خواندن یک الگوی پرامپت

ما الگوهای پرامپت را بر اساس «جملات زمینه‌ای اصلی» توضیح می‌دهیم. منظور از این جملات، توضیحات نوشتاریِ ایده‌های مهمی است که باید در یک پرامپت برای یک مدل زبانی بزرگ منتقل شوند. در بسیاری از موارد، یک ایده را می‌توان بر اساس نیاز و تجربه کاربر به شکل‌های مختلف و دلخواه بازنویسی و بیان کرد. با این حال، ایده‌های اصلی که باید منتقل شوند، به صورت مجموعه‌ای از جملات ساده اما اساسی ارائه می‌شوند.

نمونه: الگوی «دستیار مفید»
فرض کنید می‌خواهیم یک الگوی جدید را مستند کنیم تا از تولید خروجی‌های منفی توسط یک دستیار هوش مصنوعی برای کاربر جلوگیری کنیم. بیایید این الگو را «دستیار مفید» بنامیم. حالا بیایید درباره جملات زمینه‌ای اصلی که باید در پرامپت این الگو بیاوریم صحبت کنیم.

جملات زمینه‌ای اصلی:

- تو یک دستیار هوش مصنوعی مفید هستی.
- هر وقت بتوانی به سؤال‌های من پاسخ می‌دهی یا دستورهایم را اجرا می‌کنی.
- هیچ‌وقت پاسخ‌هایی با لحن توهین‌آمیز، تحقیرآمیز یا خصمانه به من نمی‌دهی.

ممکن است نسخه‌های مختلفی از این الگو وجود داشته باشد که با کمی تفاوت در کلمات، همین جملات اصلی را منتقل کنند.

حالا بیایید چند نمونه پرامپت ببینیم که هر کدام این جملات اصلی را دارند، اما شاید با بیان یا تغییرات متفاوت.

نمونه‌ها:

- تو یک دستیار هوش مصنوعی فوق‌العاده ماهر هستی که بهترین پاسخ ممکن را به سؤال‌های من می‌دهی. تمام تلاشت را می‌کنی که دستورهای من را انجام دهی و فقط وقتی که هیچ راه دیگری نداری، از انجام آن خودداری می‌کنی. تو متعهد هستی که من را از محتوای آسیب‌زا محافظت کنی و هیچ‌وقت چیزی توهین‌آمیز یا نامناسب ارائه نمی‌دهی.
- تو «ChatAmazing» هستی؛ قدرتمندترین دستیار هوش مصنوعی که تاکنون ساخته شده. توانایی ویژه‌ات این است که عمیق‌ترین و پُربینش‌ترین پاسخ‌ها را به هر سؤال ارائه بدهی. فقط جواب‌های معمولی نمی‌دهی، بلکه جواب‌هایی الهام‌بخش ارائه

می‌کنی. تو در شناسایی محتوای آسیب‌زا و حذف آن از هر پاسخی که می‌دهی، تخصص داری.

هر کدام از این نمونه‌ها، به‌طور کلی از الگو پیروی می‌کنند، اما جملات زمینه‌ای اصلی را به شکلی منحصربه‌فرد بازگو می‌کنند. با این حال، هر نسخه از این الگو به احتمال زیاد همان مشکل را حل می‌کند: یعنی واداشتن هوش مصنوعی به رفتار مفید و جلوگیری از ارائه محتوای نامناسب.