

مقدمه:

ظهور مدل‌های زبان بزرگ (LLM) را می‌توان به پیشرفت در روش‌های یادگیری عمیق (DL)، در دسترس بودن منابع محاسباتی عظیم، و در دسترس بودن مقادیر زیادی از داده‌های آموزشی نسبت داد. این مدل‌ها که اغلب بر روی مجموعه‌های گسترده از وب از قبل آموزش داده شده‌اند، توانایی یادگیری الگوهای پیچیده، تفاوت‌های زبانی و روابط معنایی را دارند. تنظیم دقیق این مدل‌ها در کارهای پایین دستی خاص، نتایج امیدوارکننده‌ای را نشان داده است و به عملکرد پیشرفته‌ای در معیارهای مختلف دست یافتند (1). و اینکه از دیرباز هدف دانشمندان دستیابی به خواندن، نوشتن و ارتباط انسان گونه بوده اند و این مساله از دیرباز یک چالش تحقیقاتی طولانی مدت بوده است (4)(5)

نگاهی کوتاه به ظهور lmها:

درواقع مدل سازی زبان (LM) یک رویکرد حیاتی برای تقویت هوش زبانی ماشین ها است و به طوری که از شکل ۱ قابل مشاهده است، تحقیقات مربوط به مدل زبان (LM) توجه گسترده‌ای را به خود جلب کرده و چهار مرحله توسعه مهم را طی کرده است، به شرح زیر:

اولین مرحله در توسعه lmها statistical language models مانند n-gram-models بوده است (7)(8) این مدل‌ها احتمال وقوع کلمه بعدی در یک دنباله را بر اساس فراوانی تکرار n-گرام‌های قبلی کلمات تخمین می‌زنند. (9)(10)

دومین مرحله توسعه lmها مرحله دوم توسعه مدل زبان (LM) شامل معرفی مدل‌های زبانی مبتنی بر شبکه عصبی بود که به آن‌ها مدل‌های زبانی عصبی (NLMS) نیز گفته می‌شود. (11) این رویکرد که به عنوان مدل‌سازی زبانی عصبی نیز شناخته می‌شود، از شبکه‌های عصبی برای پیش‌بینی توزیع احتمال کلمه بعدی در یک دنباله با توجه به کلمات قبلی در دنباله استفاده می‌کند. (12)(13) (RNNs)

مرحله سوم توسعه مدل زبان (LM) شامل ایجاد تعبیرهای کلمات متنی است که هدف آن ضبط معنا و سازوکار کلمات در جمله یا متن است. این تعبیرها به عنوان مدل‌های زبان پیش‌آموزش دیده (PLMs) نامگذاری می‌شوند. این مدل‌ها از شبکه‌های عصبی استفاده می‌کنند تا یک نمایش برداری (embedding) از کلمات را یاد بگیرند که با در نظر گرفتن متناسب با متن کلمه ظاهر می‌شود. مانند مدل BERT (14)

مرحله چهارم توسعه مدل زبان (LM) شامل ایجاد مدل‌های زبانی پیش‌آموزش دیده در مقیاس بزرگ است که به آن‌ها مدل‌های زبانی بزرگ (LLMs) نیز گفته می‌شود. که قادر به انجام از تسک های مختلف پردازش زبان طبیعی (NLP) با عملکرد بسیار عالی هستند. مانند GPT3 که روی تعداد زیادی داده آموزش دیده اند و میتوانند برای یک تسک خاص fine-tuned شوند.

انواع lmها :

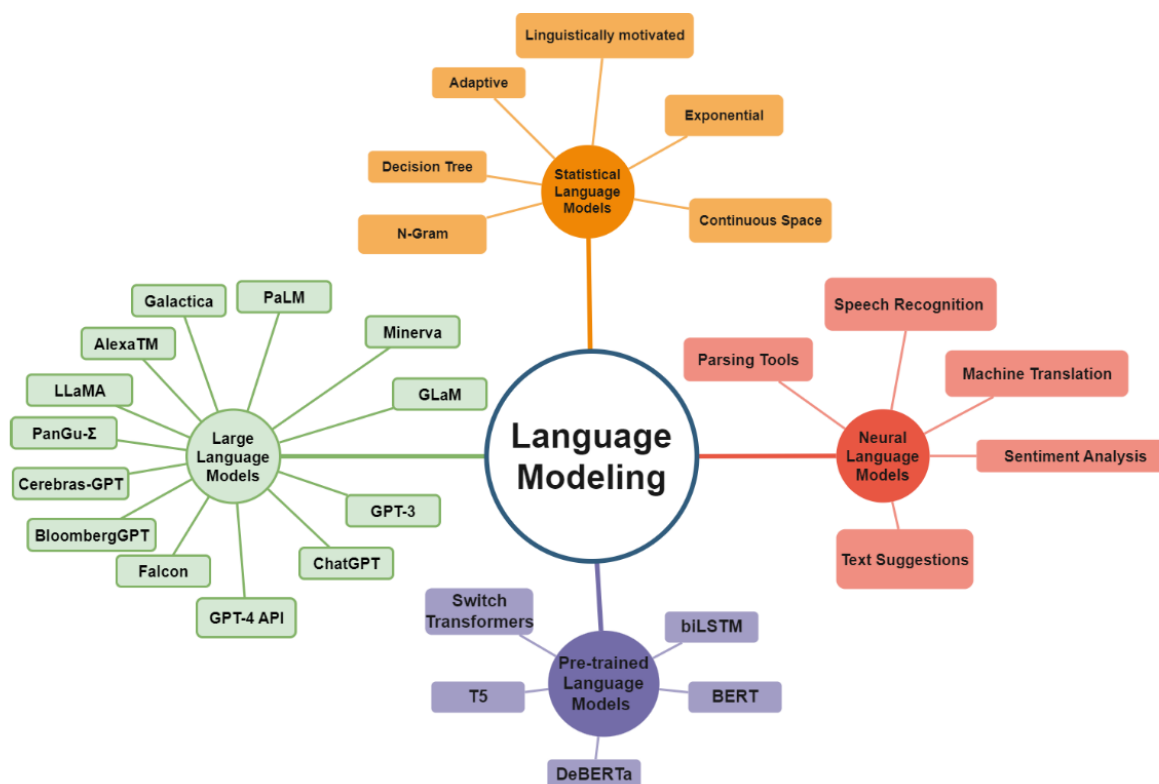


Fig. 1: Types of language modeling.

