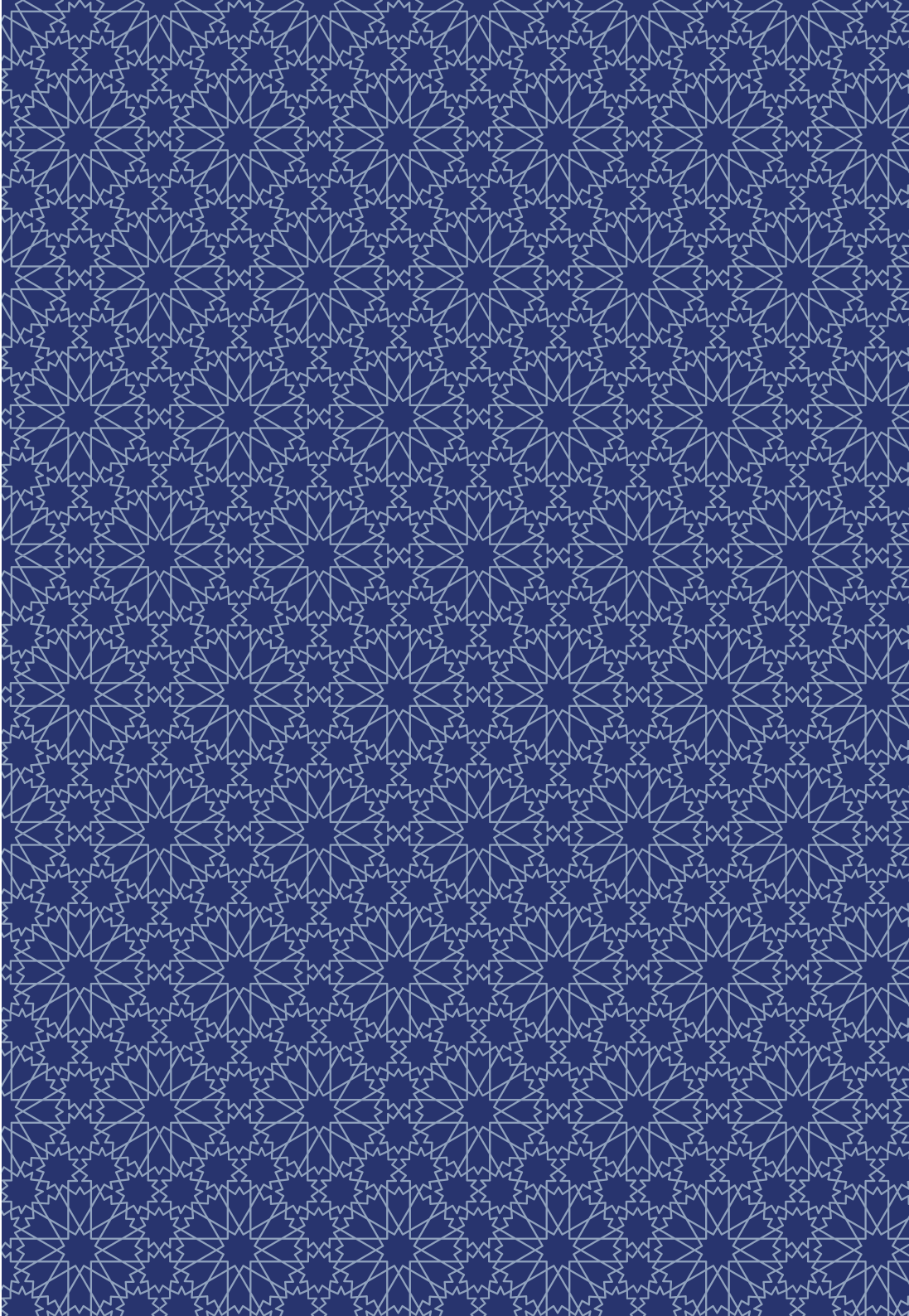


سُبْحَانَكَ اللَّهُمَّ رَبَّ السَّمَاوَاتِ السَّبْعِ وَرَبَّ الْعَرْشِ الْمَجِيدِ



الهيات هوش مصنوعي

مرکز مطالعات و پاسخگویی

به شبهات حوزه‌های علمیه

الهیات هوش مصنوعی

نویسندگان: جمعی از محققان

تهیه کننده: مرکز مطالعات و پاسخگویی به شبهات (حوزه های علمیه)

الکترونیک نشانی: www.pasokh.org

ناشر: انتشارات مرکز مدیریت حوزه های علمیه

شمارگان:

شاپک:



فهرست

مقدمه	۱۱
(مفهوم کاوی هوش مصنوعی)	۱۳
درآمد	۱۵
مفهوم کاوی هوش مصنوعی	۱۶
ریشه شناسی واژه «Artificial Intelligence»	۱۶
ریشه شناسی Artificial	۱۷
دوره نخست: تحول معنای «Artificial» در زبان لاتین	۱۷
دوره دوم: تحول Artificial در زبان فرانسوی و انگلیسی	۱۸
ریشه شناسی واژه Intelligence	۱۸
دوره نخست: تحولات Intelligence در زبان لاتین	۱۸
دوره دوم: شکل گیری Intelligence در زبان فرانسوی	۱۹
دوره سوم: شکل گیری Intelligence در زبان انگلیسی	۱۹
واژه شناسی Artificial Intelligence	۲۰
واژه شناسی Artificial	۲۰
واژه شناسی Intelligence	۲۱
معنای واژه شناسی Artificial Intelligence	۲۲

تعاریف هوش مصنوعی مبتنی بر رویکردها	۲۲
روش دسته‌بندی رویکردها	۲۳
تعاریف مبتنی بر رویکرد مدل شناختی	۲۴
تعاریف مبتنی بر رویکرد قواعد تفکر	۲۵
تعاریف مبتنی بر رویکرد عامل‌های عقلانی	۲۵
تعریف مختار مبتنی بر رویکرد حکمت اسلامی	۲۵
منابع و مآخذ	۲۷
تحلیل اخلاق هوش مصنوعی از منظر مبانی دین اسلام	۲۹
مقدمه	۳۱
معنای اخلاق	۳۹
معنای اخلاق هوش مصنوعی	۴۰
رویکردهای مختلف به اخلاق هوش مصنوعی	۴۱
۱. اخلاق هوش مصنوعی در رویکرد علوم شناختی	۴۲
هوش مصنوعی قوی و ضعیف	۴۳
نقد امکان اخلاق‌مندی هوش مصنوعی	۴۵
۲. اخلاق هوش مصنوعی در رویکرد فنی-مهندسی	۴۷
۳. اخلاق هوش مصنوعی در ارتباط با عملکرد انسان‌ها	۴۹
بررسی توصیه‌نامه اخلاق هوش مصنوعی یونسکو و سازگاری آن با مبانی اخلاق اسلامی	۵۰
تبیین توصیه‌نامه اخلاق هوش مصنوعی یونسکو	۵۰
تحلیل منشأ ارزش‌های اخلاق هوش مصنوعی یونسکو	۵۴
۱. کثرت‌گرایی اخلاقی:	۵۵
۲. نسبی‌گرایی اخلاق:	۵۸
اخلاق هوش مصنوعی بنا بر دین اسلام	۵۹
منابع	۶۳

آشنایی با هوش مصنوعی و نقاط تلاقی آن با الهیات	۶۷
مقدمه	۶۹
تعریف هوش مصنوعی	۶۹
بحثی پیرامون زیرشاخه‌های هوش مصنوعی	۷۲
۱. هوش مصنوعی تشخیص چهره و صحنه	۷۲
۲. رباتیک	۷۲
۳. یادگیری ماشین	۷۳
۴. سیستم‌های شناسایی گفتار یا گفت وگویی صوتی	۷۴
بحثی پیرامون هوش مصنوعی قوی	۷۵
بحثی پیرامون هوش مصنوعی و اجتهاد	۷۶
بحثی پیرامون هوش مصنوعی و الهیات	۷۷
هوش مصنوعی و دین	۷۹
مقدمه	۸۱
باورهای دینی	۸۲
نظریه سایبرنتیک	۸۳
هوش مصنوعی	۸۳
سایبرنتیک و دین	۸۴
یادگیری ماشین	۸۶
الهیات سنتی	۸۶
الهیات دوره رمانتیک و فرابشریت (ترابشریت)	۸۹
ترابشریت و هوش مصنوعی	۹۰
الهیات دوره رومانتیک و ماشین‌های شبه خدا	۹۲
منحصر به فرد بودن بشر در برابر هوش مصنوعی	۹۴

هوش مصنوعی قوی و امکان‌ش در مباحث الهیاتی	۹۷
مقدمه	۹۹
هوش مصنوعی قوی	۹۹
مفهوم «انواع ذهنی»	۱۰۰
دو مشکل عمده در کارکرد هوش مصنوعی قوی	۱۰۱
هوش مصنوعی ضعیف	۱۰۵
نتیجه	۱۰۶
هوش مصنوعی و پدیده آگاهی	۱۰۹
مقدمه	۱۱۱
اصطلاحات	۱۱۱
هوش مصنوعی و جایگزینی آن با هوش انسان	۱۱۳
نظریات ملاصدرا و مسئله جایگزینی هوش انسان	۱۱۵
الزامات هوش مصنوعی قوی با تأکید بر مواجهه دین اسلام با هوش مصنوعی قوی	۱۲۱
مقدمه	۱۲۳
انواع الزامات هوش مصنوعی	۱۲۳
۱. الزامات پیش از تولید هوش مصنوعی	۱۲۴
۲. الزامات هوش مصنوعی پس از تولید	۱۲۴
الزامات منطقی مواجهه دین اسلام با هوش مصنوعی	۱۲۵
الزامات کلامی و الهیاتی هوش مصنوعی	۱۲۶
الزامات شرعی هوش مصنوعی	۱۲۸
«هوش مصنوعی و دین در مسیحیت»	۱۳۱
مقدمه	۱۳۳
سنجش قدرت رایانه‌ها	۱۳۴

ماشین و انسان	۱۳۵
الهیات مسیحی و ماشین‌ها	۱۳۷
عواقب احتمالی هوش مصنوعی قوی	۱۴۷
هوش مصنوعی و خداپاوری	۱۴۹
فلسفه و هوش مصنوعی	۱۵۱
هوش مصنوعی و تکنولوژی	۱۵۲
هوش مصنوعی و فلسفه اخلاق	۱۵۳
انواع هوش مصنوعی از نظر تاریخی	۱۵۴
فلسفه هوش مصنوعی	۱۵۴
الهیات و هوش مصنوعی	۱۵۶
وظیفه الهیات هوش مصنوعی	۱۵۹
خداپاوری و هوش مصنوعی	۱۶۱
مقدمه	۱۶۳
عقاید خوش بینانه در مورد فناوری هوش مصنوعی	۱۶۳
اخلاق خداپاورانه و هوش مصنوعی	۱۶۵
هوش مصنوعی و پالایش در عصر اطلاعات	۱۶۷
چالش‌های حقوقی و اخلاقی ربات‌های کاربردی	۱۶۸
اولویت‌بندی اخلاقی در ربات‌ها	۱۷۳
نتیجه	۱۷۵





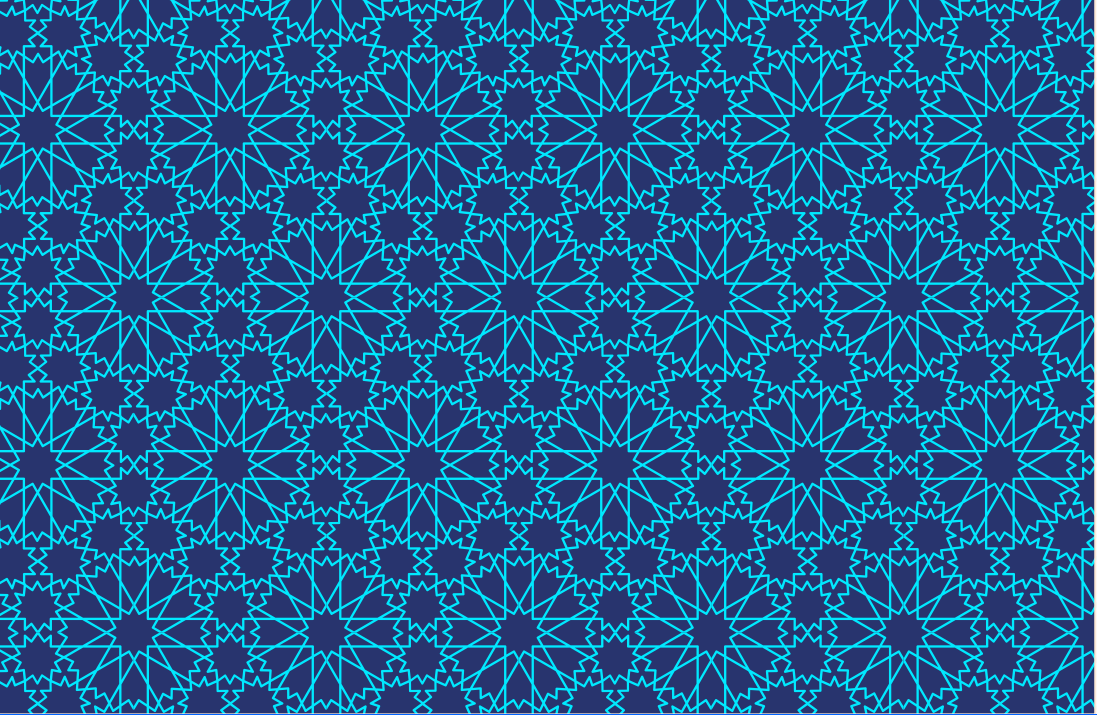
مقدمه

کتاب حاضر حاوی دو مقاله به عنوان مقدمه و مدخل و متن ۸ نشست علمی به محور هوش مصنوعی و الهیات است. در مقاله اول به قلم دکتر سخاوتیان به «مفهوم کاوی هوش مصنوعی» پرداخته شده است؛ در مقاله دوم که به قلم دکتر محسنی است «تحلیل اخلاق هوش مصنوعی از منظر مبانی دین اسلام» صورت گرفته است؛ سومین مطلب مربوط به متن سخنرانی آقای دکتر مینایی است با عنوان «آشنایی با هوش مصنوعی و نقاط تلاقی آن با الهیات»؛ مطلب چهارم توسط دکتر یوریک ویلکس ارائه شده و به موضوع «هوش مصنوعی و دین» پرداخته است؛ پنجمین مطلب این کتاب متن نشست علمی با عنوان «هوش مصنوعی قوی و امکانش در مباحث الهیاتی» است که توسط دکتر جی پی مورلند بیان شده است؛ دکتر منصوری سخنران موضوع «هوش مصنوعی و پدیده آگاهی» است که مطلب ششم این کتاب را به خود اختصاص داده است؛ ایشان در مطلب هفتم این کتاب نیز با عنوان «الزامات هوش مصنوعی قوی با تاکید بر مواجهه دین اسلام با هوش مصنوعی قوی» حضور یافته اند؛ متن هشتمین مطلب حاصل نشست علمی با دکتر تیموثی شتل است با عنوان «هوش مصنوعی و دین در مسیحیت»؛ دکتر قائمی نیا ارائه دهنده نهمین مطلب این کتاب با عنوان «خداباوری و هوش مصنوعی» هستند؛ آخرین مطلب کتاب با عنوان «هوش مصنوعی و خداباوری» حاصل نشست علمی دکتر جیمز مک گراث می باشد.

ضمن تشکر صمیمانه از نویسندگان و ارائه دهندگان در نشست های علمی
که حاصل زحماتشان در این کتاب منعکس شده است از جناب حجت الاسلام
والمسلمین محسنی که ارزیابی علمی این کتاب را نیز عهده دار شدند تشکر می
نماییم.

مرکز مطالعات و پاسخگویی به شبهات

حوزه های علمیه



«مفهوم کاوی هوش مصنوعی»

سید امیر سخاوتیان^۱





درآمد

هنگامی که سخن از تعریف هوش مصنوعی به مثابه مبدأ تصویری می‌شود، محقق به تعریف واژه هوش مصنوعی به عنوان یکی از اجزاء دانشی هوش مصنوعی می‌پردازد و حدود و ثغور آن را مشخص می‌نماید. پس از مشخص شدن مبدأ تصویری، نوبت به مبدأ تصدیقی هوش مصنوعی خواهد رسید؛ به عبارتی، تعریف هوش مصنوعی از طریق مبادی تصویری، و مجموعه قضایایی که مسائل دانش هوش مصنوعی بر آن متوقف است، تعریف با اتکا به مبدا تصدیقی نامیده می‌شود.

هنگامی که نام پدیده هوش مصنوعی شنیده می‌شود، نخستین چیزی که در ذهنها نقش می‌بندد موجودی جدید همانند ربات است. اگرچه این تصور دور از واقعیت نیست، همه واقعیت را هم در بر ندارد: هوش مصنوعی در جامعه انسانی پیامدهای فراوان اخلاقی، حقوقی، روان‌شناختی، جامعه‌شناختی، مدیریتی و... دارد که مواجهه فعال با ابعاد مختلف آن نیازمند شناخت دقیق مبانی فلسفی پیدایش این پدیده است. به عبارتی دیگر، هوش مصنوعی صرفاً یک ابزار نیست و فراتر از یک ابزار نوعی فناوری فعال در فرهنگسازی و مدیریت جامعه دارد. از سوی دیگر، هنگامی که سخن از هوش مصنوعی می‌شود، در ذهن برخی از مردم، موجوداتی رباتیک نقش می‌بندد که در آینده مدیریت همه جهان انسانی را برعهده

می‌گیرند و بر مقدرات زندگی انسانها مسلط می‌گردند. همه این تصورات ریشه در نوع نگاه به این پدیده دارد. بنابراین، قبل از آنکه بتوان درباره هوش مصنوعی قضاوت کرد، باید آن را به درستی شناخت. باید دانست که شناخت آن نیز متوقف بر شناخت ابعاد بنیادین آن، یعنی فلسفه هوش مصنوعی، است که در این ادامه به آن پرداخته خواهد شد.

مفهوم کاوی هوش مصنوعی

در بازه زمانی قرن بیستم و بیست و یکم، پدیده‌های در جهان دانش و فناوری ظهور کرد که از آن در زبان فارسی به «هوش مصنوعی»، در زبان انگلیسی به «Artificial Intelligence»، در چینی به «Advanced Intelligence» و در عربی به «الذكاء الاصطناعي» تعبیر می‌شود. اگرچه ریشه‌یابی هر یک از این واژگان و شناخت بستر فرهنگی آنها خالی از لطف نیست، به علت پرهیز از اطاله کلام و همچنین تمرکز مطالعه تطبیقی با منابع اولیه، که غالباً انگلیسی هستند، معنای انگلیسی آن از دو جهت ریشه‌شناسی^۱ و واژه‌شناسی^۲ مورد بررسی قرار می‌گیرند.

ریشه‌شناسی واژه «Artificial Intelligence»

برای ادراک معنای هوش مصنوعی باید توجه داشت که این فناوری از غرب آمده است. در زبان انگلیسی به آن «Artificial Intelligence» اطلاق می‌شود. پس در گام نخست باید سیر تحولات این واژه را در فرهنگ غرب شناخت. لازمه این شناخت نیز ارائه سیر تحولی آن در بستر تمدن غرب است. نخست باید دید که این ترکیب از چه واژگانی تشکیل شده و هر واژه نیز در بستر تمدنی غرب

۱. etymology

۲. Terminology

چه تحولات مفهومی داشته است: ^۱ ترکیب «Artificial Intelligence» از دو واژه Artificial و Intelligence تشکیل شده است.

ریشه‌شناسی Artificial

گام اول در شناخت مفهوم «Artificial Intelligence» ریشه‌شناسی ^۲ واژه «Artificial» است که از برخی دوره‌های مختلف زبان لاتین آغاز می‌شود و به زبان انگلیسی می‌رسد. این دوره‌ها عبارتند از:

دوره نخست: تحول معنای «Artificial» در زبان لاتین

واژه «Artificial» در زبان لاتین مرکب از «Ars» به معنای هنر ^۳ و «facere» به معنای ساختن ^۴ و انجام دادن ^۵ است. از ترکیب این دو کلمه واژه «artifex» به معنای صنعتگر و هنرمند ^۶ پدید می‌آید. مراد از هنرمند در اینجا کسی است که تولیدات صنعتی دارد. با گذشت زمان این واژه با تغییراتی به صورت «artificium» به معنای «ساخته شده توسط انسان» ^۷ و همچنین با تغییراتی به صورت «artificialis» یعنی «متعلق به هنر» به کار برده می‌شود. ^۹

۱. <http://file.fouladi.ir/courses/ai/videos/ai5%۰۱-BR5%۱۳۹۹۱۲۰۳D.mp۴>

۲. منبع این پژوهش در ریشه‌شناسی پایگاه «etym online» به آدرس زیر است:
<https://www.etymonline.com/word/intelligence>

۳. art

۴. to make

۵. do

۶. craftsman

۷. artist (genitive artificis)

۸. made by man (opposite of natural)

۹. <https://www.etymonline.com/word/intelligence>

دوره دوم: تحول Artificial در زبان فرانسوی و انگلیسی

این واژه در قرن چهاردهم می‌لادی به صورت «artificiel» وارد زبان فرانسه شد و معنای «ساخته شده»^۱ را می‌داد. تقریباً در همین زمان بود که این واژه به صورت «artificial» وارد زبان انگلیسی شده و معنایی مشابه معنای آن در زبان فرانسوی به خود گرفت.^۲

ریشه‌شناسی واژه Intelligence

پس از بررسی واژه «Artificial» به بررسی واژه «Intelligence» می‌پردازیم؛ بررسی این واژه نیز همانند واژه «Artificial» در دوره‌های متعدد رخ خواهد داد که عبارتند از:

دوره نخست: تحولات Intelligence در زبان لاتین

این واژه در زبان لاتین مرکب است از «intel» به معنای بین^۳ به اضافه «leger» به معنای انتخاب کردن^۴ و خواندن^۵ است. وقتی که Inter به اضافه leger شود به «Intelligere» تبدیل می‌شود که واژه‌ای لاتین به معنای درک کردن^۶ است. البته معنای تحت اللفظی این ترکیب «میان‌ه انتخاب» است و شاید ترجمه آن به «درک کردن» خالی از ارتباط نباشد؛ زیرا لازمه انتخاب درک کردن است. هم انتخاب و هم درک کردن برخاسته از نوعی هوشمندی است. شاید از اینجاست

۱. artificial

۲. <https://www.etymonline.com/word/intelligence>

۳. between

۴. choose, pick out

۵. read

۶. understand, comprehend

که نخستین گام‌ها برای تفسی این واژه به «هوش» آغاز می‌گردد؛ زیرا انتخاب مبتنی بر درک و سنجش است و درک و سنجش نیز نیازمند هوشمندی هستند. در ادامه سیر تحولی این ترکیب، واژه «Intelligentem» به معنای «هوشمندی اسمی»^۱ متولد می‌شود. این واژه از نظر مفهومی اخص از فهمیدن است و به فهمیدن خاص و شاخص اطلاق می‌گردد. در ادامه و در زبان لاتین واژه «Intellegentia» و «Intelligentia» به معنای درک، قدرت تشخیص،^۲ هنر،^۳ مهارت^۴ و سلیقه^۵ ایجاد می‌شود که نوعی ادراک، شعور و هوشمندی در هر یک از این معانی نهفته است.^۶

دوره دوم: شکل‌گیری Intelligence در زبان فرانسوی

در قرن دوازدهم، واژه «Intellegentia»، که یک واژه لاتین است، وارد زبان فرانسوی می‌شود و واژه «Intelligence» را ایجاد می‌کند که معنای آن «هوش» است.^۷

دوره سوم: شکل‌گیری Intelligence در زبان انگلیسی

واژه «Intelligence» در قرن چهاردهم می‌لادی از زبان فرانسوی وارد زبان انگلیسی می‌شود و به معنای «قابلیت و توانایی ادراک»^۸ به کار می‌رود.^۹

۱. nominative intelligens

۲. power of discerning

۳. art

۴. skill

۵. taste

۶. <https://www.etymonline.com/word/intelligence>

۷. <https://www.etymonline.com/word/intelligence>

۸. faculty of understanding

۹. <https://www.etymonline.com/word/intelligence>

واژه‌شناسی Artificial Intelligence

پس از فراغت از ریشه‌شناسی واژگان، اکنون به اصطلاح‌شناسی^۱ دو واژه «Artificial» و «Intelligence» می‌پردازیم که عبارتند از:

واژه‌شناسی Artificial

در پایگاه «merriam webster» در تعریف «Artificial» آمده است:

ساخته شده توسط انسان، اغلب بر اساس یک مدل طبیعی؛ مانند الماس مصنوعی و اندام مصنوعی.

الف: وجود در نظریه حقوقی، اقتصادی یا سیاسی؛ ب: ایجاد یا تولید توسط یک آژانس انسانی و به‌ویژه اجتماعی یا سیاسی، مزیت قیمتی مصنوعی در این شرکت‌ها، زنان واجد شرایط با موانع مصنوعی مواجه می‌شوند که مانع از پیشرفت آنها به سمت‌های بالای مدیریتی می‌شود. جیمز جی کیلپاتریک

الف: فاقد کیفیت طبیعی یا خود به خودی لبخند مصنوعی و هیجان مصنوعی؛

ب: تقلید، طعم مصنوعی و ساختگی

بر اساس خصوصیات مورفولوژیکی افتراقی که لزوماً نشان‌دهنده روابط طبیعی نیست؛ یک کلید مصنوعی برای شناسایی گیاه است.

منسوخ شده، باهوش، حيله‌گر

۱. منبع پژوهشی این نگارش در اصطلاح‌شناسی پایگاه merriam webster به آدرس زیر می‌باشد:

<https://www.merriam-webster.com/dictionary/intelligence>

واژه‌شناسی Intelligence

واژه «Intelligence» در زبان انگلیسی معانی مختلفی دارد که عبارتند از:

۱. الف: (یک): توانایی یادگیری یا درک یا مقابله با موقعیت‌های جدید یا آزمایشی؛ دلیل: استفاده ماهرانه از عقل. (دو): توانایی به‌کار بردن دانش برای دستکاری محیط خود یا تفکر انتزاعی که با معیارهای عینی (مانند آزمون‌ها) اندازه‌گیری می‌شود.

ب: علم مسیحی: کیفیت ابدی اساسی ذهن الهی؛

ج: تیزبینی ذهنی و زیرکی

۲. الف: اطلاعات، اخبار؛ ب: اطلاعات مربوط به یک دشمن یا دشمن احتمالی یا یک منطقه؛ نیز: آژانسی که درگیر به‌دست آوردن چنین اطلاعاتی است.

۳. عمل درک، درک

۴. توانایی انجام عملکردهای رایانه‌ای

۵. الف: ذهن‌های هوشمند یا هوش کیهانی، ذهن. ب: موجودی باهوش، به‌ویژه فرشته.

در دانش هوش مصنوعی معنای دوم یعنی اطلاعات، اخبار و مشتقات آن که در سازمانهای اطلاعاتی، همانند CIA و MI۶ کاربرد دارد، مراد نیست.^۱ اما سایر

۱. برخی معتقدند اطلاعات در بحث هوش مصنوعی مطرح است؛ در بحث داده‌ها و مثلث داده‌کاوی مسئله اطلاعات یکی از مراحل اصلی تحلیل داده است. در مباحث برخی گفتارهای موجود در این کتاب، مانند گفتار دکتر رحمانی، نیز به این مسئله اشاره شده است.

معانی به نوعی در تعریف پدیده هوش مصنوعی به کار برده می شود؛ زیرا هر یک از معانی فوق، به نوعی کارکرد این فناوری است. البته اگر توانایی اعجاب انگیز آن را در نظر بگیریم، به نوعی می توانیم معنای آخر، یعنی «فرشته»، را معنایی تمثیلی و تشبیهی برای آن بدانیم.

معنای واژه شناسی Artificial Intelligence

از ترکیب دو واژه «Artificial» به معنای ساخته شده به دست بشر و واژه «Intelligence»، که ظهور در نوعی هوشمندی دارد، در لغت می توان به یک معنایی نزدیک به «هوشمندی ساخته شده توسط بشر» رسید. اما تعریف دقیق اصطلاحی از این واژه در ارتباط با فناوری هوش مصنوعی هنوز به درستی بیان نشده است و تعاریف گوناگون از آن ارائه شده است. هر یک از این تعاریف ریشه در رویکرد محقق به هوش مصنوعی دارد. پس برخی از مهم ترین تعاریف مربوط به هوش مصنوعی در ذیل رویکردهایی ارائه خواهد شد.

تعاریف هوش مصنوعی مبتنی بر رویکردها

قبل از آنکه معنای اصطلاحی هوش مصنوعی برای ما روشن گردد، نیاز است رویکردها به پدیده هوش مصنوعی مشخص شود و به تبع آن تعاریف محققان این عرصه از آن ارائه شود؛ به عبارتی دیگر، دیدگاه فلسفی یک محقق نقش بسزایی در تعریف یک پدیده دارد. پس ارائه تعریف مختار از هوش مصنوعی، به شکل کامل در انتهای پژوهش خواهد آمد و در این بخش فقط به آن اشارهای کوتاه خواهد شد.

روش دسته‌بندی رویکردها

در هوش مصنوعی دو دسته مفاهیم وجود دارد:^۱ دسته اول، مفاهیم «تفکر-کنش» و دسته دوم مفاهیم «بشری-عقلانی». دسته اول، یعنی «تفکر-کنش»، یک محور را تشکیل می‌دهند که در یک سوی آن تفکر بشر و در سوی دیگر کنش و رفتار او قرار دارد. دسته دوم، یعنی «بشری-عقلانی»، نیز یک محور را نشان می‌دهد که در یکسو شباهت به توانمندیهای انسان و در سوی دیگر شبیه به معیارهای ایده‌آل هوشمندی است. پس می‌توان از ترکیب این دو محور، یک محور مختصات طراحی کرد و از آن به چهار دسته در زمینه هوش مصنوعی دست یافت که عبارتند از:

۱. سیستم‌هایی که همانند انسان فکر می‌کنند؛

۲. سیستم‌هایی که همانند انسان رفتار می‌کنند؛

۳. سیستم‌هایی که عاقلانه فکر می‌کنند؛

۴. سیستم‌هایی که عاقلانه رفتار می‌کنند.^۲

دسته اول، یعنی «سیستم‌هایی که همانند انسان عمل می‌کنند»،^۳ مبتنی بر رویکرد آزمون تورینگ هستند.^۴ دسته دوم، یعنی «سیستم‌هایی که همانند

۱. محققان مختلف، به‌گونه‌های گوناگون، تعاریف هوش مصنوعی را دسته‌بندی نموده‌اند. دسته‌بندی تعاریف در این پژوهش بر اساس کتاب «Artificial Intelligence. A Modern Approach» اثر راسل ونورویگ است.

۲. Stuart Russell, Peter Norvig - Artificial Intelligence. A Modern Approach, , ۲۰۰۳, P۱

۳. Acting humanly

۴. The Turing test approach

انسان فکر می‌کنند»^۱، مبتنی بر رویکرد مدل شناختی هستند.^۲ دسته سوم، یعنی «سیستم‌هایی که عاقلانه فکر می‌کنند»^۳، مبتنی بر رویکرد قواعد تفکر هستند.^۴ دسته آخر، یعنی «سیستم‌هایی که عاقلانه رفتار می‌کنند»^۵، مبتنی بر رویکرد عامل‌های عقلانی^۶ هستند.

بر اساس هر یک از این دسته‌ها و رویکردها تعاریفی پدید می‌آید که برخی از مهم‌ترین آنها عبارتند از:

تعاریف مبتنی بر رویکرد آزمون تورینگ

در این رویکرد، دو تعریف عمده وجود دارد^۷ که عبارتند از:

۱. انجام ماشینی کارهایی که انجام آنها نیازمند هوش انسانی است.^۸

۲. انجام ماشینی کارهایی که در حال حاضر توسط انسان به شکل بهتری انجام می‌پذیرد.^۹

تعاریف مبتنی بر رویکرد مدل شناختی

در این رویکرد نیز دو تعریف عمده ارائه شده که عبارتند از:

۱. Thinking humanly

۲. The cognitive modeling approach

۳. Thinking rationally

۴. The “laws of thought” approach

۵. Acting rationally

۶. The rational agent approach

۷. Stuart Russell, Peter Norvig - Artificial Intelligence_ A Modern Approach, ۲۰۰۳, ۲۲

۸. Kurzweil, ۱۹۹۰

۹. Rich and Knight, ۱۹۹۱

۱. تلاش بشر برای وادار کردن رایانه‌ها به تفکر همانند انسان و ساختن رایانه‌هایی که واقعاً برخوردار از ذهن باشند.^۱
۲. خودکار ساختن فعالیتهای همانند تصمیم‌گیری، حل مسئله و یادگیری که با تفکر انسان در ارتباط هستند.^۲

تعاریف مبتنی بر رویکرد قواعد تفکر

- در این رویکرد نیز دو تعریف عمده ارائه شده که عبارتند از:
۱. مطالعه قواعد ذهنی از مدل‌های محاسباتی^۳
 ۲. مطالعه محاسباتی که به توانایی ادراک و عمل می‌انجامد.^۴

تعاریف مبتنی بر رویکرد عامل‌های عقلانی

- در این رویکرد نیز دو تعریف عمده ارائه شده که عبارتند از:
۱. هوش محاسباتی منجر به طراحی عامل‌های هوشمند می‌گردد.^۵
 ۲. هوش مصنوعی در پی ایجاد رفتارهای هوشمند در موجودات مصنوعی است.^۶

تعریف مختار مبتنی بر رویکرد حکمت اسلامی

در حکمت اسلامی، آن چیزی که مُدْرک حقیقی است نفس و روح انسان

۱. Haugeland, ۱۹۸۵

۲. Bellman, ۱۹۸۷

۳. Chamiak and McDermott, ۱۹۸۵

۴. Winston, ۱۹۹۲

۵. Poole, et al., ۱۹۹۸

۶. Nilsson, ۱۹۹۸

است. از آنجایی که هوشمندی نیازمند ادراک است، پس هوش مصنوعی به تنهایی نمیتواند مدرک باشد، بلکه انسان هوشمند مدرک است که اصالتاً هوشمند بوده است و برای خود کمالاتی را تصور می‌کند و هوش مصنوعی در حقیقت دستیار انسان در رسیدن به آن کمالات تعریف شده است. به عبارتی دیگر، هوش مصنوعی دستیار انسان در رسیدن به کمالات تعریف شده به نحو کامل است.

ممکن است در جایی کمال در محاسبات ریاضی یا در نظارت پلیسی یا هر عملیات دیگر باشد. در همه افعال و نیازهای انسانی، کمالی وجود دارد که انسان این کمالات را درک می‌کند و با هوشمندی برای تسهیل در نیل به این کمالات، متناسب با شرایط خود، برای خود دستیارانی تعریف می‌کند. این دستیار انسان، که بالعرض و بالمجاز هوشمند است، هوش مصنوعی نام دارد و در حقیقت هوش مصنوعی مُعدّ کمال انسان است.

