롯데이노베이트

AWS Summit 2025 참관후기

AI기술팀 이기훈 프로

2025.05.16

Change Today,
Create Tomorrow .

AWS Summit Seoul 2025

- **Information**
  참여날짜: 2025년 5월 14일 (수)
  장소: 삼성역 코엑스

Industry Day로 기술 트렌드와 생성형 AI만을 위한 트랙을 포함해, 업종별로는 리테일 및 소비재, 금융 및 핀테크, 제조 및 하이테크, 게임, 미디어 및 엔터테인먼트, 소프트웨어 및 인터넷, 헬스케어 및 공공부문, 통신, 여행 및 숙박 트랙이 있었습니다.

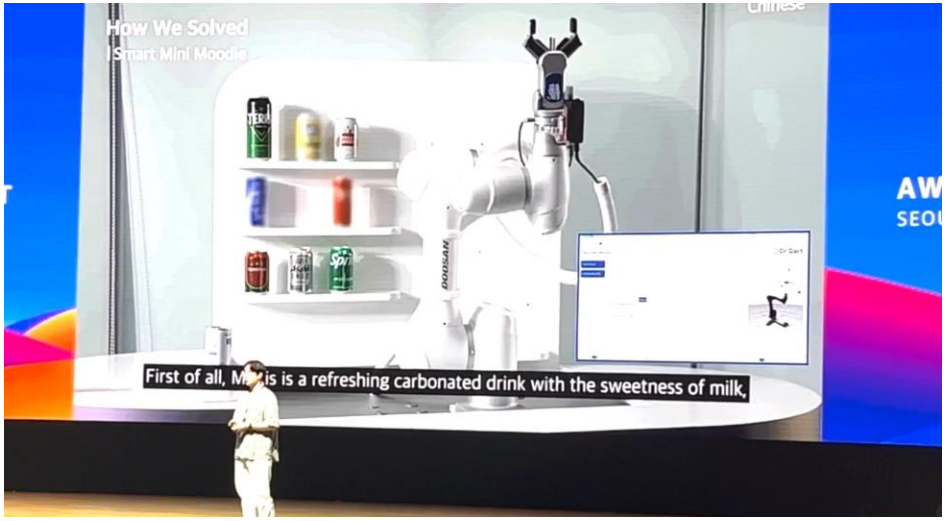
중점적으로 들은 내용은 다음과 같습니다.

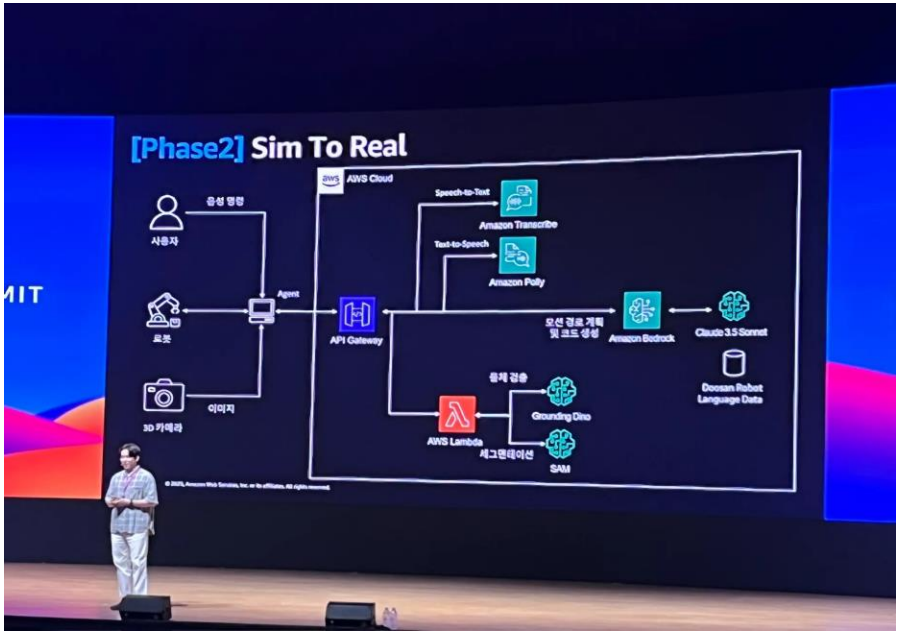
- 카카오스타일의 AI 기반 개인화 패션 추천
- 대한항공의 AI 챗봇 구축 사례
- 당근페이의 Text-to-SQL 전략
- 두산로보틱스의 멀티모달AI 기반 로봇 팔
- 데이터독의 할루시네이션 모니터링 방안
- 삼성전자와 에이블리의 운영 리소스 최소화 전략

