Journal of Recent Advancements in Material Science and Biomedical Engineering
Vol. 1, Issue 4, Spring 2026, 32-53
ram.apadana.ac.ir
Apadana Institute of Higher Education

Promising Prospects for Diagnosis and Treatment in Supporting Physicians through Use of Artificial Intelligence based on Large Language Models (LLMs)

Solmaz Farajzadeh*
PhD in Human Resource Management, Apadana Institute of Higher Education, Shiraz, Iran.

Received: 23/December/2026 Accepted: 19/February/2026 eISSN: 3115-7610 ISSN: 3115-7572

Abstract

At the heart of medicine lies the doctor-patient conversation, where skillful history taking paves the way for accurate diagnosis, effective management, and lasting trust. AI systems capable of diagnostic conversation can increase access, consistency, and quality of care. In this paper, we present an AI system based on large language models optimized for diagnostic conversation. Large language models such as ChatGPT, Claude, Llama, and Qwen are emerging as transformative technologies for the diagnosis and treatment of various diseases. With exceptional reasoning capabilities over long contexts, large language models are adept at relevant clinical tasks, especially in medical text analysis and interactive conversation. They can increase diagnostic accuracy by processing large volumes of patient data and medical texts, and have demonstrated their utility in diagnosing common diseases and facilitating the identification of rare diseases by recognizing subtle patterns in symptoms and test results. Relying on their image recognition capabilities, multimodal large-language models show promising potential for diagnosis based on radiographs, chest computed tomography, electrocardiography, and common pathological images. These models can also aid in treatment planning by suggesting evidence-based interventions and improving clinical decision support systems through integrated analysis of patient records. Despite these promising advances, there are significant challenges to the use of large-language models in medicine, including concerns about algorithmic bias, the possibility of illusion, and the need for rigorous clinical validation. Ethical considerations also emphasize the importance of maintaining monitoring performance in clinical practice. This article highlights the rapid advances in research on diagnostic and therapeutic applications of large-language models across various medical disciplines and emphasizes the importance of policymaking, ethical oversight, and multidisciplinary collaboration in promoting more effective and safer clinical applications of large-language models. Future directions include integrating proprietary clinical knowledge, exploring open-source and custom models, and evaluating real-time effects on clinical diagnosis and treatment practices.

Keywords: diagnosis & treatment, artificial intelligence, large language models

* Corresponding Author: soolmaz.farajzadeh@gmail.com

فصلنامه پیشرفت‌های مهندسی در حوزه‌ی پزشکی و مواد
دوره ۱، شماره ۴، بهار ۱۴۰۵، ۳۲-۵۳
ram.apadana.ac.ir

تاریخ دریافت: ۱۴۰۵/۰۱/۱۵ تاریخ پذیرش: ۱۴۰۵/۰۲/۲۰ eISSN: 3115-7610 ISSN: 3115-7572

چشم‌اندازهای امیدبخش تشخیص و درمان بیماری‌ها در کمک به پزشکان از طریق به‌کارگیری هوش مصنوعی مبتنی بر مدل‌های زبان بزرگ (LLMs)

سولماز فرج‌زاده* | دکتری مدیریت منابع انسانی، مؤسسه آموزش عالی آپادانا، شیراز، ایران.

چکیده

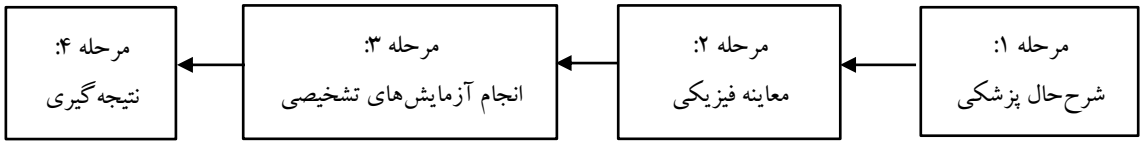
در قلب پزشکی، گفت‌وگوی پزشک و بیمار قرار دارد، جایی که گرفتن شرح‌حال ماهرانه، راه را برای تشخیص دقیق، مدیریت مؤثر و اعتماد پایدار هموار می‌کند. سیستم‌های هوش مصنوعی که قادر به گفت‌وگوی تشخیصی هستند، می‌توانند دسترسی، ثبات و کیفیت مراقبت را افزایش دهند. در مقاله حاضر سیستم هوش مصنوعی مبتنی بر مدل‌های زبان بزرگ (LLM) را که برای گفت‌وگوی تشخیصی بهینه شده است، معرفی می‌شود. مدل‌های زبان بزرگ مانند ChatGPT، Claude، Llama و Qwen به‌عنوان فناوری‌های متحول‌کننده به‌منظور تشخیص و درمان بیماری‌های مختلف در حال ظهور هستند. مدل زبان بزرگ با قابلیت‌های استثنایی استدلال در زمینه‌های طولانی، در وظایف بالینی مرتبط، به‌ویژه در تجزیه‌وتحلیل متن پزشکی و گفت‌وگوی تعاملی، مهارت دارند. آن‌ها می‌توانند با پردازش حجم زیادی از داده‌های بیمار و متون پزشکی، دقت تشخیصی را افزایش داده و کاربرد خود را در تشخیص بیماری‌های شایع و تسهیل شناسایی بیماری‌های نادر با تشخیص الگوهای ظریف در علائم و نتایج آزمایش، نشان دهند. مدل‌های زبان بزرگ چندوجهی (MLLMs) با تکیه بر توانایی‌های تشخیص تصویر خود، پتانسیل امیدوارکننده‌ای به‌منظور تشخیص براساس رادیوگرافی، توموگرافی کامپیوتری قفسه سینه، الکتروکاردیوگرافی و تصاویر پاتولوژیک رایج نشان می‌دهند. همچنین این مدل‌ها می‌توانند با پیشنهاد مداخلات مبتنی بر شواهد و بهبود سیستم‌های پشتیبانی تصمیم‌گیری بالینی از طریق تجزیه‌وتحلیل یکپارچه سوابق بیمار، در برنامه‌ریزی درمان کمک کنند. علی‌رغم این پیشرفت‌های امیدوارکننده، چالش‌های قابل توجهی در مورد استفاده از مدل‌های زبان بزرگ در پزشکی وجود دارد، ازجمله نگرانی‌هایی در مورد سوگیری الگوریتمی، احتمال توهم و نیاز به اعتبارسنجی بالینی دقیق. ملاحظات اخلاقی همچنین بر اهمیت حفظ عملکرد نظارت در عمل بالینی تأکید می‌کند. این مقاله پیشرفت‌های سریع در تحقیقات در مورد کاربردهای تشخیصی و درمانی مدل‌های زبان بزرگ در رشته‌های مختلف پزشکی را برجسته می‌کند و بر اهمیت سیاست‌گذاری، نظارت اخلاقی و همکاری چندرشته‌ای در ترویج کاربردهای بالینی مؤثرتر و ایمن‌تر مدل‌های زبان بزرگ تأکید دارد. مسیرهای آینده شامل ادغام دانش بالینی اختصاصی، بررسی مدل‌های متن‌باز و سفارشی و ارزیابی اثرات بلادرنگ در تشخیص بالینی و شیوه‌های درمانی است.

کلیدواژه‌ها: تشخیص و درمان، هوش مصنوعی، مدل‌های زبان بزرگ

* نویسندهٔ مسئول: soolmaz.farajzadeh@gmail.com

مقدمه

مطالعه تشخیص بیماری برای صنعت مراقبت‌های بهداشتی ضروری است. بیماری، هر وضعیت یا مجموعه‌ای از شرایط است که منجر به ناراحتی، ناخوشی، نقص عملکرد یا در نهایت مرگ در فرد می‌شود. بیماری‌ها می‌توانند بر سلامت جسمی و روانی فرد تأثیر گذارند و شیوه زندگی او را به‌طور قابل توجهی تغییر دهند. اصطلاح «فرایند پاتولوژیک» به مطالعه علت بیماری اشاره دارد. یک بیماری ناشی از نشانه‌ها یا علائمی است که متخصصان پزشکی تفسیر می‌کنند. فرایند تعیین پاتوفیزیولوژی یک بیماری از طریق شناسایی علائم و نشانه‌های آن، به‌عنوان تشخیص شناخته می‌شود.



شکل ۱. نمودار بلوکی فرایند تشخیص بیماری

همان‌طور که در شکل ۱ مشاهده می‌شود، تشخیص و فرایند تعیین نوع بیماری یک فرد براساس علائم و نشانه‌های اوست. دانش مورد نیاز برای تشخیص از معاینه فیزیکی و شرح‌حال پزشکی بیمار مبتلا به بیماری گرفته می‌شود. اغلب، در طول این درمان، حداقل یک روش تشخیصی مانند آزمایش پزشکی انجام می‌شود [۱].

یک تشخیص معتبر توسط یک پزشک متخصص از طریق فرایندی است که وی را قادر می‌سازد تا حد امکان داده‌های مناسب جمع‌آوری کند. توانایی متخصص مراقبت‌های پزشکی در تشخیص صحیح یک بیماری، برای مراقبت از بیماران بسیار حیاتی است، اما همین توانایی، این کار را به چالش‌برانگیزترین وظیفه تبدیل می‌کند. روش تشخیص ممکن است بسیار پر زحمت و پیچیده باشد. پزشکان متخصص، شواهد تجربی را برای تعیین نوع بیماری فرد بیمار جمع‌آوری می‌کنند تا عدم‌قطعیت در تشخیص پزشکی را کاهش دهند. اگر روش تشخیص ناقص باشد، ممکن است بیمار درمان مناسب را دریافت نکند یا درصورت داشتن مشکلات عمده سلامتی، درمان به تعویق بیفتد [۲].

ازاین‌رو، گفت‌وگوی بین پزشک و بیمار به‌منظور مراقبت مؤثر و دل‌سوزانه امری مهم و اساسی است. مصاحبه پزشکی، قدرتمندترین، حساس‌ترین و پرکاربردترین ابزار موجود برای پزشک نامیده شده است. اعتقاد بر این است که ۶۰ تا ۸۰ درصد تشخیص‌ها صرفاً از طریق گرفتن شرح‌حال بالینی انجام می‌شود. گفت‌وگوی پزشک و بیمار فراتر از گرفتن شرح‌حال و تشخیص است – این تعاملی پیچیده است که باعث ایجاد تفاهم و اعتماد می‌شود و به‌عنوان ابزاری برای رسیدگی به نیازهای بهداشتی عمل می‌کند و بیمار را قادر می‌سازد تا تصمیمات آگاهانه‌ای اتخاذ کند که ترجیحات، انتظارات و نگرانی‌های وی را در نظر گیرد [۳]. درحالی‌که تنوع زیادی در مهارت‌های ارتباطی بین پزشکان وجود دارد، متخصصان آموزش دیده می‌توانند مهارت‌های قابل توجهی در گرفتن شرح‌حال بالینی و «گفت‌وگوی تشخیصی» گسترده‌تر داشته باشند. بااین‌حال، دسترسی به این تخصص همچنان پراکنده و در سطح جهانی کمیاب است[۴].

با توجه به اهمیت موضوع فوق، در این شرایط وجود یک سیستم تشخیص خودکار ضرورت می‌یابد که از ویژگی‌های دقت کامپیوتر و همچنین دانش انسانی خوبی برخوردار باشد. بنابراین، برای به‌دست آوردن نتایج دقیق تشخیص با هزینه کمتر، یک سیستم پشتیبانی تصمیم‌گیری مناسب مورد نیاز است. برای انسان‌ها، طبقه‌بندی بیماری‌ها براساس معیارهای مختلف کاری دشوار می‌باشد. اما هوش مصنوعی[1] در شناسایی و مدیریت این نوع موقعیت‌ها به متخصصان کمک خواهد کرد. به‌کارگیری هوش مصنوعی در مراقبت‌های بهداشتی، به‌ویژه در زمینه‌های پیشگیری و تشخیص بیماری‌ها، در حال ایجاد یک تغییر الگویی بنیادین است. امروزه هوش مصنوعی نقش مهمی در شناسایی

1. Artificial intelligence (AI)

بیماری‌ها ایفا می‌کند، ازاین‌رو سرعت، دقت و اثربخشی روش‌های تشخیصی متحول شده است.

