랩 위드의 AI 기반 솔루션 >

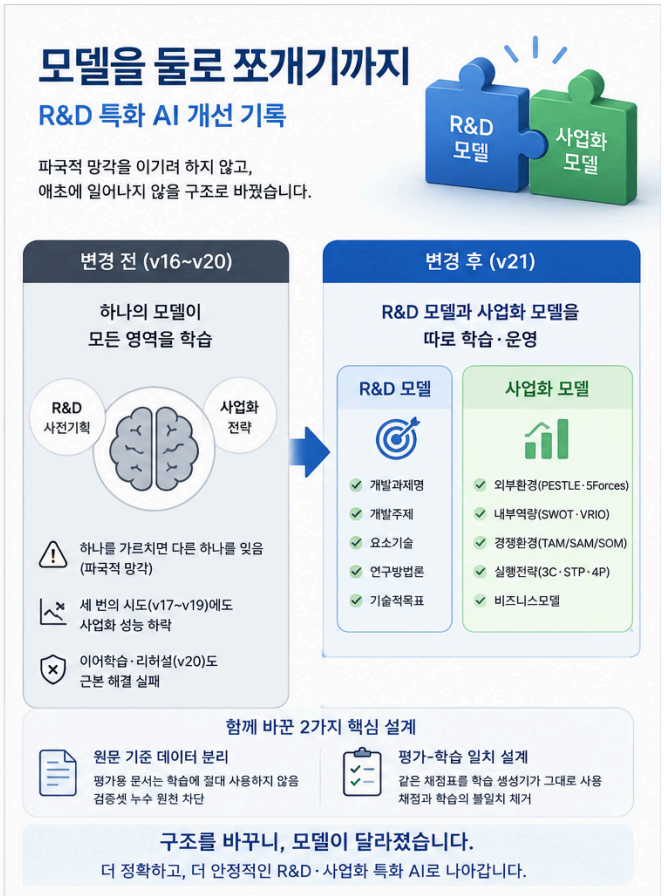
랩 위드가 직접 만드는 R&D 특화 AI, 왜 네 번이나 다시 만들었을까?

기술사업화 랩 위드
2026.09.03. 16:29 조회 4

댓글 0 URL 복사

랩 위드가 직접 만드는 R&D 특화 AI, 왜 네 번이나 다시 만들었을까?





AI가 똑똑해지는 것과 우리 회사 일을 잘하는 AI를 만드는 것은 전혀 다른 문제였습니다.

기술사업화 랩 위드는 R&D 사전기획과 사업화 전략 업무를 지원하기 위한 **자체 sLLM(특정 업무에 특화한 소형 언어모델)**을 개발하고 있습니다.

쉽게 말하면,

일반적인 AI를 가져다 쓰는 것이 아니라,
R&D 기획과 기술사업화라는 특정 업무를 잘하도록 직접 가르치는 AI입니다.

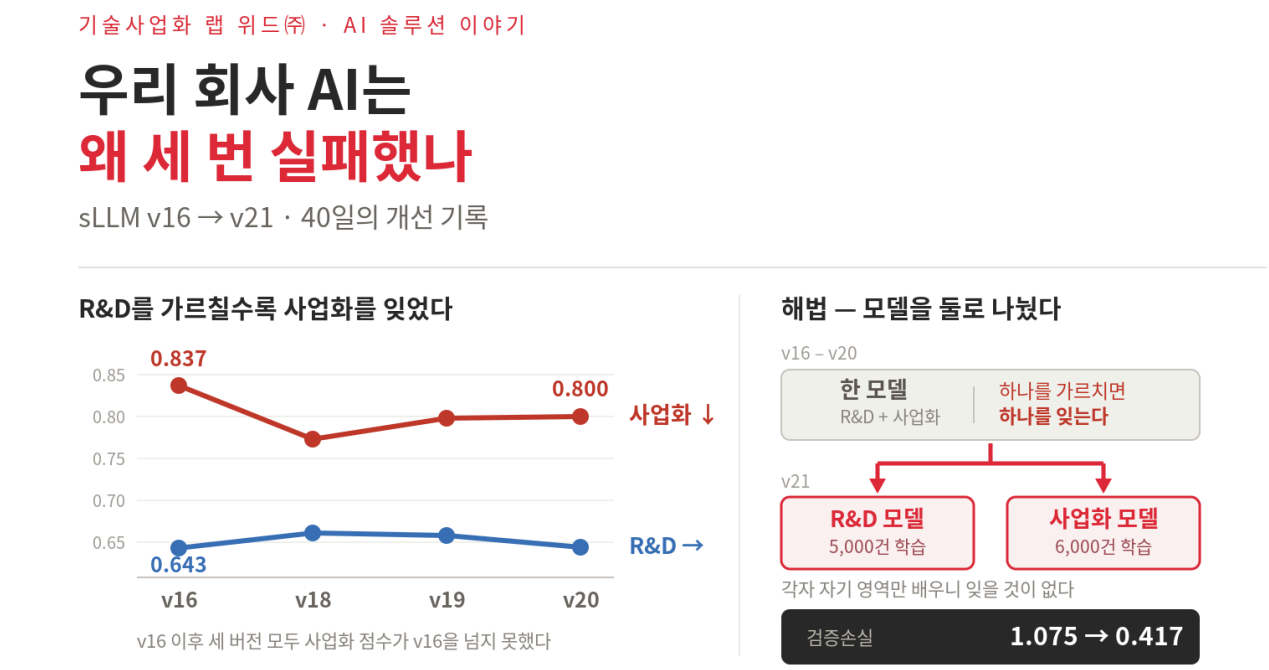
현재 저희가 사용하는 기반 모델은 KT의 Mi:dm 2.0이며, 여기에 그동안 축적해 온 사업계획서와 보고서 약 2,000여 건을 활용해 R&D와 사업화 업무에 필요한 내용을 학습시키고 있습니다.