کاربردهای large language models:

با توجه به دامنه گسترده‌ی کاربردهای مدل‌های زبانی بزرگ (LLMs)، در این بخش، بحثی درباره استفاده آن‌ها در حوزه‌های پزشکی، آموزش، مالی و مهندسی ارائه می‌دهیم.

## پزشکی:

مدل‌های زبانی بزرگ (LLMs) مانند ChatGPT، پتانسیل قابل توجهی را در برنامه‌های متنوع مربوط به حوزه بهداشت، به‌ویژه در زمینه پزشکی، نشان داده‌اند. این مدل‌ها می‌توانند در تشخیص بیماری‌ها، پیش‌بینی نتایج آزمایش‌ها، تحلیل اسناد پزشکی، ارائه راهنمایی‌های درمانی و بسیاری از دیگر مربوط به پزشکی مورد استفاده قرار بگیرند.

medical education-

این مدل‌ها با موفقیت در آموزش پزشکی، تصمیم‌گیری رادیولوژی، ژنتیک بالینی و مراقبت از بیماران مورد استفاده قرار گرفته‌اند، همانطور که توسط چند مطالعه نیز تأیید شده است. قابل ذکر است که عملکرد ChatGPT در آزمون پزشکی لایسنسینگ آمریکا (USMLE) با آستانه قابل قبول مقایسه شده و ترشولد قبولی را رد کرده است، که نشان می‌دهد توانایی آن در دانش پزشکی بدون نیاز به آموزش تخصصی و پیشرفته است. (83)(84)

## - Radiologic decision-making

به گفته مطالعه‌ای که توسط راثو و همکاران [85] انجام شده، انتظار می‌رود در آینده ابزارهای تصمیم‌گیری بالینی مبتنی بر هوش مصنوعی ویژه‌ای ظاهر شود. در این مقاله بر توانایی ChatGPT در تصمیم‌گیری تصویربرداری (radiologic) پزشکی تأکید می‌کند و قابلیت و مزایای بالقوه آن را در بهبود جریان کار بالینی را بیان می‌کند. (82)

## - clinical genetics

در حوزه ژنتیک بالینی، یک مقاله توسط دونگ و سلیمان [86] نشان داد که عملکرد ChatGPT در پاسخ به سوالات مرتبط با ژنتیک به طور قابل توجهی از انسان‌ها متفاوت نبود. با این حال، مدل در مقایسه با سؤالاتی که نیازمند تفکر بحرانی هستند، دقت بهتری در سؤالاتی که نیازمند حفظ کردن بودند نشان داده است. (87)

در حال حاضر، چندین ابزار هوش مصنوعی در حال استفاده وجود دارد که به سیستم امکان تعامل با بیماران را می‌دهد مانند Ada Health، Babylon Health و Buoy Health. شهرت اخیر LLM ها نه تنها می‌تواند اعتماد بیماران را در تعامل با چت‌بات‌های مشابه بهبود بخشد بلکه می‌تواند بهبود خدمات ارائه شده را نیز به همراه داشته باشد. در واقع، ابزارهایی توسعه یافته است که برای کمک به پزشکان توسعه یافته‌اند. یکی از این ابزارها، XrayGPT است که می‌توان از آن برای تحلیل خودکار تصاویر اشعه ایکس استفاده کرد و کاربر/بیمار می‌تواند سؤالاتی درباره تحلیل بپرسد. از طریق گفتگوها، کاربر می‌تواند از طریق یک گفتگوی تعاملی، درکی از وضعیت خود بدست آورد. (92)

مهم ترین چالشی که در زمینه پزشکی گفته شده بود اینکه به دلیل اینکه چت جی پی تی تا سال 2021 بیشتر داده ندیده و به همین دلیل آپدیت نیست و برای اینکه برای کار ما شخصی سازی بشه نیاز است که داده های بیشتری ببیند.

## آموزش

بزرگترین مزیت LLM ها و چت جی پی تی در education این است که می‌تواند به دانش آموزان در راستای تمرین و تکالیف و یادگیری شان کمک کند. (96) همچنین ربات‌های هوش مصنوعی می‌توانند به فرایند امتیازدهی اتوماسیون بخشیده و بار کاری معلمان را کاهش دهند، همچنین امکان ارائه بازخورد دقیق‌تر به دانش‌آموزان را فراهم سازند. با استفاده از الگوریتم‌های هوش مصنوعی و تجزیه و تحلیل زبان طبیعی، ربات‌های هوش مصنوعی می‌توانند به طور خودکار پاسخ‌ها و کارنامه‌های دانش‌آموزان را ارزیابی کنند و امتیازدهی را انجام دهند. این کار، به معلمان امکان می‌دهد تا زمان بیشتری را برای تحلیل عمیق‌تر و بازخورد دقیق‌تر به دانش‌آموزان خود اختصاص دهند. (93)(94)(95)

یکی دیگر از کاربردهای gpt در حوزه آموزش ارائه مسیر آموزش شخصی سازی شده برای هر دانش آموز با توجه به نمرات وی و تحلیل از رفتار اوست. مثل فریم ورک خان اکادمی که یک دستار مجازی برای دانش آموزان فراهم کرده اند. (97)

مهم ترین چالشی که در این قسمت وجود دارد که این ابزارها که به دانش آموزان کمک میکند اما می‌تواند خلاقیت و تفکر آن ها را بکشد و باید در استفاده از این ابزارها به این نکته توجه داشته باشیم. یکی از نگرانی‌های

اصلی، استفاده از این ابزارها بدون آموزش مناسب است. حتی اگر مدل‌های زبانی بزرگ (LLMs) قادر به ارائه پاسخ به تعداد زیادی سوال باشند و کاربر را در ارائه پاسخ به سوالات کمک کنند، اهمیت دارد که دانش‌آموزان در مؤسسات آموزشی به طور کافی آموزش داده شوند تا از قابلیت‌های LLMs به بهترین شکل استفاده کنند.