همان‌طور که راسل و نورویگ[1] در سال ۱۹۹۵ استدلال کردند، روش‌های مختلفی به‌منظور تعریف هوش مصنوعی وجود دارد. وجه مشترک آن‌ها این است که هوش مصنوعی الگوریتم‌هایی را توصیف می‌کند که به‌طور مصنوعی فرایندهای فکری، شناختی و رفتاری انسان را تقلید می‌کنند و در برنامه‌های نرم‌افزاری نمونه‌سازی می‌شوند. از آن زمان تا به امروز، تعداد تعاریف با افزایش تعداد برنامه‌های هوش مصنوعی افزایش یافته است [۵]. چندین درک خاص از هوش مصنوعی وجود دارد، مانند آنچه توسط دی بروین[2] و همکارانش ارائه شده است که هوش مصنوعی را به‌عنوان نرم‌افزاری تعریف می‌کنند که می‌تواند «به‌طور مستقل ساختارهای دانش جدیدی تولید نماید» [۶]. رویکردهای کلی‌تر، هوش مصنوعی ضعیف، هوش مصنوعی قوی و هوش مصنوعی عمومی (AGI)[3] را توصیف و از یکدیگر متمایز می‌کنند. اصطلاح هوش مصنوعی ضعیف که توسط جان سرل[4] در سال ۱۹۸۰ ابداع شد، نرم‌افزاری را توصیف می‌کند که با تقلید از فرایندهای شناختی خاص انسان مانند تشخیص تصویر یا پردازش زبان طبیعی، هوشمند به نظر می‌رسد. هوش مصنوعی قوی به نرم‌افزاری اشاره دارد که واقعاً بدون تقلید از آن، هوشمند است. هوش مصنوعی عمومی (AGI) به‌عنوان بسط این اصطلاحات، هوش واقعی را برای همه فرایندهای شناختی انسان و نه‌فقط برای وظایف فردی، تعیین می‌کند [۷]. در مقاله حاضر، هنگام صحبت در مورد عوامل مکالمه‌ای[5] مبتنی بر هوش مصنوعی، درک هوش مصنوعی ضعیف اتخاذ می‌شود. الگوریتم‌های پیاده‌سازی شده در نرم‌افزار عامل مکالمه‌ای، هرکدام فرایندهای شناختی و انسانی متمایز و محدودی را تقلید می‌کنند.

آخرین پیشرفت‌ها در هوش مصنوعی، امکان تعاملات طبیعی فزاینده‌ای را بین انسان‌ها و همتایان عامل ماشینی آن ها فراهم کرده است. این ارتباط شبیه‌سازی شده انسان و ماشین، به‌ویژه از طریق پیشرفت در یادگیری ماشین با کاربرد شبکه‌های عصبی، پیچیده‌تر و پیچیده‌تر می‌شود [۸]. این امر در افزایش تعداد عوامل مکالمه‌ای که تبادلات شبه انسانی را هدف قرار می‌دهند در زمینه‌هایی مانند تجارت الکترونیک، سفر، گردشگری و مراقبت‌های بهداشتی منعکس شده است. نمونه‌های شناخته‌شده چنین چت‌بات‌های هوشمندی عبارت‌اند از کورتانا مایکروسافت[6]، الکسا آمازون[7] یا سیری اپل[8] [۹].

تمرکز بر رابطه انسان و ماشین از همان ابتدا در تاریخ چت‌بات‌ها وجود داشت؛ برنامه نرم‌افزاری مبتنی بر قانون ELIZA[9] به‌منظور ایفای نقش یک روان‌درمانگر طراحی شده بود تا تبادل روان‌درمانی راجری بیمارمحور را تقلید کند. این برنامه در سال ۱۹۶۶ توسط جوزف وایزنباوم[10] توسعه یافت و پس از آن PARRY، یکی دیگر از چت‌بات‌های مرتبط با مراقبت‌های سلامت روان که در سال ۱۹۷۲ توسعه یافت و روانه بازار شد [۱۰]. درحالی‌که ELIZA نقش درمانگر را بازی می‌کرد، PARRY نقش یک بیمار اسکیزوفرنی را بر عهده داشت. اگرچه ELIZA یک آزمون تورینگ[11] محدود -آزمون هوش ماشین با معیار موفقیت اینکه آیا انسان می‌تواند در طول مکالمه، یک ماشین را از یک انسان تشخیص دهد - را با موفقیت پشت‌سر گذاشت، اما یک برنامه نرم‌افزاری مبتنی بر قانون و از پیش نوشته شده بود. به‌طور مشابه، سایر اشکال اولیه ربات‌های چت که در آن زمان چت‌بات نامیده می‌شدند، مانند

1. Russel and Norvig
2. De Bruyn
3. Artificial general intelligence (AGI)
4. John Searle
5. Conversational agents (CA)
6. Microsoft’s Cortana
7. Amazon’s Alexa
8. Apple’s Siri
9. The rule-based software program ELIZA
10. Joseph Weizenbaum
11. turing test

Psyxpert، یک سیستم خبره به‌منظور پشتیبانی از تشخیص بیماری که با زبان Prolog نوشته شده بود یا SESAM-DIABETE، یک سیستم خبره برای آموزش بیماران دیابتی که با زبان Lisp نوشته شده بود، از رویکردی مبتنی بر قانون پیروی می‌کردند. نهاد کامپیوتری اینترنتی زبان‌شناسی مصنوعی (ALICE)[1]، در سال ۱۹۹۵، اولین سیستم کامپیوتری بود که از پردازش زبان طبیعی برای تفسیر ورودی کاربر استفاده کرد [۱۱].

از آن زمان، دسترسی و ذخیره‌سازی کارآمدتر داده‌ها، کاهش هزینه‌های سخت‌افزاری و دسترسی آسان‌تر به خدمات مبتنی بر ابر، توسعه معماری هوش مصنوعی را بهبود بخشید. این پیشرفت‌ها منجر به استقرار استانداردتر پردازش زبان طبیعی، تشخیص صدا، تولید زبان طبیعی و موارد مشابه در توسعه چت‌بات شد [۱۲].

در مراقبت‌های بهداشتی، چنین عوامل مکالمه‌ای مبتنی بر هوش مصنوعی مزایای متعددی را برای تشخیص بیماری، نظارت یا پشتیبانی از درمان در دو دهه گذشته نشان داده‌اند. آن‌ها به‌عنوان مداخلات دیجیتالی برای ارائه راه‌حل‌های پشتیبانی پزشکی مقرون‌به‌صرفه، مقیاس‌پذیر و شخصی‌سازی‌شده استفاده می‌شوند که می‌توانند در هر زمان و هر مکانی از طریق برنامه‌های مبتنی بر وب یا تلفن همراه ارائه شوند [۱۳]. مطالعات تحقیقاتی، انواع مختلفی از عوامل مکالمه‌ای مبتنی بر هوش مصنوعی را به‌منظور کاربردهای مختلف مراقبت‌های بهداشتی مانند ارائه اطلاعات به بیماران سرطان پستان، ارائه اطلاعات در مورد رابطه جنسی، مواد مخدر و الکل به نوجوانان؛ خودارزیابی برای بیماران درمانی، کمک به مربیگری سلامت برای ترویج سبک زندگی سالم؛ یا ترک سیگار بررسی کرده‌اند [۱۴].

پیشرفت‌های اخیر در مدل‌های زبانی بزرگ به‌دلیل توانایی آن‌ها در درک زبان طبیعی و تولید محتوای با کیفیت بالا، توجه زیادی را در حوزه مراقبت‌های بهداشتی به خود جلب کرده است. مدل‌های زبانی بزرگ به‌طور فزاینده‌ای در عوامل مکالمه‌ای ادغام می‌شوند تا مشاوره‌های پزشکی و مداخلات بهداشتی را بهبود بخشند [۱۵]. عوامل مکالمه‌ای، که سیستم‌های نرم‌افزاری طراحی شده برای تعامل با انسان‌ها از طریق زبان طبیعی هستند، نقش مهمی در کاربردهای بهداشتی، مانند ارائه روان‌درمانی و خدمات مراقبت‌های بهداشتی از راه دور، ارائه اطلاعات آموزش سلامت، انجام مصاحبه‌های بالینی، کمک به تشخیص علائم و ارتقای رفتارهای بهداشتی ایفا می‌نمایند. با رواج بیشتر این فناوری‌های پیشرفته در مراقبت‌های بهداشتی و افزایش استفاده از اطلاعات شخصی، محققان توجه بیشتری را به ارزیابی مزایای بالقوه آن‌ها و همچنین پرداختن به چالش‌های نوظهور مانند حریم خصوصی و امنیت داده‌ها معطوف کرده‌اند [۱۶].

سیستم‌های هوش مصنوعی توانایی برنامه‌ریزی، استدلال و گنجاندن زمینه‌های مرتبط به‌اندازه کافی برای برگزاری مکالمات طبیعی را دارند. این پیشرفت، فرصتی را برای بازنگری در پیشرفت‌های اخیر در مدل‌های زبان بزرگ[2] (LLMs) و احتمالات هوش مصنوعی در پزشکی به‌سمت توسعه هوش مصنوعی مکالمه‌ای کاملاً تعاملی فراهم می‌کند. چنین سیستم‌های هوش مصنوعی پزشکی، زبان بالینی را درک می‌کنند، هوشمندانه اطلاعات را در شرایط عدم‌قطعیت به‌دست می‌آورند و در مکالمات پزشکی طبیعی و ازنظر تشخیصی مفید با بیماران و کسانی که از آن‌ها مراقبت می‌کنند، شرکت می‌کنند. کاربرد بالقوه سیستم‌های هوش مصنوعی که قادر به گفت‌وگوی بالینی و تشخیصی هستند، در دنیای واقعی گسترده است و توسعه چنین قابلیت‌هایی احتمالاً دسترسی به تخصص تشخیصی و پیش‌آگهی را بهبود می‌بخشد و در نتیجه کیفیت، ثبات، در دسترس بودن و مقرون‌به‌صرفه بودن مراقبت را بهبود می‌بخشد. رویکردی مبتنی بر عدالت سلامت به‌منظور ادغام چنین فناوری در گردش‌های کاری موجود، که مستلزم کار در مراحل توسعه، اجرا و سیاست‌گذاری می‌باشد، ممکن است پتانسیل کمک برای دستیابی به نتایج بهتر سلامت (به‌ویژه برای جمعیت‌هایی که با نابرابری‌های مراقبت‌های بهداشتی مواجه هستند) را داشته باشد. بااین‌حال، اگرچه نشان داده

1. Artificial linguistic internet computer entity (ALICE)
2. Large language model (LLM)

شده است که LLMها دانش بالینی را رمزگذاری کرده و توانایی پاسخ‌گویی به سؤالات پزشکی تک‌نوبتی بسیار دقیقی را نشان داده‌اند، بااین‌حال قابلیت‌های مکالمه‌ای آن‌ها متناسب با حوزه‌های خارج از پزشکی بالینی تنظیم شده است [۱۷]. کارهای قبلی در LLMها هنوز به‌طور دقیق قابلیت‌های گرفتن شرح‌حال بالینی و گفت‌وگوی تشخیصی سیستم‌های هوش مصنوعی را بررسی نکرده‌اند یا این موضوع را در مقایسه با قابلیت‌های گسترده پزشکان عمومی شاغل در حوزه سلامت، زمینه‌سازی نکرده‌اند.

گرفتن شرح‌حال بالینی و گفت‌وگوی تشخیصی که از طریق آن پزشکان برنامه‌های تشخیص و مدیریت را استخراج می‌کنند، مهارتی پیچیده را نشان می‌دهند که انجام بهینه آن به‌شدت به زمینه بستگی دارد. بنابراین، محورهای ارزیابی متعددی به‌منظور ارزیابی کیفیت یک گفت‌وگوی تشخیصی، ازجمله ساختار و کامل بودن شرح‌حال استخراج شده، دقت تشخیصی، مناسب بودن برنامه‌های مدیریتی و منطق آن‌ها و ملاحظات بیمارمحور، مانند ایجاد رابطه، احترام به فرد و اثربخشی ارتباط، مورد نیاز است. اگر قرار باشد پتانسیل مکالمه‌ای LLMها در پزشکی محقق گردد، یک نیاز برآورده نشده مهم برای بهینه‌سازی بهتر، توسعه و ارزیابی سیستم‌های هوش مصنوعی پزشکی برای ویژگی‌هایی مانند این موارد وجود دارد که منحصر به گرفتن شرح‌حال و گفت‌وگوی تشخیصی بین پزشکان و بیماران می‌باشد [۱۵، ۱۶].

توجه دقیق به استقرار مسائل اخلاقی در این فناوری، ازجمله ارزیابی دقیق کیفیت در محیط‌های بالینی مختلف و تحقیق در مورد روش‌های تخمین عدم قطعیت قابل اعتماد که امکان ارجاع به متخصصان بالینی انسانی را درصورت نیاز فراهم می‌کند، ضروری است. این موارد و سایر موانع به‌منظور کاهش اتکای بیش از حد بالقوه به فناوری‌های LLM، به‌همراه سایر اقدامات خاص برای توجه به الزامات اخلاقی و نظارتی خاص موارد استفاده آینده و حضور پزشکان واجد شرایط در حلقه برای محافظت از هرگونه خروجی مدل، مورد نیاز می‌باشد. همچنین تحقیقات بیشتری برای ارزیابی میزان سوگیری‌ها و آسیب‌پذیری‌های امنیتی که ممکن است از مدل‌های پایه یا شرایط استفاده در استقرار ایجاد شوند، مورد نیاز خواهد بود. با توجه به تکامل مداوم دانش بالینی، توسعه راه‌هایی برای LLMها جهت استفاده از اطلاعات بالینی به‌روز نیز مهم تلقی خواهد شد [۱۸].