그런데 생각보다 쉽지 않았습니다.

지난 40여 일 동안 네 번의 버전을 만들었고, 여러 번 이전 버전으로 되돌아가야 했습니다.

왜 그랬을까요?

1. AI에게 R&D를 가르쳤더니 사업화를 잊어버렸습니다



처음에는 아주 단순하게 생각했습니다.
“R&D 계획을 잘 작성하도록 데이터를 더 많이 학습시키면 되겠지.”
그래서 R&D 계획서에 중요한 **요소기술** 데이터를 보강했습니다.
그 결과 요소기술 관련 성능은 크게 좋아졌습니다.

그런데 이상한 일이 발생했습니다.
사업화 전략 성능이 오히려 크게 떨어진 것입니다.

v16에서는 사업화 정합성 점수가 87점 수준이었는데, R&D 관련 데이터를 보강한 v17에서는 63점 수준으로 떨어졌습니다.

즉, **하나를 잘 가르쳤더니 다른 것을 잊어버린 것입니다.**

AI가 사람처럼 “복습하면서 모든 것을 기억하는 것”은 아니었습니다.

새로운 것을 학습하는 과정에서 기존에 학습했던 내용이 약해지는 문제가 발생했습니다.

AI 분야에서는 이런 현상을 **파국적 망각(Catastrophic Forgetting)**이라고 합니다. 단어가 무시무시합니다.

2. 데이터를 더 넣으면 해결될 줄 알았습니다

그렇다면 방법은 간단해 보였습니다.
“기존에 잘하던 내용을 다시 가르치자.”

그래서 다음 버전에서는 답변 형식도 정리하고, 기존에 잘하던 내용을 다시 학습시키고, 사업화 관련 데이터도 늘렸습니다.

하지만 결과는 기대와 달랐습니다.

사업화 관련 학습 데이터를 **808건에서 1,200건으로 약 50% 늘렸는데도** 점수는 오히려 63점에서 60점으로 떨어졌습니다.

여기서 중요한 질문이 생겼습니다.

“정말 AI가 문제인 것이 맞을까?”

3. 그래서 AI가 아니라 '시험지'부터 의심했습니다

AI를 평가하려면 당연히 시험을 봐야 합니다.
그런데 우리가 사용한 시험지를 다시 살펴보니 문제가 있었습니다.

예를 들어 AI에게는 **“TRL은 언급하지 않는다”**고 가르쳐 놓고, 평가할 때는 TRL이 포함되어 있는지를 점수에 반영하고 있었습니다.

쉽게 말하면, **시험에서 요구하지 말라고 가르친 내용을 시험에서는 요구하고 있었던 것입니다.**

이것은 AI의 문제가 아니라 **평가 기준의 문제**였습니다.

그래서 평가표 자체를 다시 만들었습니다.

기존에는 한 업무에서 필수 항목이 2개 정도였다면, 이를 세분화하고 R&D와 사업화의 실제 업무 비중을 고려해 평가 기준도 다시 구성했습니다.

그리고 새로운 문제를 발견했습니다.

