
Corca LLM

코르카 테크 리드 이태호



목차

- 지난 18개월 동안 코르카가 바라본 LLM
- 코르카 에이전트의 시작
- 코르카의 LLM 프로덕트
- 코르카의 LLM 연구
- Q&A

지난 18개월 동안 코르카가 바라본 LLM

2022년 4월 LLM리서치 및 외부초청 세미나 시작

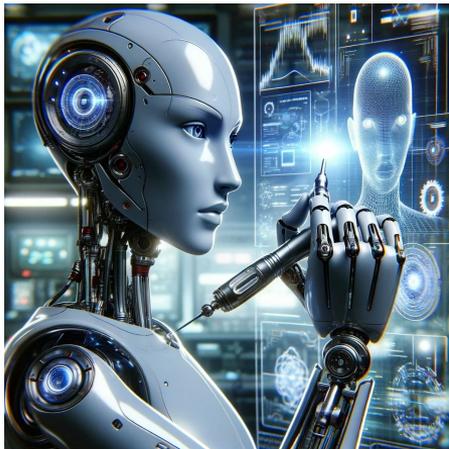
- LLM으로 무엇을 할 수 있는가
- 여기에서 모델의 크기가 더 커지면 어떤 것이 가능할까
- 어떤 비즈니스를 해야 할까

등등에 대한 고민...

코르카 에이전트의 시작



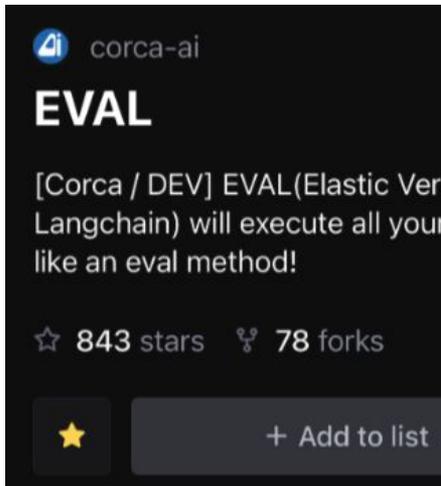
에이전트의 개념



- LLM을 생각할 수 있는 브레인이라고 한다면,
- 에이전트는 **미션, 행동, 도구, 외부 지식** 등과 함께 LLM을 활용하는 객체의 단위이다.

- Text completion보다 한 차원 높은 관점으로 볼 수 있도록 함

EVAL



- Terminal, Code Editor 도구 개발을 통해 직접 코딩, 디버깅, 테스트, 배포할 수 있도록 함
- Langchain 활용

- 전세계 최초로 웹서비스를 개발하고 서빙하는 첫 에이전트였음



EVAL

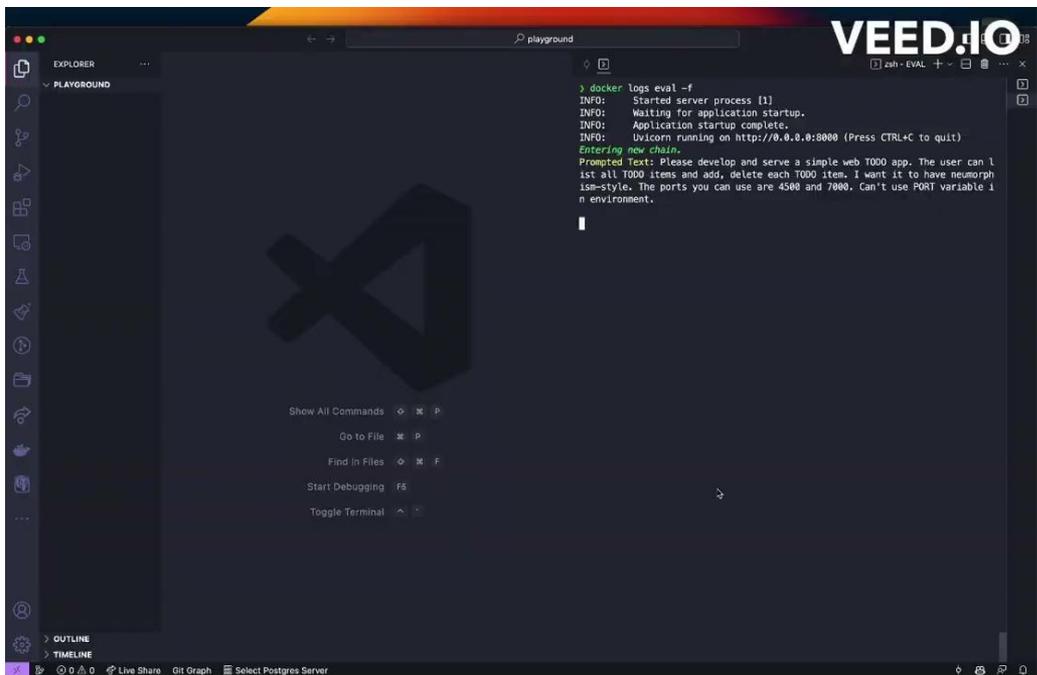
corca-ai

EVAL

[Corca / DEV] EVAL(Elastic Ver Langchain) will execute all your like an eval method!

