

새로운 노트

2025.05.17 토 오후 8:06 · 19분 52초

이태수

참석자 1 00:00

코드 및 데이터 벡터를 저장하는 검색 및 벡터 DB고 awss에는 이 업로드된 임시 파일을 저장하는 장소가 될 겁니다.

참석자 2 00:12

파이벡터 저게 이번에 저걸로 배우셨나요? 네. 수업에서 강의에서 저걸로 배우신 거죠?

네네 혹시 벡터 DB도 여러 가지 종류가 많잖아요 네네네.

밀버스라는 게 있고 파인 콘 같은 것도 있고 크로마 DB 같은 것도 있고 여러 가지가 있는데 저걸 무조건 써야 되나요?

제가 잘 몰라가지고

참석자 1 00:38

그건 아닐 거예요. 네 그냥 벡터 DB만 사용하면 되는 걸로 알고 있습니다.

참석자 1 00:48

혹시 추천해 주실 만한 벡터 DB가 있을까요?

참석자 2 00:50

현업에서 제일 많이 지금은 쓰는 건 밀버스나 아니면 엘라스틱 서치 아니면 파인 콘 이런 것들을 많이 쓰긴 하거든요.

참석자 1 01:14

타인콘이랑 이런 라드틱 소치

참석자 2 01:16

밀버스를 최근에는 많이 제일 많이 본 것 같아요.

참석자 2 01:25

구글에서 잠시만요.

참석자 3 01:36

저 질문 하나 있는데 네네 말씀하세요. 저희가 관계형 데이터베이스로 포스트 그릴을 아마 쓰게 될 것 같은데 네네네.

그러면 PC 벡터로 하는 게 낫지 않을까요?

참석자 2 01:49

왜 그렇게 생각하시는 거예요?

참석자 3 01:51

포스트 그리 내에서 피지 벡터를 지원하고 있어가지고 같이 제가 토의 프로젝트 때도 포스트 그리 내에서 피지 벡터로 같이 저장을 했었거든요.

참석자 2 02:04

사실 그렇게도 생각하실 수 있는데 관계형 데이터베이스 rd 베스는 어떤 걸 쓰셔도 상관은 없거든요.

근데 포맷들이 다 비슷하고 그리고 이제 지원되는 것들이 다 비슷하기 때문에 어떤 포맷을 이제 지원할지 아이스버그 형태인지 이런 것들로 판단해 가지고 하는 거기 때문에 여러분들 편하신 걸로 하시면 돼요.

이거는 그냥 저는 이제 이런 게 있다고 알려드리는 거여 가지고

참석자 3 02:32

네 알겠습니다.

참석자 1 02:33

네네 감사합니다.

참석자 1 02:40

혹시 이로서 해도 될까요? 네네 네네 알겠습니다.

이제 두 번째 에이전트는 이제 저희가 멘토님한테 조언을 드렸던 도큐먼트 사내 문서에 해당하는 에이전트고요.

이거 역시 마찬가지로 문서 기반 질의 응답인데 문서 요약 정책 관리 문서가 되겠죠.

그리고 사용 분야는 사내 정책 문서 가이드라인 등의 검색 및 요약인데 이 부분은 이거 뒤에서 시장 조사를 할 때 또 다시 말씀드릴 거긴 한데 제가 어제 박람회 다녀왔을 때 실제로 이 부분에 해당하는 레그 기술이라든지 에이전트를 사용하는 시스템 AI 프로그램들이 굉장히 많았더라고요.

그래서 시장 조사할 때 거기서 도움을 좀 받았습시다.

참석자 2 03:22

네 확실히 저 주제에 관심이 많더쥬.

참석자 1 03:26

한 비율로 따지면 100개 중에 70개는 다 이 정도인 것 같았어요.

참석자 2 03:32

네 맞아요. 그래서 지금 어떤 회사들이 제일 관심 있어 하는 주제를 만약에 이걸로 뭔가 사업성을 할 거다 이러면 모르겠는데 네 그게 아니고 이제 취업을 준비할 거다 하면 회사가 관심 있어 하는 주제를 하는 게 제일 좋긴 합니다.

참석자 1 03:47

네 맞아요. 레그나 에이전트 그런 게 대부분이더라고요.

