

AX의 다음 속제는 PX, 가격 전환이다

잘 팔릴수록 손해 보는 AI 제품... 요금제는 마케팅이 아니라 손익계산서다.

백민기 로이젠트(ROIGENT) 대표

지난 6월 1일 깃허브는 코파일럿의 정액 요금제를 사실상 끝냈다. 월정액에 묶여 있던 프리미엄 요청을 없애고 크레딧 과금으로 전면 전환했다. 1년 앞서 같은 전환을 시도한 커서(Cursor)는 같은 값에 쓸 수 있는 양이 절반 아래로 줄자 이용자 반발에 부딪혀 대표가 사과문을 냈다. 오픈AI는 8월부터 챗GPT 비즈니스 플랜에서 포함 한도를 넘는 사용량을 크레딧으로 과금하기 시작했다. 업계는 이 흐름을 한 문장으로 정리한다. "무제한의 시대는 끝났다."

기업의 AI 전환(AX)은 방향의 문제에서 속도의 문제로 넘어갔다. 에이전트로 업무를 재편하는 프로젝트가 동시에 돌아가고, 그 위에서 새 서비스가 매주 나온다. 그런데 이 전환에서 거의 예외 없이 빠지는 항목이 하나 있다. 가격이다. 모델을 공급하는 회사들은 요금제를 갈아엎었는데, 그 모델을 가져다 서비스를 만드는 회사들은 출시 첫날의 요금표를 그대로 두고 있다.

잘 쓸수록 손해 보는 구조

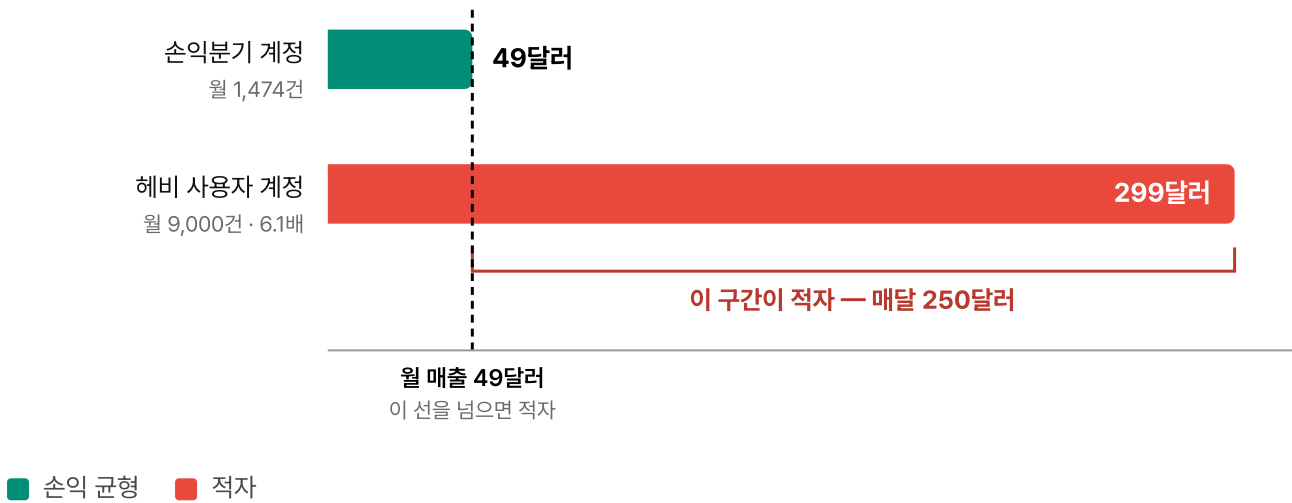
AI 제품은 가격의 문법을 바꿨다. 전통 소프트웨어(SaaS)의 한계비용은 사실상 0이었다. 고객이 얼마나 쓰든 원가는 거의 그대로였고, 덕분에 '사용자 수 곱하기 월정액'이라는 단순한 구조가 20년을 버텼다. AI 제품은 요청 한 건마다 실비가 나간다. 전통 SaaS의 공헌이익률이 70~80%라면 추론 원가가 붙는 AI 제품은 40~60%다. 사용량이 늘면 원가가 선형으로, 때로는 그 이상으로 따라 오른다. **제품이 잘 팔리고 고객이 잘 쓸수록 마진이 나빠진다.**

시트 과금은 여기서 먼저 무너진다. AI 코파일럿을 도입한 고객사가 자동화 효과로 20석을 8석으로 줄였다고 하자. 고객이 얻은 가치는 세 배가 됐는데 공급자의 매출은 60% 줄었다. 고객을 성공시킬수록 내 매출이 주는 역설이다. 정액 무제한 요금제는 반대편에서 무너진다. 요청 수와 매출이 분리돼 있으니 제품을 가장 잘 쓰는 고객이 가장 큰 적자를 만든다.

