

درآمدی بر حکمرانی هوش مصنوعی

خلاصه راهبردی از:

Allan Dafoe, AI Governance: A research
agenda , Oxford university, 2018

معاونت پژوهش‌های سیاسی - حقوقی
گروه مطالعات بنیادین حکومتی

کد موضوعی: ۳۳۰
شماره مسلسل: ۱۶۶۹۲
آبان‌ماه ۱۳۹۸

به نام خدا

فهرست مطالب

۱ خلاصه
۲ مقدمه
۴ فصل اول - سیاست هوش مصنوعی
۱۴ فصل دوم - حکمرانی مطلوب هوش مصنوعی
۱۸ نتیجه گیری
۲۰ منابع و مآخذ



درآمدی بر حکمرانی هوش مصنوعی

خلاصه راهبردی از:

**Allan Dafoe, AI Governance: A research agenda , Oxford
university, 2018**

خلاصه

پیشرفت‌های آینده بسیار سریع بوده و دانشمندان انتظار دارند که شاهد تحقق قابلیت‌های فراانسانی در حوزه‌های استراتژیک در چند دهه آینده باشند. فرصت‌های ایجاد شده توسط پیشرفت‌های تکنولوژیک بالاخص در حوزه هوش مصنوعی در حوزه‌های مختلف همانند پزشکی و سلامت، حمل‌ونقل، انرژی، آموزش، رشد اقتصادی و پایداری محیطی بسیار عظیم است. در مقابل مخاطرات ممکن نیز کم نبوده و پیشرفت‌های تکنولوژیکی ما را با چالش‌های فراوانی در حوزه حکمرانی مواجه می‌کند. از جمله این مخاطرات می‌توان به جابه‌جایی نیروی کار، نابرابری، دگرگونی در قدرت ملی، بی‌ثباتی استراتژیک و نوعی از هوش مصنوعی که می‌تواند سلامت و دیگر ارزش‌ها را قربانی خود نماید، اشاره کرد. رهبران حکومت‌های مختلف و همچنین کمپانی‌ها و شرکت‌های بزرگ در این وضعیت جدید غالباً تقاضای مشاوره برای سیاستگذاری بهتر را می‌کنند، ولی هنوز هم

توجه علمی به انقلابی که هوش مصنوعی ایجاد کرده است بسیار اندک است. به همین دلیل نیاز فوری به انجام پژوهش‌های متعدد درباره مسئله حکمرانی هوش مصنوعی^۱ وجود دارد. مسئله‌ای که خود را در بازاندیشی نهادها، سیاست‌ها و هنجارهای جهانی به‌منظور تضمین گسترش و کاربرد بهتر و سودمندانه‌تر هوش مصنوعی می‌یابد.

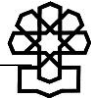
گزارش پیش‌رو با توجه به ضرورت انجام پژوهش در این باره نوشته شده است: در بخش نخست سیاست هوش مصنوعی، بر پویایی میان دولت‌ها، کمپانی‌های بزرگ، عامه مردم، محققان و دیگر کنشگران تمرکز کرده است. در بخش آخر نیز ساختارهای ممکن برای گذار به مدیریت هوش مصنوعی تحلیل شده است.^۲

مقدمه

پیش‌بینی می‌شود که هوش مصنوعی در بیشتر وظایفی که برای آن تعریف شده است به نیروی فرانسایی در این قرن بدل شود. چنین پیشرفتی می‌تواند فرصت‌ها و مخاطرات زیادی را برای بشریت در پی داشته باشد. به‌نظر متخصصان مختلف در چند دهه آینده، قابلیت‌های مختلف هوش مصنوعی خود را نشان داده و بدین‌ترتیب به شکل رادیکالی قدرت، ثروت و رفاه موجود را دگرگون می‌کند. دگرگونی‌ای که از نظر دامنه تغییرات با انقلاب صنعتی قابل مقایسه است. حکمرانی هوش مصنوعی در واقع مطالعه چگونگی مدیریت گذار به سیستم‌های هوش مصنوعی پیشرفته با تمرکز بر ابعاد گوناگون سیاسی،

1. AI Governance

۲. گزارش حاضر نگاه اثباتی و ابطالی به موضوع ندارد و فقط برای آشنایی نمایندگان مجلس ارائه شده و منابع گزارش نیز به طرح پژوهشی حکمرانی هوش مصنوعی مربوط می‌شود.



اقتصادی، نظامی، حکومتی و اخلاقی است. در واقع گزارش پیش‌رو درآمدی بر این مبحث می‌باشد.

حکمرانی هوش مصنوعی اغلب در کنار سلامت و ایمنی هوش مصنوعی قرار داده می‌شود، البته هدف هر دوی آنها کمک به بشریت برای استفاده مفید هر چه بیشتر از هوش مصنوعی است، اما با این تفاوت که حکمرانی هوش مصنوعی بر نهادها و زمینه‌هایی که هوش مصنوعی ساخته شده و استفاده می‌شود، متمرکز است.