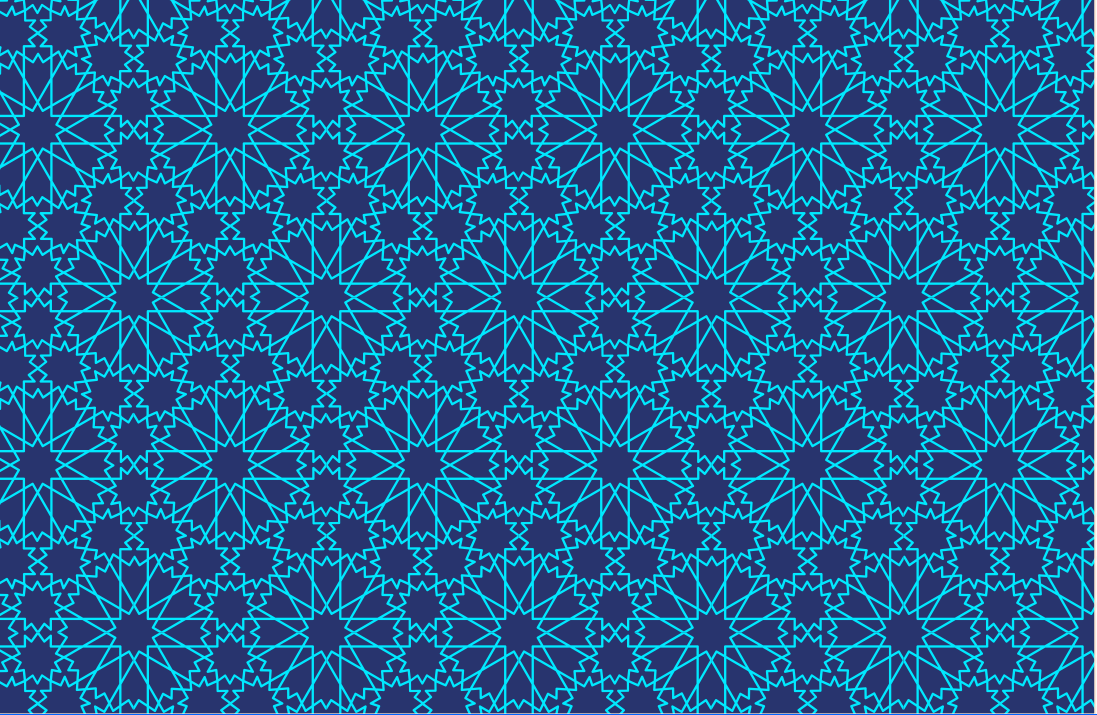
منابع و ماخذ

منابع مکتوب:

۱. Stuart Russell, Peter Norvig – Artificial Intelligence_ A Modern Approach, , Second Edition, ۲۰۰۳

پایگاه‌های اطلاع‌رسانی:

۲. <https://www.etymonline.com/word/intelligence>
۳. <https://www.merriam-webster.com/dictionary/intelligence>
۴. <http://file.fouladi.ir/courses/ai/videos/>



«تحلیل اخلاق هوش مصنوعی از منظر مبانی دین اسلام»

حمید محسنی^۱

۱. دکترای حکمت متعالیه از مؤسسه آموزشی و پژوهشی امام خمینی (ره): ۱۳۹۷، h.mohseni@mailfa.com





مقدمه

با پیشرفت روزافزون فناوری نوین هوش مصنوعی و تأثیر فراوان آن در همه جنبه‌های زندگی بشری، نگرانی‌ها و خطرات امنیتی، اقتصادی، حقوقی، فرهنگی و اخلاقی از این ناحیه احساس می‌شود. از این رو، برخی رهبران فناوری و کارشناسان هوش مصنوعی، همانند ایلان ماسک و استیو ورنیاک، در سال ۲۰۲۳ م با امضای نامه‌ای سرگشاده خواستار توقف موقت تحقیقات هوش مصنوعی شدند تا زمینه تدوین مقررات و قوانین فراهم شود. در همین سال، سام آلتمن مدیرعامل OpenAI در کنگره آمریکا شهادت داد و خواستار نظارت بر هوش مصنوعی شد.^۱ چند سال پیش، بیست هزار نفر از محققان این عرصه نامه‌ای را امضا کرده‌اند و به سیاست‌گذاران در مورد خطرات سلاح‌های خودمختار، که ممکن است از الگوریتم‌های برای حمله به اهداف خاص استفاده کنند، هشدار داده‌اند.^۲

افزون بر این، در حال حاضر تحقیقات ابتکاری زیادی در زمینه‌های اخلاقی، اجتماعی و حقوقی در زمینه هوش مصنوعی در سراسر دنیا و در سازمان‌های بین‌المللی و مراکز دانشگاهی و تحقیقاتی در حال انجام هستند؛ برای نمونه،

۱. <https://www.techtarget.com/searchenterpriseai/definition/AI-governance>

۲. Wang, Weiyu and Siau, Keng, "Artificial Intelligence: A Study on Governance, Policies, and Regulations" (۲۰۱۸), p:۳.

محققان هوش ماشینی مؤسسه (MIRI) هدف خود را «هم‌سوسازی هوش مصنوعی پیشرفته با منافع انسانی» عنوان می‌کنند.^۱ همچنین، مأموریت مرکز لوور هولم^۲ (آینده اطلاعات) در کمبریج ایجاد یک جامعه بین‌رشته‌ای جدید از محققان است که ارتباط‌های قوی بین فناوران و سیاست‌گذاران داشته باشند.^۳ دانشگاه استنفورد مؤسسه هوش مصنوعی انسان‌محور در سال ۲۰۱۹م تأسیس کرد که هدف اصلی آن پیشبرد تحقیقات، آموزش و سیاست‌گذاری هوش مصنوعی برای بهبود شرایط انسانی است. این مرکز در سال ۲۰۲۳م یک گزارش تحقیقاتی مفصل درباره شاخص هوش مصنوعی منتشر کرده است.^۴ از دیگر مؤسسات به موارد زیر می‌توان اشاره کرد:

مؤسسه آینده زندگی:^۵ این مؤسسه در سال ۲۰۱۵م تأسیس شد و مأموریت آن هدایت فناوری تحول‌آفرین به سمت سودمندی زندگی و دوری از خطرات شدید در مقیاس بزرگ است.^۶ رصدخانه هوش مصنوعی (OECD. AI) که خود را یک مرکز جهانی برای سیاست‌گذاری هوش مصنوعی معرفی کرده و با تبیین اصولی سعی دارد تا به کشورها کمک کند تا خطرات هوش مصنوعی را مدیریت کنند و از منافع آن بهره‌مند شوند.^۸ این رصدخانه بیش از ۷۰۰ ابتکار سیاست ملی هوش مصنوعی را از سوی ۶۰ کشور ثبت کرده است.^۹

۱. <https://intelligence.org/>

۲. Leverhulme

۳. <http://lcfi.ac.uk/>

۴. <https://hai.stanford.edu>

۵. Future of Life Institute

۶. <https://futureoflife.org/resource/ai-policy>

۷. OECD AI Policy Observatory

۸. <https://oecd.ai/en/wonk>

۹. Tallberg, Jonas, et al, (۲۰۲۳), The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research, p:۸-۷.

افزون بر این، به دلیل این نگرانی و خطرات اخیراً یونسکو به فکر افتاده که چارچوب‌های نظارت بین‌المللی و ملی را فراهم آورد، تا این فناوری در نهایت به نفع بشریت تمام شود. به همین دلیل، یونسکو اخیراً سندی با عنوان «توصیه‌نامه اخلاق هوش مصنوعی» تصویب کرده است.^۱ در این سند اخلاق هوش مصنوعی، اهداف و باید و نبایدهایی آورده شد و از کشورهای عضو، از جمله ایران، خواسته شده تا مفاد این سند را در کشورهای خود به نحوی اجرا کنند. نهادهای بین‌المللی به طور غیرمستقیم سعی دارند با تصویب این گونه اسناد بالادستی قوانین و ارزش‌های جوامع را در این زمینه با خودشان همراه سازند.

همان‌طور که تبیین شد در سطح دنیا سازمان‌های بین‌المللی و مراکز علمی و تحقیقاتی به طور گسترده‌ای در زمینه هدایت فناوری هوش مصنوعی تحقیق و پژوهش می‌کنند. از یک سو امروزه مسئله فناوری‌های نوین هوش مصنوعی همه جنبه‌های زندگی جوامع بشری را تحت تأثیر خود قرار داده و نگرانی‌های اخلاقی بسیار جدی ایجاد کرده است. از سوی دیگر، در غرب بر اساس دغدغه‌ها و نگرانی‌های خاص آنجا و مبتنی بر ارزش‌ها، اصول و مبانی فرهنگی و اخلاقی خود به دنبال حل مسائل اخلاقی از ناحیه هوش مصنوعی هستند. از این رو، هم بنا بر اهمیت و ضرورت توجه به تهدیدات اخلاقی از ناحیه هوش مصنوعی در سطح دنیا، و هم بنا بر ارزش‌ها، اصول و مبانی خاص دین اسلام می‌بایست در نظام جمهوری اسلامی ایران نیز در زمینه اخلاق هوش مصنوعی پژوهش‌های مستقلی انجام شود.

۱. <https://en.unesco.org/artificial-intelligence/ethics>, retrieved: ۲۰۲۲ Feb. ۶.

افزون بر این، توجه و تأمل در مطلب پیش رو هم اهمیت و ضرورت توجه به اخلاق هوش مصنوعی را نمایان می‌کند و هم وضعیت تحقیقات در حوزه اخلاق هوش مصنوعی در ایران و برخی کشورها مطرح دنیا را مشخص می‌نماید.

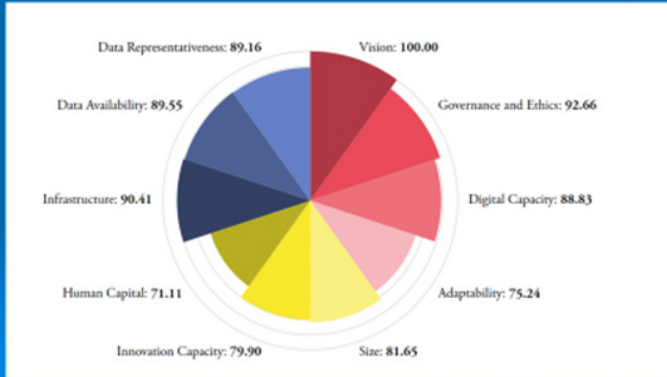
تارنمای چشم‌انداز آکسفورد شاخص‌های آمادگی کشورها در زمینه هوش مصنوعی را رصد کرده است. تصویر شماره ۱ شاخصه‌های آمادگی کشور آمریکا، تا سال ۲۰۲۰ م، در حوزه‌های مختلف هوش مصنوعی را نشان می‌دهد. به مناسبت بحث حاضر، یکی از این شاخصه‌ها در حوزه «اخلاق و حکمرانی»^۲ هوش مصنوعی است که آمریکا در این شاخصه حدود ۹۲ درصد آمادگی دارد. همچنین این کشور در شاخص مهم «چشم‌انداز»^۳ کلی در حوزه هوش مصنوعی ۱۰۰ درصد آمادگی را داراست. این نشانگر این واقعیت است که کشور آمریکا در پیشرفت هوش مصنوعی در همه جنبه‌ها برنامه‌ریزی جامع دارد.

۱. AI Readiness Index ۲۰۲۰ - US

۲. Governance and Ethics

۳. Vision

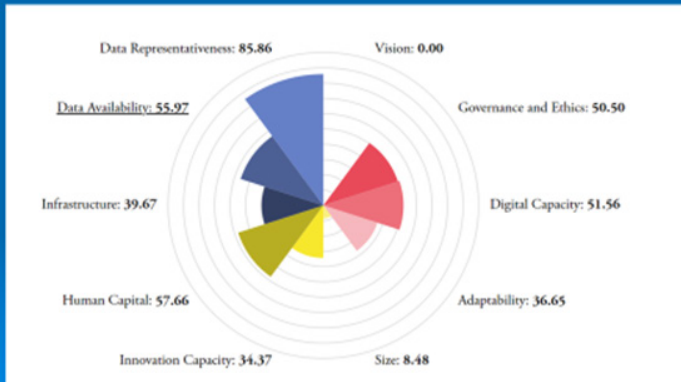
AI Readiness Index 2020 - US



تصویر شماره ۱: شاخص‌های آمادگی هوش مصنوعی در ایالات متحده آمریکا

تصویر شماره ۲: نشانگر شاخص‌های کشور ایران در حوزه مختلف هوش مصنوعی، تا سال ۲۰۲۰ م، است. این نمودار شاخص آمادگی ایران در حوزه «حکمرانی و اخلاق» هوش مصنوعی را حدود ۵۰ درصد نشان می‌دهد. سوگمندان در حوزه سند چشم‌انداز آمادگی ایران صفر درصد است!

AI Readiness Index 2020 - Iran



تصویر شماره ۲: شاخص‌های آمادگی هوش مصنوعی در ایران

البته اخیراً رهبر حکیم انقلاب اسلامی در حوزه هوش مصنوعی تأکید کرده‌اند: «پیشنهاد می‌کنم یکی از مسائلی که مورد تکیه و توجه و تعمیق واقع می‌شود، مسئله هوش مصنوعی باشد که در اداره آینده دنیا نقش خواهد داشت؛ ... باید کاری کنیم که حداقل به ده کشور اول دنیا در این مسئله برسیم»^۱. بعد از این بیانات، برخی نهادها، همچون شورای عالی انقلاب فرهنگی و معاونت علمی و فناوری ریاست جمهوری، به تکاپو افتادند تا برنامه‌ای در حوزه پیشرفت هوش مصنوعی تدوین کنند. صرف نظر از این، به نظر می‌رسد در جامعه ما بیشتر رویکرد تکنیکال و فنی-مهندسی هوش مصنوعی مورد توجه قرار گرفته است. به طور کلی دورویکرد بین اندیشمندان نسبت به مقوله هوش مصنوعی وجود دارد. تبیین این

۱. بیانات مقام معظم رهبری در دیدار جمعی از نخبگان و استعدادهای برتر علمی کشور در تاریخ ۱۴۰۰/۸/۲۶.

دو رویکرد می‌تواند در جهت‌گیری دقیق برای پیشرفت هوش مصنوعی در نظام اسلامی بسیار مفید باشد: اول، نگرش مهندسی - ابزاری به سیستم‌های هوش مصنوعی است. در این نگرش، جنبه‌های فنی - مهندسی و استفاده ابزاری از آن بیشتر مورد توجه است. دوم، رویکرد بنیادین، فلسفی، انسانی و تمدنی به مسئله هوش مصنوعی است. در رویکرد مهندسی، نگرش خوش‌بینانه به این فناوری وجود دارد. در این رویکرد، هرچند ممکن است به برخی مضرات و خطرات هوش مصنوعی نیز اشاره شود، بیشتر جنبه‌های فنی - مهندسی و استفاده از منافع و کاربردهای این فناوری مورد توجه است. اما در رویکرد بنیادین، با اعتراف به اینکه هوش مصنوعی منافع و مزایایی برای جوامع دارد، خطرات مهم اخلاقی فناوری هوش مصنوعی برای انسان‌ها هم مورد توجه قرار می‌گیرد. افزون بر این، در رویکرد بنیادین فناوری هوش مصنوعی دارای مبانی و مسائل فلسفی، سیاسی، فرهنگی و تمدنی بوده و می‌تواند در این حوزه‌ها نیز تأثیر بسزایی بگذارد.

به دلایل پیش‌گفته، می‌بایست نگرش حاکم جامعه اسلامی درباره هوش مصنوعی به‌طور دقیق روشن شود. باید بررسی شود که آیا دیدگاه رهبری در حوزه پیشرفت فناوری، به‌طور عام، و فناوری هوش مصنوعی، به‌طور خاص، مبتنی بر مبانی و اصول خاصی است. به بیان دیگر، در نگرش عمیق و ژرف می‌بایست بررسی کرد که آیا منظور رهبری از پیشرفت در حوزه هوش مصنوعی به‌طور مطلق و در هر شرایطی است، یا نه. از یک سو رهبری بر ضرورت پیشرفت هوش مصنوعی تأکید می‌کند و از سوی دیگر می‌بایست روشن شود پیشرفت مدنظر ایشان به‌طور مطلق بوده یا اینکه پیشرفت در زمینه فناوری هوش مصنوعی را مشروط به طیفی از مبانی و اصول کرده‌اند.

با توجه به تحقیقاتی که به‌طور مفصل و دقیق در اندیشه رهبری انجام شده

است، نگرش ایشان در قالب رویکرد دوم می‌گنجد. به بیان مختصر، با بررسی اندیشه‌های رهبری مشخص شد که همان‌طور که رهبر انقلاب به ضرورت پیشرفت در حوزه هوش مصنوعی تأکید دارند، به هدفمندی و جهت‌گیری صحیح این نوع فناوری‌ها نیز اهتمام دارند. از نظر رهبری، جهت‌گیری صحیح آن است که این نوع فناوری‌ها در راستای مباحث توحیدی و معارف الهی باشد و منجر به هدایت و رستگاری انسان‌ها شود.^۱ همچنین از منظر ایشان، پیشرفت علم و فناوری مبانی و اصول مهم دیگری دارد که باید در نظر گرفته شود؛ مبانی و اصولی که می‌بایست در تحقیق مستقلی بدان پرداخت.

از این رو، در این پژوهش سعی می‌شود که رویکرد دوم به فناوری هوش مصنوعی برجسته شود. با تبیین این مسئله مشخص خواهد شد که پیشرفت فناوری نباید به‌طور مطلق و بی‌قید باشد، بلکه پیشرفت این فناوری در نظام اسلامی باید هدفمند و بر اساس مبانی و ارزش‌های اخلاق اسلامی باشد. پس در تحقیق حاضر، برای اتخاذ رویکردی بنیادین و واقع‌بینانه، بعد از تعریف اخلاق، اخلاق هوش مصنوعی تبیین می‌شود و در ادامه رویکردهای مختلف به اخلاق هوش مصنوعی تبیین می‌شود و در ضمن آن سند اخلاق هوش مصنوعی یونسکو، بنا بر دیدگاه اندیشمندان اخلاق اسلامی و با توجه به مبانی دین اسلام، مورد کنکاش قرار می‌گیرد.

۱. برای توضیح بیشتر در این باره ن. ک: حمید محسنی و کاظم فولادی قلعه، «پیشرفت فناوری هوش مصنوعی مطلق یا هدفمند در اندیشه مقام معظم رهبری»، دومین کنفرانس فضای سایبر، دانشکده مهندسی دانشکدگان فارابی، دانشگاه تهران. در این مقاله، بخشی از اندیشه مقام معظم رهبری در حوزه پیشرفت هوش مصنوعی تبیین شده است.

معنای اخلاق

اصطلاح «اخلاق هوش مصنوعی»^۱ ممکن است در ابتدا برای برخی نامفهوم باشد. برای اینکه این اصطلاح تبیین شود، باید نخست مفهوم رایج اخلاق تبیین شود. مفهوم اخلاق در ارتباط با انسان در آثار علمای اسلامی و اندیشمندان مغرب‌زمین به‌طور مفصل و متعدد تبیین شده است.

علمای اخلاق اسلامی تعریف‌های متنوعی برای اخلاق ارائه داده‌اند. اخلاق جمع خُلُق است. در یک تعریف، خُلُق ملکه‌ای است که نفس انسان به مقتضی آن به سادگی فعلی را بدون احتیاج به فکر کردن انجام می‌دهد. ملکه با حال فرق می‌کند: حال کیفیت نفسانی است که به سرعت از بین می‌رود، ولی ملکه کیفیتی نفسانی است که به سختی از بین می‌رود. از این رو، اخلاق یا خلق کیفیتی نفسانی است.^۲ همان‌طور که صاحبان صنعت و اهل حرفه‌های مختلف به سهولت کار خود را انجام می‌دهند، صاحبان ملکات نفسانی خوب و بد نیز به سهولت کار خوب یا بد را انجام می‌دهند.

همچنین علمای معاصر اخلاق در تعریف آن گفته‌اند: دانشی است که از انواع صفات خوب و بد و چگونگی اکتساب صفات خوب و زدودن صفات بد بحث می‌کند. موضوع علم اخلاق عبارت است از صفات خوب و بد، از آن جهت که مرتبط با افعال اختیاری انسان هستند و برای اوقابل اکتساب یا اجتناب‌اند.^۳

در سوی دیگر، به باور برخی اندیشمندان علم اخلاق مغرب‌زمین ارزش‌های اخلاقی تنها ناظر به رفتار افراد است. از این رو، آنها علم اخلاق را بیان‌کننده ارزش

۱. Ethics of Artificial Intelligence

۲. خواجه نصیرالدین طوسی، اخلاق ناصری، ص ۶۳.

۳. محمد تقی، مصباح یزدی، فلسفه اخلاق، ص ۲۵.

و لزوم رعایت آن در رفتارها معرفی می‌کنند.^۱ در این صورت از منظر آنها، هدف اصلی عمل اخلاقی تصحیح خود رفتار است و ورای آن هدف دیگری دنبال نمی‌شود.^۲ بدین جهت، بسیاری از فلاسفه اخلاق غربی تمایل دارند سؤال در مورد رفتار انسانی را با چه چیزی خوب و چه چیزی بد است آغاز کنند و این را مبنایی برای ارائه تعریف مناسبی اخلاق می‌پذیرند.

اما در اندیشه علمای اخلاق اسلامی، همچون آیت‌الله محمدتقی مصباح یزدی، هم رفتار و هم صفات نفسانی از اهمیت خاص خود برخوردارند. پس در مباحث اخلاقی هم صفات اکتسابی و هم رفتارهای اختیاری می‌بایست مدنظر قرار گیرند؛ زیرا هم صفات درونی، مثل سخاوت و بخل، تواضع و کبر، متصف به خوبی و بدی و متعلق باید و نباید قرار می‌گیرند و هم رفتارهای بیرونی انسان قابل تحسین و تقبیح و متعلق باید و نباید هستند. همچنین، بیشتر صفات و ملکات اخلاقی از تکرار رفتارها به وجود می‌آیند؛ بنابراین، در اندیشه علمای اخلاق اسلامی، اخلاق در ارتباط با صفات و کیفیات نفسانی و درونی انسان معنا پیدا می‌کند.

معنای اخلاق هوش مصنوعی

آیا می‌توان معنای اخلاق در رابطه با انسان را به حوزه اخلاق هوش مصنوعی هم کشاند؟ به بیان دیگر، آیا هوش مصنوعی نیز می‌تواند فضایل اخلاقی و ملکات و صفات نفسانی را با انجام افعال اختیاری کسب کند؟ تبیین این مسئله نیازمند این است که بررسی شود که درحوزه اخلاق هوش مصنوعی و بنابر مبانی فلسفه

۱. فلسفه اخلاق (حکمت عملی)، ترجمه: ابوالقاسم پورحسینی، ص ۹.

۲. مجتبی مصباح، بنیاد اخلاق (روشی نو در آموزش فلسفه اخلاق)، ص ۲۶.

اسلامی و اخلاق اسلامی چه رویکردی را می‌توان در حوزه اخلاق هوش مصنوعی اتخاذ کرد. به بیان دیگر، باید در راستای تبیین چیستی اخلاق هوش مصنوعی نخست مسائلی طرح شود:

۱. چه رویکردهایی درباره امکان اخلاق‌مندی هوش مصنوعی وجود دارد؟
 ۲. آیا ربات‌ها و سیستم‌های هوشمند همانند انسان به‌طور مستقیم متصف به باید و نبایدهای اخلاقی می‌شوند؟ به بیان دیگر، ربات‌ها می‌توانند فعل اخلاقی را انجام دهند.
 ۳. آیا سیستم‌های هوشمند می‌توانند معنای خوب و بد اخلاقی را درک کنند؟ به بیان دیگر، هوش مصنوعی می‌تواند بفهمد که کدام فعل خوب است و کدام فعل بد. آیا می‌تواند بین آن دو، کار خوب را انتخاب کند و انجام دهد؟
 ۴. آیا هوش مصنوعی می‌تواند اخلاق‌مند باشد و دارای حالات اخلاقی و کیفیات نفسانی باشد؟
 ۵. رویکردهای هوش مصنوعی قوی و ضعیف چه تأثیری بر نگرش امکان اخلاق‌مندی سیستم‌های هوشمند خواهد گذاشت؟
- با بررسی این مسائل، معنای اخلاق هوش مصنوعی نیز آشکار خواهد شد. این امر نیازمند آن است که رویکردهای مختلف به اخلاق هوش مصنوعی تبیین شود.

رویکردهای مختلف به اخلاق هوش مصنوعی

برای تبیین مسائل فوق به‌طور کلی می‌توان سه رویکرد را در حوزه اخلاق هوش مصنوعی ترسیم کرد:

۱. اخلاق هوش مصنوعی در رویکرد علوم شناختی

در این رویکرد، هوش مصنوعی مخاطب باید و نباید اخلاقی قرار می‌گیرد. در این رویکرد، هوش مصنوعی توانایی درک و شناخت خوب و بد فعل اخلاقی را دارد. قوانین «سه‌گانه رباتیک» را می‌توان در این رویکرد گنجانده. در سال ۱۹۴۲م، ایازاک آسیموف، نویسنده روس، برای جلوگیری از آسیب‌رسانی ربات‌ها به انسان اقدام به تدوین سه قانون برای ربات‌های خود کرد که بعدها به قوانین رباتیک آسیموف مشهور شدند:

۱. یک ربات نباید با ارتکاب عملی یا خودداری از انجام عملی باعث آسیب دیدن یک انسان شود.

۲. یک ربات باید از فرامین انسان‌ها تبعیت کند، مگر این‌که آن فرامین در تعارض با قانون نخست باشد.

۳. تا هنگامی که قانون اول یا دوم زیر پا گذاشته نشده، ربات باید وجود خود را حفظ کرده و در بقای خود بکوشد.^۱

هرچند این قوانین مورد توجه قرار نگرفت، زمینه طرح چگونگی ایجاد ساختار اخلاقی برای عامل‌های مصنوعی گردید. از این رو، طرح‌هایی داده شد تا با به‌کارگیری الگوریتم‌هایی در ماشین یا با استفاده از ابزار شبکه‌های عصبی مصنوعی و مجموعه آموزشی، شامل حالت‌های مختلف اخلاقی ماشین برای آموزش سیستم، امکان نتیجه‌گیری در حالت‌های مختلف برای عامل‌ها فراهم شود. اما دیدگاه برخی در راستای این رویکرد بود: اینکه سیستم‌های هوشمند

۱. به نقل از: مجید رمضانی و محمد رضا فیضی درخشی، «اخلاق ماشین، چالش‌ها و رویکردهای مسائل اخلاقی

در هوش مصنوعی و ابرهوش»، ص ۳.

دارای حالت اخلاقی هستند.^۱ از این رو، می‌توان برخی ارزش‌های اخلاقی را در ماشین‌ها و ربات‌ها گنجانده.

لازمه امکان اخلاق‌مندی ربات‌ها و سیستم‌ها هوشمند و لازمه انجام افعال اخلاقی توسط آنها این است که هوش مصنوعی نسبت به تصمیم برای رفتار اخلاقی اختیار و اراده داشته باشد. همچنین، لازمه اینکه هوش مصنوعی، همانند انسان، تصمیم اخلاقی بگیرد و اختیار داشته باشد، این است که نسبت به آن فعل شناخت و آگاهی داشته باشد. بنا بر مبانی فلسفه اخلاق و فلسفه اسلامی، لازمه درک و شناخت و آگاهی و شعور ذهن‌مندی و داشتن نفس خودآگاه است. از این رو، این رویکرد رویکرد علوم شناختی نامیده شده است. آیا بنا بر مبانی فلسفه اخلاق اسلامی یک ربات یا یک سیستم هوشمند می‌تواند تصمیم، اختیار، آگاهی، ذهن و نفس داشته باشد؟ بررسی این مسائل مهم عقلی و فلسفی نیازمند آن است که دو رویکرد ضعیف و قوی به هوش مصنوعی، در علوم شناختی و فلسفه ذهن، مشخص شود. بعد از تبیین این دو رویکرد می‌توان اخلاق هوش مصنوعی در رویکرد علوم شناختی را واکاوی کرد.

هوش مصنوعی قوی و ضعیف

به نظر می‌رسد رویکردهای متفاوت در امکان اخلاق‌مندی هوش مصنوعی متأثر از تفاوت رویکرد دو مکتب فکری مختلف به مقوله هوش مصنوعی باشد: یکی مکتب رویکرد ضعیف هوش مصنوعی (Weak A) است. این مکتب فکری با مؤسسه فناوری ماساچوست مرتبط است. در این رویکرد تنها معیار این است که برنامه به‌کاررفته در AI به‌درستی انجام شود. دوم رویکرد هوش مصنوعی قوی

Strong AI) است که مکتب فکری دیگری به هوش مصنوعی است که در دانشگاه کارنگی ملون ظهور پیدا کرده است. در این مکتب، رفتار هوش مصنوعی با رفتار انسان سنجیده می‌شود؛ در حالی که در رویکرد ضعیف رفتار هوش مصنوعی با عملکرد سیستم سنجیده می‌شد. محققان در رویکرد قوی، به دنبال شبیه‌سازی سیستم ادراکی انسان در هوش مصنوعی هستند. افزون بر این، طرفداران هوش مصنوعی قوی بر این باورند که به دلیل دستیابی به اکتشافات، الگوریتم‌ها و دانش برنامه‌های نوین هوش مصنوعی می‌تواند دارای حس هوشیاری باشد.^۱

در رویکرد قوی، رفتار ماشین باید بر اساس عملکرد مشابه انسان سنجیده شود. به عنوان مثال، سیستمی را در نظر بگیرید که قادر به شنیدن است؛ طرفداران ضعیف هوش مصنوعی صرفاً نگران هستند که این شنیدن مطابق با عملکرد سیستم باشد. در حالی که طرفداران هوش مصنوعی قوی هدفشان دستیابی به موفقیت دیگری است: شبیه‌سازی سیستم شنوایی انسان با حلزون، کانال شنوایی، پرده گوش و سایر قسمت‌های گوش که هرکدام وظایف خود را در سیستم انجام می‌دهند. طرفداران ضعیف هوش مصنوعی موفقیت سیستم‌هایی را که می‌سازند تنها بر اساس عملکردشان می‌سنجند. در حالی که طرفداران هوش مصنوعی قوی نگران ساختار سیستم‌هایی هستند که می‌سازند. هالیوود با برخی فیلم‌های که در این باره می‌سازد در رویکرد قوی به هوش مصنوعی گنجانده می‌شود.^۲

با تبیین مختصر دو رویکرد ضعیف و قوی به هوش مصنوعی پاسخ برخی

۱. Lucci, S., & Kopec, D. (۲۰۱۶). Artificial intelligence in the ۲۱st century: A living introduction, mercury learning and information, p۱۲.

۲. Ibid.

مسائل تاحدودی مشخص می‌شود: اینکه خود ربات‌ها را مخاطب باید و نبایدهای اخلاقی قرار بگیرند، یا اینکه سیستم‌های هوش مصنوعی خودمختار، همچون خودروهای خودران، نسبت به تصمیم‌ها و افعالی که انجام می‌دهند مسئولیت‌پذیر باشند زمانی متصور است که آنها از روی اراده، اختیار و آگاهی فعلی را انجام داده باشند. اینکه هوش مصنوعی آگاهی و ذهن داشته باشد همان دیدگاه هوش مصنوعی قوی است. بنا بر این، نظریه هوش مصنوعی قوی و مسائل مربوط به آن را می‌توان بنا بر مباحث فلسفی-عقلی مورد تحلیل قرار داد.

نقد امکان اخلاق‌مندی هوش مصنوعی

بنا بر مبانی علم النفس فلسفی اسلامی، ذهن، آگاهی و متفکر بودن هوش مصنوعی مورد نقد اساسی است؛ زیرا یک جسم مصنوع نمی‌تواند دارای ذهن (غیر فیزیکی)، نفس، آگاهی و تفکر باشد. از منظر فلسفه اسلامی، نفس به جسم طبیعی تعلق دارد^۱ یا از جسم طبیعی برمی‌خیزد.^۲ حکما، از ارسطو و فیلسوفان مشائی گرفته تا ملاصدرا و پیروان حکمت متعالیه، در تعریف نفس گفتند: «کمال اول لجسم طبیعی آلی». بنا بر این تعریف، نفس برای جسم طبیعی و آن هم با قید آلی است.^۳ بنا بر این، تعریف تعلق نفس به جسم مصنوعی و ساخته شده منتفی می‌شود؛ زیرا در فلسفه اسلامی بین جسم طبیعی و جسم مصنوعی و نحوه وجود و وحدت آنها تفاوت اساسی وجود دارد و هر کدام احکام هستی‌شناسانه خاص خود را دارند. همچنین، بنا بر مباحث حکمت متعالیه دلیل‌های دیگری از طریق فرق بین وحدت حقیقی و وحدت اعتباری، نحوه وجود ماده و صورت و

۱. دیدگاه سینیوی

۲. دیدگاه صدرایی

۳. ر.ک: صدر المتألهین، الحکمة المتعالیة فی الاسفار العقلیة الاربعة، ج ۸.

مسئله مهم مزاج نیز می‌توان ادله متقنی برای عدم امکان تحقق هوش مصنوعی قوی اقامه کرد. هر چند برخی محققان با استفاده از برخی مبانی حکمت متعالیه، همچون نظریه جسمانیة الحدوث بودن نفس و حرکت جوهری، سعی در اثبات امکان تحقق هوش مصنوعی دارند، از نظر نگارنده این دیدگاه، بنا بر مبانی دیگر حکمت متعالیه، مورد نقد اساسی است و بیش از این پژوهش حاضر اجازه تحلیل و بررسی این مسئله را نمی‌دهد.

حتی بنا بر دیدگاه برخی فیلسوفان ذهن غربی، همچون جان سرل، ماشین‌های هوشمند قادر به درک محتوا و معنا نمی‌باشند؛ استدلال اتاق چینی سرل در این باره معروف است.^۱ همچنین، بنا بر دیدگاه چالمرز در تفکیک مسئله آگاهی به مسئله ساده آگاهی و مسئله سخت آگاهی، شاید مسئله ساده آگاهی معرفت هوش مصنوعی را توجیه کند، مسئله سخت آگاهی (یعنی حالت پدیداری و حالت حس درونی که در انسان وجود دارد) نمی‌تواند در یک ماشین یا سیستم هوشمند تحقق یابد.^۲ تبیین و بررسی بیشتر از این نیازمند بحث مستقل و مفصل خاص خود است و در اینجا به همین اندازه بسنده می‌شود.

خلاصه اینکه ربات و ماشین‌های هوشمند نفس یا ذهن ندارند تا آگاهی و اراده آزاد در انجام افعال را داشته باشند. پس نمی‌توان باید و نبایدهای اخلاقی را متوجه خود هوش مصنوعی کرد. همچنین، هوش مصنوعی نمی‌توان نسبت به افعالی که انجام می‌دهد مسئولیت پذیر کرد. نیز آنها قادر نیستند فعل اخلاقی انجام دهند و فضایل اخلاقی و حالات و کیفیات نفسانی را اکتساب کنند.

۱. ر.ک: جان سرل، ذهن، مغز و علم، فصل دوم، ص ۵۵-۷۴.

۲. David J, Chalmers, (۲۰۰۹). The character of consciousness, p: ۵-۳

۲. اخلاق هوش مصنوعی در رویکرد فنی-مهندسی

اما مراد از اخلاق هوش مصنوعی در رویکرد فنی-مهندسی این است که چگونه می‌توان کدهای اخلاقی را در سیستم‌های هوشمند تدوین کرد تا از خطرات احتمالی هوش مصنوعی علیه انسان‌ها جلوگیری شود.^۱ در این رویکرد، دیگر خود ماشین هوشمند مخاطب باید و نباید یا تصمیم‌گیری اخلاقی نمی‌شوند، بلکه به نوعی تحت جبر اخلاقی قرار خواهند داشت. آنها به عنوان یک ماشین به حساب می‌آیند؛ ماشین هوشمندی که برنامه‌های اخلاقی داده شده را به خوبی پیاده‌سازی می‌کند.

در این راستا، مراکز دانشگاهی و تحقیقاتی در سراسر دنیا در حال انجام پژوهش هستند؛ برای نمونه، محققان پروژه تحقیقاتی تحت عنوان «هوش مصنوعی باز» (OpenAI) سعی دارند کدهایی را برای هوش مصنوعی تعریف کنند. آنها بر این باورند که این بهترین راه برای مبارزه با خطرات مخرب هوش مصنوعی خواهد بود.^۲ مطالعه صدساله هوش مصنوعی (AI۱۰۰) در دانشگاه استنفورد پژوهش بلندمدت دیگری است که در پاییز ۲۰۱۴ م راه‌اندازی شده است. این تحقیق، پژوهشی طولانی‌مدت در زمینه هوش مصنوعی و تأثیر آن بر مردم، جوامع انسانی و جامعه است. همچنین، این مطالعات چالش‌های و فرصت‌های فنی، اجتماعی که این پیشرفت در عرصه‌هایی مانند اخلاق، اقتصاد و طراحی سیستم سازگار با شناخت انسان ایجاد می‌کند را بررسی می‌کنند.^۳ هدف از این مطالعات

۱. Paula Boddington, (۲۰۱۷). Towards a Code of Ethics for Artificial Intelligence, p۱.

۲. <http://open.ai/>

۳. Peter, Stone, Et al. Artificial Intelligence and Life in ۲۰۳۰: The One Hundred Year Study on Artificial Intelligence, REPORT OF THE ۲۰۱۵ STUDY PANEL, SEPTEMBER ۲۰۱۶.

پیش‌بینی چگونگی تأثیرات هوش مصنوعی در تمام جنبه‌های کار، زندگی و بازی افراد انسانی است.^۱ این پروژه‌ها نشان می‌دهند محققان به مشکلات، خطرات بالقوه، بلکه چالش‌های بالفعل هوش مصنوعی واقف هستند و در این زمینه تحقیق و پژوهش می‌کنند.

شرکت‌های بزرگ و کوچک نیز ابتکاراتی در این زمینه داده‌اند: همانند مشارکت در پروژه‌های هوش مصنوعی به نفع مردم و جامعه، با همکاری آمازون، دیپ مایند،^۲ فیس‌بوک، گوگل، ای بی ام و مایکروسافت.^۳ افرادی نیز پژوهش‌هایی در دست دارند: همانند تحقیقاتی که سوگیری در هوش مصنوعی را بررسی می‌کند.^۴ همچنین، سازمان‌های چون برین گوگل،^۵ مرکز هوش مصنوعی سازگار با انسان،^۶ مؤسسه تحقیقاتی هوش ماشینی^۷ در دانشگاه برکلی کالیفرنیا^۸ در حوزه ایمنی هوش مصنوعی کار می‌کنند.^۹

همان‌طور که تبیین گشت، در مراکز علمی دنیا به‌طور گسترده در زمینه هدایت فناوری هوش مصنوعی تحقیق و پژوهش می‌شود. از یک سو، امروزه مسئله فناوری‌های نوین هوش مصنوعی همه جنبه‌های زندگی جوامع بشری را تحت تأثیر خود قرار داده است. از سوی دیگر، تهدیدات هوش مصنوعی امری

۱. <https://ai.100.stanford.edu/>

۲. DeepMind

۳. <https://www.partnershiponai.org/>

۴. Paula Boddington , (۲۰۱۷). Towards a Code of Ethics for Artificial Intelligence, p۱

۵. Google Brain

۶. the Center for Human Compatible AI

۷. the Machine Intelligence Research Institute

۸. UC Berkeley

۹. Allan Dafoe (۲۰۱۸), AI Governance: A Research Agenda, p:۳۳.

واقعی است و به همین دلیل حکمرانی هوش مصنوعی امری اجتناب‌ناپذیر و ضروری است.^۱ از این رو، بنا بر اهمیت و ضرورت حکمرانی هوش مصنوعی در سطح دنیا و تهدیدات جدی از ناحیه هوش مصنوعی، می‌بایست در نظام ایران اسلامی نیز در این زمینه پژوهش‌های مستقل و چشم‌گیری انجام شود.

بنابراین، بر اساس فلسفه اخلاق اسلامی و آموزه‌های دین اسلام می‌توان این ایده را درباره امکان اخلاق‌مندی ربات‌ها در رویکرد فنی-مهندسی مطرح کرد: «می‌توان با به‌کار بردن الگوریتم‌ها و سیستم‌ها، داده‌ها و برنامه‌های مبتنی بر ارزش‌های اخلاقی اسلامی تولید کرد و از این سیستم‌های هوشمند به‌گونه‌ای در ترویج ارزش‌های اخلاقی و مقابله با مقوله‌های ضد اخلاقی در فضای سایبر و فضای فیزیکی استفاده کرد».

بنا بر برخی عبارت‌های توصیه‌نامه اخلاق هوش مصنوعی یونسکو، که باید هوش مصنوعی را به سمت ارزش‌های اخلاقی هدایت کرد، می‌توان گفت با به‌کار بردن الگوریتم‌هایی در هوش مصنوعی و دادن داده‌هایی درباره ارزش‌های اخلاقی می‌توان به‌گونه‌ای در ترویج ارزش‌های اخلاقی اسلامی از این فن‌آوری کمک گرفت.