전체적으로 Advanced RAG는 기본이 되었고, Text2SQL이 기존 대비 많이 사용되고 있으며, 할루시 최소화를 위해 엄청 신경쓴다는 느낌이 들었습니다. 이번 서밋을 통해 인사이트를 많이 얻었습니다. 감사합니다.

두산로보틱스의 멀티모달AI 기반 로봇 팔

- 강연명: 두산로보틱스/HL로보틱스와 함께알아보는 Physical AI와 IoT로 열어가는 새로운 로보틱스 시대





두산로보틱스에서 로봇팔을 만들고 있는데, 정적인 로직만 따르는 것이 아닌 상호작용되는 멀티모달 기반 로봇팔이 인상깊었습니다.

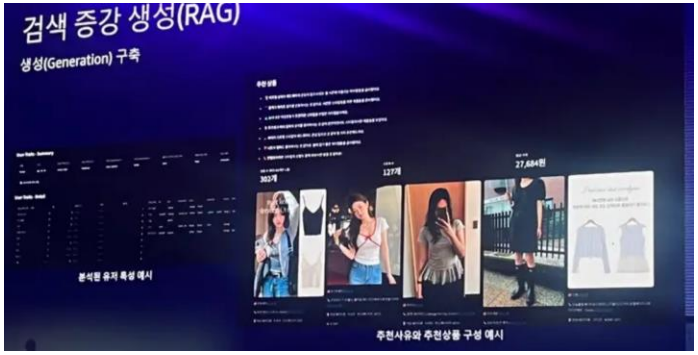
예시: “스프라이트 우측에 있는게 뭐야 그거 갖고와봐”

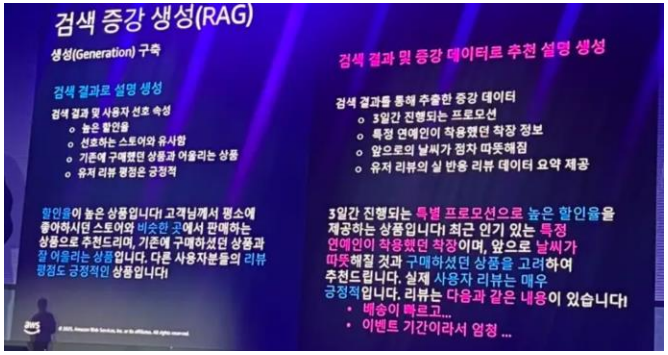
➔ Speech2Text + 이미지이해+ 세그멘테이션, 물체검출 > 멀티모달모델 > Text2Speech + 모션 경로 계획 및 Code 생성 후 실행

1단계에서는 유니티 기반의 환경에서 로봇 도메인으로 학습한 언어모델 기반으로 진행한 것 같고,
2단계에서는 더 성능이 좋은 상용 LLM 기반으로 테스트 진행한 것 같습니다.

카카오스타일의 AI 기반 개인화 패션 추천

강연명: AI가 뽑은 내 OOTD 카카오스타일의 개인화 패션 큐레이션 비하인드





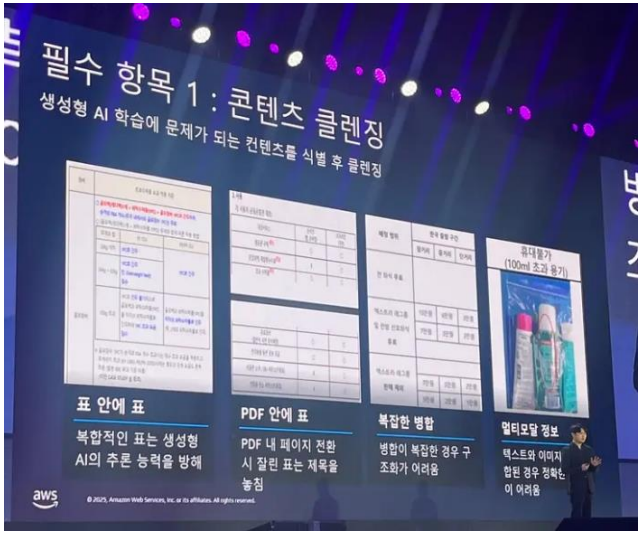
기존 1.0에서는 AWS 세이지메이커 사용해서 딥러닝 모델 사용하지 않고 ML을 사용했다고 합니다. 경량화된 메인모델 (lightGCN)을 사용해서 추천하고, 보조모델 등을 추가로 사용해서 콜드스타트 대응 모델, 개인화 재정렬 모델 등 만들어서 사용했다고 합니다.

