار نیستند و اینکه بیشتر تجارب هوشیارانه ما بیانگر خلاصه ای از پردازش های صورت گرفته بر روی مقادیر بسیار زیادی از اطلاعات ناخودآگاه هستند.
- تعمیم دادن پیش بینی های هوشیاری از مغز به ماشین ها به یک نظریه نیاز دارد. به نظر می رسد که هوشیاری نیاز به نوع خاصی از ذره یا میدان ندارد، بلکه به نوع خاصی از پردازش اطلاعات نیاز دارد که نسبتا خودمختار و یکپارچه باشد، به گونه ای که کل سیستم نسبتا خودمختار باشد، اما اجزای آن این گونه نباشند.
- هوشیاری ممکن است غیر فیزیکی به نظر برسد، زیرا از بستر خود مستقل می باشد. اگر هوشیاری به صورت شیوه درک اطلاعات در زمانی که اطلاعات به شیوه های پیچیده خاصی پردازش می شوند، تعریف شود، فقط ساختار پردازش اطلاعات مهم است، نه ساختار ماده ای که پردازش اطلاعات را انجام می دهد.
- اگر هوش مصنوعی امکان پذیر شود، فضای لازم برای تجارب ممکن هوش مصنوعی در مقایسه با آنچه که انسان ها می توانند تجربه کنند، وسیع تر خواهد شد و طیف وسیعی از کیفیات ذهنی و بازه های زمانی را در بر خواهد گرفت.

- از آنجایی که هیچ چیزی بدون هوشیاری معنا نخواهد داشت، این جهان هستی نیست که به موجودات هوشیار معنا می بخشد، بلکه این موجودات هوشیار هستند که به جهان هستی ما معنا می بخشند.
- اینکه ما انسان ها آماده تواضع در برابر ماشین های زرنکتر هوش مصنوعی هستیم، باعث می شود که با آنها کنار بیاییم و به جای تبدیل شدن به انسان های خردمند سعی کنیم به انسان هایی تبدیل شویم که بر خلاف ماشین ها از قوه درک و احساس برخوردار هستند.

### سخن آخر: داستان تیم اف ال آی

در پایان این کتاب و پس از بررسی موضوعاتی از جمله هوش، اهداف و معنا در زندگی، باید به این سوال پاسخ دهیم که چگونه می توان این ایده ها را در عمل اجرا کرد و اینکه برای بهتر ساختن هرچه بهتر آینده خود، چه کارهایی را به طور مشخص باید انجام دهیم؟ این همان سوالی است که هنگام بازگشت از کنفرانس آسیلومار که در مورد هوش مصنوعی برگزار شده بود، از خود می پرسیدم. پس اجازه دهید که افکار خود در این خصوص را در پایان کتاب با شما در میان گذارم. لازم به ذکر است که در این کنفرانس که به مدت چند روز در پورتوریکو برگزار شد، افراد مختلفی گرد هم جمع شده بودند که از میان آنها می توان به کارآفرینان، پیشگامان تحقیقات هوش مصنوعی از دانشگاه ها و شرکت هایی مانند دیپ مایند، گوگل، فیس بوک، اپل، آی بی ام، میکروسافت و بیدو، اقتصاددانان، دانشمندان علم حقوق، فلاسفه و متفکران دیگر اشاره کرد. نتایج این کنفرانس حتی انتظارات بالای من را بر طرف کرد و من احساس می کنم که در مورد آینده زندگی خوشبین تر شده ام.

### اف ال آی (موسسه آینده زندگی)<sup>۹۶</sup> متولد شد.

از زمانی که در سن چهارده سالگی چیزهایی را در مورد مسابقه تسلیحاتی هسته ای یاد گرفتم، نگران این شده ام که قدرت فن آوری ما سریعتر از توان عقلانی مان برای مدیریت آن، رشد کند. بنابراین،

تصمیم گرفتم که در کتاب خود با عنوان "جهان ریاضیاتی ما" فصلی را به بررسی این چالش اختصاص دهم. در سال جدید، با خود عهد بستم که دیگر به خود اجازه ندهم که بدون فکر و تأمل جدی در مورد کاری که شخصا می توانم انجام دهم، در مورد چیزی شکایت کنم و به مدت یک ماه بر سر عهد خود ماندم. تا اینکه پس از فکر زیاد و مشورت با همسر به این نتیجه رسیدیم که یک سازمان غیرانتفاعی تاسیس کنیم که تمرکز آن بر روی بهبود آینده زندگی از طریق نظارت بر فن آوری واقع شود. سپس، تصمیم گرفتیم که نام مثبتی را به آن اختصاص دهیم که در نهایت با نام جامع "موسسه آینده زندگی" به توافق رسیدیم و پس از صرف شام با دوست قدیمی خود، او را متقاعد کردیم که با ما هم قوا شود. دوست من یکی از داناترین و ایده آل ترین افرادی بود که می شناختم و به مدت یک دهه با من در یک سازمان غیر انتفاعی دیگر، کار کرده بود. از آنجایی که به موضوع آینده هوش مصنوعی زیاد فکر می کردم، به سراغ دمیس حسابیس در لندن رفتم و او مرا با مهربانی به دیدار اعضای هیات مدیره شرکت دیپ مایند دعوت کرد. وقتی که با آنها ملاقات کردم، از پیشرفت های آنها نسبت به دو سال پیش شگفت زده شدم و مشاهده چشم انداز وسیع اداری این شرکت که سرشار از ذهن های با استعدادی بود که به دنبال دستیابی به هدف حل مشکل هوش مصنوعی بودند، باعث شده بود که حس غریزی ام برای احتمال موفقیت در آینده تقویت شود. روز بعد با استفاده از اسکایپ با دوست دیگری صحبت کردم و در مورد چشم انداز موسسه آینده زندگی به او توضیح دادم. یک ساعت بعد، او تصمیم گرفت که به ما اعتماد کند و لذا بودجه حداکثر ۱۰۰۰۰۰ دلار در سال را برای راه اندازی این موسسه اختصاص داد. من به شدت تحت تاثیر اعتماد او قرار گرفتم و اینکه او در سال بعد به من گفت که این بهترین سرمایه گذاری او در عمرش بوده است، یک دنیا برایم ارزش داشت. پس از مشغولیت ذهنی طولانی مدت برای گذشته و آینده هوش مصنوعی، ناگهان احساس کردم که در حال قدم زدن از میان تجسمات فیزیکی افکار خود هستم. احساس کردم که توسط مجموعه ای خیالی از چیزهایی که نشان دهنده رشد دانش ما هستند، احاطه شده ام که این چیزها از لوکوموتیو استفسون گرفته تا فورد مدل تی و کاوشگر قمری به اندازه انسان آپولوی ۱۱