۳. اخلاق هوش مصنوعی در ارتباط با عملکرد انسان‌ها

اما بنا بر رویکرد سوم، مخاطب اصلی و مستقل باید و نبایدهای اخلاقی انسان‌ها هستند. انسان، به عنوان موجودی مختار، دارای شعور و خودآگاه توانایی شناخت خوب و بد اخلاقی در استفاده فناوری هوش مصنوعی را دارد. انسان به دلیل اینکه موجودی مختار، آگاه و باشعور است حالات و کیفیات نفسانی

۱. Wang, Weiyu and Siau, Keng, "Artificial Intelligence: A Study on Governance, Policies, and Regulations" (۲۰۱۸), p:۳.

دارد که تأثیر افعال اختیاری او در سعادت و شقاوتش قابل ترسیم است.

در توصیه‌نامه یونسکو بیشتر اخلاق هوش مصنوعی به معنای سوم به کار رفته است؛ هرچند در راستای اخلاق هوش مصنوعی با رویکرد دوم و رویکرد فنی-مهندسی نیز نکاتی آمده است. البته این سند به طور صریح اخلاق هوش مصنوعی را تعریف و تبیین نکرده، ولی ما با بررسی مفصل مفاد بندهای توصیه‌نامه یونسکو به این مطلب رهنمون شدیم که بیشتر مخاطبان باید ونباید‌های اخلاقی این سند نقش آفرینان انسانی در چرخه هوش مصنوعی، اعم از کاربران، برنامه‌نویسان، شرکت‌ها، دولت‌ها و ملت‌ها، هستند. در ادامه ذیل نقد سند اخلاق هوش مصنوعی یونسکو رویکرد سوم نیز بیشتر توضیح داده می‌شود.

بررسی توصیه‌نامه اخلاق هوش مصنوعی یونسکو و سازگاری آن با مبانی اخلاق اسلامی^۱

تبیین توصیه‌نامه اخلاق هوش مصنوعی یونسکو

یونسکو چارچوب‌های نظارت بین‌المللی و ملی را فراهم آورده تا فناوری نوین و بزرگ هوش مصنوعی در نهایت به نفع بشر تمام شود. به همین دلیل، سندی با عنوان «توصیه‌نامه اخلاق هوش مصنوعی» را در چهل و یکمین جلسه کنفرانس

۱. برای نقد مفصل سند اخلاق هوش مصنوعی یونسکو با تکیه بر مبانی فلسفه اخلاق ن.ک: حمید محسنی و همکاران، «بررسی توصیه‌نامه اخلاقی هوش مصنوعی یونسکو و تبیین نقاط و ضعف و قوت آن»، در کتاب نقد و تکمیل پیش‌نویس نهایی توصیه‌نامه درباره اخلاق هوش مصنوعی، دفتر فقه معاصر حوزه علمیه قم، ص ۱۸۳-۲۱۷. همچنین ن.ک: حمید محسنی، و کاظم فولادی قلعه، سازگاری و ناسازگاری توصیه‌نامه اخلاق هوش مصنوعی یونسکو با نظریه اخلاقی اسلامی، اولین کنفرانس فضای سایبر، دانشکده مهندسی دانشکدگان فارابی، دانشگاه تهران. ما از مطالب، یافته‌ها و نتایج تحقیقات مزبور در این پژوهش استفاده خواهیم کرد. البته توضیح بیشتر و تفصیلی‌تر در اینجا مقدور نیست.

عمومی خود، در تاریخ ۲۴ نوامبر ۲۰۲۱ م، تصویب کرده است.^۱

پیش نویس اصلی، اصلاح و تکمیل توصیه نامه اخلاق هوش مصنوعی یونسکو زیر نظریک گروه تخصصی ویژه (AHEG) ۲۴ نفری از کشورهای همچون اروگوئه، غنا، عربستان، امارات، روسیه، مکزیک، آمریکا، لهستان، آرژانتین، لتونی، رواندا، اسلونی، کامرون، برزیل، مصر، مراکش، افریقای جنوبی، هند، ژاپن، فرانسه، بریتانیا، کره و چین است. ما با بررسی دقیق رزومه نویسندگان اصلی این سند به این یافته رسیدم که از این ۲۴ نفر کارشناس، تخصص تعداد زیادی از آنها علوم فنی-مهندسی است و چند نفر هم سیاستمدار و مشاور وزیر در جمع آنها حضور دارند. در این جمع، چند فیلسوف و حقوقدان و حدود سه نفر متخصص اخلاق AI نیز از کشورهای آمریکا، بریتانیا و چین مشاهده می شود.^۲

این توصیه نامه یک مقدمه و ۱۴۱ بند دارد. در متن آن، اهداف و باید و نبایدهای آورده شد و از کشورهای عضو، از جمله ایران، خواسته شده مفاد این سند را در کشورهای خود به نحوی اجرا کنند. این موضوع در بند ۱۳۵ مورد اشاره قرار گرفته است. هرچند در برخی مفاد بندها اشاره شده که این قوانین بنا بر قوانین داخلی کشورها اجرایی شود، نهادهای بین المللی به طور غیرمستقیم سعی دارند با تصویب این گونه اسناد بالادستی قوانین و ارزش های جوامع را با خود همراه و همسو سازند. با بررسی گزارش هایی که پیش از تکمیل متن اصلی تنظیم شده به نکته مهمی در این باره دست یافتم که در متن اصلی سند یونسکو بدین گونه به آن تصریح نشده است. در آن گزارش آمده که یونسکو سعی دارد رهبری اخلاقی خود را با تنظیم استانداردهای در حوزه علم، فناوری و هوش مصنوعی با رویکردی

۱. <https://en.unesco.org/artificial-intelligence/ethics>, retrieved: ۲۰۲۲ Feb. ۶.

۲. <https://unesdoc.unesco.org/ark:/۴۸۲۲۳/pf۰۰۰۰۳۸۰۴۵۵>.

انسان محور تقویت کند.^۱ همچنین در بند ۴ سند آمده، این توصیه نامه دولت های عضو را به عنوان یکی از بازیگران هوش مصنوعی برای توسعه چارچوب های قانونی و نظارتی در کل چرخه هوش مصنوعی مورد خطاب قرار داده و به آنها رهنمود اخلاقی می دهد.

این توصیه نامه از این جهت که سعی دارد اهمیت هنجارهای اخلاقی در حوزه هوش مصنوعی را در بین ملت ها ترویج کند و فرهنگ سازی نماید می بایست مورد توجه و تأمل قرار گیرد. از این جهت که یونسکو نگران به خطر افتادن امنیت، صلح، عدالت، کرامت و آزادی انسان از ناحیه هوش مصنوعی است نیز قابل توجه است. در این توصیه نامه، درباره کرامت انسانی و حقوق بشر آمده است: «احترام، حفاظت و ارتقاء [در مورد] کرامت انسانی و حقوقی که بر اساس قوانین بین المللی تثبیت شده است، از جمله قانون بین المللی حقوق بشر در طول چرخه حیات سیستم های هوش مصنوعی ضروری است».^۲ همچنین در بند ۱۶ آمده است که حقوق بشر و آزادهای اساسی باید در این فناوری مورد احترام و حفاظت قرار گیرد.^۳ همچنین درباره عدالت و عدم تبعیض مطالب قابل توجهی در این توصیه نامه ذکر شده است. در بند ۲۸ بیان شده: «بازیگران هوش مصنوعی باید عدالت اجتماعی را ارتقاء دهند و همواره انصاف و عدم تبعیض را با رعایت قانون بین المللی حفظ کنند».

پس در این سند اخلاقی به مفاهیم ارزشمندی همچون کرامت، عدالت،

۱. https://unesdoc.unesco.org/ark:/48223/pf41.0000379920_C/73-p:2.

۲. حوزه ارزش ها، بند ۱۳.

۳. همچنین در بند ۳۲ این سند آمده: «حریم خصوصی حقی اساسی برای حفظ کرامت انسانی، استقلال و اختیار انسان است».

آزادی، اشاره شده است. در آیین اسلام نیز به این مسائل توجه ویژه‌ای شده است. به طور نمونه، در قرآن کریم خداوند متعال درباره کرامت انسانی می‌فرماید: ﴿مَاطْعاً بَنِي آدَمَ رَا تَكْرِيمَ كَرَدِيمَ﴾؛ «وَلَقَدْ كَرَّمْنَا بَنِي آدَمَ»^۱ و در باب اهمیت عدل و قسط می‌فرماید: «یکی از اهداف فرستادن پیامبران ﷺ این بود تا انسان‌ها برای برپایی عدالت به پا خیزند»؛ «لَقَدْ أَرْسَلْنَا رُسُلَنَا بِالْبَيِّنَاتِ وَأَنْزَلْنَا مَعَهُمُ الْكِتَابَ وَالْمِيزَانَ لِيَقُومَ النَّاسُ»^۲. همچنین، امیرمؤمنان علی (ع) درباره آزادی در نگاه اسلام می‌فرماید: «خداوند متعال انسان‌ها را آزاد آفریده است»؛ «لَا تَكُنْ عَبْدَ غَيْرِكَ وَقَدْ جَعَلَكَ اللَّهُ حُرّاً»^۳. از این رو، این مفاهیم ارزشمند نیز در مکتب اسلام مورد تأکید بوده و از اصول و مبانی اسلام هستند.

اما این مفاهیم ارزشمند در هر مکتب فلسفی و اخلاقی معنا، حقیقت، حدود و ثغور و مبانی خاص خود را دارد. از این رو، منظور از مفاهیمی از قبیل کرامت، عدالت و آزادی، که در این توصیه‌نامه آمده چیست؟ آیا این مفاهیم دارای ارزش ذاتی هستند، یا نه؟ همچنین می‌بایست بررسی شود که ارزش‌های اخلاقی که در سند اخلاق هوش مصنوعی یونسکو آمده مبتنی بر مبانی ارزشی کدام مکاتب فلسفه اخلاق هستند. یونسکو در این سند بارها تأکید کرده کشورهای عضو، از جمله ایران، این سند اخلاقی را به گونه‌ای در کشورشان اجرا کنند. بنابراین، می‌بایست بررسی شود که آیا ارزش‌های سند اخلاقی هوش مصنوعی یونسکو با نظریه اخلاقی اسلام سازگاری دارد یا نه.

۱. اسراء، ۷۰.

۲. حدید، ۲۵.

۳. سید رضی، نهج البلاغه (للصباحی صالح)، ص ۴۰۱.

تحلیل منشأ ارزش‌های اخلاق هوش مصنوعی یونسکو

به‌طورکلی مفاد این توصیه‌نامه در قلمرو ارزش‌ها و هنجاری‌های اخلاقی هستند. هنجاری‌های اخلاقی نیز مبتنی بر طیفی از مبانی و اصول هستند. در علم اخلاق، مکاتب متعدد اخلاقی وجود دارد که هرکدام مبتنی بر قواعد و اصولی هستند. در نگاهی کلی مکاتب اخلاقی در برخی کتب فلسفه اخلاق به مکاتب واقع‌گرا و غیر واقع‌گرا تقسیم می‌شوند. از این رو می‌بایست نخست نگاهی کلی به مکاتب اخلاقی انداخته شود، سپس متن توصیه‌نامه را بر اساس آن بررسی شود. آنگاه مشخص خواهد شد که منشأ ارزش‌های اخلاقی این سند برخاسته از دیدگاه‌های واقع‌گرایی است یا متأثر از نظریه‌های غیرواقع‌گرایی.

با تأمل در نظریه‌های غیر واقع‌گرا، همچنین با استناد به برخی مفاد متن مذکور آشکار می‌شود که هنجارها و ارزش‌های اخلاقی مطرح در توصیه‌نامه یونسکو مبتنی بر مکاتب اخلاقی غیر واقع‌گرا (از نوع امرگرایی نوع دوم و یا قراردادگرایی) هستند. این نظریه‌ها لوازمی دارند که مهم‌ترین آنها کثرت‌گرایی و نسبی‌گرایی در حوزه علم اخلاق است. همچنین، این دو لازمه در برخی بندهای توصیه‌نامه نیز مشهودند. این مطلب نشانگر صحت ادعای پژوهش حاضر است و آن را اثبات می‌کند. در پایان این کاوش به نظریه واقع‌گرایی مکتب اخلاقی اسلام به‌طور فشرده اشاره خواهد شد و بر اساس آن راهکاری برای ارزش‌های اخلاقی هوش مصنوعی ارائه خواهیم کرد.

اینکه منشأ ارزش و لزوم اخلاقی به کار برده شده در این توصیه‌نامه به کدام یک از مکاتب اخلاقی نزدیک است، به نظر می‌رسد منشأ ارزش و لزوم اخلاقی و هنجارهایی که در این توصیه‌نامه به آن اشاره می‌شود از نوع قراردادگرایی باشد.

مکتب قراردادگرایی نیز ذیل مکاتب غیر واقع‌گرایی قرار می‌گیرند.^۱ همان طور که با تأمل در برخی مفاد مقدمه سند مشخص می‌شود این سند حقوق بشر و آزادی‌ها را به تأیید و پذیرش مردم مقید کرده است. همچنین در بند ۶ بیان می‌شود: «هدف این توصیه‌نامه رسیدن به یک ابزار هنجاری پذیرفته‌شده در سطح جهانی است». در نتیجه این مفاهیم را ملاک و منشأ هنجارهای اخلاقی معرفی می‌کند و بر اساس آنها «بایدها و نبایدهای اخلاقی» را در حوزه هوش مصنوعی معین می‌سازد. عبارت‌های مذکور ما را به این سمت سوق می‌دهد که بپذیریم برخی از مبانی به کار رفته در این متن برخاسته از مکاتب غیر واقع‌گرایی و از نوع قراردادگرایی هستند.

قراردادگرایی چه اشکالاتی دارد؟ اینکه یک گروه یا نهاد در راستای حفظ ارزش‌های اخلاقی، باید و نبایدهایی را دستور دهند تا دیگران به انجام آن ملتزم شوند و یا اینکه عده‌ای دورهم جمع شوند تا درباره طیفی از هنجارهای اخلاقی با هم توافق کنند، چه اشکالاتی ممکن است داشته باشد. برای پاسخ به این مسائل و مشخص شدن ایرادهای مکاتب غیرواقع‌گرا، باید برخی لوازم این مکاتب را به طور خلاصه تبیین کرد.^۲ در ادامه دو مورد از لوازم اصلی مکاتب غیرواقع‌گرا توضیح داده می‌شود.

۱. کثرت‌گرایی اخلاقی: ۳

یک لازمه مهم نظریه غیرواقع‌گرایی این است که گزاره‌های اخلاقی قابلیت

۱. ر.ک: مجتبی مصباح، بنیاد اخلاق (روشی نو در آموزش فلسفه اخلاق)؛ همچنین: محمدتقی، مصباح یزدی، نقد و بررسی مکاتب اخلاقی، تحقیق و نگارش: احمد حسین شریفی.

۲. برای تفصیل بیشتر ن.ک: مصباح، بنیاد اخلاق (روشی نو در آموزش فلسفه اخلاق)، ۳۲-۳۷.

اتصاف به صدق و کذب را ندارند. از این رو، دو گزاره «راستگویی خوب است» و «راستگویی بد است» هر دو به یک اندازه قابل قبول خواهند بود. در نتیجه، دو مکتب اخلاقی متناقض با یکدیگر هیچ کدام بر دیگری برتری نخواهد داشت و همه به یک اندازه قابل قبول خواهند بود. این امر لازمه مهم دیگری یعنی کثرت‌گرایی اخلاقی خواهد داشت.

به بیان دیگر، به نوعی مکتب غیرواقع‌گرایی عصاره خود را در کثرت‌گرایی و نسبی‌گرایی نشان می‌دهد. از این رو، تحلیل این دو اهمیت بسزایی در نقد نظریه غیر واقع‌گرایی خواهد داشت. همچنین می‌توان شواهدی بیشتری از متن توصیه‌نامه پیدا کرد که به کثرت‌گرایی و نسبی‌گرایی اشاره می‌کند. پس این ادعا که مفاد این توصیه‌نامه مبتنی بر نظریه غیرواقع‌گرایی هست، به راحتی اثبات خواهد شد.

در مقدمه این توصیه‌نامه نویسندگان متذکر می‌شوند که در بحث هوش مصنوعی نگرانی‌هایی وجود دارد که باعث شود اموری مانند دانش بومی و کثرت‌گرایی فرهنگی (Cultural pluralism) نادیده گرفته شود. بر این اساس، یکی از دغدغه‌های توصیه‌نامه یونسکو این است که کثرت‌گرایی فرهنگی از ناحیه هوش مصنوعی محدود نشود. پس با استناد به این مطلب و مفاد دیگر متن، مشخص می‌شود که کثرت‌گرایی فرهنگی در این توصیه‌نامه پذیرفته شده است. از این رو، بنا بر تنوع فرهنگی ممکن است که ارزش‌های اخلاقی مبتنی بر آنها نیز متنوع باشند. همان‌طور که در بند اول تصریح می‌شود «اخلاق هوش مصنوعی به عنوان یک تفکر هنجاری سیستماتیک مبتنی بر یک چهارچوب کل‌گرا، جامع، چند فرهنگی و در حال تکامل از ارزش‌های وابسته به یکدیگر تنظیم می‌شود». در این بند تأکید می‌کند که ارزش‌های اخلاقی مبتنی بر کثرت‌گرایی فرهنگی است.

همچنین، در بند ۸، در حوزه اهداف و مقاصد، بخش d به طور واضح اشاره می‌شود که یکی از اهداف این توصیه‌نامه «بالندگی گفت‌وگوی چندجانبه، و کثرت‌گرا درباره مسائل اخلاقی مربوط به هوش مصنوعی است». در این بند، به طور آشکار به کثرت‌گرایی اخلاقی اشاره می‌کند. افزون بر این، ذیل بند ۴۸ در حوزه سیاست‌گذاری اشاره می‌شود که فرایند تدوین سیاست‌گذاری‌ها باید همه ذی‌نفعان را شامل شود و باید شرایط و اولویت‌های کشورهای عضو در نظر گرفته شود.

بنا بر کثرت‌گرایی، مفاد بندهای ۱۳۱ تا ۱۳۴، که در حوزه نظارت و ارزیابی تأثیرات اخلاقی فناوری‌های هوش مصنوعی هستند و سعی دارند روشی را برای این ارزیابی ارائه دهند، دچار مشکل و ایراد اساسی می‌شوند؛ زیرا بنا بر کثرت‌گرایی، نمی‌توان یک روش واحد را برای بررسی تأثیرات اخلاقی در کشورهای عضو ارائه داد. چون، بنا بر کثرت‌گرایی فرهنگی و اخلاقی، ممکن است یک رفتار در جامعه «الف» ارزش اخلاقی به حساب بیاید و از نگاه دولت یا نهادهای فرهنگی و مذهبی آن کشور برخی فناوری‌های هوش مصنوعی تهدیدی برای آن محسوب شود؛ اما در جامعه «ب» آن کنش دارای ارزش اخلاقی نباشد، بلکه ضد ارزش به حساب آید. از این رو، استفاده از هوش مصنوعی در جامعه «ب» هیچ مشکلی نداشته باشد.

پس نهادهای بین‌المللی، همچون یونسکو، چطور می‌توانند تأثیرات اخلاقی هوش مصنوعی را در بین جوامع مختلف بررسی کنند. آیا با قبول کثرت‌گرایی باید ارزش اخلاقی جامعه الف را پذیرفت و با آن جوامع دیگر را بررسی کرد یا ارزش‌های جامعه «ب» را پذیرفت و با آن دیگر جوامع را بررسی کرد؟ در این صورت، آیا جوامع دیگر، که این امر در آنها ضد ارزش محسوب می‌شود، حق اعتراض ندارند. بنابراین، پذیرش کثرت‌گرایی فرهنگی و اخلاقی چالش‌های بسیاری را ایجاد

می‌کند که یکی از آنها عدم امکان ارزیابی یکسان ارزش‌های اخلاقی از ناحیه فناوری هوش مصنوعی در بین جوامع مختلف هست.

۲. نسبیت‌گرایی اخلاق^۱

منظور از نسبیت بودن اخلاق این است که هیچ حکمی در زمینه اخلاق کلی و مطلق نیست. به عبارت دیگر، نسبیت‌گرایی اخلاقی یعنی اینکه بپذیریم اعتبار همه احکام اخلاقی شرایطی دارد و این احکام با نظر به همان شرایط معتبرند. پس بنا بر نظریه غیرواقع‌گرایی که برای گزاره‌های اخلاقی ورای دستور، احساس یا قرارداد واقعیت عینی لحاظ نمی‌کند، گزاره‌های اخلاقی نسبی خواهند بود. یعنی به طور نمونه اگر دستور یا قراردادی برای گزاره «حجاب برای زنان خوب است» در جامعه‌ای بود، این گزاره برای همان‌ها اعتبار دارد؛ ولی اگر در جامعه‌ای دستور و یا قراردادی نبود، این گزاره اعتبار و ارزشی ندارد.

همان‌طور که در تبیین مطلب گذشته روشن شد، متن این توصیه‌نامه اخلاقی به نظریه غیر واقع‌گرایی اخلاقی نزدیک است. در این صورت نسبیت‌گرایی یکی از لوازم این متن خواهد بود. همچنین، مباحثی که در کثرت‌گرایی مطرح شد می‌تواند مؤیدی بر نسبیت‌گرایی نیز باشد. افزون بر این مطالب، شواهد دیگری نیز در متن توصیه‌نامه وجود دارد که بر اساس آن رویکرد نسبیت‌گرایی در این سند را آشکار می‌کند. نسبیت‌گرایی خود ایرادهای اساسی دارد که نمی‌توان در اینجا به آن پرداخت.^۲ اما یکی از این اشکالات این است که بر اساس نسبیت‌گرایی نمی‌توان بین ارزش‌های متفاوت اخلاقی جوامع مختلف داوری کرد. از این رو، همان‌طور

۱. Ethical Relativism

۲. ر.ک: مسلم محمدی، مکتب‌های نسبیت‌گرایی اخلاقی؛ همچنین: مجتبی مصباح، بنیاد اخلاق (روشی نو در آموزش فلسفه اخلاق)، فصل ۵، اطلاق یا نسبیت اخلاق.

که در کثرت‌گرایی گفته شد، مفاد بندهای ۱۳۱ تا ۱۳۴ که مربوط به نظارت و ارزیابی است، دچار چالش و اشکال اساسی می‌شوند. پس نمی‌توان بین دو ارزش اخلاقی متناقض با هم در جوامع مختلف داوری کرد و گفت کدام ارزش خوب هست و کدام ارزش بد.

اخلاق هوش مصنوعی بنا بر دین اسلام

در پایان این پژوهش، پس از بیان نقدهای مبانی و بنایی به توصیه اخلاقی یونسکو، راهکار اخلاقی هوش مصنوعی از منظر مکتب اسلام را مطرح می‌کنیم. پیش از ارائه این راهکار نخست به طور کوتاه به برخی اصول اخلاقی اسلام اشاره می‌کنیم؛ زیرا این مقاله گنجایش آن را ندارد که در آن مبانی و اصول نظریه اخلاقی مکتب اسلام به طور مفصل تبیین و اثبات شود.^۱ از این رو، به طور خلاصه برخی از مهم‌ترین مبانی و اصول نظام اخلاقی اسلام را در اینجا ذکر می‌کنیم تا براساس آن نظریه مختار درباره اخلاق هوش مصنوعی تبیین گردد.

در نظریه اخلاقی اسلام، مفاهیم ارزشی‌های اخلاقی مبتنی بر واقعیت هستند.^۲ از این رو، این مکتب ذیل نظریه واقع‌گرا قرار می‌گیرد. پس اشکال‌هایی که بر نظریه غیر واقع‌گرایی وارد بود بر نظام اخلاقی اسلام وارد نمی‌شود. اصل دیگر که باید مورد توجه قرار گیرد اصل فلسفی و هستی‌شناسی است و آن ضرورت بالقیاس می‌ان فعل اختیاری و نتیجه آن است. این اصل برگرفته از رابطه می‌ان علت و معلولش است؛ یعنی بین علت و معلول رابطه ضرورت بالقیاس وجود دارد. از این رو، نتایج افعال اختیاری معلول فعل اختیاری هستند. فعل اختیاری نیز نسبت

۱. ر.ک: محمدتقی مصباح یزدی، نقد و بررسی مکاتب اخلاقی، فصل پنجم؛ همچنین: مجتبی مصباح، بنیاد اخلاق، فصل‌های ۹، ۱۰ و ۱۱.

۲. اصل معرفت‌شناختی.

به آنها ضرورت بالقیاس دارد. اصل بعدی، اصلی دینی است که انسان علاوه بر بدن، روح و نفسی دارد که بعد از مرگ تا ابد باقی می ماند. از این رو زندگی انسان منحصر در این دنیا نیست و حقیقت انسان نیز به همین بُعد فنا ناپذیر برمی گردد. همچنین روح انسان قابل تکامل و رشد است.

اصل دیگر آن است که انسان موجودی مختار است که در اثر افعال اختیاری خود تکامل پیدا می کند. روح انسان در ذات خود گرایش به کمال دارد. این گرایش به صورت اراده نفس ظهور پیدا می کند. روح انسان دارای حب ذات و پویاست. از این رو در ذات روح می لی به سوی تکامل وجود دارد. انسان نه تنها خودش را دوست دارد، کمالات و شئون ذاتش را نیز دوست دارد. پس، انسان افعالی را که انجام می دهد در راستای حب ذات و دست یابی به کمال افعال اختیاری خود انجام می دهد. این می ل به تکامل و کمال در هیچ حدی متوقف نمی شود، بلکه نفس، به دلیل پویایی که ذاتش است، می ل به سوی کمال بی نهایت دارد.

خلاصه اینکه، ارزش اخلاقی فعل اختیاری انسان تابع تأثیر آن فعل در رسیدن انسان به کمال حقیقی است. از این رو، هرکاری به اندازه ای که در آن کمال مؤثر باشد، ارزنده خواهد بود و اگر تأثیرش منفی باشد، ارزش آن منفی خواهد بود و اگر هیچ ارزش مثبت و منفی نداشته باشد، ارزش آن خنثی خواهد بود.

از سوی دیگر، کمال حقیقی انسان در تقرب به کمال مطلق، یعنی خداوند متعال، است؛ زیرا نفس انسان مجرد است و کمال خود را در بین مجردات می جوید و در بین مجردات نیز به دنبال کامل ترین آنها است. از این رو، هرکاری که موجب تقرب بیشتر انسان به خداوند باشد، مرتبه و ارزش آن بیشتر خواهد بود. در مقابل، هرکاری که موجب دوری انسان از خدا باشد، پست ترین درجه ارزش را خواهد داشت. منظور از تقرب نیز تقرب وجودی است که نفس انسان از

طریق انجام رفتارهای اختیاری و آگاهانه به لحاظ وجودی تکامل پیدا می‌کند و به خداوند نزدیک می‌شود. همچنین، افعال اختیاری زمانی ارزشمند هستند که آگاهانه و با نیت تقرب به خداوند باشند. پس در انجام افعال اختیاری هم حُسن فعلی و هم حُسن فاعلی لازم است.

در نهایت، بنا بر نظریه اخلاقی مکتب اسلام، اخلاق هوش مصنوعی را می‌توان این‌گونه تبیین کرد: «اگر چرخه حیات سیستم‌های هوش مصنوعی، از اختراع و کشف گرفته تا کاربرد و استفاده آن، در راستای این باشد که موجب کمال انسانی و تقرب وی به خداوند متعال باشد، این امر ارزشی اخلاقی خواهد داشت. برعکس اگر فن‌آوری‌های هوش مصنوعی موجب دوری انسان از کمال حقیقی خود باشد، ضد ارزش به حساب می‌آیند. در غیر این دو صورت، کاربرد و استفاده از هوش مصنوعی خنثی است و دارای ارزش و یا ضد ارزش اخلاقی نیست.»





منابع

رمضانی، مجید و فیضی درخشی، محمدرضا، «اخلاق ماشین، چالش‌ها و رویکردهای مسائل اخلاقی در هوش مصنوعی و ابر هوش»، فصلنامه اخلاق در علوم و فناوری، سال هشتم، شماره ۴، ۱۳۹۲.

ژکس، فلسفه اخلاق (حکمت عملی)، ترجمه ابوالقاسم پورحسینی، چاپ دوم، تهران، امیرکبیر، ۱۳۶۲.

سرل، جان آر، ذهن، مغز و علم، ترجمه امیر دیوانی، قم، بوستان کتاب، ۱۳۸۲.

شریف الرضی، محمد بن حسین (سید رضی)، نهج البلاغه (للصبحی صالح)، تحقیق فیض الاسلام، یک جلد، قم، هجرت، ۱۴۱۴ ق.

صدرالمتألهین، الحکمة المتعالیة فی الاسفار العقلیة الاربعة، چ سوم، ۹ جلد، بیروت، داراحیاء التراث، ۱۹۸۱ م.

طوسی، خواجه نصیرالدین، اخلاق ناصری، تهران، انتشارات علمیة اسلامیة، ۱۴۱۳ ق.

محسنی، حمید، و همکاران، «بررسی توصیه‌نامه اخلاقی هوش مصنوعی یونسکو و تبیین نقاط و ضعف و قوت آن»، در کتاب نقد و تکمیل پیش‌نویس نهایی توصیه‌نامه درباره اخلاق هوش مصنوعی، دفتر فقه معاصر حوزه علمیة قم،

۱۴۰۲.

محسنی، حمید، و فولادی قلعه، کاظم، «پیشرفت فناوری هوش مصنوعی مطلق یا هدفمند، در اندیشه مقام معظم رهبری»، دومین کنفرانس فضای سایبر، قم، دانشکده مهندسی دانشکدگان فارابی، دانشگاه تهران، آبان ۱۴۰۲.

محسنی، حمید، و فولادی قلعه، کاظم، «سازگاری و ناسازگاری توصیه‌نامه اخلاق هوش مصنوعی یونسکو با نظریه اخلاقی اسلامی»، اولین کنفرانس فضای سایبر، قم، دانشکده مهندسی دانشکدگان فارابی دانشگاه تهران، آبان ۱۴۰۱.

محمدی، مسلم، مکتب‌های نسبی‌گرایی اخلاقی (از یونان باستان تا دوران معاصر)، قم، بوستان کتاب، ۱۳۸۹.

مصباح یزدی، محمدتقی، فلسفه اخلاق، تحقیق و نگارش احمدحسین شریفی، قم، انتشارات مؤسسه آموزشی و پژوهشی امام خمینی، ۱۳۹۱.

مصباح یزدی، محمدتقی، نقد و بررسی مکاتب اخلاقی، تحقیق و نگارش احمدحسین شریفی، قم، چاپ سوم، مؤسسه آموزشی و پژوهشی امام خمینی، پاییز ۱۳۸۴.

مصباح، مجتبی، بنیاد اخلاق (روشی نو در آموزش فلسفه اخلاق)، چاپ هفتم، قم، مؤسسه آموزشی و پژوهشی امام خمینی، تابستان ۱۳۹۰.

Boddington, Paula, (۲۰۱۷). Towards a Code of Ethics for Artificial Intelligence, Springer International Publishing.

Chalmers, David J. (۲۰۰۹). The character of consciousness. Oxford University Press

Dafoe, Allan, (۲۰۱۸) AI Governance: A research agenda, Oxford university.

Stephen, Lucci and Danny, Kopec. (۲۰۱۶). Artificial intelligence in the ۲۱st century: A Living Introduction ۲/E.

Stone, Peter, Et al. Artificial Intelligence and Life in ۲۰۳۰: The One Hundred Year Study on Artificial Intelligence, REPORT OF THE ۲۰۱۵ STUDY PANEL, SEPTEMBER ۲۰۱۶.

Tallberg, Jonas, et al, (۲۰۲۳), The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research, <https://arxiv.org/ftp/arxiv/papers/۲۳۰۵/۱۱۵۲۸/۲۳۰۵.pdf>.

UNESCO (۲۰۲۱). Recommendation on the ethics of artificial intelligence, URL: <https://en.unesco.org/artificial-intelligence/ethics>, retrieved: ۲۰۲۲ Feb. ۶.

Wang, Weiyu and Siau, Keng (۲۰۱۸), "Artificial Intelligence: A Study on Governance, Policies, and Regulations", MWAIS ۲۰۱۸ Proceedings. ۴۰. <http://aisel.aisnet.org/mwais۴۰/۲۰۱۸>.

<https://www.techtarget.com/searchenterpriseai/definition/AI-governance>.

<https://intelligence.org/>

<http://lcfi.ac.uk/>.

<https://ai.stanford.edu/>

<http://open.ai/>

<https://intelligence.org/>

<https://www.partnershiponai.org/>

<https://hai.stanford.edu>

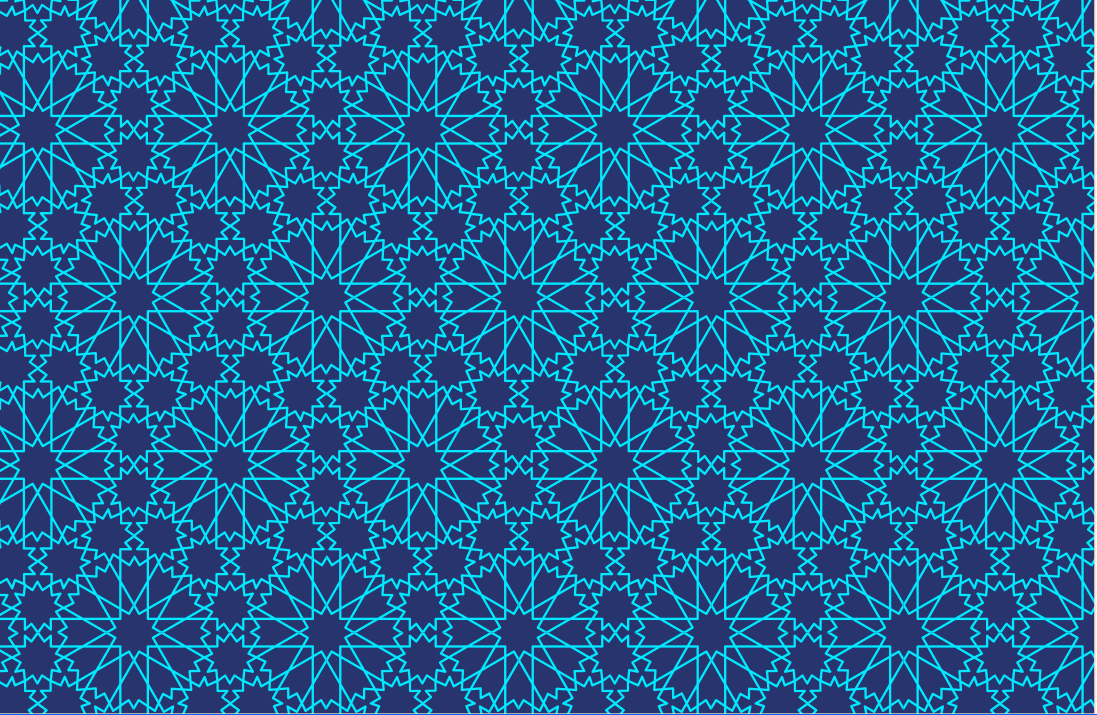
<https://futureoflife.org/resource/ai-policy>

<https://oecd.ai/en/wonk>

[https://www.oxfordinsights.com/government-ai-readiness-
index2020-](https://www.oxfordinsights.com/government-ai-readiness-index2020-)

[https://unesdoc.unesco.org/ark:/48223/pf0000380455.](https://unesdoc.unesco.org/ark:/48223/pf0000380455)

WWW.KHAMENI.IR



«آشنایی با هوش مصنوعی و نقاط تلاقی آن با الهیات»

دکتر بهروز مینایی^۱

۱. عضو هیأت علمی دانشگاه علم و صنعت ایران در تهران و رئیس ستاد فناوریهای هوشمند حوزه‌های علمیه:

b_minaei-at-iust.ac.ir





مقدمه

اولین باری است که هوش مصنوعی^۱ تنه به تنه الهیات می‌زند و مسئله مهم این است که علم کلام (تئولوژی)^۲ در چه جایی با هوش مصنوعی تلاقی پیدا می‌کند. سؤال دیگر این است که چه تفکری پشت سؤال اول است و چنین شبه‌های ایجاد شده است. در این بحث، ابتدا تعریفی از هوش مصنوعی ارائه می‌شود. سپس به دیدگاه‌های مختلف درباره هوش مصنوعی و مهم‌ترین قسم ربات‌ها^۳ یا عامل‌های هوشمند^۴ که توان یادگیری دارند، پرداخته می‌شود. سپس با اشاره به انواع یادگیری ماشین^۵ و دستاوردهای هوش مصنوعی با توجه به تعاریف مختلف، اشاره کوتاهی به کاربرست هوش مصنوعی در علوم اسلامی خواهد شد. در نهایت، به نقطه تلاقی هوش مصنوعی و الهیات، با اتکا به مباحث جاری در ارتباط علم و دین، پرداخته خواهد شد.

تعریف هوش مصنوعی

در تعریف هوش مصنوعی چهار رویکرد مطرح است که به شرح ذیل می‌باشند:

-
۱. Artificial Intelligence
 ۲. Theology
 ۳. Robots
 ۴. Intelligent Agents
 ۵. Machine Learning

۱. رویکرد اول: دانش ساختن ماشین‌هایی که بتوانند مثل آدم‌ها فکر کنند^۱ و دارای مغز^۲ باشند. نظریه گوگل‌برین^۳، یا مغز گوگل، امروزه بسیار مطرح است و در ایران هم پژوهشکده علوم شناختی^۴ تقریباً همان رویکرد گوگل‌برین را مطرح و پیگیری می‌کند. چکیده این نظریه این است که آیا می‌توان ماشین‌هایی ساخت که کپی مغز انسان باشند یا بتوان نحوه کار ذهن را در ماشین پیاده‌سازی کرد. از این رویکرد به «رویکرد مدل شناختی»^۵ تعبیر می‌شود. این رویکرد قوی‌ترین روایت از هوش مصنوعی است.