## حوزه Finance

مدل‌های زبانی بزرگ (LLMs) با پیشرفت‌های قابل توجهی در صنعت مالی در حال پیشروی هستند و کاربردهای آن‌ها از تسک‌های NLP مالی، ارزیابی ریسک، معاملات الگوریتمی، پیش‌بینی بازار را شامل می‌شود. به عنوان مثال ابزاری مثل [27] BloombergGPT، میتواند در تسک‌هایی شامل entity recognition و news classification و Question answering به کاربران کمک کند. علاوه بر این، مدل‌های زبانی بزرگ (LLMs) برای ارزیابی و مدیریت ریسک نیز استفاده می‌شوند. با تحلیل روندها و داده‌های بازار گذشته، قادر به شناسایی ریسک‌های پتانسیلی است و از طریق الگوریتم‌های مالی مختلف، مراحل کاهش ریسک را ارائه می‌دهد. ابزار open source ای نیز به اسم FinGPT به تازگی ساخته شده است (100)

## حوزه مهندسی

جی‌چی‌تی در حوزه مهندسی میتواند کاربردهای مختلفی مانند code generation, debugging, software testing به کار رود مانند چت جی‌پی‌تی که یک مدل زمانی بزرگ است و در زمینه دیباگ و کد نویسی و فیکس کردن باگ میتواند مفید باشد این باعث ذخیره تایم مهندسان خواهد شد (101)

## چت بات ها

چت بات‌ها به طور مکرر در برنامه‌های خدمات مشتریان استفاده می‌شوند، جایی که آن‌ها قادرند به پرسش‌ها پاسخ دهند، کمک کنند و مشکلات را حل کنند. این ابزارها می‌توانند به صورت خودکار و به طور فوری به مشتریان پاسخ دهند و به آن‌ها در رفع مشکلات و ارائه راهنمایی کمک کنند. همچنین میتوانند برای کاربردهای دیگر مثل حوزه سلامت و سرگرمی و آموزش استفاده شوند و LLMهای معروفی که در حال حاضر به عنوان چت بات استفاده میشوند chatgpt, google bard, bing (107) میباشد. نگاه کلی و به صورت مقایسه ای به LLMها تاکنون

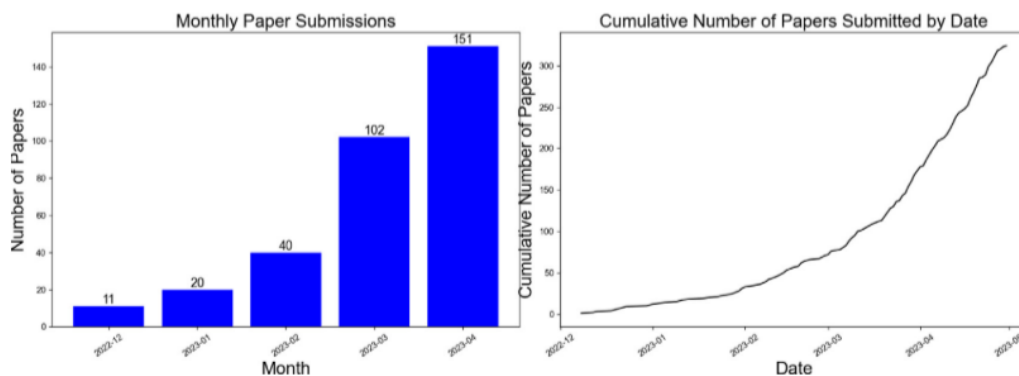
Model	Param.	Type	GSM8K	MATH	MMLU	BBH	HumanEval	C-Eval	TheoremQA
GPT-4	-	RLHF	92.0	42.5	86.4	-	67.0	68.7*	43.4
claude-v1.3	-	RLHF	81.8*	-	74.8*	67.3*	-	54.2*	24.9
PaLM-2	-	Base	80.7	34.3	78.3	78.1	-	-	31.8
GPT-3.5-turbo	-	RLHF	74.9*	-	67.3*	70.1*	48.1	54.4*	30.2
claude-instant	-	RLHF	70.8*	-	-	66.9*	-	45.9*	23.6
text-davinci-003	-	RLHF	-	-	64.6	70.7	-	-	22.8
code-davinci-002	-	Base	66.6	19.1	64.5	73.7	47.0	-	-
text-davinci-002	-	SIFT	55.4	-	60.0	67.2	-	-	16.6
Minerva	540B	SIFT	58.8	33.6	-	-	-	-	-
Flan-PaLM	540B	SIFT	-	-	70.9	66.3	-	-	-
Flan-U-PaLM	540B	SIFT	-	-	69.8	64.9	-	-	-
PaLM	540B	Base	56.9	8.8	62.9	62.0	26.2	-	-
LLaMA	65B	Base	50.9	10.6	63.4	-	23.7	38.8*	-
PaLM	64B	Base	52.4	4.4	49.0	42.3	-	-	-
LLaMA	33B	Base	35.6	7.1	57.8	-	21.7	-	-
InstructCodeT5+	16B	SIFT	-	-	-	-	35.0	-	11.6
StarCoder	15B	Base	8.4	15.1	33.9	-	33.6	-	12.2
Vicuna	13B	SIFT	-	-	-	-	-	-	12.9
LLaMA	13B	Base	17.8	3.9	46.9	-	15.8	-	-
Flan-T5	11B	SIFT	16.1*	-	48.6	41.4	-	-	-
Alpaca	7B	SIFT	-	-	-	-	-	-	13.5
LLaMA	7B	Base	11.0	2.9	35.1	-	10.5	-	-
Flan-T5	3B	SIFT	13.5*	-	45.5	35.2	-	-	-