مسئله حکمرانی هوش مصنوعی ما را با پرسش‌های جدی مواجه می‌کند: با وجود این برای به حداکثر رساندن ایمنی گذار به جهانی جدید که پیشرفت‌های رخ داده در هوش مصنوعی سهم فراوانی در شکل دادن به آن دارند چه باید کرد؟ چه اندرزهایی به آزمایشگاه‌های هوش مصنوعی، حکومت‌ها، NGOها و عامه مردم چه هم‌اکنون و چه در آینده می‌توانیم ارائه دهیم؟ چه مقررات بین‌المللی در حوزه پروتکلی، سازمانی، تکنولوژیک و... برای ممانعت از درافتادن کشورها و شرکت‌های بزرگ در دام منافع کوتاه‌بینانه لازم داریم؟

هوش مصنوعی پیشرفته می‌تواند نقش مهمی در حل مشکلات جهانی موجود از تغییرات آب‌وهوایی گرفته تا تنش‌های جهانی ایفا کند. هوش مصنوعی پیشرفته به‌طور شگفت‌انگیزی می‌تواند به ارتقای سلامت، شادی، رفاه، علم و خودفهمی کمک کند.¹ این مزایا باید در کنار ریسک‌های فاجعه‌بار احتمالی هوش مصنوعی قرار داده شود همانند تقویت تمامیت‌خواهی، جنگ هسته‌ای قدرت‌های بزرگ (که هوش مصنوعی می‌تواند چرخه بحران چنین جنگی را پیچیده‌تر و پیش‌بینی‌ناپذیرتر نماید) و ناسازگاری با ارزش‌های انسانی.

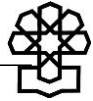
1. Nick Bostrom, Public Policy and Super intelligent AI", Future of Humanity Institute, 2018.

فصل اول - سیاست هوش مصنوعی^۱

هوش مصنوعی ماهیت ثروت و قدرت را دگرگون می‌کند. علائق و قابلیت‌های بازیگران قدرتمند عوض شده و بازیگران قدرتمند جدیدی ممکن است ظهور کنند. این بازیگران برای پیشبرد علائق و منافع‌شان به رقابت یا همکاری با همدیگر دست می‌زنند. احتمالاً مخاطرات جدی‌ای بر هوش مصنوعی مترتب باشد. ملاحظات جدی‌ای که در توسعه هوش مصنوعی دگرگون‌ساز وجود دارد به این ترتیب است: افکار عمومی انعطاف‌پذیر، ارزش‌ها و اخلاق لازم برای پژوهشگران، ثروت ملی، امنیت ملی و... . سیاست هوش مصنوعی به این نکته توجه می‌کند که چگونه چشم‌انداز فنی در حال تغییر می‌تواند سیاست داخلی، اقتصاد سیاسی بین‌المللی و امنیت بین‌المللی را دگرگون کرده و چگونه سیاست‌های بازیگران قدرتمند می‌تواند به توسعه هوش مصنوعی شکل ببخشد.

- سیاست داخلی

هوش مصنوعی این پتانسیل را دارد که هم بر سیاست داخلی تأثیر بگذارد و هم از آن تأثیر بپذیرد. به هر میزان که هوش مصنوعی و تکنولوژی‌های مربوطه توزیع قدرت در داخل کشورها را دگرگون کنند، اشکال حکومت نیز تغییر می‌کند. چنین چیزی به معنای دگرگونی در قدرت به طرف بازیگرانی است که می‌توانند سیستم‌های هوش مصنوعی را به کار بگیرند همانند نخبگان، دولت‌ها و اصناف گوناگون.



- اشکال حکومت

ساختارهای سیاسی داخلی برخواسته از کارکرد فاکتورهای پیچیده‌ای است که بعضی از آنها می‌توانند توسط هوش مصنوعی دگرگون شوند. بعضی از این فاکتورها که به نظر می‌رسد برای تعیین ویژگی خاص یک حکومت مهم است بدین ترتیب می‌باشد:

۱. توزیع نابرابر کنترل بر روی سرمایه‌های اقتصادی و ثروت،

۲. تکنولوژی‌های نظارتی،

۳. تکنولوژی‌های اقناعی،

۴. تکنولوژی‌های مشاوره شخصی،

۵. تکنولوژی‌های عمل جمعی.

هر نوع پژوهش بر روی اشکال حکومت به معنای واکاوی روندهای تحت تأثیر هوش

مصنوعی و دیگر فاکتورهاست. اهمیت چنین موضوعی به دلایل ذیل می‌باشد:

۱. روندهای موجود در حکمرانی داخلی بر نوع نظام سیاسی در بلندمدت تأثیر

می‌گذارد.

۲. ویژگی‌های بازیگران کلیدی در استراتژی هوش مصنوعی می‌تواند همانند ویژگی

دولت‌های چین و آمریکا تأثیرگذار باشد.

۳. می‌تواند خاصیت آگاهی‌بخش برای نهادهای جهانی داشته باشد.