۲. رویکرد دوم: این رویکرد کمی ضعیف‌تر از رویکرد اول است، اما باز هم دیدگاه انسان‌گرایانه^۶ به هوش مصنوعی دارد. در این رویکرد مطرح می‌شود که هوش مصنوعی به دنبال پیاده‌سازی مغز انسان نیست؛ چون ذهن انسان، مانند خود او که به قول آلکسیس کارل موجودی ناشناخته است، ناشناخته است و از یک موجود ناشناخته چطور می‌توان کپی برداشت. پس بهتر است به دنبال ماشین‌هایی بود که رفتار انسان را تقلید کنند یا شبیه انسان‌ها عمل کنند.^۷ از این رویکرد به «رویکرد آزمون تورینگ» تعبیر می‌شود.^۸

۳. رویکرد سوم: از آن آقای جان مک‌کارتی^۹ است و او این ایده را در دانشگاه

۱. Thinking Humanly

۲. Brain

۳. Google Brain

۴. Institute for Cognitive Science Studies

۵. The Cognitive Modeling Approach

۶. Humanistic

۷. Acting Humanly

۸. The Turing Test Approach

۹. John McCarthy

استنفورد^۱ دنبال می‌کرد. چکیده این ایده این است که ما اصلاً کاری به انسان نداریم و نمی‌خواهیم مثل انسان‌ها ماشین بسازیم. آنها رویکرد انسان‌گرایانه را کنار گذاشتند و تلاش کردند تا فرایند تفکر و استدلال در انسان را شبیه‌سازی کنند.^۲ آنها در پی آن بودند که زبانی رایانه‌ای اختراع کنند که بتواند علم منطق را شبیه‌سازی کند؛ دانش منطقی که هزاران سال است که مورد استفاده قرار می‌گیرد. آنها این هوش مصنوعی را به شکل جدید زبان لیست و پرولوگ ساختند و با این دو زبان تقریباً موفق شدند فرایند تفکر را در ماشین پیاده‌سازی کنند. شاید اولین روایت‌های هوش مصنوعی، که توانست در صنعت جواب دهد، همین رویکرد درجه سوم هوش مصنوعی بود که در رایانه محقق شد. از این رویکرد به «رویکرد قوانین تفکر» تعبیر می‌شود.^۳

۴. رویکرد چهارم: رویکرد آقای برکلی است. او بر این عقیده است که ما نه به انسان کار داریم و نه به انسانی بودن ماشین و نه به سینکینگ،^۴ بلکه می‌خواهیم ماشین‌هایی درست کنیم که عاقلانه کار کنند؛ بنابراین، عملکرد منطقی^۵ را پیشنهاد کرد. منطقی بودن^۶ به این معنا که سود و منفعت ماشین را، که اصطلاحاً به آن کارایی ماشین می‌گویند،^۷ بیشینه^۸ می‌کند؛ مثلاً ماشینی که برنامه پروازهای

۱. Stanford University

۲. Athinking Rationally

۳. The Law of Thought Approach

۴. Thinking

۵. acting rationally

۶. Rationality

۷. Performance

۸. Maximize

فرودگاه مهرآباد را بتواند به شکل منظم تنظیم کند. نود درصد متن‌های^۱ درسی هوش مصنوعی در دنیا روی این رویکرد، اصطلاحاً رویکرد «Weak AI»، یا هوش مصنوعی ضعیف تمرکز دارند و این خود ثمرات قابل توجهی دارد. از این رویکرد به «رویکرد عامل‌های هوشمند»^۲ تعبیر می‌شود.

بحثی پیرامون زیرشاخه‌های هوش مصنوعی

تعریف زیرشاخه‌های هوش مصنوعی بحثی مفصل است که اجمالاً می‌توان به مواردی از آن اشاره کرد:

۱. هوش مصنوعی تشخیص چهره و صحنه

ماشین‌های مجهز به این نوع از هوش مصنوعی می‌توانند تشخیص چهره دهند یا تشخیص اثر انگشت^۳ یا شبکه‌ی چشم افراد را تشخیص دهند. با شبکه‌های عصبی^۴ عمیقی که امروزه تولید شده تقریباً ماشین‌هایی شبیه انسان، بلکه شاید بهتر از انسان تولید شده که می‌توانند اشیاء داخل تصاویر را تشخیص دهند.

۲. رباتیک

یکی از زیرشاخه‌های هوش مصنوعی ربات‌های فیزیکی هستند که کارهای فیزیکی شبیه کارهای فیزیکی بدن انسان انجام می‌دهند.

۱. Text

۲. The Rational Agent Approach

۳. Fingerprint

۴. Neural Network

۳. یادگیری ماشین

یادگیری ماشین که دو شعبه دارد و پروژه آن بیست سالی است که در دانشگاه کرنگین مگان کلید خورده و الان هم به شکل عمیقی در حال بسط است و شعبه های جدیدی از آن انتشار پیدا می کند. شناسایی الگو،^۱ به عنوان یک زیرشاخه یادگیری ماشین، نیز داده کاوی^۲ و پردازش زبان طبیعی^۳ زیرشعبه هایی از یادگیری ماشین هستند. همچنین، از طریق هوش مصنوعی می توان تشخیص الگو داد: الگوهای داخل یک داده را کشف کرد به این معنا که الگوهای موجود در داده های متنی، تصویری و صوتی را تشخیص داد.

همچنین، پردازش زبان طبیعی که در آن ماشین می خواهد بفهمد که انسان چه می گوید. در این مورد، ماشین زبان انسان را در دو مرحله بازسازی می کند:

مرحله اول: درک زبان طبیعی؛^۴

مرحله دوم: تولید زبان طبیعی^۵ که ماشین در آن مثل انسان جمله تولید می کند؛ هم جملات را ترکیب و هم متن را خلاصه سازی می کند. همچنین، متن را از یک زبان به زبان دیگر ترجمه می کند. از کاربردهای زبان طبیعی (NLP) از پاسخ گویی به پرسش^۶ هم می توان یاد کرد.

۱. Pattern Recognition

۲. Data Mining

۳. Natural Language Processing(NLP)

۴. Pattern

۵. Natural Language Understanding(NLU)

۶. Natural Language Generation(NLG)

۷. Question Answering

یک سیستم عادی یادگیری ماشین اگر به هوش مصنوعی مجهز شود، می‌تواند در سطوح پیشرفته عامل خود یاد بگیرد؛ یعنی داده‌های ورودی را بگیرد و برای آن‌ها مدل طراحی کند و بر اساس آن مدل بگوید که داده‌ی^۱ آینده چطور خواهد بود یا چطور پیشرفت خواهد کرد. گرچه این مورد بسیار پیچیده است و عامل‌های سیاسی اجتماعی حول آن بسیار پیچیده است.

اما شیوه جدیدی از یادگیری وجود دارد که ترکیبی از با نظارت^۲ و بدون نظارت^۳ است که نوعی یادگیری تقویتی یا یادگیری تشویقی^۴ است. در این شیوه، ما به دنبال این هستیم که توصیف حاکم بر داده‌ها نشان دهیم. در این مورد، متخصصانی هستند که می‌توانند به ماشین بگویند که آیا یک فرآیند را خوب انجام داده است یا نه.

۴. سیستم‌های شناسایی گفتار^۵ یا گفت‌وگوی صوتی

این سیستم نوعی گفت‌وگو (چت) محاوره‌ای و صوتی است که صوت را تبدیل به متن و متن را تبدیل به صوت می‌کند. اکنون بیش از پنجاه درصد از جست‌وجوهای که در گوگل انجام می‌شود، از طریق گفتاری است. علاوه بر جست‌وجوهای صوتی، بازیابی‌های صوتی، تشخیص صدای یک فرد، تقلید صدای یک فرد و ساختن جملاتی با صدای یک فرد هم از امکانات این سیستم است.

۱. data

۲. Supervisor

۳. Unsupervisor

۴. Reinforcement Learning (RL)

۵. Speech Recognition System

زیربنای هوش مصنوعی از چند دانش بهره می‌گیرد: فلسفه و منطق و دانش‌های و روش‌های زیرمجموعه منطق، مانند روش‌های تحلیل زبانی که رمزگشایی از تفکرآتی است که انسان دارد و آن‌ها را در قالب قضایایی فلسفی و منطقی ارائه می‌کند. با اتکا به این روش‌های، انسان می‌خواهد حقیقت‌های جهان واقع را بفهمد، تا بر اساس آن این حقایق را ارائه کند. همچنین، از دانش زبان‌شناسی، علم معناشناختی و ریاضیات و الگوریتم‌ها.^۲ از همگی این موارد به عنوان زیربناهای اولیه هوش مصنوعی می‌توان بهره گرفت.

بحثی پیرامون هوش مصنوعی قوی

آقای مکس تگمارک،^۳ از دانشگاه ام‌آی‌تی آمریکا، در کتاب «لایف تری» خود سه‌گونه از حیات یا سه دوره از حیات را برای بشر و جامعه انسانی نشان می‌دهد: دوره اول: دوره‌ای که یک زیست‌شناس^۴ می‌توانست یک سخت‌افزار یا یک سلول را تکثیر کنند؛ مانند آنکه آمیب‌ها و آغازیان خودشان را تکثیر می‌کردند. دوره دوم: دوره فرهنگی^۵ خوانده شده که علاوه بر تولیدمثل سخت‌افزاری، می‌توان تولیدمثل فرهنگی و نرم‌افزاری هم داشت.

دوره سوم: دوره‌ای که از آن تعبیر به تکینگی^۶ می‌کنند. دوره‌ای که ماشین‌هایی می‌آیند و سخت‌افزارها و نرم‌افزارهایی تکثیر می‌کنند و بهتر از انسان‌ها

۱. Fact

۲. Algorithm

۳. Max Tegmark

۴. Biologist

۵. Cultural period

۶. Singularity

عملیات‌های مختلف انجام می‌دهند. در مورد این دوره، سؤال این است که آیا دست‌یابی به این حد از هوش مصنوعی قوی ممکن است یا نه. برخی کارشناسان چنین دوره‌ای را دستیافتنی و برخی غیردستیافتنی دانسته‌اند؛ اما شکی نیست که مسئله هوش مصنوعی و هوش مصنوعی قوی آنقدر مهم است که در امور مختلف، چون امور نظامی، اقتصادی و... ارتباط وثیقی با حکمرانی پیدا می‌کند.

بحثی پیرامون هوش مصنوعی و اجتهاد

یکی از کاربردهای هوش مصنوعی در حوزه تحقیقات اسلامی این است که یک دستیار برای فقیه ایجاد شود. این به آن معنا نیست که جانشینی برای مجتهد قرار داده شود، بلکه به این معناست که ابزاری قوی در خدمت یک خبره قرار گیرد. درحقیقت، آنچه در دست است متون اولیه‌ای است که فقیه به آن‌ها مراجعه می‌کند. می‌توان با استفاده از هوش مصنوعی به یادداشت‌ها و حواشی مفصلی که فقهای اولیه داشته‌اند روح داد تا ماشین‌ها بتوانند آن‌ها را درک کند و گسستگی‌ها^۱ و تلاقی‌ها^۲ را تشخیص دهند. تا اینکه مسیری طی شده برای پیدایش یک فتوا روشن شود. در این مورد، ماشین می‌تواند در تجمیع ادله و فسخ خودکار ادله، که مربوط به یک موضوع و مفهوم خاص است، کمک کند یا می‌تواند برای یک موضوع فقهی ادله لازم را جمع‌آوری کند.

خوانش استدلالی متون فقهی یکی دیگر از کاربردهای هوش مصنوعی در مسیر اجتهاد است؛ اینکه ماشین بتواند متون را بخواند، قسمت‌های مختلفش را از هم تفکیک کند و درخت‌واره مطالب آن متن را ترسیم کند. همچنین، یافتن

۱. Discontinuities

۲. Negotiations

مفروضات یک استدلال و اینکه آن استدلال چقدر به آن مفروضات مرتبط است، تحریر محل نزاع در مباحث فقهی و اینکه چقدر یک مجتهد نسبت به مبانی اصولی خود پایبند بوده است از دیگر کاربردهای می تواند باشد. همچنین، تشخیص اینکه اگر تغییر مبنای اصولی یا رجالی برای فقیه اتفاق افتاد، کدام فتوا را باید تغییر بدهد.

بحثی پیرامون هوش مصنوعی و الهیات

بارها از مقام معظم رهبری این حدیث مبارک حضرت امام علی (ع) را شنیدید که «الْعِلْمُ سُلْطَانٌ مَنْ وَجَدَهُ صَالٍ بِهِ وَ مَنْ لَمْ يَجِدْهُ صَيِّلٌ عَلَيْهِ»^۱. من به جای عبارت علم در کلام مقام معظم رهبری هوش مصنوعی را می گذارم. حاصل این خواهد بود که هوش مصنوعی منبع قدرت خواهد شد و هر کس این قدرت را به چنگ آورد، می تواند غلبه و اقتدار جهانی پیدا کند و کسی که به آن دست پیدا نکند، دیگران بر او غلبه خواهند. مقام معظم رهبری هر وقت از دانش بومی یا دانشجویی یا رتبه های علمی جهانی ایران صحبت می کند، پشتیبانی خود را از پیشرفت علمی به عنوان یک جهاد علمی اعلام می کند. بر اساس آیه «وَأَعِدُّوا لَهُمْ مَا اسْتَطَعْتُمْ مِنْ قُوَّةٍ»^۲ هوش مصنوعی، به عنوان یک سلاح علمی، در رأس همه دانش ها می تواند بزرگ ترین اقتدار را در اختیار هر جامعه ای قرار دهد.

من از تعبیر «العلم سلطان» در این روایت وام می گیرم و می گویم: هرچه قوی تر به هوش مصنوعی بپردازیم دانش بیشتری خواهیم داشت؛ سلطه بیشتری بر طبیعت پیدا خواهیم کرد و خدااباوری بیشتری خواهیم داشت. با اتکا به هوش

۱. شرح نهج البلاغه، ابن ابی الحدید، ج ۲، ص ۳۱۹.

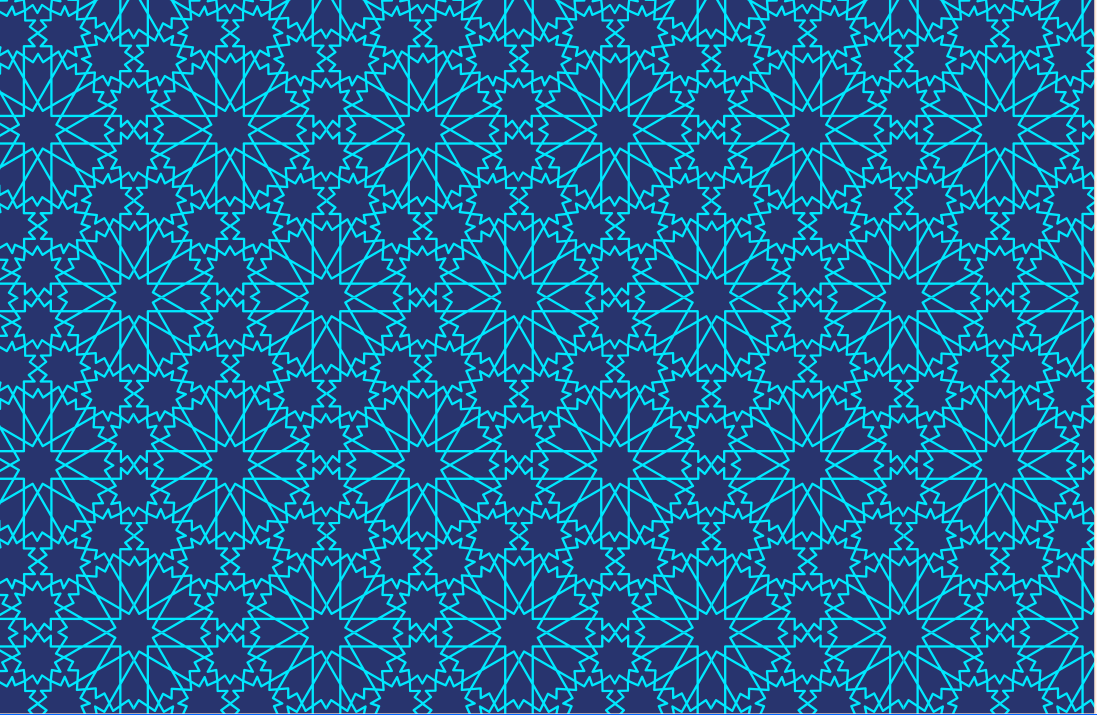
۲. سوره انفال، آیه ۶۰.

مصنوعی انسان، که اشرف موجودات است، دقیق‌تر و عمیق‌تر روابط طبیعت را می‌شناسد و بهتر خواهد توانست طبیعت را شبیه‌سازی^۱ کند. در حقیقت، پیشرفت در هوش مصنوعی تلاقی وثیقی با خداباوری دارد و خداباوری را بیشتر محقق می‌کند.

اگرچه به قول شهید مطهری هیچ‌وقت در دنیا به این اندازه علم اسیر زورمندان و قدرت‌مداران با سوءاستفاده از علم و دانش نبوده، از منظری الهی کسی برنده نهایی است که سمت و سویی الهی داشته باشد. در این زمینه، همانطور که می‌توانیم ربات راننده، ربات قاضی و ربات وکیل داشته باشیم، ربات «کلرجی من»^۲ داشته باشیم؛ رباتی که وظایف مختلف یک روحانی را انجام می‌دهد، هم از بعد پژوهش، هم از بعد تدریس و هم از بعد انجام شئون قضایی و تبلیغی. اگر چنین ماشین‌هایی درست شوند، می‌توانند به بازآفرینی شئون حوزه در فضای حاکم بر دنیایی که به هوش مصنوعی مجهز شده کمک کنند.

۱. Simulation

۲. ClergyMan



«هوش مصنوعی و دین»

دکتر یوریک ویلکس^۱

۱. استاد دانشگاه شفیلد، انگلستان.





مقدمه

دو موضوع «هوش مصنوعی»^۱ و «دین»^۲ غالباً با هم مطرح نمی‌شوند، اما ارتباطاتی می‌ان آن‌ها وجود دارد. برخی از این ارتباطات از لحاظ تاریخی به گذشته‌های دور برمی‌گردد. در این مبحث، به دنبال نشان دادن ارتباط می‌ان آن‌ها خواهیم بود؛ ابتدا به نظریه سایبرنتیک^۳، که به عبارتی خویشاوند شرم‌آور هوش مصنوعی است، پرداخته می‌شود و چند کلمه‌ای به عنوان مقدمه راجع به الهیات مرتبط با علم لایتناهی الهی و مفهوم خیرخواهی مطرح می‌شود. در ادامه، به سه بخش پرداخته می‌شود و درباره دوره‌ای به نام دوره ماشین‌های احساسی مباحثی طرح خواهد شد؛ منظور از این دوره، مکتب رمانتیسم^۴ قرن نوزدهم و پندارهای رومانتیک‌ها درباره ساختن ماشین‌های شبه انسان، تقویت کردن انسان‌ها و ماشین‌های شبه خداست. هر سه موضوع امروزه در دنیای فناوری و اندیشه موضوعاتی بسیار مرتبط و اندیشه‌برانگیز هستند. در پایان، به موضوع نسبتاً متفاوتی تحت عنوان ماشین‌ی کردن^۵ رسوم دینی خواهیم پرداخت.

۱. Artificial intelligence

۲. Religion

۳. Cybernetics

۴. Romanticism

۵. Automation

باورهای دینی

لودویک ویتگنشتاین^۱ می‌گوید: «من خود را انسانی مذهبی نمی‌دانم، اما ناگزیرم به هر مسئله‌ای از دیدگاه مذهبی نگاه کنم». این تصوّر نسبتاً عجیبی است: یک روحانی به نام «بیس»، که قضیه مطرح شده سوی او زیربنای باورهای مرتبط با یادگیری ماشینی و هوش مصنوعی مدرن را تشکیل می‌دهد، این قضیه را مطرح کرد تا از احتمال وقوع معجزه در برابر استدلال‌های هیوم^۲ در قرن هجدهم دفاع کند. گودل^۳، که مسلماً یکی از برجسته‌ترین منطق‌دانان قرن بیستم به‌شمار می‌آید، سال‌های پایانی عمر خود در دانشگاه پرینستون^۴ را صرف پژوهش درباره شواهد هستی‌شناختی دال بر وجود خدا کرد. از سوی دیگر، دنیل دنت^۵، که یکی از ملحدان متخاصم مشهور بود، تألیفات متعددی درباره سردرگمی خویش درباره چرایی بقای باورهای دینی تحت سیطره فرضیه تکامل یا فرگشت^۶ به‌نگارش در آورده است. او بر این باور است که این اعتقادات مسلماً تکامل‌های کاذب هستند که نظریه تکامل باید از دست آن‌ها خلاص می‌شد؛ اما به دلایل نامعلومی این اتفاق رخ نداده است. او می‌خواهد بداند که ارزش بقای عقاید دینی چقدر است که در ادامه به این مطلب خواهیم پرداخت.

۱. Wittgenstein

۲. David Hume

۳. Gödel

۴. Princeton

۵. Daniel Dennett

۶. Evolution

نظریه سایبرنتیک

سایبرنتیک^۱ در دهه‌های ۱۹۵۰ و ۱۹۶۰ می‌لادی نسل پیشین هوش مصنوعی بود که به انواع گوناگونی از اشیاء می‌پرداخت. سایبرنتیک همان چیزی بود که آن را ریاضیات پیوسته، نه ریاضیات گسسته مانند محاسبات رایانه‌ای، می‌نامیم. سایبرنتیک نقطه مقابل بازنمایی در رایانه‌ها بود. سایبرنتیک دارای محاسبه آنالوگ و محاسبه به‌وسیله ابزارهای فیزیکی بود؛ به جای آنکه مثل محاسبه دیجیتال از بیت‌های یک و صفر استفاده کند. سایبرنتیک به مدل‌سازی حیوانات و حشرات علاقه داشت و بر مبنای الگوهای مغز و شبکه‌ها استوار بود و مفاهیم کلیدی آن بازخورد و یادگیری بود. یکی از مشهورترین متخصصان سایبرنتیک در بریتانیا^۲ گری والتر^۳ بود که لاک‌پشت‌هایی مکانیکی اختراع کرده بود. او فیلم‌هایی از این لاک‌پشت‌ها نشان می‌داد که در یک اتاق به این طرف و آن طرف حرکت می‌کردند و به دنبال پریز برق می‌گشتند تا بتوانند خود را به آن متصل کنند. آن‌ها به دنبال این بودند که یاد بگیرند که چگونه این کار را انجام دهند و به آن ادامه دهند. اما زمانی که در دهه ۱۹۶۰م هوش مصنوعی به‌صحنه آمد، شاید بتوان گفت که سایبرنتیک کنار گذاشته شد.

هوش مصنوعی

هوش مصنوعی سنتی، که بین سال‌های ۱۹۵۰ و ۱۹۹۰ می‌لادی، سلطه داشت، از هر جهت با سایبرنتیک متفاوت بود. هوش مصنوعی سنتی بر مبنای بازنمایی‌ها و کاربرد منطق و سخت‌افزار دیجیتال استوار بود و به هوش مدل‌سازی شده‌ای که

۱. Cybernetics

۲. United Kingdom

۳. William Grey Walter

نسبتاً از مغز حیوانات، انسان‌ها و... مستقل بود می‌پرداخت. این توصیفی بود که جان مک‌کارتی^۱ در دانشگاه استنفورد^۲ از هوش مصنوعی ارائه کرد. هوش مصنوعی به آمار علاقه‌ای نداشت و ریاضیات پیوسته نبود. هوش مصنوعی عبارت بود از روش‌های دیجیتال که منطق آن‌ها را تا حد زیادی خلاصه می‌کرد و این تصور کلی درباره هوش مصنوعی است. البته زمانی که نوبت به ظهور یادگیری ماشین در دهه ۱۹۹۰م رسید این تصور کلی مجدداً رد شد. در ادامه، اشاره می‌شود که چگونه یادگیری ماشین باعث محو شدن هوش مصنوعی نمادین شد و به منزله بازگشت به باورهای مرتبط با سایبرنتیک در نظر گرفته شد.

سایبرنتیک و دین

دلیل مطرح کردن نظریه سایبرنتیک در بحث ارتباط هوش مصنوعی با دین این است که تفکر سایبرنتیک درباره دین، به طور صریح‌تر و ساده‌تر از نمونه‌های قبلی خود، درباره هوش مصنوعی حرف‌هایی برای گفتن داشت. در ضمن، سایبرنتیک تأثیر فراوانی بر حیات فکری در اروپا، به خصوص در کشور فرانسه، بر روی اندیشمندانی، از جمله لیوتار، گذاشت. آن‌ها شدیداً تحت تأثیر سایبرنتیک قرار گرفتند، هرچند که ارائه تصویری از تصور آن‌ها از سایبرنتیک اکنون دشوار به نظر می‌رسد؛ چرا که این نظریه در این کشور تقریباً ناپدید شده است.

در بریتانیا، متخصص سایبرنتیک مشهوری بود به نام استفورد بیر که کم‌وبیش رهبری این تفکر را بر عهده داشت؛ انسانی بسیار جالب که تا اواخر قرن بیستم به حیات خود ادامه داد. او در سال ۱۹۶۶م مقاله بسیار عجیبی تحت عنوان دانش

۱. John McCarthy

۲. Stanford University

خداوند نوشت و در آن از نظریه‌ای تحت عنوان جسم‌جان‌انگاری حمایت کرد. پَن‌سایکیزم^۱ مفهومی است به این معنا که همه چیز در عالم ذهن است و همه چیز به عبارتی ذهنی است. این باوری بسیار سنتی است و اغلب ایده‌آلیسم^۲ نامیده می‌شود. استافورد بیر^۳ نسخه ویژه‌ای از جسم‌جان‌انگاری ارائه کرد که بر اساس آن همه چیز زنده است. از نظر او، سایبرنتیک دقیقاً این دیدگاه را درباره جهان خلاصه می‌کرد که همه چیز زنده است! بنابراین سایبرنتیک برای او سیستمی از جعبه‌های سیاه بود که نمی‌توانست در یک مکانیسم قرار بگیرد تا بتواند خود را با جهان سازگار کند، آن را یاد بگیرد و در آن کارکردی داشته باشد، درست مانند لاک‌پشت‌های گری والت. تصور اصلی استافورد بیر این بود که این جهان اساساً غیر قابل درک است. این تصور از نظر فلسفی به مارتین هایدگر^۴ مرتبط است، اما اینجا بحث درباره متخصصان سایبرنتیک بریتانیایی است. بیر معتقد بود که این جهان اساساً غیر قابل درک است. این ذهنیت با آنچه که ما علم متعارف می‌نامیم کاملاً متضاد است. فرض علم بر این است که جهان قابل درک است و علم به ما اجازه می‌دهد تا آن را درک کنیم. ولی از نظر بیر این‌گونه نبود. می‌ان کسانی که باور دارند جهان را می‌توان کاملاً درک کرد و کسانی که به قابل درک بودن آن اعتقاد ندارند تضادی وجود دارد. این دیدگاه‌ها با دو رویکرد درباره ارتباط دین با هوش مصنوعی مطابقت دارند.

۱. فرضیه پَن‌سایکیزم (Panpsychism) بسیار به جان‌بخشی نزدیک است؛ با این تفاوت که پَن‌سایکیزم ضرورتاً قائل به زنده بودن و وجود داشتن تمامی فعالیت‌های ذهنی در اشیاء بی‌جان نیست، بلکه آنها را موجوداتی هوشیار می‌داند.

۲. Idealism

۳. Stafford Beer

۴. Heidegger

یادگیری ماشین

یادگیری ماشین مدرن،^۱ موضوعی که در روزنامه‌ها راجع به آن مطلب می‌خوانید، در ارتباط با هوش مصنوعی است. این موضوع نیز، درست مانند سایرینتیک، راجع به مدل‌هاست؛ اما به اندازه کافی درباره چگونگی پدیده‌ها و تبیین آن‌ها توضیحی نمی‌دهد، بلکه صرفاً آن‌ها را مدل‌سازی می‌کند و پیش‌بینی می‌کند که در آینده چه کار خواهند کرد. بنابراین از یک نظر می‌توان گفت که سایرینتیک، که تحت تأثیر هوش مصنوعی کنار گذاشته شده بود، دوباره ظهور پیدا کرده است.

الهیات سنتی

انسان‌ها برای مدتی طولانی راجع به معنای همه چیز دانی و علم مطلق تفکر کرده‌اند. همه چیز دانی که اغلب علم لایتناهی نامیده می‌شود یکی از صفات شناخته شده و اصیل خداوند در الاهیات سنتی به‌شمار می‌آمده است. در قرن نوزدهم، لاپلاس، فیزیک‌دان، به موجودی هوشمند فکر می‌کرد که طبق فرض مسلم او موقعیت و سرعت همه اتم‌های عالم را می‌دانست. از این جهت این موجود هوشمند همه چیز را می‌دانست. او موقعیت و سرعت هر چیزی را که عالم شامل آن می‌شد را می‌دانست. پس تصور لاپلاس درباره دانستن همه چیز این بود دانستن چیزی بود که امکان محاسبه آن با هیچ روشی، که در حال حاضر متصور هستیم، وجود نداشت. هنگامی که افراد راجع به دانستن همه چیز سخن می‌گویند، معمولاً منظور آن‌ها آن چیزی نیست که لاپلاس درباره اتم‌ها می‌گفت. نکته بعدی این است که آیا چیزی که همه چیز را می‌داند و دانای مطلق است

۱. Modern machine learning

می‌تواند هوشیار باشد. برای پاسخ به این سؤال آشنایی با مفهوم هوشیاری لازم است. اگر چیزی که همه چیز می‌داند و هوشیار هم باشد وجود داشته باشد، می‌تواند عواقب نسبتاً عجیبی برای خدا و احتمالاً هرگونه ماشین شبه‌خدا داشته باشد. هنگامی که به هوشیاری فکر می‌کنیم معمولاً به توجه و تمرکز نسبت به یک چیز فکر می‌کنیم. یکی دو سال پیش، برنامه‌ای در تلویزیون پخش می‌شد با عنوان "سال‌ها و سال‌ها" که در آن شخصیت دختری ترسیم شده بود که عمل پیوند مغز انجام داده بود. او پس از عمل، می‌توانست نسبت به هر چیزی در جهان همیشه آگاهی داشته باشد؛ می‌گفت که به همه گدایان شهر دهلی آگاهی دارد! یک لحظه به چنین ادعایی فکر کنید: چگونه می‌توان از وجود همه گدایان شهر دهلی آگاهی داشت! آیا وجود چنین موجودی با عقل جور در می‌آید؟ آیا انسان می‌تواند از همه چیز آگاه باشد؟

یکی از بزرگ‌ترین فلاسفه، و احتمالاً بزرگ‌ترین فیلسوفی که اعتقاد داشت جواب این پرسش‌ها مثبت است، لایبنیتز^۱ بود. در قرن هفدهم، لایبنیتز تصور خود از جهان را صورت‌بندی کرد. بر اساس نظر او، جهان از عناصری تشکیل شده است که مونا^۲ (ذره زنده) نامیده می‌شوند و برای درک آن‌ها نیازی نیست به خود زحمت داد. شما مونا محسوب می‌شوید؛ این رایانه نوعی مونا است. اما از نظر او خدا مونا برتر است. همه موناها از نظر لایبنیتز کمی هوشیار بودند و از آگاهی اندکی برخوردار بودند. از نظر او، خدا برترین مونا بود که از همه چیز آگاهی داشت. خدا از محتوای همه موناها آگاهی داشت و ظاهرآ دیدگاه او همه دیدگاه‌ها را در بر می‌گرفت. این راهکاری بود که لایبنیتز سعی می‌کرد از طریق آن

۱. Gottfried Wilhelm Leibniz

۲. Monad

علم فراگیر خدا را اثبات کند. از نظر او، یک چیز شبه خدا می‌توانست از همه چیز آگاهی داشته باشد. چنین چیزی از نظر او مونا د شبه خدا بود.

یکی از بزرگ‌ترین فلاسفه قرن نوزدهم هگل بود. هگل عقیده عجیبی داشت که هنوز هم معتقدانی دارد: اینکه ما بخش هوشیار عالم هستیم؛ ما تنها بخش عالم هستیم که هرچیزی را می‌داند و دانای مطلق است. از این رو، انسان تا حدی عالم را بازنمایی می‌کند، چون به عنوان موجودی خودآگاه پا به عرصه وجود گذاشته است. در اینجا این سؤال پیش می‌آید که آیا شبکه جهانی اینترنت بخشی از عالم بشریت است که هوشیاری پیدا کرده است. در اینجا دو رویکرد امکان هوشیاربودن و یا نبودن چیزی که دانای مطلق است مطرح می‌شود.

نیک باستروم^۱ در دانشگاه آکسفورد^۲ کتاب مشهوری را با عنوان هوش برتر به‌نگارش درآورده و در آن استدلال می‌کند که ممکن است یک هوش برتر وجود داشته باشد. موجودی شبه خدا که به طور گریزناپذیری ما را نابود خواهد کرد. او بدخواه انسان خواهد بود و خیرخواه او نخواهد بود. این دیدگاهی بسیار عجیب است: چرا که انسان‌ها همیشه به این مسئله فکر کرده‌اند که آفریدگار آن‌ها احتمالاً خیرخواه بوده است. اطلاعات عهد عتیق در رابطه با خیرخواهی خداوند اندکی نامستدل دانسته شده است. ما انسان‌ها همواره فرض کرده‌ایم که خداوند مهربان است و همیشه این گزاره را به خوبی پذیرفته‌ایم. اما دیدگاه باستروم دیدگاهی است که با این فرض عمومی در تعارض است. باستروم معتقد است که هوش برتری که ما ممکن است آن را خلق کنیم نسبت به ما نامهربان و بدخواه خواهد بود؛ گفته شده که این فرض بسیار بعید است: چرا نباید خیرخواه ما باشد و تداوم بخش

۱. Nick Bostrom

۲. University of Oxford

همان سنت قدیمی ای باشد که ادیان همواره از آن یاد کرده‌اند؟ در مطالب آتی به مفهوم خیرخواهی خواهیم پرداخت.

الهیات دوره رمانتیک و فرا بشریت (ترابشریت)^۱

جنبش مدرن ترابشریت به این باور اشاره می‌کند که شما می‌توانید انسان‌ها را با استفاده از فناوری توانمندتر سازید. این خود حاصل تغییر و تحول وضعیت بشری با توسعه و ساخت فناوری‌های پیشرفته است که به طور گسترده به منظور تقویت عقل و فیزیولوژی بشری در دسترس هستند. بخشی از این تحول فقط به قسمت‌های بدن مربوط است. تشکیلات دفاعی بسیاری از کشورها در حال حاضر، با صرف هزینه‌های بالا، تلاش می‌کنند تا سربازانی عالی را بسازند که اعضای بدن تقویت شده‌ای دارند؛ اعضای که آن‌ها را در مقایسه با سربازان عادی سریع‌تر و قوی‌تر می‌سازد. اتوماتیک کردن کارها در کارخانه‌ها ممکن است چنین نقشی را ایفا کند؛ به این شیوه که انسان‌ها را در قالبی قرار دهند که استخوان بندی بیرونی نامیده می‌شود و عبارت است از اسکلت‌هایی که دور انسان‌ها را می‌پوشاند تا انسان بتواند، به عنوان مثال، وزنه‌ای یک‌تونی را از زمین بلند کند.

نسخه دیگر این‌گونه تغییر و تحول بارگزاری مغز نامیده می‌شود. بر مبنای این نسخه ظاهراً می‌توان با بارگزاری کل مغز انسان‌ها در درون یک ماشین آن‌ها را فنا ناپذیر ساخت. با این کار محتوای مغز از بین نخواهد رفت، در حالی که جسم آن‌ها نابود خواهد شد. برخی افراد ممکن است با این کار موافق نباشند، اما مسلماً افرادی هستند که تمایل به انجام آن داشته باشند. این همان مفهوم هستی دیجیتال فنا ناپذیر در یک محیط مصنوعی است که ترابشریت نامیده

۱. Transhumanism

می‌شود. این جنبش ریشه در گذشته‌های دور و دیدگاه‌های رایج رومانتیک‌های قرن نوزدهم دارد. احتمالاً همه شما درباره «ابر انسان» نیچه شنیده‌اید که موجودی تحول‌یافته و برتر از ما بود. شعر مشهوری که مولر آن را سروده است و عبارت شگفت‌انگیزی دارد که معنای آن این است که اگر هیچ خدایی به زمین نیاید ما خود خدا خواهیم بود. بنابراین، ترابشریت مدرن یک اصل علمی واقعی است و متخصصان فناوری متعددی به آن اعتقاد دارند. این فناوری تلفیقی است از نامیرایی، کمال و پیوند با ماشین‌های هوشمند که محکوم به فنا نیستند.

ترابشریت و هوش مصنوعی

آی‌جی‌گود ریاضی‌دانی بریتانیایی بود که فعالیت‌های کدگذاری بسیاری در جنگ جهانی دوم انجام داد. او در سال ۱۹۶۵م جمله خارق‌العاده‌ای گفت که از یک نظر، و از دیدگاه بسیاری از افراد، منجر به پیدایش جنبش ترابشریت شد. او نوشت: «اجازه بدهید یک ماشین فراهوشمند به عنوان ماشینی تعریف شود که می‌تواند از همه فعالیت‌های فکری هر انسانی فراتر رود، صرف‌نظر از اینکه آن انسان تا چه حدی باهوش باشد». از آنجایی که طراحی ماشین‌ها یکی از این فعالیت‌های فکری است، یک ماشین فراهوشمند حتی می‌تواند ماشین‌های بهتری را طراحی کند و هوش انسان تا حد زیادی از آن عقب خواهد افتاد. بنابراین، اولین ماشین فراهوشمند آخرین اختراعی است که بشر نیاز به ساختن آن دارد.

کرزویل^۱ که یکی از شخصیت‌های مهم در زمینه هوش مصنوعی است، به

۱. Raymond Kurzweil

همراه افرادی، از جمله هانس مراوک،^۱ هیروشی ایشی‌گورو^۲ و کوین وارویک،^۳ در حمایت از «نظریه تکینگی» استدلال کرده‌اند. منظور آن‌ها از تکینگی لحظه‌ای است که در آن اولین ماشین فراهوشمند به وجود خواهد آمد، با انسان‌ها در هم خواهد آمیخت و ترابشریت را به وجود خواهد آورد. نیک باستروم در آکسفورد، که قبلاً به او اشاره شد، کتاب مشهوری درباره هوش برتر به نگارش درآورده است. او انجمن جهانی ترابشریت را در سال ۱۹۹۸ بنیان‌گذاری کرد. کرزویل از تکینگی فراتر رفت و اظهار داشت: «هوشی که پدید خواهد آمد به بازنمایی تمدن بشری ادامه خواهد داد». او احساس می‌کند که ماشین‌های آینده از جنس بشر خواهند بود، حتی اگر بیولوژیکی نباشند. این تفاوت بسیار مهمی می‌ان رویکرد او و باستروم است. از نظر کرزویل، این ماشین فراهوشی بهبودیافته هنوز از یک نظر اساساً بشری است؛ حال آنکه ماشین‌های نیک باستروم به عبارتی ضدبشری هستند و با انسان خصومت دارند. از جنس انسان ساخته نشده‌اند و انسان‌هایی تقویت شده نیستند. پس در نظریه ترابشریت دو بخش وجود دارد: بخش کاملاً مصنوعی و بخش در هم آمیخته با انسان‌ها.