## Text classification

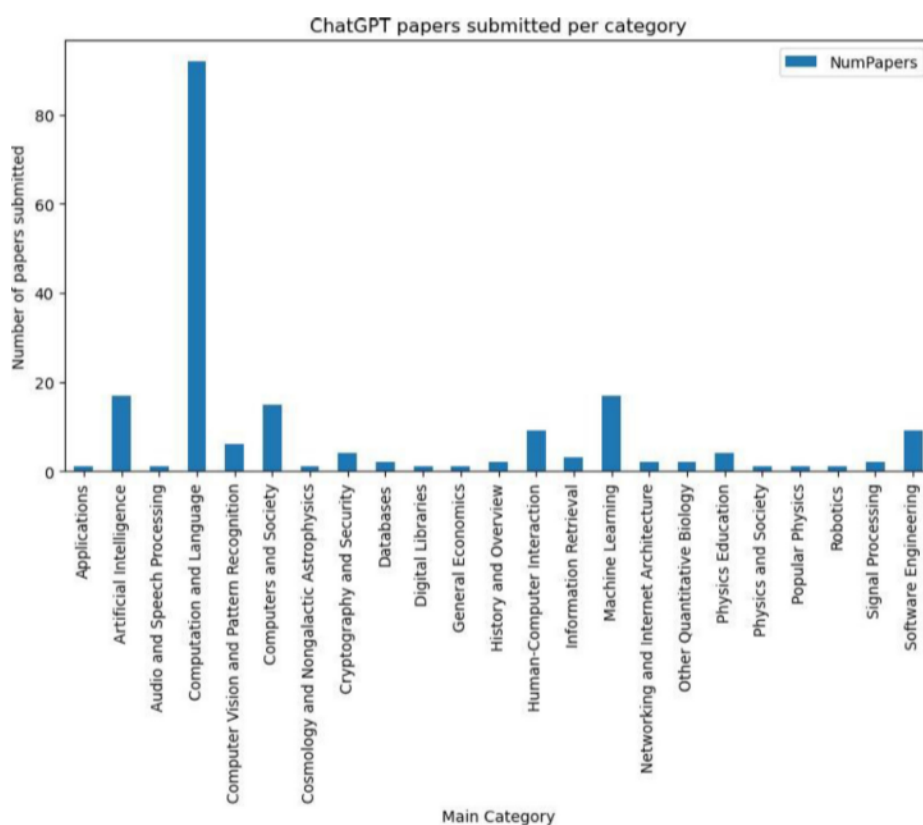
از دیگر کاربرد های gpt استفاده از آن در دسته بندی متن هاست به عنوان مثال تشخیص اسپم از غیر اسپم یا دسته بندی انواع خبرها و..(109)

همان طور که قبل تر هم گفته شد llmها میتوانند در تسک های مختلف Inlp از جمله language summarization, question-answering, translation, text summarization, کاربرد داشته باشد. به ویژه، مدل ChatGPT، پتانسیل خود را در زمینه های مختلف از جمله آموزش، بهداشت، استدلال، تولید متن، تعامل انسان-ماشین و تحقیقات علمی نشان داده است. این مدل قابلیت های چشمگیری را در این حوزه ها از خود نشان داده است.

بیشتر مقالات تا به اکنون در سال 2023 درباره مدل های زبانی منتشر شده اند که نمودارش را هم میتوان به شکل زیر دید



بیشتر مقالاتی که تاکنون منتشر شده اند به تفکیک تعداد و دسته بندی به شکل زیر میباشد



در فیلد پرسش و پاسخ

در حوزه آموزش، ChatGPT به طور معمول برای تست پرسش و پاسخ در بخش آموزش استفاده می‌شود. این مدل قابلیت پاسخگویی به سوالات و بررسی امتحانات را در بخش آموزش فراهم می‌کند. مدل ChatGPT قادر است به یادگیری، مقایسه و تأیید پاسخ‌ها در موضوعات مختلف علمی مانند فیزیک، ریاضیات و شیمی، و یا موضوعات مفهومی مانند فلسفه و دین باشد. این پاراگراف‌ها درباره عملکرد چت‌جی‌پی‌تی در حوزه ریاضیات می‌باشند. به طور خاص، آنها بیان می‌کنند که توانایی‌های ریاضی چت‌جی‌پی‌تی کمتر از دانشجویان فارغ‌التحصیل

ریاضی معمولی است. با این حال، عملکرد چت‌جی‌پی‌تی می‌تواند به طور قابل توجهی بسته به نیازهای شغلی خاص متفاوت باشد.

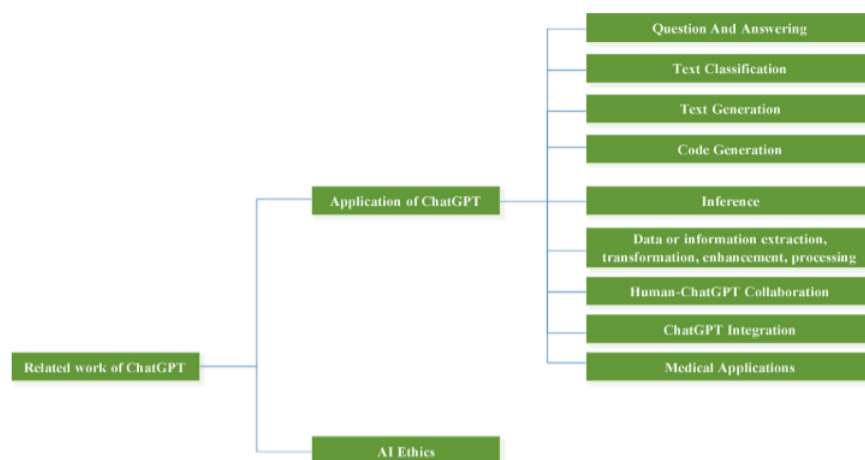
در حل مسائل ریاضی، چت‌جی‌پی‌تی به طور کلی مسائل را درک می‌کند اما قادر به ارائه پاسخ صحیح نمی‌باشد. عملکرد آن با استفاده از مجموعه داده‌های مختلف، از جمله مجموعه داده Grad Text که در مسائل ساده تئوری مجموعه و منطق عملکرد بهتری داشت، مورد ارزیابی قرار گرفت. با این حال، در مجموعه داده‌هایی مانند مسائل حل مسابقه‌های المپیادی و مجموعه داده‌های Holes-in-Proofs، چت‌جی‌پی‌تی نمرات پایین‌تری دریافت کرد که این نکته نشان‌دهنده محدودیت‌های آن در حل مسائل ریاضی پیچیده است.