رپلیکا و همچنین از ماشین حساب تفاضلی بایج گرفته تا سخت افزارهای رایانه ای امروزی را شامل می شد. همچنین، تاریخچه درک ما انسان ها از ذهن خود که از آزمایش پای قورباغه گالوانی گرفته تا کشف نوروں ها، نوار مغزی و تصویربرداری تشدید مغناطیسی کارکردی را شامل می شد، در افکارم به نمایش گذاشته شده بود. من به ندرت گریه می کنم، اما در موقع برگشتن از شهر لندن و رسیدن به زیرگذر پر از عابر پیاده ایستگاه مترو، احساس کردم که گریه ام گرفته است. تمام این عابران پیاده بدون آگاهی از آنچه که به آن فکر می کردم، با خشنودی فراوان سرگرم زندگی خویش بودند. ما انسان ها برای اولین بار کشف کردیم که چگونه کارها را توسط ماشین ها انجام دهیم و با استفاده از آنها انرژی بادی، برقی و مکانیکی تولید کنیم. به تدریج و با کشف سلول های عصبی مرز میان بدن و ذهن محو شد و ما شروع به ساختن ماشین هایی کردیم که علاوه بر کارهای بدنی، کارهای ذهنی را نیز می توانستند بهتر از ما انجام دهند. آیا به موازات اینکه ما انسان ها هویت خود را کشف می کنیم، به ناچار خودمان را منسوخ کرده و از رده خارج خواهیم ساخت؟ در این صورت، چنین اتفاقی به طور شاعرانه ای غم انگیز خواهد بود. هر چند که فکر کردن به این موضوع مرا می ترساند، اما عزم راسخ من را برای پابندی به عهد خود در سال جدید تقویت می کرد. من احساس کردم که برای تکمیل تیم بنیانگذاران "موسسه آینده زندگی" که رهبری تیمی از داوطلبان جوان آرمانگرا را به عهده بگیرد، به یک شخص دیگر نیز نیاز داریم. به همین دلیل، در بین گزینه های مختلف، شخصی را انتخاب کردیم که یکی از فارغ التحصیلان با استعداد دانشگاه هاروارد و دارنده مدال نقره در المپیاد بین المللی ریاضی بود و دانشگاه سیتادل را برای تعداد زیادی از جوانان آرمانگرا که به دنبال ایفای نقش بزرگی در جهان و زندگی خود بودند، تاسیس کرده بود. به همین دلیل، من و همسرم، او را به منزل خود دعوت کردیم تا چشم انداز خود را برای او بازگو کنیم. به این ترتیب، موسسه آینده زندگی متولد شد.

**ماجرای کنفرانس پورتوریکو**

این آغاز ماجرای شگفت انگیزی بود که همچنان ادامه دارد. همانطور که در فصل اول گفته شد، ما در خانه خود، جلسات منظم طوفان فکری را با تعداد زیادی از دانشجویان آرمانگرا، استادان و سایر متفکرین داخلی برگزار کردیم که ایده های برتر آنها تبدیل به پروژه شدند. به موازات قدم های آهسته ای که برای تاسیس یک سازمان جدید (مانند شرکت، شورای عالی و استخدای و راه اندازی یک وبگاه جدید) برداشتیم، یک مراسم رونمایی را در مقابل جمعیت زیادی از حضار در سالن اجتماعات دانشگاه ام آی تی برگزار کردیم که در آن راجع به آینده فن آوری با کارشناسان پیشرو در زمینه آینده فن آوری بحث کردیم. ما بقیه سال را صرف رسیدگی به کنفرانس پورتوریکو کردیم که همانطور که در فصل اول بررسی شد، هدفش به کارگیری محققان پیشروی جهان در زمینه هوش مصنوعی بود تا برای تداوم بهره برداری از مزایای هوش مصنوعی، بحث کنند. هدف ما این بود که مکالمات مربوط به ایمنی هوش مصنوعی را از مکالمات نگران کننده به مکالمات موثر و کاربردی و از مشاجره در مورد نگرانی های خود، به توافق بر سر پروژه های تحقیقاتی عملی که برای حداکثر سازی احتمال دستیابی به یک نتیجه خوب می توانستند از همین حالا شروع شوند، تغییر جهت دهیم. برای آمادگی جهت انجام این پروژه ها، اقدام به جمع آوری ایده های تحقیقاتی امیدوار کننده ای جهت ایمنی هوش مصنوعی از سراسر جهان کردیم و به دنبال بازخوردهای جمعی در خصوص فهرست پروژه های ممکن گشتیم و به کمک گروهی از داوطلبان جوان سخت کوش، اولویت های تحقیقاتی را در قالب سندی که قرار بود در کنفرانس مورد بحث واقع شود، بیرون کشیدیم. ما امیدوار بودیم که با ایجاد وفاق عمومی بر سر این موضوع که تحقیقات با ارزش زیادی در مورد ایمنی هوش مصنوعی وجود دارد، افراد را تشویق کنیم که شروع به انجام چنین تحقیقاتی کنند. پیروزی نهایی در پروژه جاه طلبانه ما زمانی حاصل می شد که بتوانیم شخصی را متقاعد کنیم که منابع مالی این پروژه را تامین کند، زیرا تاکنون هیچ حمایت مالی ای را برای تامین بودجه چنین پروژه ای از موسسات دولتی دریافت نکرده بودیم. تا اینکه سرمایه گذاری به نام ایلون ماسک از راه رسید و گفت: " کتاب هوش مصنوعی فوق هوشمندانه نوشته شده توسط بوستروم ارزش خواندن را دارد. ما

باید فوق العاده مراقب هوش مصنوعی باشیم، زیرا هوش مصنوعی به لحاظ بالقوه از بمب اتمی نیز خطرناک تر است." چند هفته بعد، فرصت پیدا کردم تا تلفنی با او صحبت کنم و اگرچه کاملاً احساس عصبانیت می کردم و شوکه شده بودم، اما نتیجه عالی ای حاصل شد. او با پیوستن به هیأت مشاوره علمی موسسه آینده زندگی، حضور در کنفرانس ما و تامین مالی برنامه تحقیقاتی ایمنی هوش مصنوعی موافقت کرد. موافقت او باعث شد که تمام اعضای تیم ما در موسسه آینده زندگی به وجد بیایند و تلاش های خود را برای برگزاری یک کنفرانس باشکوه و شناسایی موضوعات تحقیقاتی امیدوار کننده و حمایت اجتماعی از آنها مضاعف کنند. من بالاخره فرصت پیدا کردم تا برای برنامه ریزی بیشتر با ایلون ملاقات خصوصی داشته باشم. تنها شدن با او در یک اتاق استراحت کوچک لحظاتی بعد از اینکه مانند یک ستاره راک، هزاران دانشجوی ام آی تی را شیفته خود کرده بود، بسیار عجیب به نظر می رسید. اما بعد از چند دقیقه فقط روی پروژه مشترکمان متمرکز شدم. من فوری عاشق او شدم. او پرتوی از صداقت را به سمت دیگران می افکند و توجه فراوان او به آینده طولانی مدت بشریت و چگونگی تبدیل شجاعانه آرزوی او به عمل، برایم الهام بخش بود. او می خواست که عالم هستی را کشف کند و در آن سکنی گزیند و برای این منظور، یک شرکت فضایی تاسیس کرد. او می خواست که انرژی پایدار داشته باشد و به همین دلیل، یک شرکت بهره برداری از انرژی خورشیدی و یک شرکت تولید خودروی برقی را تاسیس کرد. به راحتی می توان فهمید که چرا افراد به حرف های او گوش می کردند، زیرا او بلندقد، خوش چهره، فصیح و دارای معلومات زیادی بود. متأسفانه، این رویداد برگزار شده در دانشگاه ام آی تی به من آموخت که رسانه ها چقدر می توانند ایجاد نگرانی کنند و تفرقه انگیز باشند. ایفای نقش ایلون در صحنه شامل بحث جالب یک ساعته در مورد کشف فضا می شد. در پایان جلسه، یکی از دانشجویان، سوالی را در مورد هوش مصنوعی از من پرسید که بی ربط به موضوع اصلی صحبت بود. پاسخ او شامل این جمله می شد: "ما با هوش مصنوعی در حال احضار کردن شیطان هستیم." این جمله بی ربط، تنها چیزی بود که بیشتر رسانه ها گزارش کردند. آنچه که مرا به فکر واداشت این بود که بسیاری از روزنامه نگارها ناخواسته