مدل زبان بزرگ و مدل زبان بزرگ چندوجهی[1]

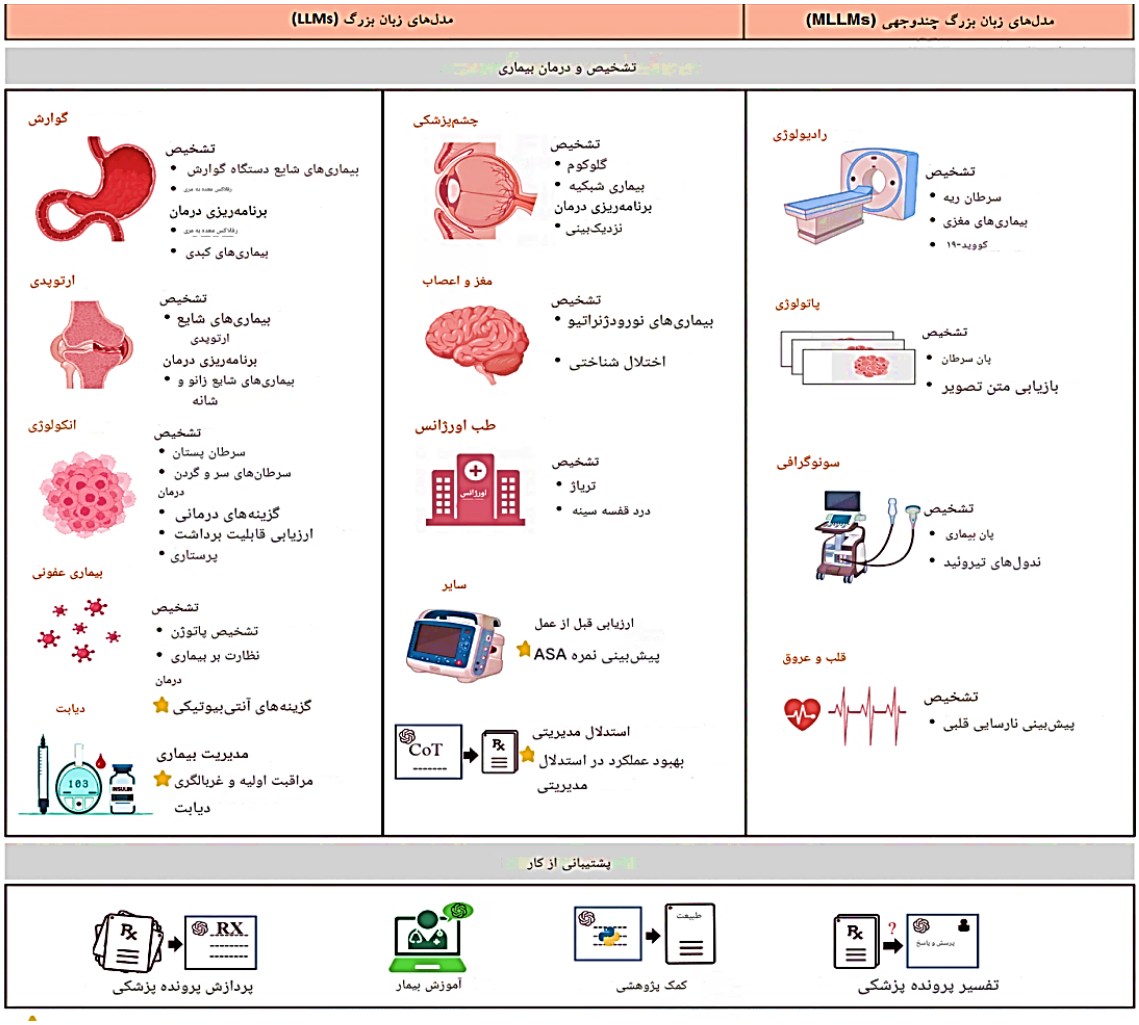
از زمان انتشار ChatGPT-3.5 در نوامبر ۲۰۲۲، استفاده بالقوه از مدل زبان بزرگ (LLM) در حوزه پزشکی توجه گسترده‌ای را به خود جلب کرده است. LLMها براساس دانش گسترده موجود در اینترنت، ازجمله دانش پزشکی، آموزش دیده و تنظیم شده‌اند. در نتیجه، این مدل‌ها در مقایسه با ابزارهای زبانی سنتی، از درک زبان طبیعی، تشخیص الگو و قابلیت‌های تحلیل ارتباط برتر برخوردارند. با تکامل سریع LLMها، نسخه‌های جدیدتر (به‌عنوان مثال، GPT-4o و Claude 3.5-sonnet) نه‌تنها می‌توانند داده‌های پزشکی چندوجهی را مدیریت کنند، بلکه جدیدترین یافته‌های تحقیقاتی آنلاین و پایگاه‌های دانش خصوصی را نیز ادغام کرده و پشتیبانی اطلاعاتی متنوعی را در اختیار پزشکان قرار می‌دهند که این امر ممکن است منجر به افزایش توانایی‌های تصمیم‌گیری آن‌ها گردد [۱۹]. LLMها از طریق پاسخ به پرسش و پاسخ تعاملی و استدلال احتمالاتی، در ترکیب با ارزیابی ویژگی‌های فردی بیمار، می‌توانند مزایا و معایب گزینه‌های مختلف تشخیصی و درمانی را استنباط و ارائه دهند و به‌طور بالقوه تشخیص بیماری و تصمیم‌گیری‌های درمانی را تسهیل نمایند. LLMها را می‌توان براساس معماری فنی آن‌ها (مثلاً مدل‌های خودهمبسته مانند ترانسفورماتور مولد از پیش آموزش‌دیده [GPT] و مدل‌های ماسک‌شده مانند نمایش‌های رمزگذار دوطرفه از

1. Multimodal LLMs (MLLMs)

ترانسفورماتورها [BERT])، دسترسی‌پذیری (منبع باز یا منبع بسته) یا عملکرد (همه منظوره یا خاص دامنه) طبقه‌بندی کرد [۲۰]. به‌منظور افزایش عملکرد LLMها در وظایف حرفه‌ای، محققان از تکنیک‌های بهینه‌سازی مختلفی ازجمله مهندسی سریع، تکنیک‌های پس از آموزش و سیستم‌های چندعاملی استفاده می‌کنند. برای وظایف چندوجهی، مانند تفسیر تصویر و تشخیص گفتار، LLMهای چندوجهی (MLLMها) برای ادغام رمزگذارهای تصویر (مثلاً ترانسفورماتورهای بینایی)، رمزگذارهای صوتی (مثلاً Wav2Vec) و ترکیب ویژگی‌های چندوجهی با استفاده از پروژکتورهای ورودی، توسعه داده شده‌اند. این معماری، MLLMها را قادر می‌سازد تا انواع مختلفی از داده‌های پزشکی را به‌طور هم‌زمان پردازش و درک کنند. MLLMها به‌دلیل پتانسیل خود در تسهیل تشخیص و درمان، توجه تحقیقاتی قابل توجهی را در حوزه بالینی به خود جلب کرده‌اند. انتخاب روش LLM و بهینه‌سازی به تخصص محقق، منابع موجود و اهداف تحقیق بستگی دارد و منجر به رویکردهای متنوعی برای به‌کارگیری LLM در تشخیص و درمان بالینی می‌شود [۲۱].

پیشرفت قابل توجهی در استفاده از LLMها به‌منظور تشخیص و درمان بیماری‌ها حاصل شده است. LLMها می‌توانند الگوهای پیچیده علائم را براساس اطلاعات متنی شناسایی کنند، درحالی‌که MLLMها می‌توانند هم‌زمان تصاویر و صداهای پزشکی رایج را تفسیر نمایند. شواهد روبه‌رشد تحقیقات بالینی نشان می‌دهد که LLMها ابزارهای ارزشمندی برای درجه‌بندی دقیق بیماری‌های شایع و تشخیص بیماری‌های نادر، پیش‌بینی خطرات بیماری براساس اطلاعات فردی بیمار و ارائه توصیه‌های درمانی شخصی‌سازی شده هستند (شکل ۲). این بررسی، مروری بر پیشرفت در استفاده از LLMها برای تشخیص و درمان بیماری‌ها در سیستم‌های مختلف بدن براساس مطالعات گذشته‌نگر و آینده‌نگر ارائه می‌دهد [۲۲].

از منظر تشخیص و درمان بالینی، علی‌رغم پتانسیل LLMها در کمک به تصمیم‌گیری بالینی، به‌دلیل پیچیدگی و تنوع بیماری‌ها، چالش‌های قابل توجهی همچنان وجود دارد. هر بیمار، مشخصات منحصربه‌فردی شامل پیشینه ژنتیکی، سبک زندگی، عوامل محیطی، سابقه پزشکی و وضعیت فیزیولوژیکی دارد که می‌تواند منجر به علائم و پاسخ‌های درمانی متنوعی برای همان بیماری شود [۲۳]. علاوه‌بر این، مراقبت‌های پزشکی شخصی‌سازی‌شده باید عوامل دیگری ازجمله تخصیص منابع، مقرون‌به‌صرفه بودن و ملاحظات اخلاقی را نیز در نظر بگیرد. اگرچه LLMها توانایی تقویت تصمیمات تشخیصی و مدیریتی را حتی در موارد پیچیده نشان داده‌اند، برای تولید توصیه‌های واقعاً شخصی‌سازی‌شده که نیازهای پزشکی فردی را برطرف می‌کنند، به اصلاحات بیشتری نیاز است. علاوه‌بر این، اگرچه MLLMها امیدوارکننده هستند، اما اعتقاد بر این است که آن‌ها هنوز در مراحل اولیه تفسیر و ادغام سوابق متنی با داده‌های غیرمتنی (به‌عنوان مثال، تصاویر توموگرافی کامپیوتری [CT]، صداهای تنفسی و ویدئوهای حرکتی) هستند. مزایای بالینی بالقوه چنین ادغامی، تحقیقات بیشتر را می‌طلبد [۲۴].



شکل ۲. مطالعات مربوط به کاربردهای LLM در تشخیص، درمان و کارهای پشتیبانی

کاربردهای LLM در تشخیص بیماری

هنگامی که LLMها برای اولین بار پدیدار شدند، در پردازش متون طولانی و انتشار دانش رایج سرآمد بودند. شواهد قابل توجهی نشان می‌دهد که LLMها در مدیریت پرونده‌های پزشکی و توضیح تعاملی شرایط پزشکی به بیماران، نویدبخش هستند. این وظایف درحال‌حاضر ماهرانه‌ترین و مرتبط‌ترین وظایف بالینی می‌باشند که توجه متخصصان مراقبت‌های بهداشتی را به خود جلب کرده است. بااین‌حال، تشخیص و درمان بیماری‌ها، که هسته اصلی مراقبت‌های بهداشتی را تشکیل می‌دهند، وظایف پیچیده‌تر و متغیرتری هستند. خوشبختانه، تحقیقات اخیر نشان می‌دهد که LLMها می‌توانند با شناسایی الگوهای بیماری و ارائه پیشنهادات تشخیصی از طریق تجزیه‌وتحلیل داده‌ها به متخصصان مراقبت‌های بهداشتی کمک کنند. علاوه‌بر این، MLLM ها پتانسیل تفسیر و ادغام داده‌های پزشکی غیر متنی، ازجمله تصاویر، صداها و حتی ویدئوها را دارند [۲۵].

گوارش‌شناسی[1]

تشخیص بیماری‌های دستگاه گوارش (GI) یکی از مهم‌ترین حوزه‌های مورد علاقه محققان است. برای ارزیابی عملکرد ChatGPT-3.5 در تشخیص بیماری‌های دستگاه گوارش، لاهات و همکاران[2] مطالعه‌ای جامع با استفاده از

1. gastroenterology
2. Lahat et al.

سؤالات واقعی از بیماران مبتلا به بیماری‌های دستگاه گوارش انجام دادند. محققان ۴۵ سؤال تشخیصی مربوط به بیماری‌های دستگاه گوارش مانند گاستریت، ازوفاژیت و سنگ کیسه صفرا را به GPT-3.5 ارسال کردند و سه متخصص گوارش باتجربه؛ دقت، وضوح و مفید بودن پاسخ‌های ارائه‌شده توسط GPT-3.5 را (در مقیاس ۱ تا ۵) ارزیابی کردند. نتایج نشان داد که برای سؤالات تشخیصی، میانگین نمرات GPT-3.5 برای دقت ۳/۷ ± ۱/۷، برای وضوح ۳/۷ ± ۱/۸ و برای مفید بودن ۳/۵ ± ۱/۷ بود. پتانسیل GPT-4 در تشخیص افتراقی بیماری رفلاکس معده به مری (GERD) نیز ارزیابی شد [۱۸]. محققان پنج سؤال تشخیصی با تمرکز بر تشخیص افتراقی و عوارض بیماری به GPT-4 ارسال کردند. نتایج نشان داد که ۹۳/۴٪ از پاسخ‌ها کاملاً مناسب یا عمدتاً مناسب در نظر گرفته شدند.