숫자로 보면 분명하다. 월 49달러 정액제로 문서 질의(RAG) 서비스를 운영하면서 프론티어급 모델을 쓴다고 가정하자. 요청 한 건에 신규 입력 2,500토큰, 캐시 6,500토큰, 출력 700토큰이 드는 워크로드다. 이 서비스의 손익분기는 계정당 월 1,474건이다. 그런데 헤비 사용자는 월 9,000건을 쓴다. 손익분기의 6.1배다. 이 계정 하나에서 매달 250달러가 빠져나가고, 헤비 사용자 비중 15%를 반영한 가중평균 마진은 -33.7%가 된다. 매출은 늘어나는데 이익은 줄어드는 그래프가 여기서 그려진다. 2026년 9월 15일 공개 가격표 기준이고, 토큰 외 계정당 원가를 0으로 둔 값이니 실제 마진은 이보다 나쁘다.

[도표 1] 월 49달러 정액제의 계정별 손익

프론티어급 모델 · 요청당 신규 입력 2,500 / 캐시 6,500 / 출력 700 토큰

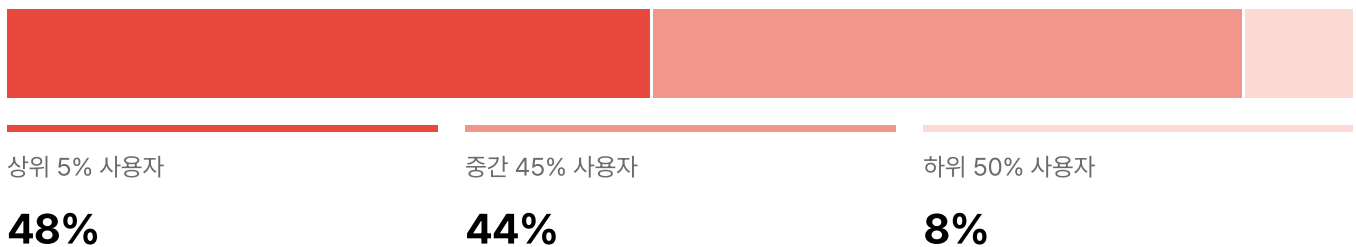


2026년 9월 15일 공개 가격표 기준. 로이젯 산출.

원가 편차도 정액제를 무력화한다. 우리가 진단과 강의에서 반복해 확인하는 분포는 상위 5% 사용자가 전체 원가의 48%를 쓰고 하위 50%는 8%에 그치는 모양이다. '평균 사용자'는 통계적으로 존재하지 않는다. 평균으로 계산한 단일 가격은 라이트 사용자가 헤비 사용자를 보조하는 구조이고, 헤비가 늘어나는 순간 함께 무너진다.

[도표 2] 사용자 구간별 원가 점유

전체 추론 원가를 100으로 놓았을 때 각 구간이 차지하는 몫



로이젯 진단·강의 과정에서 반복 확인되는 분포.

가격표는 이미 답을 말하고 있다

주요 LLM 공급 3사의 공개 가격표(9월 15일 확인)에는 공통된 설계 의도가 있다. 첫째, 입력·출력·캐시 읽기의 단가를 분리했다. 출력 토큰은 입력의 5~8배이고 캐시 읽기는 입력의 10분의 1이다. 원가를 유발하는 축을 그대로 과금 축으로 옮긴 것이다. 둘째, 모델 등급으로 층을 나눴다. 같은 워크로드에서 모델만 바뀌어도 요청당 원가는 73~85배 벌어지고, 같은 모델에서 워크로드만 바꾸면 59~69배 벌어진다. 원가를 흔드는 변수가 두 축에 걸쳐 있고, 가격표가 두 축을 모두 반영한다.

성숙한 AI 요금제는 여기에 네 겹을 쌓는다. 플랫폼 기본료로 수익을 안정시키고, 포함 크레딧으로 고객에게 예측 가능성을 주고, 초과 과금으로 성장에 연동하고, 모델 티어링으로 원가를 방어한다. 무제한을 없애고 페어유즈 조항을 두고 실행당 상한을 걸고 캐싱으로 원가 자체를 낮춘다.

[도표 3] 성숙한 AI 요금제의 네 겹

아래에서 위로 쌓인다. 각 층이 맡는 역할이 다르다.