کاری را انجام می دادند که دقیقا بر عکس آن چیزی بود که ما در کنفرانس پورتوریکو در پی دستیابی به آن بودیم. در حالی که ما می خواستیم با تاکید ورزیدن بر وجوه اشتراک، به یک وفاق جمعی برسیم، انگیزه رسانه ها این بود که روی اختلافات تاکید کنند. هرچه قدر که آنها اختلافات بیشتری را گزارش می کردند، درآمد تبلیغاتی بیشتری کسب می کردند. علاوه بر این، در حالی که ما می خواستیم به افراد کمک کنیم تا دور هم جمع شوند و عقاید خود را بیان کنند، با یکدیگر کنار بیایند و یکدیگر را بیشتر درک کنند، پوشش رسانه ها ناخواسته باعث می شد که افرادی که عقاید متفاوتی داشتند، یکدیگر را ناراحت و پریشان کنند و تنها با انتشار نقل قول های بی ربط و تحریک آمیز، به سوء تفاهمات دامن زنند. به همین دلیل، تصمیم گرفتیم که از ورود روزنامه نگارها به جلسه پورتوریکو جلوگیری کنیم و قاعده چاتم هاوس<sup>۹۷</sup> را برای شرکت کنندگان در جلسه اعمال کنیم. اگرچه کنفرانس پورتوریکو با موفقیت پایان یافت، اما بدون زحمت انجام نشد و به کار آماده سازی جدی ای نیاز داشت. به عنوان مثال، تلفن زدن به تعداد زیادی از محققان هوش مصنوعی برای شرکت در کنفرانس و جذب حضار دیگر برای شرکت در این کنفرانس، کار مشکلی بود. ایلون نگران بود که برنامه تحقیقاتی ایمنی هوش مصنوعی باعث ایجاد حس کاذب امنیت شود و محققان بی باک فقط به ظاهر در مورد امنیت صحبت کنند و وعده های خود را عملی نکنند. ما با وجود قطع شدن مداوم صدا، در مورد مزایای فراوان ترویج موضوع ایمنی هوش مصنوعی و دعوت از محققان هوش مصنوعی بیشتری برای کار کردن در این زمینه صحبت کردیم و ایلون پس از بررسی اسناد به ما قول داد که ظرف سه سال، بودجه پنج میلیون تا ده میلیون دلاری را برای حمایت از تحقیقاتمان اختصاص دهد. سال نو شروع خوبی را برای من و همسرم داشت و باعث شد که قبل از تشکیل جلسه استراحت مختصری داشته باشیم. کنفرانس نیز شروع خوبی داشت. اتفاق نظر قابل توجهی بر سر این موضوع که

---

قانونی برای محرمانه ماندن اطلاعات مبادله شده در طول جلسه و پنل های بحث و مناظره در مورد مسائل<sup>۹۷</sup>

بحث برانگیز است که به نام ستاد مرکزی مؤسسه سلطنتی امور بین المللی (چتم هاوس) لندن که این قاعده از آنجا نشأت گرفت نامگذاری شده است.

به محققین ایمنی هوش مصنوعی بیشتری نیاز است، حاصل شد و بر اساس داده های ورودی شرکت کنندگان حاضر در کنفرانس، سند اولویت های تحقیقاتی که به سختی روی آن کار کرده بودیم، بهبود پیدا کرد و نهایی شد. سپس، نامه سرگشاده حمایت از تحقیقات ایمنی هوش مصنوعی دست به دست گردانده شد و ما از اینکه تقریباً همه آن را امضا کردند، خوشحال بودیم. من و همسر در هتل اقامت خود جلسه شگفت انگیزی با ایلون داشتیم و همسر تحت تاثیر تواضع و راستگویی ایلون در زندگی شخصی اش و علاقه ای که به ما داشت، قرار گرفته بود. روز بعد، ما از مصاحبه خود با او در مورد ایمنی هوش مصنوعی و دلایل حمایتش از آن، فیلمبرداری کردیم و به نظر می رسید که همه چیز طبق برنامه پیش می رود. نقطه اوج کنفرانس که اعلامیه ایلون در مورد اعطای کمک مالی به تحقیقات ما بود، برای ساعت هفت صبح روز چهارم ژانویه سال ۲۰۱۵ زمان بندی شده بود و من شب قبل از کنفرانس به قدری هیجان زده بودم که نتوانستم بخوابم. درست پانزده دقیقه قبل از اینکه وارد جلسه ای بشویم که قرار بود اتفاق بیفتد، به مشکلی برخورد کردیم. دستیار ایلون مرا صدا زد و گفت که به نظر می رسد ایلون ممکن است نتواند اعلامیه خود را بیان کند. همسر گفت که تا به حال این قدر مرا نگران و ناامید ندیده بود. ایلون بالاخره از راه رسید و وقتی که آنجا نشسته بودم، می توانستم ثانیه های شماره معکوس برای شروع جلسه را بشنوم. او توضیح داد که آنها از زمان مقرر برای پرتاب موشک اسپیس ایکس به فضا دو روز عقب هستند و امیدوارند که برای اولین بار شاهد فرود موفقیت آمیز آن بر روی یک شناور بدون سرنشین باشند. به همین دلیل، تیم اسپیس ایکس نمی خواستند که اخبار همزمان پخش شده در رسانه ها در مورد ایلون، باعث حواس پرتی شان شوند و یکی از اعضای تیم با خونسردی می گفت که خود ایلون و جامعه هوش مصنوعی هم نمی خواهند توجه رسانه ها را به خود جلب کنند. ما چند دقیقه دیرتر به جلسه رسیدیم. اما نقشه ای داشتیم و می خواستیم صحبتی از میزان کمک مالی نکنیم تا اعلامیه ارزش خبری نداشته باشد و برای حفظ راز ایلون به مدت نه روز از قاعده چاتام هاوس استفاده کردیم تا در طی این مدت، موشک با موفقیت به ایستگاه فضایی برسد، البته اگر موشک در حین پرتاب منفجر می شد، به زمان بیشتری نیاز می داشتیم. سرانجام، شمارش