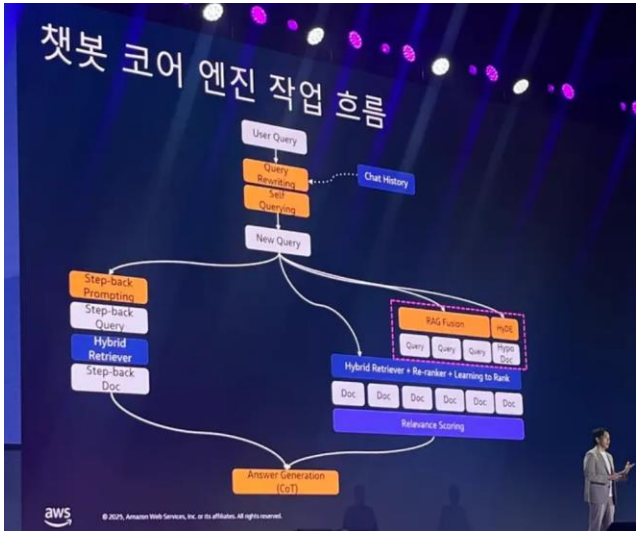
- 실시간성 부족(콜드스타트 문제, 최신정보 부족), 개인화 한계(과거 데이터에만 포커싱하다보니 정적이고 과도한 버블필터 발생)

2.0 에서는 콘텐츠 기반 추천으로 바꾸어 멀티모달, 실시간 개인화 추천, RAG를 진행하고 있다고 합니다.
상품 및 유저 속성 분석(카테고리, 스타일, 리뷰, 가격, 색상 등) 후 임베딩 진행 (이미지, 설명, 상품, 사용자 등)하여 유사도 매칭
클릭 등의 이벤트를 통해 필터를 즉시 넣어 기존 추천 강화 혹은 약화를 실시간으로 결정한다고 합니다.
리뷰같은 내용은 유저들 데이터만 사용하면 너무 반복적이고 특정 단어가 많이나오기 때문에 리뷰 스타일을 풍부하게 증대해서 쓴다고 합니다.