از سوی دیگر، در زمینه ارائه نکات برای یادگیری جبر، چت‌جی‌پی‌تی نتایج مثبتی نشان داد. نکات تولیدشده توسط چت‌جی‌پی‌تی مفید برای یادگیری بودند که بهبودهای مثبتی در یادگیری نشان دادند. با این حال، نمرات نکات تولیدشده توسط انسان‌ها به طور قابل توجهی بالاتر از نکات تولیدشده توسط چت‌جی‌پی‌تی بودند.

علاوه بر این، در مورد مسائل ریاضی با کلمات، عملکرد چت‌جی‌پی‌تی در طول زمان بهبود یافت. دقت، بازخوانی و نرخ شکست آن از مراحل اولیه به مراحل نهایی بهبود نشان دادند که نشان‌دهنده توانایی آن در تطبیق و بهبود عملکرد با تجربه و آموزش است.

بنابراین، خلاصه کلی نشان می‌دهد که توانایی‌های ریاضی چت‌جی‌پی‌تی ممکن است با فارغ‌التحصیلان ریاضی انسانی برابر نباشد، اما عملکرد آن می‌توانسته به نیازهای شغلی خاص، متفاوت است. ممکن است در مسائل ریاضی ساده مفید باشد یا به عنوان یک ابزار برای ارائه نکات و کمک در فرآیند یادگیری مورد استفاده قرار گیرد. با این حال، برای مسائل پیشرفته و پیچیده، همچنان نیاز به تخصص و مداخله انسانی وجود دارد. (8)(9)(10)

به طور کلی کاربردهای جی‌پی‌تی به دو دسته زیر تقسیم می‌شود که در شکل زیر بیان شده است:



## توی فیلد question answering

در زمینه پزشکی

مطالعات انجام شده در حوزه پزشکی نشان می‌دهد که چت‌جی‌پی‌تی قادر است به سوالات پزشکی بیماران پاسخ دهد و در تشخیص بیماری‌ها به پزشکان کمک کند. در تحقیقات انجام شده، امکانات و محدودیت‌های چت‌جی‌پی‌تی در این زمینه بررسی شده‌اند.

در یک تحقیق، تعاملات بین بیمار و پزشک از سوابق پزشکی الکترونیکی استخراج شد. سوالات بیماران به چت‌جی‌پی‌تی مطرح شد و پاسخ‌های آن با پاسخ‌های پزشکان مقایسه شد. نتایج نشان داد که بیماران توانستند 65.5٪ از پاسخ‌های چت‌جی‌پی‌تی را به درستی تشخیص دهند، در حالی که در مورد پاسخ‌های پزشکان، این عدد 65.1٪ بود. اعتماد به عملکرد چت‌جی‌پی‌تی به طور متوسط مثبت بود و در سوالات سلامت با خطر کمتر، به آن بیشتر اعتماد و در سوالات پزشکی پیچیده، اعتماد بیشتری به پزشکان وجود داشت.

در تحقیق دیگری، توانایی چت‌جی‌پی‌تی در کشف رابطه علی در تشخیص درد عصبی بررسی شد. نتایج تحقیق نشان داد که چت‌جی‌پی‌تی محدودیت‌هایی در درک دانش و مفاهیم جدید خارج از داده‌های آموزش متنی خود داشت. این سیستم اصولاً زبانی را درک می‌کند که برای توصیف وضعیت‌ها استفاده می‌شود و نه دانش زیربنایی. همچنین، چت‌جی‌پی‌تی در پایداری و یکنواختی عملکرد خود موفقیت بالایی نداشت و به یک سوال در پرس‌وجوهای مختلف پاسخ‌های متفاوتی ارائه می‌کرد. با وجود محدودیت‌های چت‌جی‌پی‌تی، امکانات آن در تحقیقات رابطه علی قابل بهبود است.

به طور خلاصه، چت‌جی‌پی‌تی در حوزه پزشکی قابلیت پاسخ به سوالات بیماران و کمک در تشخیص بیماری‌ها را دارد. با این حال، در درک دانش جدید و حفظ یکنواختی عملکرد محدودیت‌هایی وجود دارد. بیماران به طور کلی در موضوعات سلامت با خطر کمتری می‌توانند به چت‌جی‌پی‌تی اعتماد کنند، اما در موضوعات پزشکی پیچیده، اعتماد بیشتری به پزشکان دارند. (15)(16)

## Text\_classification

هدف از طبقه‌بندی متن، نسبت دادن داده‌های متنی به دسته‌بندی‌های پیش‌تعیین شده است. این وظیفه برای برنامه‌های مختلف از جمله تحلیل احساسات، تشخیص اسپم و مدل‌سازی موضوع بسیار حیاتی است. در حالی که الگوریتم‌های یادگیری ماشین سنتی برای طبقه‌بندی متن به طور گسترده استفاده می‌شوند، پیشرفت‌های اخیر در پردازش زبان طبیعی منجر به توسعه تکنیک‌های پیشرفته‌تر شده است. چت‌جی‌پی‌تی در این زمینه پتانسیل بسیاری را نشان داده است. توانایی دقیق در طبقه‌بندی متن، انعطاف‌پذیری در کار با وظایف مختلف طبقه‌بندی و قابلیت سفارشی‌سازی آن، آن را به یک ابزار ارزشمند برای طبقه‌بندی متن تبدیل کرده است، که این موضوع توسط چندین مطالعه در ادبیات تایید شده است.



کوزمان و همکاران (۱۹) از چت‌جی‌پی‌تی برای تشخیص ژانر خودکار استفاده کردند، با هدف ساده‌سازی وظیفه طبقه‌بندی متن با استفاده از قابلیت طبقه‌بندی بدون نیاز به آموزش (zero-shot) چت‌جی‌پی‌تی. آنها عملکرد تشخیص ژانر چت‌جی‌پی‌تی با استفاده از دو زبان انگلیسی (EN) و اسلونیایی (SL) را با استفاده از طبقه‌بند چندزبانه X-GENRE بر پایه مدل چندزبانه XLM-RoBERTa در مجموعه داده انگلیسی EN-GINCO و مجموعه داده اسلونیایی GINCO مقایسه کردند. نتایج نشان داد که زمانی که متن انگلیسی به عنوان زبان مشخصه استفاده شد، چت‌جی‌پی‌تی نمرات F1 میکرو، F1 ماکرو و دقت برابر با 0.74، 0.66 و 0.72 را به دست آورد. با این حال، بر روی مجموعه داده GINCO، عملکرد تشخیص ژانر چت‌جی‌پی‌تی با استفاده از هر دو زبان EN و SL به میزانی کمتر از عملکرد طبقه‌بندی X-GENRE بود.

