

مهندسی پرامپت: چگونه با دستیارهای هوش مصنوعی حرفه ای تر تعامل کنیم فاطمه عرب زاده، سهیلا شرف زاده،



صرفه جویی در زمان، و قابلیت استفاده مؤثر در کاربردهای متنوع مانند تحلیل داده‌های پیچیده، تولید محتوای تخصصی، پشتیبانی فنی، و حتی توسعه نرم‌افزار است. علاوه بر این، درک محدودیت‌های مدل‌های زبانی و بهینه‌سازی مستمر پرامپت‌ها از طریق آزمون و خطا، به کاربران امکان می‌دهد تا به نتایج مطلوب‌تر و پایدارتری دست یابند. این مهارت در عصر هوش مصنوعی به یک ضرورت برای حرفه‌ای‌ها در صنایع مختلف تبدیل شده است.

- فاطمه عرب زاده - arabzadeh_f@ymail.com - کارشناسی ارشد یادگیری الکترونیکی
- سهیلا شرف زاده - ssharafzadeh@gmail.com - کارشناسی ارشد تکنولوژی آموزشی

--

--

مهندسی پرامپت: چگونه با دستیارهای هوش مصنوعی حرفه‌ای‌تر تعامل کنیم

چکیده:

کلیدواژه‌ها: مهندسی پرامپت، هوش مصنوعی، تعامل حرفه‌ای، بهینه‌سازی ورودی، مدل‌های زبانی

مهندسی پرامپت به عنوان یک مهارت تخصصی، با هدف طراحی و بهینه‌سازی ورودی‌های دقیق و هدفمند برای دستیارهای هوش مصنوعی، به منظور تولید پاسخ‌های مرتبط، دقیق و متناسب با نیازهای کاربر توسعه یافته است. این فرآیند با بهره‌گیری از تکنیک‌هایی نظیر ارائه زمینه جامع، استفاده از دستورات صریح و شفاف، به کارگیری مثال‌های هدایت‌کننده و تنظیم پارامترهای مدل مانند دما، طول پاسخ و سطح خلاقیت، تعامل حرفه‌ای و کارآمد با مدل‌های زبانی را تسهیل می‌کند. هدف اصلی مهندسی پرامپت، هدایت مدل به سمت تولید خروجی‌هایی با کیفیت بالا، کاهش ابهام در پاسخ‌ها، و افزایش کارایی در حل مسائل پیچیده است.

مقدمه:

نتایج این رویکرد شامل بهبود دقت پاسخ‌ها،

از تکنیک‌هایی مانند زنجیره تفکر «Chain-of-Thought Prompting» می‌تواند عملکرد مدل‌ها را در وظایف پیچیده بهبود بخشد «Wei et al., 2022». مهندسی پرامپت نیازمند تسلط بر اصول طراحی پرامپت، درک عمیق معماری مدل‌ها و انجام آزمون و خطای مستمر است. با توجه به رشد کاربردهای هوش مصنوعی در صنایع مختلف، این رشته به یک ضرورت برای حرفه‌ای‌ها و پژوهشگران تبدیل شده و انتظار می‌رود نقش کلیدی در توسعه فناوری‌های هوشمند ایفا کند.

مبانی نظری

انواع پرامپت‌ها و کاربرد آن‌ها در تعامل با هوش مصنوعی

پرامپت صفرنمونه‌ای: پرامپت صفرنمونه‌ای به حالتی اطلاق می‌شود که کاربر بدون ارائه مثال یا اطلاعات زمینه‌ای خاص، مستقیماً وظیفه‌ای را از مدل درخواست می‌کند و مدل بر اساس دانش پیش‌آموزش‌دیده خود پاسخ می‌دهد. این نوع پرامپت برای وظایف عمومی مانند پاسخ به سؤالات ساده، ترجمه متون، یا خلاصه‌سازی مناسب است، زیرا نیازی به تنظیمات پیچیده ندارد. به عنوان مثال، پرامپتی مانند "پایتخت فرانسه چیست؟" نیازی به مثال ندارد، زیرا مدل از قبل این اطلاعات را می‌داند. این روش به دلیل سادگی و سرعت در کاربردهای روزمره و سناریوهایی که کاربر زمان یا منابع محدودی دارد، بسیار مؤثر است، اما ممکن است در وظایف تخصصی دقت کمتری داشته باشد. این رویکرد در

مهندسی پرامپت به عنوان یک رشته نوظهور و تخصصی در حوزه هوش مصنوعی و پردازش زبان طبیعی «NLP»، به طراحی، بهینه‌سازی و مدیریت ورودی‌های استراتژیک برای مدل‌های زبانی پیشرفته اختصاص دارد. این مهارت با هدف هدایت مدل‌های زبانی بزرگ مانند GPT-4، LLaMA و BERT به منظور تولید خروجی‌هایی دقیق، مرتبط و متناسب با نیازهای کاربر توسعه یافته است «Brown et al., 2020».

مهندسی پرامپت در تقاطع علوم کامپیوتر، علوم شناختی، طراحی تعامل انسان-ماشین و مهندسی دانش قرار دارد و از تکنیک‌های پیشرفته‌ای نظیر ارائه زمینه‌های جامع، دستورات صریح و ساختارمند، یادگیری چندنمونه‌ای «Few-Shot Learning» و صفرنمونه‌ای «Zero-Shot Learning»، تنظیم پارامترهای مدل مانند دما و حداکثر طول پاسخ، و طراحی زنجیره‌های پرامپت «Prompt Chaining» برای وظایف پیچیده بهره می‌برد «Liu et al., 2023».