معکوس برای شروع اعلامیه به صفر رسید. کارشناسان هوش مصنوعی فوق بشری که تحت مدیریت من بودند، بر روی صندلی های خود نشسته بودند و قبل از اینکه تشویق شرکت کنندگان در جلسه تمام شود، از جای خود بلند شدند. این همان لحظه ای بود که منتظرش بودم. من گفتم که بسیار خوشحال هستم که اعضای جلسه به توافق بر سر این موضوع رسیده اند که برای استفاده مفید از هوش مصنوعی به تحقیقات بیشتری نیاز است و گفتم که دستورالعمل های تحقیقاتی واقعی ای وجود دارند که از همین حالا می توانیم روی آنها کار کنیم و اضافه کردم که به دلیل صحبت از ریسک های جدی ای که در این جلسه می شود، بهتر است روحیه شما را بالا ببرم و از شما دعوت کنم که در ضیافت شامی که برایتان تدارک دیده شده شرکت کنید. سپس، بلندگو را به ایلون ماسک دادم. همین که او بلندگو را از دستم گرفت، اعلام کرد که پول زیادی را برای انجام تحقیقات مربوط به ایمنی هوش مصنوعی اعطا خواهد کرد. پس از پایان کنفرانس، من و همسرم برای ملاقات والدین خود به سوئد و رومانی رفتیم و رویداد پرتاب موشک به فضا را در استکهلم به صورت زنده از اینترنت مشاهده کردیم. متأسفانه، تلاش موشک برای فرود موفقیت آمیز نبود و پانزده ماه طول کشید که فرود آن بر اقیانوس با موفقیت انجام شود. با این حال، کلیه ماهواره ها با موفقیت در مدار قرار گرفتند و ما هم موفق شدیم که منابع مالی لازم برای انجام پروژه "موسسه آینده زندگی" را کسب کنیم.

### **ترویج تحقیقات مربوط به ایمنی هوش مصنوعی**

یکی از اهداف کلیدی کنفرانس پورتوریکو، ترویج تحقیقات مربوط به ایمنی هوش مصنوعی بود و جالب بود که این کار در چند مرحله انجام می شد. اولین گام به خود جلسه مربوط می شد که در آن، بسیاری از محققان وقتی که می فهمیدند بخشی از جامعه همتایان خود هستند، احساس راحتی می کردند و درگیر موضوع مورد بحث می شدند. من به شدت تحت تاثیر تشویق های انجام شده توسط شرکت کنندگان جلسه قرار گرفتم. به عنوان مثال، بارت سلمان، استاد هوش مصنوعی دانشگاه کورنل به من می گفت: "صادقانه می گویم که هرگز جلسه ای علمی با این میزان از هیجان انگیزی

و سازماندهی و تحریک فکری، ندیده بودم." گام بعدی وقتی شروع شد که ایلون در پیام متنی خود نوشت: "توسعه دهندگان برتر هوش مصنوعی در جهان نامه های سرگشاده ای را امضا کردند و در آن خواستار انجام تحقیقات مربوط به ایمنی هوش مصنوعی شدند." از میان هشت هزار امضا کننده در صفحه ثبت نام، اسامی بسیاری از برجسته ترین تولید کننده های هوش مصنوعی در جهان نیز دیده می شد. دیگر نمی شد ادعا کرد که افرادی که نگران امنیت هوش مصنوعی هستند، نمی دانند که راجع به چه چیزی صحبت می کنند. این نامه سرگشاده توسط رسانه ها به سراسر جهان گزارش شد و باعث شد که ما شاکر باشیم که جلوی ورود روزنامه نگاران به کنفرانس را گرفته بودیم. با وجود اینکه جنجال برانگیزترین واژه موجود در نامه عبارت "خطر غیر منتظره" بود، اما، این نامه منجر به این شد که عناوین خبری تحریک آمیزی مانند اینکه: "ایلون ماسک و استفن هاو کینگ نامه سرگشاده ای را به امید جلوگیری از شورش ربات ها امضا کردند." منتشر شود. از میان صدها مقاله ای که ما مشاهده کردیم، در یکی از عنوان های تمسخر آمیز آنها نوشته شده بود: "ربات های انسان نما مجموعه های انسان ها را زیر پای خود، لگدمال می کنند و این نشان می دهد که فن آوری پیچیده و تحول آفرین، به وسیله ای برای برپایی نمایش جشن و شادی تبدیل خواهد شد." خوشبختانه، بسیاری از مقالات خبری دیگری نیز وجود داشتند که سنجیده و معقولانه بودند. البته ما با چالش دیگری نیز مواجه بودیم که عبارت از همگام شدن با سیل بزرگی از امضاهای جدید بود که برای حفظ اعتبارمان و رها شدن از شوخی های زننده باید به صورت دستی به آنها رسیدگی می کردیم و به همین دلیل، گروهی از افراد داوطلب را مسئول بررسی امضاها کردیم. گام ترویجی سوم، چهار روز بعد و هنگامی شروع شد که ایلون پیام داد که ده میلیون دلار را برای تحقیقات مربوط به ایمنی هوش مصنوعی اعطا خواهد کرد. یک هفته بعد، سایتی را در اینترنت راه اندازی کردیم که در آن، محققینی از سراسر جهان می توانستند برای تامین بودجه تحقیقاتمان، ثبت نام و با هم رقابت کنند. مسئول یک پروژه آزاد خیرخواهانه در کالیفرنیا که روی اعطای کمک های مالی بسیار تاثیرگذار تمرکز می کرد، از روی سخاوت قبول کرد که پول بیشتری از هدیه ایلون را به ما اعطا کند. با توجه

