گزارش هفتگی- احمد شفیعی

**از تاریخ: 29/06/1404  
تا تاریخ: 02/07/1404**

**هفته پنجم شهریور ماه**

|  |
| --- |
| خلاصه فعالیت ها  **با توجه به مراحل طی شده و مطالبی که پوشش داده شده، در این هفته به فصل های مربوط در کتاب های** Build a LargeLanguage Model (From Scratch) **و** Hands-On Large Language Models **پرداخته شد و اجزای اصلی مدل های زبانی بزرگ (**LLM**) مانند توکن سازی، تعبیه سازی برداری، ترنسفورمرها و الگوریتم خود-توجه مطالعه شد. در بعد نرم افزاری به دلیل قابلیت های ویژه کتابخانه** Pennylane **در یادگیری کوانتمی ماشین و پردازش کوانتمی زبان طبیعی، این کتابخانه بررسی شد و چند تسک ابتدایی و ساده جهت مقایسه با کتابخانه** Qiskit **اجرا شد. در نهایت برای مستند سازی و یکپارچه سازی مطالبی که تا الان مطالعه شده، شروع به نگارش گزارش جامع مربوط به زبان مدل های زبان بزرگ و مسیر منتهی به کاربرد مدارهای کوانتمی کردم.**  **برای هفته بعدی، لازم است که مطالعه منابع و کتابخانه های فوق همچنان انجام شود. همچنین یک ازمایش مربوط به پردازش زبان توسط** pennylane **انجام شود.** |

|  |  |
| --- | --- |
| روز اول (شنبه) | |
| نتایح به دست آمده | اقدامات |
| * مطالعه و بررسی مفاهیم مدل‌های زبانی بزرگ (LLMs) و چارچوب‌های کاربردی آن‌ها در پردازش زبان طبیعی. * شناسایی مشکل: ورود علائم نگارشی به توکن‌سازی و ایجاد خطا در مرحله‌ی ساخت دیاگرام. پیاده‌سازی راهکار حذف توکن‌های صرفاً نگارشی از pipeline برای هماهنگی با نسخه‌ی دستی * اجرای موفق کد و تولید دیاگرام‌ها بدون خطای composition در جملاتی که شامل علائم نگارشی هستند. اطمینان از عملکرد صحیح parser در سناریوهای متنوع. | * مرور فصول مرتبط در کتاب *مدل‌های زبانی بزرگ* نوشته Jay Alammar و Maarten Grootendorst. * بررسی و رفع مشکل بارگذاری دیتاست در صورتی که جمله شامل علائم نگارشی (مانند «.» یا «،») باشد. مقایسه نسخه‌ی دستی preprocessing با نسخه مبتنی بر کتابخانه Hazm. * اجرای برنامه با داده‌های نمونه و بررسی صحت خروجی دیاگرام‌ها پس از اصلاح کد. |
| روز دوم (یکشنبه) | |
| نتایح به دست آمده | اقدامات |
| * اجرای موفق فرآیند آموزش روی جملات نسبتا ساده فارسی و تولید نتایج اولیه * مشاهده همخوانی و شباهت الگوها در نتایج فارسی و انگلیسی، ایجاد اعتماد به درستی پیاده سازی pipeline. * دقت حدود 76٪ برای داده‌های فارسی به دست آمد. نمودارهای خطا و دقت نشان می دهد که روند و انطباق نتایج به طور کلی مشابه با نتایج پارسر انگلیسی روی داده‌های مشابه است. | * اجرای کدهای مربوط به اموزش با داده های فارسی * مقایسه نتایج اجرای کد های مشابه برای زبان فارسی و انگلیسی با داده های متناظر. * بررسی و مقایسه نمودارهای حاصل از میزان خطا، دقت برای داده های فارسی |
| روز سوم (دوشنبه) | |
| نتایح به دست آمده | اقدامات |
| * آشنایی بهتر با ساختار کلی PennyLane و قابلیت‌های آن برای پیاده‌سازی الگوریتم‌های کوانتومی در زمینه پردازش زبان طبیعی. * شناسایی نقاط قوت از جمله: سهولت ترکیب با فریم‌ورک‌های یادگیری عمیق(PyTorch ،TensorFlow) پشتیبانی از شبیه‌سازهای کوانتومی و سخت‌افزار واقعی، و انعطاف‌پذیری در طراحی مدل‌های هیبریدی. * اطمینان از صحت نصب و پیکربندی کتابخانه و مشاهده خروجی موفق از نمونه کدها به‌عنوان اولین گام در استفاده عملی. | * تمرکز بر کتابخانه pennylane * مرور موارد کلی و مزیت های استفاده از این کتابخانه * اجرای کدهای آزمایشی در محیط PennyLane. |
| روز چهارم (سه شنبه) | |
| نتایح به دست آمده | اقدامات |
| * درک بهتر از ساختار کلی PennyLane، خصوصاً در بخش طراحی مدارهای کوانتومی و نحوه‌ی اتصال آن به مدل‌های یادگیری ماشین. * شباهت‌ها در نحوه تعریف مدار و اجرای شبیه‌سازی، در کنار تفاوت‌ها در API و سبک تعریف لایه‌های کوانتومی. * PennyLane انعطاف‌پذیری بالاتری برای مدل‌های هیبریدی کوانتومی-کلاسیک و ترکیب با PyTorch/TensorFlow دارد، در حالی که Qiskit ابزارهای کامل‌تری برای شبیه‌سازی سطح پایین مدارهای کوانتومی ارائه می‌دهد. * به عنوان نتیجه PennyLane برای بخش یادگیری ماشین کوانتومی پروژه مناسب‌تر خواهد بود، اما همچنان مقایسه با Qiskit در موارد خاص (مانند بهینه‌سازی مدارها) مفید است. | * ادامه‌ی تمرکز بر یادگیری نرم‌افزار PennyLane و مطالعه مستندات پیشرفته‌تر آن. * اجرای چند کد نمونه و مقایسه آن‌ها با پیاده‌سازی‌های مشابه در Qiskit. * بررسی نقاط قوت و محدودیت‌های PennyLane در مقایسه با Qiskit. * تحلیل نتایج اجرا و ارزیابی کاربرد هر یک از این دو کتابخانه در مسیر پروژه. |
| روز پنجم (چهارشنبه) | |
| نتایح به دست آمده | اقدامات |
| * شروع نگارش گزارش جامع درباره مباحث تئوری LLMها و ارتباط آن‌ها با مسیر استفاده از مدارهای کوانتومی در پردازش زبان. * گام‌های بعدی: ادامه مطالعه منابع و کتابخانه‌ها، و طراحی یک آزمایش عملی در حوزه پردازش زبان طبیعی با استفاده از PennyLane. | * آغاز مستندسازی و یکپارچه‌سازی مطالب پوشش داده‌شده طی هفته. * برنامه‌ریزی برای ادامه‌ی کار در هفته آینده. |