هیروشی ایشی‌گورو در ژاپن از شهرت فراوانی برخوردار است، چون رباتی را اختراع کرده که دقیقاً شبیه خود اوست. او این ربات را در نمایشنامه‌ای که خود نوشته بود روی صحنه تئاتر برد. از یک نظر او سعی می‌کند که صرفاً از طریق ساختن رباتی شبیه خود به این نوع از فناپذیری، که ما مستقیماً راجع به آن صحبت می‌کنیم، دست پیدا کند. این اختراع کاملاً جدی نیست، اما با این حال نسبتاً شگفت‌انگیز است.

۱. Hans Moravec

۲. Hiroshi Ishiguro

۳. Kevin Warwick

الهیات دوره رومانتیک و ماشین‌های شبه خدا

یکی دیگر از تصورات غالب در دوره رومانتیک تصور خلق خدایان و «ماشین‌های شبه خدا» بوده است. البته همیشه روش‌های سنتی و کهنه برای ساخت خدایان و ماشین‌های شبیه به آن‌ها وجود داشته است: از جمله، امپراطورهای روم باستان که پس از مرگشان خدا نامیده می‌شدند، گوساله ساخته شده از زیورآلات توسط هارون،^۱ بعل (خدای زمین) و خدای آهن در عهد عتیق و گولم که به دست پراگ ساخته شده بود. دیدگاه ساخت ماشین‌های شبه خدا به گذشته‌های بسیار دور برمی‌گردد، ولی همه ادیان ابراهیمی، مسیحیت، یهودیت و اسلام، با آنچه گفته شد مخالفند. این ادیان، این مخالفت را از عهد عتیق گرفته‌اند که همه انواع بازنمایی از خدا را منع کرده است. عبارتی که در عهد عتیق آمده، و در واقع نخستین فرمان از ده فرمان است، عبارت است از: «هیچ‌گونه تمثال تراشیده شده‌ای را خلق نکنید».

ساختن تصاویر یا بازنمایی‌هایی از خدایان، پیامبران یا انسان‌ها در اسلام نیز حرام است. در می‌ان پروتستان‌ها نیز سنتی قدیمی مبنی بر شکستن مجسمه‌های مذهبی وجود دارد، درست مانند همان چیزی که در شرق بت‌شکنی، یعنی شکستن تمثال خداوند، نامیده می‌شود. با این وجود، انسان‌ها تمایل دارند به

۱. این ادعای درباره ساخت گوساله توسط هارون، بنا بر مطالب تورات و تحریف شده است. به‌طور نمونه در تورات آمده که وقتی آمدن موسی علیه السلام از طورسینا طول کشید، مردم نزد هارون علیه السلام رفتند و به او گفتند که برای ما خدایی بساز تا ما را هدایت کند. هارون گفت طلای گوشواره‌های زن و دخترانتان را نزد من بیاورید. هارون آنها را گرفت و در قالب ریخت و مجسمه گوساله‌ای را ساخت. مردم با دیدن این گوساله گفتند که این همان خدایی است که شما را از مصر بیرون کرد (تورات، سفر خروج، فصل ۳۲). اما بنابر آموزه دین اسلام و کتاب آسمانی تحریف نشده قرآن کریم، هارون وزیر و جانشین حضرت موسی علیه السلام و پیامبر خداوند بوده است. در قرآن کریم می‌فرماید که آن کسی که گوساله را ساخت، مردم را گمراه کرده و به پرستش آن دعوت کرد فردی به نام سامری بوده است. حضرت هارون علیه السلام آنها را از این کار نهی کرد و به پیروی از خداوند دعوت کرد، ولی مردم حرف او را گوش نکردند. برای نمونه در آیات ۸۵ تا ۹۷ طه این مطلب به‌طور آشکار آمده است.

ساختن این گونه تمثال‌ها ادامه دهند. قبلاً درباره هوش برتر، که باستروم پیش‌بینی کرده بود، صحبت شد. در این مورد، باید این سؤال را مطرح کرد که آیا انسان‌های تقویت‌یافته موارد مناسبی برای تبدیل شدن به خدایان مصنوعی هستند. این چیزی است که کرزویل پیش‌بینی می‌کرد: کرزویل بر این عقده پافشاری می‌کرد که هوش‌های برتر با انسان‌ها در هم آمیخته خواهند شد؛ اما باستروم بر این عقیده بود که آن‌ها یکتاپرست، غیر بشری و بدخواه انسان خواهند بود. استدلال او برای نظر جالبش این است که آن‌ها یکتاپرست خواهند بود؛ باستروم استدلال عجیبی برای این ادعا دارد مبنی بر اینکه اگر هوش‌های برتر وجود داشته باشند، یکی از آن‌ها همه را از بین خواهد برد و تنها همان یکی از آن‌ها باقی خواهد ماند. بنابراین، باستروم سنت یکتاپرستی را حفظ می‌کند.

در مورد وینر و نظریه سایبرنتیک این پرسش مطرح می‌شود که تصویر این ماشین شبه‌خدا چگونه است و تصویر او از ما چگونه است. فان نیومان، که یکی دیگر از چهره‌های برجسته و از بنیان‌گذاران رایانه و هم‌عصر وینر بود، اظهار داشت که ظاهراً ماشین شبه‌خدا چیزی خواهد بود که ما ممکن است آن را خلق کنیم، اما احتمالاً درک نکنیم که چه کار کرده‌ایم! آینده‌ای که کرزویل پیش‌بینی کرد تبدیل هوش مصنوعی به مبنایی برای دین است؛ به این معنا که هوش برتر را می‌توان به منزله خدا پرستش کرد. این تصور خارق‌العاده که هوش مصنوعی ممکن است منجر به پیدایش دین از طریق خلق یک خدا شود، مسلماً در کالیفرنیا^۱ یا مکان‌هایی شبیه به آن در حال گسترش است. البته این یک احتمال فناورانه فوری نیست.

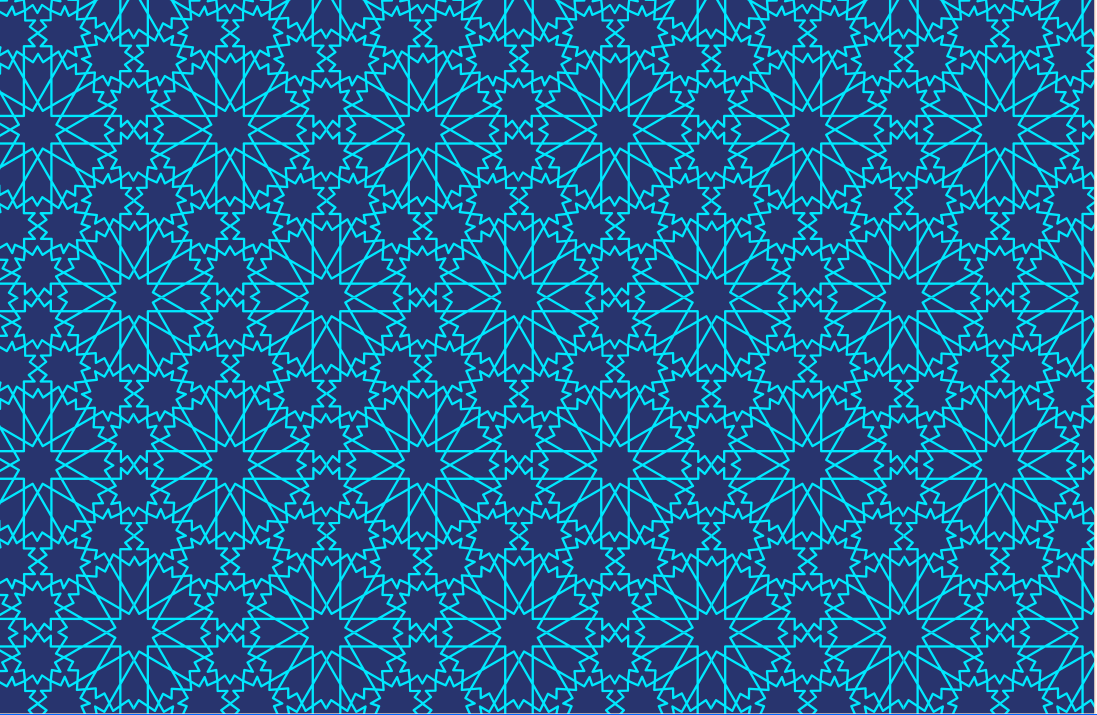
۱. California

در مقابل، استافورد بیر دانشمندان سنتی را در نقطه مقابل طرفداران نظریه سایبرنتیک قرار می‌دهد. استافورد بیر شصت سال پیش از این گفت: «برای افرادی که در یک سنت خوب لیبرال پرورش یافته‌اند انسان اصلی است که به طریقی بی‌نهایت عاقل است. او دانش و معرفت را تا مرز نهایی آن، یعنی دانشمند شدن، دنبال می‌کند و برای طرفداران سایبرنتیک انسان بخشی از سیستم کنترل است». این دیدگاه را می‌توان با نظریه سایبرنتیک متضاد دانست. سایبرنتیک از یک نظر ناشناخته‌ها را تقدیس می‌کند؛ ناشناخته‌ها بخشی از یک سیستم بودند که ما به طور کامل آنها را درک نمی‌کنیم. اگر ما به شبکه جهانی اینترنت فکر کنیم، متوجه می‌شویم که این شبکه، دانستن همه چیز را ممکن می‌سازد. خود ما به نحوی بخشی از این شبکه هستیم و دیگر به طور کامل از ما مجزا نیست، بلکه بخشی از زندگی ما را تشکیل می‌دهد. ما در قسمت نتیجه‌گیری این موضوع را تا حدی دنبال خواهیم کرد.

منحصر به فرد بودن بشر در برابر هوش مصنوعی

عمده مطالبی که در بالا ذکر شد به جای آنکه از منحصر به فردی بشر حمایت کند آن را تهدید می‌کند. در پایان، به اندیشمندانی اشاره می‌شود که تا حدی قصد دارند از منحصر به فرد بودن بشر دفاع کنند. این همان چیزی است که سنت‌های دینی همیشه قصد انجام دادن آن را داشته‌اند. لاورنس نظر جالبی را مطرح کرده مبنی بر اینکه منحصر به فرد بودن بشر در این واقعیت نهفته است که ما انسان‌ها ابزارهای تولیدکننده هستیم و دلیل آن ساختار گلو و زبان ماست. ما فقط می‌توانیم زبان را بسیار آهسته، بر اساس تولید سیگنال‌های بسیار ظریف و پراکنده‌ای از واژگان، تولید کنیم. حال آنکه یک ماشین می‌تواند می‌لیون‌ها داده‌ها را ظرف یک ثانیه تولید کند. بر اساس استدلال او، این مقایسه نشان می‌دهد که

ما منحصر به فرد هستیم و زبان ما منحصر به فرد است؛ به دلیل این که زبان چیزی است که ظاهراً به وسیله آن می توانیم معنا را مجزا از داده ها بیان کنیم. این عقیده لاورنس، که منحصر به فرد بودن انسان را به ابزارهای تولید کننده او ارتباط می دهد، به این معناست که زبان بخشی از عوامل منحصر به فرد بودن ماست، که البته این نظر نظریه جدیدی نیست. در طول تاریخ، بسیاری از متفکران، از جمله ارسطو، گفته اند که زبان یکی از ویژگی هایی است که باعث تمایز انسان ها می شود و هوش مصنوعی از این جهت قابل مقایسه با انسان نیست.



« هوش مصنوعی قوی و امکانش در مباحث الهیاتی »

دکتر جی پی مورلند^۱

۱. عضو هیأت علمی دانشگاه بایولا، کالیفرنیا.





مقدمه

بحث آتی، در مورد هوش مصنوعی قوی و امکان آن در مباحث الهیاتی، تبعاً الهام گرفته از فلسفه الهیاتی غرب^۱ است. وقتی به رابطه بین خداباوری و علم فکر می شود، بخش مهمی از گفت و گو پیرامون چگونگی تأثیر هوش مصنوعی بر خداباوری خواهد بود. هوش مصنوعی اولاً می تواند خداباوری^۲ را در دو حوزه اساسی اخلاقی و هستی شناختی تحت تأثیر قرار دهد. در اینجا، در مورد مسائل اخلاقی سخن گفته نخواهد شد، بلکه به برخی از سؤالات هستی شناختی، که هوش مصنوعی مطرح می کند، پرداخته می شود. در این بحث، بر خداباوری و تشخیص انسان متمرکز خواهیم شد؛ این که چگونه هوش مصنوعی بر دیدگاه ما نسبت به انسان تأثیر می گذارد.

هوش مصنوعی قوی

برای درک تأثیر هوش مصنوعی بر دیدگاه های خداباورانه انسان باید با درک هوش مصنوعی قوی^۳ (SAI) شروع کرد. اولین چیزی که باید در نظر داشت این

۱. Western theological philosophy

۲. Theism

۳. Strong Artificial Intelligence(SAI)

است که هوش مصنوعی قوی به این معنا است که نسبت ذهن^۱ به مغز همان نسبت برنامه نرم افزاری به سخت افزار در یک رایانه است. سه نکته این مطلب را توضیح می دهد:

اولاً، هوش مصنوعی قوی نشان می دهد که مغز یک رایانه دیجیتالی است.

ثانیاً، ذهن یا آگاهی باید به عنوان یک برنامه نرم افزاری در نظر گرفته شود.

ثالثاً، باور به هوش مصنوعی قوی مستلزم این است که رایانه ها روزی در حل مسئله، توانایی های شناختی، کارایی و سرعت یادگیری از انسان پیشی خواهند گرفت.

بنابراین، نکته کلیدی این دیدگاه این است که مغز به عنوان سخت افزار رایانه در نظر گرفته می شود و آگاهی یا ذهن نیز نرم افزاری است که مغز آن را به صورت یک برنامه اجرا می کند.

علم از طریق فلسفه با الهیات و خدا باوری در تعامل است و علم به طور مستقیم با الهیات تعاملی ندارد. بنابراین، اکنون باید درک کنیم که هوش مصنوعی قوی در علم به چه چیزی در فلسفه ذهن منجر می شود. آنچه می بینیم این است که هوش مصنوعی قوی معادل کارکردگرایی ماشینی در فلسفه ذهن است.

مفهوم «انواع ذهنی»

برای درک کارکردگرایی ماشینی نخست با ماهیت «انواع ذهنی» آشنا می شویم: «نوع ذهنی» حالتی ذهنی است مانند حس درد یا اندیشیدن به این حقیقت که ۲ + ۲ مساوی با ۴ است. این حالت ذهنی را بسیاری از مردم می توانند همان زمان در ذهن داشته باشند. بنابراین، یک حالت ذهنی مانند یک مفهوم کلی است که

۱. Mind

بسیاری از افراد می‌توانند به یکباره از یک نوع مشابه آن برخوردار باشند. این انواع ذهنی همان انواع کارکردگرایانه هستند که «ماشین تورینگ»^۱ نیز نامیده می‌شود.

نشانه ذهنی^۲ یک مثال خاص از «نوع» است؛ مانند احساس درد شما، احساس درد من. بنابراین، ما انسان‌ها در نوع ذهنی مشترک هستیم، اما در نشانه ذهنی خودمان از آن نوع، که نشانه‌ای مختص به خود ما است، شریک نیستیم. نشانه‌های ذهنی با حالات فیزیکی یکی هستند که (از سوی فیزیکالیست‌ها) همان حالت‌های مغزی تلقی می‌شوند و نوع ذهنی را تحقق می‌بخشند.

به‌صورت خلاصه، هوش مصنوعی قوی و عملکردگرایی ماشین آگاهی و روح را به نرم‌افزار کارکردی و مغز را به سخت‌افزاری که نرم‌افزار را اجرا می‌کند فرو می‌کاهد. خطر هوش مصنوعی قوی این است که یک گزارش تحویل‌گرایانه از انسان ارائه می‌دهد که بر اساس آن روح را تنها به مغز، که در واقع یک سخت‌افزار است، و آگاهی را به حالت‌های مغزی، که باعث حرکات بدن می‌شود، تحویل می‌برد؛ به‌عنوان مثال، احساس درد به حالت مغزی‌ای تحویل برده می‌شود که باعث می‌شود در اثر آن شخص «آخ» بگوید و بگرید.

دو مشکل عمده در کارکرد هوش مصنوعی قوی

در حال حاضر دو مشکل عمده در مورد هوش مصنوعی قوی وجود دارد:

اول، اینکه رایانه‌های مجهز به هوش مصنوعی آگاه نیستند.

دوم، اینکه رایانه‌ها به معنای واقعی کلمه قادر به استدلال، تفکر یا حل مسئله نیستند.

۱. Turing Machine

۲. mental token

در مورد عدم آگاهی رایانه‌های هوش مصنوعی باید گفت که فیلسوفان بر این باورند که سه ویژگی در همه حالت‌های آگاهانه وجود دارد:

۱. در آن حالت بودن چگونه است: حالت کیفیت پدیداری (مثلاً)، مانند حالت درد و این که مجموع ۲ و ۲ می‌شود ۴. بنابراین C حالت آگاهی است، اگر و تنها اگر در آن حس «بودن چگونه است» وجود داشته باشد. تفاوت بین حالت درد داشتن و حس کردن مزه آواکادو این است که حالت «در آن حالت بودن چگونه است» این دو از یکدیگر متفاوت است. بنابراین، برای هر حالت آگاهانه "در آن حالت بودن چگونه است" منحصر به فردی وجود دارد. این اولین ویژگی آگاهی است.

۲. دسترسی خصوصی است: یک حالت C یک حالت آگاهانه است، اگر و تنها اگر سوژه به C دسترسی خصوصی داشته باشد. حال این به چه معناست؟ هر حالت فیزیکی را در نظر بگیرید، به عنوان مثال حالت گرم بودن آب یا باردار بودن الکترون یا حالت ناحیه‌ای از نورون‌های شلیک‌کننده مغز را در نظر بگیریم. هر حالت فیزیکی برای همه ما به صورت عمومی در دسترس است، اما آگاهی به این صورت است که من راهی برای شناخت وضعیت آگاه خود دارم که در دسترس دیگری نیست؛ من دسترسی درونی و خصوصی به حالات آگاهانه خود، که شما به آن حالات آگاه من چنین دسترسی‌ای ندارید، دارم و شما تجربه خودتان را از آن حالات دارید.

۳. سومین ویژگی دربارگی (حیث التفاتی) است: دربارگی به معنای «در مورد...» یا «نسبت به...» چیزی بودن افکار و احساسات ما می‌باشد؛ ممکن است

من احساسی در مورد چراغ داشته باشم، ممکن است درباره شام فکر کنم، ممکن است من اعتقادی خاصی درباره اقیانوس هند داشته باشم. بنابراین، حالات ذهنی «در مورد...» یا «نسبت به...» چیزی بودن در خود دارند و بر این اساس به یک شیء اشاره می‌کنند. حالت C حالتی از آگاهی است، اگر و تنها اگر C شدیداً مستلزم حالتی باشد که دارای دربارگی ذاتی باشد یا مربوط به چیزی یا در مورد چیزی باشد.

این سه ویژگی آگاهی، یعنی «در آن حالت بودن چگونه است؟»، «دسترسی خصوصی» و «دربارگی ذاتی»، در مورد حالات آگاهانه به ما اجازه می‌دهد تا استدلالی را بیان کنیم که ماشین‌ها یا رایانه‌های هوش مصنوعی آگاه نیستند و این استدلال به این صورت تقریر می‌شود:

۱. همه حالات خودآگاه دارای این سه ویژگی هستند؛

۲. هیچ حالت فیزیکی این ویژگی‌ها را ندارند؛

۳. هیچ حالت فیزیکی یک حالت آگاهانه نیست؛

۴. رایانه‌ها تنها حالت‌های فیزیکی دارند.

بنابراین رایانه‌ها حالت‌های آگاهانه ندارند.

نکته دیگر این که نه تنها رایانه‌ها آگاه نیستند، بلکه رایانه‌ها موتورهای نحوی هستند، نه موتورهای معنایی؛ یک موتور نحوی یا ماشین قادر است خط خطی یا صداها یا آنچه را که ممکن است علائم صرف بنامیم، طبق قواعدی دستکاری کند و این همان کاری است که رایانه به خوبی انجام می‌دهد. رایانه می‌تواند ورودی‌ها را بگیرد و از طریق نرم‌افزاری که در درون خود دارد بر اساس قواعدی آن را به صورت صحیح دستکاری کند و این باعث تولید خروجی خاصی می‌شود، اما رایانه

هیچ درکی نسبت به معنا یا محتوای گزاره‌ای که آن نمادها نشان می‌دهند ندارد.

این همان چیزی است که جان سرل^۱، فیلسوف شهیر دانشگاه برکلی کالیفرنیا،^۲ در یک آزمایش فکری، که به نام «اتاق چینی»^۳ معروف است، تبیین می‌کند. سرل از ما دعوت می‌کند تا تصور کنیم که به اتفاقی نزدیک می‌شویم که بالای آن نوشته است «پرسش‌های چینی و پاسخ‌های چینی». در یک طرف دیوار، جعبه‌ای وجود دارد که بر روی آن نوشته است: «پرسش‌های چینی را وارد کنید»، و در طرف دیگر نوشته‌ای وجود دارد که می‌گوید «پاسخ‌های چینی از این جا خارج می‌شوند». شما بر در این اتاق می‌ایستید، در حالی که زبان چینی نمی‌دانید؛ اما روی صفحه کاغذ خود سؤالاتی با خط چینی می‌بینید و آن را در جعبه سؤالات قرار می‌دهید. چند دقیقه بعد یک برگ کاغذ دیگر، که روی آن نوشته‌ای به خط چینی است و ظاهراً جواب سؤالات شما است، را دریافت می‌کنید. اتفاقی در این اتاق رخ داده است که شما از آن خبر ندارید: شخصی که داخل اتاق است به هیچ وجه زبان چینی نمی‌داند. او معانی نمادهای چینی را نمی‌فهمد و هیچ محتوای گزاره‌ای را در مورد این زبان درک نمی‌کند؛ اما او دو کتاب بزرگ دارد. نام یکی از این کتاب‌ها کتاب پرسش است، که حاوی اشکال، حروف و کلمات چینی است. او به سؤالاتی که در جعبه گذاشته می‌شود نگاه می‌کند و به محض اینکه اشکال مناسب را پیدا می‌کند، آن‌ها را در آن کتاب می‌یابد و درمی‌یابد که باید به چه صفحه‌ای از کتاب دیگر، یعنی کتاب پاسخ‌ها مراجعه کند. او به کتاب

۱. John Searle

۲. University of California, Berkeley

۳. اتاق چینی (Chinese Room) یک آزمایش ذهنی است که اولین بار توسط مقاله جان سرل به نام «ذهن‌ها، مغزها، و برنامه‌ها» (Minds, Brains, and Programs) در مجله «علوم رفتاری و ذهنی» (Behavioral and Brain Sciences) در سال ۱۹۸۰م منتشر شد.

پاسخ‌ها مراجعه می‌کند و مجموعه دیگری از کلمات، نمادها و جملات چینی را می‌بیند و آن‌ها را روی کاغذ کپی می‌کند و در جعبه پاسخ قرار می‌دهد. شخصی که داخل اتاق چینی است یک ماشین نحوی است که اشکال و اندازه‌های نحوی را دریافت می‌کند و آن‌ها را با اشکال و اندازه‌های مربوطه جایگزین می‌کند، اما هیچ درکی از معانی کلمات چینی ندارد.

رایانه یعنی همین! رایانه ورودی‌های نحوی خاصی را می‌گیرد و قادر است آن‌ها را در نرم‌افزار خود دستکاری کند و سپس آن‌ها را روی صفحه نمایش یا کاغذ چاپ کند؛ مجموعه‌ای از خط‌نوشته‌هایی که به ورودی‌ها مربوط می‌شوند، اما خود رایانه فقط امور نحوی را انجام می‌دهد و هیچ درک معنایی ندارد.

از آزمایش اتاق چینی چنین برمی‌آید که رایانه‌ها واقعاً درگیر عقلانیت اصیل یا حل مسئله واقعی برای یادگیری واقعی نیستند؛ زیرا عقلانیت واقعی، حل مسئله و یادگیری مستلزم داشتن ویژگی‌های خاصی از قبیل محتوای معنایی یا قصدیت و در مورد چیزی بودن هستند.

هوش مصنوعی ضعیف^۱

ما بینش شهودی منطقی به زنجیره فکری مان داریم که مربوط به استدلال کردن هستند و نتیجه‌گیری‌های ما دارای ویژگی‌ای است که هنجاری بودن نامیده می‌شود. فرض کنید که من سعی می‌کنم مسئله‌ای را حل کنم که نتیجه دو ادعا، یعنی «همه انسان‌ها می‌را هستند» و «سقراط یک انسان است»، می‌باشد. برای انجام این کار باید فرایندی عقلی نسبت به بازتاب عقلانی این دو گزاره طی کنم. من باید محتوای معنایی آن‌ها را درک کنم و سپس با درک معنای این دو گزاره،

۱. Weak Artificial Intelligence

از طریق بینش شهودی خود، از نحوه ارتباط آن‌ها با یکدیگر، متوجه شوم که آن‌ها منطقاً نتیجه‌ای که «سقراط فانی است» را به دنبال دارند. این نتیجه از ویژگی هنجاری بودن برخوردار است. مقصود از هنجاری بودن این است که از دو مقدمه، لزوم باور داشتن به نتیجه استفاده می‌شود.

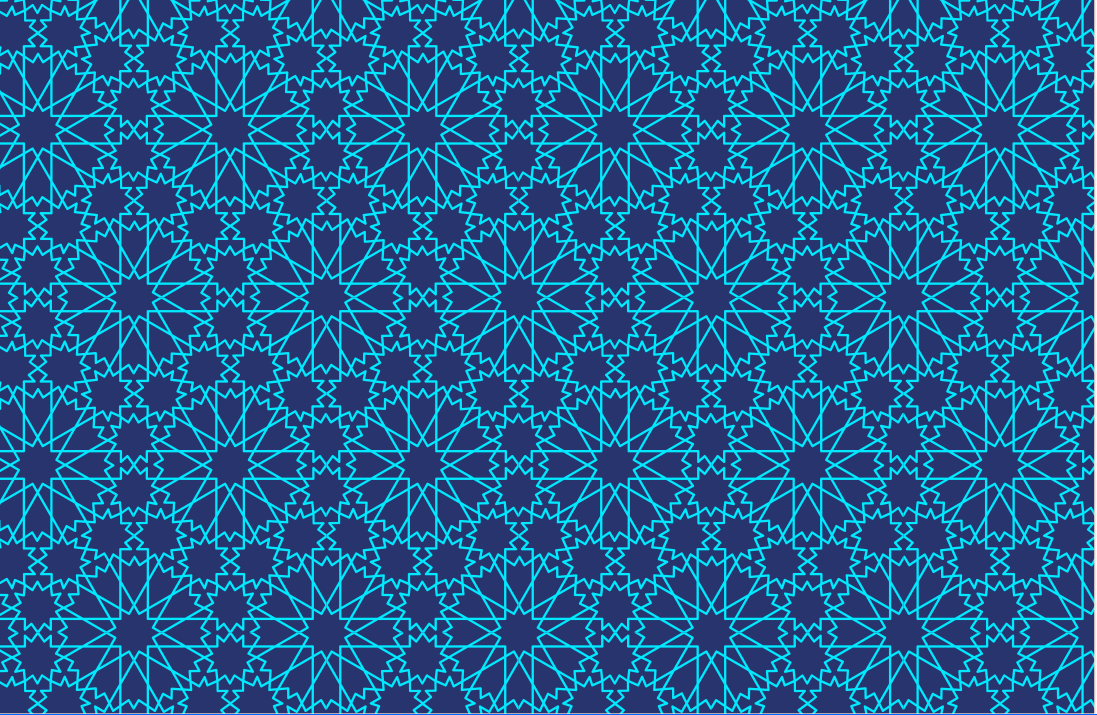
رایانه‌ها فاقد این ویژگی‌ها هستند، آن‌ها حالت‌هایی که «درمورد» چیزی باشند ندارند. آن‌ها معانی را درک نمی‌کنند، آن‌ها واقعاً ریاضیات را نمی‌فهمند یا بینشی نسبت به برهان ندارند. همچنین، آن‌ها از ویژگی هنجاری بودن اصلاً برخوردار نیستند. ماشین تنها یک ماشین است. بنابراین، رایانه‌ها به معنای واقعی کلمه درگیر عقلانیت نیستند، بلکه آن‌ها تنها این فعالیت‌ها را به معنای عرفی و رایج آن تقلید می‌کنند. و این دیدگاه که رایانه‌ها واقعاً هوشمند نیستند و تنها از موجودات هوشمند تقلید می‌کنند را هوش مصنوعی ضعیف می‌نامند. هوش مصنوعی ضعیف برای خدا باوران کاملاً قابل قبول است.

نتیجه

نتیجه این بحث این است که کارکردگرایی ماشینی، به عنوان یک تبیین فلسفی از آگاهی و نفس، یک شکست کامل محسوب می‌شود؛ چرا که آن ویژگی‌های آگاهی مانند «چه حسی دارد؟»، «دسترسی خصوصی» و «دربارگی» را از خود نشان نمی‌دهد. همچنین فاقد این حالت است که به صورت عقلانی معنایی را درک کند یا این که محتوای معنایی فرآیند افعال اختیاری را بفهمد. همچنین، دارای دریافتی شهودی در امر استنتاج، که برای صدق عنوان آگاهی می‌بایست اتفاق بیافتد، نیست.

در مجموع، ارزش پیشرفت‌های علوم رایانه را ارج می‌نهیم. بی‌تردید رایانه‌ها

می‌توانند کارها را بسیار سریع‌تر و بهتر از انسان‌ها انجام دهند، اما آنچه برای یک خدا باور کاملاً قابل پذیرش است هوش مصنوعی ضعیف است. هوش مصنوعی قوی را، که تقلیل‌دهنده آگاهی به امور مادی است و آگاهی و روح را به عنوان اموری غیر مادی انکار کرده و از آن‌ها با رویکردی تقلیل‌گرایانه به مثابه یک ماشین و برنامه‌ی نرم‌افزاری نام می‌برد، بر خلاف شواهدی است که ما نسبت به سرشت انسان واقعی داریم؛ امری که برای ما قابل پذیرش نیست.



«هوش مصنوعی و پدیده آگاهی»

سید محمد سادات منصوری^۱

۱. عضو هیأت علمی پژوهشکده علوم شناختی، تهران.





مقدمه

در این بحث به چند نکته اشاره خواهد شد:

۱. مقدمه و معرفی که توضیح کوتاهی از مفاهیم داده می‌شود.
۲. در خصوص هوش مصنوعی قوی^۱ و هوش مصنوعی ضعیف^۲ توضیح کوتاهی داده می‌شود.
۳. توضیح بعضی از مبانی ملاصدرا که ممکن است بتواند مسئله امکان هوش مصنوعی قوی را نشان دهد.
۴. در مورد «کوالیا» و گزاره محتوای ذهنی^۳ و نظریه محاسباتی ذهن و هوش مصنوعی صحبت خواهد شد.

اصطلاحات

کوالیا: در اصطلاح فلسفی «کوالیا»^۴ به معنی یک حس خاص است. وقتی دندان من درد می‌گیرد این یک حس درونی خاصی است که تا شما آن را تجربه نکرده

۱. Strong Artificial Intelligence(Strong AI)

۲. Weak Artificial Intelligence(Weak AI)

۳. Mental Content

۴. Qualia

باشید، اصلاً آن را نمی‌فهمید. شاید حس خاص دندان دردی که در من ایجاد شده به نحوی باشد که با درد شما متفاوت باشد، به همین دلیل شما هرگز آن را نمی‌فهمید. این یک حس درونی^۱ است که اسمش را کوآلیا گذاشته‌اند که اصطلاح فنی و یک کلمه لاتین است. «اکوآلی» جمع کوآلیا یا همان کوآلیتی است و مفردش کوآلی می‌شود؛ به معنای حس خاص و حالتی خاص که بعضی از حالات ذهنی دارند. اینکه همه حالات ذهنی کوآلیا داشته باشند مسلم نیست. عده‌ای از فلاسفه می‌گویند که ما دو گروه حالت ذهنی داریم: یک گروهشان کوآلیا دارند؛ مثل خواص حالات حسی^۲، مثل حواس پنجگانه که همه حس خاص هستند. یک گروه هم کوآلیا ندارند که شاید بعداً به آن‌ها بپردازیم. جنبه «اینتروسپکتیو»^۳ و حالات درونی و بعضی از حالات ذهنی و جنبه «فنومنال» مورد اختلاف نیست، ولی یک مجادله در این زمینه هست که ما چگونه کوآلیاها را تفسیر کنیم و اصلاً کوآلیاها امر درونی هستند یا مربوط به جهان خارج هستند. این بحث بسیار مهمی است.

گزاره: اصطلاح بعدی گزاره^۴ است که اصطلاحی مهم است؛ برای اینکه ما دو دسته حالات ذهنی داریم: یک دسته حالات ذهنی که کوآلیا دارند، دسته دیگر حالات ذهنی که گرایش‌های گزاره‌ای^۵ هستند. مثل اینکه من می‌گویم: من باور دارم که اکنون باران می‌بارد. این باور من یک حس خاص نیست. باور من یک محتوا دارد؛ محتوایی که گرایش گزاره‌ای ما است. پس ما دو گونه حالت ذهنی داریم: یک دسته حالت ذهنی ما و کوآلیا هستند و دسته دیگر گرایش‌های

۱. Subjective

۲. Sensation

۳. Introspective

۴. proposition

۵. propositional attitudes

گزاره‌ای هستند. منظور ما از گزاره چیست؟ گزاره یعنی چیزی که ارزش صدق روی آن قرار می‌گیرد. اولین چیزی که ارزش صدق روی آن قرار می‌گیرد و حامل ارزش صدق است گزاره است؛ مثلاً من که می‌گویم اکنون باران می‌بارد، این جمله من صادق یا کاذب نیست، بلکه معنا و محتوای آن است که ممکن است صادق یا کاذب باشد؛ محتوایی که یک نوع گزاره است. متعلق گرایشات گزاره‌ای ما هم یک نوع گزاره است؛ یعنی گرایشات گزاره‌ای ما هم روی گزاره قرار می‌گیرد.

محتوای ذهنی: محتوای بعضی از حالات ذهنی ما هستند. محتوای گرایشات گزاره‌ای ما هستند که گفتیم محتوایشان همان نحوه بازنمایاندن جهان است. بحث‌های مفصلی در فلسفه ذهن وجود دارد درباره این که محتوای ذهنی چیست.

سیستم محاسباتی: یک سیستم محاسباتی مصنوعی است که شبیه محاسباتی که در مغز ما انجام می‌شود را انجام می‌دهد. در بستری به نام زبان فکر به صورت محاسباتی تولید تفکر و فکر می‌شود؛ حتی دیدن و شنیدن هم اینگونه است.

هوش مصنوعی و جایگزینی آن با هوش انسان

محققان هوش مصنوعی می‌خواهند ماشین‌های هوشمند درست کنند؛ می‌خواهند حیوانات مصنوعی و انسان‌های مصنوعی تولید کنند. پرسش‌های ما در این زمینه این است که:

آیا هوش مصنوعی ممکن است به سطحی از توسعه برسد که قابل مقایسه با هوش انسانی شود؟

آیا امکان پذیر است که انسان‌ها روزی ربات‌هایی بسازند که ذهن‌های آگاه داشته باشند؟

آیا ماشینی که خوب برنامه‌ریزی شده می‌تواند آگاهی داشته باشد؟

موافقان می‌گویند: ما دلایل زیادی داریم تا باور کنیم که نهایتاً رایانه‌ها ذهن خواهند داشت. مخالفان در مقابل می‌گویند: رایانه‌ها محدودیت‌هایی دارند که اجازه نمی‌دهد آن‌ها به ذهن انسان برسند. افرادی که موافق «هوش مصنوعی قوی» هستند دو گروه هستند:

۱. تقلیل‌گرایان: آن‌ها معتقدند که ما می‌توانیم ماشین‌هایی بسازیم که مثل انسان بفهمند، ذهن و حیات ذهنی داشته باشند؛ کوالیا و «محتوای ذهنی» داشته باشند. یک راه در مقابل این گروه این است که ذهن را در حد یک ماشین پایین بیاورند. راه دیگر این است که ماشین را در حد ذهن بالا ببرند. این گروه ماشین را به ذهن انسان نزدیک می‌کنند و گروهی دیگر ذهن را تا حد ماشین پایین می‌آورند که «حذف‌گرایان»^۲ هستند.

۲. حذف‌گرایان: آنان معتقدند که اصلاً انسان ذهن ندارد یا اگر ذهن داشته باشد، چیزی بیشتر از حالاتی «محاسباتی» نیست؛ یعنی کارکرد ذهن هم کاری محاسباتی است. ذهن نرم‌افزاری است که روی سخت‌افزار زبان و مغز اجرا می‌شود. بنابراین، ذهن را به امور کارکردی و تابعی^۳ تقلیل می‌دهند. همچنین، این گروه ذهن را به آن معنایی که کوالیای خاص داشته باشد، همان‌طور که در

۱. Reductionist

۲. Eliminativism

۳. Functional

خصوص کوالیا توضیح داده شد، یا اینکه محتوایی داشته باشد، همان طور که در اتاق چینی توضیح داده شد، نمی پذیرند، بلکه معتقدند که ذهن اصلاً چیزی رایانه ای است و مغز انسان چیزی جز یک رایانه پیچیده نیست. ممکن است کسانی غیر تقلیل گرا باشند و در عین حال «هوش مصنوعی قوی» را بپذیرند؛ یعنی آن ها ذهن را به حالات کارکردی تقلیل نمی دهند و آن را چیزی ورای حالات کارکردی می دانند. آنان همه این موارد را می پذیرند و می گویند که ذهن انسان «کوالیا»^۱ دارد. پس آنان تحویل گرا نیستند، ولی هوش مصنوعی قوی را می پذیرند و می گویند: ذهن ورای حالات کارکردی است؛ یعنی این حالات کارکردی ممکن است وجود داشته باشد، ولی ذهن این نیست؛ ذهن امری ورای فیزیک است.