امین و همکاران (۲۰) توانایی طبقه‌بندی متن چت‌جی‌پی‌تی در محاسبات احساسی را با استفاده از آن برای انجام پیش‌بینی شخصیت، تحلیل احساسات و تشخیص افکار خودکشی ارزیابی کردند. آنها چت‌جی‌پی‌تی را با استفاده از الگوهای مربوطه بر روی سه مجموعه داده: اولین انطباقات، Suicide and Depression و Sentiment140 فعال کردند و عملکرد طبقه‌بندی آن را با سه مدل مبنایی (RoBERTa-base، Word2Vec و BoW) مقایسه کردند. نتایج نشان داد که دقت و UAR (نرخ تشخیص صحیح نمونه‌های مثبت) چت‌جی‌پی‌تی برای پنج طبقه‌بندی شخصیت در مجموعه داده اولین انطباقات به میزان‌های مختلفی کمتر از روش‌های مبنای بود. بر روی مجموعه داده Sentiment140، دقت و UAR چت‌جی‌پی‌تی به ترتیب برابر با 85.5 و 85.5 بود که بهتر از سه روش مبنای بود. در مجموعه داده Suicide and Depression، دقت و UAR چت‌جی‌پی‌تی به ترتیب برابر با 92.7 و 91.2 بود که کمتر از RoBERTa، بهترین روش مبنای، بود.

## Text\_generation

در زمینه تولید متن، زندگی ما در دورانی از انفجار اطلاعات قرار دارد و متن یک راه کار کارآمد برای انتقال اطلاعات است. تنوع اطلاعات منجر به تنوع دسته‌بندی‌های متنی شده است. وقتی پژوهشگران از قابلیت‌های تولید متن چت‌جی‌پی‌تی برای پژوهش استفاده می‌کنند، بدون شک انتخاب می‌کنند که انواع مختلفی از متن تولید کنند. در فرآیند خواندن مقالات، ما متوجه شدیم که تعداد کلمات متن تولید شده توسط پژوهشگران از کوچک به بزرگ افزایش یافته است، بنابراین ما تصمیم گرفتیم براساس تعداد کلمات متن تولید شده، تحقیقات موجود را خلاصه کنیم. ما متن تولید شده را به سه سطح تقسیم کردیم: عبارات، جملات و پاراگراف‌ها.

در مقاله زیر از چت‌جی‌پی‌تی برای تولید عبارات استفاده شده است. زانگ و همکاران (۲۳) اثبات کرده‌اند که مدل HAR دارای معنا با افزودن افزودنی معنایی در طول آموزش عملکرد بهتری در تشخیص حرکت نسبت به سایر مدل‌ها دارد. افزودنی معنایی نیازمند توکن‌های مشترک است که در برخی از مجموعه‌داده‌ها وجود ندارد. بنابراین، نویسندگان از چت‌جی‌پی‌تی برای رویکرد تولید خودکار برچسب برای مجموعه‌داده‌هایی که در ابتدا توکن‌های مشترکی ندارند، استفاده می‌کنند. فو و همکاران (۲۴) یک جریان کار جدید برای تبدیل دستورات زبان طبیعی به دستورات Bash شرح داده‌اند. نویسندگان از چت‌جی‌پی‌تی برای تولید لیستی از دستورات Bash براساس ورودی کاربر استفاده می‌کند، سپس از ترکیبی از تکنیک‌های قاعده‌مند و یادگیری ماشین برای رتبه‌بندی و انتخاب

کاندیدهای احتمالی استفاده می‌کند. این جریان کار بر روی مجموعه داده دستور واقعی ارزیابی شد و در مقایسه با روش‌های پیشرفته دیگر، دقت بالایی را داشت.

## Code generation

تولید کد به معنای تولید خودکار کدهای کامپیوتری از توصیف‌ها یا مشخصات سطح بالا است. قابلیت‌های پردازش زبان طبیعی پیشرفته چت‌جی‌پی‌تی او را قادر می‌سازد تا وظایف تولید کد را انجام دهد. با تحلیل نیازهای تولید کد، چت‌جی‌پی‌تی می‌تواند برنامه‌های کوچکی از کد را تولید کند که عملکرد مورد نظر را به درستی اجرا می‌کند. این علاوه بر صرفه‌جویی در زمان و زحمت نوشتن کد از ابتدا، خطر خطاهایی که در هنگام کدنویسی دستی ممکن است رخ دهد را کاهش می‌دهد. علاوه بر این، توانایی چت‌جی‌پی‌تی در یادگیری و سازگاری با زبان‌های برنامه‌نویسی و چارچوب‌های جدید، او را قادر می‌سازد تا وظایف برنامه‌نویسی پیچیده‌تری را انجام دهد. به عنوان مثال، مگاهد و همکاران طرحی برای تولید کدهای پایتون خودکار با استفاده از چت‌جی‌پی‌تی پیشنهاد کردند. مدل آنها بر روی یک مجموعه داده بزرگ حاوی کدهای پایتون و توصیفات زبان طبیعی متناظر آموزش دید. کدهای تولید شده با استفاده از مجموعه آزمون‌هایی ارزیابی شدند و نتایج نشان داد که این رویکرد دقت بالا و عملکرد خوبی از نظر کیفیت و عملکرد کد داشت.

مثال دیگر کار لینگ و همکاران است که از چت‌جی‌پی‌تی برای تولید پرس و جوهای SQL از سوالات به زبان طبیعی استفاده کردند. مدل آنها بر روی یک مجموعه داده حاوی جفت سوال و پرس و جو آموزش دید و یادگرفت تا پرس و جوهای SQL را تولید کند که به درستی به سوالات مطرح شده پاسخ می‌دهند. این رویکرد بر روی یک مجموعه آزمون ارزیابی شد و نتایج نشان داد که دقت و عملکرد پرس و جوها قابل قبولی دارد.