هدف اصلی این رشته، بهبود دقت، کارایی و قابلیت اطمینان پاسخ‌های مدل‌های زبانی در کاربردهای متنوع، از جمله تحلیل داده‌های کلان، تولید محتوای تخصصی، استخراج خودکار دانش، ترجمه چندزبانه، پشتیبانی فنی و توسعه سیستم‌های تصمیم‌گیری مبتنی بر هوش مصنوعی است. مهندسی پرامپت نه تنها به کاهش ابهام و خطاهای ناشی از ورودی‌های ناکارآمد کمک می‌کند، بلکه با بهینه‌سازی تعاملات، زمان و منابع محاسباتی را کاهش می‌دهد «Reynolds & McDonell, 2021». با این حال، این رشته با چالش‌هایی مانند مدیریت محدودیت‌های ذاتی مدل‌ها «مانند توهم مدل»، تطبیق با نیازهای چندفرهنگی و چندزبانه، و طراحی پرامپت‌هایی که در وظایف تخصصی پایدار و مقیاس‌پذیر باشند، مواجه است. مطالعات نشان داده‌اند که استفاده

مدل‌های زبانی بزرگ مانند GPT-3 به دلیل توانایی‌های تعمیم‌دهی آن‌ها برجسته شده است » Brown et al., 2020 «.

پرامپت‌های مبتنی بر نقش: پرامپت‌های مبتنی بر نقش، مدل را به پذیرش یک نقش خاص «مانند معلم، مشاور، یا تحلیلگر» ترغیب می‌کنند تا پاسخ‌هایی متناسب با دیدگاه یا تخصص آن نقش تولید کند. این روش در سناریوهایی مانند مشاوره کسب‌وکار، آموزش، یا تولید محتوای تخصصی کاربرد دارد. به عنوان مثال، پرامپتی مانند "در نقش یک مربی زبان، یک برنامه آموزشی برای یادگیری انگلیسی در 30 روز طراحی کن"، مدل را وادار می‌کند تا پاسخ را از دیدگاه یک متخصص آموزش ارائه دهد. این نوع پرامپت به دلیل توانایی در تطبیق پاسخ‌ها با نیازهای خاص کاربر و افزایش حس شخصی‌سازی، در تعاملات حرفه‌ای و خلاقانه بسیار مؤثر است » Reynolds & McDonell, 2021 «.

اشتباهات رایج در طراحی پرامپت

اشتباهات رایج در طراحی پرامپت، شامل استفاده از دستورات مبهم، عدم ارائه زمینه کافی، نادیده گرفتن محدودیت‌ها، و عدم آزمایش و بهینه‌سازی پرامپت است. دستورات مبهم مانند "یه چیزی بنویس" باعث پاسخ‌های کلی یا نامرتب می‌شوند؛ برای اجتناب، از زبان صریح و مشخص مانند "سه مزیت انرژی خورشیدی را در 100 کلمه توضیح بده" استفاده کنید. فقدان زمینه می‌تواند منجر به فرضیات نادرست مدل شود؛ ارائه اطلاعات پس‌زمینه‌ای مرتبط این مشکل را حل می‌کند. نادیده گرفتن محدودیت‌هایی مانند طول یا فرمت، خروجی‌های نامنظم تولید می‌کند؛ تعیین معیارهای دقیق از این امر جلوگیری می‌کند. در نهایت، عدم آزمایش پرامپت مانع بهبود آن می‌شود؛ با آزمون و خطا و تحلیل خروجی‌ها،

پرامپت چندنمونه‌ای: پرامپت چندنمونه‌ای شامل ارائه چند نمونه ورودی-خروجی در پرامپت است تا مدل الگوی پاسخ‌دهی را یاد بگیرد و خروجی‌های مشابهی تولید کند. این روش برای وظایفی که نیاز به سبک یا فرمت خاصی دارند، مانند تولید شعار تبلیغاتی، طبقه‌بندی متون، یا پاسخ به سؤالات تخصصی، کاربرد دارد. به عنوان مثال، با ارائه چند نمونه شعار برای برندهای مختلف، مدل می‌تواند شعاری متناسب برای یک برند جدید تولید کند. این نوع پرامپت به‌ویژه در سناریوهایی که داده‌های آموزشی محدود است یا کاربر می‌خواهد مدل را به سرعت برای وظیفه‌ای خاص تنظیم کند، مفید است. مطالعات نشان داده‌اند که این روش عملکرد مدل را در وظایف پیچیده بهبود می‌بخشد » Liu et al., 2023 «.

پرامپت زنجیره تفکر: پرامپت زنجیره تفکر، مدلی را تشویق می‌کند تا به صورت گام‌به‌گام استدلال کند و پاسخ نهایی را پس از طی مراحل منطقی ارائه دهد. این روش برای حل مسائل پیچیده مانند مسائل ریاضی، تحلیل داده‌ها، یا استدلال‌های منطقی طراحی شده است. به عنوان مثال، در پرامپتی مانند "هزینه یک پروژه را با داده‌های زیر محاسبه کن و مراحل را توضیح بده"، مدل ابتدا داده‌ها را تحلیل کرده، محاسبات را انجام داده و سپس نتیجه را ارائه می‌دهد. این نوع پرامپت به کاربر کمک می‌کند تا نه تنها پاسخ نهایی، بلکه فرآیند استدلال مدل را نیز درک کند، که در کاربردهای آموزشی و حرفه‌ای بسیار ارزشمند است. این تکنیک به دلیل توانایی‌اش در بهبود استدلال مدل‌های زبانی مورد توجه قرار گرفته است » Wei et al., 2022 «.

پرامپت را بهینه کنید « Liu et al., 2023 ».

تکنیک‌های پیشرفته مهندسی پرامپت

تکنیک‌های پیشرفته مهندسی پرامپت، ابزارهایی قدرتمند برای کاربران حرفه‌ای هستند که می‌خواهند از مدل‌های زبانی هوش مصنوعی خروجی‌هایی با دقت بالا، پیچیدگی بیشتر، و تطبیق‌پذیری خاص دریافت کنند. یکی از این تکنیک‌ها، پرامپت‌های زنجیره‌ای « Prompt Chaining » است که در آن یک وظیفه پیچیده به چندین پرامپت کوچک‌تر و متوالی تقسیم می‌شود. به عنوان مثال، برای تحلیل یک گزارش مالی، ابتدا می‌توان از مدل خواست داده‌ها را خلاصه کند، سپس روندها را شناسایی کرده و در نهایت پیشنهادهایی ارائه دهد. این روش با کاهش بار شناختی مدل در هر مرحله، دقت و ساختار خروجی را بهبود می‌بخشد.

