وزارت ارتباطات و فناوری اطلاعات
پژوهشگاه ارتباطات و فناوری اطلاعات

تنظیم‌گری هوش مصنوعی
مولد: ربات چت ChatGPT

تهیه‌کننده: اعظم صادق‌زاده

عنوان گزارش: تنظیم‌گری هوش مصنوعی مولد: ربات چت ChatGPT

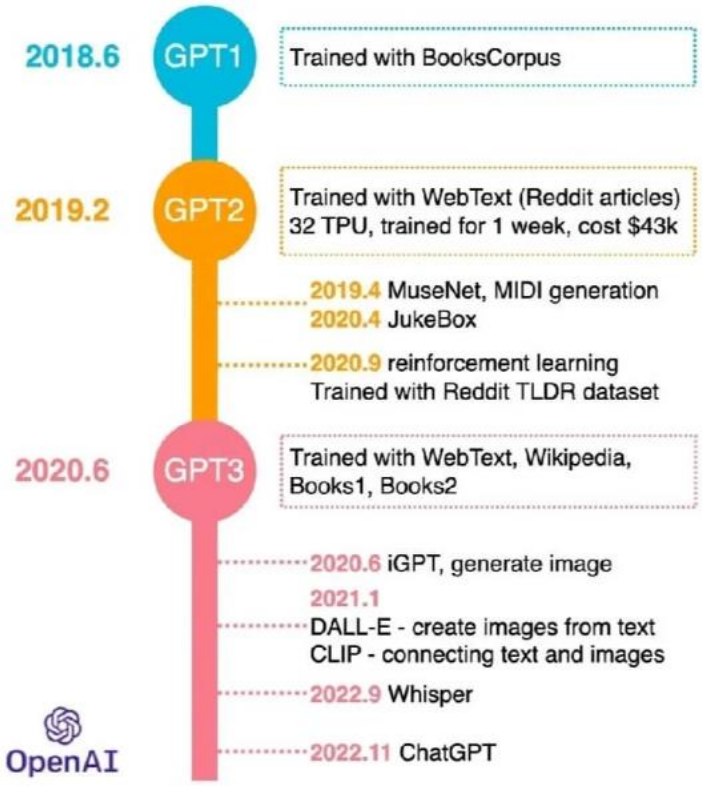
تهیه کننده: اعظم صادق زاده

گروه پژوهشی: معاونت پژوهش و توسعه ارتباطات علمی

تاریخ نشر: ۱۴۰۲

مقدمه

هوش مصنوعی مولد به مجموعه ای از فناوری های هوش مصنوعی اطلاق می شود که محتوای جدید را براساس درخواست کاربران تولید می کند. ChatGPT مخفف Generative Pre-trained Transformer یک ابزار هوش مصنوعی مولد و مدل زبانی است که از یادگیری خلاقانه برای تولید متون شبیه آنچه انسان واقعی انجام می دهد استفاده و پاسخ ها و تعامل انسانی را تقلید می کند. این برنامه در ابتدا برای پاسخگویی به سوالات کاربران اینترنت ساخته شد، اما به سرعت به ابزار جادویی اینترنت تبدیل شد که به هر کسی این امکان را داده است تا با استفاده از یک شکل بسیار منحصر به فرد از هوش مصنوعی و با یک کلیک، قابلیت ها و امکانات فوق العاده ای داشته باشد. یک کتاب یا مقاله بنویسید یا حتی یک برنامه کامپیوتری طراحی کند. ChatGPT توسط Open AI توسعه یافته است که ارزش تخمینی آن در سال ۲۰۲۳ حدود ۲۹ میلیارد دلار برآورد شده است.



شکل ۱: روند توسعه زمانی قابلیت ها و ویژگی های ربات چت ChatGPT

۱. چالش ها و مخاطرات توسعه کاربرد ChatGPT

سرعت نفوذ ربات چت ChatGPT در مقابل سایر خدمات فناورانه نشان دهنده ظهور یک انقلاب فناوری است. دستیابی به یک میلیون کاربر در ChatGPT طی ۵ روز مطابق نمودار زیر در مقایسه با بسیاری از خدمات محبوب دیجیتالی مانند اینستاگرام و یا خدمات کاربردی و یا جذاب مانند Airbnb وNetflix ، شتاب و سرعت توسعه هوش مصنوعی مولد را نمایش می دهد.