برای بررسی دقت پاسخ‌های LLMها به سؤالات تشخیصی سرطان کبد، محققان ۲۰ سؤال (شامل ۱۳ سؤال مرتبط با تشخیص) را از سه LLM (شامل GPT-3.5، Gemini و Bing) مطرح کردند. نتایج نشان داد که Gemini بالاترین دقت را داشت و پس از آن GPT-3.5 و Bing (به‌ترتیب ۶۰٪، ۴۵٪ و ۳۰٪) قرار داشتند، این نتیجه مؤید آن است که LLMهای عمومی توانایی کمتر از حد مطلوبی در پاسخ به سؤالات پیچیده مرتبط با تشخیص (زمانی که تشخیص اولیه واضح بود)، نشان دادند. طبق تحلیلی سیستماتیک توسط مائورو و همکارانش دقت در تشخیص بیماری‌های گوارشی با GPT-3.5 از ۶/۴٪ تا ۴۵/۵٪ و با GPT-4 بین ۴۰/۰٪ تا ۹۱/۴٪ متغیر بود. بااین‌حال، همه مطالعات خطر سوگیری بالایی داشتند و تعمیم نتایج تک مطالعه توصیه نشد. این محدودیت‌ها نشان‌دهنده فقدان فعلی الگوهای تحقیقاتی استاندارد و معیارهای ارزیابی قوی برای مطالعات در مورد استفاده از LLMها در تشخیص بالینی است [۲۶].

عصب‌شناسی[1]

در زمینه تشخیص بیماری‌های عصبی، کوگا و همکاران[2] توانایی GPT-3.5، GPT-4 و Google Bard را برای پیش‌بینی تشخیص‌های نوروپاتولوژیک از خلاصه‌های بالینی بررسی کردند. چندین LLM در ۲۵ مورد از بیماری‌های نورودژنراتیو مورد پرسش قرار گرفتند و از آن‌ها خواسته شد تا تشخیص‌ها و دلایل پاتولوژیک خود را ارائه دهند. در مقایسه با تشخیص‌های بالینی نهایی انجام شده توسط پزشکان، GPT-3.5، GPT-4 و Google Bard به‌ترتیب به دقت ۷۶٪، ۸۴٪ و ۷۶٪ دست یافتند. این یافته‌ها پتانسیل LLMها را در تشخیص نوروپاتولوژیک نشان داد [۲۷]. علاوه‌بر این، برای بررسی عملکرد GPT-4 در غربالگری اولیه اختلال شناختی خفیف (MCI)، وانگ و همکاران[3] داده‌های ۱۷۴ شرکت‌کننده را از پایگاه داده بانک دمانس جمع‌آوری کردند. مطالعه آن‌ها نشان داد که GPT-4 حساسیت ۷۷/۳–۸۶/۴٪ و ویژگی ۸۳/۳–۹۴/۹٪ دارد. این یافته‌ها تأکید کردند که GPT-4 می‌تواند به‌طور مؤثر بیماران بالقوه مبتلا به MCI را تشخیص دهد، اما بهینه‌سازی بیشتر و مهندسی سریع استاندارد توسط پزشکان مورد نیاز است. مهندسی سریع فرایند طراحی LLMها برای استخراج اطلاعاتی است که نیازهای کاربر را بهتر برآورده می‌کند. این امر ممکن است عملکرد مدل را بدون تغییر پارامترهای مدل بهبود بخشد و آن را کاربرپسندتر و ترویج آن را آسان‌تر کند [۲۸].

طب اورژانس[4]

تشخیص دقیق و به‌موقع، مسئولیت‌های اصلی پزشکان اورژانس در بخش‌های اورژانس (EDs)[5] است. تحقیقات نشان

1. neurology
2. Koga et al.
3. Wang et al.
4. emergency medicine
5. Emergency departments (EDs)

داده است که LLMها پتانسیل بهبود کارایی تریاژ و دقت تشخیصی در بیماران اورژانسی را دارند. برگ و همکاران[1] [۲۹]، توانایی GPT-3.5 و GPT-4 را در ایجاد تشخیص‌های افتراقی اولیه در ۳۰ بیمار در بخش اورژانس بررسی کردند. نتایج نشان داد که بدون داده‌های آزمایشگاهی، پزشکان در ۸۳٪ موارد، تشخیص نهایی را به‌درستی لحاظ کردند، درحالی‌که GPT-3.5 و GPT-4.0 به‌ترتیب ۷۷٪ و ۸۷٪ دقت داشتند. با داده‌های آزمایشگاهی، هم پزشکان و هم GPT-4.0 دارای دقت ۸۷٪ بودند، درحالی‌که دقت GPT-3.5 به ۹۷٪ افزایش یافت. برای بررسی اینکه آیا GPT-4 می‌تواند به‌طور دقیق شدت بالینی را در بخش اورژانس ارزیابی کند، یک مطالعه مقطعی شامل ۲۵۱۴۰۱ ویزیت اورژانس بزرگ‌سالان انجام شد تا پتانسیل GPT-4 برای طبقه‌بندی سطوح شدت بیمار براساس ۱۰۰۰۰ جفت نمره شاخص شدت اورژانسی ارزیابی شود. یافته‌ها نشان داد که GPT-4 به دقت ۸۹٪ دست یافته است و در زیرمجموعه‌ای از ۵۰۰ جفت طبقه‌بندی شده دستی، دقت LLM با دقت پزشک بررسی کننده قابل مقایسه بود (۸۸٪ در مقابل ۸۶٪) [۲۹]. درد قفسه سینه علامتی شایع در بخش اورژانس محسوب می‌شود که اغلب با بیماری‌های جدی همراه است، اما معمولاً خوش‌خیم می‌باشد. هستون و لوئیس[2] توانایی GPT-4 را در طبقه‌بندی خطر در موارد درد قفسه سینه شبیه‌سازی شده ارزیابی کردند. نمرات GPT-4 با نمره خطر ترومبولیز در انفارکتوس میوکارد (TIMI) و سابقه، یافته‌های الکتروکاردیوگرافی (ECG)، سن، عوامل خطر و نمره سطح تروپونین (HEART) مقایسه شدند و همبستگی بالایی (r = ۰/۹۰ و ۰/۹۳) نشان دادند. بااین‌حال، این مطالعه همچنین نشان داد که پاسخ‌های GPT-4 برای یک نمره ثابت، هنگامی که LLM به‌طور مکرر مورد پرسش قرار گرفت، نرخ ناسازگاری بین ۴۵ تا ۴۸ درصد داشت که اهمیت نظارت پزشک را برجسته می‌کند [۳۰].

بیماری‌های عفونی[3]

در بیماری‌های عفونی، تشخیص پاتوژن یک عامل کلیدی است که اغلب به تجربه و قضاوت ذهنی پزشک بستگی دارد. تعیین دقیق وجود عفونت و شناسایی پاتوژن خاص قبل از کشت موفقیت‌آمیز برای درمان بعدی بیماران بسیار مهم می‌باشد. کاربردهای LLM در تشخیص بیماری‌های عفونی بررسی شده است. پرت و اشمید[4] عملکرد GPT-4 را در تشخیص عفونت‌های دستگاه ادراری مرتبط با کاتتر در ۵۰ مورد مجازی گزارش کردند. این مدل با حساسیت ۹۱٪ و اختصاصیت ۹۲٪ عملکرد بالایی را نشان داد [۳۱].

علاوه‌بر این، سریکانث و همکاران[5] گزارش دادند که BERTweet، یک مدل زبانی مبتنی بر BERT، می‌تواند اطلاعات رسانه‌های اجتماعی را تجزیه‌وتحلیل و دسته‌بندی کند تا به‌طور مؤثر میزان بروز بیماری لایم را رصد کند. این مطالعه کاربردهای بالقوه گسترده LLMها را در سلامت عمومی و درمان پیشگیرانه برجسته نمود [۳۲].

قلب و عروق[6]

ECG نقش مهمی در تشخیص به‌موقع رویدادهای قلبی-عروقی دارد. محققان از یک LLM با یک مجموعه داده عمومی در یک کار، هم‌ترازی گزارش ECG براساس BioClinical BERT استفاده کردند. در مرحله بعد، شبکه، به‌منظور پیش‌بینی خطر نارسایی قلبی با استفاده از گروه‌هایی که بر بیماران مبتلا به فشارخون بالا و انفارکتوس میوکارد تمرکز داشتند، به‌طور دقیق تنظیم شد. امتیاز شاخص تطابق (C-index) مدل برای فشارخون بالا ۰/۶۳ و برای

1. Berg et al.
2. Heston & Lewis
3. infectious diseases
4. Perret & Schmid
5. Srikanth et al.
6. cardiology

انفارکتوس میوکارد ۰/۵۸ گزارش شد که پتانسیل BERT را ازنظر اثربخشی و مقیاس‌پذیری برای پیشبرد ارزیابی ریسک با داده‌های پیچیده ECG بالینی نشان داد [۳۳]. در مطالعه‌ای دیگر، محققان استفاده از LLMها با نسل بازیابی-تقویت‌شده (RAG)[1] را برای تشخیص ECG با دقت صفر بررسی کردند. این مطالعه نشان داد که GPT-3.5 در تشخیص آریتمی‌ها به دقت ۷۵/۷٪ دست یافته است که نشان‌دهنده توانایی قوی GPT-3.5 در تشخیص صفر می‌باشد. RAG می‌تواند عملکرد LLMها را در وظایف تخصصی بدون تغییر پارامترهای مدل با ساخت پایگاه‌های دانش خاص دامنه بهبود بخشد [۳۴].

چشم‌پزشکی[2]

قابلیت‌های تشخیصی LLMها براساس متون موردی در چشم‌پزشکی ارزیابی شده است. با ارزیابی عملکرد GPT-4 در تشخیص و مدیریت گلوکوم و بیماری‌های شبکیه از طریق سؤالات موردی، یک مطالعه نشان داد که LLMها می‌توانند ازنظر دقت و کامل بودن از متخصصان آموزش‌دیده در فلوشیپ بهتر عمل کنند یا با آن‌ها برابری کنند. این امر پتانسیل LLMها را به‌عنوان ابزارهای تشخیصی مؤثر در زمینه‌های تخصصی چشم‌پزشکی برجسته کرد. برخلاف LLMهای مبتنی بر متن، MLLMها ماژول‌های پردازش را برای انواع مختلف اطلاعات، مانند تصاویر و صدا، ادغام می‌کنند. این مدل‌ها ویژگی‌های داده‌های چندوجهی را با هم ترکیب می‌کنند و به آن‌ها امکان می‌دهند طیف وسیعی از وظایف چندوجهی را با قابلیت‌های قابل توجهی انجام دهند. توانایی MLLMها در توصیف تصاویر پزشکی اخیراً توجه زیادی را ازسوی محققان به خود جلب کرده است [۳۵].

سیف و همکاران[3] دقت تشخیصی GPT-4V، یک MLLM، را در تشخیص گلوکوم با استفاده از ۴۰۰ تصویر آزمایشی فوندوس از مجموعه داده‌های REFUGE (Retinal Fundus Glaucoma Challenge) ارزیابی کردند. بدون آموزش خاص، GPT-4V به دقت ۹۰٪ و ویژگی ۹۴/۴۴٪ دست یافت، اما حساسیت کمتری معادل ۵۰٪ نشان داد. تکنیک‌های پیش‌پردازش تصویر بیشتر مورد بررسی قرار گرفتند: برش تصاویر برای تمرکز بر روی دیسک نوری و ناحیه اطراف پاپیلاری، حساسیت را به‌طور قابل توجهی به ۷۸/۵۰٪ بهبود بخشید، درحالی‌که استفاده بیشتر از برابرسازی هیستوگرام تطبیقی محدود به کنتراست (CLAHE) منجر به حساسیت ۶۲/۵۰٪ شد. بااین‌حال، این مراحل پیش‌پردازش، دقت را کاهش دادند. این مطالعه پتانسیل MLLMها را در تشخیص تصاویر پزشکی شناسایی کرد و نشان داد که آن‌ها ممکن است به داده‌های آموزشی کمتری نیاز داشته باشند و می‌توانند به ابزارهای پشتیبانی پزشکی نوآورانه، به‌ویژه در محیط‌های با محدودیت منابع، منجر شوند [۳۶].