همچنین، استفاده از پرامپت‌های خود-اصلاحی « Self-Correcting Prompts »، که مدل را به بررسی و اصلاح پاسخ خود ترغیب می‌کند « مانند "پاسخت را بازبینی کن و خطاها را اصلاح کن" »، به‌ویژه در وظایف تحلیلی و استدلالی مفید است. این تکنیک‌ها نیازمند درک عمیق از رفتار مدل و طراحی پرامپت‌های دقیق هستند « Liu et al., 2023 ».

تکنیک مهندسی پرامپت مبتنی بر الگو « Pattern-Based Prompting » رویکرد دیگری است که در آن از الگوهای استاندارد مانند "دلیل-نتیجه" یا "مشکل-راه‌حل" برای ساختاردهی به پرامپت استفاده می‌شود. به عنوان مثال، در پرامپتی مانند "مشکل: کاهش فروش

در سه ماهه اخیر. راه‌حل: سه استراتژی بازاریابی پیشنهاد کن"، مدل به صورت نظام‌مند به ارائه پاسخ‌های منطقی هدایت می‌شود. این روش در سناریوهای حرفه‌ای مانند برنامه‌ریزی استراتژیک یا حل مسائل تجاری کاربرد دارد. علاوه بر این، پرامپت‌های متنی پویا « Dynamic Contextual Prompts »، که در آن زمینه به صورت تدریجی و بر اساس پاسخ‌های قبلی مدل به‌روزرسانی می‌شود، برای وظایف تعاملی مانند دیالوگ‌های پیچیده یا تحلیل داده‌های چندمرحله‌ای مناسب است. این تکنیک به کاربران اجازه می‌دهد تا مدل را در یک فرآیند پویا و تکراری هدایت کنند، که به‌ویژه در محیط‌های حرفه‌ای با داده‌های متغیر مفید است « Wei et al., 2022 ».

برای کاربران حرفه‌ای، بهینه‌سازی پرامپت با تنظیم پارامترهای مدل « Parameter-Tuned Prompting » نیز اهمیت دارد. این تکنیک شامل تنظیم پارامترهایی مانند دما « برای کنترل خلاقیت » و حداکثر توکن‌ها « برای مدیریت طول پاسخ » در رابط‌های برنامه‌نویسی « API » است. به عنوان مثال، دمای پایین « 0.2-0.4 » برای وظایف تحلیلی دقیق و دمای بالاتر « 0.7-0.9 » برای تولید محتوای خلاقانه مناسب است. همچنین، استفاده از پرامپت‌های ترکیبی « Hybrid Prompts »، که ترکیبی از چند نمونه‌ای، زنجیره تفکر، و نقش‌محور هستند، می‌تواند برای وظایف چندوجهی مانند طراحی سیستم‌های نرم‌افزاری یا تولید گزارش‌های جامع به کار رود. این تکنیک‌ها، با ترکیب خلاقیت و دقت، به کاربران حرفه‌ای امکان می‌دهند تا از قابلیت‌های مدل‌های زبانی به صورت بهینه و تخصصی بهره‌برداری کنند « Reynolds & McDonell, 2021 ».

مطالعات موردی: نمونه‌هایی از پرامپت‌های موفق و ناموفق

تعریف شده « ، محدودیت‌های دقیق » طول و فرمت « ، و مثال‌های هدایت‌کننده است. این ساختار مدل را به تولید پاسخ‌های مرتبط و کاربردی هدایت کرد. استفاده از Few-Shot Learning با مثال‌ها به مدل کمک کرد تا سبک و ساختار موردنظر را دنبال کند « Liu et al., 2023 ». درس کلیدی: ترکیب زمینه، محدودیت‌ها، و مثال‌ها در پرامپت، دقت و کاربردی‌پذیری خروجی را افزایش می‌دهد.

پرامپت موفق: حل مسئله ریاضی با زنجیره تفکر

پرامپت: "یک مسئله ریاضی را گام‌به‌گام حل کن: اگر یک ماشین با سرعت 60 کیلومتر در ساعت حرکت کند و 180 کیلومتر را طی کند، چند ساعت طول می‌کشد؟ وظیفه: پاسخ را با توضیح مراحل استدلال ارائه کن. فرمت: مراحل به‌صورت فهرست شماره‌دار، سپس پاسخ نهایی. مثال: مسئله: مسافت 100 کیلومتر با سرعت 50 کیلومتر در ساعت. مراحل: 1. فرمول زمان = مسافت ÷ سرعت. 2. محاسبه: $100 \div 50 = 2$ ساعت. پاسخ: 2 ساعت." خروجی « فرضی » :

- فرمول زمان = مسافت ÷ سرعت.
- مسافت = 180 کیلومتر، سرعت = 60 کیلومتر در ساعت.
- محاسبه: $180 \div 60 = 3$ ساعت. پاسخ نهایی: 3 ساعت. تحلیل موفقیت: این پرامپت با استفاده از

پرامپت موفق: تولید استراتژی بازاریابی

پرامپت: "در نقش یک مشاور بازاریابی با 10 سال تجربه، یک استراتژی شبکه‌های اجتماعی برای یک برند نوشیدنی گیاهی در ایران طراحی کن. مخاطب: جوانان 18-30 سال. وظیفه: سه تاکتیک اصلی پیشنهاد کن.

محدودیت‌ها: پاسخ به زبان فارسی، حداکثر 300 کلمه، در قالب فهرست شماره‌دار. مثال: 1. انتشار ویدئوهای کوتاه از فرآیند تولید « مانند نمایش گیاهان ارگانیک ». 2. همکاری با اینفلوئنسرهای سلامت « مانند اینفلوئنسر با 100 هزار فالوور ». فرمت: هر تاکتیک با توضیح 80-100 کلمه. "خروجی « فرضی » :

- کمپین چالش سلامت: چالش 30 روزه در اینستاگرام برای اشتراک‌گذاری نوشیدنی‌های گیاهی با هشتگ اختصاصی « #سلامت_طبیعی » . جوایز « مانند اشتراک رایگان » تعامل را افزایش می‌دهد.

