

제 2회 LLM 밋업

LLM이 흔들어 놓았던 업스테이지

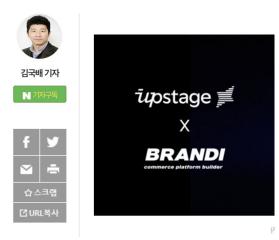


1. LLM으로 가는 길

LLM 등장 전… (추천 & 검색)

업스테이지, 패션 플랫폼 '브랜디'에 추천AI 제공

"앱 첫 페이지부터 초개인화된 제품 추천"



IT-과학

우수민 기자

입력: 2022-07-04 15:19:01

`재능발달 추천 AI 솔루션 계약`

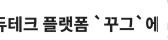
"스타트업 AI 생태계 확대 노력"

*iy*pstage

aipack | RecSys

클래스 추천 AI 공급

업스테이지, 에듀테크 플랫폼 `꾸그`에



() 꾸그



롯데온, 업스테이지 추천 AI 서비스 도입

입력 2023.06.05. 오전 9:07 기사원문













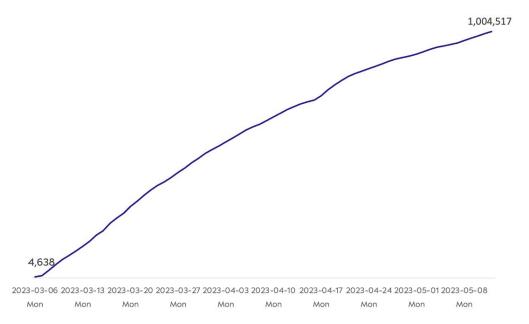
7일부터 개인화 추천 서비스 영역에 AI 기술 적용 롯데온-업스테이지, 약 2년간 기술 도입 준비

[이데일리 백주아 기자] 롯데온은 오는 7일부터 글로벌 인공지능(AI) 기업인 업스테이지의 상품 추천 AI를 도 입한다고 5일 밝혔다.



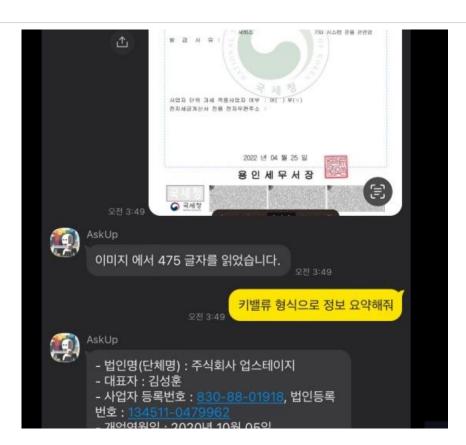
ChatGPT 등장 - AskUp 출시

Over 1,000,000 Users in 2 months

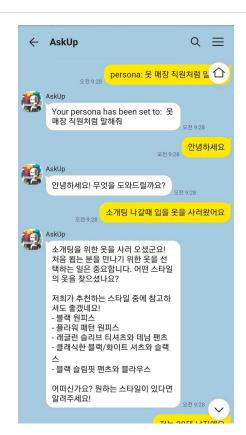


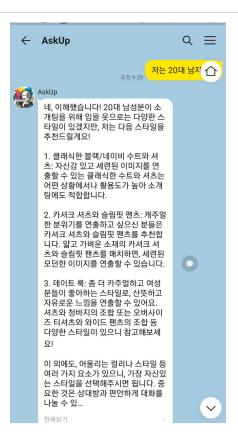


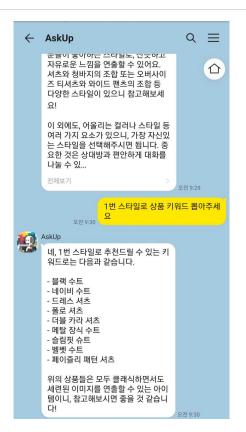
LLM + OCR - 눈달린 ChatGPT



오… 추천에도 써볼 수 있겠는데?