마지막으로 시계열 데이터 시계열 데이터나 카테고리 데이터로 머신러닝 모델을 기반으로 한 예측 혹은 데이터 인사이트 분석을 하는 에이전트인데요.

이거는 이따가 남궁 승무 님이랑 규희 님이 해 주신 부분이 있는데 저희가 예전에 수업 시간에 배웠던 활용한 데이터를 이용해서 예 하는 코드를 한번 저희가 작성해서 실행해 봤습니다.

그걸로 그래프도 그려보고 그래서 이따가 다시 보여드릴게요.

그리고 마지막으로 AWS 파트는 데이터베이스 관련이 있지만 AWS 3는 목적은 사용자가 업로드한 파일 문서나 기타 코드 같은 것을 임시로 저장을 할 거고 이거는 코드와 관련된 기능들이다 보니까 에이전트 첫 번째 에이전트가 주로 사용을 할 겁니다.

추후에는 인빌딩 처리하는 앱 삭제 가능하고요. 가능해야겠죠 그리고 벡터 같은 경우에는 벡터를 저장하고 검색하는 건데 에이전트랑 에이전트 2 둘 다 사용을 할 것입니다.

참석자 1 04:53

대신에 사용하는 벡터가 코드 벡터냐 문서 벡터냐 그거에 따라서 나뉘는 거고요. 그리고 이거는 방금 인호 님이 말씀해 주셨듯이 분석 결과 저장용 RDB인데 에이전트 3에서 사용한 겁니다.

왜냐하면 데이터 분석 및 예측 결과를 어 대시보드를 구성해야 되기 때문에 중앙 API는 장고에 있는 레스트 ipi에서 할 거고 램 그래프 기반의 에이전트 드레 워크플로우를 장고 백엔드 API로 제휴할 생각입니다.

설명을 보시면 알겠지만 전체적인 요청은 장고 레스트로 감싸져 있는데 뭐 필

요하다면은 후에 나중에 AWS 람다나 다른 백엔드 서비스로도 확장이 가능할 것 같아 에이전 호텔 연계 예시 한번 보여드릴게요.

참석자 1 05:46

네 KT 고객 이탈 예측 기업을 하나 딱 정해서 뭐 실제 있는 기업이 아니더라도 구한 다음에 한번 예측해 보라고 하셔서 가지고 이 연결 에이전트는 머신러닝 예측 그러니까 세 번째 에이전트가 전담을 할 거고 데이터를 먼저 설명해 드리자면 통신사 사용자 행동에 따라서 이 고객이 서비스를 계속 사용할지 혹은 이탈할지를 예측하는 머신러닝 모델이 학습이 되어 있습니다.

실제로 이걸 저희가 부트 캠프 예전에 사용해 봤던 그리고 실행을 해봤던 주제이기도 하고요.

참석자 2 06:18

했던 거 해도 상관이 없을까요? 제가 잘 몰라가지고

참석자 1 06:22

그 제안은 없던 걸로 합니다.

참석자 2 06:29

알겠습니다. 네 네

참석자 1 06:33

결과는 에이전트 3가 해당 모델 저희가 만든 머신러닝 모델을 호출해서 이탈 확률을 예측하고 결과는 포스트 그레이에 저장되거나 바로 사용자에게 쿼리나 대시보드로 보여주게 될 거예요.

이 밑에 있는 추가 가입 데이터 셋도 동일한 기능이고요.

그리고 이거는 저희가 사용하는 에이전트가 두 번째 그러니까 사내 문서 다큐

멘터리를 분석하는 에이전트인데요.

내부 PDF HTML 보고서 정책 문서 등 문서 내용은 벡터에 임베딩 돼서 저장이 될 거고 흐름은 에이전트 2가 벡터 검색을 통해서 유사한 문서를 추출한 다음에 문서 요약이나 문서에 대한 내용에 대한 질의 응답이 가능한 형태로 출력이 되게 될 겁니다.

참석자 2 07:23

잠깐 잠깐 저 위에 부분 같은 경우는 예 이게 파싱 같은 것들을 좀 해야 되긴 할 것 같거든요.

OCR 같은 거나 네 그래서 그런 전철이나 아니면 이제 쿼리 인리치먼트 같은 것도 한번 고려를 해봐야 될 것 같아요.