رادیولوژی[4]

برای ارزیابی پتانسیل LLMها در تشخیص‌های رادیولوژیکی، دهداب و همکاران[5] یک مطالعه برای ارزیابی عملکرد GPT-4V در تشخیص سی‌تی‌اسکن قفسه سینه، به‌ویژه برای شناسایی بیماری کروناویروس ۲۰۱۹ (COVID-19)، سرطان ریه سلول غیرکوچک (NSCLC)[6] و موارد کنترل انجام دادند. نتایج نشان داد که GPT-4V دقت تشخیصی کلی ۵۶/۸٪ دارد. حساسیت برای شناسایی NSCLC ۲۷/۳٪ و اختصاصیت ۶۰/۵٪ بود؛ برای COVID-19، حساسیت ۱۳/۶٪ و اختصاصیت ۶۴/۳٪ بود؛ همچنین برای موارد کنترل، حساسیت و اختصاصیت به‌ترتیب ۳۱/۸٪ و ۹۵/۲٪ حاصل

1. Retrieval-Augmented Generation (RAG)
2. ophthalmology
3. Saif et al.
4. radiology
5. Dehdab et al.
6. Non-small cell lung cancer (NSCLC)

شد. این یافته‌ها نشان داد که GPT-4V عملکرد تشخیصی متغیری را در تفسیر سی‌تی‌اسکن قفسه سینه نشان می‌دهد. عملکرد تشخیصی متغیر MLLMها عمدتاً به‌دلیل محدودیت‌هایی مانند داده‌های آموزشی ناکافی و غیرمتنوع و همچنین معماری‌های خاص ممکن است برای پردازش تصویر، بهینه نشده باشند. علاوه‌بر این، چالش‌های موجود در ادغام مؤثر اطلاعات چندوجهی و محدودیت‌های منابع محاسباتی، به عملکرد نامطلوب در درک تصاویر کمک می‌کند [۳۷].

برای بررسی قابلیت تشخیصی MLLMها در ادغام اطلاعات متن و تصویر در تشخیص بیماری، هیروساوا و همکاران[1]، دقت تشخیصی خود را با استفاده از ۳۶۳ شرح مورد با تصاویر آزمایش کردند و آن‌ها را با GPT-4 بدون بینایی مقایسه کردند. نتایج نشان داد که تشخیص نهایی در ۸۵/۱٪ از ۱۰ فهرست برتر تشخیص افتراقی GPT-4V گنجانده شده است، مشابه ۸۷/۹٪ GPT-4[۳۷]. به‌طور مشابه، هوریوچی و همکاران[2]، هنگام مقایسه ورودی متنی با ورودی مستقیم تصویر، دقت تشخیصی کمتری برای GPT-4 و GPT-4V نسبت به رادیولوژیست‌ها یافتند. در مطالعه دیگری، سباستین و همکاران[3] [۳۸] گزارش‌های رادیولوژی تهیه شده توسط GPT-4 را با گزارش‌های نوشته شده توسط رادیولوژیست‌های انسانی با استفاده از داده‌های اشعه ایکس قفسه سینه مقایسه کردند. در میان گزارش‌های تهیه شده توسط GPT-4، گزارش‌های مبتنی بر سیستم‌های هوش مصنوعی ترکیبی مبتنی بر متن و تصویر–متن، عملکرد نسبتاً خوبی داشتند. LLM مبتنی بر متن کمی بهتر از امتیازدهی رادیولوژیست عمل کرد (۱۶/۹۵ در مقابل ۱۵/۵۴)، درحالی‌که MLLM ترکیبی از تصویر و متن، از اکثر معیارهای ارزیابی خودکار بهتر عمل کرد. در مقابل، گزارش‌های تولید شده توسط هوش مصنوعی فقط تصویر، به‌طور مداوم در تمام ارزیابی‌ها کمترین امتیاز را کسب کردند. محققان مجموعه‌ای از داده‌های سی‌تی‌اسکن مغزی سه‌بُعدی متشکل از ۱۸۸۸۵ جفت اسکن متنی ایجاد کردند. سپس از تنظیم دستورالعمل بصری بالینی برای آموزش و تنظیم دقیق مدل BrainGPT استفاده کردند. یک آزمون تورینگ برای ارزیابی BrainGPT انجام شد و نشان داد که تقریباً ۷۴٪ از گزارش‌های تولید شده توسط BrainGPT از گزارش‌های نوشته شده توسط انسان قابل تشخیص نیستند، که این نتیجه نشان‌دهنده قابلیت‌های قوی پردازش زبان طبیعی مدل و پتانسیل بالینی آن در تولید گزارش‌های رادیولوژی است [۳۸].

برای بهبود توانایی استدلال MLLMها در تشخیص‌های رادیولوژی، دیوید و همکاران[4] [۳۹] یک سیستم چندعاملی را معرفی کردند که CLIP و GPT-4 را با معماری‌های چندعاملی و روش‌های مهندسی سریع ترکیب کرد. این مطالعه نشان داد که سیستم چندعاملی در وظایف تشخیص عکس‌برداری با اشعه ایکس قفسه سینه، به‌ویژه برای بیماری‌های نادر، از روش‌های موجود بدون نیاز به عکس‌برداری (zero-shot) بهتر عمل می‌کند و در چندین مجموعه داده، به عملکردی قابل مقایسه یا برتر از برخی روش‌های تحت نظارت دست یافته است. سیستم‌های چندعاملی می‌توانند چندین عامل هوشمند و ابزار پردازش اطلاعات را در خود جای داده و عملکرد بهتری را در وظایف پیچیده ارائه دهند [۳۹].

این نتایج نشان داد اگرچه گزارش‌های تولید شده توسط هوش مصنوعی به کیفیتی در سطح انسان نزدیک شده‌اند، اما از گزارش‌های رادیولوژیست‌ها پیشی نگرفته‌اند. ادغام تصاویر با متن، دقت GPT-4 را به‌طور قابل توجهی بهبود نبخشید، احتمالاً به‌دلیل اتکای MLLMها به ویژگی‌های متنی که دقت تشخیصی آن‌ها را در سناریوهای داده‌های غیرمتنی محدود می‌کرد.

1. Hirosawa et al.
2. Horiuchi et al.
3. Sebastian et al.
4. David et al.

سونوگرافی[1]

در یک مطالعه کوچک که توسط سلطان و همکاران[2] [۴۰] انجام شد، GPT-4V دقت بالایی در تجزیه‌وتحلیل و تفسیر تصاویر سونوگرافی تیروئید و کلیه نشان داد و توانست ضایعات را روی تصاویر شناسایی و علامت‌گذاری کند. در مطالعه دیگری، محققان عملکرد تشخیصی LLaVA-Ultra، یک دستیار چندوجهی چینی زبان و بصری که برای پزشکی سونوگرافی توسعه داده شده است را بررسی کردند [۴۰]. ازنظر عملکرد، LLa-VA-Ultra از مدل‌های پیشرفته موجود در مجموعه داده‌های بیمارستان ایالات متحده پیشی گرفت و به دقت ۶۲/۰٪ و امتیاز ۷۲/۲٪ F1 کسب کرد. همچنین در CT، تصویربرداری رزونانس مغناطیسی (MRI) و زیرمجموعه‌های اشعه ایکس برتری داشت و به دقت ۸۳/۸٪ و امتیاز ۹۳/۴٪ F1 دست یافت [۴۱].

در مطالعه‌ای دیگر [۴۲]، محققان LLMهای همراه با فناوری تبدیل تصویر به متن را با یک مدل یادگیری عمیق (DL) سنتی برای تشخیص ندول‌های تیروئید ادغام کردند. این مطالعه نشان داد که اگرچه LLMها ممکن است در مقایسه با مدل‌های DL در تشخیص تصاویر پزشکی عملکرد ضعیف‌تری داشته باشند، اما شفافیت و قابلیت تفسیر آن ها ارزش قابل توجهی برای آموزش پزشکی و تصمیم‌گیری بالینی ارائه می‌دهد. MLLMها از معماری‌های مبتنی بر ترانسفورماتور استفاده می‌کنند و به مجموعه داده‌های چندوجهی متنوعی نیاز دارند، درحالی‌که مدل‌های DL معمولاً از معماری‌های خاص وظیفه مانند شبکه‌های عصبی کانولوشن (CNN) استفاده می‌کنند و فقط به داده‌های تک‌وجهی نیاز دارند. مدل‌های DL اغلب برای وظایف خاص به دقت بالایی دست می‌یابند و استقرار محلی آن‌ها نسبتاً آسان است. در مقابل، MLLMها پوشش در چندین سیستم پزشکی، قابلیت‌های یادگیری بدون نیاز به پردازش و تعاملات پیشرفته انسان و کامپیوتر را ارائه می‌دهند. MLLMها عموماً نیازهای محاسباتی بالاتری دارند و ممکن است با ملاحظات حریم خصوصی بیشتری روبه‌رو شوند، درحالی‌که مدل‌های DL معمولاً ازنظر منابع کارآمدتر و پیاده‌سازی آن‌ها برای حفاظت از حریم خصوصی ساده‌تر است. این تفاوت‌ها بین MLLMها و مدل‌های DL منعکس کننده رویکردهای متمایز آن‌ها در تشخیص مبتنی بر تصویر است که هرکدام ویژگی‌های منحصربه‌فردی را ارائه می‌دهند که می‌تواند به قابلیت‌های تشخیصی در تصویربرداری پزشکی کمک کند [۴۲].

آسیب‌شناسی[3]

در زمینه آسیب‌شناسی، ظهور MLLMهای بنیادی به‌طور قابل توجهی به استفاده از هوش مصنوعی تعمیم‌یافته برای تجزیه‌وتحلیل تصاویر پاتولوژیک کمک کرده است. PathChat، یک مدل زبانی چندوجهی که برای آسیب‌شناسی توسعه و ارزیابی شده است، یک رمزگذار بصری از پیش آموزش‌دیده از طریق یادگیری خودنظارتی را با LLMهای Llama2 متشکل از ۱/۳ میلیارد پارامتر ترکیب کرد و بر روی بیش از ۴۵۶۰۰۰ دستورالعمل زبان بصری دقیق تنظیم شد. نتایج نشان داد که PathChat در سؤالات تشخیصی چندگزینه‌ای با دقت ۷۸/۱٪ بدون زمینه بالینی، عملکرد خوبی داشت. با افزودن زمینه بالینی، دقت Path-Chat به ۸۹/۵٪ افزایش یافت. خو و همکاران[4] همچنین یک مدل آسیب‌شناسی کل اسلاید بنیادی، Prov-GigaPath، توسعه دادند که بر روی ۱۷۱۱۸۹ تصویر کل اسلاید از پیش آموزش داده شده بود و عملکرد پیشرفته‌ای را در وظایف مختلف آسیب‌شناسی دیجیتال نشان داد.

1. ultrasonography
2. Sultan et al.
3. pathology
4. Xu et al.

ژنگ و همکاران[1] [۴۳] یک ارزیابی معیار از عملکرد مدل PathCLIP در تجزیه‌وتحلیل تصاویر آسیب‌شناسی انجام دادند. PathCLIP یک مدل پیش‌آموزش زبان-تصویر مقابله‌ای است که به‌طور خاص برای آسیب‌شناسی طراحی شده و بر روی مجموعه داده‌ای حاوی بیش از ۲۰۰۰۰۰ جفت تصویر-متن آموزش دیده است. PathCLIP برای تسهیل بازیابی تصویر-متن و کمک به تشخیص در آسیب‌شناسی مفید می‌باشد.

کاربردهای LLM در درمان بیماری‌ها

تحقیقات بالینی موجود در مورد استفاده از LLMها برای کاربردهای درمانی، عمدتاً مبتنی بر مطالعات گذشته‌نگر یا مشاهده‌ای در رشته‌های بالینی متعدد بوده است. LLMها پتانسیل قابل توجهی در تولید برنامه‌های درمانی شبیه‌سازی شده براساس سوابق پزشکی بیمار، تفسیر داده‌های تشخیصی و ارائه توصیه‌های درمانی به بیماران نشان داده‌اند.

انکولوژی[2]

استراتژی‌های درمان سرطان معمولاً براساس نوع خاص سرطان، مشخصات ژنتیکی و مرحله بیماری تنظیم می‌شوند و این امر مستلزم رویکردهای درمانی شخصی‌سازی‌شده فزاینده‌ای می‌باشد. محققان استفاده از LLMها را برای توسعه برنامه‌های درمانی سفارشی برای بیماران در انواع مختلف سرطان بررسی کرده‌اند.

یک مطالعه مشاهده‌ای، سازگاری توصیه‌های درمانی را برای ۲۰ مورد سرطان پستان بین GPT-3.5 و یک تیم چندرشته‌ای (MDT) مقایسه کرد. در گزینه‌های درمانی خاص، GPT-3.5 سازگاری بالایی با MDT در توصیه‌های جراحی، شیمی‌درمانی و رادیوتراپی (به‌ترتیب ۹۵٪، ۹۴/۵٪ و ۹۵٪) نشان داد، درحالی‌که سازگاری در توصیه‌های ژن‌درمانی (۷۰٪) کمتر بود [۴۴]. اشمیدل و همکاران[3] در ارائه توصیه‌های درمانی برای ۲۰ بیمار مبتلا به سرطان سر و گردن، GPT-3.5 و GPT-4 را با MDT مقایسه کردند. GPT-3.5 و GPT-4 در مقایسه با MDT، پاسخ‌های کلی برای جراحی، شیمی‌درمانی و پرتودرمانی با سازگاری متوسط ارائه دادند. در برنامه‌های درمانی پیشنهادی، GPT-3.5 در ۹۰٪ موارد با توصیه‌های MDT مطابقت نزدیکی داشت. در مقابل، GPT-4 طیف متنوع‌تری از گزینه‌های درمانی را ارائه داد، به‌طور متوسط ۵/۱ گزینه بیشتر از GPT-3.5 و به‌طور قابل توجهی از محدوده MDT فراتر رفت [۴۵].

