

기업 AX 워크프로세스 설계

한쪽 데이터만으로는 왜 안 크는지 못 말한다

브랜드가 가진 맥락과 무신사가 가진 플랫폼 신호를 같은 형식으로 맞춘 뒤에야 성장
병목을 말하는 Codex 플러그인입니다.

musinsa-partner-growth

8 / 8

합성 fixture 병목 정확도

12 / 12

채점 기준 적용 — 사전 확정 임계값 10 / 12

7

병목 유형 · 판정 순서 고정

pass

하드코딩 경계 검사

무신사 채용 페이지와 뉴스룸만 썼습니다. 내부 지표는 쓰지 않았고, 공개 자료에 없는 수치는 문서 어디에도 넣지 않았습니다. 이 여섯 건이 문제를 좁히는 데 실제로 쓰인 전부입니다.

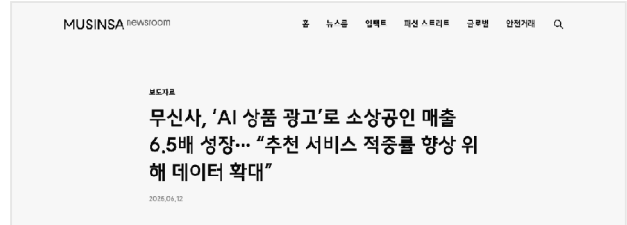


채용 · 조직 소개

무신사 파트너그로스비즈 — 브랜드의 성장을 설계합니다

가장 많이 참조한 자료. 브랜드 파트너십의 시작부터 성과까지 데이터로 성장 솔루션을 제공하는 조직이 실재한다는 것 — 즉 이 문제가 실제 업무 표면이라는 근거.

musinsacareers.com



뉴스룸 · 2025.06

무신사, 'AI 상품 광고'로 소상공인 매출 6.5배 성장

AI 광고가 넓게 도입됐지만 참여하지 않는 브랜드가 여전히 많다는 점. '광고 준비 병목' 유형을 규칙에 넣은 이유.

newsroom.musinsa.com

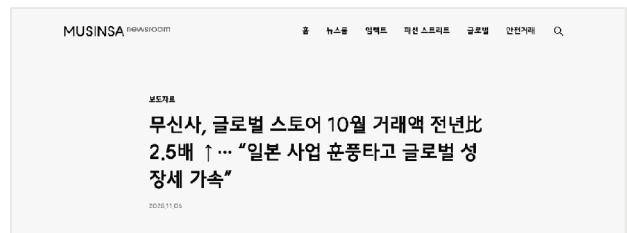


뉴스룸 · 2026.06

무신사, 'AI 트렌드 큐레이션' 도입

스타일·속성 태깅이 플랫폼 쪽 자신임을 확인. '포지셔닝 불일치'를 브랜드 선언 스타일과 플랫폼 반응 클러스터의 차이로 정의한 근거.

newsroom.musinsa.com

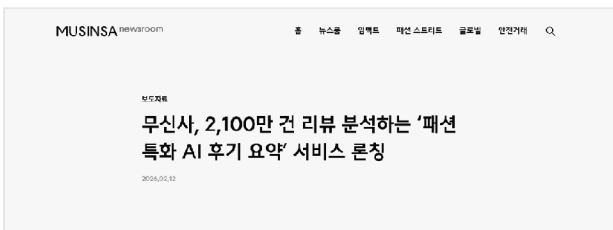


뉴스룸 · 2025.11

무신사, 글로벌 스토어 10월 거래액 전년비 2.5배 증가

글로벌 확장이 실제 성장 축임을 확인. '오프라인·글로벌 준비 병목' 유형의 근거.