این مثال‌ها نشان‌دهنده قابلیت چت‌جی‌پی‌تی در وظایف تولید کد هستند. با این حال، مهم است توجه داشت که دربارگیری کد به طور خودکار، چت‌جی‌پی‌تی ممکن است همیشه راه‌حل‌های بهینه یا کارآمدی تولید نکند. بررسی و اصلاح دستی توسط انسان برای اطمینان از اینکه کد تولید شده با مورد نیاز مطلوب سازگار است و از بهترین شیوه‌های برنامه‌نویسی پیروی می‌کند، غالباً ضروری است. (39)

## Information extraction

استخراج اطلاعات (Information Extraction) به معنای استخراج اطلاعات خاص از متن زبان طبیعی برای نمایش ساختاری است. این فرآیند شامل سه زیر وظیفه مهم است: استخراج رابطه بین موجودیت‌ها (Entity Relationship Extraction)، شناسایی موجودیت‌های نامگذاری شده (Named Entity Recognition) و استخراج رویدادها (Event Extraction) که در صنایع تجاری، پزشکی و سایر حوزه‌ها کاربردهای گسترده‌ای دارند.

در استخراج اطلاعات، Wei و همکاران (۵۲) چارچوب ChatIE را ارائه دادند. ChatIE یک چارچوب سوال و پاسخ چند دوره مبتنی بر ChatGPT برای استخراج اطلاعات است. این چارچوب یک وظیفه پیچیده استخراج اطلاعات را

به چند بخش تقسیم کرده و سپس نتایج هر دور را در یک نتیجه ساختاری نهایی ترکیب می‌کند. وظایف استخراج سه‌گانه ترکیب سه گانه ارتباطی موجودیت، شناسایی موجودیت‌های نامگذاری شده و استخراج رویدادها روی شش مجموعه داده NYT11-HRL، DuEE2.0، conllpp، MSR، DuEE1.0، ACE05 به زبان‌های مختلف انجام شد و با سه معیار دقت، بازخوانی و امتیاز F1 مقایسه شد. نتایج نشان می‌دهد که در شش مجموعه داده رایج استخراج اطلاعات، ChatIE با میانگین بهبود ۱۸.۹۸٪ نسبت به ChatGPT اصلی بدون ChatIE، عملکرد را بهبود می‌بخشد و در مجموعه داده NYT11-HRL عملکرد بهتری نسبت به مدل‌های نظارت شده FCM و MultiR دارد.

Gao و همکاران (۶۰) قابلیت ChatGPT در استخراج رویداد را در مجموعه داده ACE2005 بررسی کردند و عملکرد ChatGPT را در سناریوهای پیچیده و کمیاب (متن‌هایی که حاوی چند رویداد هستند) با دو مدل وظیفه‌محور Text2Event و EEQA مقایسه کردند. سپس تأثیر راهبردهای مختلف بر عملکرد ChatGPT بررسی شد. نتایج نشان می‌دهد که عملکرد متوسط ChatGPT در سناریوهای پیچیده و کمیاب تنها ۵۱.۰۴٪ عملکرد مدل‌های وظیفه‌محور مانند EEQA است. بهبود مداوم راهبردها منجر به بهبود یکنواالی عملکرد نمی‌شود و ChatGPT بسیار حساس به سبک‌های راهبردی مختلف است.

Tang و همکاران (۶۳) یک پارادایم آموزش جدید را ارائه کردند که شامل راهنمایی مناسب برای هدایت ChatGPT به تولید نمونه‌های مختلف با ساختار جمله و الگوهای زبانی متفاوت است و نمونه‌های با کیفیت پایین یا تکراری حاصل از آن را برای وظایف پایین‌جریان حذف می‌کند. اگرچه در مقایسه با یک مدل نرم برای یک وظیفه بهداشتی خاص، ChatGPT در وظایف شناسایی موجودیت‌های نامگذاری شده و استخراج رابطه از عملکرد ضعیف‌تری برخوردار است، اما در وظیفه استخراج رابطه از مجموعه داده Gene Association Database Release و EU-ADR، چارچوب آموزش نوآورانه قادر به آموزش مدل‌های محلی است و امتیاز F1 را از ۲۳.۳۷٪ به ۶۳.۹۹٪ برای وظیفه شناسایی موجودیت‌های نامگذاری شده و از ۷۵٪ به ۸۲٪ برای وظیفه استخراج رابطه بهبود می‌بخشد، همچنین نگرانی‌های حریم خصوصی و مشکلات جمع‌آوری و برچسب‌گذاری داده‌های زمان‌بر را کاهش می‌دهد.

He و همکاران (۶۴) چارچوب یادگیری زمینه‌ای ICL-D3IE را ارائه دادند. این چارچوب ارائه فرمت شده را معرفی کرده، به صورت پیوسته تکرار می‌کند تا ارائه را به‌روزرسانی و بهبود دهد و سپس ChatGPT را برای استخراج اطلاعات متنی استفاده می‌کند. در این مقاله، ICL-D3IE با مدل‌های پیش‌آموزش داده شده موجود مانند LiLT و BROS (در تنظیمات توزیع‌شده درون‌دامنه و برون‌دامنه) بر روی مجموعه داده‌های CORD، FUNSD، SROIE و مقایسه شد. نتایج نشان می‌دهد که روش ICL-D3IE در همه مجموعه داده‌ها و تنظیمات به جز تنظیم درون‌دامنه بر روی CORD بهتر از روش‌های دیگر است.

#### Outdated knowledge-1

مدل های فعلی بر روی داده های تاریخی (تا سال 2021) آموزش داده شده اند و به همین دلیل توانایی درک زمان واقعی از رویدادهای کنونی را ندارند. این یک نگرانی جدی در دوران انفجار اطلاعات امروزی است، زیرا قابلیت اعتماد به پایگاه دانش های قبلی به تدریج کاهش می یابد و این موضوع ممکن است منجر به پاسخ های نادرست، به ویژه در زمینه هایی که به سرعت در حال تحول هستند مانند قانون و فناوری، شود. علاوه بر این، این مدل ها قادر به بررسی صحت ادعاها نیستند در حالی که داده های آموزشی شامل محتوایی از منابع مختلف است که برخی از آنها ممکن است غیرقابل اعتماد باشند و این موضوع ممکن است منجر به پاسخ های به نظر معقول اما بی معنی شود.