- مشاوران و عمل جمعی

هوش مصنوعی می‌تواند به توانمندسازی شهروندان کمک کند. هوش مصنوعی‌هایی که

نقش مشاور شخصی را ایفا می‌کنند به شهروندان این امکان را می‌دهند که به شکلی

آگاهانه‌تر در فرآیندهای سیاسی مشارکت کنند. سیستم‌های هوش مصنوعی می‌توانند عمل جمعی را تسهیل کرده و به ایجاد ائتلاف‌های سیاسی کمک نمایند. با کمک گرفتن از هوش مصنوعی افراد می‌توانند به شکلی **پیچیده‌تر** استراتژی‌های سیاسی پیچیده را بیان کرده و به شیوه‌ای بهتر با هم همکاری کنند.

ترجمه بهتر توسط هوش مصنوعی می‌تواند به تسهیل ارتباطات بین‌فرهنگی کمک کند. فیلترهای هوش مصنوعی به کاربران کمک کرده تا از اخبار نامعتبر دوری نمایند. همچنین با کمک هوش مصنوعی می‌توان دروغ‌های موجود در فایل‌های صوتی و تصویری را شناسایی کرد و بدین ترتیب راستی را از دروغ تمییز داد.

- نابرابری، جابه‌جایی شغلی و بازتوزیع

به نظر می‌رسد که هوش مصنوعی به افزایش نابرابری میان مردم و احتمالاً کشورها می‌انجامد. هوش مصنوعی باعث جابه‌جایی در شغل‌های مربوط به طبقه متوسط شده و حداقل روند کوتاه‌مدت بر از بین رفتن بعضی از مشاغل با درآمد پایین حکایت دارد. سهم کار در درآمدهای ملی رو به کاهش است و هوش مصنوعی احتمالاً این روند را تقویت می‌کند.

به نظر می‌رسد هوش مصنوعی باعث نوعی انحصارطلبی جهانی می‌شود و عملاً تنها یک موتور جستجوی خوب (گوگل)، یک شبکه خدمات اجتماعی (فیس‌بوک) و یک بازار آنلاین (آمازون) وجود دارد. احتمالاً این انحصارطلبی‌های ناشی از هوش مصنوعی و نابرابری‌های هم‌پیوند با آن از خواسته‌های اصلی تقاضاهای بازتوزیعی در آینده خواهد بود. به همین منظور هر نوع پژوهش در این زمینه باید به محاسبه و فهم تأثیرات هوش



مصنوعی بر نابرابری و جابه‌جایی‌های شغلی بپردازد. پرسش اصلی در این باره چنین است: پیامدهای تأثیرگذاری هوش مصنوعی بر بازار اشتغال و نابرابری، بر روی تقاضاهای عامه مردم و نیز مشروعیت مدل‌های حکمرانی متفاوت چیست؟ راه‌حل‌های بالقوه برای مدیریت این وضعیت در شیوه‌های مختلف حکمرانی چه خواهد بود؟

- اقتصاد سیاسی بین‌المللی

اگر روند موجود که در آن تمرکز بر منافع حاصل از هوش مصنوعی عمدتاً در اختیار کمپانی‌های بزرگ (همانند گوگل) است ادامه یابد احتمالاً شاهد نوعی از انحصارگرایی در به کارگیری و استفاده در حوزه هوش مصنوعی در سطح جهانی خواهیم بود.

کشورهایی که فاقد صنایع هوش مصنوعی هستند نگرانند که از چرخه منافع جهانی حاصل از هوش مصنوعی به حاشیه رانده شوند. کسانی در آمریکا و اروپا نگران آنند که چین قدرت اقتصادی خود را به طوری افزایش دهد که به شکلی استراتژیک جلوتر از شرکت‌های غربی از نظر دانش تکنولوژیک بیفتد. چنین نگرانی احتمالاً انگیزه اصلی ترامپ برای جنگ تجاری با چین است.^۱ این نگرانی‌ها ممکن است به نوعی ملی‌گرایی هوش مصنوعی و حمایت‌گرایی هوش مصنوعی منجر شود.^۲ به طوری که کشورها منابع زیادی را برای توسعه هوش مصنوعی صرف کرده و از قهرمانان داخلی در حوزه هوش مصنوعی به شدت حمایت کنند.