newsroom.musinsa.com

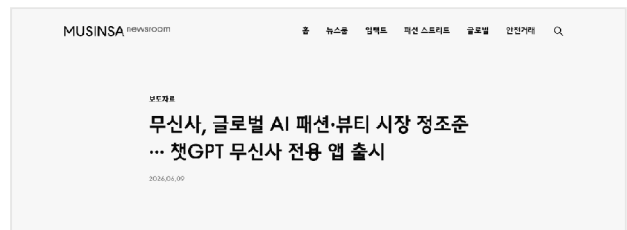


뉴스룸 · 2026.02

무신사, 2,100만 건 리뷰 분석하는 '패션 특화 AI 후기 요약' 론칭

첫 후보였던 '리뷰 요약 품질 검증' 방향의 근거였고, 동시에 그 방향을 버린 이유. 이미 출시된 기능을 흉내내지 않기 위해 파트너 그로스 쪽으로 방향을 바꿈.

newsroom.musinsa.com



뉴스룸 · 2026.06

무신사, 챗GPT 무신사 전용 앱 출시

이미 에이전트 인터페이스 쪽으로 움직이고 있음을 확인. 새 앱을 만드는 대신 플러그인 레이아웃을 택한 판단의 배경.

newsroom.musinsa.com

처음 세운 가설은 검증 과정에서 스스로 무너졌습니다. 무너진 이유가 “틀렸다”가 아니라 “이미 있다”였기 때문에, 가설을 고치는 대신 후보 전체를 다시 세워야 했습니다.

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

가설	내용	어디서 확인했나	결과
H1	AI 후기 요약의 품질을 검수하는 도구가 필요하다	뉴스룸에서 2,100만 건 리뷰 기반 AI 후기 요약이 이미 베타로 출시됐음을 확인	□ 폐기
H2	성장 진단에 필요한 데이터가 브랜드와 플랫폼 두 조직에 나뉘어 있다	파트너그로스비즈 조직이 실재하고, 브랜드에 데이터 기반 성장 솔루션을 제공한다고 공개	■ 확인
H3	신규·중소 브랜드는 그 진단을 상대적으로 덜 받는다	무신사가 제한된 리소스 탓에 맞춤 지원이 일부 대형·핵심 파트너에 집중돼 왔다고 공개적으로 언급	■ 확인
H4	‘광고를 할 준비가 안 된 상태’가 병목의 한 유형이다	뉴스룸 — AI 상품 광고로 소상공인 매출 6.5배, 그러나 미참여 브랜드가 여전히 많음	■ 확인
H5	브랜드가 선언한 스타일과 플랫폼이 관측한 반응이 어긋난다	뉴스룸 — AI 트렌드 큐레이션. 스타일·속성 태깅이 플랫폼 쪽 자산임을 확인	■ 확인
H6	오프라인·글로벌 확장도 별도의 준비 병목이다	뉴스룸 — 글로벌 스토어 10월 거래액 전년比 2.5배 증가	■ 확인

되돌리기 · 후보 전면 재랭킹

이미 출시된 기능을 흉내 내는 방향을 버리고, 다섯 개 후보를 다시 줄 세웠다

첫 가설은 후기 요약 품질 검수였습니다. 그런데 뉴스룸을 확인하는 과정에서 무신사가 이미 그 기능을 출시했다는 것을 알게 됐습니다. 남의 기능을 다시 만드는 제출물이 되는 순간이었습니다. 그래서 후보를 다섯 개로 다시 세우고(후기요약 QA · 글로벌 로컬라이제이션 QA · 광고 온보딩 · 트렌드·속성 매핑 QA · 온·오프라인 연계 사이즈 데이터), 공개 근거의 두께와 사설 데이터 없이 검증 가능한지로 다시 정렬했습니다. 살아남은 것이 파트너 그로스였습니다.

후기 요약 QA → 이미 출시됨 → 후보 5개 재랭킹 → 공개 근거 두께로 정렬 → 파트너 그로스 → 테스트 8건으로 재검증

최종 문제 정의 — “무신사가 브랜드를 알아서 진단해 준다”가 아니라, 양쪽이 최소한의 맥락을 같은 형식으로 교환하게 만든다.

브랜드가 가진 것과 플랫폼이 가진 것은 서로 보이지 않습니다. “왜 성장하지 못하는가”는 어느 쪽도 혼자서는 답할 수 없는 영역입니다.

브랜드 쪽에만

제약을 아는 데이터

판매 · 재고 · 생산능력 · 마진 · 할인 여력 · 예산
· 스타일 · 타겟 · 목표. 플랫폼은 볼 수 없습니다.