شکل ۲: مقایسه زمان دستیابی به ۱ میلیون کاربر در خدمات فناورانه

در حالیکه ChatGPT مزایای بی شماری دارد و برای کاربردهای بسیاری می تواند مفید باشد، توجه به این نکته مهم است که ChatGPT فقط یک ابزار است و نباید به طور کامل به آن اعتماد کرد. متن تولید شده توسط هوش مصنوعی ممکن است همیشه دقیق یا قابل اعتماد نباشد، همچنانکه تکیه بر هوش مصنوعی برای تصمیم گیری در مورد استخدام، مراقبت های بهداشتی یا سیستم های قضایی نیز می تواند عواقب جدی داشته باشد. بنابراین لازم است جنبه های خاکستری و تاریک این فناوری شناسایی و به آن پرداخته شود.

- **اطلاعات نادرست**: یکی از بزرگترین نگرانی ها در مورد مدل های زبانی مانند ChatGPT، احتمال گسترش اطلاعات نادرست است. از آنجایی که ChatGPT بر روی مقادیر زیادی داده از اینترنت آموزش دیده است، توانایی ایجاد پاسخ به سوالاتی را دارد که ممکن است گمراه کننده یا نادرست باشد. این مساله می تواند منجر به انتشار

اطلاعات نادرست شود و حتی در شرایط خاص و برخی حوزه ها مانند سلامت، مخاطرات گسترده ای به همراه داشته باشد.

- **مالکیت فکری**: یکی از ابعاد منفی توسعه ChatGPT ایجاد چالش در زمینه حقوق مالکیت فکری است، زیرا آثار یا عناصر تولید شده توسط Chat GPT به واسطه عدم وجود عنصر خلاقیت انسانی، ممکن است از حمایت قانونی مالکیت فکری برخوردار نباشند و این مساله زمینه ساز چالش های حقوقی بسیاری خواهد شد.
- **کاهش تعامل انسانی**: یکی از آشکارترین عوارض جانبی استفاده بیش از حد ChatGPT، وابستگی فزاینده به هوش مصنوعی برای کارهای روزمره و ارتباطات است. همزمان با وابستگی بیشتر کاربران به ChatGPT، کاهش تدریجی در تعاملات انسانی بیشتر تجربه می شود. این کاهش در ارتباط شخصی می تواند منجر به احساس انزوا، تنهایی و تضعیف مهارت های فردی شود. انسان ها به عنوان موجودات اجتماعی نیازمند تعامل واقعی برای حفظ رفاه عاطفی و ذهنی خود هستند. بنابراین، اتکای بیش از حد به هوش مصنوعی، سلامت اجتماعی کاربران را به خطر می اندازد.
- **از دست دادن تفکر انتقادی و خلاقیت**: استفاده گسترده از ChatGPT منجر به کاهش مهارت کاربران در تفکر انتقادی و خلاقیت می شود. سیستم هوش مصنوعی می‌تواند پاسخ‌ها و پیشنهادهایی را بر اساس داده‌های از قبل موجود ایجاد کند، اما ممکن است همیشه ایده‌های اصلی یا نوآورانه ارائه نکند. با تکیه بر ابزار برای کارهای خلاقانه، کاربران ممکن است به طور ناخواسته با تضعیف خلاقیت، توانایی خود را برای تفکر خارج از چارچوب محدود کنند.
- **حفظ حریم خصوصی و امنیت**: استفاده بسیار از ChatGPT کاربران را در معرض خطرات نقض حریم خصوصی و چالش های امنیتی قرار می دهد. همانطور که کاربران داده های بیشتری را به سیستم وارد می کنند، به طور بالقوه فرصت های بیشتری نیز برای دستیابی به اطلاعات حساس در اختیار هکرها قرار می گیرد.
- **فرسایش مهارت های زبانی و کاهش کیفیت ارتباطات**: اگر کاربران برای فعالیت هایی که به درک و بیان زبان پیچیده نیاز دارند به ابزار ChatGPT تکیه کنند، پیامد آن می تواند به کاهش تدریجی مهارت های زبانی کاربران منجر شود.

