

기업 AX 워크프로세스 설계

취소 직후 추천, 내보내도 되는가

예약을 취소한 사용자에게 대체 상품을 다시 보여주는 정책을, 배포하기 전에 검수하는
Codex 플러그인입니다. 추천 엔진이 아닙니다.

myrealtrip-cancel-recovery-eval

3종

후보별 판정 — 내보냄 · 사람 확인 · 막음

6

리스크 규칙

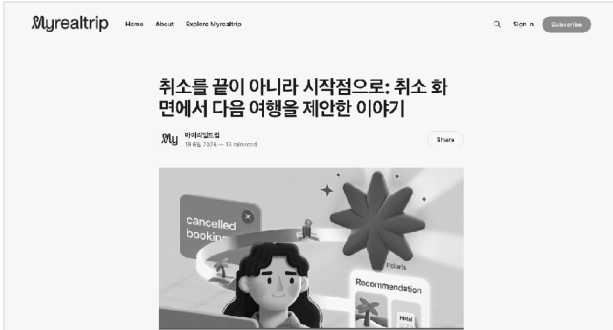
5 / 5

단위 테스트 통과

hold

실제 상품 데이터를 넣었을 때의 정책
판정

마이리얼트립이 스스로 공개한 기술 회고와 API 문서만 썼습니다. 내부 고객 데이터·결제 정보·매출 로그는 쓰지 않았습니다. 각 자료에서 정확히 무엇을 가져왔는지를 아래에 적었습니다.

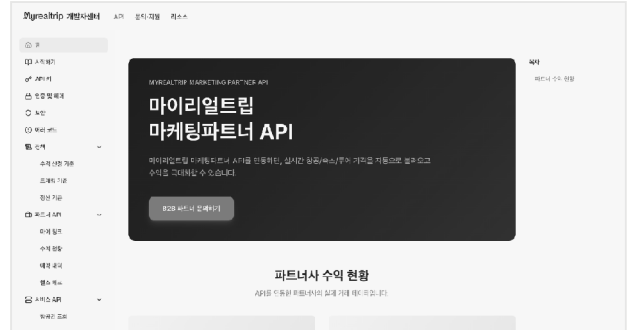


기술블로그 · 마이리얼트립

취소를 끝이 아니라 시작점으로 — 취소 화면에서 다음 여행을 제안한 이야기

문제 정의의 출발점. 취소 완료 화면의 다음 여행 제안과 7일 재예약 신호를 근거로 삼되, 회고가 스스로 관찰 결과임을 밝힌 점을 인과 주장 제한의 근거로 사용.

blog.myrealtrip.com

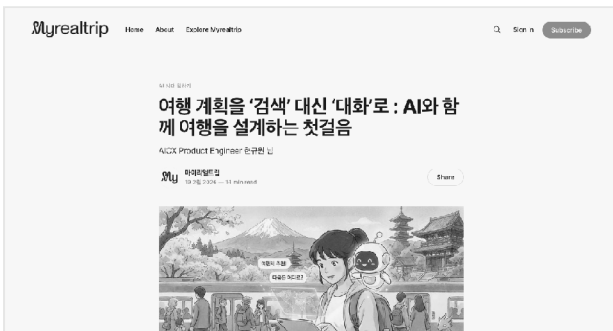


공개 API 문서

마이리얼트립 API 사용 가이드

투어·액티비티 상품의 검색·상세·옵션·예약가능 확인이 각각 별도 단계임을 파악. 후보 상품이 갖춰야 할 필수 항목(가격·옵션 확정 여부·취소정책)를 여기서 도출.

docs.myrealtrip.com



기술블로그 · 마이리얼트립

여행 계획을 '검색' 대신 '대화'로: AI와 함께 여행을 설계하는 첫걸음

AI 여행 탐색이 실제 재고와 연결되어야 한다는 요구를 확인. 예약 가능성·옵션 가격 검증을 판정 규칙에 넣은 근거.

blog.myrealtrip.com



기술블로그 · 마이리얼트립

AI가 평균을 높일수록, '나다운 여행'은 귀해진다

개인 취향과 최저가 사이의 긴장을 확인. 세그먼트별로 노출 조건을 다르게 두는 정책 설계의 배경.

blog.myrealtrip.com

자료를 읽고 네 개의 가설을 세운 뒤, 하나씩 공개 자료로 확인했습니다. 확인되지 않은 가설이 하나 남았고, 그 지점에서 문제 정의로 되돌아갔습니다.