플랫폼 쪽에만

반응을 아는 데이터

노출 · 검색추천 노출 · 클릭 · 장바구니 · 전환 ·
찜 · 리뷰 · 카테고리 벤치마크. 브랜드는 볼 수 없습니다.

그래서

생기는 일

한쪽 데이터로 내린 진단은 반대쪽 제약을 모릅니다. 공급이 막혀 있는데 노출을 늘리라고 말하게 됩니다.

이 도구를 쓰는 사람 — 한 명이 아니라 두 조직

브랜드 담당자

신규·중소 입점 브랜드 실무자. 브랜드 정체성, 상품, 재고, 생산능력, 마진, 목표를 넣어 브랜드 데이터 묶음을 만듭니다.

무신사 담당자

노출, 퍼널, 광고, 리뷰, 카테고리 벤치마크를 넣어 플랫폼 데이터 묶음을 만듭니다.

공동 검토자

결합 진단 결과를 보고 다음 액션과 추가로 필요한 데이터를 확인합니다.

플러그인은 무엇으로 이루어져 있나

쓰는 사람

브랜드 담당자

자기 쪽 데이터를 넣습니다.

무신사 담당자

플랫폼 쪽 데이터를 넣습니다.

Codex 앱 대화창

두 사람이 각자 자기 자리에서 같은 도구를 씁니다.

플러그인 — 이 프로젝트가 만든 것

플러그인 진입점

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

사용 설명서 (스킬)

skills/partner-growth-diagnosis/SKILL.md
세 명령을 언제 어떤 순서로 쓰는지 적어 둔 문서.

진단 스크립트

scripts/partner_growth.py
데이터 정리 두 개와 진단 하나, 세 명령이 모두 여기 있습니다.

병목 규칙 · 데이터 형식

rules/ · schemas/
임계값과 데이터 모양을 코드 밖에 빼 뒀, 조직이 자기 숫자로 갈아 끼울 수 있게 했습니다.

드나드는 데이터

들어오는 것

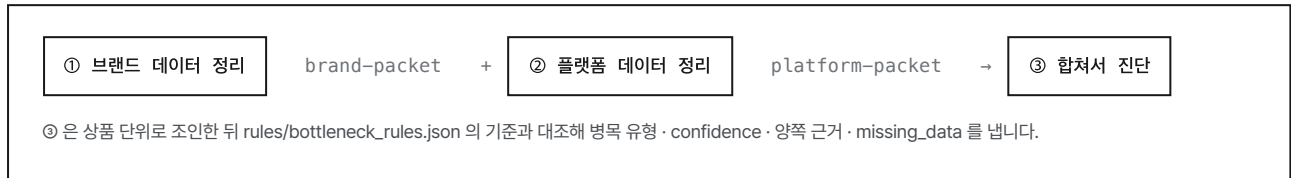
브랜드 쪽 — 판매·재고·생산능력·마진
·목표. 무신사 쪽 — 노출·퍼널·광고·
리뷰·벤치마크.

나가는 것

병목 유형과 확신도, 어느 필드가
근거인지, 아직 없는 데이터 목록,
다음 액션이 담긴 진단서.

두 조직이 각자 자기 데이터만 넣고, 합치는 일은 도구가 합니다. 한쪽만 들어오면 진단 대신 질문이 나옵니다.

브랜드가 자기 맥락을 스키마로 포장하고, 무신사가 플랫폼 데이터를 붙인 뒤에야 병목을 말합니다. 필수 필드가 없으면 강한 결론 대신 “무엇이 더 필요한지”를 되돌립니다.