대한항공의 AI 챗봇 구축 사례

- 강연명: 대한항공의 AI 대전환: Amazon Bedrock 기반 AI 컨택센터 지식 검색 챗봇 구축 사례



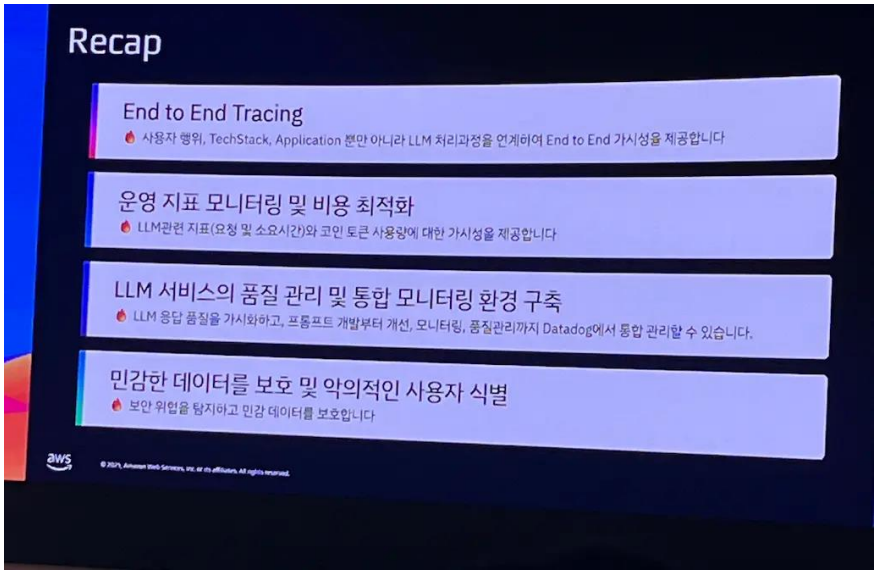


대한항공 상담원 지원 챗봇(QRS)에 대해 구축 과정에 대해 설명하는 강연이였습니다.
구축 과정은 일반적인 RAG기반 챗봇과 매우 유사하였고 특별한 점은 크게 없다고 느껴졌습니다.

저희도 한 때 많이 고민했던 문서 파싱 문제를 언급했는데요, 좌측 사진처럼 PDF나 Word 문서 안의 다중 중첩 표와 같은 복잡한 구조에 대해 문서 이렇게 만들지 말라는 최소 규칙을 정의하고 전부 리팩토링하고 가이드를 제공했다고 합니다.
또한, “이런 질문은 어렵습니다”라는 FAQ를 포함한 가이드를 제작해 상담원 교육을 진행했다고 합니다.
대한항공도 사용자 사전, 유의어 사전, 금기어 사전을 별도로 관리한다고 하며 key-vallue 형식으로 관리한다 합니다.
‘반려동물’과 ‘수하물’ 카테고리로 커스텀 데이터셋을 만들어 난이도별 llm-as-a-judge를 진행하며 성능을 확인한다고 합니다.

데이터독의 할루시네이션 모니터링 방안

- 강연명: LLM Observability: LLM의 잘못된 응답 (Hallucination)을 잡아내는 법

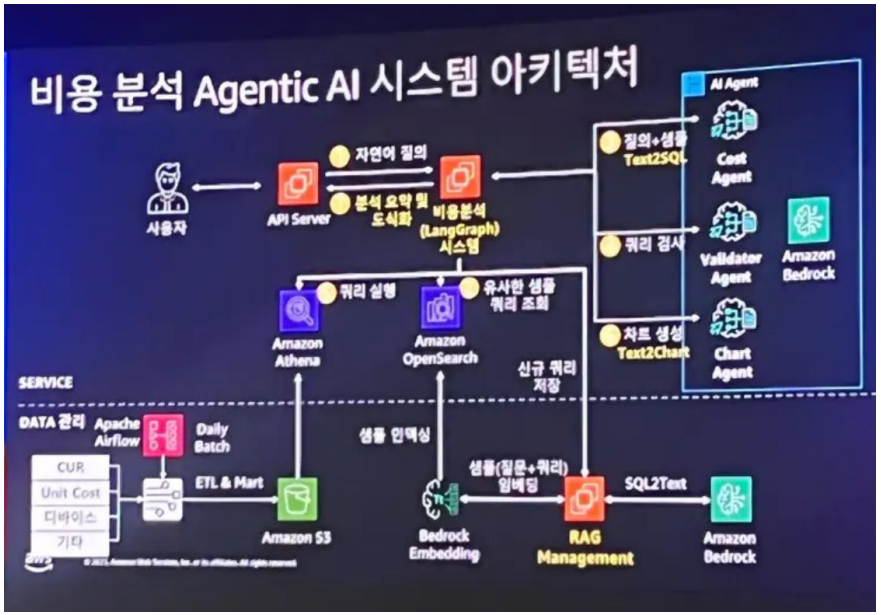


데이터독에서 제안하는 할루시네이션을 포함한 답변 품질을 모니터링하는 방법은 다음과 같습니다.
“운영에 영향을 주지 않는 별도의 모델을 하나 띄워서 사용자 질의와 답변을 입력해서 퀄리티 점검하는 방안”
이 방안은 보안위협 탐지 및 대응이 가능하고, 개인정보 마스킹되었는지 등도 확인 가능하며,
데이터독이 말하는 할루시네이션은 “RAG등을 통해 제대로 가져 왔는데 답변만 이상하다는 가정”을 통해 ‘답변이 컨텍스트에 내용과 맞는지’에 대한 검증을 통해 틀리면 할루시라 판단한다고 합니다.