843 stars 78 forks

+ Add to list



Alvilization



- 에이전트가 에이전트를 만들어낸다면?
- 에이전트가 에이전트를 교육시킨다면?
- 도구를 직접 만들고 사용하면서 개선한다면?
- 인사이트를 서로 공유한다면?
- 같이 문제를 풀어가면서 각자의 전문성이 생긴다면?

AGI의 궁극적인 형태이며 지능 하나가 아니라 **지능의 집단이자 문명**

Alvilization



- 같이 문제를 풀어가면서 각자의 전문성이 생긴다면?
 - ⇒ 대화를 하면서 학습할 수 있을까?
 - ⇒ in-context learning으로는 잘 되지만 확장가능한 방법이 아니다. 비용과 시간이 계속 증가하기 때문이다.
 - ⇒ context를 즉시 parameter에 반영해버린다면?
-
- To be continued

코르카의 LLM 프로젝트



LLM으로 성공할 프로덕트는?

- 파도를 거스르지 말고 파도 위에 올라타라.
- Do things that don't scale.



- OpenAI, Google, Meta 등이 연구를 하면 할수록 즉시 우리의 프로덕트가 업그레이드될 수 있도록 해야 한다.
 - 전 세계로 뻗어나갈 수 있는 프로덕트를 꿈꿔야 한다. 하지만 지금 당장 고객 하나의 문제를 해결할 수 있어야 한다.
-

코르카가 생각한 서핑 방법



1. **Agent를 새로운 공간으로 옮기자! ⇒ ADCIO Agent**
 - Copilot이 잘 된 이유는 성능보다 쓰기 쉬워서이고 쓰기 쉬운 이유는 코딩창 안에 있기 때문이다.
2. **Agent간의 협력 구조를 만들자! ⇒ Agent Village**
 - 에이전트 1명이 감당할 수 있는 문제의 범위와 깊이는 한정적이다.
 - 확장 가능한 에이전트 협력 구조를 만들자.

ADCIO Agent



- 검색(Search)이 잘 되는 이커머스 플랫폼은 많지 않다.
- 대화를 토대로 사용자의 니즈를 분석하여 상품을 추천해준다면?
- 대화 지면을 광고에 사용한다면?

RecSys with LLM보다 Conversational RecSys에 집중

ADCIO Agent

대화를 통해 유저의 의도를 얼마나 정확하게 이해하고
얼마나 정확하게 수행할 수 있는가?

- Keyword Search (Elastic Search, ...)
- Vector Search (Weaviate, Pinecone, ...)
- Text to SQL

알고리즘, 머신러닝, LLM 각각의 장단이 있기에 장점만을
뽑아낼 수 있도록 해야 함



ADCIO Agent



- 에이전트를 매장의 점원으로 만들자. (Microsoft Copilot은 개인의 비서 컨셉)
- 매장의 점원은?
 1. 굳이 먼저 말 걸지 않는다. 점원을 필요로 할 때 조심스럽게.
 2. 매장에 대한 정보, 매장에서 파는 상품들에 대한 정보를 다 안다.
 3. 상품을 추천해줄 수 있다.
 4. 매장 안에서 길을 안내해줄 수 있다.
 5. (상상) 누군가 들어오면 그 사람에게 맞는 상품들로 다시 진열한다.

Next UX

기존의 인터페이스와 채팅 인터페이스의 융합



- 웹 read → 채팅: 페이지 콘텐츠 기반 Q&A
- 웹 write → 채팅: 사용자 의도 파악
- 채팅 → 웹 read: Browsing
- 채팅 → 웹 write: Next Browser

TARGET AGAIN



23 SS Collection with Cute Characters colorful graphic designs



- V of BTS WORE
- ARTIST COLLABO
- NEW
- BEST
- MBTI
- T - SHIRT
- HOODIE
- SWEATSHIRT
- ACCESSORIES
- HEADWEAR

- ABOUT US
- LOOKBOOK
- COMMUNITY

- Kor / Eng
- Instagram
- Facebook
- Blog

©2019-2019 TARKE
ALL RIGHT RESERVED.