■ 확인 — 공개 자료로 뒷받침됨 ■ 근거 부족 — 공개 자료만으로는 확인 불가 □ 폐기 — 가설 자체를 버림

가설	내용	어디서 확인했나	결과
H1	취소 직후에도 여행 의도는 남아 있어, 대체 상품 노출이 회복 기회가 된다	기술블로그가 취소 완료 화면의 다음 여행 제안과 7일 재예약 신호를 공개	■ 확인
H2	그 신호를 인과라고 부르면 잘못된 정책이 근거 없이 확대된다	같은 회고가 스스로 “관찰된 신호이지 인과가 아니다”라고 밝힘	■ 확인 · 강화
H3	후보 상품의 예약 가능성 근거가 비면, 추천 후 재실패로 CS·환불 리스크가 커진다	공개 API 문서에서 옵션·예약가능·취소정책이 각각 별도 확인 단계임을 확인	■ 확인
H4	맥락을 잘못 읽은 노출(날씨 취소 뒤 다시 야외 상품)이 실제로 신뢰를 깎는다	공개 자료에 실패 사례가 없음. 내부 로그와 A/B 배정 없이는 확인 불가	■ 근거 부족
H4'	보충 — 근거가 빈 후보는 mock이 아니라 실데이터에서도 실제로 나온다	마이리얼트립이 공개한 상품 조회 창구(MCP)에서 가져온 나트랑 투어·액티비티 상품 3건 중 2건이 날씨 리스크 반복 + 스냅샷 84.3시간 경과, 1건은 취소정책 미상 → 정책 전체 hold	■ 재검증 통과

되돌리기 · 문제 정의로 회귀

증명할 수 없는 효과를 주장하는 대신, 검수할 수 있는 리스크로 문제를 바꿨다

H4를 공개 자료만으로 확인할 수 없다는 것이 분기점이었습니다. “추천을 더 잘하자”는 방향은 내부 로그와 A/B 배정 없이는 어떤 주장도 증명할 수 없습니다. 그래서 문제를 한 단계 앞으로 옮겼습니다. 증명할 수 없는 효과를 여섯 개의 검수 가능한 리스크 규칙으로 바꾸고, 그 규칙이 실데이터에서도 걸리는지를 다시 확인했습니다.

추천 엔진 → 증명 불가 → 문제 정의로 회귀 → 노출 정책 검수 → 리스크 규칙 6개 → 실데이터 재검증

최종 문제 정의 — 대체 상품 노출 정책을 배포하기 전에, 사용자 맥락 · 상품 근거 · 노출 조건 · 성과 해석 가능성을 검수한다.

이 플러그인의 사용자는 여행자가 아닙니다. 취소 직후 대체 상품 노출 정책을 운영하는 담당자입니다. 사용자를 여기까지 좁힌 것이 이 프로젝트에서 가장 먼저 한 결정이었습니다.

상황 1

정책을 켜기 직전

어떤 후보를 어떤 조건으로 노출할지 정해두고, 배포 버튼을 누르기 전 구간.

상황 2

이미 돌린 결과를 해석할 때

노출 후 재예약률이 올랐을 때, 그것을 인과라고 불러도 되는지 판단해야 하는 구간.

상황 3

실데이터가 들어왔을 때

mock이 아니라 실제 상품 후보가 들어와도 같은 기준으로 막히는지 확인해야 하는 구간.

이 도구를 쓰는 사람

그로스

대체 상품 노출 정책의 조건과 세그먼트를 정하는 사람. 무엇을 언제 누구에게 보여줄지의 판단을 소유합니다.

프로덕트

취소 완료 화면의 경험과 노출 규칙을 소유한 사람.

운영

잘못된 추천이 만든 CS 문의와 환불을 실제로 받는 사람.

플러그인은 무엇으로 이루어져 있나

쓰는 사람

플러그인 — 이 프로젝트가 만든 것

드나드는 데이터

그로스 · 프로덕트 · 운영 담당자

노출 정책을 켜기 전에 검수를 요청합니다.

Codex 앱 대화창

따로 만든 화면이 없습니다. 담당자가 보는 화면은 Codex 대화창 그대로입니다.

플러그인 진입점

src/.codex-plugin/plugin.json
Codex가 이 폴더를 플러그인으로 알아보게 하는 파일.

사용 설명서 (스킬)

skills/cancel-recovery-eval/SKILL.md
언제 이 도구를 써야 하는지, 어떤 주장을 하면 안 되는지를 적어 둔 문서. LLM은 여기까지만 읽습니다.

판정 스크립트

scripts/evaluate_policy.py
실제 판정을 내리는 곳. LLM이 아니라 이 파일이 결론을 냅니다.

리스크 규칙 · 입력 형식 정의

policy/risk_rules.yaml · schemas/
무엇을 막을지와 어떤 모양의 데이터를 받을지를 코드 밖에 빼 둔 파일.

들어오는 것

취소 이벤트 · 후보 상품 목록 · 노출 정책. 가상 데이터이거나, 마이리얼트립이 공개한 상품 조회 창구에서 그때그때 받아온 실제 상품.

나가는 것

후보별 판정과 이유, 정책 전체 판정이 담긴 리포트 두 벌 — 사람이 읽는 문서와 기계가 읽는 데이터.

웹 서비스가 아닙니다. 화면은 Codex 대화창, 실행은 로컬 파이썬 스크립트이고 데이터베이스는 없습니다.