به جدید بودن این موضوع و کوتاهی موعد مقرر برای ثبت نام، ما نمی دانستیم که چه تعداد متقاضی خواهیم داشت. تعداد زیاد پاسخ دهندگان ما را شگفت زده کرد، به طوریکه حدود سیصد تیم از سراسر جهان حاضر بودند که حدود ۱۰۰ میلیون دلار به ما اعطا کنند. گروهی از استادان هوش مصنوعی و محققین دیگر، پیشنهادات ما را با دقت بررسی کردند و ۳۷ تیم برنده را برای اعطای کمک مالی به ما در ظرف سه سال، انتخاب کردند. وقتی که ما فهرست برندگان را اعلام کردیم، این اولین باری بود که پاسخ رسانه ها به فعالیت های ما نسبتاً متفاوت و خالی از تصویرهای ربات های کشنده بود. در پایان به خوبی درک شد که بحث ایمنی هوش مصنوعی یک گفتگوی پوچ نیست و اینکه کارهای عملی واقعی و مفیدی را می توان در این زمینه انجام داد و به این ترتیب، بسیاری از تیم های تحقیقاتی بزرگ آماده پیوستن به تلاش های ما شدند. گام ترویجی چهارم در طی دو سال بعد و با انتشار تعداد زیادی از نشریات فنی و کارگاه های آموزشی در خصوص ایمنی هوش مصنوعی در سراسر جهان اتفاق افتاد که باعث ترویج موضوعات مطرح شده در کنفرانس هوش مصنوعی شد. افراد سخت کوش طی سال های زیادی تلاش کردند که با کمک اعضای جامعه هوش مصنوعی درگیر موضوع تحقیقات ایمنی هوش مصنوعی شوند. اما موفقیت چندانی نداشتند. اما امروزه پیشرفت های زیادی در این زمینه حاصل شده است. بودجه بسیاری از این نشریه ها از طریق طرح کمک های مالی موسسه ما تامین می شود و ما در موسسه آینده زندگی تمام تلاش خود را کرده ایم تا به سازماندهی و تامین بودجه کارگاه های آموزشی کمک کنیم. با این حال، بخش فزاینده ای از آنها با سرمایه گذاری وقت و منابع خود محققان هوش مصنوعی انجام می شوند. در نتیجه، محققان می توانند چیزهای زیادی را در مورد تحقیقات ایمنی هوش مصنوعی از همکاران خود بیاموزند و این کار نه تنها مفید است، بلکه یک تفریح محسوب می شود و باعث می شود که آنها فکر خود را روی مسائل ریاضی و محاسباتی جالب توجه متمرکز کنند. البته، معادلات پیچیده برای همه مایه تفریح نیست. دو سال بعد از برگزاری کنفرانس پورتوریکو و قبل از برگزاری کنفرانس آسیلومار<sup>۹۸</sup>، یک

---

شهری در کالیفرنیا<sup>۹۸</sup>

کارگاه آموزشی فنی را تشکیل دادیم که در آن برندگان طرح کمک مالی به موسسه آینده زندگی می توانستند تحقیقات خود را با استفاده از اسلایدهایی نمایش دهند و در این اسلایدها افراد می توانستند نمادهای ریاضی را روی یک صفحه بزرگ مشاهده کنند که این باعث خستگی آنها می شد. رشد چشمگیر تحقیقات ایمنی هوش مصنوعی فقط به محیط های دانشجویی محدود نمی شود، بلکه شرکت هایی مانند آمازون، دیپماینند، فیسبوک، گوگل، آی بی ام و میکروسافت همکاری های صنعتی ای را برای استفاده مفید از هوش مصنوعی آغاز کرده اند. کمک های عمده مالی جدید برای تحقیقات مربوط به ایمنی هوش مصنوعی باعث انجام تحقیقات گسترده ای در شرکت های وابسته غیرانتفاعی بزرگ و جدیدی از جمله موسسه تحقیقات هوش ماشینی در برکلی، موسسه آینده بشریت در آکسفورد و مرکز مطالعات خطر نابودی (مربوط به چیزهایی که هستی انسان ها را به خطر می اندازند) در کمبریج انگلستان شده است. کمک های مالی ۱۰ میلیون دلاری و بیشتر، باعث شروع به کار فعالیت های مفید هوش مصنوعی در شرکت هایی از جمله مرکز لورهلوم برای آینده هوش مصنوعی در کمبریج، شرکت موقوفه کی اند ال گیتس در زمینه اخلاق و فن آوری های محاسباتی هوش مصنوعی در پترزبورگ و شرکت سرمایه گذاری در زمینه اخلاق و مدیریت هوش مصنوعی در میامی شده است. در آخر این را هم بگویم که ایلون ماسک با قول اعطای کمک مالی میلیارد دلاری برای راه اندازی یک شرکت غیر انتفاعی در سان فرانسیسکو که در زمینه استفاده مفید از هوش مصنوعی فعالیت خواهد کرد، با شرکت های کارآفرینی دیگر وارد همکاری شد. همگام با موج تحقیقات مربوط به ایمنی هوش مصنوعی، موجی از عقاید ابراز شده فردی و گروهی نیز به وقوع پیوست. همکارانی از بخش صنعت که در زمینه هوش مصنوعی کار می کردند، اصول زیربنایی خود را منتشر کردند و گزارش های طولانی ای که شامل فهرستی از پیشنهادات می شد، توسط دولت آمریکا، دانشگاه استنفورد و سازمان IEEE (بزرگترین سازمان جهان از لحاظ شغل های فنی) منتشر شد که شامل تعداد زیادی از گزارش های وضعیت از جاهای دیگر نیز می شد. ما مشتاق این بودیم که بحث های هدفمند صورت گرفته بین حاضرین کنفرانس آسیمولار را تسهیل بخشیم و در مورد

موضوعات مورد توافق این جمعیت دارای فرهنگ های گوناگون، چیزهایی را یاد بگیریم. یکی از همکاران ما وظیفه شجاعانه خواندن تمام اسناد دریافت شده و بیرون کشیدن تمام نظرات مطرح شده در آنها را به عهده گرفت. تیم ما (موسسه آینده زندگی) تلاش می کرد که نظرات مشابه را در کنار هم قرار دهد و درازگویی های اضافی اداری را به طور کامل حذف کند تا به فهرست منتخبی از اصول مختصر دست پیدا کند و برای تولید این فهرست منتخب، سعی می کرد که از عقاید منتشر نشده و در عین حال تاثیرگذاری که به صورت غیر رسمی در گفتگوها و جاهای دیگر، ابراز می شدند نیز استفاده کند. با این حال، این فهرست هنوز دارای ابهام ها و تناقض های زیادی بود و نیاز به تفسیرهای بیشتری داشت. به همین دلیل، یک ماه قبل از برگزاری کنفرانس، ما آن را با شرکت کنندگان در میان گذاشتیم و نظرات و پیشنهادات آنها را برای رسیدن به اصول جدید و بهبود یافته جمع آوری کردیم و به اصول به شدت اصلاح شده ای رسیدیم که قابل استفاده شدن در کنفرانس بودند. در کنفرانس آسیلومار شاهد بهبود این فهرست در طی دو گام بودیم. در گام اول، گروه های کوچکی در مورد اصول مورد علاقه خود بحث و گفتگو کردند و با دریافت بازخوردها به طور مفصل آن را بهسازی کردند و اصول جدیدی را ایجاد کردند و در گام بعد، از همه حضار خواسته شد که میزان حمایت خود از هر یک از نسخه های اصول مطرح شده را تعیین کنند. این فرایند جمعی هم جامع و هم خسته کننده بود و من و همسرم در تلاش برای جمع آوری به موقع اطلاعات لازم برای گام های بعدی، زمان خواب و غذا خوردن در کنفرانس را مختصر کردیم. اما این فرایند، هیجان انگیز هم بود. پس از بحث های مفصل، پر دردسر و گاهی ستیزه جویانه و طیف وسیعی از بازخوردهای دریافتی، از میزان بالای توافق حاصل شده میان حضار در مورد برخی از اصول که بیش از ۹۷ درصد بود، شگفت زده شدم. این توافق باعث شد که توقعاتمان در مورد اصول گنجانده شده در فهرست نهایی بالا رود. ما فقط اصولی را حفظ کردیم که توافق حاصل شده بر سر آنها بیش از ۹۰ درصد بود. هرچند که برخی از اصول مورد علاقه شخصی من در دقیقه آخر از قلم افتاد، اما این باعث شد که بیشتر شرکت کنندگان از تائید شدن اصول مورد علاقه خود، احساس خوبی پیدا کنند.