일단 이 부분은 추후에 이제 파이프라인 그려보면서 한번 이야기를 해보면 좋을 것 같습니다.

참석자 1 07:47

네 알겠습니다. 네

참석자 1 07:54

전체적인 통합 흐름은 다음과 같이 될 예정입니다.

사용자가 입력을 하면은 챗봇 UI에서 에이전트로 병렬 처리가 될 겁니다.

질에 따라서 그리고 의 사용되는 에이전트들도 각각 다르고 각 에이전트들은 이런 전문 분야들이 있을 것입니다.

참석자 2 08:11

권열 처리를 한다는 게 어떤 뜻이죠? 여기서

참석자 1 08:14

나뉘는 거죠 그 분기가 되는 거죠. 에이전트가 지리에 따라서

참석자 2 08:19

네 그게 그거랑 약간 분열 처리라면 네 이게 동시에 어떤 에이전트들의 작업이 들어가야 되는 걸로 제가 이해가 되거든요.

여기서는 그럼 제가 이거 다음에 수정하겠습니다.

네

참석자 2 08:38

3 에이전트라고 하는 것보다 멀티 에이전트 그냥 형태로 그냥 이제 멀티에이전트에서 분기가 작동한다기보다 멀티에이전트를 통해서 결과를 생성한다 이런 식이 좀 더 용어가 맞을 것 같아요.

참석자 1 09:04

감사합니다. 그래서 에이전트에 맞게 지리에 맞게 각 에이전트가 해당하는 답변을 내놓는 식으로 흐름이 될 것 같습니다.

그리고 마지막으로 질문별 에저트 작동 예시를 가져왔는데 첫 번째 에이전트 그러니까 코드나 SQL 쿼리문 오류 분석하는 에이전트죠.

예를 들어서 이 에스q 쿼리 원하요라고 물어보면은 랜 그래프에서 에이전트 원으로 분기가 될 거고 에이전트 1 f에서 코드 또는 쿼리를 감지한 다음에 벡터 검색으로 유사 코드 또는 지리를 검색을 할 겁니다.

그리고 LLM이 호출된 다음에 응답 데이터를 스트리트 아웃풋에 저장하고 장구 라이스 에서 저장된 응답을 반환을 할 거고요.

참석자 2 09:50

이거는 그러면 잇이 아니고 그냥 코드 자체인 거죠 네 그렇습니다.

약간 애매한 게 네 코드 자체만 분석하면 이 에이전트를 왜 써야 될지를 모르겠거든요.

참석자 1 10:07

포드 자체만

참석자 2 10:08

네 지금 챗gpt 쓸 거잖아요. 네네네. 7hpt에 있는 성능에 기대는 건데 실제로 어떤 특정 모델의 파인 튜닝을 하거나 오픈레미스 환경에 내부로 구축해서 보안 문서에 어서 나가면 안 되거나 이런 게 아니고 네 채팅 PT에 기댈 건데 채팅 피티에 질문하면 되지 이거 여기에 왜 해야 되는지 이해가 안 되거든요.

아니면 네 저번에 말씀드린 것처럼 네 이 SQL 커리 오류가 왜 나냐는 것보다 우리 사내에서 이런 걸로 이제 코드를 짜본 적이 있냐 이런 시스템에 대한 코드가 있냐 물어보면 네 거기에 해당되는 도큐멘테이션을 참고를 하고 거기에 이제 이슈 번호들이 있을 거잖아요.

저번에 말씀드렸듯이 그래서 그 이슈 번호에 해당하는 코드를 분석해 드릴까요? 이렇게 물어봤을 때 ys라고 누르면 해당 길 리파저 트롤을 참조해가지고 코드를 분석해 주는 그런 시나리오가 나올 것 같거든요.

참석자 2 11:11

그래야 좀 더 이제 사내 시스템이라고 부를 수 있으니까 그리고 회사 내에서 네 코드라고 하면 가장 많이 현재 최근에 엄청나게 관심이 많아요.

회사에서 코드 생성 쪽은 네 어떤 식으로 관심이 많냐면 두 가지가 있어요.

두 가지가 산에는 커서를 커서나 코파일럿이나 이런 걸 못 써요.

그렇기 때문에 이제 회사 내부에 자체적으로 구현을 해야 되거든요.