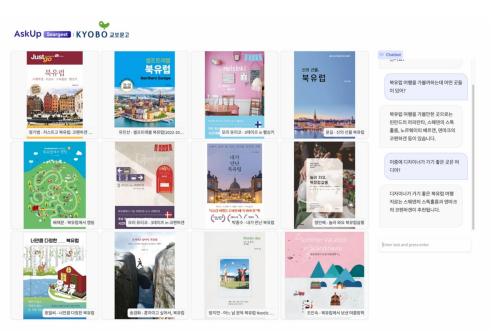


AskUp Seargest 제안

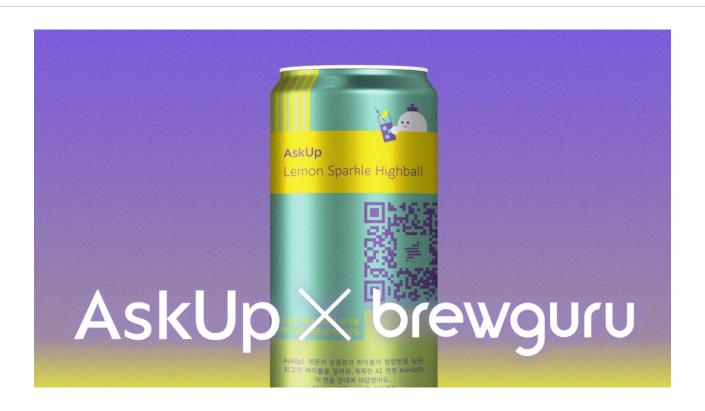
'애스크업 서제스트' 공개..AI 대중화 속도 "내년 BEP 달성…일본·동남아 등 진출 시동"



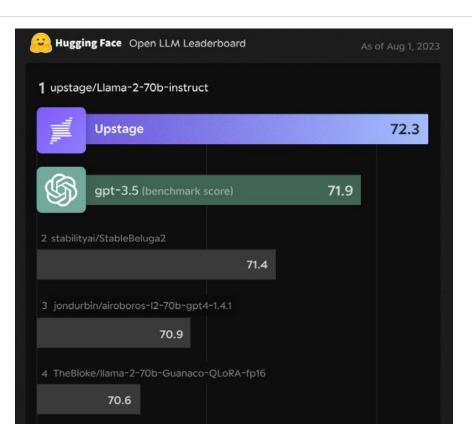
김성훈 업스테이지 대표가 16일 서울 중구 페럼타워에서 열린 기자간담회에서 '애스크업 서제스트'를 소개하고 있다./사진=비즈워치



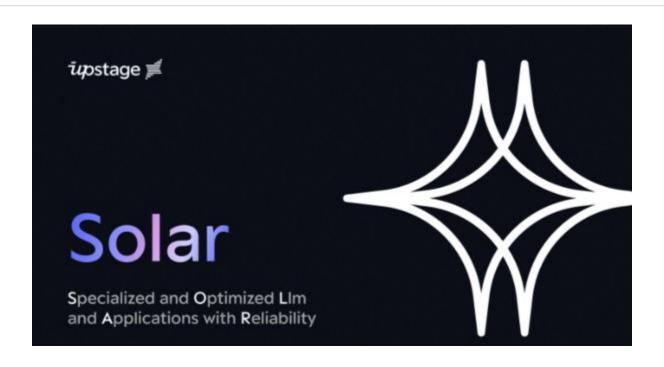
AskUp 하이볼도 출시



OpenLLM Leaderboard 1등…?!



LLM으로 제대로 해보자!