## اصول مربوط به هوش مصنوعی در کنفرانس آسیلومار

هوش مصنوعی تاکنون ابزارهای مفیدی را فراهم کرده که هر روز توسط مردم سراسر جهان استفاده می شوند. ادامه توسعه هوش مصنوعی که توسط اصول زیر هدایت می شود، فرصت های شگفت آوری را برای کمک به افراد و توانمند ساختن آنها در دهه ها و قرن های پیش رو فراهم خواهد ساخت.

### مسائل مربوط به تحقیقات

۱- **هدف تحقیقات:** هدف از تحقیقات هوش مصنوعی باید ایجاد هوش مفید باشد، نه هوش هدایت نشده.

۲- **بودجه تحقیقات:** سرمایه گذاری در زمینه هوش مصنوعی باید با تامین بودجه تحقیقات آن همراه شود تا از استفاده شدن مفید از آن مطمئن شویم. در این خصوص، باید به سوالات دشواری در زمینه های علم رایانه، اقتصاد، حقوق، اخلاقیات و مطالعات اجتماعی پاسخ دهیم. به عنوان مثال:

- چگونه می توانیم سیستم های هوش مصنوعی آینده را به شدت مقاوم سازیم تا آنچه که از آنها می خواهیم را بدون درست کار نکردن یا برگشتن به وضع قبلی، انجام دهند؟

- چگونه می توانیم در عین حفظ اهداف و منابع انسانی، از خود کار سازی برای افزایش رفاه انسان ها استفاده کنیم؟

- چگونه می توانیم سیستم های حقوقی را به نحوی به روز رسانی کنیم که کاراتر و منصفانه تر شوند، با پیشرفت های هوش مصنوعی همگام شوند و ریسک های مربوط به هوش مصنوعی را مدیریت کنند؟



- هوش مصنوعی باید با چه مجموعه از ارزش هایی همسو شود و از چه جایگاه حقوقی و اخلاقی ای باید برخوردار باشد؟

۳- **پیوند علم با سیاست:** لازم است که تبادل نظر سازنده و سالمی بین محققان هوش مصنوعی و سیاستگزاران برقرار شود.

۴- **فرهنگ رایج برای انجام تحقیقات:** لازم است که فرهنگ همکاری، اعتماد و شفافیت بین محققان و توسعه دهندگان هوش مصنوعی رواج پیدا کند.

۵- **اجتناب از رقابت:** تیم های توسعه دهنده سیستم های هوش مصنوعی باید همکاری فعالی را با یکدیگر داشته باشند و از دور زدن قوانین مربوط به استانداردهای ایمنی اجتناب کنند.

## اخلاقیات و ارزش ها

۶- **ایمنی:** سیستم های هوش مصنوعی باید در سراسر طول عمر کاری خود، از ایمنی و امنیت کافی برخوردار باشند و در جایی که تحقق چنین امنیتی امکان پذیر و شدنی است، آن را به اثبات برسانند.

۷- **شفافیت خرابی:** اگر یک سیستم هوش مصنوعی خسارتی وارد می کند، باید دلیل خرابی آن را معلوم کند.

۸- **شفافیت قضایی:** هر گونه مداخله یک سیستم خودمختار در تصمیم گیری های قضایی باید با ارائه یک توضیح قانع کننده انجام شود که توسط مقامات ذی صلاح انسانی قابل رسیدگی باشد.

۹- **مسئولیت پذیری:** طراحان و سازندگان سیستم های پیشرفته هوش مصنوعی در مورد پیامدهای اخلاقی استفاده از این سیستم و اقدامات خود، مسئول هستند و این پیامدها را باید مشخص کنند.

۱۰- **همسویی ارزش ها:** سیستم های به شدت خودمختار هوش مصنوعی باید به گونه ای طراحی شوند که بتوان مطمئن شد که اهداف و رفتارهایشان با ارزش های انسانی همسو می باشد.

۱۱- **ارزش های انسانی:** سیستم های هوش مصنوعی باید به گونه ای طراحی و ساخته شوند که با ایده آل هایی مانند شرافت انسانی، حقوق انسانی، آزادی و تنوع فرهنگی، سازگار باشند.

۱۲- **حریم شخصی:** افراد باید این حق را داشته باشند که دسترسی به داده های تولید شده خود را که توسط سیستم های هوش مصنوعی مورد تجزیه و تحلیل و بهره برداری واقع می شود، مدیریت و کنترل کنند.

۱۳- **آزادی و حریم خصوصی:** استفاده از هوش مصنوعی برای تجزیه و تحلیل داده های شخصی نباید بی دلیل باعث محدود شدن آزادی واقعی و محسوس افراد شود.

۱۴- **منافع عمومی:** فن آوری های هوش مصنوعی باید در جهت منافع و توانمندسازی تعداد هر چه بیشتری از افراد باشند.

۱۵- **رفاه عمومی:** رفاه اقتصادی ایجاد شده توسط هوش مصنوعی باید در جهت سودمندی برای تمام بشریت، گسترش پیدا کند.

۱۶- **کنترل انسانی:** برای دستیابی به اهداف انسانی، لازم است که انسان ها در مورد چگونگی واگذاری تصمیمات خود به سیستم های هوش مصنوعی حق انتخاب داشته باشند.

۱۷- **مقابله با خرابکاری:** از قدرت اعطا شده به واسطه کنترل سیستم های هوش مصنوعی به شدت پیشرفته باید در جهت تکریم و بهبود فرایندهای اجتماعی و مدنی که سلامت جامعه به آنها وابسته است، استفاده شود، نه در جهت تخریب آنها.

۱۸- **رقابت تسلیحاتی ایجاد شده توسط هوش مصنوعی:** باید از رقابت تسلیحاتی در زمینه توسعه سلاح های خودکار کشنده، جلوگیری به عمل آید.

## مسائل بلندمدت تر

۱۹- **احتیاط در مورد توانمندی هوش مصنوعی:** در جایی که هیچ توافقی وجود ندارد،

نباید حد بالایی را برای توانمندی های هوش مصنوعی در آینده تصور کنیم.

۲۰- **اهمیت:** هوش مصنوعی پیشرفته باید بیانگر تغییر عمیقی در تاریخ زندگی بشر بر روی زمین

باشد و باید با مراقبت و تامین منابع مورد نیاز، طراحی و مدیریت شود.