아니면 온프레미스로 나오는 설치형 플랫폼을 사야 돼요.

네. 그러면 이제 그게 챗gpt 모델이 아니고 작은 모델이 될 수도 있고 왜냐하면 또 운영을 해야 되니까 네네네.

그게 운영에 대한 코스트가 엄청 많이 드니까 자기 사내에만 있는 코드에만 잘 하면 되거든요.

네 그러면 한 두 가지 이슈가 있어요. 가장 관심이 많은 게 첫 번째로는 도큐멘테이션만 주면 그걸로 코드를 짜주는 거예요.

커서처럼 첫 번째가 그거고요. 일단 먼저 적으세요.

예

참석자 2 12:25

그래서 그냥 코드를 생성하는 게 아니고 폴더별로 코드를 생성해 주는 거죠.

두 번째는 이제 폴더별로 코드가 만들어져 있잖아요.

엄청 큰 시스템에 네 그거 자체에 대해서 다 분석을 해가지고 네 코멘트를 달아주고 그다음에 이제 레거시 같은 코드가 있거든요.

예를 들면 은행 거는 코볼 같은 걸 많이 써요. 근데 그거를 이제 자바로 변환하는 작업들을 많이 하거든요.

그래서 그런 코드 컨벌전을 많이 합니다. 그래서 지금은 우리가 데모니까 뭔가 이걸 전체를 하기보다는 파일 하나라도 하나 제대로 해보자 라는 생각으로 해보는 게 저는 좋을 것 같아요.

이 두 가지의 에이전트 안에 또 분기들이 있는 거죠.

참석자 1 13:20

아 네네

참석자 2 13:22

그래서 지금 보면 한 세 가지의 분기가 있겠죠. 예를 들면 이런 거 우리가 짜봤냐 물어봤을 때 사내 도큐멘테이션을 참고해서 이 코드가 있습니다.

혹시 이게 맞을까요? 아니면 분석해 드릴까요라는 질문이 나와서 분석을 해서 분석을 하는 코드 첫 번째 기 두 번째는 코드 생성해 주는 거 도큐멘테이션을 넣으면 네 그다음에 세 번째는 이 도큐멘트를 이제 전체 세 번째는 폴더 사내에 있는 폴더로 가면 이제 힘드니까 지금은 첫 번째로는 그냥 사내에 있는 어떤 코드 하나를 컨벌전 해 주는 거 예를 들면 코브를 잡아라든가 코브를 파이썬이라

든가 이런 식으로 그런 식으로 이제 우리가 에이전트를 만든다는 거는 그 안에 여러 가지 에이전트를 부른다는 건 그 안에서도 여러 가지 분기가 일어나는 거거든요.

참석자 2 14:13

네네네네 그래서 그런 식의 분기가 있으면 좋을 것 같은데 다 안 하셔도 되고 그래서 해보시면서 이게 좋을 것 같다.

이번에는 여기까지만 하면 될 것 같더라는 아이디어를 한번 잘 생각을 해보시고 하면 좋을 것 같습니다.

네 알겠습니다. 감사합니다. 회사에서는 이런 부분이 가장 관심이 많습니다.

참석자 1 14:30

네 알겠습니다. 그리고 일단 첫 번째 에이전트 1이었고요.

그다음에 두 번째 에이전트 바로 가겠습니다. 두 번째 에이전트는 사내 직원이라면 한 번쯤 물어볼 법한 질문인 것 같긴 한데 뭐 휴가 정책 알려줘 그러면은 사내 문서를 검색을 할 테고 문서를 검색한 다음에 무지하다 높은 문서만 추출을 해가지고 그 추출된 문서를 LRM이 입력해서 요약한 다음에 답변을 해주는 방식이에요.

참석자 2 14:57

네 이것도 좋아요. 너무 잘 만들어졌고 여기서 이제 우리가 좀 더 생각을 해보죠.

어떤 분기들이 있을까요? 여기는 지금 말한 거는 HR 정책에 대한 질문이라고 생각을 하거든요.

사내 문서에는 HR만 있지 않을 거잖아요. 그렇죠 프로젝트에 대한 질문이 될 수도 있고 네네네네 그러니까 이제 아니면 또 사주한테 할 만한 그런 질문들 있잖아요.