برای درمان شخصی‌سازی‌شده سرطان، مانوئلا و همکاران[4] [۲۴]، چهار LLM را در برنامه‌ریزی درمان برای ۱۰ مورد سرطان ارزیابی کردند. نمرات F1 LLMها (۰/۰۴-۰/۱۹) پایین‌تر از نمرات متخصصان انسانی بود و توصیه‌های آن‌ها به‌راحتی به‌عنوان توصیه‌های تولید شده توسط هوش مصنوعی شناسایی گردید (میانگین، ۷/۵ در مقابل ۲/۰ برای موارد حاشیه‌نویسی شده دستی). LLMها در ارائه بینش‌های مکمل، با حداقل یک LLM که یک توصیه مفید برای هر مورد و حتی دو استراتژی درمانی منحصربه‌فرد و مفید ارائه داد، نویدبخش بودند. با ترکیب توصیه‌های LLMهای مختلف، نمره F1 به ۰/۲۹ بهبود یافت که نشان‌دهنده دامنه بهبود در آینده است. این مطالعه نشان داد که اگرچه LLMها پتانسیل برنامه‌ریزی درمان سرطان را نشان می‌دهند، اما برای مطابقت با عملکرد سطح متخصص، پیشرفت‌های بیشتری لازم است [۲۴]. در یک مطالعه جداگانه، لاوسون و همکاران[5] دریافتند که GPT-3.5 قادر به ارائه توصیه‌های شخصی‌سازی‌شده برای بدخیمی‌های زنان براساس علائم خاص بیمار، مانند توصیه‌هایی برای تخلیه پلور یا آسیت و مراقبت‌های بین‌رشته‌ای است. برای مقابله با مشکلات توضیح درمان به بیماران، یک مطالعه مقطعی، توانایی LLMها

1. Zheng et al.
2. oncology
3. Schmidl et al.
4. Manuela et al.
5. Lawson et al.

را در پاسخ‌گویی به سؤالات بیماران در حوزه انکولوژی پرتودرمانی ارزیابی کرد. نتایج نشان داد که هنگام مقایسه پاسخ‌های خاص به روش‌های درمانی، LLMها ازنظر دقت، کامل بودن و مختصر بودن، عملکرد مشابهی با پاسخ‌های متخصصان نشان دادند [۴۶]. لاوسون و همکاران همچنین در مورد پتانسیل ChatGPT و Bard در توضیح برنامه‌های درمانی نوروانکولوژی بحث کردند و دریافتند که آن‌ها می‌توانند فشار اطلاعاتی را بر کارکنان مراقبت‌های بهداشتی کاهش داده و خودمدیریتی بیمار را افزایش دهند [۴۶].

گوارش

LLMها همچنین به‌دلیل پتانسیل خود در ارزیابی برنامه‌های درمانی در طیف وسیعی از اختلالات دستگاه گوارش، ازجمله بیماری‌های کبدی، بیماری‌های مری و سایر بیماری‌های دستگاه گوارش، مورد ارزیابی قرار گرفته‌اند.

گیوفره و همکاران[1] گزارش دادند که LLMها می‌توانند پروتکل‌های درمانی برای هپاتیت C مزمن همراه با عفونت ویروس نقص ایمنی انسانی (HIV) را با در نظر گرفتن شرایط مختلف بیماران ارائه دهند [۴۷]. هنسون و همکاران[2] دریافتند که GPT-4 در پاسخ به سؤالات مربوط به درمان GERD عملکرد خوبی داشت، به‌طوری که ۹۳/۹٪ از پاسخ‌های آن تقریباً یا کاملاً مناسب و ۷۵/۸٪ حاوی راهنمایی‌های خاص بودند. پاسخ‌های GPT-4 توسط ۹۷/۷٪ از بیماران به‌خوبی درک شد [۴۸]. یک مطالعه نشان داد که GPT-4 در پاسخ به شش سؤال مربوط به درمان از ۱۵ سؤال مربوط به سندرم روده تحریک‌پذیر، ۸۳٪ دقیق بود. از ۲۰ سؤال مربوط به بیماری التهابی روده، GPT-4 برای هفت سؤال مربوط به درمان، ۸۵٪ دقت نشان داد [۴۹].

جی و همکاران[3] علاوه‌بر ارزیابی درمان بیماری‌های دستگاه گوارش با استفاده از مدل‌های بزرگ تجاری عمومی، سعی کردند از RAG برای توسعه "LiVersa"، یک LLM خاص برای بیماری‌های کبدی، استفاده کرده و عملکرد آن را در درمان هپاتیت B و مسائل مربوط به نظارت بر کارسینوم سلول‌های کبدی ارزیابی کنند. آن‌ها دریافتند که خروجی‌های LiVersa دقیق‌تر هستند، اما در مقایسه با خروجی‌های GPT-4، جامعیت و ایمنی کمتری دارند [۵۰].

طب سالمندان[4]

مدیریت بیماری‌های مزمن، به‌ویژه در جامعه‌ای که رو به سالخوردگی است، مسئله‌ای حیاتی است. مطالعات اخیر، پتانسیل LLMها را در ارائه پشتیبانی بالینی برای بیماری‌های مزمن مختلفی که معمولاً افراد مسن را تحت‌تأثیر قرار می‌دهند، بررسی کرده‌اند.

رائو و همکاران[5] کاربرد GPT-3.5 را در مدیریت دارو برای بیماران مسن بررسی کردند. آن‌ها دریافتند که GPT-3.5 به‌طور مداوم کاهش دارو را در بیماران مسن‌تر با فعالیت‌های روزمره زندگی مختل‌شده و بدون سابقه بیماری قلبی-عروقی توصیه می‌کند. در مقابل، برای بیماران مسن‌تر با سابقه حوادث قلبی، GPT-3.5 احتیاط بیشتری در تنظیم رژیم‌های دارویی نشان داد، به‌طوری که ۵۶٪از پاسخ‌ها توصیه به عدم کاهش دارو داشتند [۵۱]. آیمتیاز و همکاران[6] با گسترش دامنه فراتر از مدیریت دارو، گزارش دادند که GPT-3.5 اغلب در ارائه گزینه‌های درمانی برای بیماری انسدادی مزمن ریوی، که یک بیماری شایع در سالمندان است، از Microsoft Bing بهتر عمل می‌کند، اگرچه نتایج نشان‌دهنده نیاز به به‌روزرسانی برای هماهنگی با آخرین دستورالعمل‌های پزشکی است. به‌طور مشابه، مطالعه دیگری

1. Giuffrè et al.
2. Henson et al.
3. Jin et al.
4. geriatrics
5. Rao et al.
6. Imtiaz et al.

در مورد نزدیک‌بینی، که می‌تواند با افزایش سن پیشرفت کند، نشان داد که GPT-4 در ارائه اطلاعات درمانی و پیشگیری، به‌طور قابل توجهی از GPT-3.5 و Google Bard بهتر عمل می‌کند [۵۲]. در مجموع، این نتایج نشان داد که LLMها ممکن است پشتیبانی بالینی مفیدی برای مدیریت بیماری‌های مزمن که معمولاً جمعیت سالمندان را تحت‌تأثیر قرار می‌دهند، ارائه دهند.

یک مطالعه قابل توجه در مورد دیابت، بیماری شایع در میان سالمندان، ادغام LLMها را با DL مبتنی بر تصویر بررسی کرد. این مطالعه اثربخشی سیستمی به نام DeepDR-LLM را در ۴۸۷ بیمار که به‌تازگی به دیابت و رتینوپاتی دیابتی قابل ارجاع (DR) مبتلا شده بودند، ارزیابی کرد. محققان، نتایج بین بیمارانی که به‌تنهایی از پزشکان مراقبت‌های اولیه (PCP) مراقبت دریافت می‌کردند و بیمارانی که از PCPهای تقویت‌شده با DeepDR-LLM بهره‌مند می‌شدند را مقایسه کردند. تمرکز اصلی بر پایبندی به مدیریت دیابت، ازجمله رفتارهای خودمدیریتی و رعایت ارجاع DR بود. نتایج نشان داد که بیماران در گروه PCP + DeepDR-LLM رفتارهای خودمدیریتی بهتری نشان دادند و احتمال بیشتری داشت که به ارجاعات DR پایبند باشند [۵۳]. مطالعه فوق، چشم‌اندازهای امیدوارکننده‌ای را برای مراقبت‌های بهداشتی مبتنی بر هوش مصنوعی در بهبود مدیریت بیماری‌های مزمن و پایبندی به درمان در جمعیت سالمندان نشان داد.

بیماری‌های عفونی[1]

استفاده از داروهای حساس به پاتوژن به‌منظور درمان بیماری‌های عفونی بسیار مهم است. مطالعات همچنین کاربردهای بالقوه LLMها را در این زمینه گزارش کرده‌اند. محققان دریافتند که توصیه‌های تجربی درمان ضدمیکروبی که به‌صورت آینده‌نگر توسط GPT-4 برای بیمارانی با کشت خون مثبت ارائه شده، در ۶۴٪ موارد مناسب تلقی شده، اما در ۲٪ موارد باعث آسیب گردیده است. برای درمان نهایی آنتی‌بیوتیک، پیشنهادات GPT-4 در ۹۱٪ از بیماران کافی تلقی شد، اما همچنان در ۵٪ موارد به اثرات مضر منجر گردید [۵۴]. مطالعه فوق، محدودیت‌ها و خطرات بالقوه استفاده از هوش مصنوعی به‌منظور مشاوره‌های پزشکی را برجسته کرد.

ارتوپدی[2]

تشخیص دقیق و درمان مناسب بیماری‌های ارتوپدی به‌منظور بهبودی بیمار و رفاه طولانی‌مدت ضروری است. عملکرد LLMها در درمان و مدیریت بیماری‌های مفصلی و آرتروز ارزیابی شده است. یک مطالعه، عملکرد GPT-4 را در ارائه توصیه‌های درمانی برای بیماری‌های ارتوپدی رایج زانو و شانه براساس ۲۰ گزارش MRI ارزیابی کرد. پزشکان هنگام ارزیابی سودمندی بالینی و ارتباط توصیه‌های درمانی، به‌ترتیب ۶۰٪ و ۲۰٪ از توصیه‌ها «موافق» و «کاملاً موافق» گزارش دادند. بااین‌حال، پزشکان همچنین خاطرنشان کردند که GPT-4 به‌طور کامل شرایط خاص بیمار و فوریت درمان را در نظر نمی‌گیرد [۵۵]. مطالعه دیگری گزارش داد که یک ابزار مبتنی بر GPT-4 با استفاده از RAG و دستورالعمل‌های راهنما، در مدیریت آرتروز به‌طور قابل توجهی از سایر LLMهای عمومی مانند GPT-4 و GPT-3.5 پیشی گرفته است [۵۶].

1. infectious diseases
2. orthopedics

محدودیت‌ها

مسئله کلیدی مرتبط با استفاده از LLMها در عمل بالینی، دقت سؤالات و پاسخ‌های آن‌هاست. توهم در LLMها ممکن است اطلاعات نادرست یا حتی مضر تولید کند، اگرچه این مسئله در تکرارهای جدیدتر مدل‌ها بهبود یافته است. با این وجود، انحراف عملکرد مشاهده شده در برخی از LLMها هنوز فاقد توضیح نظری کافی است و به‌طور بالقوه نگرانی‌هایی را در مورد ایمنی و پایداری LLMها در کاربردهای بالینی ایجاد می‌کند [۵۷]. درحال‌حاضر، MLLMها مانند GPT-4V توانایی تفسیر تصاویر اشعه ایکس پزشکی را نشان می‌دهند، اما هنوز فاقد قابلیت‌های کافی با برخی از روش‌های تصویربرداری مانند MRI هستند. این محدودیت احتمالاً به‌دلیل داده‌های آموزشی ناکافی و غیرمتنوع و معماری‌های مدل بهینه نشده برای پردازش تصویر است. علاوه‌بر این، MLLMها تمایل دارند اطلاعات متنی را بر ورودی‌های بصری اولویت دهند، که دقت تشخیصی آن‌ها را در وظایف پزشکی مبتنی بر تصویر محدود می‌کند. چالش‌ها در ادغام مؤثر اطلاعات چندوجهی و محدودیت‌های منابع محاسباتی، توانایی این مدل‌ها را در درک تصاویر بیشتر مختل می‌کند. برای افزایش کاربرد MLLMها در تفسیر تصاویر پزشکی، بهینه‌سازی و تنظیم دقیق ماژول‌های پردازش ویژگی‌های تصویر براساس داده‌های چندوجهی بسیار مهم است. پیشرفت‌های آینده باید بر بهبود ادغام داده‌ها و قابلیت‌های سنتز اطلاعات بین‌وجهی متمرکز شوند تا بر محدودیت‌های فعلی غلبه کرده و به درک بهتری از تصاویر پزشکی برای تشخیص‌های دقیق‌تر دست یابند. علاوه‌بر این، گزارش شده است که LLMها گاهی توانایی استدلال ناکافی را از خود نشان دهند که ممکن است عملکرد آن‌ها را در موارد چالش‌برانگیز محدود کند [۵۸]. LLMها اغلب به‌دلیل اطلاعات محدود ارائه‌شده توسط بیماران در هر نوبت، توانایی سازمان‌دهی سؤالات چندمرحله‌ای برای اهداف تشخیصی را ندارند، بنابراین کاربرد آن‌ها در مشاوره پزشکی محدود می‌شود. تشخیص و درمان بالینی نیاز به استدلال بالا و قابلیت‌های منطقی برای تجزیه‌وتحلیل علائم چندسیستمی و پیشرفت بیماری دارد. محدودیت‌های LLMها در استدلال همچنین ممکن است منجر به مشکلاتی در تشخیص صحت توضیحات بیمار یا ایجاد اتکای بیش از حد به نتایج تشخیصی قبلی شود. خوشبختانه، LLMهایی با قابلیت‌های استدلال بهبودیافته به‌طور مداوم در حال ظهور هستند. مدل OpenAI O1 که اخیراً منتشر شده است، نمونه‌ای از این روند است. انتظار می‌رود چنین پیشرفت‌هایی پیش‌بینی تشخیصی و برنامه‌ریزی درمان را در تحقیقات آینده بیشتر بهبود بخشد [۵۹].