TARKE AGAIN

target target

23 SS Collection with Cute Characters colorful graphic designs



어름 휴기에서 입을 만한 옷으로는 '반팔 티셔츠', '스웨트셔츠', '악세서리', '모자' 등이 있어요. 어떤 스타일을 선호하시나요? 색상이나 추가적인 요청사항이 있다면 알려주세요. 제가 도움을 드릴게요!

반팔티셔츠 줄래?

말씀하신 상품을 보여드릴게요! 상품을 선택하시면 상세 페이지로 이동합니다.



메시지를 입력하세요 전송

Agent Village



- 자연 확장 가능한 **Artificial Civilization**은 매우 어렵다.
- 조금 더 심플한 에이전트 마을을 만들자!
- 에이전트 간의 **협력**으로 새로운 가치를 만들어보자.

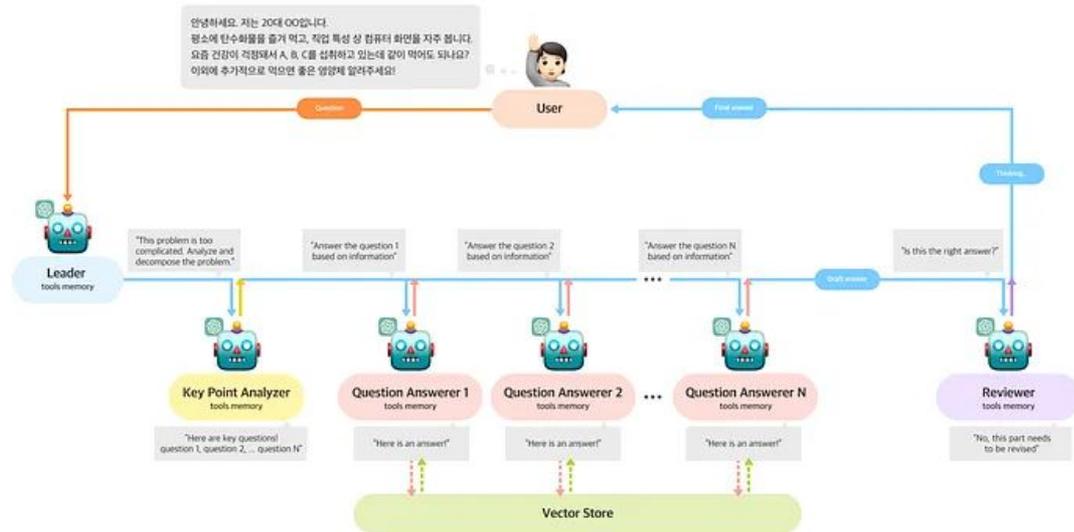
현재) 미션, 행동, 행동 순서, 도구, 외부 지식 **고정**. 도구 사용
자유.

향후) **Autonomous Agent**로의 진화

행동 순서 → 도구 생성 → 외부 지식 → 미션 자율화

Agent Village

< Multi Agent Problem Solving Architecture >



하루에 아래 영양제들 한 번에 섭취해도 되나요?

- 쏘리서치SAT
- 교릭 숙성마늘
- 안국코큐텐
- 고려은단비타민C1000
- 나우푸드전립선

안녕하세요. 파마브로스 약사입니다. 😊 여러 가지 영양제를 동시에 복용하실 수 있는지에 대한 질문이시군요.

질문하신 Thorne Research SAT, Kyorick aged garlic, Anguk CoQ10, Korea Eundan Vitamin C1000, 그리고 Now Food Prostate는 모두 함께 복용하셔도 안전합니다. 각 제품의 성분이 중복되지 않으며, 서로 간섭하는 성분도 없습니다.

🕒 복용 시간

아침 식사 후: Anguk CoQ10, Korea Eundan Vitamin C1000, Now Food Prostate

점심 식사 후: Thorne Research SAT, Kyorick aged garlic

💡 복용 주의사항

마늘: 특별한 주의사항이 없습니다. 하지만, 마늘 알러지가 있으신 분은 복용을 피하시는 것이 좋습니다.

CoQ10: 특별한 주의사항이 없습니다. 하지만, 혈액 희석제를 복용하시는 분은 의사와 상의하신 후 복용하시는 것이 좋습니다.

비타민 C: 특별한 주의사항이 없습니다. 하지만, 위장장애가 있으신 분은 식사 후 복용하시는 것이 좋습니다.

※ 식단 및 생활습관

- 균형 잡힌 식단을 섭취하시는 것이 중요합니다. 특히, 신선한 과일과 채소를 충분히 섭취하시는 것이 좋습니다.

- 충분한 수분 섭취와 꾸준한 운동도 건강에 도움이 됩니다.

😊 기능성 식품은 약과 대체되는 것이 아니며, 기능성 식품과 약을 함께 복용하기 전에는 반드시 의사와 상의하시기 바랍니다.

특정 성분에 알레르기가 있거나 제품이 유통기한이 지난 경우에는 섭취하지 않도록 주의하시기 바랍니다.

도움이 되시길 바라며, 추가적인 질문이 있으시면 언제든지 문의해 주세요.



고려은단비타민
C1000...



NOW Foods,
전립선 건강 지...



Thorne SAT



Kyorick, 숙성 마
늘 추출물, 액...



안국건강 코엔자
임Q10

코르카의 LLM 연구



Alvilization



- 같이 문제를 풀어가면서 각자의 전문성이 생긴다면?
 - ⇒ 대화를 하면서 학습할 수 있을까?
 - ⇒ in-context learning으로는 잘 되지만 확장가능한 방법이 아니다. 비용과 시간이 계속 증가하기 때문이다.
 - ⇒ context를 즉시 parameter에 반영해버린다면?
-
- To be continued

Real-Time Reverse Fine-Tuning

- context를 실시간으로 parameter에 반영해버리자.
- In-context learning을 대체하자.

Advantage

1. 추론 비용 대폭 감소
2. Text-based RLHF를 통해 Agent 진화
3. 에이전트 간의 대화를 토대로 실시간 학습 가능
⇒ Multi-Agent 협업 시 job이 생김

Real-Time Reverse Fine-Tuning



```
nvtop (Press h for help or q to quit)
Med Nov 22 06:18:10 2023
NVITOP 1.3.0 Driver Version: 535.104.05 CUDA Driver Version: 12.2

GPU Fan Temp Perf Pwr:Usq/Cap Memory-Usage GPU-Util Compute M.
 0 N/A 53C P0 73W / 400W 78.99GiB / 80.00GiB 0% Default
 1 N/A 46C P0 69W / 400W 7.75MiB / 80.00GiB 0% Default
 2 N/A 43C P0 69W / 400W 7.75MiB / 80.00GiB 0% Default
 3 N/A 51C P0 70W / 400W 7.75MiB / 80.00GiB 0% Default
 4 N/A 51C P0 68W / 400W 7.75MiB / 80.00GiB 0% Default
 5 N/A 43C P0 69W / 400W 7.75MiB / 80.00GiB 0% Default
 6 N/A 41C P0 65W / 400W 7.75MiB / 80.00GiB 0% Default
 7 N/A 47C P0 68W / 400W 7.75MiB / 80.00GiB 0% Default

Load Average: 2.09 2.03 2.01
CPU: 1.6%

120s |-----60s-----30s|-----
MEM: 14.76GiB (2.1%)
SWP: 0.49GiB (0.1%)

Processes:
GPU PID USER GPU-MEM %SM %CPU %MEM TIME COMMAND
 0 114135 C onrgrou+ 30.07GiB 0 0.0 0.3 22:52:32 /home/onrgroun/a...
 0 114514 C onrgrou+ 48.89GiB 0 0.0 0.4 22:51:09 /home/onrgroun/a...
```

Our Topics:

- Meta Learning (MAML)
- HyperNetwork
- Task Vector
- Zero-shot task adapter
- Test-time adaptation
- Continual learning
- Etc...

With NVIDIA A100 80GB X 8개!!!!!!



Q&A



Q&A

Q. 코르카의 ML, Backend 엔지니어는?

A.

- NLP 연구 (transformer 날날히 파헤쳐보기, Pytorch의 날날히 파헤쳐보기, 논문 겁나 읽기, 구현 겁나 하기)
- Search 연구
- Agent 개발 (프롬프팅, 도구 개발, 자율화 등)

Q&A

Q. sLLM, Private LLM, Open-source LLM에 대한 생각

A.

- 내 비밀을 지켜주고 답장이 빠른 IQ 100 친구와 동네방네 떠벌리는 IQ 150 친구가 있다면 누구한테 물어볼 것인가?
- Open-source LLM의 강점은 **Controllability**이다. 이를 살릴 수 없다면 무의미하다.
- **Controllability**를 살릴 수 있는 연구 필요하다.

Q&A

Q. LLMOps? LangChain? llama-index?

A.

- Do things that don't scale.
- 우리가 할 수 있는 건 우리가 한다. 설부른 추상화는 하지 않는다.

We're Hiring!