- **اعتیاد دیجیتال و چالش های سلامت روان**: ساعت های طولانی تعامل با هوش مصنوعی ممکن است منجر به وابستگی ناسالم به دنیای دیجیتال شود و موجب شود کاربران از روابط و مسئولیت های واقعی خود غافل شوند.
- **تاثیر بر اشتغال**: یکی دیگر از تأثیرات منفی مدل های زبانی مانند ChatGPT، تهدید اشتغال است. همانطور که ChatGPT به پیشرفت خود ادامه می دهد، این پتانسیل را دارد که فعالیت هایی را که توسط انسان قابل انجام است، خودکار شود. این مساله می تواند با حذف مشاغل بر اقتصاد تأثیر منفی بگذارد.

۲. تنظیم گری هوش مصنوعی در کشورهای مختلف

توسعه سریع هوش مصنوعی و برخی پیامدهای آشکار و پنهان آن، درخواست برای تنظیم گری این فناوری را افزایش داده است. در این رابطه برخی کشورها اقداماتی را اجرا کرده و برخی دیگر به دنبال پیگیری موضوع هستند. اما سرعت پیشرفت فناوری به حدی است که برای دولت ها، مواجهه موثر، دشوار است. هوش مصنوعی مولد اکنون می تواند هنر واقع گرایانه خلق کند، مقاله کامل بنویسد یا حتی خطوط کد را درعرض چند ثانیه تولید کند. در مقابل، تنظیم گران نگران چالش هایی هستند که هوش مصنوعی برای امنیت شغلی، حریم خصوصی داده ها و برابری ایجاد می کند. همچنانکه دستکاری هوش مصنوعی پیشرفته در حوزه های سیاسی از طریق تولید اطلاعات نادرست، مخاطرات گسترده ای را به همراه خواهد داشت. در اینجا آخرین اقداماتی که نهادهای حاکم ملی و بین المللی و سازمان ها برای تنظیم ابزارهای هوش مصنوعی (مولد) انجام داده اند، اشاره شده است.

❖ *ایتالیا*

ایتالیا اولین کشور غربی است که ChatGPT را ممنوع کرده است. ChatGPT با قابلیت های خود علاوه بر محققین، قانونگذاران و کارشناسان را در مورد پیامدهای منفی برای جامعه نگران کرده و تحت تاثیر قرارداده است. توسعه ChatGPT و چالش های ناشی از آن موجب شده است که ایتالیا، ربات چت محبوب استارتاپ Open AI آمریکا را ممنوع کند. در همین رابطه، سازمان ناظر حفاظت از داده های ایتالیا به Open AI دستور داد تا به طور موقت پردازش داده های کاربران ایتالیایی را به دلیل آنچه انجام تحقیقات در مورد نقض قوانین سختگیرانه حریم

خصوصی اروپا (GDPR) عنوان شده، متوقف کند. یکی از موارد نقض داده در open AI این است که به کاربران اجازه می دهد عناوین مکالماتی را که سایر کاربران با ربات چت داشته اند، مشاهده نمایند. علاوه بر این، نهاد ناظر در ایتالیا معتقد است هیچ مبنای قانونی که پشتوانه جمع آوری و پردازش گسترش داده های شخصی به منظور آموزش الگوریتم های پلتفرم است، وجود ندارد. عدم محدودیت سنی در کاربرد ChatGPT و ارائه اطلاعات نادرست در پاسخ های ربات چت نیز از جمله نگرانی های دولت ایتالیا اعلام شده است. با توجه به این موارد، Open AI که توسط مایکروسافت پشتیبانی می شود، در صورتیکه طی مدت ۲۰ روز راه حلی برای این وضعیت ارائه نکند، با پرداخت جریمه ۲۰ میلیون یورو یا ۴ درصد از درآمد سالانه خود مواجه خواهدشد.

ایتالیا تنها کشوری نیست که با سرعت پیشرفت هوش مصنوعی و پیامدهای آن برای جامعه درگیر است. کشورهای دیگر نیز قوانین خود را برای هوش مصنوعی ارائه کرده اند. اگرچه اغلب این تنظیم گری بدون اشاره به هوش مصنوعی مولد است، ولیکن بدون شک روی آن تاثیر خواهد گذاشت.