별도 모델을 통해 판단해서 모니터링 하는 방안은 괜찮아 보이나 리소스를 고려해봐야할 것 같습니다.

삼성전자의 Text2chart 방안

- 강연명: 생성형 AI 시대의 클라우드 혁신: 삼성전자와 에이블리의 FinOps에서 Agentic AI까지



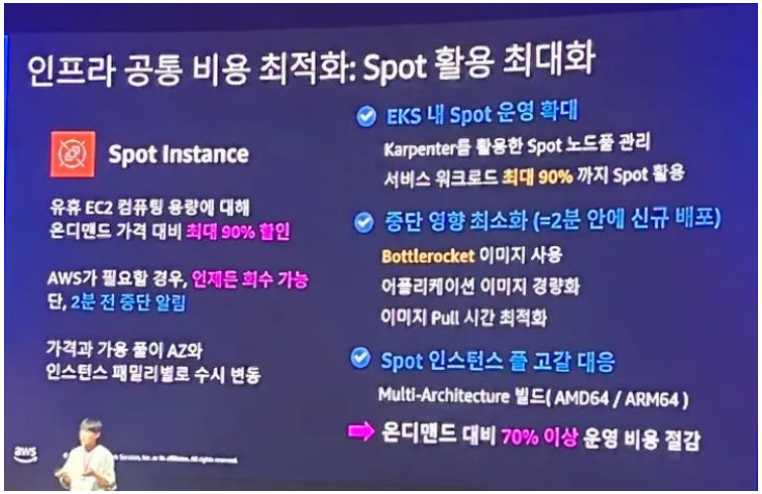
삼성전자는 비용 관련한 내용을 질문하면 답변 + 차트가 나오는 내부용 대화형 BI가 있는 것 같습니다.

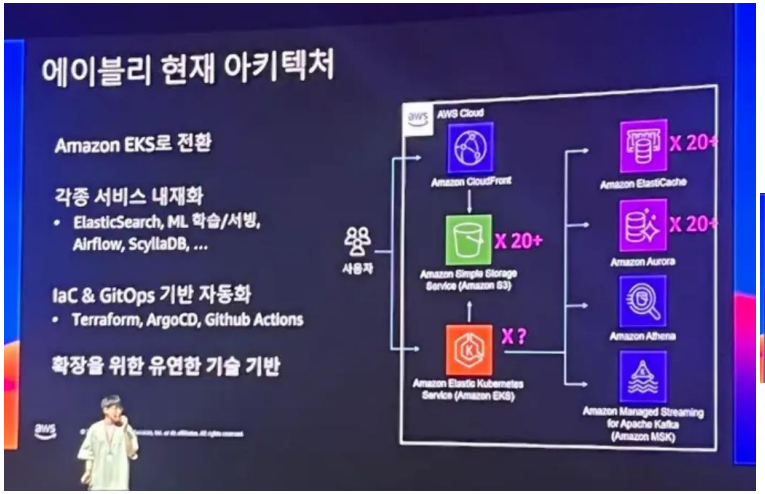
아키텍처와 시연영상을 보니 SQL Query와 Results를 입력으로 사용해서 모델을 태워 matplotlib 기반의 Chart를 만드는 파이썬 코드를 생성하라고 지시한 다음, 코드를 실행하여 이미지로 만든 다음 이를 대화창에 출력하는 방식을 사용하는 것으로 파악했습니다.

다른 내용은 일반적인 대화형BI 아키텍처를 따르는 것으로 확인하였고,
아이맴버에도 Text2Chart 기능을 추가하면 좋을거같다는 생각을 하였습니다.