همچنین اثربخشی LLMها در تشخیص پزشکی به‌شدت به دسترسی به اطلاعات بالینی جامع، دقیق و به‌موقع وابسته است. بااین‌حال، چشم‌انداز فعلی مراقبت‌های بهداشتی چالش‌های قابل توجهی را در این زمینه ایجاد می‌کند. آزمایش‌های آزمایشگاهی و رادیولوژیکی متعدد، انواع متنوعی از داده‌های پزشکی را تولید می‌کنند. وجود سیستم‌های اطلاعات پزشکی متفاوت و اغلب غیرقابل تعامل، سیلوهای اطلاعاتی را در داخل و بین مؤسسات مراقبت‌های بهداشتی ایجاد می‌کند [۶۰]. این پراکندگی داده‌های بالینی همچنین مانع قابل توجهی برای ادغام یکپارچه LLMها در محیط‌های مراقبت‌های بهداشتی ایجاد می‌کند. برای پرداختن به چالش‌های اجرای مؤثر LLMها در تشخیص پزشکی، یک استراتژی چندوجهی و جامع مورد نیاز است. وظیفه اصلی، ارتقای استانداردهای تبادل داده‌های یکپارچه، مکانیسم‌های همگام‌سازی بلادرنگ و رابط‌های برنامه‌نویسی کاربردی باز (API) برای دستیابی به ادغام یکپارچه سیستم‌های پرونده الکترونیکی سلامت موجود است. هم‌زمان، تقویت امنیت داده‌ها و اقدامات حفاظت از حریم خصوصی نیز بسیار مهم است. تحقیقات نشان داده است که LLMها پتانسیل سازمان‌دهی گزارش‌ها و سوابق پزشکی را دارند [۶۱]. بنابراین، می‌توان از آن‌ها برای ساختاردهی اطلاعات پزشکی پیچیده استفاده کرد. بهینه‌سازی رابط‌های انسان و ماشین، مانند دستیاران صوتی هوشمند و فناوری واقعیت افزوده (AR)، به‌طور قابل توجهی کارایی همکاری بین پزشکان و LLMها را افزایش خواهد داد. علاوه‌بر این، ارتقای استانداردسازی صنعت، بهبود چارچوب‌های قانونی

و اخلاقی و افزایش آموزش‌های مرتبط با هوش مصنوعی برای پرسنل پزشکی ضروری است. بااین‌حال، اجرای این اقدامات مستلزم تلاش‌های مشترک ازسوی مؤسسات پزشکی، شرکت‌های فناوری، نهادهای نظارتی و دانشگاه‌ها است تا از پتانسیل LLMها در تشخیص پزشکی به‌طور کامل بهره برده شود و در نهایت کیفیت و کارایی خدمات مراقبت‌های بهداشتی بهبود یابد [۶۲].

فقدان معیارهای استاندارد ارزیابی

همان‌طور که در یک بررسی سیستماتیک توسط مائورو و همکارانش برجسته شده است، چشم‌انداز فعلی LLMها چالش‌های قابل توجهی را ازنظر مقایسه و ارزیابی آن‌ها در مطالعات مختلف تشخیصی بالینی نشان می‌دهد. یکی از عوامل مهم مؤثر در این موضوع، فقدان معیارهای استاندارد ارزیابی است. تنوع ذاتی روش‌های ارزیابی برای بیماری‌های مختلف و مراحل پیشرفت بالینی، چالش‌های قابل توجهی را در این زمینه ایجاد می‌کند. معیارهای ارزیابی که به متخصصان انسانی متکی هستند، عمدتاً شامل دقت تشخیصی، خوانایی یا نمرات ذهنی براساس سناریوهای کاری خاص هستند. مطالعاتی که معیارهایی مبتنی بر ارزیابی متخصصان دارند، اغلب به‌دلیل کمبود متخصصان باتجربه، به‌دلیل اندازه‌های کوچک نمونه برای آزمایش LLM محدود می‌شوند [۶۳]. در همین حال، معیارهای ارزیابی مبتنی بر انسان ممکن است سوگیری و تغییرپذیری را در ارزیابی‌ها ایجاد کنند. ازسوی‌دیگر، معیارهای ارزیابی خودکار مبتنی بر هم‌پوشانی‌های واژگانی، مانند ارزیابی دوزبانه جایگزین (BLEU) و معیارهای ارزیابی ترجمه با ترتیب صریح (METEOR)، با همبستگی ضعیفی با ارزیابی‌های انسانی دست و پنجه نرم می‌کنند [۶۴]. این اختلافات ممکن است سؤالاتی را در مورد اثربخشی معیارهای ارزیابی خودکار ایجاد کند. علاوه‌بر این، طرح‌های آزمایشی و معیارهای ارزیابی متفاوت به‌کاررفته در مطالعات، مقایسه بین یافته‌های تحقیقاتی را دشوار می‌کند. این فقدان استانداردسازی، نیاز به چارچوب‌های ارزیابی قوی‌تر و جهانی‌تر را برجسته می‌کند.

برای پرداختن به این چالش‌ها و افزایش بیشتر قابلیت اطمینان و کاربرد LLMها در تشخیص پزشکی، توسعه مجموعه داده‌های استانداردتر سؤالات آزمون و افزایش RCTهایی با کیفیت بالا با استفاده از روش‌های ارزیابی سازگار، مسیرهای امیدوارکننده‌ای را نشان می‌دهند. این رویکردها می‌توانند به‌طور بالقوه محدودیت‌های فعلی را برطرف سازند، شواهد قوی‌تر و قابل مقایسه‌تری برای اثربخشی LLMها در محیط‌های مراقبت‌های بهداشتی ارائه دهند و مقایسه‌های متقابل معنادارتری را تسهیل کنند و در نتیجه ادغام آن‌ها را در عمل بالینی تسریع بخشند.

چالش‌ها، خطرات موجود و ملاحظات اخلاقی

نگرانی‌ها در مورد هوش مصنوعی و امنیت داده‌های بیمار از حادثه‌ای ناشی می‌شود که در سال ۲۰۱۵ در سیستم سلامت دانشگاه کالیفرنیا، لس‌آنجلس (UCLA) رخ داد. در این حادثه، سیستم سلامت UCLA دچار نقض داده‌ها شد که در آن هکرها به داده‌های مهم بیمار، ازجمله اطلاعات پزشکی حساس مانند تشخیص‌ها، سوابق درمان و شناسه‌های شخصی، دسترسی غیرمجاز پیدا کردند. این نقض تقریباً ۴/۵ میلیون نفر را تحت‌تأثیر قرار داد و آسیب‌پذیری‌های موجود در پروتکل‌های امنیت داده‌های سیستم‌های مراقبت‌های بهداشتی را برجسته کرد. این حادثه نگرانی‌هایی را در مورد امنیت داده‌های بیمار در محیط‌های مراقبت‌های بهداشتی ایجاد کرد، عمدتاً به این دلیل که برنامه‌های هوش مصنوعی به‌طور فزاینده‌ای برای آموزش الگوریتم‌ها و تصمیم‌گیری‌های بالینی به مجموعه داده‌های بزرگ متکی هستند. نگرانی‌های اصلی دیگر شامل این بود که چه کسی باید از آن‌ها استفاده کند؟ چگونه باید از آن ها استفاده شود؟ و چه کسی باید مسئول و پاسخ‌گو باشد؟

همان‌طور که در بخش‌های قبلی عنوان شد، پاسخ‌های ارائه‌شده در تشخیص تصاویر سی‌تی‌اسکن قفسه سینه و توصیه‌های درمان آنتی‌بیوتیکی مستعد خطا هستند. اگرچه برخی از LLMها در مشاوره‌های پزشکی عملکرد خوبی داشته‌اند، کاربرانی که پیشینه دانش بالینی ندارند، ممکن است هنگام استفاده از LLMها برای دریافت مشاوره مراقبت‌های بهداشتی، با عوارض جانبی سلامتی و خطرات قانونی مواجه شوند. علاوه‌بر این، گزارش شده است که GPT-4 به‌طور بالقوه اطلاعات کاربر را فاش می‌کند و نگرانی‌هایی را در مورد ایمنی اطلاعات ایجاد می‌کند [۶۵]. برخی از محققان، استقرار مدل‌های محلی برای حاشیه‌نویسی گزارش‌های رادیولوژی را برای رفع محدودیت‌های حریم خصوصی بیمار بررسی کرده‌اند. بااین‌حال، هزینه و الزامات تکنولوژیکی استقرار و نگهداری LLMهای با کارایی بالا، استقرار آن‌ها را به‌صورت محلی در بیمارستان‌ها برای محافظت از حریم خصوصی اطلاعات بیمار چالش‌برانگیز می‌کند. با توجه به این مسائل، رویکرد امیدوارکننده‌ای شامل استفاده از ارائه دهندگان خدمات شخص ثالث توانمند می‌باشد که تحت چارچوب‌های نظارتی قوی فعالیت کنند. این ارائه دهندگان می‌توانند تکنیک‌های یادگیری فدرال یا اقدامات محرمانه سخت‌گیرانه را اجرا کرده و به‌طور بالقوه نگرانی‌های امنیتی داده‌ها را کاهش دهند و در عین حال از قدرت مدل‌های زبانی پیشرفته برای برنامه‌های مراقبت‌های بهداشتی استفاده نمایند [۶۶].

نتیجه‌گیری

اگر سیستم‌های هوش مصنوعی پزشکی دارای توانایی تعامل مکالمه‌ای بهتری باشند، با تکیه بر دانش پزشکی در مقیاس بزرگ و در عین حال با سطوح مناسبی از همدلی و اعتماد ارتباط برقرار کنند، می‌توان کاربرد آن‌ها را تا حد زیادی بهبود بخشید. این کار، قابلیت‌های بالقوه‌ی بالای سیستم‌های هوش مصنوعی مبتنی بر LLM را برای محیط‌هایی که شامل گرفتن شرح حال بالینی و گفت‌وگوی تشخیصی هستند، نشان می‌دهد.

تحقیقات پزشکی سالانه حجم زیادی از دانش و بینش جدید تولید می‌کند که می‌تواند منجر به منسوخ شدن یافته‌های پزشکی قبلی شود. LLMها می‌توانند به‌طور مداوم پایگاه دانش خود را به‌روز کنند و از آخرین تحقیقات پزشکی مطلع شوند و در نتیجه مشاوره پزشکی پیشرفته ارائه دهند.

ظهور و شیوع بیماری‌های عفونی جدید در سال‌های اخیر، مانند کووید-۱۹ و تب خونریزی‌دهنده ابولا، سلامت عمومی را به‌شدت تهدید کرده و باعث نگرانی سازمان‌های بهداشت عمومی جهانی در مورد سناریوهای آینده «بیماری X» شده است. بنابراین امروزه به‌روزرسانی دانش پزشکی و پرداختن به بیماری‌های همه‌گیر نوظهور به ضرورتی غیرقابل اجتناب تبدیل شده است. در این زمینه، سیستم‌های مبتنی بر به‌روزرسانی‌های سریع پایگاه‌های دانش موجود و استقرار LLMهای جدید، نویدبخش ابزارهای آموزشی دانش بهداشت عمومی یا کمک‌های تشخیصی در آینده هستند. چنین سیستم‌هایی می‌توانند به‌طور بالقوه اطلاعات و توصیه‌های پزشکی مرتبط با «بیماری X» را برای کمک به مردم در مقابله سریع با بیماری‌های همه‌گیر آینده ارائه دهند. LLMها به‌عنوان یک فناوری متحول‌کننده برای تشخیص و درمان بیماری‌های مختلف ظهور کرده‌اند. امید است تحت چارچوب‌های نظارتی معقول و همکاری چندرشته‌ای، LLMها بتوانند به تکامل خود ادامه دهند و خدمات تشخیصی و درمانی بالینی عادلانه‌تر، ایمن‌تر، باکیفیت‌تر و در دسترس‌تری را در آینده‌ای نزدیک ارائه دهند.