2. LLM Application 팀으로

보험사 사례

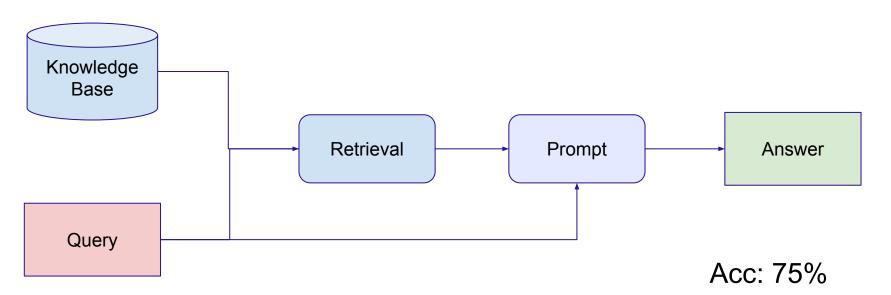
질문 예시	답변 예시
내가 대장암에 걸렸는데, 수술자금 보장 받을 수있어?	제공된 보험 증권 정보에 따르면 대장암 수술자금에 대한 보장 내용을 포함하고 있습니다. 상세 정보는 아래 내용을 확인해주세요.
	▲암수술자금 ▲암다빈치로봇수술S(갱)

보험금의 지급사유 (암수술자금) - 보장 여부 판단 근거

보험기간 중 피보험자가 보장개시일(암(기타피부암, 갑상선암, 대장점막내암 제외) 또는 갑상선암은 암보장개시일) 이후에 암 (기타피부암, 갑상선암, 대장점막내암 제외), 기타피부암, 갑상선암, 대장점막내암, 제자리암 또는 경계성종양으로 진단이 확정되고, 보험기간 중 그 치료를 직접적인 목적으로 "수술분류표"(약관 참조)에서 정한 수술을 받았을 경우

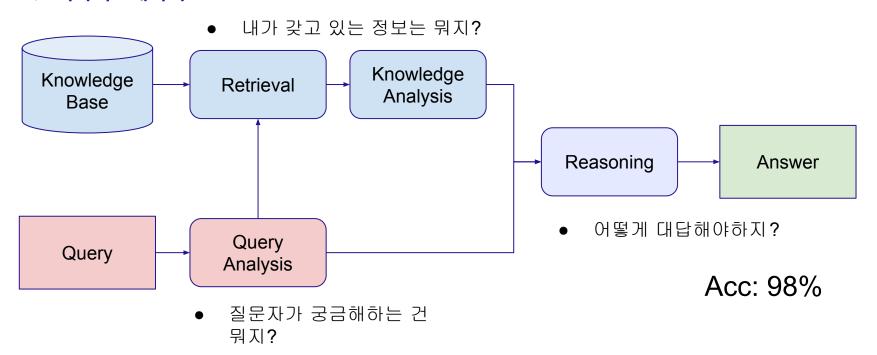
보험사 사례

Prompt에 관련 정보를 모두 넣고 LLM의 판단에 맡기는 Case (a.k.a 그냥 RAG)



보험사 사례

QA 시나리오에서의 RAG



보험사 사례: RAG 생각보다 어렵다

느낀점1

- 처음에는 **단순 RAG**로 문제가 풀릴 것이라고 생각했음.
- 실제 문제를 풀 때는, **복잡한 RAG**가 필요함.
- 복잡한 RAG란? -> **Human Mental Model**을 Following하도록 설계한 Pipeline

느낀 점 2

- 고객의 Tipping point는 정확도로 정의되지 않는다
- 정확도가 높다고 과연 많이 Buy할 것인가?

아티스트의 정체성, 말투, 대화 학습해 실제 같은 대화 가능한 '페르소나 AI'