1. David Barboza, How This U.S Tech Giant is Backing China's Tech Ambitions, New York Times, August 4, 2017.

2. Ian Hogarth, AI Nationalism, Ian Hogarth (blog), June 13, 2018.

کشورها و کمپانی‌های مختلف در حال جستجو برای یافتن راه‌های مختلفند تا از لحاظ اقتصادی، به سود هرچه بیشتر از هوش مصنوعی دست یابند. آنها حتی در تلاشند که نمونه‌های موفق شرکت‌هایی که از هوش مصنوعی استفاده می‌کنند را در داخل کشور خودشان تولید کنند یا حداقل از نمونه‌های داخلی حمایت بیشتری نمایند، پرسش‌های اساسی در این باره چنین است: چه مزایا و منافعی می‌توان برای کشوری همانند چین (در مقایسه با دیگر کشورها) در توانایی‌اش برای مدیریت اقتصاد بزرگش و حذف رقابیش متصور شد؟ امکانات و دینامیک‌های احتمالی اقتصاد جهانی مبتنی بر هوش مصنوعی چیست؟ آیا حمایت کشورها از کمپانی‌های هوش مصنوعی داخلی‌شان برای رقابت بهتر با صنایع تراز اول جهانی هوش مصنوعی، قابل قبول است؟ به هر حال واضح است جابه‌جایی‌های تکنولوژیکی به‌طور متفاوتی بر کشورها تأثیر می‌گذارد و کشورهای مختلف نیز در واکنش به آن، سیاست‌های مختلفی را وضع می‌کنند. آنچه که از نظر امنیت بین‌المللی حائز اهمیت است احتمال وقوع رقابت‌های شدید میان کشورها بر سر هوش مصنوعی است. از این‌رو بررسی دقیق‌تر و جامع‌تر سناریوهای احتمالی چنین رقابت‌هایی در آینده پیامدها و راه‌های مدیریت آن می‌تواند موضوع پژوهش‌های آینده در حوزه امنیت و اقتصاد بین‌المللی باشد.

– امنیت بین‌المللی

احتمالاً هوش مصنوعی و تکنولوژی‌های هم‌پیوند با آن پیامدهای مهمی برای امنیت ملی و بین‌المللی دارند. به‌نظر می‌رسد هوش مصنوعی پیامدهای نظام استراتژیک و دگرگون‌سازی در آینده کوتاه‌مدت و میان‌مدت داشته باشد. در بلندمدت نیز هوش



مصنوعی به سرمایه‌ای اساسی برای کشورها بدل می‌شود و ممکن است که دولت‌ها بکوشند تا هر آنچه به بخش تحقیق و توسعه (R&D) هوش مصنوعی مربوط می‌شود را کنترل و امنیتی نمایند. همچنین مزایای استراتژیک و نظامی هوش مصنوعی ممکن است بر آتش مسابقات تسلیحاتی در جهان بدمد. به همین دلیل فهم منطق چین رقابتی و راه‌های اجتناب از آن و در صورت وقوع، خاتمه دادن به آن ضروری است.

- چالش‌های کوتاه‌مدت هوش مصنوعی

در سال‌های آتی هوش مصنوعی ما را با چالش‌های جدید امنیتی روبه‌رو می‌کند. این چالش‌ها شامل استفاده‌های داخلی و بین‌المللی از بمب‌های خودران، عملیات سایبری با استفاده از هوش مصنوعی و کمپین‌های سیاسی تأثیرگذار با کمک هوش مصنوعی است. بسیاری از این چالش‌ها بالقوه ماهیتی دگرگون‌ساز دارند. به همین دلیل ضروری است که از این چالش‌ها باخبر بوده و از نحوه مواجهه با آنها آگاه باشیم.

- کنترل، انسداد و امنیتی شدن

امروزه تحقیقات اساسی مربوط به هوش مصنوعی به شکلی آزاد انجام می‌شود: محققان علاقه زیادی به انتشار عمومی یافته‌هایشان برای کسب اعتبار دارند، البته بعضی از تحقیقات مربوط به هوش مصنوعی نیمه‌بسته بوده و به‌طور خصوصی و برای منافع خصوصی انجام می‌شود، البته خصوصی بودن تحقیقات مربوط به هوش مصنوعی امری رایج نیست. در صورتی که هوش مصنوعی به‌عنوان دانش خطرناک و یا از نظر اقتصادی و نظامی استراتژیک تلقی شود، احتمالاً چین روندی تغییر خواهد کرد. به هر میزان که

احتمال فاجعه‌بار بودن گسترش عمومی هوش مصنوعی برود احتمال دارد که سرآمدان و دانشمندان هوش مصنوعی بخواهند که از انتشار قابلیت‌های هوش مصنوعی برای دیگران ممانعت به عمل آورند. از این رو پرسش‌های اساسی که باید به آن اندیشید چنین است: مدل‌های متفاوت محدود نمودن تحقیقات مربوط به هوش مصنوعی چیست؟ تا چه میزان دولت‌ها باید در چنین مواردی دخالت کنند؟ مقررات قانونی احتمالی که دولت‌ها برای اجرای کنترل بر روی کمپانی‌های هوش مصنوعی باید به کار ببرند چیست؟ در چه شرایطی می‌توان به پژوهش‌های مربوط به هوش مصنوعی و گسترش این دانش نو پدیده به شکل امنیتی نگریست؟ چه گزینه‌های حقوقی برای کمپانی‌های بزرگ وجود دارد تا از دخالت دولت‌ها جلوگیری کنند؟ و آیا می‌توانیم از مطالعه تلاش‌های انجام شده برای کنترل دیگر تکنولوژی‌ها بیاموزیم؟

- رقابت بر سر هوش مصنوعی

هوش مصنوعی پیشرفته می‌تواند ثروت و قدرت فراوانی را در اختیار صاحبان آن قرار دهد. اگر چنین چیزی درست باشد و مخصوصاً احساس شود که منافع اقتصادی و نظامی استراتژیکی برداشتن هوش مصنوعی مترتب است پس باید انتظار وقوع رقابتی بین‌المللی را بر سر هوش مصنوعی داشت.