❖ بریتانیا

طی ماه های اخیر، بریتانیا نیز برنامه های خود را برای تنظیم مقررات هوش مصنوعی اعلام کرده است. در این کشور به جای ایجاد مقررات جدید، دولت از تنظیم کننده ها در بخش های مختلف خواسته است تا مقررات موجود را برای هوش مصنوعی تطبیق دهند. در پیشنهادات بریتانیا بدون ذکر نام ChatGPT، برخی از اصول کلیدی مانند ایمنی، شفافیت، انصاف، مسئولیت پذیری و مشروعیت که شرکت ها هنگام استفاده از هوش مصنوعی در محصولات خود ملزم به رعایت آن هستند را مشخص کرده است. بریتانیا قصد دارد اطمینان حاصل کند که شرکت ها در توسعه و استفاده مسئولانه از ابزارهای هوش مصنوعی متعهد هستند و به کاربران، اطلاعات کافی در مورد چرایی و چگونگی تصمیم گیری های خود ارائه می دهند. به نظر می رسد، اولویت اصلی رویکرد بریتانیا پرداختن به استفاده خوب از هوش مصنوعی است.

هرچند بریتانیا در این مرحله، محدودیتی برای ChatGPT یا هر نوع ابزار هوش مصنوعی پیشنهاد نکرده است، ولیکن وزیر دیجیتال این کشور اعلام کرده است که محبوبیت ناگهانی هوش مصنوعی مولد نشان می دهد که خطرات

و فرصت های پیرامون این فناوری با سرعت خارق العاده ای درحال ظهور هستند و دولت ها باید با مداخله بیشتر نسبت به اتخاذ رویکردهای غیرقانونی در پیشرفت های هوش مصنوعی پاسخ دهند.

❖ اتحادیه اروپا

اتحادیه اروپا که اغلب در خط مقدم مقررات گذاری فنی مطرح می شود، قانون پیشگامانه ای در مورد هوش مصنوعی پیشنهاد کرده است. این قانون که به عنوان قانون هوش مصنوعی اروپا شناخته می شود، استفاده از هوش مصنوعی را در زیرساخت های حیاتی، آموزش، اجرای قانون و سیستم قضایی به شدت محدود می کند. این قانون در راستای مقررات عمومی حفاظت از داده اتحادیه اروپا، نحوه پردازش و ذخیره سازی داده های شخصی توسط شرکت ها را تنظیم می کند. به گزارش رویترز، پیش نویس قوانین اتحادیه اروپا، ChatGPT را نوعی هوش مصنوعی با هدف عمومی می داند که در برنامه های پرخطر استفاده می شود. سیستم های هوش مصنوعی پرخطر به عنوان سیستم هایی تعریف می شوند که می توانند بر حقوق اساسی و ایمنی افراد تاثیر گذارند.

در حالیکه بروکسل، قوانین مربوط به هوش مصنوعی را اصلاح می کند، برخی کشورهای اتحادیه اروپا در حال بررسی اقدامات ایتالیا در مورد ChatGPT هستند. کمیسر فدرال حفاظت از داده های آلمان گفته است که اجرای رویه مشابه در آلمان نیز امکان پذیر است. تنظیم کننده های حریم خصوصی فرانسوی و ایرلندی نیز با همتایان خود در ایتالیا در حال بررسی موضوع هستند.

❖ آمریکا

ایالات متحده هنوز هیچ قانون رسمی برای نظارت بر فناوری هوش مصنوعی پیشنهاد نکرده است. موسسه ملی علم و فناوری این کشور با ارائه یک چارچوب ملی به شرکت هایی که سیستم های هوش منصوعی را استفاده، طراحی یا استقرار می دهند، در مورد ریسک ها و آسیب های احتمالی مشاوره می دهد. البته این مشاوره ها به صورت داوطلبانه اجرا می شود و شرکت ها با عدم رعایت قوانین، متحمل هیچگونه عواقبی نمی شوند. برهمین اساس نیز تا کنون هیچ اقدامی برای محدودکردن ChatGPT در ایالات متحده انجام نشده است.

البته دریافت شکایت از سوی یک گروه تحقیقاتی غیرانتفاعی طی ماه گذشته درخصوص تهدید حریم خصوصی و امنیت عمومی توسط GPT-۴ آخرین مدل زبان Open AI و نقض دستورالعمل های هوش مصنوعی آژانس، می تواند منجر به تحقیق در مورد Open AI و احتمالا تعلیق استقرار تجاری مدل های زبان بزرگ شود.