클릭했는지, 7일 안에 재예약했는지는 추천하는 순간에는 알 수 없는 정보입니다. 이걸 판단에 섞으면 실험실에서만 좋아 보이는 평가가 됩니다. 위아래를 가르는 실선이 이 도구의 핵심 설계입니다.

판정 시점 - 추천하는 그 순간에 알 수 있는 것만

취소 이벤트

후보 상품 목록

노출 정책

6개 리스크 규칙

내보내도 됨

사람이 봐야 함

막아야 함

후보별 판정을 모아 정책 전체를 진행(ship) · 수정(revise) · 보류(hold) 중 하나로 정리합니다.

노출 판단에 쓰는 것은 여기까지

노출 이후 - 사후 해석 전용, 판정에 섞지 않음

노출 후 성과 기록

성과 해석 가능성 검토

관찰된 신호 ≠ 인과 효과

클릭 여부 · 7일 내 재예약 · 노출 후 거래액은 판정 입력에서 아예 뺍니다 (사후 정보가 판단에 새어 들어오는 것을 차단).

무엇을 잡아내는 규칙인가

심각도

이 후보를

예약 가능 여부-옵션이 확정되지 않음

missing_confirmed_availability

높음

막는다

날씨 때문에 취소했는데 또 같은 날씨 리스크

same_weather_risk_after_weather_cancel

높음

막는다

판정 입력에 노출 이후 데이터가 섞임

outcome_leakage

높음

정책 전체 보류

가격에 민감한 사용자에게 과도한 가격 상승

price_spike_for_price_sensitive_user

중간

사람이 본다

후보 상품의 취소 정책을 알 수 없음

unknown_cancellation_policy

중간

사람이 본다

사용자가 왜 취소했는지 알 수 없음

unknown_cancel_reason

중간

사람이 본다

설계하면서 내린 판단

판정은 LLM이 아니라 스크립트

판정 스크립트가 정해진 형식의 데이터를 읽고 명시적 규칙으로 수행합니다. 같은 입력이면 같은 결과가 나와야 하고, 왜 막았는지 설명할 수 있어야 하기 때문입니다.

근거가 없으면 통과가 아니다

취소 정책을 알 수 없거나 옵션 가격이 확정되지 않았으면, 자동 통과시키지 않고 사람 검토로 되돌립니다.

스냅샷 신선도까지 본다

상품 정보가 판정 시점에서 몇 시간 떨어져 있는지도 리스크로 기록합니다. 실데이터 3건 중 2건이 여기서 걸렸습니다.

LLM이 맡은 부분

언제 이 도구를 쓸지 판단, 비정형 입력의 정규화, 리포트 서술, 그리고 과장된 성과 주장 차단.

평가 항목	결과	의미
단위 테스트	5 / 5	핵심 리스크 규칙 회귀 검증
날씨 취소 · 나트랑 (mock)	ship	실내 머드스파 expose / 수상 호핑투어 · 프리미엄 요트 · 문화쇼 suppress
가격 변동 · 오사카 (mock)	revise	저가 패스 expose / 기존 결제액 2.00배 후보 needs_review
공개 상품 조회 창구 실데이터 · 나트랑 투어·액티비티 상품 3건	hold	2건 suppress — 날씨 리스크 반복 + 스냅샷 84.3시간 경과, 1건 needs_review
Codex 플러그인 인식	확인	로컬 Codex 앱에서 skill이 실제 응답에 사용됨

기대할 수 있는 것 — 검증된 성과가 아니라 이 도구가 노리는 지점

재취소 · CS 리스크 감소

취소 사유를 반복하는 후보를 배포 전에 거른다.

2차 실패 방지

재고·옵션 근거가 빈 후보를 되돌려, 추천 후 다시 실패하는 상황을 막는다.

운영 가능한 평가

사후 데이터 누수를 차단해 실험실에서만 좋아 보이는 평가를 방지한다.

정책 확대 통제

관찰 신호와 인과 효과를 구분해 근거 없는 확대를 막는다.

주장하지 않은 것

매출·재예약을 상승을 증명하지 않았다

내부 로그와 A/B 배정 없이는 말할 수 없습니다.

인과라고 부르지 않았다

무작위 배정·균형 검사·가드레일 지표가 없으면 observational_signal 로 표기합니다.

내부 데이터를 쓰지 않았다

고객 데이터·결제 정보·내부 로그·API 키 미사용. 후보 상품은 공개 조회 창구에서 받아온 것과 가상 데이터뿐입니다.

운영 배포를 하지 않았다

검증한 것은 로컬 Codex 앱에서 플러그인이 인식되고 skill이 응답에 쓰인다는 것까지.

설계하면서 느낀 점

제일 먼저 한 일은 만들 것을 정하는 게 아니라, 사용자를 여행자에서 정책 담당자로 좁히는 일이었습니다. 그 다음에 알게 된 건 공개 자료로 증명할 수 없는 효과를 주장하려 들면 프로젝트 전체가 흔들린다는 것이었습니다. 증명할 수 없는 것을 빼고 나니 남은 게 검수 규칙이었고, 그게 결국 이 도구의 정체가 됐습니다.