به دلیل مزایای فراوان در اختیار داشتن هوش مصنوعی پیشرفته احتمال می‌رود که بازیگران مختلف چندان به فکر ایجاد توازن میان در اختیار داشتن آن با ارزش‌هایی همچون سلامت، شفافیت، مسئولیت‌پذیری و مردم‌سالاری نباشند. مخصوصاً نگرانی اصلی اینجاست که در کاربرد سیستم‌های پیشرفته هوش مصنوعی به احتیاط‌های لازم توجه



نکرده و وضعیت فوق پرسش‌های متعددی را فراروی قانونگذاران قرار می‌دهد: چه ابزارهایی برای اجتناب از وقوع چنین رقابت مهلکی در اختیار داریم؟ چه قوانینی (داخلی و جهانی) می‌توان برای جلوگیری از چنین رقابتی وضع کرد؟ نمونه‌های تاریخی گذشته و قیاسی با موارد مشابه همانند قراردادهای حقوقی بسته شده برای ممانعت از مسابقات تسلیحاتی اتمی می‌تواند به ما در مواجهه با این وضعیت جدید کمک کند. همچنین می‌توان به تئوریزه کردن مراحل احتمالی رقابت شدید بر سر هوش مصنوعی و ویژگی‌های آن پرداخت همانند شتاب، خطر و پیامدهای امنیتی. از این رو باید به‌طور جدی تلاش کرد تا پاسخی به این پرسش‌ها یافت: آیا می‌توان وضعیت کنونی ما را در مواجهه با هوش مصنوعی پیشرفته در چنین چارچوبی ریخت؟ تا چه حد سیاستگذاران و قانونگذاران ما نسبت به منطق چنین رقابت خطرناکی آگاهند؟ چه نوع حوادثی می‌تواند جرقه نبردی سهمگین بر سر هوش مصنوعی را زده و تنش بر سر آن را افزایش دهد؟

- احتمال نبرد بر سر هوش مصنوعی

با بالا رفتن احتمال وقوع خطر رقابت‌های شدید بر سر هوش مصنوعی لازم است که راه‌های ممکن اجتناب از چنین رقابت‌هایی اندیشیده شده و یا در صورت وقوع برای خاتمه آن برنامه‌ریزی کرد. راه‌حل سیاسی معمول برای مواجهه با مخاطرات سیاسی در جهان معمولاً در افزایش شفافیت از یک طرف و وضع هنجارها، قراردادهای و یا امضای توافقنامه‌ها است که می‌تواند دوطرفه، چندطرفه و یا در سطح جهانی باشد. هنجارها دربرگیرنده فهم دو طرف درباره رفتارهای غیرقابل پذیرش و واکنش‌های تحمیل شده در پاسخ به آن است.

هنجارهای پنهان این مزیت را دارند که بدون رضایت آشکار وضع می‌شوند، البته در مقابل این ایراد را نیز می‌توان بر آنها وارد دانست که می‌توانند ظالمانه بوده و اغلب ناکافی باشند.^۱ از اشکال اصلی هنجارهای بین‌المللی، قانون‌های رایج و مرسوم‌اند که البته در فقدان دادگاه و مرجع عالی حقوقی بین‌المللی برای قدرت‌های بزرگ موضوعیت چندانی ندارد.

توافقات دیپلماتیک و قراردادهای در صورتی مؤثرند که به‌خوبی در سطح جزئیات تعریف شده باشند، البته در هر صورت تأثیرگذاری آنها منوط به همکاری همه طرفین است. نهادهایی همانند سازمان تجارت جهانی نوعی از بروکراسی را ایجاد کرده‌اند که توانایی روشن کردن موارد مبهم را داشته و می‌تواند مذاکرات آتی را تسهیل کند. همکاری‌های بین‌المللی نیز اغلب با هنجارها شروع شده، به قراردادهای منطقه‌ای و چندجانبه می‌انجامد و در نهایت باعث ایجاد و تثبیت نهادها می‌شود.

احتمال تصمیم به همکاری در سطح جهانی درباره هوش مصنوعی به این فرضیات برمی‌گردد:

۱. طرف‌های درگیر به این نتیجه برسند که منفعت زیادی در دستیابی به توافقی موفق دارند.
۲. وقتی که طرف‌های درگیر رابطه‌ای مبتنی بر اعتماد داشته باشند.
۳. وقتی که اجماع کافی در این باره وجود داشته باشد که توافق به چه شکلی باشد.
- احتمال تحقق این مورد زمانی بالاتر است که ساده، گیرا، محکم و استوار باشد.
۴. هنگامی که قرارداد بسته شده به آسانی، سریعاً و به شکل عمومی قابل تأیید باشد.

1. Noseph Nye, A Normative Approach To Preventing Cyberwarfar, Project Syndicate, March 2017.



۵. انگیزه‌های به شکست کشاندن قرارداد پایین باشد.