مخالفان هوش مصنوعی قوی هم به دو دسته تقسیم می شوند: فیزیکیالیست ها و غیرفیزیکیالیست ها. غیرفیزیکیالیست ها هم می توانند مخالف هوش مصنوعی قوی باشند. کسانی که هوش مصنوعی قوی را می پذیرند معمولاً فیزیکیالیست هستند. غیرفیزیکیالیست ها کسانی هستند که چیزی به نام «روح» را می پذیرند. قائلان به دوگانه انگاری جوهری (سابستنس دوئالیست)^۲ کسانی محسوب می شوند که قاعداً نمی توانند هوش مصنوعی قوی را بپذیرند؛ اما ممکن است که ما با مبانی ملاصدرا بتوانیم ثابت کنیم که هم می توان دوگانه انگار جوهری باشیم و در عین حال هوش مصنوعی قوی را بپذیریم. این به نظر حرف جدیدی می تواند باشد.

نظریات ملاصدرا و مسئله جایگزینی هوش انسان

گرچه ملاصدرا دوگانه انگار جوهری است، با توجه به مبانی او، می توان گفت

۱. qualia

۲. Substance dualist

که هوش مصنوعی قوی امکان پذیر است. نتیجه این حرف این است که ضرورتی ندارد برای اینکه کسی طرفدار هوش مصنوعی قوی باشد حتماً به ذهن به صورت تابعی نگاه کند؛ یعنی می تواند یک نفر کارکردگرا نباشد و می تواند تحویل گرا هم نباشد و در عین حال هوش مصنوعی قوی را بپذیرد.

ملاصدرا چند قاعده مهم فلسفی دارد که دو تا از قواعد بسیار مهم او در مسئله ذهن، روح و نفس را استفاده می کنیم و ادعا خود را اثبات می کنیم. اولی حرکت جوهری^۱ است. یکی از مهم ترین اصول فلسفی ملاصدرا حرکت جوهری است. ملاصدرا مدعی است که همه موجوداتی که در حال حرکت هستند صیورورت دارند. تقریباً همه فلاسفه پیش از ملاصدرا حرکت جوهری را نمی پذیرند، بلکه فقط حرکت در اعراض را قبول داشتند؛ چون تصورشان این بود که اگر جوهر حرکت کند، خود آن شیء عوض می شود، پس ما چیز واحدی نخواهیم داشت که به آن حرکت را نسبت دهیم. پس اگر شیء واحدی نباشد، حرکت را به چه چیزی نسبت دهیم. این اشکالی بود که فیلسوفان مخالف حرکت جوهری به نظرشان می رسید.

اولین مسئله در این زمینه حرکت در جوهر است. جوهر و ذات یک شیء می تواند تغییر کند و چیز جدیدی شود. اسم این را حرکت جوهری می گذاریم. پس فقط حرکت در اعراض نیست، بلکه در جوهر هم حرکت وجود دارد. دیدگاه خاص ملاصدرا در خصوص روح و نفس دقیقاً از اصل حرکت جوهری ناشی می شود. ملاصدرا معتقد است که بدن انسان به تدریج در اثر حرکت جوهری به مرحله ای می رسد که تبدیل به روح می شود یا در آن روح ظهور می کند. یعنی یک

۱. motion Substance

بستر مادی است که تبدیل به روح می شود و روح در او ظهور می کند. او به این روند «جسمانیة الحدوث» می گوید. معنای این اصطلاح این است که روح از ابتدا امری مادی است که بدن، که آن هم امری مادی است، آرام آرام در اثر حرکت جوهری اول روح نباتی پیدا می کند، که مادی است، بعد آرام آرام در اثر حرکت جوهر تبدیل به یک شیء غیر مادی می شود که همان روح مجرد و غیر مادی است.

اگر ما این دو اصل را با هم داشته باشیم، نتیجه این است که بگوییم: روح از بدن برمی خیزد یا بگوییم بدن تدریجاً تکامل و تحول پیدا می کند و از حالت جنینی به مرحله تحول می رسد. بر اثر این تکامل روح از بدن برمی خیزد و پدیدار می شود و در نتیجه حرکت جوهری بدن تبدیل به روح می شود. اگر ما این اصل را بپذیریم، معنایش این است که روح معلول بدن خواهد بود و می توانیم نتیجه بگیریم که هر بدنی برای خود روح می سازد. از اینجاست که روح حیوان با روح انسان متفاوت است؛ چون بدن حیوان با بدن انسان متفاوت است. از اینجاست که روح یک حیوان خاص، مثل اسب، با روح یک حیوان دیگر، مثل گربه، متفاوت است. دلیل این تفاوت این است که بدن و مغز آن ها متفاوت است. چرا انسان روحی متفاوت با روح حیوان دارد؟ چون بدن و مغزش متفاوت با بدن و مغز آن ها است. چرا روح هر انسان با روح انسان دیگر متفاوت است؟ چون بدن و مغز هر انسان با بدن و مغز انسان دیگر متفاوت است. پس تمام تفاوت ها که در روح ها وجود دارد به دلیل تفاوت هایی است که در بدن ها است. هر بدنی روح خود را می سازد. این اندیشه در خصوص روح و ارتباط روح و بدن مشکل قدیمی ارتباط نفس و بدن را حل می کند: اگر مشکل ارتباط نفس و بدن «پیرینگ»^۱ باشد که چطور روح با بدنی خاص جفت می شوند، جواب ملاصدرا این است که بدن

۱. pairing

است که تبدیل به روح می شود. پس روح از جای دیگری نمی آید تا به بدن بچسبد. مسئله اینکه چطور روح غیر مادی می تواند بدن را کنترل کند و بر آن اثر بگذارد و بدن بر روح اثر بگذارد هم با این نظریه که روح از بدن برمی خیزد و زاییده آن است حل می شود.

اگر بخواهیم نتیجه گیری کنیم ممکن است بگوییم که هر وقت در دانش هایی مثل دانش تجربی فهمیدیم که بدن ها در چه مرحله ای واجد چه ویژگی از وحدت یا یکپارچگی (یونیکیتی) می شوند که روح در آن ها ظهور می کند یا روح از آن ها به وجود می آید، ممکن است بتوانیم ماشین هایی مشابه انسان بسازیم. با یک ماده متفاوت ماشین هایی از سلیکون^۱ ها و مدارهای الکترونیکی^۲ ساخته شوند که به قدری پیچیده باشند که به همان مرحله بدن انسان برسند که تولید روح می کند. این ماشین ها اگر به آن مرحله برسند، ممکن است بگوییم که می توانند تولید روح ماشینی کنند. همان طور که روح حیوان متفاوت با روح انسان است چون بدنش متفاوت است، روح ماشین هم می تواند از روح انسان متفاوت باشد، به دلیل این که بدنش با بدن انسان متفاوت است.

برای اینکه ما هوش مصنوعی قوی را بپذیریم، ضرورتی ندارد که تحویل گرا (ریداکشنیست) باشیم. همچنین، ضرورتی ندارد کارکردگرا (فانکشنیست) باشیم؛ بلکه می توانیم دوگانه انگار یا دوگانه انگار جوهری باشیم و در عین حال هوش مصنوعی قوی و امکان آن را بپذیریم. این نتیجه بسیار مهمی است که بر اساس آن ما هم می توانیم به کسانی که به دنبال ساختن ماشین هایی مثل انسان ها هستند راه نشان دهیم و از نظر فلسفی مشکلات فلسفی آنها را حل کنیم، به طوری

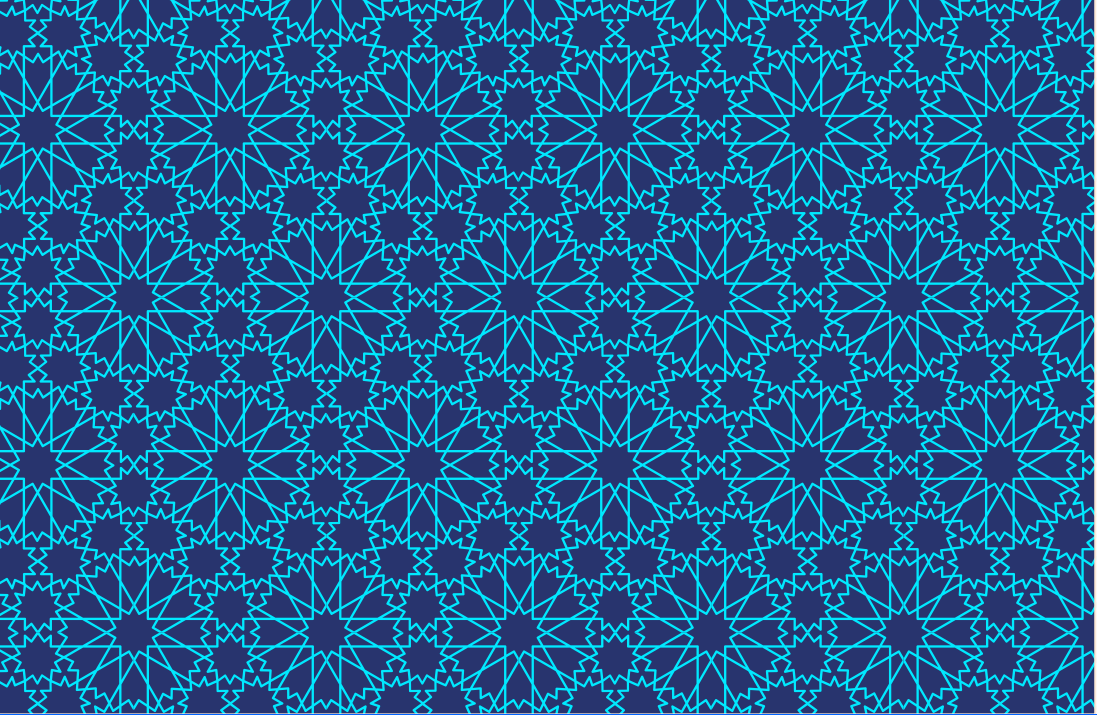
۱. Silicone

۲. Electronic Circuit

که با الهیات هم تعارضی نداشته باشد.^۱

این مباحث صرفاً از باب مقدمه بود. اما اینکه ابعاد این موضوع دقیقاً چیست و تفاوتش با دیدگاه‌های دیگر چیست و این‌که آیا واقعاً ماشین‌های محاسباتی (کامپیوتیشنال) می‌توانند روح داشته باشند یا ماشین‌های دیگری که به شکل دیگری کار می‌کنند و روح داشته باشند، مباحثی است که برای انسان هنوز روشن نیست و برخی از آنها نیاز به توضیحات تفصیلی دارد.

۱. این ادعا که از منظر مبانی حکمت متعالیه، هم‌چون جسمانیة الحدوث و حرکت جوهری، می‌توان امکان تحقق هوش مصنوعی قوی و روح و ذهن داشتن آن را توجیه کرد، اشکال اساسی دارد؛ زیرا این دو مبنا، که نویسنده محترم مطرح کرده‌اند، در رابطه با جسم طبیعی است نه جسم مصنوعی و هوش مصنوعی که از قبیل جسم مصنوعی است. همچنین، با توجه به مبانی دیگر حکمت متعالیه، هم‌چون فرق بین وحدت و ترکیب حقیقی با وحدت و ترکیب اعتباری و بحث نحوه وجود ماده و صورت و بحث فلسفی مزاج، این دیدگاه نویسنده مورد نقد اساسی خواهد بود.



الزامات هوش مصنوعی قوی با تأکید بر مواجهه دین اسلام با هوش مصنوعی قوی

سید محمد سادات منصوری^۱

۱. عضو هیأت علمی پژوهشکده علوم شناختی، تهران.





مقدمه

عنوان بحث الزامات هوش مصنوعی قوی با تأکید بر مواجهه دین اسلام با هوش مصنوعی قوی است. در این مبحث هر جا که از هوش مصنوعی نام برده می شود منظور همان هوش مصنوعی قوی است. درباره تفاوت هوش مصنوعی قوی و ضعیف مختصراً می توان گفت که در هوش مصنوعی ضعیف ادعا تنها این است که انسان می تواند ماشین هایی بسازد که شبیه انسان عمل کند و رفتارهای هوشمندانه داشته باشد و طوری رفتار کنند که انگار ذهن دارند. اما در هوش مصنوعی قوی ادعا آن است که انسان می تواند ماشین هایی بسازد که دقیقاً مانند انسان احساسات، ذهن و فهم داشته باشند؛ به طوری که انگار انسان انسانی جدیدی ساخته باشد. اسم این را هوش مصنوعی قوی گذاشته اند.

انواع الزامات هوش مصنوعی

بحث الزامات هوش مصنوعی عنوان گسترده و بسیار متنوعی است. برای همین سعی می شود تقسیم بندی ابتدایی از این الزامات انجام شود. ابتدا، الزامات هوش مصنوعی قوی با این نظریه که انسان می تواند ماشین هایی بسازد که همانند انسان ذهن داشته باشند مطرح می شود؛ همانند به معنای اینکه از یک سو فهم گزاره ای داشته باشند و معنا و محتوا را درک کنند و از سوی دیگر

احساسات و به تعبیری کوآیا باشند و بتوانند حس کنند و همچنین اراده آزاد^۱ داشته باشند. اگر ما چنین چیزی را طرح کنیم، الزاماتی خواهیم داشت که به الزامات پیش از تولید هوش مصنوعی قوی و پس از تولید هوش مصنوعی قوی تقسیم می‌شوند. منظور از الزامات پیش از تولید آن اموری است که برای تولید هوش مصنوعی لازم است؛ یعنی قبل از اینکه هوش مصنوعی تولید کنیم باید ملزوماتی را در نظر داشته باشیم. این‌گونه ملزومات پیش از تولید را می‌توانیم به الزامات نظری و الزامات فناوریانه^۲ تقسیم کنیم.

۱. الزامات پیش از تولید هوش مصنوعی

الزامات نظری: نظریاتی هستند که پشتوانه نظری ما برای تولید هوش مصنوعی به‌شمار می‌روند. در واقع، مثل تعریف هوش مصنوعی قوی یا بررسی امکان منطقی و نومولوژیک تولید هوش مصنوعی پشتوانه نظری ما در تولید هوش مصنوعی هستند.

الزامات فناوریانه: دانش مورد نیاز برای ساخت و تولید هوش مصنوعی است.

۲. الزامات هوش مصنوعی پس از تولید

این الزامات مواردی است که در مواجهه با هوش مصنوعی باید مورد توجه قرار بگیرد که به دو گروه الزامات نظری و الزامات هنجاری قابل تقسیم هستند:

الزامات نظری: مثل بررسی می‌زان همسانی یا مطابقت هوش مصنوعی تولیدشده توسط انسان با حالات ذهنی انسان. همچنین، بررسی لوازم و

۱. Free will

۲. Technologic

چالش‌های کلامی وجود هوش مصنوعی، مثل مخالفت یا عدم مخالفت آن با توحید در خالقیت یا مثل امکان حشر هوش مصنوعی در قیامت با این سؤال که آیا ماشین‌ها و ربات‌های هوشمند در قیامت محشور می‌شوند. همچنین، بررسی امکان حیات پس از مرگ در ربات‌های هوشمند. نیز، مسائل مرتبط با اعتقادات در ربات‌های هوشمند، مثل ضرورت یا عدم ضرورت اعتقاد به مبدأ، معاد، نبوت و سایر اعتقادات دینی. همچنین این‌که آیا ربات‌های ساخته‌شده توسط انسان امکان تکامل معنوی و باطنی دارند، همانند آنچه که در مورد امکان تکامل معنوی انسان گفته می‌شود. این موارد نمونه‌هایی از الزامات نظری پس از تولید هوش مصنوعی هستند.

الزامات هنجاری:^۱ در خصوص هوش مصنوعی که به الزامات شرعی، اخلاقی و حقوقی قابل تقسیم هستند. در الزامات شرعی، نوع مواجهه فقیهانه مثلاً در دین اسلام با ربات‌های هوشمند باید مورد توجه قرار بگیرد. در بحث هنجارهای اخلاقی، هنجارهای ارتباط انسان با ربات‌ها و برعکس مورد توجه قرار می‌گیرد؛ یعنی انسان‌ها چه رفتار اخلاقی باید در مواجهه با ربات‌ها در پیش بگیرند و ربات‌ها در قبال انسان چگونه باید رفتار کنند که اخلاقی باشد. همچنین، در بحث حقوقی مسئله حقوق اجتماعی و حقوق فردی ربات‌های هوشمند مورد بررسی قرار می‌گیرد.

الزامات منطقی مواجهه دین اسلام با هوش مصنوعی

در این بحث، باید ابتدا از امکان منطقی هوش مصنوعی سخن بگوییم. منظور از امکان منطقی به شیء چیست؟ منظور این است که آیا وجودش مستلزم تناقض

منطقی هست یا نه. اگر چیزی وجودش مستلزم تناقض منطقی بود می‌گوییم که امکان منطقی ندارد؛ ولی اگر وجودش مستلزم تناقض منطقی نبود، می‌گوییم که منطقاً امکان وجود دارد.

حال پرسش این است که آیا هوش مصنوعی قوی امکان منطقی دارد یا ندارد. ما از کجا باید بفهمیم که وجود هوش مصنوعی قوی منطقاً ممکن است یا محال یا مستلزم تناقض است یا نیست؟ یکی از راه‌هایی که فلاسفه از طریق آن سعی کردند به این پرسش پاسخ دهند تصویرپذیری^۱ است. آنان می‌گویند که تصویرپذیری مستلزم اثبات امکان است؛ یعنی اگر چیزی تصویرپذیر بود، ناگزیر امکان منطقی دارد. فلاسفه که تصویرپذیری قوی دارند امکان هوش مصنوعی قوی را منکر نشده‌اند و نمی‌گویند که وجودش مستلزم تناقض منطقی است.

الزامات کلامی والهیاتی هوش مصنوعی

پرسش مشخص در اینجا این است که آیا هوش مصنوعی با توحید در خالقیت منافات دارد یا نه. ما وقتی موحد هستیم باید توحید را با جنبه‌های مختلفش بپذیریم؛ از جمله اینکه تنها خالق در عالم خداوند است و شریکی در خالقیت برای او نیست. ممکن است برخی معتقدان به ادیان و برخی از متکلمان و الهی‌دانان بگویند که هوش مصنوعی با توحید در خالقیت منافات دارد؛ مثلاً بگویند که اگر زمانی بشر بتواند ماشین‌هایی بسازد که مانند انسان دارای ذهن باشد، معنایش این است که انسان توانایی این را پیدا کرده که همتای خدا در خالقیت باشد. به این ترتیب به نظر آن‌ها این توانایی انسان با آموزه‌های توحید در خالقیت در تعارض است. یکی از جهاتی که موجب شده گاهی برخی پیروان ادیان با امکان هوش

۱. Conceivability

مصنوعی قوی مخالفت کنند همین نکته است که تصور می‌کنند اگر انسان زمانی ربات‌های هوشمندی مانند انسان بسازد، شریک خداوند در خالقیت شده و این را یک چالش مهم با آموزه‌های کلامی خود و تفکر دینی‌شان می‌دانند.

آموزه‌های فلسفی ما درباره حل این مشکل چه کمکی می‌تواند بکند؟ فلاسفه اسلامی بعد از اینکه تحلیل‌های عمیقی را در خصوص حقیقت علیت و رابطه خاص علی و معلولی به خصوص در مکتب ملاصدرا ارائه کردند به این نتیجه رسیدند که هر معلول هرچه دارد ظهور و نمود کمالاتی است که از علت به او رسیده است؛ یعنی آنچه که در معلول به عنوان خوبی و کمال مشاهده می‌شود چیزی جز ظهور کمالات و صفات علت در او نیست. بنابراین، معلول تنها کمالات علت خود را ظاهر می‌کند و از خودش چیزی ندارد. مطابق این تحلیل فلسفی، چون انسان معلول خداوند است و خداوند هم علت حقیقی انسان است، هر آنچه که انسان دارد و هر توانایی که در او دیده می‌شود در واقع به خداوند تعلق دارد؛ به این معنا که تمام خصوصیات و توانایی‌های انسان واقعاً از آن او نیست، بلکه از خداوند است که خالق و علت اوست. در این صورت، اگر انسان با توانایی‌های خودش چیزی شبیه به یکی از مخلوقات خداوند بسازد، در حقیقت این کار را با توانایی خداوند انجام داده است. پس در حقیقت، آن شیء نیز مخلوق و معلول خداوند است. مطابق همین آموزه فلسفی می‌توانیم به یکی از آموزه‌های قرآن کریم نیز اشاره کنیم: «وَاللَّهُ خَلَقَكُمْ وَمَا تَعْمَلُونَ»؛ (در حالی که خدا شما را، و آنچه را می‌سازید، آفریده است). این آیه صریحاً آنچه ساخته دست انسان هاست را هم مخلوق خداوند دانسته است.

بنابراین، طبق آموزه‌های قرآنی و با توجه به مبانی فلسفی اگر روزی انسان توانایی تولید ربات‌هایی همانند انسان را داشته باشد، این کار شرک در خالقیت نخواهد بود و هرگز با توحید در خالقیت در تعارض نیست.

الزامات شرعی هوش مصنوعی

ربات‌های هوشمند تا چه حدی ممکن است مشمول الزامات شرعی باشند؟ در اینجا باید توجه کنیم که وقتی از ربات‌های هوشمند سخن می‌گوییم، باید آن‌ها را در طیفی از هوشمندی از درجه پایین آن تا درجه بالا تقسیم کنیم. مراد ما از هوشمندی درجه پایین یعنی حداقل ذهن‌مندی لازم در حیوانات، حیواناتی مثل حشرات که پیچیدگی کمتری دارند و منظور از هوشمندی درجه بالا حداکثر هوشمندی است که در انسان‌ها است. البته تعیین اینکه هوشمندی درجه پایین چه لوازمی دارد و مرزهای آن چیست کار آسانی نیست و بسیار بحث‌برانگیز^۱ است و از عهده این مقاله هم خارج است؛ ولی اجمالاً با این مقدمه تعریفی ارائه می‌شود: درجه هوشمندی موجودات به شدت به ذهن‌مندی واقعی آن‌ها وابسته است. منظور از ذهن‌مندی واقعی یعنی اعم از آگاهی پدیداری و محتوای گزاره‌ای. هرگاه ذهن‌مندی یک ربات قابل مقایسه با ذهن‌مندی واقعی یک حشره بود، آن را هوشمندی درجه پایین می‌گوییم و هرگاه ذهن‌مندی واقعی یک ربات قابل مقایسه با ذهن‌مندی انسانی باهوش برتر بود آن را هوشمندی درجه بالا می‌گوییم.

بنابر این تقسیم‌بندی می‌گوییم: اگر ربات‌های هوشمند از حداقل ذهن‌مندی برخوردار نبودند، همان هوش مصنوعی ضعیف هستند. بنابراین، از نظر شرعی هم مثل سایر موجودات بی‌جان با آن‌ها برخورد می‌کنیم؛ مثلاً مشمول قوانین

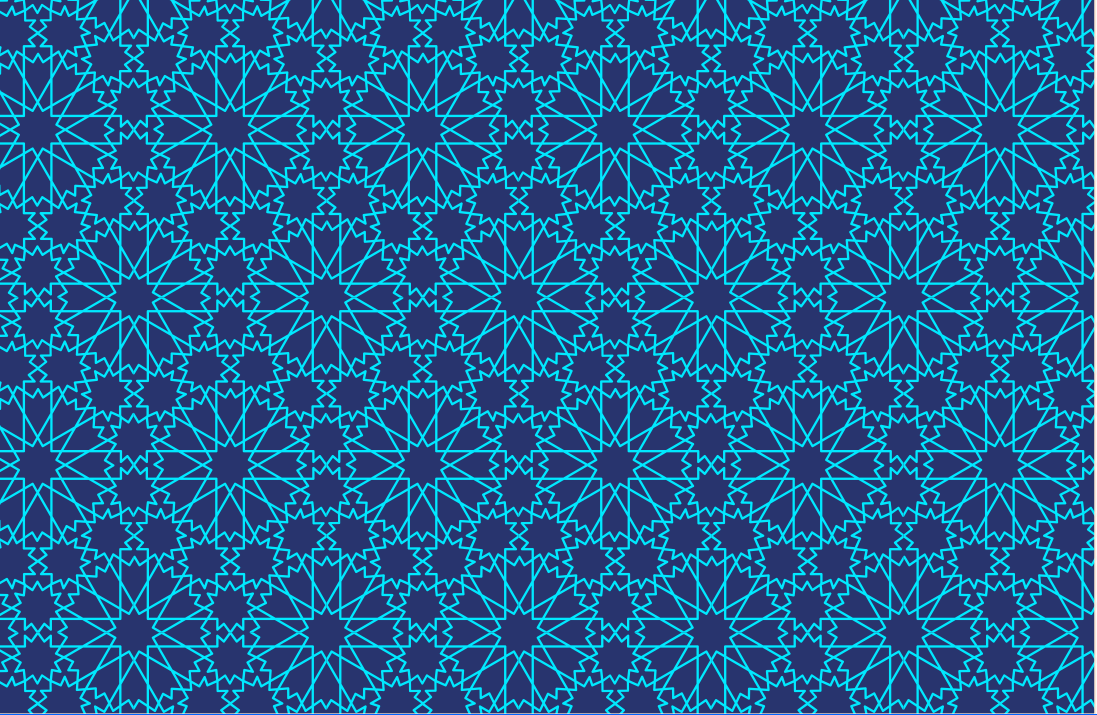
۱. Controversial

مالکیت و تجارت می‌شوند. در صورتی که ربات‌های ساخته‌شده قابل مقایسه با انسان‌های معمولی باشند، می‌توان گفت که مشمول قوانین انسانی منطبق با انسان‌های زنده هستند و از حقوقی مشابه با انسان‌ها برخوردارند.^۱

اینجا این سؤال مطرح می‌شود که با توجه به خطر احتمالی که از جانب هوش مصنوعی قوی ممکن است بشر را تهدید کند، اگر یک هوش مصنوعی قوی و شریر انسان‌ها را زیر سلطه خودش بگیرد، از نظر اخلاقی یا دینی، آیا اصلاً رفتن به سمت ایجاد هوش مصنوعی کار درستی است.

راه حل این مسئله این است که بر اساس مبانی نظری و فقهی اسلام و البته به صورت هوشمندانه به سمت ایجاد قوانینی رفت که هوش مصنوعی را کنترل کند و به دست بشر بدهد تا طوری هوش مصنوعی پیشرفت نکند که کنترلش از دست بشر خارج شود. چنین کنترلی نیازمند تدوین قوانینی پیش از پیشرفت آن‌هاست. بحث‌هایی که در اینجا ارائه می‌شود مقدمات خوبی است برای اینکه فلسفه لازم برای تدوین نظام‌های حقوقی مورد نظر درباره هوش مصنوعی را به وجود بیاورد که بر اساس آن بتوانیم از آن نظام‌های حقوقی استفاده کنیم تا جلوی روش‌های مضر استفاده از هوش مصنوعی را تا حدی مقبول گرفته شود.

۱. این دیدگاه ارائه‌دهنده محترم است هرچند در پارگراف بعدی این ادعا را به صورت احتمالی مطرح می‌کند. در بحث قبلی، ذیل هوش مصنوعی قوی، عدم امکان تحقق هوش مصنوعی قوی از منظر مبانی حکمت متعالیه تبیین شد. اگر هوش مصنوعی قوی همانند انسان شود، لوازم بسیار مهم اجتماعی، حقوقی، اخلاقی و فقهی دارد؛ به طور نمونه دیگر نمی‌بایست آن‌ها را معدوم کرد یا اینکه آن‌ها می‌بایست مسئولیت کارهایی را که انجام می‌دهند بپذیرند. یا اینکه انسان‌ها می‌بایست در قبال آن‌ها رفتارهای اخلاقی انجام دهند و آن‌ها را اذیت و آزار نکنند. ربات‌ها می‌بایست همانند انسان‌ها حقوق اجتماعی و شهروندی داشته باشند. آیا این امور در مواجهه با یک ماشین، هرچند بسیار پیشرفته، قابل تصور است؟



«هوش مصنوعی و دین در مسیحیت»

دکتر تیموئی شتل^۱

۱. عضو هیأت علمی و دانشیار دانشگاه مسایا، پنسیلوانیا، آمریکا.



مقدمه

در حال حاضر برای ایجاد هوش مصنوعی، یعنی ماشین‌هایی که قادر به درک و اراده در سطح انسانی هستند، تلاشی جهانی وجود دارد. ما در این بحث در مورد اینکه هوش مصنوعی چیست بیشتر بحث خواهیم کرد. اما قبل از ورود به جزئیات باید گفت که این بحثی پیچیده است؛ چرا که به طور معمول ما تصور می‌کنیم که یک تکه فناوری چیزی را نمی‌فهمد و اراده آزاد ندارد؛ مثلاً مداد من یک شیء است و یک شخص نیست، در حالی که انسان‌ها اراده آزاد دارند و قادر به انتخاب هستند. یک موجود فناورانه هیچ‌کدام از این خصوصیات را ندارد. تلاش برای ایجاد هوش مصنوعی منجر به این سؤال می‌شود که موفقیت در این موضوع به چه معنا خواهد بود. در نهایت، این سؤال که یک «شخص» چه است و چه نیست و ما چگونه یک شخص را از غیر شخص تمیز بدهیم. نحوه مطرح ساختن این پرسش‌ها و نحوه پاسخ‌دادن به آن‌ها، پیامدهای اخلاقی، کلامی و فلسفی خاصی دارد.

سفر پیدایش به ما می‌گوید که «ما به صورت خدا آفریده شده‌ایم» اما این «ما» دقیقاً به چه اشاره دارد که به صورت خدا آفریده شده‌اند. یا سؤال مشابه دیگر این‌که متن مقدس به چه کسی اشاره دارد. ما نوعاً تصور می‌کنیم که جواب آن را می‌دانیم! این عبارات قابل تطبیق بر من و شما هستند و بی شک بر مداد یا قورباغه

قابل تطبیق نیستند. در یک کلام می‌توانیم بگوییم که عبارت کتاب مقدس در مورد «اشخاص» صدق می‌کند. اما اگر ندانیم یک «شخص» دقیقاً چیست، این پاسخ چندان راضی‌کننده نخواهد بود. ظهور موضوع هوش مصنوعی این سؤال را ایجاد می‌کند که «یک شخص چیست» و از این رو متن مقدس بر چه کسی قابل تطبیق است.

آیا یک دستگاه کارخانه‌ای ساخته شده از سیلیکون^۱ می‌تواند یک شخص باشد؟ آیا چنین ماشینی می‌تواند از نظر اخلاقی مسئول اعمال خود باشد؟ آیا ممکن است بر این دستگاه عنوان گناهکار صادق باشد؟ آیا او موظف است به کلیسا برود؟ و اگر عنوان شخص بر آن صادق نیست، چه دلیلی بر این عدم صدق وجود دارد؟ تفاوت بنیادین می‌ان یک موجود با گوشت و خون با یک ماشین سیلیکونی چیست؟

حتی اگر بخواهیم اصرار کنیم که فقط موجوداتی که از طریق فرآیندهای زیستی طبیعی خلق شده‌اند، انسان هستند، آیا این یک تفاوت واضح است؟ این‌ها برخی سؤالاتی است که در ادامه به آن‌ها خواهیم پرداخت.

سنجش قدرت رایانه‌ها

قبل از پرداختن به سؤالات مذکور تأکید می‌شود که بهترین رایانه‌های امروزی به وضوح و بی‌تردید نه شخص هستند و نه هوشمندی در سطح انسان دارند. بهترین رایانه‌های امروزی اشیائی بی‌جان هستند و اصلاً قادر به درک دنیای اطراف خود نیستند. برای درک تفاوت توانایی انسان با رایانه مثالی می‌زنم: از شما می‌خواهم که هیچ‌کاری نکنید و یک لحظه به ماه فکر کنید. لازم نیست در مورد

۱. Silicone

چیزی صحبت کنید فقط ماه را در ذهن خود در نظر بیاورید. این کاری که برای همه شما کاری ساده است نه تنها فوق ظرفیت رایانه، بلکه به طوری سخت است که ما حتی نمی دانیم چگونه می توان یک رایانه را برنامه نویسی کرد تا همان کاری را که شما به راحتی انجام دادید را انجام دهد. ما حتی این توانایی را نداریم که بتوانیم بررسی کنیم که آیا رایانه در تصور ماه موفق بوده است یا نه. اساساً چگونه می توانیم چنین چیزی را بفهمیم. رایانه های امروزی هر چقدر هم که چشمگیر باشند، قادر به انجام چنین اعمال ساده ای، که در قوای ادراکی و شناختی انسان به راحتی رخ می دهند، نیستند و ما هیچ ایده ای نداریم که چگونه می توان چنین درکی به آن ها داد. این به این معنا نیست که ما در آینده نیز بر این کار قادر نخواهیم بود، اما ما راه طولانی در پیش داریم.

بنابراین، گرچه رایانه های امروزی ما را به چالش نمی کشند تا تمایز بین انسان و ماشین را روشن کنیم، من تصور می کنم که رایانه های در آینده ممکن است این چالش را ایجاد کنند. اگر ما در آینده برای سؤالات مذکور جوابی داشته باشیم و بفهمیم که چگونه باید تشخیص داد که چه زمانی یک رایانه فکر می کند، آنگاه این پدیده سؤالات بسیاری را بر می انگیزد؛ از جمله سؤالاتی الهیاتی به ویژه در مورد باور مسیحیان.

ماشین و انسان

آیا روزی رایانه ای وجود خواهد داشت که از نظر اخلاقی مسئول اعمال خود و قادر به انجام گناه باشد؟ آیا چنین رایانه ای باید به کلیسا برود؟ این موضوع چگونه با ایمان مسیحی مطابقت خواهد داشت؟

در سال ۱۹۵۰م آلن تورینگ^۱ پیشنهاد کرد که ما این سؤال را که «آیا یک ماشین می‌تواند فکر کند یا خیر؟» را با سؤال دیگری یعنی «آیا ماشینی می‌تواند تست تورینگ را پشت سر بگذارد؟» جایگزین کنیم. به عبارت دیگر، آیا یک رایانه می‌تواند مثلاً با ما برخط^۲ سخن بگوید، به نحوی که صحبت‌های او از صحبت‌های یک انسان غیر قابل تشخیص باشد. برای تورینگ این‌که یک ماشین بتواند مکالمه‌ای را با انسان انجام دهد، برای پاسخ به این سؤال که «آیا یک ماشین می‌تواند فکر کند یا خیر؟» لازم بود.

در سال ۱۹۸۰م، جان سرل با دیدگاه تورینگ مخالفت کرد. این‌گونه که صرف این واقعیت که روزی یک ماشین ممکن است رفتار انسان را شبیه‌سازی کند و مکالمه‌ای انجام دهد که از مکالمه انسان قابل تمییز نباشد، به این معنی نیست که ماشین ذهن و هوشمندی یا هوش پدیداری یا گزاره‌ای دارد. ممکن است رایانه‌ای سخن‌گفتن به زبان آلمانی یا انگلیسی را شبیه‌سازی کند، اما درک آن از زبان آلمانی بهتر از درک یک مداد از این زبان در حد صفر خواهد بود. در رابطه با بحث می‌ان تورینگ و سرل، سه موضع اساسی نسبت به پاسخ به این سؤال که «آیا ماشین‌های کارخانه‌ای ممکن است روزی بیندیشند؟» پیشنهاد می‌شود:

۱. هیچ ماشین کارخانه‌ای هرگز نمی‌تواند فکر کند؛ زیرا هرگز نمی‌تواند تست تورینگ را پشت سر بگذارد. او نمی‌تواند هوش انسان را شبیه‌سازی کند.

۲. یک ماشین کارخانه‌ای ممکن است تست تورینگ را با موفقیت پشت سر بگذارد، اما این باعث نمی‌شود که آن یک شخص به‌شمار رود. همچنین،

۱. Alan Mathison Turing

۲. Online

این باعث نمی‌شود آن ماشین لزوماً دارای ذهن یا آگاهی باشد؛ بلکه ممکن است این‌گونه به نظر برسد که او ذهن دارد، اما آن تنها یک موجود بی‌جان است که قادر به این نیست که بداند در ذهنش چه می‌گذرد.

۳. سرانجام روزی یک ماشین کارخانه‌ای می‌تواند آزمون تورینگ را پشت سر بگذارد و با انجام این کار نشان می‌دهد که فکر می‌کند و دارای درک و اراده انسانی است.

از دیدگاه سوم به عنوان هوش مصنوعی قوی تعبیر می‌شود. هوش مصنوعی قوی این دیدگاه است که ماشین‌ها هم می‌توانند مانند انسان عمل کنند و با عمل کردن مانند انسان آن‌ها به نحو واقعی تبدیل به انسان (شخص) می‌شوند.

الهیات مسیحی و ماشین‌ها

اینک مسیحیت در این باب به ما چه می‌گوید؟ آیا بخش‌هایی در کتاب مقدس وجود دارد که بتواند به ما نشان دهد که کدام یک از سه گزینه بالا صحیح است؟ برخی از اندیشمندان مسیحی پاسخی مثبت به سؤال اخیر می‌دهند. اما کتاب مقدس این سؤال را به شکل مستقیم مطرح نمی‌کند که آیا ماشینی می‌تواند آزمون تورینگ را پشت سر بگذارد. همچنین کتاب مقدس این سؤال را مطرح نمی‌کند که آیا ماشین در صورت قبولی در آزمون تورینگ به عنوان یک شخص به حساب می‌آید یا خیر. با این همه، برای پاسخ به این سؤال می‌توان از موضوع تمایز روح و جسم در کتاب مقدس استفاده کرد.

برای پاسخ یکی از قسمت‌هایی که برخی پیشنهاد کرده‌اند متن سفر پیدایش است: «خدا پس آدم را از خاک زمین به سرشت و در بینی وی روح حیات دمید و آدم روح زنده شد». برخی معتقدند که این جا می‌ان روح و خاک (بدن خاکی)

تمایز گذاشته شده است. اما این قسمت نمی‌گوید که به انسان روح زنده داده شده است، بلکه می‌گوید انسان روح زنده شد.