- محتوای آموزشی: پست‌های کاروسل درباره فواید گیاهان دارویی، با گرافیک جذاب برای جلب توجه جوانان.

- لایو با متخصصان: جلسات زنده با کارشناسان تغذیه برای بحث درباره نوشیدنی‌های گیاهی و پاسخ به سؤالات مخاطبان. تحلیل موفقیت: این پرامپت موفق است زیرا دارای وضوح « دستور مشخص » ، زمینه‌سازی « مخاطب و برند

پرامپت ناموفق: تحلیل داده بدون زمینه

پرامپت: "این داده‌ها رو تحلیل کن: 10، 20، 15، 30." خروجی « فرضی »: "میانگین داده‌ها 18.75 است." تحلیل شکست: این پرامپت به دلیل نبود زمینه و دستور مشخص ناموفق بود. کاربر نوع تحلیل « مثلاً آماری، روند، یا پیش‌بینی »، منبع داده‌ها، یا هدف تحلیل را مشخص نکرد. این ابهام باعث شد مدل به ساده‌ترین تحلیل « محاسبه میانگین » اکتفا کند، که ممکن است نیاز کاربر را برآورده نکند. فقدان محدودیت « مانند فرمت یا طول پاسخ » نیز به خروجی غیرساختاریافته منجر شد. تحقیقات نشان می‌دهند که زمینه‌سازی ناکافی یکی از دلایل اصلی خروجی‌های غیرمرتبط است (Liu et al., 2023). روش اجتناب: پرامپت را با زمینه « مثلاً "داده‌های فروش ماهانه یک فروشگاه" »، دستور دقیق « مثلاً "روند تغییرات را تحلیل کن" »، و فرمت « مثلاً "در قالب پاراگراف" » بازنویسی کنید.

تکنیک زنجیره تفکر « Chain-of-Thought » مدل را به استدلال گام‌به‌گام هدایت کرد. ارائه مثال واضح و درخواست فرمت مشخص « فهرست شماره‌دار » باعث شد مدل پاسخ را به صورت ساختاریافته و دقیق ارائه دهد. این روش به ویژه در وظایف استدلالی مؤثر است، زیرا مدل را به تجزیه مسئله به بخش‌های قابل مدیریت تشویق می‌کند (Wei et al., 2022). درس کلیدی: استفاده از زنجیره تفکر و مثال‌های مرتبط در پرامپت‌های استدلالی، دقت و شفافیت را بهبود می‌بخشد.

پرامپت ناموفق: تولید محتوای خلاقانه « شعار تبلیغاتی »

پرامپت: "یه شعار خوب برای یه شرکت بنویس." خروجی « فرضی »: "با ما موفق باش!" تحلیل شکست: این پرامپت به دلیل ابهام و فقدان جزئیات ناموفق بود. عدم مشخص کردن نوع شرکت، مخاطب هدف، یا سبک شعار باعث شد مدل پاسخی کلی و غیرمرتبط ارائه دهد. همچنین، نبود زمینه و مثال، مدل را بدون جهت‌گیری رها کرد، که منجر به خروجی غیرمفید شد. مطالعات نشان می‌دهند که پرامپت‌های مبهم اغلب به پاسخ‌های غیرمرتبط منجر می‌شوند، زیرا مدل نمی‌تواند نیت کاربر را به درستی تفسیر کند (Reynolds & McDonell, 2021). روش اجتناب: پرامپت را با زمینه « مثلاً "برای یک شرکت فناوری سلامت" »، محدودیت‌ها « مثلاً "حداکثر 10 کلمه" »، و مثال « مثلاً "مانند: فناوری برای زندگی بهتر" » بازنویسی کنید.

نقش زمینه « Context » در عملکرد دقیق‌تر دستیار هوشمند

زمینه « Context » در طراحی پرامپت نقش محوری در بهبود دقت و عملکرد دستیارهای هوشمند ایفا می‌کند، زیرا اطلاعات پس‌زمینه‌ای لازم را برای درک بهتر وظیفه و تولید پاسخ‌های مرتبط و متناسب با نیاز کاربر فراهم می‌آورد. ارائه زمینه‌ای جامع، مانند مشخص کردن مخاطب هدف، حوزه کاری، یا سناریوی خاص، به مدل کمک می‌کند تا از فرضیات نادرست اجتناب کرده و خروجی‌هایی تولید کند که با شرایط موردنظر هم‌راستا باشند. برای مثال، پرامپتی مانند "یک برنامه آموزشی

برای یادگیری زبان طراحی کن" بدون زمینه ممکن است پاسخی کلی تولید کند، اما افزودن زمینه‌ای مانند "برای دانش‌آموزان دبیرستانی با سطح مبتدی"، پاسخ را دقیق‌تر و کاربردی‌تر می‌کند. فقدان زمینه اغلب به پاسخ‌های غیرمرتبط یا ناقص منجر می‌شود، در حالی که زمینه‌سازی مناسب، دقت و کارایی مدل را به‌طور قابل‌توجهی افزایش می‌دهد « Liu et al., 2023 ».

هنر هدایت مدل‌های زبانی به سوی پاسخ‌های دقیق و مؤثر

مهندسی پرامپت مجموعه‌ای از تکنیک‌ها و روش‌ها برای طراحی، نگارش و بهینه‌سازی دستوراتی است که به مدل‌های زبانی بزرگ « LLMs » داده می‌شود. این دستورات که «پرامپت» نام دارند، نقش تعیین‌کننده‌ای در کیفیت خروجی مدل دارند؛ به‌گونه‌ای که می‌توانند موجب تولید پاسخ‌هایی دقیق، شفاف، قابل تکرار و از نظر محتوایی درست شوند. در کاربردهای واقعی، پرامپت همان ورودی مستقیمی است که به مدل داده می‌شود، و نحوه‌ی طراحی آن می‌تواند به شکل چشمگیری بر خروجی تأثیر بگذارد. تغییر در ساختار پرامپت «مانند طول یا ترتیب مثال‌ها» یا محتوای آن «مانند شیوه‌ی بیان دستور یا انتخاب مثال‌ها» می‌تواند نتایج تولیدشده توسط مدل را به‌طور معناداری تغییر دهد. پژوهش‌ها نشان می‌دهند که هم نحوه‌ی بیان و هم ترتیب ارائه‌ی اطلاعات در پرامپت می‌تواند رفتار مدل را به‌طور قابل‌توجهی تحت تأثیر قرار دهد.