확신도는 모델의 자신감이 아니라 데이터가 얼마나 갖춰졌는지를 뜻한다

순서	병목	성립 조건	하지 말아야 할 말
1	공급	전환·찜은 강한데 재고·생산능력이 낮음	재고를 풀기 전에 노출을 늘리라고 하지 말 것
2	전환	조회는 충분한데 구매 전환이 벤치마크 미만	페이지 제약을 보기 전에 노출을 더 주라고 하지 말 것
3	클릭	노출은 충분한데 클릭률이 벤치마크 미만	노출이 이미 충분한데 수요가 없다고 말하지 말 것
4	노출	재고는 준비됐는데 노출·검색추천이 벤치마크 미만	노출이 적다는 이유만으로 제품 실패라고 말하지 말 것
5	광고 준비	유료 성장을 원하지만 마진·할인여력·예산이 부족	브랜드 준비 부족을 플랫폼 문제로 돌리지 말 것
6	포지셔닝	선언한 스타일과 플랫폼 반응 클러스터가 다름	신호 하나로 포지셔닝을 다시 쓰지 말 것
7	확장 준비	확장 목표는 있는데 재고 깊이·콘텐츠·운영이 부족	수요 신호만으로 확장 가능하다고 말하지 말 것

설계하면서 내린 판단

판정 순서를 고정했다

과적합을 피하려고 유형을 작게 유지하고 순서를 박았습니다. 공급이 막혀 있으면 노출을 늘리라고 말하지 않기 위해서입니다.

금지 목록을 같이 출력한다

모든 진단에 ‘하지 말아야 할 말’ 목록을 붙여, 그럴듯하지만 데이터가 지지하지 않는 결론을 명시적으로 막습니다.

confidence 의 정의를 바꿨다

모델 자신감이 아니라 데이터 완전성입니다. 양쪽 필드가 다 있고 신호가 일치하면 high, 한쪽이 부분적이면 medium, 한쪽이 없거나 신호가 충돌하면 low.

하드코딩 경계 자체를 테스트했다

관측값·브랜드명·근거를 코드에 심지 않았는지를 평가 항목에 넣었습니다. 합성 데이터는 가상 데이터 폴더에, 임계값은 규칙 파일에만 둡니다.

평가 항목	결과	의미
합성 테스트 케이스 8건 병목 정확도	8 / 8	7개 병목 유형 + “진단 불가” 1건
채점 기준 적용	12 / 12	정확도 · 근거 · 양방향 추론 · 누락 규율 · confidence 보정 · 실행가능성, 8건 전부 만점
사전 확정 통과 임계값	10 / 12	구현을 시작하기 전에 확정
하드코딩 경계 검사	pass	관측값은 입력 데이터에서만, 임계값은 규칙 파일에서만
8번 케이스 — 정보 부족	low	병목 후보 0건 + 누락 필드 목록 반환이 정답

기대할 수 있는 것 — 검증된 성과가 아니라 이 도구가 노리는 지점

진단 경로의 평등

신규·중소 브랜드도 구조화된 진단을 받을 수 있다.

합의 가능한 근거

담당자 경험 편차 대신 필드 단위 근거로 두 조직이 같은 화면을 본다.

데이터 요청 루프

누락 데이터를 질문으로 되돌려 ‘요청 → 재진단’ 흐름이 생긴다.

조직 맞춤 교체

임계값이 외부화되어 실제 벤치마크로 갈아 끼울 수 있다.

주장하지 않은 것

남의 기능을 깎아내리지 않았다

무신사의 기존 AI 기능이 부실하다거나 내부 QA가 없다고 주장하지 않았습니다.

내부 지표를 안다고 하지 않았다

모든 관측값은 직접 만든 가상 데이터이며, 공개 자료에 없는 수치는 쓰지 않았습니다.

브랜드를 대신 판단하지 않는다

한쪽 데이터만 있으면 진단이 아니라 증상 감지와 질문까지만 냅니다.

운영 배포를 하지 않았다

로컬 Codex 앱에서 플러그인 인식과 skill 동작까지 검증한 상태입니다.

설계하면서 느낀 점

첫 방향을 버린 게 이 프로젝트에서 제일 잘한 판단이었습니다. 뉴스룸을 끝까지 읽지 않았다면 이미 있는 기능을 다시 만든 제출물이 됐을 겁니다. 그리고 테스트 케이스를 짜면서 알게 된 건, 답을 내는 케이스보다 답을 못 내는 게 정답인 케이스를 넣는 쪽이 훨씬 어렵고 훨씬 중요하다는 것이었습니다.