در مقایسه با دیگر حوزه‌ها به نظر می‌رسد هوش مصنوعی کمتر پذیرای توافقات بین‌المللی است. به همین دلیل تنها در شرایطی می‌توان بر سر هوش مصنوعی به انجام توافقات بین‌المللی پرداخت که:

الف) طرف‌های درگیر به این نتیجه برسند که از رقابت‌های شدید و لجام‌گسیخته بر سر هوش مصنوعی دچار مخاطرات جدی می‌شوند و نیز منافع بسیاری در همکاری متقابل قابل تصور است.

ب) چین و غرب بتوانند روابط نسبتاً همکاری‌جویانه فعلی را به توافق‌های جدید بر سر جلوگیری از مسابقات تسلیحاتی و رقابت‌های شدید بر سر هوش مصنوعی تسری دهند.

- استانداردهای طرف سوم تأیید، اجرا و کنترل

از راهکارهایی که برای اجتناب از وقوع رقابت‌های شدید و متخاصمانه بر سر هوش مصنوعی می‌توان به کار گرفت استفاده از وضع استانداردها، تأیید، اجرا و کنترل توسط طرف سوم است. مدل نخستینی که می‌توان به کار گرفت و با قاطعیت می‌توان گفت قابلیت اجرایی را دارد، تأسیس آژانس بین‌المللی برای تعریف استانداردهای ایمنی در حوزه هوش مصنوعی و الزام به اجرای آن است. چنین قدمی به‌عنوان مرحله‌ای اساسی در دستیابی به دانش مشترک درباره آن چیزی است که می‌توان توافقنامه در زمینه هوش مصنوعی نامید. مدل دیگر، ایجاد سازمانی است که خود قابلیت‌های وضع همکاری و ایجاد توافق‌های جدید را داشته باشد. مدل آخر نیز ایجاد چارچوبی جهانی برای هوش مصنوعی

است که خارج از کنترل قدرت‌های بزرگ بوده و انحصار تکنولوژی استراتژیک (نظامی) را در اختیار داشته باشد، البته این مدل نیازمند دقت فراوان در طراحی سازمانی است تا انگیزه‌های همکاری جویانه در میان همه طرف‌ها را تضمین کند.

این مدل‌های پیشنهادی با خود پرسش‌های متعددی را درباره چگونگی عملکرد سازمان‌ها و نهادها در پی دارد: پیامدهای اعطای اختیارات کافی توسط قدرت‌های جهانی به سازمانی بین‌المللی که متکفل امور مربوط به هوش مصنوعی باشد چیست؟ ابزارها، توافقات و سناریوهای ممکن که می‌تواند در بهبود عملکرد چنین سازمانی مؤثر باشند چه خواهد بود؟ چه درس‌هایی می‌توان از نمونه‌های هم موفق و ناموفق قراردادهای بین‌المللی برای جلوگیری از رقابت‌های تسلیحاتی گرفت؟

فصل دوم - حکمرانی مطلوب هوش مصنوعی

فصل اول (سیاست هوش مصنوعی) در پی فهم این نکته بود که بازیگران متفاوت چگونه برای دسترسی به اهدافشان در رابطه با هوش مصنوعی قدرتمند با همدیگر به همکاری یا رقابت می‌پردازند. این فصل برای امکان همکاری متمرکز است: اگر ما بتوانیم به‌طور مناسبی به همکاری با یکدیگر پردازیم، چه چیزی را می‌توانیم براساس همکاری‌هایمان ایجاد کنیم؟ حکمرانی مطلوب هوش مصنوعی به بررسی ترتیبات جهانی بالقوه برای پاسخ به این پرسش می‌پردازد که انواع مختلف هوش مصنوعی چه بوده و چه کسانی و با چه اهدافی و در چه دشواری‌هایی کاربردی شده و توسعه یافته‌اند؟ همچنین حکمرانی مطلوب هوش مصنوعی دربرگیرنده تلاش‌های انجام شده برای شناسایی مدل‌های ایدئال حکمرانی



جهانی نیز است. ارزش‌های مشترک انسانی ما چیست و به چه ترتیبی می‌توانیم به بهترین وجه به اهداف متفاوت خودمان برسیم؟ چه اصول سازمانی و مکانیسم‌های نهادی برای تحقق هدف ذکر شده وجود دارد؟

تمرکز بر مسئله حکمرانی جهانی مطلوب در نسبت با هوش مصنوعی به دو دلیل

است:

۱. اندیشیدن به آنچه که می‌خواهیم به آن برسیم بسیار ضروری است.

۲. با همفکری درباره بصیرت مشترک می‌توان به تسهیل همکاری پرداخت.

در همین راستا ضروری است که به گسترش هر چه بیشتر ابزارهای مؤثر ایجاد ارتباط با همدیگر و منافع بالقوه همکاری مشترک را مرتب گوشزد کنیم و در مقابل خطرات عمیق ناشی از رفتار متخاصمه‌جویانه را نیز در نظر داشته باشیم. همچنین نیازمند آن هستیم که مطمئن باشیم از نگرانی‌های متفاوت و جهان‌بینی‌های مردم و نخبگان با خبر هستیم.