۲۱- **ریسک ها:** در مورد ریسک های مصیبت بار و دارای خطر نابودی ناشی از توسعه سیستم

های هوش مصنوعی، لازم است که برنامه ریزی و مدیریت ریسک متناسب با اثر مورد انتظار

را انجام دهیم.

۲۲- **خودبهبودی بازگشتی:** سیستم های هوش مصنوعی طراحی شده برای خودبهبودی که

منجر به افزایش سریع کیفیت و کمیت خواهد شد، باید در معرض اقدامات ایمنی و کنترلی

قرار گیرند.

۲۳- **خیر عمومی:** هوش مصنوعی فوق بشری باید در جهت خدمت رسانی به ایده آل های

اخلاقی عمومی گسترش پیدا کند و به نفع تمام بشریت باشد، نه یک کشور یا سازمان خاص.

پس از اینکه ما این اصول را به صورت آنلاین به اشتراک گذاشتیم، فهرست امضاها به طور

چشمگیری افزایش یافت و این فهرست تاکنون شامل امضای بیش از هزاران محقق هوش مصنوعی و

متفکران برجسته دیگر می باشد. ما نه فقط از میزان توافق حاصل شده در مورد اصول، بلکه از قدرت

این اصول نیز، شگفت زده شدیم. برخی از پیام ها مانند اینکه: "صلح، عشق و مادری با ارزش

هستند." این اصول را تأیید می کنند و برخی از پیام ها حاکی از نقض این اصول هستند. به عنوان

مثال، این پیام که: "توسعه هوش مصنوعی فوق بشری در آینده، غیر ممکن است." اصل ۱۹ را نقض

می کند یا این پیام که: "انجام تحقیقات برای کاهش خطر نابودی ناشی از توسعه هوش مصنوعی،

وقت تلف کردن است" اصل ۲۱ را نقض می کند. در واقع، بسیاری از همکاران ما در موسسه آینده

زندگی که در گروه مباحثاتی ما شرکت داشتند، با این موضوع که هوش مصنوعی فوق بشری احتمالا توسعه پیدا خواهد کرد و اینکه تحقیقات در زمینه امنیت هوش مصنوعی مهم است، موافقت کردند. من امیدوارم که اصول هوش مصنوعی مطرح شده در کنفرانس آسیلومار، نقطه شروعی برای بحث های مفصل تر باشد و در نهایت، منجر به توسعه راهبردها و سیاست های معقولی شود. در این راستا، مدیر رسانه ای موسسه آینده زندگی به همراه اعضای تیم خود اقدام به مصاحبه با محققان پیشرو در زمینه اصول هوش مصنوعی و تفسیر و ترجمه آنها زبان های مهم دنیا کردند.

### خوش بینی نسبت به آینده

همانطور که در ابتدای این فصل گفتم، در مورد آینده زندگی، خوش بینی بیشتری را نسبت به زمان های گذشته در خود احساس می کنم. در اینجا داستان شخصی خود را برایتان نقل می کنم. تجربه من در طی چندسال گذشته به دو دلیل جداگانه باعث افزایش خوش بینی ام شده بود. دلیل اول این بود که شاهد بودم که جامعه هوش مصنوعی به شکل جالب توجهی دور هم جمع شده بودند تا مسئولیت چالش های پیش رو را به شکل سازنده ای به عهده بگیرند و برای این منظور مایل بودند که با متفکرانی از حوزه های دیگر همکاری کنند. پس از جلسه آسیلومار، ایلان به من گفت که تعجب می کند که چگونه موضوع ایمنی هوش مصنوعی فقط در ظرف چند سال، از یک موضوع حاشیه ای به یک جریان فکری غالب تبدیل شده است و این برای خود من هم عجیب بوده است. اکنون فقط موضوعات کوتاه مدت مطرح شده در فصل سوم نیستند که به عنوان موضوع بحث، قابل احترام هستند و موضوعات مطرح شده در کنفرانس آسیلومار در مورد ماشین های هوش مصنوعی فوق بشری و خطر نابودی بشر توسط آنها نیز حائز اهمیت هستند. اگر این اصول در زمان دو سال پیش در کنفرانس پورتوریکو مطرح می شد، تنها واژه ای که در نامه سرگشاده، ترسناک به نظر می رسید، عبارت "خطر نهفته" بود. من از تماشای مردم لذت می برم. به همین دلیل، در صبح روز کنفرانس آسیلومار، کنار حضار ایستادم و به تماشای شرکت کنندگانی پرداختم که به بحث های مربوط به هوش مصنوعی گوش می دادند. با کمال تعجب، احساس محبت نامعلومی وجودم را فرا گرفت و

ناگهان منقلب شدم. این احساس با آنچه که در پورتوریکو تجربه کرده بودم، کاملاً فرق می کرد. به خاطر دارم که در کنفرانس پورتوریکو جمعیت هوش مصنوعی را با احساس ترس آمیخته با احترام نگاه می کردم. من به آنها به عنوان یک تیم حریف نگاه نمی کردم، بلکه به عنوان گروهی نگاه می کردم که باید آنها را قانع کنم. اما اکنون نگاهم این است که همه ما اعضای یک تیم واحد هستیم. همان طور که شما با خواندن این کتاب احتمالاً متوجه شده اید، من هنوز پاسخی برای چگونگی ایجاد یک آینده عالی با هوش مصنوعی ندارم. بنابراین، همه ما اعضای از یک جامعه رو به رشدی هستیم که به دنبال یافتن پاسخی برای سوالات خود هستیم. دلیل دوم برای خوش بینی من این است که تجربه موسسه آینده زندگی برایم دلگرم کننده بوده است. در گذشته چاره ای برای برخی از مسائل وجود نداشت و تصور می شد که آینده نگران کننده ای در حال وقوع است و کاری راجع به آن نمی توان کرد. اما با گذشت سه سال از آن زمان، دلتنگی اجباری من به تدریج بر طرف شد. حتی اگر گروهی از افراد طبقات پایین اجتماع به طور داوطلبانه بتوانند تاثیر مثبتی را بر مهمترین گفتگوی عصر ما به جای بگذارند، تصور کنید که از همکاری آنها چه کارهایی را می توانیم انجام دهیم. در گفتگوهای مطرح شده در آسیلومار صحبت از دو نوع خوش بینی می شد. نوع اول که بی قید و شرط است مانند انتظار مثبتی است که ما از طلوع خورشید در صبح فردا داریم. نوع دوم خوش بینی این است که اگر ما با دقت برنامه ریزی کنیم و به سختی تلاش کنیم، چیزهای خوبی در آینده اتفاق خواهند افتاد و خوش بینی من در مورد آینده زندگی، از این نوع بوده است. اکنون این سوال مطرح می شود که ما با ورود به عصر هوش مصنوعی، چه کاری را برای ایجاد تاثیر مثبت در آینده زندگی می توانیم انجام دهیم؟ به دلایلی که به زودی توضیح خواهم داد، فکر می کنم که گام بزرگ اول این است که اگر خوش بین نیستیم، روی خوش بینی خود نسبت به آینده کار کنیم. برای این منظور لازم است که چشم انداز مثبتی نسبت به آینده داشته باشیم. وقتی که دانشجویان دانشگاه ام آی تی برای مشاوره شغلی به دفتر من مراجعه می کنند، از آنها می پرسم که ظرف یک دهه بعد خود را در چه جایگاهی می بینند. اگر یکی از دانشجویان پاسخ دهد که: "پس از تصادف با یک اتوبوس در