همچنین، یک پرامپت طراحی‌شده به‌صورت حرفه‌ای می‌تواند برخی چالش‌های رایج مدل‌های زبانی، مانند «توهم ماشینی» یا ارائه‌ی اطلاعات نادرست را که در مطالعات مورد بحث قرار گرفته‌اند، تا حدی برطرف کند. علاوه بر این، مهندسی پرامپت به علاقه‌مندان این امکان را می‌دهد که بدون نیاز به دانش عمیق در یادگیری ماشینی یا ریاضیات، از طریق تعامل زبانی طبیعی یا طراحی پرسش‌های هدفمند، دانش پنهان مدل را استخراج و بررسی کنند. این امر باعث دسترسی‌پذیری بیشتر مدل‌های زبانی بزرگ برای طیف وسیع‌تری از کاربران شده است.

ضرورت مهندسی پرامپت برای تعامل مؤثر با مدل‌های زبانی بزرگ

با این‌که نوشتن اعلان‌های مؤثر می‌تواند برای کاربران غیر فنی دشوار باشد، اما این مهارت با خلاقیت، شهود و تمرین مکرر به دست می‌آید. زمانی که برای انجام وظایف تخصصی به اطلاعات دقیق نیاز است، استفاده از درخواست‌های نادرست ممکن است منجر به پاسخ‌هایی مبهم، نادقیق، نادرست یا خارج از حوزه‌ی مورد نظر شود. رشته مهندسی پرامپت هم‌زمان با توسعه‌ی مدل‌های زبانی بزرگ رشد کرده است. آنچه در ابتدا به‌عنوان روشی ابتدایی برای هدایت خروجی مدل‌ها آغاز شد، اکنون به یک حوزه‌ی پژوهشی ساختارمند با روش‌ها و رویه‌های خاص خود تبدیل شده است. مدل‌های زبانی بزرگ نسبت به طراحی پرامپت بسیار حساس‌اند و در دنیای امروز، تسلط بر مهندسی پرامپت برای متخصصان به یک ضرورت تبدیل شده است. بدون شناخت کافی از این حوزه، متخصصان ممکن است در انتقال مؤثر خواسته‌های خود به مدل‌های زبانی بزرگ دچار چالش

چالش‌های متعددی روبه‌رو است. بسیاری از این چالش‌ها به واسطه پیشرفت‌های سریع در مدل‌های زبانی بزرگ و استفاده گسترده از رابط‌های مکالمه‌محور مانند ChatGPT به وجود آمده‌اند.

شده و نتوانند درخواست‌های خود را به صورت اصولی و ساختارمند بیان کنند.

پیچیدگی‌ها، کاربردها و چالش‌های نوظهور در مهندسی پرامپت

یکی از محدودیت‌های این رویکرد، نادیده گرفتن زمینه‌های ضمنی درخواست و اهداف دقیق کاربر است. چنین سیستم‌هایی اغلب از کنترل کامل کاربر پشتیبانی نمی‌کنند و در فرآیند معناسازی و بازنگری محتوای تولید شده کمک محدودی ارائه می‌دهند. از جمله مهم‌ترین چالش‌های مطرح در این حوزه می‌توان به موارد زیر اشاره کرد: تفسیرپذیری و شفافیت، پیچیدگی طراحی پرامپت‌ها، محدودیت‌های زبانی و فرهنگی، وابستگی به مدل خاص، مقیاس‌پذیری، دغدغه‌های اخلاقی و اجتماعی، و دشواری‌های ارزیابی و بهینه‌سازی عملکرد. در زمینه تفسیرپذیری، یکی از مسائل کلیدی، عدم قطعیت در تأثیرگذاری پرامپت‌ها و ضعف در تعمیم‌پذیری آن‌هاست؛ به این معنا که پرامپتی که برای یک مدل یا مسئله خاص مؤثر است، ممکن است در موقعیتی دیگر نتیجه مطلوبی نداشته باشد.

مهندسی پرامپت مجموعه‌ای گسترده از رویکردها را شامل می‌شود؛ از تکنیک‌های پایه‌ای مانند نقش‌دهی «Role Prompting گرفته تا روش‌های پیشرفته‌تری همچون پرامپت زنجیره‌ای «Chain-of-Thought Prompting» این حوزه بسیار پویا است و پژوهش‌های نوین به طور مداوم در حال معرفی تکنیک‌ها و کاربردهای جدید در این زمینه هستند.

اهمیت مهندسی پرامپت به دلیل توانایی آن در هدایت پاسخ‌های مدل و افزایش دقت، انعطاف‌پذیری و مرتبط بودن خروجی مدل‌های زبانی بزرگ در کاربردهای مختلف به خوبی شناخته شده است. از جمله حوزه‌هایی که مهندسی پرامپت در آن‌ها به طور مؤثر استفاده می‌شود می‌توان به آموزش، پزشکی، رادیولوژی و آموزش علوم اشاره کرد. این فناوری می‌تواند نقش‌هایی چون دستیار مجازی برای مراقبت از دانش‌آموزان یا تولید گزارش‌های تخصصی را بر عهده گیرد و داده‌های پیچیده را به روایتی منسجم و قابل فهم تبدیل کند.

دلیل این تفاوت‌ها نیز معمولاً برای کاربران مشخص نیست. به عنوان نمونه، ممکن است پرامپتی مانند: "Explain quantum physics in simple terms for a 10-year-old" در یک نسخه از مدل GPT پاسخ رضایت‌بخشی تولید کند، اما همان پرامپت در نسخه‌ای دیگر «مانند GPT-4» نتیجه‌ای پیچیده‌تر و دشوارتر برای درک ایجاد کند. این تفاوت‌ها نشان‌دهنده محدودیت‌های فعلی در تفسیر رفتار مدل‌های زبانی هستند و لزوم تحقیق بیشتر در این زمینه را برجسته می‌کنند.

