

شرح مسئله

با پیشرفت مدل‌های زبانی بزرگ (LLMs)، مانند GPT، Claude، LLaMA و دیگر مدل‌ها، اکنون امکان ساخت عامل‌های هوشمند خودمختار فراهم شده است؛ عامل‌هایی که می‌توانند وظایف پیچیده را با استفاده از زبان طبیعی، تعامل با کاربر، و بهره‌گیری از ابزارهای مختلف انجام دهند.

در این چالش، شرکت کنندگان باید یک عامل هوشمند مبتنی بر LLM طراحی کنند که بتواند یک یا چند وظیفه مشخص را به صورت خودکار انجام دهد.

پیشنهاد ما: معمولاً در هر سایتی یک بخش برای پاسخ به سوالات وجود دارد. در این مساله قصد داریم یک عامل هوشمند درست کنیم که ابتدا یک سایت را کامل پیمایش کند و اطلاعات آن را به نحو مناسبی ذخیره کند. سپس به پرسش کاربر طبق محتوای سایت پاسخ دهد. مثلاً این سایت می‌تواند یک سایت خبری باشد یا می‌تواند یک فروشگاه آنلاین باشد. فرآیند indexing باید به طور دوره ای اتفاق بیفتد تا پایگاه داده به روز باشد.

به طور دقیق تر ابتدا یک خزشگر هوشمند داشت که بر روی سایت خزش کند و اطلاعات را به نحو قابل فهم استخراج کند. سپس این اطلاعات به نحو شایسته ای ذخیره شود (مثلاً گراف دانش برای سایت های خبری یا پایگاه داده های برداری برای ذخیره اطلاعات embedding های متنی و تصویری محصولات فروشگاه های آنلاین)، سپس برای پاسخ به پرسش کاربر یک agentic RAG استفاده شود که از این اطلاعات طبقه بندی شده استفاده لازم را بکند.

توجه کنید که ممکن است محتوای سایت شامل عکس یا فیلم هم باشد. اگر در راه حل شما به این موضوع هم پرداخته شود امتیاز ویژه ای دارد. مثلاً باعث می شود که بتوانیم قابلیت جستجوی تصویری هم در بین مطالب سایت مورد نظر داشته باشیم.

اهداف مسئله

- خزش هوشمند روی سایت منبع
- استخراج مطالب از هر صفحه شامل متن یا تصویر
- ساخت گراف دانش منطبق با مطالب و تصاویر سایت منبع
- ساخت عامل (Agent) مبتنی بر LLM که توانایی درک وظایف، برنامه ریزی، اجرای اقدامات و تعامل چندمرحله‌ای با کاربران مطابق با اطلاعات index شده
- کار کردن با embedding های متنی و تصویری

خروجی مورد انتظار

- نمونه اولیه قابل اجرا (API، CLI، Web App) یک سیستم که بتواند با دریافت آدرس یک وبسایت، پردازش های اولیه و indexing را آغاز کند و در ادامه بتواند مانند یک راهنمای سایت به گفتگو با کاربر بپردازد و به سوالات پاسخ دهد.
- ویدئوی دمو: نمایش عملکرد سیستم، از ورودی گرفتن آدرس سایت تا انجام مکالمه و نمایش نتایج.
- توضیحات **README:** شامل جزئیات معماری ایجنت، توضیحات در مورد نحوه تعامل اجزای مختلف سیستم (LLM، ابزارها، API ها، و منطق عامل). مراحل پیاده سازی و آزمایش های که طی مسیر طی شده به علاوه راهنمای استفاده.

ابزارها و منابع پیشنهادی

- LLM APIs: GPT-4، Claude، Mistral، Gemini و غیره
- Vector Graph DB: Neo4J، Nebula
- Vector DBs: Milvus، Chroma
- UI tools: Streamlit، Gradio، Flask
- استفاده از ابزارهای مختلف خزنگری مانند selenium
- Embedding های متنی و تصویری
- Graphiti برای ساخت و تغییر online گراف دانش

نکات کلیدی داوری

به دنبال راه حل هایی هستیم که:

- تا جای ممکن فرآیندها را به صورت خودکار و بدون دخالت انسانی انجام دهند. پشتیبانی از تصویر و صوت امتیاز مثبت محسوب می شود.
- خلاقانه باشند: نوآوری در نحوه indexing یک سایت و ایجاد گراف دانش و پاسخ دهی طبق آن با استفاده از LLM ها.
- تأثیر عملی یا اجتماعی واقعی داشته باشند: پتانسیل ساده سازی و تسریع فرآیندهای پاسخ دهی به کاربران