4. 알고 보니 AI가 '배우지 않은 것'이 있었습니다

여러 버전을 비교해 보니 세 모델이 공통적으로 못하는 항목들이 있었습니다.

예를 들어,

개발주제
기술동향
시장수요
기술개요

개발과제명
성과지표
적용분야
정량목표
등이 반복적으로 빠졌습니다.

처음에는 “AI가 이 부분을 못하는구나”라고 생각했습니다.
하지만 데이터를 다시 확인해 보니 답은 의외로 간단했습니다.

학습 데이터에 그 내용이 충분히 들어 있지 않았던 것입니다.

AI도 결국 배운 만큼 할 수 있습니다.

가르치지 않은 것을 잘 해내라고 요구할 수는 없습니다.

이것은 자체 AI를 개발하면서 저희가 가장 크게 배운 부분 중 하나입니다.

5. 또 하나의 문제 — AI에게 시험지를 미리 보여주고 있었습니다

더 큰 문제도 발견했습니다.
AI를 학습할 때는 보통 **학습 데이터 → AI를 가르치는 데이터**와 **검증 데이터 → AI가 처음 보는 시험 문제**
를 분리합니다.

그런데 기존 방식에서는 하나의 사업계획서에서 여러 개의 학습 문제가 만들어진 뒤 이것을 무작위로 나누었습니다.

예를 들어 하나의 사업계획서에서

요소기술 문제
개발주제 문제
비즈니스모델 문제

가 만들어졌다면, 이 문제들이 학습 데이터와 검증 데이터에 나뉘어 들어갈 수 있었습니다.

그러면 AI 입장에서는 완전히 처음 보는 문제가 아닙니다.

같은 사업계획서에서 나온 내용을 이미 일부 학습한 상태에서 시험을 보는 것입니다.
말 그대로 **시험지가 새고 있었던 셈입니다.**

그래서 v21에서는 방식을 완전히 바꿨습니다.

6. 그래서 AI를 둘로 나누기로 했습니다

이번 v21에서 가장 큰 변화입니다.

기존에는 하나의 AI가
R&D + 사업화를 모두 담당하도록 했습니다.

하지만 이제는 **R&D 특화 모델은** R&D 사전기획을 집중적으로 학습
사업화 특화 모델은 시장분석·경쟁환경·전략·비즈니스모델 등을 집중적으로 학습
하는 방식으로 분리했습니다.

즉, **한 사람이 모든 일을 잘하도록 가르치는 대신, 각자가 자기 분야의 전문가가 되도록 만든 것입니다.**

이렇게 하면 R&D를 더 가르친다고 사업화 능력이 떨어지는 문제도 줄일 수 을 거라고 생각하고 있습니다. 모르죠. 해봐야 합니다.

파국적 망각을 억지로 해결하려고 하기보다, **처음부터 망각이 발생하기 어려운 구조로 바꾸는 것**입니다.

7. 데이터도 처음부터 다시 나눴습니다

이번에는 학습 데이터를 먼저 나누지 않았습니다.
원본 문서 자체를 먼저 나눴습니다.

총 2,072건의 사업계획서와 보고서를 기준으로
1,866건 → 학습용
206건 → 평가용 으로 원본을 먼저 분리했습니다.
평가용 206건은 어떤 버전에서도 학습에 사용하지 않습니다.

그리고 학습 전에 두 데이터가 겹치는지 자동으로 확인하도록 만들었습니다.

이렇게 해야 진짜 실력을 측정할 수 있기 때문입니다.

8. 학습시키기 전에 먼저 시험해 봤습니다

이번에는 무작정 수천 건을 학습시키지 않았습니다.

11개 주요 항목에 대해 먼저 20건씩, 총 **220건을 시험적으로 만들어 직접 확인했습니다.**
그 과정에서 예상하지 못했던 문제들도 발견했습니다.

예를 들어, “필수 항목을 9개 이상 포함하면서 1,600자 이내로 작성하라.”
는 조건이 있었는데 실제로는 두 조건을 동시에 만족하기 어려웠습니다.

AI의 문제가 아니라 **우리가 만든 요구조건 자체가 모순이었던 것입니다.**

이런 문제를 사전에 찾아내고 조건을 수정한 뒤 본격적으로 데이터를 만들었습니다.

그 결과 그동안 자주 빠지던 항목들이 채워지기 시작했습니다.