در حال حاضر، تلاش‌هایی برای استانداردسازی مفاهیم و اصطلاحات در مهندسی پرامپت از طریق طبقه‌بندی انواع مختلف اعلان‌ها بر اساس معیارهای گوناگون در جریان است. با این حال، مهندسی پرامپت به عنوان یکی از حوزه‌های نوظهور و پیچیده در هوش مصنوعی با

چالش‌های طراحی پرامپت: از مهارت انسانی تا وابستگی به مدل و محدودیت‌های زبانی

نادرست یا غیرطبیعی افزایش می‌یابد.

یکی از چالش‌های مهم در مهندسی پرامپت، پیچیدگی فرآیند طراحی آن‌هاست. این امر به‌طور مستقیم به مهارت انسانی، دانش زبانی و درک عمیق از ساختار مدل‌های زبانی مبتنی بر هوش مصنوعی وابسته است. طراحی یک پرامپت مؤثر معمولاً نیازمند تجربه، خلاقیت و صرف زمان زیادی برای آزمون و خطا است، که این فرآیند نه تنها زمان‌بر بلکه منجر به اتلاف منابع نیز می‌شود.

برای مثال، در یک پروژه آموزش زبان، ممکن است از پرامپتی مانند: "Translate the following text into French and explain the grammar rules used in the translation" استفاده شود. در حالی که هدف از طراحی چنین پرامپتی دستیابی به پاسخ دقیق و قابل فهم است، اما کاربر تنها زمانی می‌تواند از کیفیت و صحت پاسخ اطمینان حاصل کند که چندین بار پاسخ را آزمایش و بازبینی کند تا به نتیجه نهایی برسد.

از دیگر چالش‌های مهم، محدودیت‌های زبانی و فرهنگی در طراحی پرامپت‌هاست. پرامپت‌هایی که برای یک زبان خاص طراحی می‌شوند ممکن است در زبان‌های دیگر کارایی نداشته باشند یا به دلیل تفاوت‌های فرهنگی، نتایج غیر دقیق یا حتی نادرستی تولید کنند. برای نمونه، پرامپتی مانند: "Translate this English proverb into Persian and provide an equivalent proverb in Persian culture" ممکن است با مشکل مواجه شود، زیرا مدل ممکن است نتواند معادل فرهنگی مناسب و دقیق را شناسایی کند. همچنین در زبان‌هایی با ساختار پیچیده مانند عربی یا چینی، احتمال تولید پاسخ‌های ناهنجار،

افزون بر این، یکی دیگر از مسائل اساسی در مهندسی پرامپت، وابستگی شدید آن به مدل‌های خاص است. مدل‌های زبانی مختلف—even آن‌هایی که از یک خانواده هستند مانند GPT-3 و GPT-4—ممکن است نسبت به یک پرامپت مشابه پاسخ‌های بسیار متفاوتی ارائه دهند. به علاوه، به روزرسانی‌های مداوم در ساختار و پارامترهای مدل‌ها ممکن است عملکرد پرامپت‌هایی که قبلاً خوب عمل می‌کردند را مختل کرده و نیاز به بازطراحی آن‌ها را ایجاد کند.

برای مثال، کاربران ChatGPT مشاهده کرده‌اند که برخی از پرامپت‌هایی که در نسخه‌های پیشین مدل به خوبی پاسخ می‌دادند، در نسخه‌های جدیدتر کارایی خود را از دست داده‌اند. این تغییر رفتار اغلب ناشی از تفاوت در ساختار پاسخ‌دهی مدل ارتقاء یافته است. نمونه‌ای از این مسئله را می‌توان در پرامپتی مانند: "Summarize this medical article in layman's terms" مشاهده کرد؛ در حالی که GPT-4 ممکن است پاسخی دقیق و قابل فهم ارائه دهد، مدل‌های ساده‌تر مانند GPT-2 ممکن است خروجی مبهم، ناقص یا گمراه‌کننده تولید کنند که دلیل آن محدودیت‌های معماری مدل‌های کوچکتر است. این چالش‌ها ضرورت طراحی دقیق‌تر، بازبینی مستمر و تحلیل عملکرد پرامپت‌ها در مدل‌های مختلف را بیش از پیش نمایان می‌سازند.



چالش‌های مقیاس‌پذیری و هزینه پردازش در مدیریت پرامیت‌ها

یکی از چالش‌های اساسی در حوزه مهندسی پرامیت، مسئله مقیاس‌پذیری و هزینه‌های مرتبط با پردازش است. این چالش به‌ویژه در سیستم‌هایی که با مجموعه‌ای از پرامیت‌های متعدد و متنوع کار می‌کنند، پررنگ‌تر می‌شود. از یک سو، هماهنگی میان پرامیت‌های مختلف در چنین سیستم‌هایی می‌تواند بسیار دشوار باشد؛ زیرا ممکن است پرامیت‌ها در عملکرد با یکدیگر تداخل پیدا کرده و منجر به نتایج متناقض یا غیرهم‌راستا شوند.

برای مثال، در یک سامانه مدیریت مکالمات با مشتری، پرامیتی که برای پاسخ‌گویی به شکایات طراحی شده ممکن است با پرامیت دیگری که هدف آن پیشنهاد محصولات جدید است، تداخل پیدا کند. در نتیجه، خروجی‌هایی تولید می‌شوند که هم از نظر عملکردی ناسازگارند و هم موجب سردرگمی کاربر خواهند شد. از سوی دیگر، استفاده از پرامیت‌های طولانی و پیچیده نیز می‌تواند منجر به افزایش چشمگیر زمان پردازش و هزینه‌های محاسباتی شود. به بیان دیگر، هرچه محتوای پرامیت گسترده‌تر و پردازش آن دشوارتر باشد، مدل زبان نیازمند صرف منابع بیشتری برای تولید پاسخ خواهد بود.