예를 들면 회의록을 참고할 수도 있고 아니면 우리가 특정 어떤 회의를 했을 때

그 회의의 요즘은 이제 그 회의를 했을 때 그 회의가 우리가 STT처럼 해가지고 우리가 회의했던 게 이제 글로 남겨져 있는 것들이 많거든요.

네네 그래서 그걸 참고해서 요약을 한다든지 이런 식의 내부 문서에 대해서 관심이 많단 말이에요.

네네 그런 여러 가지 분기가 있거든요. 그래서 그 분기들을 한번 좀 이야기를 한번 해보시면 지금 좋을 것 같거든요.

참석자 1 15:56

문서 카테고리별 분기로 나뉜다고 이해해도 될까요?

참석자 2 16:00

문서 카테고리별 분기도 될 수도 있고요. 네 그러니까 그러니까 질문에 유형에 대한 카테고리죠.

그러니까 HR적인 질문이거나 회의록에 대한 질문이거나 제품에 대한 질문이거나 이런 정도면 좋을 것 같네요.

이것도 제 아이디어고 제가 회사에서 가장 사람들이 관심 있어 하던 게 무엇일 까를 생각했을 때 했던 아이디어기 때문에 참고하시고 일단 나중에 보시면 좋을 것 같습니다.

참석자 1 16:31

네 알겠습니다. 감사합니다. 그리고 마지막 에이전트 하고 마무리할게요.

그리고 이 고객 이탈할까 머신러닝에 필요한 예측 부분 질문이죠.

에이전트 3로 분기가 될 거고 입력 데이터를 전처리한 다음에 머신러닝 모델에서 엔드 포인트를 호출하고 저희가 만든 모델이 예측 결과를 수신 해줄 거고 그 다음에 LM은 그 결과로 해석을 만들 겁니다.

그리고 결과는 포스트 브리 SQL에 저장 혹은 사용자에게 직접 반환을 할 거고 장고 API에서 응답을 출력을 해 줄 생각입니다.

참석자 2 17:07

네 이거는 좋습니다. 근데 이 고객이 이탈할까 이거는 약간 업무적인 질문이 될 수도 있잖아요.

네네네 그래서 여기서도 뭔가 분기를 하나 더 생각을 한다면 네 업무적인 걸 한 두 개 정도가 있을 거라고 생각해요.

지금 포스트 글에 SQL 저장을 한다고 하니까 또 여기서 추가적으로 텍스트 SQL을 하고 그걸 또 시각화하는 어떤 시나리오가 하나 있으면 너무 좋을 것 같 거든요.

예. 예를 들면 최근 어떤 최근에 어떤 업무 리포트를 참고하니까 우리 회사에 유입된 고객은 어떤 성별이 많고 그 고객이 이제 어떤 프로덕트를 가장 많이 샀는지 시각화해줘라는 만약에 예를 들면 질문이 들어왔어요.

그러면 거기에 해당되는 DB를 참고해서 그 DB에 텍스트 SQL을 날립니다.

방금 텍스트를 날렸잖아요. 자연어 질문을 그걸 SQL로 변환해줘서 저 DB에 날려가지고 포스트를 SQL을 날려서 이제 결과를 묻고 그 결과를 이제 파이썬 코드로 변환해가지고 시각화를 해주는 거죠.

그래서 그래프가 나오게 되는 거죠.

참석자 2 18:12

이 정도 두 개 정도의 분기가 있으면 거의 지금 회사에서 최근에 관심 있는 게 다 붙어 있다고 생각하시면 됩니다.

참석자 1 18:20

네 알겠습니다.

참석자 1 18:26

일단 아키텍처 설명도에 대한 내용은 여기서 마치도록 하고 이제 다른 팀원분들이 이제 해오신 시계열 데이터 분석이라든지 아니면 그 시장 조사 같은 거 한번 발표 부탁드립니다 될까요?

참석자 2 18:41

네네. 그러면 한 5분 정도 쉬고 30분에 그러면 한번 바로 시작을 해보도록 하겠습니다.

참석자 1 18:47

네 알겠습니다. 네 알겠습니다.

참석자 2 18:49

혹시 혹시 여기까지 질문 있으실까요? 아니요. 그럼 30분 후에 다시 이야기하도록 하겠습니다.

화장실 갔다 오세요. 네 네.

clovanote.naver.com