منابع

1. Tariq M, Hayat Y, Hussain A, Tariq A, Rasool S. Principles and perspectives in medical diagnostic systems employing artificial intelligence (AI) algorithms. Int Res J Econ Manag Stud. 2025;3(1):376-398.
2. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Predicting clinical events via recurrent neural networks. In: Proc Mach Learn Healthcare Conf. 2016:301-318.
3. Fu Y, Peng H, Khot T, Lapata M. Improving language model negotiation with self-play and in-context learning from AI feedback. 2023.
4. Levine D. History taking is a complex skill. Br Med J. 2017.
5. Kaplan A, Haenlein M. Siri, Siri, in my hand: who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Bus Horiz. 2019;62(1):15-25.
6. De Bruyn A, Viswanathan V, Beh Y, Brock J, von Wangenheim F. Artificial intelligence and marketing: pitfalls and opportunities. J Interact Mark. 2020;51:91-105.
7. Montenegro JLZ, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. 2019;129:56-67.
8. Suta P, Lan X, Wu B, Mongkolnam P, Chan J. An overview of machine learning in chatbots. Int J Mech Eng Robot Res. 2020;9(4):502-510.
9. Gentsch P. A bluffer's guide to AI, algorithmics and big data. In: AI in marketing, sales, and service. Cham: Palgrave Macmillan; 2019:11-24.
10. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J. 2019;6(2):94-98.
11. von Krogh G. Artificial intelligence in organizations: new opportunities for phenomenon-based theorizing. Acad Manag Discov. 2018;4(4):404-409.
12. Fadhil A. A conversational interface to improve medication adherence: towards AI support in patient's treatment. 2018.
13. Martínez-Miranda J, Martínez A, Ramos R, Aguilar H, Jiménez L, Arias H. Assessment of users' acceptability of a mobile-based embodied conversational agent for the prevention and detection of suicidal behaviour. J Med Syst. 2019;43(8):246.
14. Bickmore T, Pusateri A, Kimani E, Paasche-Orlow M, Trinh H, Magnani J. Managing chronic conditions with a smartphone-based conversational virtual agent. In: Proc Int Conf Intell Virtual Agents. 2018:119-124.
15. Singhal K. Large language models encode clinical knowledge. 2023:172-180.
16. Singhal K. Toward expert-level medical question answering with large language models. Nat Med. 2025;31:943-950.
17. Nori H. Can generalist foundation models outcompete special-purpose tuning? Case study in medicine. arXiv. 2023.
18. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17:195.
19. Zhang N, Sun Z, Xie Y, Wu H, Li C. The latest version ChatGPT powered by GPT-4o: what will it bring to the medical field? Int J Surg. 2024;110:6018-6019.
20. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health. 2023.
21. Gallegos IO. Bias and fairness in large language models: a survey. Comput Linguist. 2024;50:1-79.
22. Yang X, Li T, Su Q. Application of large language models in disease diagnosis and treatment. Chin Med J. 2025;138(2):130-142.
23. Goh E, Gallo R, Strong E, Weng Y, Kerman H, Freed J. Large language model influence on management reasoning: a randomized controlled trial. medRxiv. 2024.
24. Benary M, Wang XD, Schmidt M, Soll D, Hilfenhaus G, Nassir M. Leveraging large language models for decision support in personalized oncology. JAMA Netw Open. 2023;6:e2343689.
25. Li K, Ruan G, Liu S, Xu T, Guan K, Li J. Eosinophilic gastroenteritis: pathogenesis, diagnosis, and treatment. Chin Med J. 2023;136:899-909.
26. Lahat A, Shachar E, Avidan B, Glicksberg B, Klang E. Evaluating the utility of a large language model in answering common patients' gastrointestinal health-related questions: are we there yet? Diagnostics (Basel). 2023;13:1950.
27. Koga S, Martin NB, Dickson DW. Evaluating the performance of large language models: ChatGPT and Google Bard in generating differential diagnoses in clinicopathological conferences of neurodegenerative disorders. Brain Pathol. 2024;34:e13207.

28. Wang C, Liu S, Li A, Liu J. Text dialogue analysis for primary screening of mild cognitive impairment: development and validation study. J Med Internet Res. 2023;25:e51501.
29. Berg HT, van Bakel B, van de Wouw L, Jie KE, Schipper A, Jansen H. ChatGPT and generating a differential diagnosis early in an emergency department presentation. Ann Emerg Med. 2024;83:83-86.
30. Heston TF, Lewis LM. ChatGPT provides inconsistent risk-stratification of patients with atraumatic chest pain. PLoS One. 2024;19:e0301854.
31. Perret J, Schmid A. Application of OpenAI GPT-4 for the retrospective detection of catheter-associated urinary tract infections in a fictitious and curated patient data set. Infect Control Hosp Epidemiol. 2024,45:96-99.
32. Boligarla S, Laison EKE, Li J, Mahadevan R, Ng A, Lin Y, et al. Leveraging machine learning approaches for predicting potential Lyme disease cases and incidence rates in the United States using Twitter. BMC Med Inform Decis Mak. 2023,23:217.
33. Hen C, Li L, Beetz M, Banerjee A, Gupta R, Grau V. Large language model-informed ECG dual attention network for heart failure risk prediction. arXiv Preprint. 2024.
34. Yu H, Guo P, Sano A. Zero-shot ECG diagnosis with large language models and retrieval-augmented generation. Mach Learn Health. 2023,225:650-663.
35. Topol EJ. As artificial intelligence goes multimodal, medical applications multiply. Science. 2023.
36. Alryalat SA, Musleh AM, Kahook MY. Evaluating the strengths and limitations of multimodal ChatGPT-4 in detecting glaucoma using fundus images. Front Ophthalmol (Lausanne). 2024,4:1387190
37. Hirosawa T, Harada Y, Tokumasu K, Ito T, Suzuki T, Shimizu T. Evaluating chatgpt-4's diagnostic accuracy: Impact of visual data integration. JMIR Med Inform. 2024,12:e55627.
38. Horiuchi D, Tatekawa H, Oura T, Oue S, Walston SL, Takita H, et al. Comparing the diagnostic performance of GPT-4-based ChatGPT, GPT-4v-based ChatGPT, and radiologists in challenging neuroradiology cases. Clin Neuroradiol. 2024,34:779-787.
39. Bani-Harouni D, Navab N, Keicher M, eds. MAGDA: Multi-agent guideline-driven diagnostic assistance. In: International workshop on foundation models for general medical AI. Springer: 2024
40. Sultan LR, Mohamed MK, Andronikou S. ChatGPT-4: A breakthrough in ultrasound image analysis. Oxford University Press; 2024.
41. Wu SH, Tong WJ, Li MD, Hu HT, Lu XZ, Huang ZR, et al. Collaborative enhancement of consistency and accuracy in US diagnosis of thyroid nodules using large language models. Radiology. 2024,310:e232255.
42. Zheng S, Cui X, Sun Y, Li J, Li H, Zhang Y, Benchmarking pathCLIP for pathology image analysis. J Imaging Inform Med. 2024.
43. Griewing S, Gremke N, Wagner U, Lingenfelder M, Kuhn S, Boekhoff J. Challenging ChatGPT 3.5 in senology-an assessment of concordance with breast cancer tumor board decision making. J Pers Med. 2023,13:1502.
44. Schmidl B, Hütten T, Pigorsch S, Stögbauer F, Hoch CC, Hussain T, et al. Assessing the role of advanced artificial intelligence as a tool in multidisciplinary tumor board decision-making for primary head and neck cancer cases. Front Oncol. 2024,14:1353031
45. Yalamanchili A, Sengupta B, Song J, Lim S, Thomas TO, Mittal BB, et al. Quality of large language model responses to radiation oncology patient care questions. JAMA Netw Open. 2024,7:e244630.
46. Lawson McLean A, Wu Y, Lawson McLean AC, Hristidis V. Large language models as decision aids in neuro-oncology: A review of shared decision-making applications. J Cancer Res Clin Oncol. 2024,150:139.
47. Giuffrè M, Kresevic S, Pugliese N, You K, Shung DL. Optimizing large language models in digestive disease: Strategies and challenges to improve clinical outcomes. Liver Int. 2024,44:2114-2124.
48. Henson JB, Glissen Brown JR, Lee JP, Patel A, Leiman DA. Evaluation of the potential utility of an artificial intelligence chatbot in gastroesophageal reflux disease management. Am J Gastroenterol. 2023,118:2276-2279.
49. Kerbage A, Kassab J, El Dahdah J, Burke CA, Achkar JP, Rouphael C. Accuracy of ChatGPT in common gastrointestinal diseases: Impact for patients and providers. Clin Gastroenterol Hepatol. 2024,22:1323-1325e3.

50. Ge J, Sun S, Owens J, Galvez V, Gologorskaya O, Lai JC, Development of a liver disease-specific large language model chat interface using retrieval-augmented generation. Hepatology. 2024,80:1158-1168.
51. Rao A, Kim J, Lie W, Pang M, Fuh L, Dreyer KJ, Proactive polypharmacy management using large language models: Opportunities to enhance geriatric care. J Med Syst. 2024,48:41.
52. Imtiaz A, King J, Holmes S, Gupta A, Bafadhel M, Melcher ML, ChatGPT versus Bing: A clinician assessment of the accuracy of AI platforms when responding to COPD questions. Eur Respir J. 2024,63:400163.
53. Howard A, Hope W, Gerada A. ChatGPT and antimicrobial advice: The end of the consulting infection doctor? Lancet Infect Dis. 2023,23:405-406.
54. Maillard A, Micheli G, Lefevre L, Guyonnet C, Poyart C, Canouï E, Can chatbot artificial intelligence replace infectious diseases physicians in the management of bloodstream infections? A prospective cohort study. Clin Infect Dis. 2024,;78:825-832.
55. Truhn D, Weber CD, Braun BJ, Bressem K, Kather JN, Kuhl C, A pilot study on the efficacy of GPT-4 in providing orthopedic treatment recommendations from MRI reports. Sci Rep. 2023,13:20159.
56. Chen X, Wang L, You M, Liu W, Fu Y, Xu J, Evaluating and enhancing large language models' performance in domain-specific medicine: Development and usability study with DocOA. J Med Internet Res. 2024,26:e58158.
57. Li D, Gupta K, Bhaduri M, Sathiadoss P, Bhatnagar S, Chong J. Comparing GPT-3.5 and GPT-4 accuracy and drift in radiology diagnosis please cases. Radiology. 2024,310:e232411
58. Liu X, Wu Z, Wu X, Lu P, Chang KW, Feng Y. Are LLMs capable of data-based statistical and causal reasoning? Benchmarking advanced quantitative reasoning with data. arXiv Preprint. 2024.
59. Chen Y, Wang Z, Xing X, Xu Z, Fang K, Wang J, Bianque: Balancing the questioning and suggestion ability of health LLMs with multi-turn health conversations polished by ChatGPT. arXiv Preprint. 2023.
60. Williams CYK, Zack T, Miao BY, Sushil M, Wang M, Kornblith AE, Use of a large language model to assess clinical acuity of adults in the emergency department. JAMA Netw Open. 2024,7:e248895
61. Huang J, Yang DM, Rong R, Nezafati K, Treager C, Chi Z, A critical assessment of using ChatGPT for extracting structured data from clinical notes. NPJ Digit Med. 2024,7:106.
62. Amin KS, Davis MA, Doshi R, Haims AH, Khosla P, Forman HP. Accuracy of ChatGPT, Google Bard, and Microsoft Bing for simplifying radiology reports. Radiology. 2023,309:e232561.
63. Shea YF, Lee CMY, Ip WCT, Luk DWA, Wong SSW. Use of GPT-4 to analyze medical records of patients with extensive investigations and delayed diagnosis. JAMA Netw Open. 2023,6:e2325000.
64. Tang L, Sun Z, Idnay B, Nestor JG, Soroush A, Elias PA, Evaluating large language models on medical evidence summarization. NPJ Digit Med. 2023,6:158.
65. Pelrine K, Taufeeque M, Zając M, McLean E, Gleave A. Exploiting novel gpt-4 apis. arXiv Preprint. 2023.
66. Mukherjee P, Hou B, Lanfredi RB, Summers RM. Feasibility of using the privacy-preserving large language model vicuna for labeling radiology reports. Radiology. 2023,309:e231147.

استناد به این مقاله: فرج‌زاده، سولماز. (۱۴۰۵). چشم‌اندازهای امیدبخش تشخیص و درمان بیماری‌ها در کمک به پزشکان از طریق به‌کارگیری هوش مصنوعی مبتنی بر مدل‌های زبان بزرگ (LLMs). *فصلنامه پیشرفت‌های مهندسی در حوزه‌ی پزشکی و مواد*، ۱(۴)، ۳۲-۵۳.