❖ چین

ChatGPT در چین و کشورهای مختلف که با سانسور شدید اینترنت مواجه هستند مانند کره شمالی و روسیه در دسترس نیست. ChatGPT به طور رسمی در این کشور مسدود نشده است اما Open AI به کاربران در این کشور اجازه ثبت نام نمی دهد. چندین شرکت بزرگ فناوری در چین در حال توسعه خدمت جایگزین هستند. بایدو، علی بابا، JD.com و برخی از بزرگترین شرکت های فناوری چین، برنامه هایی را برای رقبای ChatGPT اعلام کرده اند. چین مشتاق است اطمینان حاصل کند که غول های فناوری این کشور محصولاتی را مطابق با قوانین سختگیرانه خود توسعه می دهند.

تنظیم‌کننده فضای سایبری چین در ماه آوریل ضمن ارائه پیش‌نویس اقدامات مدیریت خدمات هوش مصنوعی مولد، اعلام کرد که از شرکت‌ها می‌خواهد ارزیابی‌های امنیتی خود را قبل از ارائه خدمات به عموم به مقامات ارائه کنند. دفتر اقتصاد و فناوری اطلاعات این کشور نیز در ماه فوریه اعلام کرد، از شرکت‌های پیشرو در ساخت مدل‌های هوش مصنوعی که می‌توانند ChatGPT را به چالش بکشند، حمایت خواهد کرد.

ماه گذشته پکن اولین مقررات در نوع خود را در مورد جعل عمیق تصاویر، ویدئوها یا متن های ساخته شده با هوش مصنوعی که به طور مصنوعی تولید و یا تغییر داده شده، معرفی کرد. رگولاتوری چین، قبلا قوانینی برای نحوه عملکرد شرکت ها با الگوریتم ها معرفی کرده بود. در همین رابطه، شرکت ها باید جزئیات الگوریتم های خود را به رگولاتور فضای مجازی ارسال کنند. چنین مقرراتی می تواند برای هر نوع فناوری به سبک ChatGPT نیز اعمال شود.

❖ فرانسه

به گزارش سازمان نظارت بر حریم خصوصی فرانسه CNIL، پس از ممنوعیت موقت ChatGPT در ایتالیا به دلیل مشکوک بودن به نقض قوانین حفظ حریم خصوصی، چندین شکایت از ChatGPT در این کشور دریافت کرده است.

البته مجلس ملی فرانسه در ماه مارس استفاده از نظارت تصویری هوش مصنوعی را در بازی‌های المپیک ۲۰۲۴ پاریس، بدون توجه به هشدارهای گروه‌های حقوق مدنی مبنی بر اینکه این فناوری تهدیدی برای آزادی‌های مدنی است، تصویب کرد.

Science Po، یکی از برترین دانشگاه های فرانسه استفاده از ChatGPT را در ژانویه امسال ممنوع اعلام کرد. دلیل این تصمیم گیری، افزایش سرقت محتوای آنلاین توسط دانشجویان اعلام شده است.

❖ ژاپن

وزیر تحول دیجیتال این کشور اعلام کرد، طی برگزاری نشست وزرای دیجیتال گروه G۷ در تاریخ ۲۹ تا ۳۰ آوریل، فن‌آوری‌های هوش مصنوعی از جمله ChatGPT مورد بحث قرارگرفته و پیام یکپارچه ای در این خصوص توسط این گروه صادر می شود.

❖ اسپانیا

آژانس حفاظت از داده های اسپانیا در ماه آوریل اعلام کرد که در حال آغاز تحقیقات اولیه در مورد نقض احتمالی داده ها توسط ChatGPT است. این آژانس اعلام کرده است که همچنین از سازمان نظارت بر حریم خصوصی اتحادیه اروپا خواسته است تا نگرانی های مربوط به حریم خصوصی ChatGPT را ارزیابی کند.