서우원 메타버스엔터테인먼트 대표(왼쪽), 김성훈 업스테이지 대표가 협약식에서 기념 촬영을 하고 있다. <출처=넷마블>





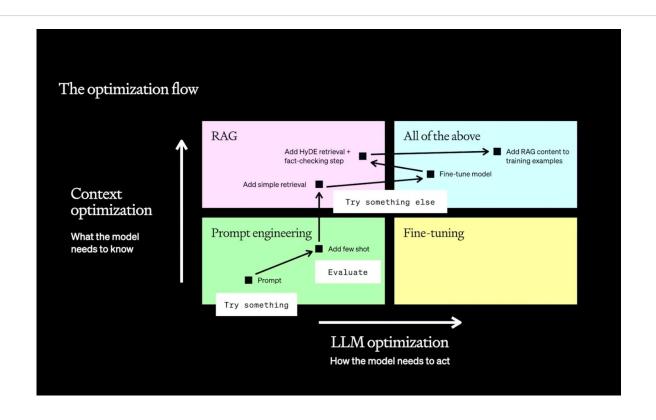
페르소나AI: Finetuning까지 할 수 있다면 더 Powerful

느낀점1

- 처음에는 RAG로 자연스러운 말투를 정의하려고 접근함.
- 잘 안돼서 Fine-tuning을 사용해서 접근함.
- 최종적으로, Fine-tuning + RAG를 사용함
- 자연스러운 말투 / 캐릭터 설정 기억 / 민감발언 대응 / 사용자에 대한 기억

느낀점2

- RAG에 사용할 데이터를 만드는 것이 어렵다
- 데이터를 만드는 유저 피드백 Pipeline을 어떻게 구성할 것인가?



3. 다시 One Team으로

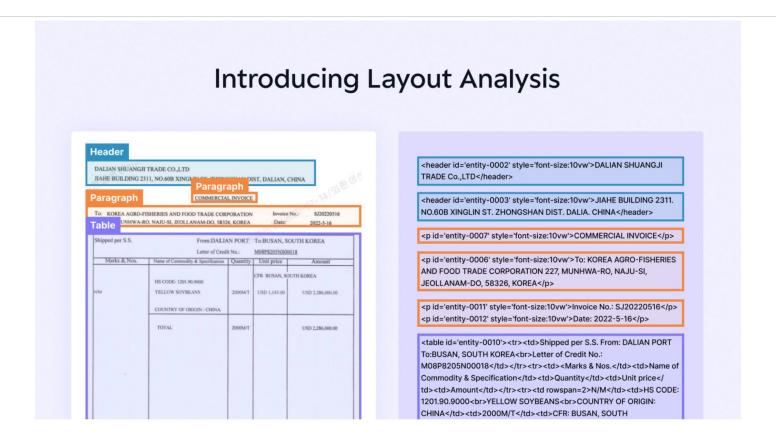
SOLAR의 성능 업





지난달 29일(현지시간) 이활석 업스테이지 최고기술책임자(CTO)는 클라우드 컨퍼런스인 '아마 존웹서비스(AWS) 리인벤트(re:Invent) 2023'이 열린 미국 라스베이거스에서 기자들과 만나 "연내 파라미터 수가 각각 70억개, 107억개인 생성형 AI의 거대언어모델(LLM) 파운데이션모델 (FM·기반 모델)을 출시할 예정"이라며 "내년 1분기엔 300억개 규모의 FM도 선보일 계획"이라고 밝혔다. 파라미터 수는 인공 신경망의 규모를 의미한다. 규모가 클수록 많은 양의 데이터를 학습, 처리할 수 있다.

Layout Analysis for RAG



4. 배우고 느낀 점

LLM B2B를 하면서 배우고 느낀점!

B2B 시나리오에서는 "야 신기하다~ 재밌다~"로는 어렵다.

• 고객이 실제로 Buy 하려면 진짜 고객이 어려움을 겪고 있는 문제를 풀어야 한다.

하지만 많은 고객들이 LLM을 도깨비 방망이처럼 생각한다.

● LLM이 풀기 어려운 문제, LLM으로 풀 필요가 없는 문제도 모두 LLM으로 풀어보자고 한다.

그래서 앞으로 중요한 것은 <u>"문제를 정의하는 능력!"</u> 일 것이다

- 문제를 예전보다 정말 쉽게, 빠르게 풀 수 있는 도구가 나왔다.
- 예전에는 문제를 푸는 능력이 중요했다면, 앞으로는 풀어야할 문제를 정의하는 능력이 중요해질 것이다.