به همین منظور باید به شکلی عملگرایانه درباره ویژگی‌های نهادی چنین شیوه‌ای از حکمرانی بیندیشیم. پایه‌های حقوقی چنین نهادی چگونه باید باشد؟ چه کسی و در چه شرایطی تصمیم‌گیری می‌کند؟ پاسخ دقیق به این پرسش‌ها و پرسش‌های مشابه برای تضمین اینکه هر نوع نگاه مطلوب قابل پذیرش به حکمرانی از نظر سیاسی باثبات باشد ضروری است.

یافته‌های این بخش برای هر نوع اندیشیدن جدی به‌مثابه حکمرانی مطلوب هوش

مصنوعی به دلایل زیر اساسی است:

۱. مسائل مربوط به حکمرانی که امروزه با آن مواجهیم و در آینده نیز مواجه خواهیم شد عمدتاً با یکدیگر همپوشانی دارند به استثنای:

الف) دامنه‌ای مختلف از علائق و منافع که باید بیان شود.

ب) نیاز بالقوه به رقابت در بعضی قلمروهای اقتصادی و نظامی بزرگ‌تر.

برای توضیح بیشتر بعضی از این تشابهات، می‌توان به این نکته توجه کرد که چگونه مدیریت یک ائتلاف جهانی در حوزه هوش مصنوعی به شکل ایدئالی باید تعهداتی اساسی به نفع عمومی داشته و دربرگیرنده سازوکارهای نهادی برای تضمین اصولی همچون عدم تهدید دیگران و... باشد. درواقع اگر بتوانیم مدلی جذاب و واقع‌گرا از هوش مصنوعی ارائه کنیم، چنین مدلی می‌تواند به الگوی راهنمای ما برای بیرون آمدن از موازنه‌های مبتنی بر رقابت و تخاصم کنونی باشد.

۲. ما معمولاً عادت داریم که به چارچوب‌های کنونی حکمرانی چسبیده و از آنها دست نکشیم.

- ارزش‌ها و اصول

حکمرانی مطلوب هوش مصنوعی در پی اندیشیدن به راه‌حل‌های نهادی ایدئال برای چالش‌هایی است که حکمرانی هوش مصنوعی در مقابل انسان‌ها قرار می‌دهد. اصول و ارزش‌های مشترکی که انسان‌ها می‌توانند براساس آنها با همدیگر به همکاری بپردازند چیست؟ خواست‌های عامه مردم، گروه‌های فرهنگی، محققان هوش مصنوعی، نخبگان، حکومت‌ها و اقشار گوناگون در کوتاه‌مدت و بلندمدت از هوش مصنوعی چیست؟ بهترین راه‌های مصالحه میان گروه‌های متفاوت و ارزش‌های متضاد چیست؟ براساس آنچه تاکنون



درباره خودمان و ارزش‌هایمان یاد گرفته‌ایم آیا می‌توان انتظار مسیری مشخص را داشت که ارزش‌هایمان در جهت آن حرکت کنند؟

محققان مختلف مجموعه‌ای از کمبودها در حوزه سیاستگذاری را مشخص کرده‌اند که اهمیت زیادی در هوش مصنوعی فوق هوشمند دارد که شامل این موارد می‌شود: پیشرفت سریع، ایمنی هوش مصنوعی، ثبات نسبی، نفع عمومی، تداوم، اندیشیدن درباره اصول اولیه، سرعت و سرنوشت، خرد و انطباق‌پذیری.

به‌نظر می‌رسد دو اصل مهم در جهان وجود دارد که در تضاد با یکدیگرند:

۱. امنیت (شامل ایمنی هوش مصنوعی، ثبات نسبی)،

۲. خودبنیادی (آزادی، تداوم، حاکمیت).

- نهادها و مکانیسم‌ها

انسان‌ها خواستار فراهم کردن امنیت و تضمین ایمنی هوش مصنوعی از طرف حکومت‌هایشان هستند تا گسترش تکنولوژی‌های نوین آنها را با مخاطرات جدی روبه‌رو نکند. چنین دغدغه‌ای نیازمند کنترل تمرکزگرا بر توسعه هوش مصنوعی و یا نظارت گسترده بر روی پروژه‌های مختلف هوش مصنوعی است. چنین امنیتی احتمالاً از طریق همکاری گروه‌های مختلف در جهان امکان‌پذیر می‌شود و احتمالاً نیازمند تمرکز قدرت و ثروت است.

مسئله چگونگی ایجاد نهادهایی برای مدیریت امور سیاسی هسته اصلی علوم سیاسی و اقتصاد سیاسی است. دانشمندان سیاسی بالاخص در رشته سیاست تطبیقی به‌طور گسترده‌ای ویژگی‌های نظامی سیاسی گوناگون را مطالعه کرده‌اند. دانشمندان رشته علوم سیاسی و نظریه‌های سیاسی نیز به مطالعه قانون‌های اساسی گوناگون روی آورده‌اند.