قبرستان خواهم بود." از او به شدت انتقاد خواهم کرد. اینکه ما فقط یک آینده منفی را تصور کنیم، رویکرد بسیار بدی برای برنامه ریزی مسیر شغلی مان خواهد بود. اختصاص دادن ۱۰۰ درصد تلاش هایمان برای اجتناب از بیماری ها و تصادفات، فقط دستورالعمل خوبی برای افراد بدبین و خود بیمار انگار می باشد و به شادی منجر نمی شود. در عوض، من دوست دارم که دانشجویانم اهداف خود را با اشتیاق توصیف کنند و پس از تعیین اهداف خود، در مورد راهبردهای دستیابی به آنها و اجتناب از خطرهای پیش رو گفتگو کنند. اریک می گوید که طبق نظریه بازی ها، چشم اندازهای مثبت، مبنایی را برای بخش بزرگی از همکاری ها در جهان مانند ازدواج یا ادغام شرکت ها تشکیل می دهند. به هر حال، اگر نتوانید تصور کنید که سود بیشتری در نتیجه همکاری حاصل خواهد شد، چرا باید چیزی که دارید را قربانی کنید؟ این بدین معناست که ما باید آینده های مثبتی را نه تنها برای خودمان، بلکه برای جامعه و خود بشریت تصور کنیم. به عبارت دیگر، ما نیاز به امید بیشتری برای هستی خود داریم. با این حال، همانطور که همسر من به من خاطر نشان می کند، در فیلم های مانند فرانکشتاین یا ترمیناتور، چشم اندازهای آینده غالباً توصیف کننده ویران شهر هستند و این در حالی است که ما به عنوان یک جامعه باید برای آینده ای قوی برنامه ریزی کنیم و به همین دلیل، نیاز به خوش بینی بیشتری داریم و به همین خاطر است که در سراسر این کتاب شما را تشویق می کنم که به جای آینده ای که از آن می ترسید، به آینده ای که می خواهید، فکر کنید و برای رسیدن به اهداف آینده خود، برنامه ریزی و کار کنید. ما در سراسر این کتاب دیدیم که هوش مصنوعی چگونه می تواند هم فرصت های بزرگ و هم چالش های سختی را برایمان فراهم کند. یکی از راهبردهایی که به ما کمک می کند تا در مقابل چالش های واقعی هوش مصنوعی مقاومت کنیم، این است که سریعاً دست به کار شویم و قبل از اینکه هوش مصنوعی به طور کامل پیشرفت کند، جامعه انسانی خود را اصلاح کنیم و قبل از اینکه به نیروهای جوان قدرت زیادی را واگذار کنیم، به آنها آموزش دهیم که از فن آوری به شکل مفید استفاده کنند و قبل از اینکه فن آوری باعث منسوخ شدن قوانین شود، قوانین خود را به روز رسانی کنیم و قبل از اینکه مناقشات بین المللی به رقابت تسلیحاتی در زمینه توسعه سلاح های خود کار

منتهی شود، آنها را حل کنیم و قبل از اینکه هوش مصنوعی به صورت بالقوه باعث تشدید بی عدالتی ها شود، اقتصادی را ایجاد کنیم که رفاه را برای همه تضمین کند و باید تلاش کنیم تا جامعه ای را بسازیم که نتایج حاصل از تحقیقات ایمنی هوش مصنوعی به جای نادیده گرفته شدن، در آن اجرا شوند و برای مقابله با چالش های پیش روی ماشین هوش عمومی فوق بشری، بهتر است که قبل از یاد دادن استانداردهای اخلاقی به ماشین های قوی، با آنها در مورد برخی از استانداردهای اخلاقی اولیه، به توافق برسیم. در دنیای دو قطبی و نابسامان امروزی، افرادی که از قدرت استفاده از هوش مصنوعی برای اهداف بدخواهانه برخوردار هستند، انگیزه و توانایی بیشتری برای این کار دارند و تیم هایی که برای ساختن ماشین های هوش مصنوعی فوق بشری با یکدیگر در حال رقابت هستند، به جای همکاری، فشار بیشتری را برای دور زدن قوانین ایمنی در خود احساس می کنند. خلاصه اینکه، اگر ما بتوانیم جامعه بشری سازگارتری را ایجاد کنیم که به واسطه همکاری در جهت اهداف مشترک توصیف شود، چشم انداز ما که همان عاقبت به خیر شدن انقلاب هوش مصنوعی است، بهبود پیدا خواهد کرد. به عبارت دیگر، "بهبود فردا"، یکی از بهترین راه ها برای بهبود آینده زندگی می باشد. این کار را می توان به شیوه های مختلفی انجام داد. شما می توانید نظر خود را در مورد موضوعاتی مانند آموزش، حریم شخصی، سلاح های خودکار کشنده، بیکاری ناشی از توسعه فن آوری و مسائل دیگر، وارد صندوق رای کنید و به اطلاع سیاستمداران برسانید. شما همچنین در زندگی روزمره می توانید از قدرت اختیار خود برای انتخاب آنچه که می خواهید بخرید، اخباری که می خواهید گوش کنید، آنچه که می خواهید با دیگران به اشتراک بگذارید و الگویی که می خواهید برای دیگران باشید، استفاده کنید. آیا می خواهید شخصی باشید که صحبت دیگران را قطع می کند یا اینکه می خواهید شخصی باشید که با استفاده از فن آوری به شیوه برنامه ریزی شده و آگاهانه، احساس توانمندی می کند؟ آیا شما می خواهید که مالک فن آوری باشید یا می خواهید که فن آوری مالک شما باشد؟ نظرتان راجع به انسان ماندن در عصر هوش مصنوعی چیست؟ در مورد این موضوعات با اطرافیانتان گفتگو کنید. این مکالمات هم مهم و هم جذاب خواهند بود. ما انسان ها

از طریق شکل دادن به عصر هوش مصنوعی، از آینده زندگی محافظت می کنیم. اکنون احساس می کنم که هر چند که آینده به شکل اجتناب ناپذیری رخ خواهد داد، اما ما می توانیم آن را بهبود دهیم و تاثیر مثبت بر روی آن بگذاریم. آینده زندگی تغییر ناپذیر نیست و این ما هستیم که باید آن را ایجاد کنیم. پس اجازه دهید که با کمک یکدیگر، آینده امیدوار کننده ای را ایجاد کنیم.