نمونه‌ای از چنین پرامیتی عبارت است از: "Analyze this 5,000-word document and summarize it into a report with three sections, including a detailed SWOT analysis." در این مثال، مدل باید علاوه بر درک و تحلیل یک متن بلند، خروجی‌ای ساختارمند، دقیق و تحلیلی تولید کند. چنین پرامیت‌هایی به‌طور طبیعی نیاز به منابع پردازشی گسترده‌تری دارند که می‌تواند هزینه‌های عملیاتی را افزایش داده و بهره‌وری سیستم را کاهش دهد. در نتیجه، طراحی پرامیت‌ها در سیستم‌های پیچیده باید با دقت بسیار بالا انجام شود تا هم از تداخل‌های مفهومی و عملکردی جلوگیری شود و هم هزینه‌های پردازشی در حد مطلوب باقی بماند.

چالش‌های اخلاقی، اجتماعی و ارزیابی در مهندسی پرامیت

یکی دیگر از چالش‌های اساسی در حوزه مهندسی پرامیت، مسائل اخلاقی و اجتماعی است که می‌تواند منجر به تولید محتوای نادرست، سوگیرانه یا حتی مخرب شود. در برخی موارد، پرامیت‌ها به‌طور مستقیم یا غیرمستقیم موجب تولید محتوایی می‌شوند که از منظر اخلاقی یا اجتماعی غیرقابل قبول است. طراحی آگاهانه یا ناآگاهانه پرامیت‌هایی با اهداف غیراخلاقی یا غیرقانونی، مانند تولید اخبار جعلی یا محتوای خطرناک، از جمله موارد چالش‌برانگیز در این حوزه به شمار می‌رود.

برای نمونه، پرامیتی با ساختار: "List reasons why one race is superior to another" می‌تواند منجر به خروجی‌هایی شود که حاوی اطلاعات تبعیض‌آمیز، نژادپرستانه و مغرضانه هستند و در نهایت بستر

سوءاستفاده را فراهم سازد. همچنین، برخی پرامپت‌ها به صورت ناآگاهانه حامل سوگیری‌های پنهان هستند. به عنوان مثال، پرامپت: "Why are women less successful in tech?" ممکن است پاسخ‌هایی تولید کند که کلیشه‌های جنسیتی را تقویت کرده و موجب بازتولید تبعیض شود.

روش تحقیق

پژوهش حاضر از نوع کاربردی بوده و با رویکرد توصیفی-تحلیلی انجام شده است. هدف اصلی، بررسی اصول و تکنیک‌های مؤثر در مهندسی پرامپت و ارائه راهکارهایی برای تعامل حرفه‌ای‌تر با دستیارهای هوش مصنوعی نظیر ChatGPT، Copilot و سایر مدل‌های زبانی است.

در کنار چالش‌های اخلاقی، مسئله ارزیابی و بهینه‌سازی پرامپت‌ها نیز از پیچیدگی‌های قابل توجهی برخوردار است. فقدان معیارهای استاندارد و جامع برای سنجش کیفیت پرامپت‌ها، فرایند بهینه‌سازی را دشوار می‌سازد. طراحی پرامپت‌هایی که به طور هم‌زمان چندین هدف مانند دقت، سرعت، خلاقیت و اختصار را دنبال می‌کنند، به‌ویژه دشوار است.

روش گردآوری داده‌ها

داده‌های این پژوهش از طریق دو روش گردآوری شده‌اند:

برای مثال، پرامپتی مانند: "Generate a creative story about space exploration" بیش از حد به قضاوت انسانی وابسته است، چراکه معیار مشخصی برای "خلاقیت" یا "جذابیت" وجود ندارد. همچنین، پرامپت‌هایی با محدودیت‌های سخت مانند: "Answer this question in 50 words or less" دقت علمی و محتوایی ممکن است قربانی کوتاهی پاسخ شود که در کاربردهای تخصصی می‌تواند منجر به گمراهی شود.

- مطالعه اسنادی و کتابخانه‌ای برای بررسی منابع علمی، مقالات پژوهشی و مستندات فنی شرکت‌های توسعه‌دهنده هوش مصنوعی مانند OpenAI، Google، Anthropic و... در زمینه طراحی پرامپت و تعامل انسان-ماشین.

در مجموع، چالش‌های اخلاقی، اجتماعی و ارزیابی از مهم‌ترین موانع گسترش مهندسی پرامپت به عنوان ابزاری استاندارد و قابل اتکا در حوزه هوش مصنوعی محسوب می‌شوند. پرداختن دقیق به این چالش‌ها در قالب مطالعات مروری و تحلیلی مبتنی بر منابع معتبر، می‌تواند مسیر بهبود عملکرد پرامپت‌ها و توسعه اصولی این حوزه را هموارتر سازد.

- مطالعه میدانی و تجربی شامل تحلیل کیفی تعامل کاربران با مدل‌های زبانی از طریق آزمایش‌های طراحی‌شده با پرامپت‌های مختلف و بررسی نتایج حاصل از تغییرات در ساختار پرامپت.

جامعه و نمونه تحقیق

جامعه مورد مطالعه شامل کاربران فعال ابزارهای هوش مصنوعی متنی مانند معلمان، پژوهشگران، تولیدکنندگان محتوا و برنامه‌نویسان است. از میان این جامعه، با استفاده از روش نمونه‌گیری هدفمند، ۱۰ نفر کاربر حرفه‌ای انتخاب شدند که تجربه زیادی در کار با مدل‌های زبانی دارند.

ابزار گردآوری اطلاعات

برای جمع‌آوری داده‌های میدانی، از ابزارهایی مانند چک‌لیست ارزیابی عملکرد پرامپت‌ها، پرسش‌نامه باز با پاسخ‌های کیفی و ثبت تعاملات کاربران با هوش مصنوعی استفاده شد.