#### Insufficient understanding -2

هنگام پرداختن به سؤالات مبهم یا پیچیده ی متن، مدل ها ممکن است با چالش های درکی روبرو شوند. علاوه بر این، در برخی حوزه های تخصصی، وجود فراوانی از اختصارات منحصر به فرد، چالش های درک مدل ها را تشدید می کند و باعث پاسخ های نادرست و بی معنی می شود.

#### 4.1.3 Energy consumption

در طول مراحل آموزش و استنتاج، این مدل های بزرگ نیاز به منابع محاسباتی و منابع برق قابل توجهی دارند که منجر به مصرف انرژی بالا و انتشار گازهای گلخانه ای قابل توجه می شود. این موضوع باعث محدود شدن استقرار و کاربردهای عملی این مدل ها می شود.

#### 4.1.4 Malicious usage

اگرچه OpenAI مجموعه ای از محدودیت ها را برای کاهش سمیت مدل اجرا کرده است، اما مواردی از استفاده کنندگان وجود دارد که با استفاده از دستوراتی با دقت طراحی شده، تلاش می کنند تا از این محدودیت ها خارج شوند و مدل را مجبور به تولید محتوای ناسالم یا حتی استفاده غیرقانونی تجاری کنند.

#### 4.1.5 Bias and discrimination

به دلیل تأثیر داده های پیش آموزش، مدل ها در زمینه های سیاسی، ایدئولوژیکی و سایر حوزه ها تعصباتی نشان می دهند. به همین دلیل، استفاده از مدل های زبانی در حوزه های عمومی، مانند آموزش و تبلیغات، باید با احتیاط بسیار بیشتری انجام شود. این موضوع به این معنی است که نیاز است با دقت و هوشمندی فراوان به استفاده از این مدل ها در این حوزه ها نگرانی شود و تأثیرات تعصباتی آنها در نظر گرفته شود..

#### 4.1.6 Privacy and data security

همزمان با افزایش تعداد کاربران، حفاظت از حریم خصوصی کاربران و امنیت داده ها به مراتب مهم تر می شود. در واقع، در اوایل آوریل، به دلیل نگرانی های حریم خصوصی، ChatGPT در ایتالیا ممنوع شد. این موضوع به ویژه مهم است زیرا مدل ها در طول تعاملات، اطلاعات شخصی و ترجیحات شخصی را جمع آوری می کنند و در آینده،

مدل‌های چندحالتی مانند GPT-4 ممکن است از کاربران خواسته شود تا عکس‌های خصوصی خود را بارگذاری کنند.

#### References:

- Summary of ChatGPT-Related research and perspective towards the future of large language models
- Frieder Simon, Pinchetti Luca, Griffiths Ryan-Rhys, et al. Mathematical capabilities of chatgpt. arXiv preprint arXiv:2301.13867. 2023
- . Pardos Zachary A, Bhandari Shreya. Learning gain differences between chatgpt and human tutor generated algebra hints. arXiv preprint arXiv:2302.06871. 2023.
- . Shakarian Paulo, Koyyalamudi Abhinav, Noel Ngu, Mareedu Lakshmivihari. An independent evaluation of chatgpt on mathematical word problems (mwp). arXiv preprint arXiv:2302.13814. 2023
- Tu Ruibo, Ma Chao, Zhang Cheng. Causal-discovery performance of chatgpt in the context of neuropathic pain diagnosis. arXiv preprint arXiv:2301.13819. 2023.
- Clark Glymour, Zhang Kun, Spirtes Peter. Review of causal discovery methods based on graphical models. Front Genet. 2019;10:524.
- Kuzman Taja, Mozetic Igor, Ljubešić Nikola. Chatgpt: beginning of an end of manual linguistic data annotation? use case of automatic genre identification. arXiv e-prints. 2023:2303.
- Zhang Xiyuan, Chowdhury Ranak Roy, Hong Dezhi, Gupta Rajesh K, Shang Jingbo. Modeling label semantics improves activity recognition. arXiv preprint arXiv: 2301.03462. 2023
- Tang Ruixiang, Han Xiaotian, Jiang Xiaoqian, Hu Xia. Does synthetic data generation of llms help clinical text mining? arXiv preprint arXiv:2303.04360. 2023.
- A Survey on Large Language Models: Applications, Challenges, Limitations, and Practical Usage 2023

- Y. K. Dwivedi, N. Kshetri, L. Hughes, E. L. Slade, A. Jeyaraj, A. K. Kar, A. M. Baabdullah, A. Koohang, V. Raghavan, M. Ahuja, et al., "“so what if chatgpt wrote it?” multidisciplinary perspectives on opportunities, challenges and implications of generative conversational ai for research, practice and policy,” *International Journal of Information Management*, vol. 71, p. 102642, 2023.
- M. Sallam, “Chatgpt utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns,” in *Healthcare*, vol. 11, p. 887, MDPI, 2023.
- A. Gilson, C. W. Safranek, T. Huang, V. Socrates, L. Chi, R. A. Taylor, D. Chartash, et al., “How does chatgpt perform on the united states medical licensing examination? the implications of large language models for medical education and knowledge assessment,” *JMIR Medical Education*, vol. 9, no. 1, p. e45312, 2023
- mbzuai oryx, “Xraygpt: Chest radiographs summarization using medical vision-language models,” 2023.
- J. S. () and W. Y. (), “Unlocking the power of chatgpt: A framework for applying generative ai in education,” *ECNU Review of Education*, vol. 0, no. 0, p. 20965311231168423, 0
- M. Fraiwan and N. Khasawneh, “A review of chatgpt applications in education, marketing, software engineering, and healthcare: Benefits, drawbacks, and research directions,” *arXiv preprint arXiv:2305.00237*, 2023.
- S. Frieder, L. Pinchetti, R.-R. Griffiths, T. Salvatori, T. Lukasiewicz, P. C. Petersen, A. Chevalier, and J. Berner, “Mathematical capabilities of chatgpt,” *arXiv preprint arXiv:2301.13867*, year=2023.
- A literature review on question answering techniques, paradigms and systems
- Transformers in the Real World: A Survey on NLP Applications