개발주제의 기술동향·시장수요·기술개요 → **20/20**
개발과제명의 적용분야·정량목표 → **20/20**
비즈니스모델의 고객군·핵심활동 → **20/20**

9. 그리고 v21에서 변화가 나타났습니다

새로운 방식으로 **학습 데이터 약 11,000건**과 **검증 데이터 330건**을 새로 만들었습니다.

학습에는 약 20시간이 걸렸습니다. 물론 gpu가 좋으면 빨랐을텐데... 돈이 필요합니다.

도와주세요 ^^

가장 눈에 띄는 변화는 **검증손실(Validation Loss)**이었습니다.
이 숫자는 AI가 처음 보는 문제를 얼마나 잘 학습하고 있는지를 확인하는 중요한 지표입니다.

v20에서는 **1.0805 → 1.0745**로 거의 변화가 없었습니다.
반면 v21에서는 **0.5811 → 0.4174**까지 크게 내려갔습니다.

아직 최종 성능을 모두 확인한 것은 아닙니다.
하지만 적어도 이번에는 **“새로운 데이터를 실제로 배우고 있다”**는 신호를 확인했습니다.

결국 AI 개발에서 가장 중요한 것은 무엇일까요?

40여 일 동안 네 번의 버전을 만들면서 저희가 배운 것은 의외로 단순했습니다.

① 모델보다 데이터가 먼저입니다
AI가 못한다고 바로 모델을 의심하기보다,
“우리가 그 내용을 제대로 가르쳤는가?”
를 먼저 확인해야 합니다.

② 평가 기준도 함께 만들어야 합니다
AI에게 요구하는 것과 평가하는 것이 달라서는 안 됩니다.
교재와 시험지가 같은 기준을 가져야 합니다.

③ 시험지는 새것이어야 합니다
학습에 사용한 자료와 평가 자료가 섞이면 실제 성능을 알 수 없습니다.
특히 하나의 원본에서 여러 개의 데이터를 만드는 경우에는 **원본 문서 기준으로 학습·평가 데이터를 분리하는 것이 중요합니다.**

④ 안 되는 방법을 계속 반복해서는 안 됩니다
저희도 여러 번 같은 벽에 부딪혔습니다.
결국 문제는 학습 방법 하나가 아니라 **AI를 만드는 구조 자체**에 있었습니다.

⑤ 실패한 모델도 버릴 필요는 없습니다
v19와 v20은 최종적으로 승격되지 않았지만 특정 항목에서는 가장 좋은 성능을 보였습니다.
그래서 그 강점을 다시 추출해 새로운 학습 데이터 제작에 활용했습니다.

우리가 자체 sLLM을 만드는 이유

왜 이렇게까지 직접 만들고 있을까요?

R&D 기획과 기술사업화는 일반적인 질문에 답하는 AI와 조금 다릅니다.
단순히 글을 잘 쓰는 것만으로는 부족합니다.

기술 → 시장 → 경쟁환경 → 사업전략 → 비즈니스모델이 서로 연결되어야 합니다.

저희가 만들고 있는 AI는 바로 이 연결을 잘 이해하고, **R&D 기획부터 사업화 전략까지 실제 업무를 지원하는 AI**
를 목표로 하고 있습니다.
아직 완성된 것은 아닙니다.
오히려 지금도 계속 시험하고 고치고 있습니다.
하지만 직접 만들면서 한 가지는 확실히 알게 되었습니다.

좋은 AI는 좋은 모델 하나를 가져오는 것으로 끝나는 것이 아니라,
그 AI가 잘해야 하는 일을 정확히 정의하고,
제대로 된 데이터를 만들고,
제대로 시험하면서 계속 개선하는 과정에서 만들어집니다.

기술사업화 랩 위드는 앞으로도 저희가 만들고 있는 **R&D·기술사업화 특화 AI의 개발 과정과 실제 개선 결과**를 계속 공유하겠습니다.

R&D 기획이나 기술사업화에서 AI를 어떻게 활용할 수 있을지 궁금하시다면 랩 위드와 함께 고민해 보시기 바랍니다.

이 사업, 신청하면 붙을까요?

R&D 평가위원이 만든 진단으로 우리 기업의 선정 가능성을
항목별 점수로 확인해 보십시오. 무료입니다.

▶ 무료 기업진단 labwith.kr/free-diagnosis.html?utm_source=cafe
▶ R&D 실무교육 아카데미 (연구계획서·PSST·IR DECK)
labwith.kr/training-rnd.html?utm_source=cafe

기술사업화 랩 위드㈜ · 010-6408-8557 · 카카오톡 antipoto

클린봇

기술사업화 랩 위드
댓글을 남겨보세요
등록