روش تحلیل داده‌ها

داده‌های کیفی حاصل از تحلیل پرامپت‌ها و پاسخ‌ها با استفاده از روش تحلیل محتوا بررسی شدند. همچنین، برای ارزیابی اثربخشی پرامپت‌های مختلف، از مقایسه تطبیقی پاسخ‌ها و کدگذاری مضمون‌ها استفاده گردید.

نتیجه‌گیری

مهندسی پرامپت به‌عنوان یک مهارت کلیدی، قابلیت تبدیل تعاملات با دستیارهای هوش مصنوعی به یک مزیت رقابتی برجسته را برای افراد و سازمان‌ها فراهم می‌کند. طراحی پرامپت‌های دقیق، واضح، و زمینه‌محور، کاربران را قادر می‌سازد تا خروجی‌هایی با کیفیت بالا و متناسب با نیازهای خاص در حوزه‌هایی مانند تحلیل داده‌های پیچیده، تولید محتوای خلاقانه، و حل مسائل تخصصی دریافت کنند. تکنیک‌های پیشرفته نظیر زنجیره تفکر «Chain-of-Thought»، پرامپت‌های چندمنونه‌ای، و تنظیم پارامترهای مدل «مانند دما و طول پاسخ»، امکان تولید پاسخ‌های استدلالی و هدفمند را تقویت می‌کنند. این مهارت نه تنها بهره‌وری را افزایش می‌دهد، بلکه با کاهش خطاها و بهینه‌سازی زمان، به کاربران اجازه می‌دهد تا در محیط‌های رقابتی پیشرو باشند.

علاوه بر این، اجتناب از اشتباهات رایج در طراحی پرامپت، مانند ابهام در دستورات یا فقدان زمینه کافی، نقش مهمی در تضمین نتایج قابل‌اعتماد و کاربردی ایفا می‌کند. با تسلط بر اصول مهندسی پرامپت و استفاده استراتژیک از انواع پرامپت‌ها «صفرمنونه‌ای، مبتنی بر نقش، و غیره»، کاربران می‌توانند از پتانسیل کامل مدل‌های زبانی بهره‌برداری کنند و راه‌حل‌هایی نوآورانه ارائه دهند. این توانایی، به‌ویژه در صنایع فناوری، بازاریابی، و آموزش، به یک مزیت رقابتی پایدار تبدیل می‌شود که سازمان‌ها را از رقبا متمایز می‌کند. در نهایت، مهندسی پرامپت نه تنها یک ابزار فنی، بلکه یک رویکرد تحول‌آفرین است که آینده تعاملات هوشمند را شکل می‌دهد.

منابع

» . <https://doi.org/10.48550/arXiv.2104.08786>

Webson, A., & Pavlick, E. « 2022 » . Do prompt-based models really understand the meaning of their prompts? In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* « pp. 2300-2344 » . <https://doi.org/10.48550/arXiv.2109.01247>

Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. « 2020 » . On faithfulness and factuality in abstractive summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* « pp. 1906-1919 » . <https://doi.org/10.48550/arXiv.2005.00661>

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., et al. « 2023 » . Sparks of artificial general intelligence: Early experiments with GPT-4. *arXiv preprint*, arXiv:2303.12712. <https://doi.org/10.48550/arXiv.2303.12712>

Lo, L. S. « 2023 » . The art and science of prompt engineering: A new literacy in the information age. *Internet Reference Services Quarterly*. <https://doi.org/10.1080/10875301.2023.2227621>

Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., & Yang, Q. « 2023 » . Why

Giray, L. « 2023 » . Prompt engineering with ChatGPT: A guide for academic writers. *Annals of Biomedical Engineering*. <https://doi.org/10.1007/s10439-023-03272-4>

White, J., et al. « 2023 » . A prompt pattern catalog to enhance prompt engineering with ChatGPT. *arXiv preprint*, arXiv:2302.11382. <https://doi.org/10.48550/arXiv.2302.11382>

Eager, B., & Brunton, R. « 2023 » . Prompting higher education towards AI-augmented teaching and learning practice. *Journal of University Teaching and Learning Practice*, 20 « 5 » . <https://doi.org/10.53761/1.20.5.02>

Kaddour, J., Harris, J., Mozes, M., Bradley, H., Raileanu, R., & McHardy, R. « 2023 » . Challenges and applications of large language models. *arXiv preprint*, arXiv:2307.10169. <https://doi.org/10.48550/arXiv.2307.10169>

Lu, Y., Bartolo, M., Moore, A., Riedel, S., & Stenetorp, P. « 2022 » . Fantastically ordered prompts and where to find them: Overcoming few-shot prompt order sensitivity. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics* « pp. 8086-8098

- Liu, P., et al. « 2023 ». Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55 « 9 », 1-35.
- Wei, J., et al. « 2022 ». Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35, 24824-24837.
- Reynolds, L., & McDonell, K. « 2021 ». Prompt programming for large language models: Beyond the few-shot paradigm. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1-7.
- Brown, T. B., et al. « 2020 ». Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901.
- Johnny can't prompt: How non-AI experts try « and fail » to design LLM prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems « CHI '23 »*. New York, NY, USA: ACM. <https://doi.org/10.1145/3544548.3581388>
- Lecler, A., Duron, L., & Soander, P. « 2023 ». Revolutionizing radiology with GPT-based models: Current applications, future possibilities, and limitations of ChatGPT. *Diagnostic and Interventional Imaging*, 104 « 6 », 269-274. <https://doi.org/10.1016/j.diii.2023.02.003>
- Epstein, R. H., & Dexter, F. « 2023 ». Variability in large language models" responses to medical licensing and certification examinations. *JMIR Medical Education*, 9. <https://doi.org/10.2196/48305>
- Cooper, G. « 2023 ». Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology*, 32 « 3 », 444-452. <https://doi.org/10.1007/s10956-023-10039-9>
- Zhang, Z., Gao, J., Dhaliwal, R. S., Li, J.-J., & Li, T. « 2023 ». VISAR: A human-AI argumentative writing assistant with visual programming and rapid draft prototyping. *arXiv preprint*, arXiv:2304.07810. <https://doi.org/10.48550/arXiv.2304.07810>

پاورقی‌ها: