

에이전틱 Loop를 활용한 LLM 성능 향상

AI Creatrip을 만들면서 배운 점들

크리에이트립 서종학

자기소개

크리에이트립 개발팀 리드 (얼마 안 됨)

궂은일 담당

Creatrip 소개

전 세계 사용자를 위한 서비스

한국에 대한 모든 것을 제공

- 예약, 숙박, 체험, 컨텐츠 등등

서비스 제공 국가 20개국 이상

- 지원 언어 14개

다국어 지원 화면 예시

The image displays three side-by-side screenshots of the creatrip website, illustrating its multilingual capabilities. Each screenshot shows a different language version of a travel listing for Changdeokgung Palace.

Left Screenshot (English Version):

- Header: TRAVEL, STAYS, creatrip, TRENDS, LANGUAGE, search icon, location icon, heart icon, user icon.
- Image: Seoul Jongno, Changdeokgung Moonlight Festival.
- Title: 영어
- Text: [Creatrip Only] 2025 Changdeokgung Moonlight Festival Foreigners Only
- Rating: ★★★★★ 5 - 79K+ reviews
- Price: 21.1 USD
- Info: More than 882 Global travelers have booked this place! Instant Book: Payment instantly confirms the reservation.
- Point: Maximum 0.47 Points

Middle Screenshot (Japanese Version):

- Header: 韓国旅行, 韩国宿泊, creatrip, 韩国购物, 韩国新闻, 语言学校, search icon, location icon, heart icon, user icon.
- Image: Seoul Jongno, Changdeokgung Moonlight Festival.
- Title: 일본어
- Text: 2025 昌徳宮「月明かり紀行」(外国人チケット独占)
- Rating: ★★★★★ 5 - 79K+ reviews
- Price: ¥ 3,178
- Info: 日本の旅行者がこの場所を 206 件以上予約しました! この商品を旅行プランに追加しているユーザーが増えています。即時確定 支払いが即時に予約を確定します。
- Point: 最大72 ポイント

Right Screenshot (Korean Version):

- Header: 韩国旅游, 韩国住宿, creatrip, 韩国购物, 韩国新闻, 语言学校, search icon, location icon, heart icon, user icon.
- Image: Seoul Jongno, Changdeokgung Moonlight Festival.
- Title: 대만 번체
- Text: 獨家發售! 2025 昌德宮《月光之旅(月光紀行)》門票 (9/28、10/5、10/12、10/19)
- Rating: ★★★★★ 5 - 79K+ reviews
- Price: TWD 646
- Info: 臺灣旅客已經預訂這個地點超過 192 次! 此商品近期人氣上升中。立即確認 付款後立即確認訂單。
- Point: 最多可得 15 回贈金
- Buttons: 確認行程 (Confirm Trip), 選擇日期 (Select Date), 收藏起來之後好找 (Save for later), 加入旅遊計畫 (Add to Travel Plan).

AI Creatrip 소개

GraphQL 스키마 검색 챗봇 토이프로젝트로 시작

- <https://vercel.com/templates/next.js/nextjs-ai-chatbot>

전사 AI 리터러시 확산 + 우리만의 유즈케이스

현재는 매우 다양한 기능 제공

12월: 첫 런칭, GraphQL 스키마 검색, GraphQL 요청 발송 기능

1월: pdf, txt, csv, MS Office 파일 기반 대화 기능, 웹 검색 기능

2월: 개인화 메모리 기능

3월: 프로젝트(Like GPTs) 기능

4월: 회의록 전사-요약, 이미지 생성/수정, 로컬 MCP 연동 기능

5월: Slack, Google, Smithery, Airtable 연동, 딥 리서치, 오디오 대화 기능

6월: 도구 호출에 HITL 추가, Jira, Github 연동

7월: Notion 연동, Data / Github 에이전트

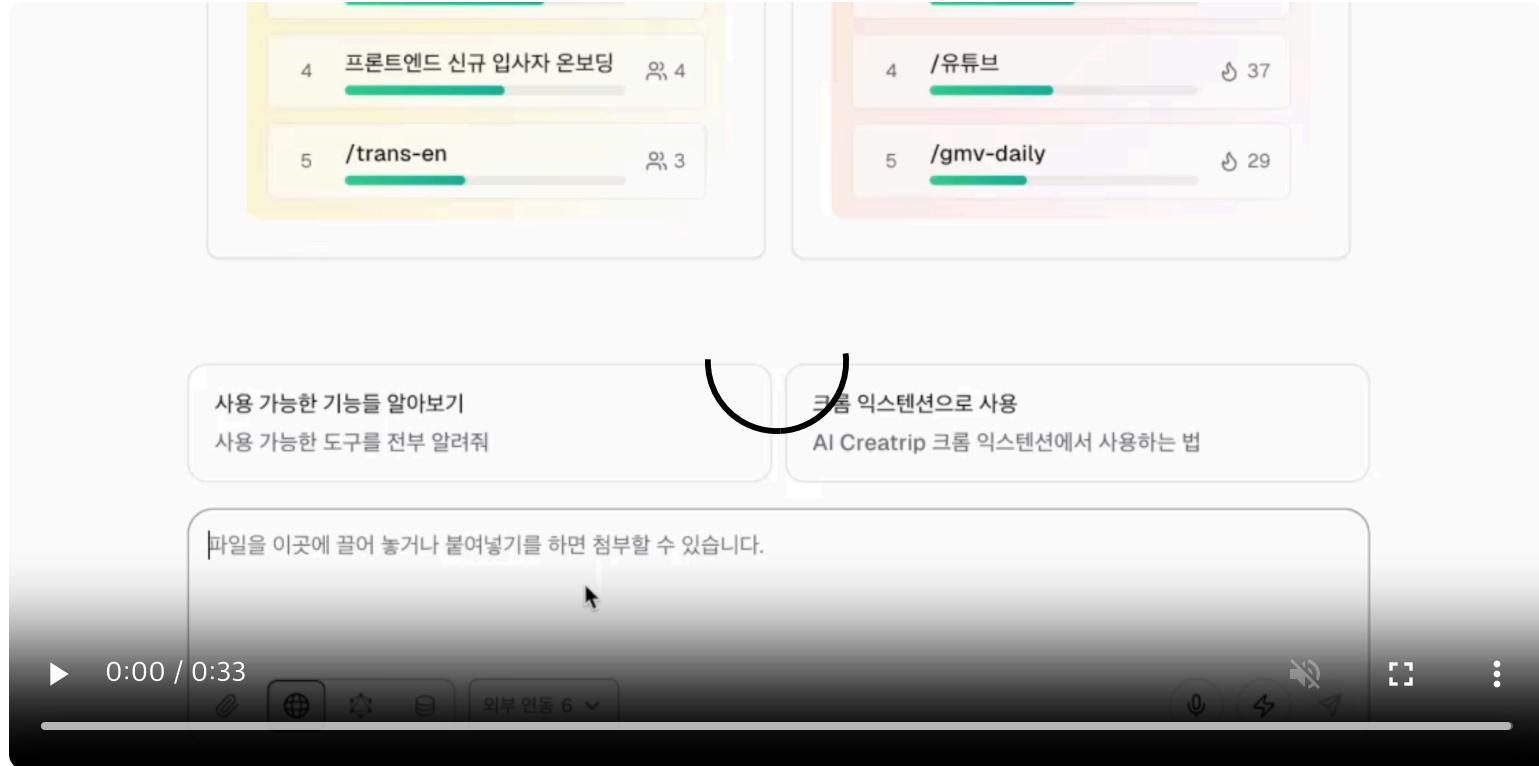
8월: Youtube 조회 기능 추가

9월: 마크다운 기반 텍스트 에디터 추가, 메시지 큐 추가

10월: MCP 엔드포인트 지원, 커스텀 UI DSL 추가

맛보기 1

스킬, 도구, 메시지 큐



맛보기 2

데이터 에이전트

목적: 최근 30일간 여행플래너 신규 생성자 중 '가입 1주 이내' vs '1주 이상' 비율 조회. 입력: 기간=최근 30일, 그룹=계정나이(<=7일, >7일).

데이터 구조 조회 완료

생각중 ...

데이터 에이전트 실행

진행상황

0:00 / 0:40

연동 서비스 + AI Creatrip 전용 도구 활성화

외부 서비스 연동

외부 서비스와 연동하여 더 많은 기능을 사용할 수 있습니다.

슬랙 연동
워크스페이스 메시징 On >

구글 연동
캘린더, 메일, 드라이브 등의 기능을 사용할 수 있습니다. On >

에어테이블 연동
베이스, 테이블, 레코드 관리 등의 기능을 사용할 수 있습니다. On >

Smithery 연동
다양한 MCP 연동 >

Figma 연동
디자인 컨텍스트를 MCP 도구로 불러올 수 있습니다. On >

GitHub 연동
저장소 관리, 이슈 트래킹, 코드 리뷰 등의 기능을 사용할 수 있습니다. On >

Jira 연동
이슈 관리, 프로젝트 트래킹, 워크플로우 관리 등의 기능을 사용할 수 있습니다. On >

파일을 이곳에 끌어 놓거나 붙여넣기를 하면 첨부할 수 있습니다.

- Worked for 3s -

- creatrip 관련 도구들은 다음과 같습니다:
- creatrip__GA_getCustomDimensionsAndMetrics
- creatrip__GAGetPropertyDetails
- creatrip__GA_runAnalyticsReport
- creatrip__airtable_createRecord
- creatrip__airtable_deleteRecords
- creatrip__airtable_getBaseSchema
- creatrip__airtable_getRecord
- creatrip__airtable_listBases
- creatrip__airtable_listRecords
- creatrip__airtable_searchRecords
- creatrip__airtable_updateRecords
- creatrip__chat_getChatDetails
- creatrip__chat_moveChatProject
- creatrip__chat_searchChats
- creatrip__download_figma_images
- creatrip__get_figma_data
- creatrip__google_createCalendarEvent
- creatrip__google_deleteCalendarEvent
- creatrip__google_listCalendarEvents
- creatrip__google_listCalendars
- creatrip__google_listEmails
- creatrip__google_searchEmails
- creatrip__google_updateCalendarEvent
- creatrip__jira_addComment
- creatrip__jira_createIssue
- creatrip__jira_getAllProjects

오늘의 주제: 에이전트의 심장

데이터, 웹, Github 에이전트

구현하며 배운 점

초기 접근의 문제점

BigQuery 연동과 한계

SQL과 데이터 구조에 대한 높은 이해도 요구

회의론

- 개발자: DB에서 직접 쿼리하면 되는데?
- 비개발자: SQL 모르면 못 쓰겠는데?

심지어 당시 SQL 세미나를 통해 여러 팀에서 직접 쿼리 날리는 중

진짜 문제: 컨텍스트 부족

LLM은 똑똑하지만,
답안지(컨텍스트)를 주지 않으면 좋은 답변을 할 수 없다

해결책: 생각의 사슬(CoT)과 Agentic Loop

성능 향상의 확실한 방법으로 제시된 CoT(Chain of Thought)

Vercel AI SDK의 Reasoning Steps에서 영감

Sequential Thinking MCP 경량화 버전과 유사한 구현

steps-reasoning@vercel

[api/chat/route.ts](#)

Agent Loop 첫 구현

1. 전체 데이터 구조 파악 (어떤 테이블들이 있는지, 각 테이블은 뭘 하는지)
2. 문제 해결에 필요한 테이블 선택
3. 테이블 정보(들) 조회 및 쿼리 생성 (Dry Run)
4. 결과 확인 및 루프 종료 결정
5. 완료

새로운 즐거운 문제들

비용/지연: 매번 전체 테이블 목록을 조회하여 비효율적

컨텍스트 초과: 테이블 정보가 너무 많음 (예시 row에서 HTML 컨텐츠 포함)

무한 루프: LLM이 루프를 종료하지 않고 계속 `final`

결과 처리: 방대한 결과 데이터로 인한 컨텍스트 폭발 및 연속 질문 불가

해결 과정 (1): 입력 데이터 최적화

캐싱

- 전체 데이터 구조(테이블 목록)
- 테이블 메타데이터

데이터 정제

- 메타데이터를 '인간 친화적'으로 정리
- 컨텍스트 초과에 대한 안전장치 (출력값이 초과되어서 요기까지밖에 못 보여줌)

해결 과정 (2): 루프 제어와 결과 요약

강제 종료

- nextStep: 'final' 일 경우 루프를 강제로 종료하여 무한 루프 방지

Summary 레이어

- 별도의 LLM이 최종 보고서를 생성하여 메인 LLM의 컨텍스트 오염을 막고 연속 질문을 가능하게 함
- 메인 LLM(유저와 대화하는)에서는 데이터 에이전트의 생각, 호출 쿼리, 최종 보고서라는, 상대적으로 적은 컨텍스트 오염으로 필요 결과를 다 얻을 수 있음

이걸로 행복해졌을까?

NO

가장 중요한건, Agent Loop의 Loop가 닫힌 상태면 한계가 있음

루프의 성능을 극대화하려면 외부의 장/단기 피드백이 반드시 필요

단기적 피드백: Inside Loop

도구 호출 결과: 쿼리의 문법적/도메인적 정합성(Dry Run), 성공한 쿼리 결과 등

LLM의 자체 교정: '결과가 0개네? 내가 뭘 놓쳤지?'

확장된 사고 단계:

- `alternativePlan` (플랜 B)
- `critic` (비판)
- `previousResultSummary` (이전 결과 요약)

Reasoning Step

```
{  
  "reasoningTitle": string,  
  "previousResultSummary": string,  
  "reasoningContent": string,  
  "critic": string,  
  "alternativePlan": string,  
  "tool": {  
    "toolName": string,  
    "params": object  
  } | null,  
  "nextStep": "continue" | "final"  
}
```

장기적 피드백: Real-world context injection

Real-world Rules

- 에이전트가 알 수 없는 것 (데이터 구조에 드러나지 않는 사실)
- 우리 회사의 규칙 (주간 매출액 산정의 Week Date 기준)
- 현재 연월일시분초 인지 - 별거 아니지만 정말정말 중요
 - 모든 LLM 프로덕트의 시스템 프롬프트 첫 줄

현실적인 어려움 (1): 귀찮음

유저

- "이거 이상해요" 라고 개발자에게 말하기 귀찮음.
- '"25년 기준으로 다시 찾아줘" 라고 그냥 내가 다시 말하고 말지...'

개발자

- 잖은 프롬프트 수정 - 잖은 배포
- LLM의 비결정성으로 인한 테스트의 어려움

현실적인 어려움 (2): 프롬프트 관리

모든 비즈니스 규칙을 넣으면 프롬프트가 너무 길어짐

제한된 상황에만 필요한 프롬프트가 오히려 환각(Hallucination)을 유발

- 프롬프트가 상충된다면?

해결책: 동적 프롬프트 주입

컨텍스트 엔지니어링

- 정적인 프롬프트 대신, 상황에 맞는 지침(규칙)들을 동적으로 주입하는 방식
 - 여러 지침/지식들 - 맥락에 필요한 지식을 기억해내듯 주입
- 컨텍스트 엔지니어링: 우리가 제어할 수 있는 유일한 `input` 인 프롬프트를 동적으로 제어