در حالیکه برخی کشورها در بهره گیری و استفاده از ChatGPT و ابزارهای هوش مصنوعی عجله دارند، برخی کشورها نیر به شدت به مقررات متکی هستند و برخی دیگر استفاده از آن را به طور کامل ممنوع اعلام کرده اند. همانطور که اشاره شد، کشورها اغلب براساس نگرانی های مربوط به حفظ حریم خصوصی این برنامه را ممنوع اعلام کرده اند و برخی دیگر به ویژه کره شمالی، چین و روسیه ادعا می کنند که ایالات متحده از ChatGPT برای انتشار اطلاعات نادرست استفاده خواهد کرد. در کنار ممنوعیت های کشورها، توسعه دهنده ChatGPT Open AI نیز استفاده از ChatGPT را در برخی کشورها مسدود کرده است. برخی از این کشورها در زمان نگارش این متن عبارتند از: روسیه، چین، کره شمالی، کوبا، سوریه، ایتالیا و ایران. Open AI همچنین قبلا اکراین را به دلیل ناتوانی در مسدودسازی کریمه که در اشغال روسیه است، ممنوع کرده بود. با این حال این ممنوعیت اخیراً برطرف شده است.

علاوه بر دولت ها، شرکت های بزرگ فناوری نیز در این رابطه دارای تنظیم گری هستند. به گزارش بلومبرگ، استفاده کارکنان شرکت سامسونگ از ابزارهای مولد هوش مصنوعی مانند ChatGPT و Bard به دلیل نگرانی از مخاطرات امنیتی، متوقف و ممنوع شده است. این تصمیم به دنبال انتشار داده های حساس مربوط به نیمه هادی ها توسط کارکنان این شرکت در ChatGPT اتخاذ شده است. ChatGPT مانند سایر ابزارهای هوش مصنوعی مولد، اکثرا بر روی داده های ارسالی آموزش داده شده است. بنابراین هر چیزی که توسط کارکنان سامسونگ به آن داده شده است، در پاسخ به درخواست کاربر دیگری در هر کجای دنیا می تواند نمایش داده شود. اگرچه توسعه دهنده ChatGPT Open AI اخیرا یک حالت ناشناس اضافه کرده است که به کاربران اجازه می دهد از ورودی های خود برای استفاده از آموزش مدل هوش مصنوعی جلوگیری کنند.

سامسونگ در حال حاضر، بررسی اقدامات امنیتی برای ایجاد محیطی امن و استفاده ایمن از هوش مصنوعی مولد برای افزایش بهره وری و کارایی کارکنان را آغاز کرده است و تا زمانی که اقدامات مورد توافق قرار نگیرد، این شرکت استفاده از هوش مصنوعی مولد را محدود می کند. علاوه بر این، سامسونگ در حال توسعه پلتفرم هوش مصنوعی خود برای ترجمه، خلاصه‌سازی اسناد و توسعه نرم‌افزار است.

سامسونگ تنها کسب و کاری نیست که به دلیل نگرانی های امنیتی علیه این فناوری، موضع گیری کرده است. بانک ها و موسساتی مانند Bank of America Corp, JPMorgan Chase & Co. و Citigroup Inc. نیز اقداماتی را جهت ممنوعیت یا محدود کردن استفاده از هوش مصنوعی مولد اتخاذ کرده اند.

۳. موضوعات پیش رو

ربات چت هوش مصنوعی ChatGPT علاوه بر تهدیدهایی که اشاره شد، با کاهش زمان توسعه و فراهم نمودن قابلیت های ویژه به کاربران، ظرفیت های گسترده ای در فعالیت ها و کسب وکارهای مختلف از جمله حوزه های مالی، زنجیره تامین، آموزش، پشتیبانی خدمات و ... فراهم کرده است که قطعا تا به امروز فقط برخی از آن ها آشکار شده است. با توجه به اینکه دامنه مقررات گذاری فناوری براساس مجموعه کاربردها و مخاطرات تنظیم می شود،

توسعه آزمایشگاه های هوش مصنوعی مولد جهت شناسایی ارزش افزوده و فرصت های فناوری این حوزه از هوش مصنوعی و نیازسنجی توانمندسازهای این فناوری در بخش های مختلف کشور، به طور موثر قابل بهره برداری است.

نشانی: تهران، انتهای کارگر شمالی، پژوهشگاه ارتباطات و فناوری اطلاعات، معاونت پژوهش و توسعه ارتباطات علمی

تلفن: ۰۲۱-۸۸۶۳۰۳۵۵

نمابر: ۰۲۱-۸۸۶۳۰۳۵۶