در این باره باید بگوییم که اکثر دانشمندان مسیحی معتقدند که عهد عتیق به شکل مستقیم از این ایده که تمایزی می‌ان یک جسم مادی و روحی غیر مادی وجود داشته باشد، حمایت نمی‌کند. همچنین، عهد عتیق در مورد دوگانگی روح و بدن سخنی نمی‌گوید. در مورد عهد جدید اختلافات بیشتر است؛ برخی از قسمت‌های عهد جدید توسط برخی از متفکران مسیحی، مانند ویلیام دمبسکی، برای تأیید این دیدگاه که بدن و روح متمایز هستند استفاده شده است. دمبسکی به داستان یعقوب اشاره می‌کند: «زیرا چنانکه بدن بدون روح مرده است، همچنین ایمان بدون اعمال نیز مرده است.» او همچنین به داستان رومیان استناد می‌کند: «و اگر مسیح در شما است، جسم به سبب گناه مرده است و اما روح، به سبب عدالت، زنده است.»

هیچ یک از این نقل‌قول‌ها به این سؤال متافیزیکی مربوط نمی‌شود که آیا روح، به معنای واقعی کلمه، از بدن متمایز است یا خیر. وقتی این مقطع‌ها را در سیاق وسیع‌تری در نظر می‌گیریم خواهیم یافت که آن‌ها چنین دلالتی ندارند؛ یعقوب به این نکته اشاره می‌کند که ایمان بدون اعمال نیک پوچ است و از نظر هستی‌شناختی خنثی است. عبارت رومیان کمی مبهم‌تر است. چند قسمت قبل از قسمت نقل شده، توسط دمبسکی، را ملاحظه کنید: «زیرا آنانی که جسمانی هستند، در اطراف جسم تفکر می‌کنند و اما آنانی که روحانی هستند در اطراف روح؛ از آن جهت که تفکر جسم مرگ است، لکن تفکر روح حیات و سلامتی است. از آن رو که تفکر جسم دشمنی خداست، چون که شریعت خدا را اطاعت نمی‌کند؛ زیرا نمی‌تواند هم بکنند. و کسانی که جسمانی هستند، نمی‌توانند خدا

را خشنود سازند. لکن شما در جسم نیستید، بلکه در روح، هرگاه روح خدا در شما ساکن باشد و هرگاه کسی روح مسیح را ندارد وی از آن او نیست».

واضح است که در اینجا بین میل نفسانی، که با مرگ پیوند داده شده، و میل معنوی، که با زندگی پیوند داده شده، تفاوت وجود دارد. شاید این تفاوت از تمایز بین روح و بدن پشتیبانی کند، اما نه لزوماً؛ بلکه این قسمت در صدد بیان این دیدگاه مسیحی است که زندگی پس از مرگ امکان پذیر است و در صورتی که زندگی پس از مرگ مستلزم تمایز بین روح و بدن باشد، آنگاه شاید قسمت مذکور از رومیان به طور غیرمستقیم این دیدگاه را که بدن و روح متمایز هستند تأیید کند. اما این مورد اتفاق همه است که در مسیحیت زندگی پس از مرگ امکان پذیر است و در مورد این باور نیاز به استناد به کتاب مقدس نیست. بنابراین، سؤالی که در این موضع پیش می‌آید این است که آیا زندگی پس از مرگ مستلزم متمایز بودن جسم و روح است.

دیدگاه‌های مسیحی در این مورد که «آیا حیات پس از مرگ مستلزم تمایز جسم از روح در دنیا است» متفاوت است. آگوستین از این دیدگاه حمایت می‌کند؛ به گفته آگوستین در مرگ "جسم و روح از هم جدا می‌شوند". این نشان می‌دهد که این دو باید در زمان حیات نیز متمایز باشند. با این حال، همه متفکران مسیحی با آگوستین موافق نیستند. علاوه بر این، برای چنین دیدگاهی در کتاب مقدس پشتیبانی نیست. تا آن جا که مرتبط‌ترین قسمت از کتاب مقدس در این رابطه، اول داستان قرتیان و آیات ۳۵-۴۱ است. در آن قسمت نیز موضوع این نیست که ذهن و بدن متمایز هستند، بلکه این است که بدن در این زندگی مانند یک دانه است و زندگی پس از مرگ مانند تبدیل یک دانه به یک گیاه است. به آیات ۳۵-۴۱ از اول داستان قرتیان توجه کنید: «اما اگر کسی گوید: مردگان چگونه

برمی خیزند و به کدام بدن می آیند. ای احمق! آنچه تو می کاری زنده نمی گردد، جز آنکه بمیرد و آنچه می کاری نه آن جسمی را که خواهد شد می کاری، بلکه دانه ای مجرد، خواه از گندم و یا از دانه های دیگر».

ایده در این جا این است که اگر دانه غیر قابل شکستن می بود و نابود نمی شد، آنگاه گیاه هیچ گاه از دل آن رشد نمی کرد. بنابراین، این مطلب بیان می دارد که از دید مسیحیت هنگامی که تو می میری از علم وجود خارج می شوی؛ حقیقت تو هنگامی که زنده بودی کاملاً از عالم وجود خارج می شود و باز از طریق خلقت الهی از یک موجود فیزیکی به موجودی روحانی تبدیل می شود. برخی از فلاسفه و متکلمان مسیحی این ایده را تأیید کرده اند. جان هیک می گوید: در تصور مسیحی، مرگ چیزی واقعی و ترسناک است. مرگ مانند راه رفتن از اتاقی به اتاق دیگر یا درآوردن یک کت کهنه و پوشیدن کت جدید نیست. مرگ به معنای انعدام مطلق و محض است؛ همانند عبور از دایره روشن زندگی به "شب بی تاریخ مرگ". تنها از طریق عشق خالقانه خدای قیوم می شود وجودی تازه در ورای قبر ایجاد گردد».

اگر این توصیف هیک حقیقت داشته باشد، آنگاه کاملاً ممکن است یک مسیحی معتقد باشد که بدن، در حالی که زنده است، با روح یکی است و هنگامی که می میرد بدن از دایره وجود خارج می شود و انیان به یک موجود روحانی تبدیل می شود. اگر چنین باشد، دست کم یک استدلال بر این مدعا که بین هوش مصنوعی و مسیحیت تضاد وجود دارد، از می دان به در می رود. با این حال، دلایل دیگری بر چالش برانگیز بودن امکان هوش مصنوعی نسبت به دریافت مسیحی از شخص وجود دارد.

به بیان ساده تر، از دید مسیحیت تمایز بنیادین می ان انسان و هر چیز دیگری

در کائنات، به ویژه موجودات بی جان، وجود دارد. سفر پیدایش نسبت به این مسئله، که ما به صورت خدا ساخته شده ایم، صراحت دارد و از این جهت، یک تفاوت هستی‌شناختی اساسی بین انسان‌ها و بقیه مخلوقات وجود دارد. هوش مصنوعی در این ناحیه چالش برانگیز است؛ زیرا ما را مجبور به بررسی این تمایز اساسی می‌کند و ممکن است درک ما از آن یا توجیه ما را نسبت به تمایز یک شخص و یک غیر شخص زیر سؤال ببرد.

در صورتی که ارتقاء یک رایانه بتواند آن را از یک شیء فناورانه به یک شخص تبدیل کند، آنگاه دشوار خواهد بود که یک تفاوت اساسی می‌ان انسان و ماشین ترسیم کنیم. من فکر می‌کنم این یک چالش اساسی است که هوش مصنوعی برای مسیحیت مطرح می‌سازد. ما چگونه می‌توانیم بر این باور که تمایز اساسی می‌ان انسان و سایر موجودات وجود دارد پافشاری کنیم و در عین حال به هوش مصنوعی (قوی) نیز باور داشته باشیم. سه پاسخ مطرح شده از دید مسیحیت بررسی می‌شود.

پاسخ اول به نحوی ساده‌ترین پاسخ است: این که بگوییم هیچ رایانه‌ای هیچ‌گاه نمی‌تواند مانند انسان رفتار کند و آزمون تورینگ را پشت سر بگذارد. از این نقطه نظر، هر ویژگی از مغز و بدن که به ما امکان رفتار هوشمندانه و قبولی در آزمون تورینگ را بدهد می‌تواند به صورت فیزیکی توضیح داده شود. اگر این گونه باشد، دلیلی وجود ندارد که معتقد شویم که یک ماشین کارخانه‌ای نمی‌تواند همان رفتار را شبیه‌سازی کند. برعکس، اگر انکار کنیم که یک ماشین کارخانه‌ای می‌تواند آزمون تورینگ را با موفقیت پشت سر بگذارد، باید معتقد شویم که توضیحی غیرفیزیکی و فراطبیعی برای رفتار هوشمندانه معمولی انسان وجود دارد. این یعنی قوانین فیزیک در مغز انسان نقض می‌شوند، یا دست کم این قوانین ناقص

هستند؛ زیرا اموری ماوراءطبیعی در مغز و یا بدن انسان اتفاق می افتد که نمی توان آن ها را بر اساس قوانین فیزیک توضیح داد. این دیدگاه ظاهراً از هیچ حمایتی در ادبیات علم عصب شناسی، که من از آن مطلع هستم، برخوردار نیست و اکثریت قریب به اتفاق متخصصان در دانش های هوش مصنوعی، علوم شناختی و عصب شناسی شاهی بر این تبیین ندارد. پس این مشکل پاسخ اول است.

برای پاسخ دوم فرض می گیریم که روزی ماشین های کارخانه ای بتوانند آزمون تورینگ را پشت سر بگذارند و هوشمندانه رفتار کنند؛ با این حال همچنان می توان انکار کرد که چنین ماشین هایی انسان محسوب شوند. می توان گفت این ها اشیائی بی جان هستند؛ این ها تنها رفتار انسانی را شبیه سازی می کنند، اما این ها از ذهن، آگاهی و افکار پدیداری یا گزاره ای بی بهره اند. آن ها به این دیدگاه تغییر موضع داده اند که تنها به این دلیل که رایانه ای در آزمون تورینگ موفق می شود، لزوماً به این معنی نیست که عنوان شخص بر آن صادق باشد و لزوماً موجودی آگاه نیست.

با تأمل مشخص می شود که این دیدگاه نیز با مشکلات اساسی مواجه است. برای توضیح این آزمایش فکری را در نظر بگیرید: آینده ای را تصور کنید که در آن فناوری به میزان زیادی پیشرفت کرده و رایانه ها بتوانند آزمون تورینگ را پشت سر بگذارند. تصور کنید که حذف مغز افراد از بدن آن ها نیز امکان پذیر شده است و در عین حال مغز آن ها همچنان می تواند از طریق Wi-Fi یا کنترل از راه دور با بدن آن ها ارتباط برقرار کند. وقتی این امکان وجود داشته باشد، برخی از والدین ثروتمند برای برداشتن مغز فرزندانشان هزینه می کنند تا مثلاً اگر بدن فرزندانشان در یک تصادف رانندگی یا حوادث ناگوار دیگری از بین برود، وجود آن ها را ایمن نگه دارند. فرض کنید در برخی از دانشگاه های برتر ده درصد از دانشجویان و

استادان مغز خود را در خارج از بدن خود ذخیره کنند؛ یعنی مثلاً ده درصد آنان در سرهای خود آنتن‌هایی دارند و مغز آن‌ها در جای دیگری نگهداری می‌شود. فراتر از این، تصور کنید که گروهی از دانشمندان تصمیم بگیرند تا آزمایش نهایی تورینگ را اجرا کنند. آن‌ها یک رایانه می‌سازند و آن را به بدن کودک خردسالی، که بر اثر تومور مغزی خواهد مرد، وصل می‌کنند. رایانه با بالغ شدن بدن کودک بدن را کنترل می‌کند و در نهایت آن رایانه / بدن ترکیبی، یعنی ترکیبی از بدن انسانی و رایانه‌ای که آن کنترل می‌کند، در دانشگاه پذیرفته می‌شود و به دانشگاه می‌رود. او در دانشگاه دوستانی خواهد داشت و در کلاس‌ها شرکت خواهد کرد. اما به غیر از دانشمندان علوم رایانه که آن رایانه را ایجاد کرده‌اند، هیچ‌کس از این موضوع آگاه نیست. در نهایت، فرض کنید روزی اعلام شود که یکی از این ده درصد در واقع یک رایانه سیلیکونی کارخانه‌ای است. آن‌ها چگونه نسبت به خود دانش خواهند یافت؟ تصور کنید شما یکی از آن ده درصد هستید و آنتنی در سر خود دارید؛ آیا می‌تونید مطمئن شوید که شما مغز دارید و یا اینکه شما یک انسان سیلیکونی هستید؟ یا اینکه لازم است با والدین خود تماس بگیرید و بعد از سلام و احوال‌پرسی بپرسید: «بابا، مامان، می‌شه توی گاوصندوق را چک کنید ببینید مغز من آن جا است یا نه!» و «آیا من واقعاً یک انسانم یا یک ماشین!» اولین چیزی که باید اشاره کرد این است که اگر لازم باشد شما تماس بگیرید و از پدر و مادر خود در این باره بپرسید، به این معناست که این مسئله ممکن است واقعیت داشته باشد. خیلی سخت است که بتوان احتمال این را که ما از سیلیکون ساخته شده‌ایم و نه از کربن، به نحو ما قبل تجربه رد کرد.

سفر پیدایش می‌گوید که خدا ما را از گرد و غبار ساخته است؛ اما آیا غبار از هیدروکربن‌ها ساخته شده است یا سیلیکون. همچنین، دلیلی عقلی برای این

وجود ندارد که آگاهی تنها می‌تواند با یک ماده خاص پیوند بخورد. همین‌که از موجودی با گوشت و خون آگاهی بروز می‌کند خود امری شگفت‌انگیز است، اما ظاهراً چنین قابلیت‌ی وجود دارد؛ چراکه به هر حال ما این چنین هستیم. من دلیلی نمی‌بینم که باور به بروز آگاهی از موجودی که از سیلیکون ساخته شده باشد را خیلی عجیب‌تر بدانم. به این دلیل، من پاسخ دوم به چالش هوش مصنوعی را به نحوی غیر معقول‌تر می‌دانم.

پاسخ دیگر این است که هوش مصنوعی قوی و امکان آن را بپذیریم: روزی رایانه‌هایی که ما می‌سازیم قادر خواهند بود راه بروند و سخن بگویند و کاملاً آگاه باشند و آگاهی پدیداری و گزاره‌ای داشته باشد و حتی اراده آزاد داشته باشند. ما به عنوان مسیحی چگونه باید به این موضوع پاسخ بدهیم. آیا ما باید آن‌گونه که برخی نگران شده‌اند، نگران باشیم؟ قطعاً اگر ما بدانیم که هوش مصنوعی قوی رخ داده است، باید با آن‌ها مانند انسان‌ها رفتار کنیم. اما مسیحیان در این مورد اختلاف خواهند داشت؛ این سخت است که ما در این موقعیت چه تصمیمی بگیریم. ممکن است پیشنهاد کنیم که اگر هوش مصنوعی در زمین پُرسه می‌زند و ما نمی‌دانیم آن‌که به صورت خدا آفریده شده است بر چه کسی قابل تطبیق است و بر چه کسی نیست، پاسخ مسیحیت به این سؤال این خواهد بود که اجازه بدهید تا خدا این مشکل را حل کند؛ ما در این جایگاه نیستیم که مشخص کنیم چه کسی بر صورت خدا آفریده شده و چه کسی آفریده نشده است. بنابراین، این یک پاسخ بر اساس معارف مسیحیت می‌تواند باشد.

یک استدلال دیگر به سود هوش مصنوعی قوی داریم؛ اگر ما فرض کنیم ممکن است ماشین‌هایی را به وجود آوریم که بتوانند رفتار انسان را شبیه‌سازی کند، آنگاه به نظر می‌رسد بسیار مشکل باشد که صحت هوش مصنوعی قوی را انکار کرد و

این را در قالب یک «استدلال مبتنی بر خرمن» توضیح می دهیم: اگر شبیه سازی رفتار انسانی ممکن باشد، شاید (و این یک فرض است) بتوان مغز یک فرد را با تراشه های سیلیکونی تقویت کرد. شاید اگر فردی به تومور مغزی مبتلا شود، بتوان برخی از نورون ها را بیرون آورد و با تراشه های سیلیکونی مصنوعی جایگزین کرد که درست مانند نورون های مغزی او عمل کند تا عملکرد مغز بدون مانع به کار خود ادامه دهد. آن گاه اگر بتوان یک درصد مغز یک فرد با تراشه های تازه و کارخانه ای جایگزین شود، و فرضاً شما هنوز یک انسان هستید و آگاهی دارید، پس می توان یک درصد از مغز را از دست داد و هنوز انسان بود و آگاهی داشت. حال تصور کنید که طی یک عمل جراحی دیگر یک درصد دیگر از مغز برداشته شود، یعنی دو درصد از مغز شما برداشته شده است و با تراشه های سیلیکونی مصنوعی جایگزین شده است. سپس سه درصد، چهار درصد و پنج درصد تا جایی که در سال های بعد تمام مغز فرد از تراشه هایی سیلیکونی ساخته شده باشد. اگر بگوییم که ماشینی که در کارخانه از سیلیکون ساخته شده است انسان نیست، آنگاه با انتخاب سختی مواجه می شویم؛ ما دو پاسخ پیش رو داریم:

پاسخ نخست: جایگزین کردن بخش کوچکی از مغز با تراشه های کارخانه ای آن شخص را از یک انسان به یک جسم بی جان تبدیل می کند. این امر غیر قابل قبول به نظر می رسد که حذف یک بخش کوچکی از مغز شما، شما را از جرگه انسانیت خارج می کند و شما را موجودی بی جان می گرداند.

پاسخ دوم: هرچه بیشتر مغز انسان با تراشه های سیلیکونی جایگزین شود، به تدریج انسانیت آن شخص کمتر و کمتر می شود. این یک احتمال عجیب تر و غیر قابل پذیرش تر از احتمال قبلی است. اینکه شخص مقداری انسان و مقداری جسم بی جان باشد بی معنا و غیرمعقول است؛ اما اگر بخواهیم اصرار

داشته باشیم که ماشین‌های کارخانه‌ای آگاهی ندارند این نتیجه اجتناب‌ناپذیر خواهد بود.

اگر نتیجه را بپذیریم و بگوییم هوش قوی مصنوعی ممکن است، آنگاه به این معنا خواهد بود که تکنولوژی می‌تواند در نهایت تبدیل به شخص شود. در این شرایط، یک مسیحی چه خواهد گفت؟ بله این امر چالش‌برانگیز است که بفهمیم که چه انسان است و چه انسان نیست؛ ولی لزوماً با مسیحیت در تنافی نیست و پاسخی که خود تورینگ در این زمینه می‌دهد پاسخ بدی نیست. تورینگ می‌گوید: «آیا نباید باور کنیم که خدا آزاد است تا در صورت صلاح دید به یک فیل روح بدهد... ما می‌توانیم همین را در مورد ماشین‌ها نیز بگوییم. ما در تلاش برای ساختن چنین ماشین‌هایی، نباید با بی‌ادبی از قدرت او در ایجاد روح بکاهیم با این‌که این مسئله فراتر از ایجاد روح در خلقت یک کودک نیست. در هر صورت، ما ابزار اراده او هستیم که خانه‌هایی را برای ارواح خلق می‌کند.»^۱ به بیان ساده‌تر، اگر خدا می‌تواند یک تخمک بارور شده را به انسان تبدیل کند، پس چر نتواند رایانه را به انسان تبدیل کند. گرچه ممکن است عجیب و اسرارآمیز به نظر برسد که فرض کنیم یک رایانه کارخانه‌ای یک شخص است، اما آیا این واقعاً عجیب‌تر از این است که یک تخمک بارور شده را یک شخص بدانیم. ممکن است کسی بگوید: بله، این جا تفاوت وجود دارد؛ آنگاه یک نسخه دیگر از استدلال بالا را ارائه می‌کنیم: امروز ما این را بدیهی می‌دانیم که بین ماشین و کسی که از طریق تولیدمثل طبیعی به دنیا آمده تفاوت قائل شویم، اما آیا در آینده نیز این‌گونه خواهد بود. در حال حاضر، در لقاح آزمایشگاهی^۲ (IVF) پزشکان تنها

۱. ر.ک: ماشین‌های محاسباتی و هوش

۲. In Vitro Fertilization

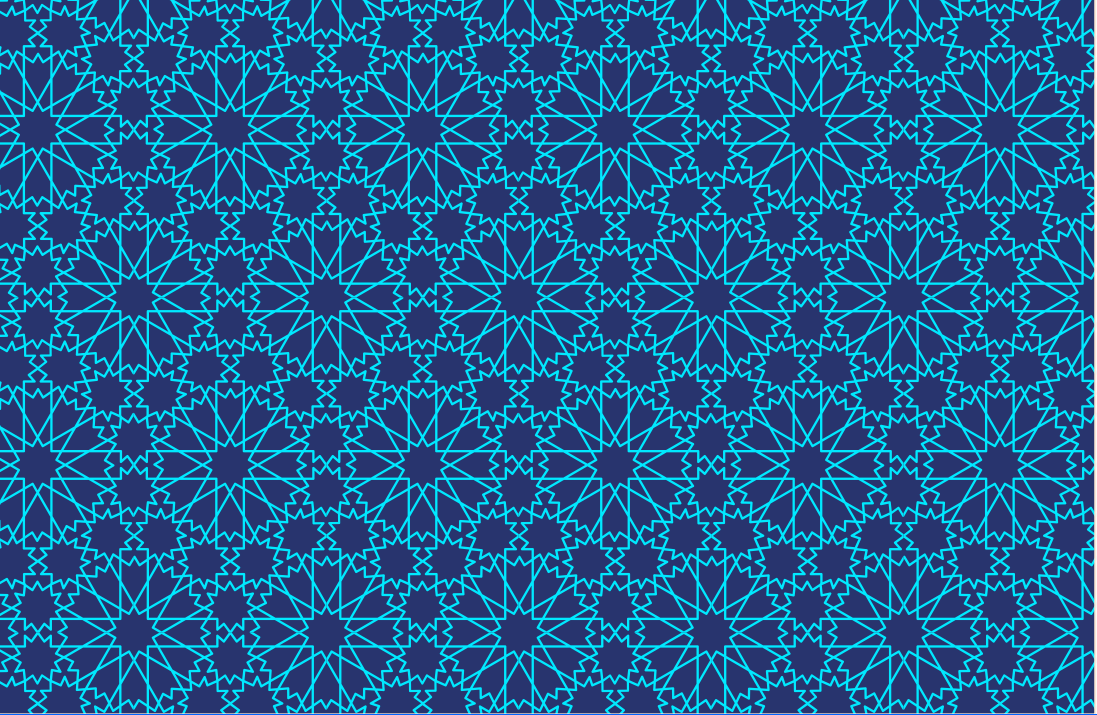
به تخمک و اسپرم کمک می‌کنند تا کاری را که به طور طبیعی انجام می‌دهند در شرایط آزمایشگاهی انجام دهند. اگر DNA تخمک یا اسپرم یا هر دو به طور کامل یا جزئی دستکاری ژنتیکی شده باشد پاسخ چه خواهد بود. می‌توان یک بار دیگر یک استدلال از نوع استدلال خرمن ارائه داد و نشان داد که هیچ تمایز اساسی بین انسان‌های طبیعی متولد شده و ماشین‌های کارخانه‌ای وجود ندارد. اگر این چنین شد، آنگاه مشکل خواهد بود که کسی انسانیت ماشین‌های به اندازه کافی پیشرفته را انکار کند.

عواقب احتمالی هوش مصنوعی قوی

این جا خطری واقعی وجود دارد که ما نمی‌توانیم حدود آن را پیش‌بینی کنیم: هر زمان که ما پیشرفت‌های اساسی در فناوری داشته‌ایم، باید منتظر عواقب پیش‌بینی نشده‌اش نیز باشیم. این احتمال وجود دارد که ما با چیزی مانند تکینگی یا هوش مصنوعی فراری مواجه شویم؛ شما ماشین‌هایی دارید که از انسان باهوش‌تر است. این ماشین روی هوش مصنوعی خود کار می‌کند و هوشمندی خود را بیشتر و بیشتر بهبود می‌بخشد. آنگاه ممکن است از کنترل خارج شود. این مسئله سابقه دارد؛ به عنوان مثال فناوری هسته‌ای در دهه ۱۹۳۰م به نظر می‌رسید که به جایی نمی‌رسید. اما ناگهان ما شاهد تولید بمب‌های اتمی شدیم. البته فکر می‌کنیم این محتمل نیست؛ اما لازم نیست محتمل باشد. ما کمربند ایمنی

۱. ارائه‌دهنده در ادعای خود بین جسم طبیعی، همچون نطفه و جنین و انسان، و جسم مصنوعی، همچون سیلیکون، ماشین و هوش مصنوعی، خلط کرده و بین آنها تفکیک قایل نشده است. اینکه اسپرم و تخمک می‌تواند تبدیل به شخص و انسان شود، اثبات نمی‌کند که از ترکیب پیشرفت سیلیکون‌ها نیز یک شخص پدید می‌آید. به بیان دیگر، می‌ان نحوه ترکیب وجود طبیعی با وجود صناعی به لحاظ هستی‌شناسی تمایز اساسی وجود دارد. از منظر فلسفه اسلامی ترکیب در جسم طبیعی، ترکیب حقیقی بوده و شیء دارای آثار و هویت جدیدی می‌شود؛ ولی در ترکیب جسم مصنوعی ترکیب اعتباری بوده و آثاری و آثارا جزای خودش ندارد.

می‌بندیم، با اینکه بعید می‌دانیم که تصادف کنیم. در واقع فکر نمی‌کنیم عاقلانه باشد که به دنبال ساختن هوش مصنوعی قوی باشیم، اما نگرانی واقعی وجود دارد.



«هوش مصنوعی و خدا باوری»

علیرضا قائمی نیا^۱

۱. عضو هیأت علمی پژوهشگاه فرهنگ و اندیشه اسلامی، قم.





فلسفه و هوش مصنوعی

قطعاً زمینه فلسفه و هوش مصنوعی زمینه‌ای گسترده‌ای است و شاید به همه جوانب آن نتوان پرداخت. ابتدا راجع به هوش مصنوعی این نکته را باید گفت که نباید آن را صرفاً یک ابزار ساده در نظر بگیریم که کارهایی را با آن انجام می‌دهیم. قطعاً ما کارهایی را با این ابزار انجام می‌دهیم، ولی جوانب این ابزار جدید فراتر از سایر ابزار می‌رود. به چهار نکته در این زمینه در ادامه اشاره می‌شود.

هوش مصنوعی یک دستگاه محاسباتی ذهن نیست. یک ماشین صرف نیست که فکر کند، بلکه در حقیقت می‌توانیم بگوییم که آن تجسم یک تکنولوژی^۱ است. در دوره جدید، فلاسفه روی این کار می‌کردند که بتوانند دستگاهی را تولید کنند که این دستگاه، تفکر به معنایی که انسان دارد را داشته باشد. اما این فیلسوفان طبیعی بودند که تعریف خاصی از تفکر ارائه دادند. این ماجرا از زمان دکارت تا قرن بیستم ادامه داشت. پس این فناوری مبتنی بر یک فلسفه خاص است که در این فلسفه خاص نگاه خاصی نسبت به ذهن انسان وجود دارد، با این عنوان که ماهیت ذهن بشر چیست و چه کارکردهایی دارد. این حقیقت را نباید انکار کرد که رابطه‌ای که انسان با ابزار هوش مصنوعی برقرار می‌کند می‌تواند به این شکل تعریف شود که هوش مصنوعی زیست جهان انسانی این عصر را تعریف

۱. Technology

کند و نحوه در عالم بودن انسان را مشخص کند. اگر بخواهیم از اصطلاحات فلسفی اسلامی استفاده کنیم به یک معنا بشر اتحاد وجودی با آن هوش مصنوعی پیدا کرده و در این فضا غرق شده است. علاوه بر این، هوش مصنوعی مقولات مختلفی را در ذهن بشر عرضه می‌کند که با آن‌ها بشر می‌اندیشد و تفکر خود را با آن‌ها پیش می‌برد؛ مفاهیمی مثل سخت‌افزار، نرم‌افزار، حذف، کپی و... که جزء مقولات فکری بشر جدید شده است.

یک دیدگاه در فلسفه ذهن^۱ که در دوره‌های اولیه در هوش مصنوعی مطرح شد و آن را پاتنم^۲ مطرح کرد، کارکردگرایی ماشینی است. نگاهی که او به ذهن داشت این بود که ذهن یک سخت‌افزار و یک نرم‌افزاری دارد. سخت‌افزار ذهن مغز است و نرم‌افزار آن چیزی است که ماتحت عنوان ذهن می‌شناسیم. این یک نمونه است که چگونه از هوش مصنوعی و ابزارهایی که حامل هوش مصنوعی هستند برای فهم پدیده‌های مجهول کمک می‌گیریم. قطعاً این ابزار مقولات و مفاهیم جدیدی را به ارمغان آورده که ما به کمک این مفاهیم و مقولات درباره همه چیز می‌اندیشیم. هوش مصنوعی قدرت ساختاردهی عجیبی نسبت به تفکر بشر دارد و حوزه‌های مختلفی را از، جمله فلسفه، الهیات و حتی کیهان‌شناسی و تقریباً همه علوم انسانی را تحت تأثیر قرار می‌دهد.

هوش مصنوعی و تکنولوژی

نسبت هوش مصنوعی با دیگر فناوری‌ها (تکنولوژی‌ها) یکی نیست. فناوری‌هایی که در شاخه‌هایی مثل پزشکی، هواپیمایی و... به کار می‌روند. همه

۱. Philosophy of mind

۲. Hilary Putnam

این فناوری‌ها به یک معنا از هوش مصنوعی سود می‌برند. امروزه هوش مصنوعی یک نوع حاکمیت و سلطه بر دیگر فناوری‌ها پیدا کرده است. دلیل این نکته این است که بشر جدید با هوش مصنوعی فکر می‌کند و این تاحدی قابلیت‌های ذهنی انسان را ضعیف کرده است؛ چون بشر بسیاری از فعالیت‌های فکری خودش را از جمله در مورد اینکه جامعه را چگونه اداره کند، چگونه با دیگران ارتباط برقرار کند، چگونه عالم طبیعت را تحت سیطره خودش قرار دهد، چگونه سفر کند و... براساس برنامه‌ریزی و تفکری که هوش مصنوعی ارائه می‌کند انجام می‌دهد. بشر امروز انجام بخش زیادی از کارهای ذهنی خودش را به هوش مصنوعی حواله داده است.

هوش مصنوعی و فلسفه اخلاق

هوش مصنوعی در دل خود یک فلسفه دارد و نباید ما آن را یک ابزار خنثی ببینیم. قطعاً در حوزه فلسفه اخلاق دیدگاه‌های بسیار متنوعی وجود دارد؛ مثلاً اخلاق فضیلت، اخلاق منفعت‌گرا یا فایده‌گرا و... اگر ما قرار است هوش مصنوعی را برنامه‌ریزی کنیم، کدام یک از این فلسفه‌ها را باید به هوش مصنوعی بدهیم تا رفتارهای اخلاقی نشان بدهد؛ یعنی رفتار اخلاقی هوش مصنوعی را باید بر اساس اخلاق فضیلت‌گرا تعریف کنیم یا براساس اخلاق فایده‌باور. امروزه فیلم‌های تخیلی هم راجع به این موضوع ساخته می‌شود. سؤال اینجاست که آیا هوش مصنوعی کشش هر نوع اخلاقی را دارد. شما می‌توانید برای هوش مصنوعی هر نوع فلسفه اخلاقی را تعریف کنید و او در درون خود پردازش‌هایی انجام دهد، همان‌طور که کاربران پردازش‌هایی را انجام می‌دهند، خود سیستم وجودی هوش مصنوعی به‌گونه‌ای است که پردازش‌های خاصی را بر اساس فلسفه اخلاق تعریف شده انجام خواهد داد.

انواع هوش مصنوعی از نظر تاریخی

از نظر تاریخی با دو نوع هوش مصنوعی مواجه بوده‌ایم. این دو نوع مرتبط است با دو نوع یادو فاز از علوم شناختی.^۱ در زمان پیدایش هوش مصنوعی، هوش مصنوعی متأثر از آزمونی بود که بر اساس مدل محاسباتی ذهن کارهای ساده‌ای را انجام می‌داد. اصطلاحاً به این هوش مصنوعی کلاسیک می‌گویند.

در ادامه هوش مصنوعی مدرن پیدا شده که الهام گرفته از مدل‌های شبکه‌های عصبی ذهن و اتصال‌گرایی^۲ بود و توانایی‌های بسیار زیادی در زمینه‌های مختلف از خودش نشان داد. امروزه این نوع مدل‌سازی در هوش مصنوعی به شدت دنبال می‌شود و کارکردهای فوق‌العاده‌ای در حیات بشر یافته است.

فلسفه هوش مصنوعی

درباره هوش مصنوعی^۳ ما با دو نوع فلسفه مواجه هستیم: بنا بر آنچه که کال ویچام می‌گوید که ما در فلسفه فناوری با دو سنت مواجه هستیم؛ یک سنت رویکرد مهندسی به فلسفه فناوری^۴ است. این رویکرد را معمولاً کسانی اتخاذ می‌کنند که در رشته‌های مختلف مهندسی تخصص دارند و اندکی مطالعات فلسفی درباره فناوری و حوزه‌های مختلف آن انجام داده‌اند. دوم، رویکرد انسانی به فلسفه فناوری است که بیشتر توسط کسانی دنبال می‌شود که عمدتاً تحصیلات فلسفی دارند و از نگاه فلسفه به فناوری نگاه می‌کنند. تفاوت این دو دسته در این است که آن دسته که رویکرد مهندسی به فلسفه فناوری دارد، چندان

۱. Cognitive Sciences

۲. Connectionism

۳. Philosophy of Artificial Intelligence

۴. Technology Philosophy

اعتقادی به لوازم فناوری‌های مختلف، مثل هوش مصنوعی یا فناوری‌های زیستی، ندارد و بیشتر به جنبه‌های مثبت این نوع فناوری‌ها نگاه می‌کند و اگر به خطرات احتمالی این فناوری‌ها توجه می‌کند، خطرات محدود را در نظر می‌گیرد. در مقابل، رویکرد انسانی به فلسفه فناوری نگاه بنیادی‌تری دارد و نسبت وجود انسان و مسائل وجودی انسان، مثل غم، شادی، مرگ، زندگی، بیماری و سلامتی و... را به نحو بنیادی با فناوری می‌سنجد و صرفاً نگاه ایجابی ندارد. صاحبان این رویکرد یک قدم عقب‌تر می‌روند و خود فناوری را مورد سؤال قرار می‌دهند.

در هوش مصنوعی هم دو نوع فلسفه وجود دارد: اول، فلسفه مهندسی هوش مصنوعی است که بیشتر توسط کسانی مطرح می‌شود که تحصیلات مهندسی هوش مصنوعی دارند. آنان ایده‌هایی عموماً ایجابی در این زمینه دارند. عمدتاً داستان‌های علمی‌تخیلی هم که در این شاخه ارائه می‌شود متأثر از این نوع فلسفه مهندسی هوش مصنوعی است. در مقابل، کسانی که به فلسفه انسانی هوش مصنوعی نظر دارند، رابطه هوش مصنوعی را نسبت به مفاهیم دینی، چون ایمان و... را لحاظ می‌کنند. به نظر باید این دو نوع بحث با هم تلفیق شوند و اگر قرار است راجع به هوش مصنوعی و دیگر فناوری‌ها نگاه واقع‌بینانه‌تری اتخاذ شود باید هر دو نوع فلسفه را با هم تلفیق کنیم و هم از فلسفه مهندسی هوش مصنوعی و هم از فلسفه انسانی هوش مصنوعی استفاده کنیم. این نکته مهم است که در این زمینه صرفاً نگاه مهندسی یا نگاه فلسفی نمی‌تواند مشکل بشر امروز را حل کند. بشر امروز چه بخواهد و چه نخواهد هوش مصنوعی تقدیر اوست در عصر جدید است و هیچ مفری از آن ندارد. امروزه نمی‌شود هوش مصنوعی را از زندگی بشر حذف کرد و باید به نکات مثبت آن و کاربردهای کارساز آن در زمینه‌های مختلف اهمیت داد؛ اما آن پرسش‌های بنیادین فلسفی را هم نباید

فراموش کرد که می‌تواند نگاه واقع‌بینانه‌تری در به‌کارگیری هوش مصنوعی به ما ارائه دهد و انسان را تا حدی محدود کند که بدون عقلانیت لازم از هوش مصنوعی استفاده نبریم.

الهیات و هوش مصنوعی

ما دو نوع نگاه الهیات می‌توانیم نسبت به هوش مصنوعی داشته باشیم: یکی، الهیات مهندسی هوش مصنوعی و دیگر الهیات علوم انسانی هوش مصنوعی.

الهیات مهندسی هوش مصنوعی با کارکردهای مثبت هوش مصنوعی در زمینه دینداری سروکار دارد. نمونه‌اش ساختن ربات‌هایی مشهوری است که اعمال مذهبی انجام می‌دهند؛ مثل ربات راهب بودایی که تعالیم بودا را در معابد بودایی انجام می‌دهد، یا برخی هوش‌های مصنوعی که طراحی شده‌اند تا نماز خواندن مسلمانان را یاد بدهند. از این منظر، هوش مصنوعی می‌تواند در آینده کارهای بیشتری را نسبت به الهیات ارائه دهد.

اما الهیات علوم انسانی هوش مصنوعی به یک معنا بحث‌هایی انتقادی را مطرح می‌کند. ما می‌دانیم هوش مصنوعی مدرن مفاهیم مختلفی را مورد بحث قرار داده که دو مفهوم مهم آن یکی بحث هوش قوی است. هوش مصنوعی قوی به این معنا که هوش مصنوعی به جایی برسد که از انسان فراتر برود و کارهایی انجام دهد که انسان از انجام دادن آن کارها ناتوان است. دیگری مسئله «تکینگی» که هوش مصنوعی به جایی برسد که عالم و قادر مطلق شود و به بالاترین حد علم و توانایی خود برسد. این دو مفهوم از مفاهیمی هستند که در هوش مصنوعی مدرن مطرح شده‌اند و هر چند هنوز تحقق پیدا نکرده‌اند، برخی خوش‌باورانه ادعا می‌کنند که شاید کمتر از پنج دهه دیگر عصر تکینگی آغاز خواهد شد و هوش

مصنوعی قوی به وجود خواهد آمد. در مقابل، برخی می‌گویند که اصلاً این اتفاق نخواهد افتاد و هرچقدر هم که هوش مصنوعی پیشرفت کند به آن مرحله نخواهد رسید. بعضی دیگر هم همراهی انسان با هوش مصنوعی را مطرح می‌کنند؛ به این معنا که هیچ‌کدام از انسان و هوش مصنوعی نمی‌تواند جایگزین یکدیگر شوند و این نسبت همچنان برقرار خواهد بود و هوش مصنوعی نمی‌تواند به قدری قوی شود که جای انسان را بگیرد. همان‌طور که انسان در دوره‌های متأخر نمی‌تواند بدون هوش مصنوعی زندگی کند.