컨텍스트 수집

피드백 저장 모드: 유저가 "이 규칙 기억해"라고 하면, 에이전트가 규칙을 구조화하여 저장

UI 제공: 귀찮음을 줄이기 위해 UI에서 직접 컨텍스트를 조회하고 추가/수정 가능

구조화된 데이터: 연관 테이블, 사용 맥락, 요약, 전체 내용, 라벨 등

에이전트 피드백 시스템

The screenshot shows the 'Agent Feedback System' interface. On the left, there's a main panel with a header '데이터 에이전트 실행' (Data Agent Execution) and a status indicator '실행 상태: 완료' (Execution Status: Completed). Below this are sections for '최종 결론' (Final Conclusion) and '타임라인 보기 (5단계)' (Timeline View (5 steps)). A note section contains the text: '분석 중 알게 된 규칙이나 팁을 적어 주시면 에이전트가 더 정확해져요.' (Please write down any rules or tips you know during analysis so the agent can become more accurate.) and an example: '예: reserve.status=5(부분환불)은 취소가 아니므로 예약 취소건 집계에서 제외하지 않는다. 오직 status=4(취소)만 취소로 간주해 제외한다.' (Example: reserve.status=5 (partial refund) is not a cancellation, so it is not excluded from the count. Only status=4 (cancellation) is considered a cancellation and excluded.) with a '피드백 저장' (Save Feedback) button. A blue arrow points from this note area to a tooltip on the right.

저장된 데이터 컨텍스트
데이터 에이전트가 수집한 규칙과 참고 정보를 확인해 보세요.

Default weekly aggregation uses ISO WEEK 업무 규칙 2025. 10. 08. 오후 06:17
조건: When weekly GMV/metrics requested
Unless users specify a different week definition or date range, aggregate weekly metrics (e.g., GMV) using ISO 8601 ISO WEEK (Mon–Sun). Apply EXTRACT(ISOYEAR/ISO WEEK) in BigQuery for grouping.
GMV forsum_gmv dashboard.SpotSalesNew weekly report Creatrip

Creatrip spot user page URL format 업무 규칙 2025. 10. 08. 오후 04:47
조건: When generating user page links
Use <https://creatrip.com/{locale}/spot/{code}> for spot detail links. If {locale} is omitted, it defaults to the Taiwan locale page.
Creatrip spot user page link

컨텍스트의 동적 활용

매 사고 단계마다 LLM의 생각을 기반으로 필요한 컨텍스트를 동적으로 검색

방법 1: 쿼리 테이블의 이름 등으로 정적 매핑

방법 2: 별도 LLM에게 "이런 생각을 하는데, 어떤 컨텍스트가 필요할까?"라고 물어보기

현재는 방법 2를 사용 중이며, 컨텍스트 양에 따라 RAG 등 다른 방식으로 확장 가능

그래서 성능은? (1)

성능 측정의 어려움

LLM의 성능을 정량적으로 테스트하기는 매우 복잡함

구조 리팩토링, 모델 변경, 프롬프트 수정 등 실험을 계속하며 개선 중

인터넷 제품이므로 내부 구성원의 피드백을 체감하는 것이 가장 큰 척도

그래서 성능은? (2)

체감되는 성능 향상

숙련도 하락

- 데이터를 요청하는 사람(직군)을 덜 타게 됨
- 궁금한 점을 비동기로 물어보고 답변 확인하는 경험이 매우 편리

컨텍스트 효율성: 수십 번의 에이전트 호출에도 메인 컨텍스트가 안정적으로 유지

결론

Agent Loop 개발은 단순히 기능을 만드는 것 이상의 많은 인사이트를 줌

이 경험을 바탕으로 사내 AI Cell에서 다양한 시도 중

Agent Loop와 피드백 구조를 통해 Github PR 리뷰 액션도 크게 개선

- 입사 2개월차 신규 입사자분이 주도!

저희 이런거 만들어요!

여행 정보 MCP: 전세계 여행객에게 전문적인 한국 여행 정보 제공

CX Agent: CX팀이 단순 문의를 넘어 한국 여행 컨시어지가 되도록 지원

그 밖에도... 한국에 대해 잘 알기 때문에 제공해 줄 수 있는 독점적인 경험(with AI)

사람을 찾습니다 (Node.js)



감사합니다

Q&A

궁금하신 점을 자유롭게 질문해 주세요!



서종학 (jonghak.seo@creatrip.com)