에이블리의 운영 리소스 최소화 전략

- 강연명: 생성형 AI 시대의 클라우드 혁신: 삼성전자와 에이블리의 FinOps에서 Agentic AI까지



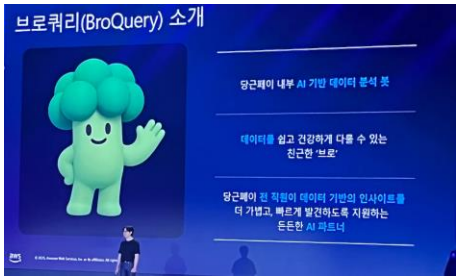


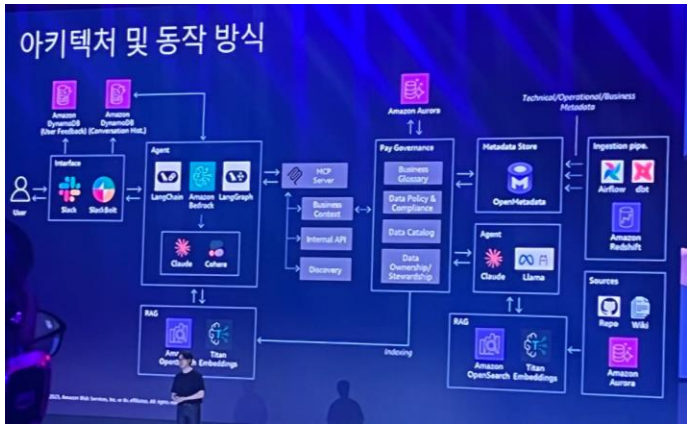
에이블리는 간단한 아키텍처에서 시작했는데, 점점 사용자가 증대되면서 비용이 왜 올라가는지 트래킹이 힘들었다고 합니다. (단순히 유저가 늘어나서인지, 특정 로직에서 잘못된건지, 복합적인 이유가 있는지 등등)

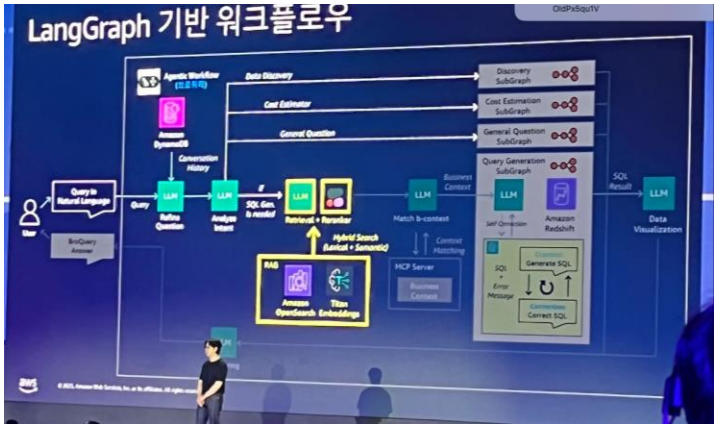
자체 모니터링도 진행하는데 Spot Instance을 활용하도록 바꾼게 비용절감에 가장 효과적이였다고 합니다.
이미지 생성도 spot 비용 낮은데 공급 많은 미국 오리건 gpu쓰고 있다고 합니다. (100% spot 인스턴스 사용)

당근페이의 Text-to-SQL 전략

- 강연명: Amazon Bedrock 기반 Text-to-SQL로 완성하는 데이터 혁신: 당근페이의 핀테크 성공 전략







당근페이 내부 임직원들 타겟으로 사용되는 Text2SQL 기반 슬랙봇 브로쿼리 내용입니다.

슬랙에서 사용자의 질문이 들어오면, DB에서 과거 대화 히스토리를 조회하고, 질문의 의도를 파악하는 에이전트가 동작합니다.
질문이 일상 대화인지 데이터 분석 or DB 조회가 필요한 질문인지 등을 분류해서 MCP로 넘어갑니다.

SQL 쿼리 생성 과정에서 오류가 발생하면 에러 코드와 함께 쿼리를 다시 전달해 재생성하는 재시도 로직을 사용한다고 합니다.
DDL, 스키마, 메타데이터, 컬럼 설명 등을 포함해 만들고 있고, 좋은 샘플 쿼리를 퓨샷으로 넣어 정확도를 높이고 있으며,
모든 유형의 샘플 쿼리를 만드는 것이 어렵기 때문에 별도의 용어사전을 구축해 이를 기준으로 대응하고 있다고 합니다.
추가적으로, RAG 할 때는 하이브리드 검색과 리랭커를 한 다음 사용한다고 합니다.

MISSION

사랑과 신뢰를 받는
제품과 서비스를 제공하여
인류의 풍요로운 삶에 기여한다

We enrich people's lives by providing
superior products and services that
our customers love and trust

롯데이노베이트