به‌هر حال اگرچه دانشمندان رشته علوم سیاسی به مطالعه موضوعات گوناگونی پرداخته‌اند و دانش ما را درباره آنها وسعت بخشیده‌اند اما تاکنون تحقیقات جامعی درباره حکمرانی هوش مصنوعی انجام نشده است و گزارش حاضر نیز بنا به کمبود مطالعات جامع در این‌باره تنها درآمدی بر طرح مسئله حکمرانی هوش مصنوعی می‌باشد.

نتیجه‌گیری

تاکنون بررسی علمی جامع درباره تأثیرات هوش مصنوعی بر روند حکمرانی در جهان انجام نشده است. در حالی که بیشتر دانشمندان علم سیاست و بالاخص سیاست تطبیقی بر تأثیر هوش مصنوعی بر فرآیندهای حکمرانی اتفاق نظر دارند هنوز هم دامنه، ابعاد و پیامدهای این تأثیرگذاری به‌طور کامل واکاوی نشده است. این گزارش با توجه به ضرورت انجام پژوهش در این‌باره نوشته شده است. این گزارش تنها درآمد مقدماتی بر طرح مسئله حکمرانی هوش مصنوعی بود و هنوز به پژوهش‌های بیشتری در این‌باره نیاز است.

در فصل اول گزارش، سیاست هوش مصنوعی یعنی پویایی میان دولت‌ها، کمپانی‌های بزرگ، عامه مردم، محققان و دیگر کنشگران بررسی شده است. همچنین بر نقش ایجابی هوش مصنوعی در حل مشکلات جهانی بالاخص کنش‌های بین‌المللی تأکید شده است. همچنان‌که ملاحظه شد سیاست بین‌المللی به این نکته توجه دارد که چگونه چشم‌انداز فنی در حال تغییر می‌تواند سیاست داخلی، اقتصاد سیاسی بین‌المللی و امنیت بین‌المللی را دگرگون کند.



در بخش حکمرانی مطلوب هوش مصنوعی، ساختارهای ممکن برای گذار به مدیریت هوش مصنوعی تحلیل شده است. در همین راستا بر گسترش هر چه بیشتر ابزارهای مؤثر ایجاد ارتباط با همدیگر و منافع بالقوه در همکاری‌های مشترک تأکید شده است. همچنین ضرورت ائتلاف جهانی در حوزه هوش مصنوعی که به شکل ایدئالی تعهداتی اساسی به‌نفع عمومی داشته و دربرگیرنده مکانیسم‌های نهادی برای تضمین اصولی همچون عدم تهدید دیگران باشد، گوشزد شده است. گزارش مزبور همچنین پرسش‌های متعدد اساسی درباره حکمرانی هوش مصنوعی طرح می‌کند بدون آنکه ضرورتاً پاسخ واضحی به آنها بدهد. به همین دلیل نوشتن گزارش حاضر تنها درآمد مقدماتی بر طرح مسئله حکمرانی هوش مصنوعی بوده و در آینده با پیشرفت هوش مصنوعی احتمالاً شاهد انتشار گزارش‌های جامع‌تری در این‌باره خواهیم بود.

منابع و مأخذ

1. Ian Hogarth, AI Nationalism, Ian Hogarth (blog), Tunc 13, 2018.
2. Allan Dafoe, AI Governance: A research agenda , Oxford university, 2018.
3. David Barboza, How This U.S Tech Giant is Backing China's Tech Ambitions, New York Times, August 4, 2017.
4. Noseph Nye, A Normative Approach to preventing cyberwarfar, Project Syndicate, March 2017.
5. Korinek, Anton, and Joseph E. Stiglitz. "Artificial intelligence and its implications for income distribution and unemployment." No. w24174. National Bureau of Economic Research, 2017.
6. Dorn, David, Lawrence F. Katz, Christina Patterson, and John Van Reenen. "Concentrating on the Fall of the Labor Share." American Economic Review 107, no. 5 (2017): 180-85.
7. Ryan, Amy and Gary Keeley. "Sputnik and US Intelligence: The Warning Record." Studies in Intelligence 61:3 (2017).
8. Bostrom, Nick. "Strategic Implications of Openness in AI Development." *Global Policy* , February 2017.
9. Hörner, Johannes. "A Perpetual Race to Stay Ahead." Review of Economic Studies (2004) 71.



مرکز پژوهش‌ها
مجلس شورای اسلامی

شماره مسلسل: ۱۶۶۹۲

شناسنامه گزارش

عنوان گزارش: درآمدی بر حکمرانی هوش مصنوعی خلاصه راهبردی از:

Allan Dafoe, AI Governance: A research agenda , Oxford university, 2018

نام دفتر: مطالعات سیاسی (گروه بنیادین حکومتی)

تهیه و تدوین: خالد شیخ‌الاسلامی

ناظر علمی: سیدیونس ادیانی

ویراستار تخصصی: —

ویراستار ادبی: —

واژه‌های کلیدی:

۱. هوش
۲. هوش مصنوعی
۳. سیاست هوش مصنوعی
۴. حکمرانی هوش مصنوعی
۵. حکمرانی مطلوب هوش مصنوعی
۶. چالش‌های هوش مصنوعی



تاریخ انتشار: ۱۳۹۸/۸/۱۱