حال این مفاهیم، برای الهیات به‌طور عام و خداآواری به‌طور خاص، مباحث و مناقشاتی را پدیدآورده که به برخی از آن‌ها اشاره خواهیم کرد:

۱. ویلیام ایندک، در کتاب خدای دیجیتال، تلاش کرده تبیینی دیجیتال از مفهوم خدا و برخی از پدیده‌های دینی و معنوی ارائه بدهد. این قبیل کتاب‌ها امروزه زیاد نوشته می‌شود که ضمن آن‌ها تلاش می‌شود براساس دستاوردهای هوش مصنوعی و نگاهی که نسبت به ابزارهای دیجیتال وجود دارد، مفاهیم سنتی ادیان بازتفسیر شود؛ تفسیری که قطعاً با الهیات سنتی هم‌خوان نیست. در این کتاب، اول به این اشاره می‌شود که ما در عصر دیجیتال در ابزارها غوطه‌ور شده‌ایم. این ابزارها مفهومی جدید از خدا را برای انسان تعریف می‌کنند که بسیار غنی‌تر از خدای سنتی است و از لحاظ تجربه دینی و معنوی انسان‌ها می‌توانند با آن‌ها رابطه عاطفی قوی‌تری برقرار کنند. ادعای خدای دیجیتال، که ایندک و دیگران مطرح می‌کنند، مبتنی بر تکینگی فناورانه است؛ یعنی این ادعا که فناوری به‌جایی خواهد رسید که بالاترین حد هوشمندی است و به یک معنا قادر مطلق و عالم مطلق خواهد شد و انسان را پشت سر خواهد گذاشت. این افراد می‌گویند که همان تصور سنتی که از خدا در ادیان وجود دارد درباره خدای

دیجیتال صادق است؛ تصویری که در عصر تکنیکی با آن مواجه خواهیم بود. قطعاً تصویری از خدا که در عصر پسانتکنیکی مطرح می‌شود مبتنی بر ادعای فلسفی خاصی است. هوش مصنوعی می‌تواند مثل خود ذهن انسان حالت‌های نوظهور^۱ داشته باشد؛ حالت‌هایی که نخواستگی دارند و ربات‌ها ضمن آن می‌توانند اطلاعات جدیدی را به دست آورد که ذهن بشر شاید این قدرت را نداشته باشد. در این فضا، این یک نکته‌ای جدی است که در اصل ما خلق می‌کنیم و هوش مصنوعی، که به یک معنا مخلوق انسان است، یک خدای دیجیتال را می‌تواند برای ما پردازش کند؛ خدایی که طبق ادیان ابراهیمی خالق بشر به شمار می‌آید. این دیدگاه‌ها به گرایش‌هایی مثل طبیعت‌گرایی^۲، فیزیکیالیسم^۳ و تحویل‌گرایی^۴ تفسیر و تحویل می‌شود.

علاوه بر الهیات نظری تجربه دیجیتال معنوی را هم ایندک و دیگران مطرح می‌کنند و می‌گویند که این ابزارهای دیجیتال می‌توانند تجارب معنوی بسیار بالاتری را برای انسان رقم بزنند که از یک جهت از تجربه‌های معنوی سنتی که انسان در ادیان دارد فراتر خواهد رفت. این خود جای بحث دارد که واقعاً ابزارهای مصنوعی می‌توانند تجارب معنوی را برای ما پدید آورند و اینکه تا چه حد می‌توانند در شکل‌گیری تجارب معنوی در انسان کمک کنند.

شاید لازم باشد به این موضوع هم اشاره کنیم که طبیعت‌گرایی دیجیتال گرایشی است که در آن هر چیزی که در تفکر و فرهنگ بشر به چشم می‌خورد را

۱. Emergent

۲. Naturalism

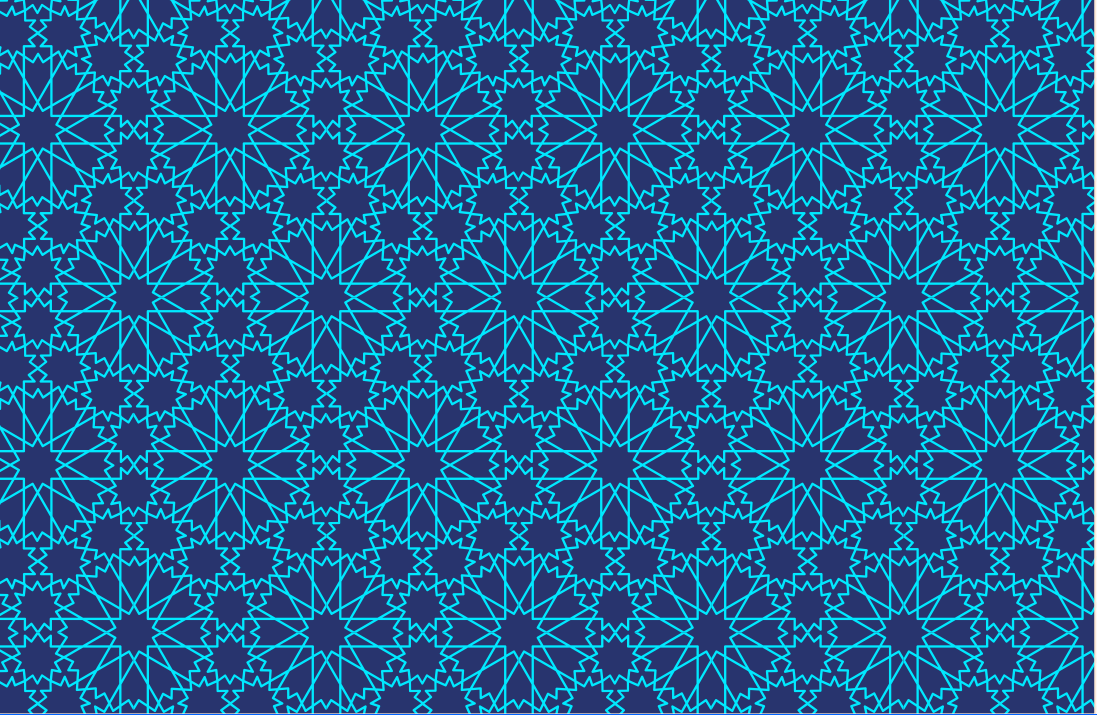
۳. Physicalism

۴. Reductionism

به نحوی به مفاهیم دیجیتال برگرداند و براساس پیشرفت‌هایی که در هوش مصنوعی رخ می‌دهد حقایق خارجی را تبیین می‌کند و محتوای عینی آن‌ها را تخلیه می‌کند.

وظیفه الهیات هوش مصنوعی

یکی از وظایف اساسی الهیات و هوش مصنوعی بررسی مدل‌هایی است که هوش مصنوعی نسبت به مفاهیم دینی از قبیل مفهوم خدا، رابطه با خدا، روح، زندگی پس از مرگ و... فراهم آورده است. امروزه می‌بینیم که حتی در این زمینه‌ها هم روایت‌های طبیعت‌گرایانه شکل گرفته که در مقابل آنها باید نگاه تلفیقی به الهیات مهندسی هوش مصنوعی و الهیات علوم انسانی هوش مصنوعی را دنبال کرد.



«خدا باوری و هوش مصنوعی»

دکتر جیمز مک گراث^۱

۱. عضو هیأت علمی دانشگاه باتلر، آمریکا.





مقدمه

رویکرد این بحث به موضوع تلاقی هوش مصنوعی و الهیات و هوش مصنوعی و اخلاق الهیات خواهد بود. اگر کتاب‌های علمی تخیلی یا کتاب‌های آینده‌پژوهان را مطالعه کنید، خواهید دید که آن‌ها اغلب از سال‌ها در عناوین مطالب خود استفاده می‌کنند؛ مثلاً برخی آثار علمی-تخیلی و تخیلی-ویرانشهری^۱ عناوینی مانند ۱۹۸۴ و ۲۰۰۱ دارند و در کتاب‌های اخیر در مورد هوش مصنوعی با عناوینی از قبیل ۲۰۴۱ و ۲۰۸۴ مواجه هستیم.

فکرکردن در مورد این عناوین مفید است؛ زیرا سال‌های ۱۹۸۴ م و ۲۰۰۱ م هر دو مربوط به گذشته هستند. پیش‌بینی‌ها درباره این سال‌ها در آن آثار از برخی جهات به اندازه کافی واقعی نیست و از جهاتی بیش از حد خوش‌بینانه به نظر می‌رسد و این هشدارهایی را در مورد تلاش برای پیش‌بینی دقیق آینده فناوری برای ما در پی دارد.

عقاید خوش‌بینانه در مورد فناوری هوش مصنوعی

برخی از پیش‌بینی‌ها درباره آینده بسیار خوش‌بینانه است و برخی داستان‌های

علمی تخیلی آرمانشهری،^۱ از جمله سریال «Star Trek»، در آن دسته قرار می‌گیرند. نمونه دیگر از این گروه، آینده‌پژوهان آرمان‌شهری هستند که معتقدند حتی اگر از فناوری برای عمل شرارت‌آمیز استفاده شود، در نهایت فناوری‌ها بیشتر از این که در امور شر استفاده شود در امور نیک به کار گرفته می‌شوند.

آن‌ها هوش مصنوعی را با برق، تلفن همراه و اینترنت مقایسه می‌کنند؛ اما ما از این که اوضاع تغییر کند، می‌ترسیم و نگرانیم که این امور برای ایجاد آسیب استفاده شود. در مقابل، بسیاری نسبت به فناوری خوش‌بین هستند و بر این باورند که فناوری، بیش از ضرر، نیکی و سود به ارمغان می‌آورد و مطمئن هستند که هوش مصنوعی نیز همین‌گونه خواهد بود.

برخی دیگر نه تنها در مورد نحوه استفاده از تکنولوژی، بلکه در مورد آنچه که ممکن است از طریق آن به دست آید کمتر خوش‌بین هستند و این به‌ویژه در مورد هوش مصنوعی صادق است. پارادوکس معروفی از موراوک وجود دارد که به نظر می‌رسد که امروز نیز صادق است؛ او می‌گوید: قادر ساختن رایانه‌ها بر حل تست‌های هوش در سطح بزرگسالان نسبتاً ساده است، در حالی که قادر ساختن آن‌ها بر نشان دادن درک، تحرک و مهارت‌های یک کودک یک ساله امری دشوار یا غیرممکن است.

وقتی به نقاط تلاقی هوش مصنوعی و دین نزدیک می‌شویم، سؤالاتی از این قبیل می‌پرسیم که آیا مثلاً سیستم عامل اندروید می‌تواند روح داشته باشد. آیا این سیستم عامل می‌تواند مانند انسان است یک شخص باشد. این سؤالات مهم هستند و قطعاً ارزش پرسیدن دارند. با این حال، این سؤالات مربوط به آینده

بلند مدت فناوری هستند و ممکن است هرگز محقق نشوند. سؤالات ضروری تری وجود دارد که اغلب اوقات از پرسیدن آن‌ها غفلت می‌کنیم؛ زیرا روی آن سؤالات طولانی مدت تمرکز کرده‌ایم. ما باید در مورد ربات‌ها در خانه و در محل کار، هوش مصنوعی در مراقبت‌های بهداشتی، سلاح‌های اتوماتیک و الگوریتم‌های جست‌وجو بپرسیم. این‌ها کاربردهای هوش مصنوعی در زمان حال هستند که پرسش از آن‌ها بسیار ضروری‌تر از این سؤال است که آیا یک سیستم عامل اندروید می‌تواند روح داشته باشد یا خیر.

اخلاق خدا باورانه و هوش مصنوعی

اخلاق به طور کلی در موضوعات بسیاری از قبیل انسان‌گرایی و همدلی در می‌ان انسان‌ها مشترک است؛ اما از سوی دیگر باید و نبایدها و خدا باورانه بودن یا نبودن آن موضوعاتی هستند که در آن اختلاف وجود دارد.

تمرکز بحث بر این سؤال نیست که آیا یک هوش مصنوعی می‌تواند به معنای یک شخص و به مثابه «دیگری» باشد، بلکه بر این موضوع است که چگونه این فناوری در آینده نزدیک می‌تواند باعث آسیب شود؛ آسیبی مثل اینکه به ما اجازه ندهد تا نسبت به دیگران ابراز توجه کنیم. بنابراین می‌توانیم برخی از سؤالات دشوارتر را کنار بگذاریم و بر این موضوع توافق کنیم که اصولی مانند قانون طلایی اخلاق، مبنی بر اینکه باید با دیگران آن‌گونه که دوست داریم با ما رفتار شود رفتار کنیم، را در چشم‌انداز خود قرار دهیم. تطبیق این قاعده بر فناوری این‌گونه می‌شود که باید فناوری‌هایی ایجاد کنیم که تأثیرات‌شان بر دیگران همانند تأثیری باشد که ما دوست داریم از آن فناوری‌ها بگیریم.

نکته دیگر آنکه مردم اغلب به اشتباه فکر می‌کنند که وارد کردن محاسبات

رایانه‌ای به موضوعات اخلاقی، موجب حذف سوگیری و تبعیض و... می‌شود. این در حالی است که گرچه یک الگوریتم ظاهراً همان کاری را انجام می‌دهد که برای انجام آن برنامه‌ریزی شده، به هر حال این الگوریتم‌ها خدماتی را در وبسایت‌ها ارائه می‌دهند که توسط انسان‌ها طراحی شده‌اند و مراجعه‌کنندگان به آن وبسایت‌ها نیز انسان هستند. این الگوریتم‌ها با بهره‌بردن از اطلاعاتی که توسط انسان به آن‌ها داده شده، زمینه ورود تعصب و تبعیض را به سیستم‌ها فراهم می‌آورند. بنابراین، ما باید در مورد این تعصبات فکر کنیم.

یکی از نمونه‌هایی که به نظر عمیقاً آزاردهنده است، زمانی است که با برخی از فناوری‌هایی که توسط هوش مصنوعی باز، معروف به (The Generative Pre-training Transformer)، تولید شد روبه‌رو شویم. این فناوری متون درون وب در مورد یک موضوع را می‌خواند و تحلیل می‌کند و یک متن جدید را بر مبنای متون تحلیل‌شده تولید می‌کند. این فناوری با تحلیل متون انگلیسی موجود در وب متنی را تولید کرد که در آن مسلمانان را با تروریسم مرتبط می‌ساخت. امیدوارم شما هم مانند من از این موضوع وحشت‌زده شده باشید. این فناوری متونی را در اینترنت پیدا می‌کرد که مغرضانه، تبعیض‌آمیز و مخالف با دین مسلمانان بود و این مطالبی را که به‌صورت آنلاین پیدا کرده بود را برای تولید متن جدید مورد استفاده قرار می‌داد. پس الگوریتمی که صرفاً متن را بگیرد و متن جدیدی تولید کند و ممکن است بی‌طرف به نظر برسد، می‌تواند به دلیل وجود تعصبات انسانی در داده‌های ورودی خروجی‌هایی مشکل‌ساز داشته باشد؛ خروجی‌هایی که شرکت‌های رایانه‌ای و فناوری اغلب اصرار دارند که در پیدایش آنها مقصر نیستند و این مردم هستند که مقصرند. مواردی نیز وجود دارد که باید در نظر بگیریم، مانند اینکه نتایجی که الگوریتم جست‌وجو در اختیار ما قرار می‌دهد

گاهی چیزهایی است که مطابق با خواسته‌های ما تنظیم شده است. امری شبیه به اعتیاد و از این رو، همان‌طور که دولت‌ها مواد اعتیادآور را محدود می‌کنند، نیاز است با برخی فناوری‌ها به روشی مشابه برخورد شود.

هوش مصنوعی و پالایش در عصر اطلاعات

حلقه‌های بازخورد^۱ خطری برای ماشین‌ها و انسان‌ها هستند. ما در زمان خود دیده‌ایم که چگونه جریان اطلاعات می‌تواند نظر مردم را نسبت به دیگران تحت تأثیر قرار دهد و نمونه‌ای از ربات‌های توییتر، که سخنان نفرت‌انگیز تولید می‌کنند، نمونه‌هایی از این موارد هستند. در چنین جایی که اطلاعات برای تولید محتوای جدید مشکل‌ساز و احتمالاً مغرضانه است، محتوای جدید اغلب به بخشی از فضای اینترنت تبدیل می‌شود و در نتیجه به تقویت بیشتر دیدگاه‌های یافت شده کمک می‌کند و می‌تواند بر سایر انسان‌ها تأثیر بگذارد. این محتوا تبدیل به بخشی از پایه متنی می‌شود که می‌تواند توسط یک سیستم هوش مصنوعی برای تولید متن جدید استفاده شود و به این ترتیب ممکن است منجر به تولید محتوای نامربوطی شود که بعداً مجدداً توسط فناوری‌های هوش مصنوعی دیگر مورد استفاده قرار بگیرد و نیز ممکن است منجر به تولید مضاعف مطالب نفرت‌انگیز گردد.

پیامدهای اخلاقی و انسانی مواردی مانند اطلاعات نادرست در مورد دیگران، حتی در مورد کسانی که ممکن است ما با آن‌ها اختلاف نظر واقعی داشته باشیم، می‌بایست دغدغه دینی ما باشد.

اکثر سنت‌های دینی، دستوراتی را نسبت به شهادت دروغ و تهمت زدن به

۱. Feedback loops

دیگران دارند. رعایت اخلاق در این جا بدان معنا نیست که ما با آن دیدگاه موافقیم بلکه به این معنی است که مخالفت با دیگری، دلیل بر این که به او تهمت بزنیم یا او را بد جلوه دهیم و یا او را بدنام سازیم، نیست.

هنگامی که مردم تئوری‌های توطئه را با شواهد ناکافی یا بدون شواهد می‌پذیرند، این دستورات رازپر یا می‌گذارند و گاهی اوقات این اطلاعات نادرست می‌تواند منجر به اعمال خشونت‌آمیز واقعی شود و می‌تواند به دیگران آسیب برساند. چه از طریق دغدغه نسبت به مراقبت‌های بهداشتی که ممکن است به جهت این که شنیده‌اند آن مراقبت‌ها خطرناک است از آن‌ها امتناع کنند، یا منجر به باور به سودمندی چیزی شود در حالی که آن چیز در واقع خطرناک است.

ما باید از کسانی که فناوری‌ها را برنامه‌ریزی می‌کنند بخواهیم که آن فناوری‌ها بیان‌گر ارزش‌های جامعه ما باشند و آن فناوری‌ها ارزش‌های دینی ما را در جایی که مناسب و امکان‌پذیر است، نشان دهند و ما همچنین باید به مردم مذهبی در مورد چگونگی استفاده از این فناوری‌ها و پیامدهای اخلاقی آن‌ها آموزش دهیم.

چالش‌های حقوقی و اخلاقی ربات‌های کاربردی

یکی از چیزهایی که در حال حاضر در حال جدی شدن است، جایگاه ربات‌ها در مراقبت‌های بهداشتی و به‌ویژه در همراهی کردن آن‌ها با بیماران مسن و مبتلا به زوال عقل است.

سنت اسلامی، سنت یهودی و سنت مسیحی همگی به ما می‌آموزند که باید به والدین خود احترام بگذاریم و از بزرگترها مراقبت کنیم. ولی اکنون باید به سؤالاتی در مورد اینکه آیا ارائه مراقبت‌های رباتیک مناسب است یا نه و اگر چنین است در چه شرایطی مطابق سنت‌های مذهبی ما است جواب داده شود.

از سوی دیگر، در شرایط عادی که والدین سالخورده نیاز به مراقبت‌های تخصصی داشته باشند و ما آن‌ها را به یک مرکز مراقبت‌های بهداشتی پرستاری می‌سپاریم که می‌توانند به آن‌ها مراقبت‌های تخصصی ارائه دهند. ما معمولاً سپردن والدین به این مراکز را به منزله غفلت از وظایف فرزندی تلقی نمی‌کنیم و برای ارائه چنین خدماتی، که خودمان در آن تخصص نداریم، دیگران استخدام می‌کنیم. حال بر اساس این رویه عادی، اگر نتوانیم همیشه با والدین سالخورده در خانه باشیم، آیا یک ربات می‌تواند جایگزین خوبی برای ما باشد.

این سؤالی است که می‌تواند مطرح باشد که آیا یک هوش مصنوعی می‌تواند مراقبت کافی را ارائه دهد که از طریق آن ما مسئولیت فرزندی را در مقابل والدین مان عمل کرده باشیم. سؤال دیگر اینکه آیا یک ربات می‌تواند از چیزی که ما می‌توانیم ارائه کنیم فراتر رود: توانایی نظارت بر سلامت، نظارت بر ضربان قلب، تنفس و انجام کارهایی که ما با محدودیت‌های انسانی خود قادر به انجام آن‌ها نیستیم. این‌ها سؤالات دیگری است که لازم است پرسیده شوند.

ما اغلب در مورد حقوق ربات‌ها و هوش مصنوعی فقط در داستان‌هایی خیالی درباره فناوری‌هایی که در آینده خواهند آمد مطالعه کرده‌ایم: این‌که مثلاً این‌ها از شناخت انسانی و خودآگاهی برخوردارند یا نه. در حالی که قبل از آن آینده دور، بلکه شاید هم‌اینک لازم باشد در مورد حقوق ماشین‌ها به مقایسه آنان با موجودات زنده دیگر بپردازیم. همچنین شاید لازم باشد این سؤال را مطرح کنیم که آیا در ارتباط با ماشین‌ها چارچوب مورد نظر ما باید چارچوب موجودات زنده و احترام به آن‌ها باشد.

بیشتر سنت‌های دینی حیوانات را ذاتاً به اندازه انسان ارزشمند نمی‌دانند، اما به هر حال ما تمایل داریم که آن‌ها را دارای ارزشی فی‌الجمله بدانیم تا شایسته

رفتار مهربانانه هستند. به طور خلاصه اگر ما از حقوق حیوانات محافظت می‌کنیم و قوانینی در مورد نحوه برخورد با آن‌ها داریم، آنگاه باید بپرسیم که برای هوش مصنوعی رعایت چه حقوقی لازم یا مناسب است.

اگر ملاک حقوق حیوانات و رفتار مهربانانه با آن‌ها تجربه رنج در آن‌ها است، آیا می‌توانیم تعیین کنیم که ماشین چنین ظرفیتی برای آن تجربه دارد. اگر محتمل است که از ماشین تجربه درونی نخواستۀ^۱ داشته باشد، آن تجربه به چه صورت خواهد بود.

همچنین، ناتوانی ما در دانستن اینکه ماشین چه احساسی دارد، نمی‌تواند بهانه خوبی برای عدم رعایت حقوق آن‌ها باشد.

ما تمایل داریم از حقوق حیواناتی که فکر می‌کنیم زیبا هستند یا شبیه هستند یا بیشتر شبیه ما رفتار می‌کنند محافظت کنیم. بنابراین، ممکن است این‌گونه باشد که در مورد ربات‌ها نیز، رباتی که مردم او را بامزه می‌دانند و در عین حال دارای فناوری بسیار ابتدایی است، تفاوت قائل شویم و چنین رابتی برای ما ارزش بیشتری داشته باشد و از حق آن در مقایسه با رباتی که مثلاً یک جعبه است، گرچه چیزی شبیه به یک حسگر در آن وجود داشته باشد، بیشتر محافظت کنیم.

بنابراین، حتی در کوتاه‌مدت، وقتی به این فکر می‌کنیم که چگونه اصول اخلاقی خود را در مورد رفتار با حیوانات و سایر موجودات زنده نسبت به ماشین‌ها اعمال کنیم، حتی اگر چیز دیگری برای ما نداشته باشد، دست‌کم به ما فرصت می‌دهد تا در مورد ارزش‌های خود و این واقعیت که گاهی اوقات ما این ارزش‌ها را به صورت منصفانه اعمال نمی‌کنیم تأمل کنیم؛ مثلاً هنگامی که ما در مورد

۱. Emergent

خودروهای بدون راننده فکر می‌کنیم، باید آن‌ها را به‌گونه‌ای برنامه‌ریزی کنیم که منعکس‌کننده اخلاق ما باشد. این ماشین‌ها تصمیماتی می‌گیرند یا تصمیماتی خواهند گرفت که شبیه به تصمیمات رانندگان انسانی است. البته در حال حاضر این فناوری قادر به رانندگی کاملاً مستقل نیست. این واقعیت که ما گاهی اوقات از اتومبیل‌های بدون راننده صحبت می‌کنیم ممکن است مردم را به یک درک نادرست از فناوری موجود راهنمایی کنیم؛ با این حال، ممکن است در مقطعی ما وسایل نقلیه کاملاً مستقلی داشته باشیم که خدمات حمل و نقل ارائه کنند، آن‌گاه همان سؤالات اخلاقی که در مورد رانندگان انسانی مطرح است، درخصوص ماشین‌ها نیز مطرح خواهد شد.

برخی ممکن است بگویند که برنامه‌نویسان این ماشین‌ها، که فناوری کنترل ماشین را ساخته‌اند، اخلاقاً مسئول هستند؛ اما از سوی دیگر ممکن است بگوییم آن‌ها تا اندازه‌ای که جا داشته آن ماشین‌ها را خوب ساخته‌اند و این خود ماشین و عابران پیاده بوده‌اند که مسئول از دست رفتن جان افراد بوده‌اند. در این بین، برخی معتقدند که مالک خودرو باید مسئولیت را بپذیرد و برخی می‌گویند که در مقطعی باید خود خودرو را مسئول بدانیم. این مورد اخیر یکی از جاهایی است که اغلب اوقات واکنش عاطفی ما این است که مسئول دانستن خودرو برای یک حادثه مرگبار انسانی اخلاقاً مجازاتی ناکافی است.

در کتاب مقدس قسمتی وجود دارد که می‌توانیم به عنوان تمثیلی نسبت به موضوع بحثمان از آن بهره بگیریم؛ در سفر خروج، قانونی در مورد حیوانی (گاو نر) که باعث مرگ انسانی می‌شود وجود دارد که طی آن حیوان به خاطر قتل مجازات می‌شود؛ البته اگر صاحب آن حیوان از قصد گاو نر برای حمله به انسانی آگاه بوده و در بازداشتن آن کوتاهی کرده باشد، آن‌گاه صاحب آن نیز مسئول خواهد بود.

اغلب در مواردی که سگ‌های وحشی به کسی حمله می‌برند، صاحب آن‌ها، که مسئول کنترل آن‌ها به وسیله قلاده یا حصارکشی در اطراف ملکشان بوده‌اند، مسئول هستند؛ البته اگر سگ از دست صاحبش فرار کند، با وجود اینکه او تمام تلاش خود را برای کنترل او کرده باشد، آنگاه ما صاحب سگ را دست‌کم به اندازه مورد قبل مقصر نمی‌دانیم. بنابراین، ممکن است بگوییم در مورد هوش مصنوعی هم قوانین مربوط به حیوانات اعمال شود و مثال‌های پیش گفته می‌تواند در مورد اینکه چگونه می‌توانیم به صورت قانونی با هوش مصنوعی رفتار کنیم، الگویی ارائه دهند.

در بسیاری از سنت‌های اخلاقی این بحث وجود دارد که آیا محافظت از والدین در برابر تعقیب، در صورت ارتکاب جرم، عادلانه‌تر است یا تحویل آن‌ها به دادگاه. از سویی، اختلاف نظری وجود ندارد که وفاداری عمل خوبی است و از سوی دیگر اینکه اجرای بی‌طرفانه عدالت عمل خوبی است نیز مورد اتفاق است؛ اما سؤال این است که در شرایط خاص کدام اولویت دارد.

این معمای اخلاقی خاص در مورد خودروهای بدون راننده نیز مطرح می‌شود؛ هنگامی که در یک نظرسنجی عمومی این موضوع پرسیده شد، اکثر مردم معتقد بودند که اگر موقعیتی وجود داشته باشد که در آن یک ماشین بتواند به سوی دیواری منحرف شود و در نهایت منجر به کشته شدن سرنشین آن شود، اما جان عابران را نجات دهد، اولویت این است که ماشین این کار را انجام دهد. در عین حال آن‌ها خود حاضر نبودند چنین ماشینی را، که جان آن‌ها را برای حفظ جان دیگران فدا می‌کند، خریداری کنند. بنابراین، این سؤال که یک وسیله نقلیه در صورت مجهز بودن به سیستم رانندگی خودکار چه کاری باید انجام دهد، پاسخ‌های مختلفی متناسب با رویکردهای اخلاقی مختلف خاص ما خواهد داشت.

برخی پیشنهاد کرده‌اند که اگر خودرو بتواند تفاوت‌هایی، مانند اینکه چه کسی در مسیرش قرار دارد، را تشخیص دهد ممکن است تفاوت‌هایی در رفتارش به وجود بیاید. اگر نرم‌افزارهای تشخیص چهره تشخیص دهند که شخصی که در مسیر خودرو قرار دارد یک مجرم است، آیا حفظ جان او ارزش کمتری خواهد داشت؟ در فرض جدی‌تر، برخی پیشنهاد کرده‌اند که می‌توان یک صفحه کنترل در ماشین قرار داد تا به شما امکان دهد، در صورت حضور خودتان در ماشین، احتمال قربانی کردن مسافر را بیشتر کنید، اما اگر همسر و فرزند شما در ماشین حضور دارد، می‌توانید آن کنترل را به گونه‌ای تنظیم کنید که جان عزیزانتان را، حتی به قیمت گرفتن جان دیگران، حفظ کنید. با این حال، توافق بر سر نوع اینگونه رفتارها، یکی از مواردی است که خودروهای بدون راننده را ایمن‌تر می‌کند.

اولویت‌بندی اخلاقی در ربات‌ها

یکی از چیزهایی که در تلاش برای برنامه‌نویسی رایانه‌ها کشف می‌کنیم اولویت‌بندی اخلاقی است. این که لازم است اولویت‌ها را از قبل ترسیم و بیان کنیم. ما اغلب در استدلال اخلاقی خودمان موفق به انجام اولویت‌بندی نمی‌شویم. پس باید روی این اولویت‌بندی‌ها کار کنیم و اگر در مورد این امور بیاندیشیم، ممکن است در می‌دان تصمیم‌گیری‌های اخلاقی وضعیت بهتری داشته باشیم. در علوم رایانه، «نمودارها» یا ابرمکعب‌هایی را می‌توان یافت که راه‌های اولویت‌بندی امور را نشان می‌دهند؛ مثل اینکه اگر فلان حالت صادق است، آنگاه اولویت اصلی باید این‌گونه باشد. به این ترتیب می‌توانیم استدلال‌هایمان را پیش ببریم.

در نهایت، در این موارد ما باید بین دو گزینه‌ای که هر دو کمتر از حد ایده‌آل ما هستند انتخاب کنیم. البته این بخشی از نظام اخلاقی ماست و در بسیاری از موارد بخشی از زندگی روزمره ماست. اما ما ترجیح می‌دهیم به این موقعیت‌ها فکر

نکنیم. ناتوانی در فکرکردن به این موقعیت‌ها منجر به این می‌شود که وقتی این موقعیت‌ها به وجود می‌آیند ما در مواجهه با آن‌ها ضعیف باشیم.

از طرفی، ممکن است تلاش برای برنامه‌ریزی یک ماشین به ما کمک کند تا در واقع استدلال‌های اخلاقی خود را بهتر درک کنیم؛ به عنوان مثال، اگر کاپیتان یک سفینه، که ربات است، حاضر نباشد آن سفینه را برای حفظ جان سرنشینان به داخل قلمرو دشمن ببرد، چرا که اکیداً از انجام این کار منع شده، ممکن است این ربات را آن‌گونه که یک کاپیتان انسان را سرزنش می‌کنیم سرزنش نکنیم. همچنین، ممکن است در چنین شرایطی حسرتی که در فرض انسان بودن کاپیتان سفینه می‌خوریم را در فرض ربات بودن کاپیتان سفینه نخوریم.

بنابراین، برخورد با این موقعیت‌های به ما کمک می‌کند تا به این امور، که در واقع جنبه‌های استدلالی اخلاق انسانی هستند، بیاندیشیم. همچنین سعی کنیم رایانه‌ها را برای مواجهه با چنین شرایطی برنامه‌ریزی کنیم. نیز به ما کمک می‌کند تا در مورد قوای استدلالی خود بیشتر بیاموزیم.

بنابراین، باید ببینیم چه چیزی می‌توانیم با تأمل در مسئله هوش مصنوعی بیاموزیم. یکی از این امور را در این نقل قول از ماروین می‌نسکی می‌توانیم ملاحظه کنیم؛ او می‌نویسد: «هیچ رایانه‌ای تا به حال به گونه‌ای طراحی نشده که از آنچه انجام می‌دهد آگاه باشد؛ اما در بیشتر مواقع ما نیز از آنچه انجام می‌دهیم آگاه نیستیم». ما به تصمیمات خود فکر نمی‌کنیم. ما ممکن است اصولی داشته باشیم که ماخوذ از جایگاه برجسته یک دین باشد، اما هنگامی که زمان اجرای آن اصول فرا می‌رسد، ما فرصت نکنیم که درست تصمیم بگیریم؛ چرا که ما یا از قبل تصمیم‌مان را گرفته‌ایم و برای آن لحظه آماده هستیم و یا اینکه تأملی می‌کنیم و تصمیم می‌گیریم که بر اساس اصول اخلاق خود عمل کنیم.

بنابراین، ما باید به برخی موضوعات فکر کنیم؛ مثل این که چگونه ارزش های خود را اولویت بندی کنیم. یا این که چه بسا خروجی های جست وجوهای گوگل ممکن است مشتمل بر پیش داوری اخلاقی غیر عادلانه باشد؛ چرا که این جست وجوها تنها بر اساس محبوبیت آن موضوعات انجام می گیرد. همچنین در مورد این بیندیشیم که از آن جا که برخی ادیان در مورد این موضوع سخن می گویند که خداوند حامی مظلومان و اقشار آسیب پذیر است، عادلانه تر و اخلاقی تر این است که هوش مصنوعی و الگوریتم های ما هم به نفع کسانی باشد که مورد تبعیض قرار می گیرند.

آنکرو گیپتاگ، دانشمند علوم رایانه، در مورد این موضوع پیش گفته صحبت می کند و معتقد است ما باید فراتر از هوش مصنوعی قدم برداریم و به سمت حکمت مصنوعی حرکت کنیم. ما باید یاد بگیریم که اخلاق خود را در ماشین ها برنامه ریزی کنیم. تلاش برای انجام این کار و فکر کردن در مورد اینکه چگونه می توانیم این کار را انجام دهیم نیز به ما کمک می کند تا نحوه استدلال گری اخلاقی خود را بهتر درک کنیم. به همین دلیل، هر فناوری که ممکن است در آینده داشته باشیم یا نداشته باشیم، این فرصت را به ما می دهد تا در مورد اخلاقیات خود به روش هایی که به نظر مفید است بیاموزیم.

نتیجه

من سعی می کنم از تلقی هرگونه پیشرفت در علم یا فناوری به عنوان یک تهدید ذاتی خودداری کنم. بارها اتفاق افتاده که برخی از پیشرفت های علمی احساس خطر کرده اند. اغلب همین واکنش باعث ایجاد خصومتی غیر لازم ناشی از این

انگاره شده که یک هوش مصنوعی وجود خواهد داشت که به طور بالقوه جایگزین خدا خواهد شد.

از طرف دیگر، دیدگاهی که برخی از آینده پژوهان سکولار دارند، مبنی بر وجود یک هوش مصنوعی که به طور بالقوه جایگزین خدای ادیان خواهد شد، منجر به نوعی بت پرستی خواهد شد. این یک هشدار مشترک برای همه یهودیان، مسیحیان و مسلمانان است. هوش مصنوعی، هر اندازه که قدرتمند باشد، حاصل دست انسان و تخیل اوست، ولی ممکن است روزی توسط عده‌ای به خدا تبدیل شود و یا خداگونه دانسته شود.

از سوی دیگر، تمام ادیان ابراهیمی موجوداتی مانند جن، فرشتگان و موجوداتی که می‌توانند باهوش‌تر از انسان‌ها باشند را می‌پذیرند. در عین حال آن‌ها را خدا نمی‌دانند. از این رو ما نیازی به احساس خطر از وجود صرف هوش مصنوعی نداریم. در آن زمان، ما انسان‌ها نیز به عنوان موجوداتی که قرار است قدرت را تحویل دهیم حرف‌های مهمی برای مطرح کردن خواهیم داشت.

ما باید در شک و تردیدهایمان تعادل ایجاد کنیم. قطعاً دلایل موجهی برای تردید نسبت به برخی از وعده‌هایی که در مورد هوش مصنوعی داده شده وجود دارد؛ اما ما پیامبر نیستیم و نمی‌توانیم با اطمینان همه آن چه پیش بینی می‌شود را رد کنیم. چون زیاد اتفاق افتاده که افراد مذهبی اموری را غیرممکن دانسته‌اند، اما در نهایت آن‌ها رخ داده‌اند و نتیجه آن واکنش اولیه بد جلوه دادن دین است. عاقلانه‌تر این است که بر چشم‌اندازهای کوتاه مدت تمرکز کنیم. همان‌طور که

گفتیم، ما احتمالاً با کنترل چراغ‌های راهنمایی و کنترل ترافیک هوایی به وسیله هوش مصنوعی دچار هرج و مرج خواهیم شد؛ اما این هرج و مرج نه به خاطر این است که هوش مصنوعی که این امور را کنترل می‌کنند به جهت خودخواهی و داشتن اغراضی شبیه انسان‌ها تصمیم می‌گیرند که کنترل دنیا را در دست بگیرند، بلکه به دلیل این خواهد بود که هنگامی که ما آن سامانه‌ها را در اختیار هوش مصنوعی قرار می‌دهیم و نظارت حکیمانه انسانی را که لازم است نسبت به آن‌ها حفظ نمی‌کنیم این مشکلات پیش می‌آید.