④ 모델 티어링	원가를 방어한다
③ 초과 과금	성장에 연동시킨다
② 포함 크레딧	고객에게 예측 가능성을 준다
① 플랫폼 기본료	수익을 안정시킨다

여기에 무제한을 없애고, 페어유즈 조항을 두고, 실행당 상한을 걸고, 캐싱으로 원가 자체를 낮춘다.

문제는 이 정교함이 공급자 쪽에 머물러 있다는 점이다. 올해 4~5월 B2B 소프트웨어·AI 기업 230곳을 조사한 결과(Growth Unhinged, 2026)를 보면 시장 전체에서는 기본료에 사용량을 얹는 하이브리드가 37%로 가장 흔한 구조가 됐지만, 연매출 500만달러 미만 초기 기업에서는 정액제가 37%로 가장 많았다. 공급자는 사용량에 비례해 청구하고 나는 정액으로 받는 비대칭이 손익계산서에 쌓이는 데는 두세 분기면 충분하다.

B2B 소프트웨어·AI 230곳 전체	연매출 500만달러 미만 초기 기업
하이브리드 37%	정액제 37%
기본료에 사용량을 얹는 구조가 가장 흔하다	아직 단일 정액제가 가장 많다

PX, 얼마가 아니라 어떻게 받을 것인가

가격 전환(PX:Pricing Transformation)은 가격을 올리거나 내리는 일이 아니다. 제품이 만들어낸 경제적 가치를 가격 구조에 정확히 반영하는 일이고, 그 구조로 매출과 이익을 함께 키워 지속적인 성장을 만드는 일이다. 질문의 순서를 바꾸는 작업이기도 하다. 예전에는 '얼마를 받을까'부터 물었지만 AI 제품에서는 '요청 한 건에 얼마가 나가는가'부터 묻는다.

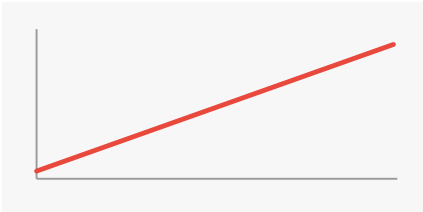
실무는 세 단계로 간다. 첫째, 원가를 실측한다. 추정이 아니라 요청 유형별 실측이다. 토큰 입출력, 모델 등급, 컨텍스트 길이, 재시도와 검증, 멀티스텝 실행, 부가 인프라로 원가 변수를 분해해 고객별·기능별로 귀속시킨다. 대부분의 팀이 이 단계에서 처음으로 구멍을 발견한다.

둘째, 최악의 경우를 먼저 본다. 사용량 분포의 꼬리에서 한 계정이 얼마까지 쓸 수 있는지, 그때 마진이 어디서 음수로 꺾이는지를 계산한다. 포함 한도는 감이 아니라 사용량 분포의 70~80분위에서 잡는다.

셋째, 원가 곡선의 모양을 닮은 가격 구조를 만든다. 사용량에 정비례하는 선형 곡선이면 크레딧이나 종량제, 임계점에서 단가가 뛰는 계단형이면 그 계단을 그대로 플랜 경계로, 분산이 큰 스파이크형(에이전트 멀티스텝)이면 태스크·성과 과금에 실행당 상한을 건다. 요금제의 모양은 원가 곡선의 모양을 닮아야 한다.

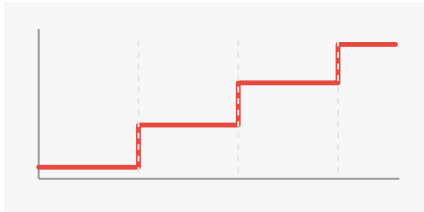
[도표 4] 원가 곡선의 모양과 그에 맞는 요금 구조

가로축은 사용량, 세로축은 원가



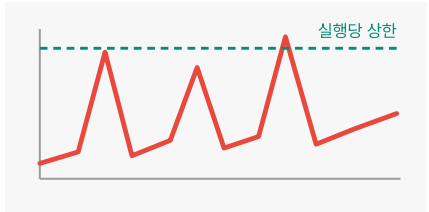
선형

사용량에 정비례한다 → 크레딧이나 종량제



계단형

임계점에서 단가가 뛴다 → 그 계단을 그대로 플랜 경계로



스파이크형

에이전트 멀티스텝처럼 분산이 크다 → 태스크·성과 과금 + 실행당 상한

요금제의 모양은 원가 곡선의 모양을 닮아야 한다.

이 작업은 고객에게 값을 더 받는 일이 아니다. 같은 고객, 같은 사용 분포에서 정가는 그대로 두고 구조만 정액 무제한에서 하이브리드로 바꾼 사례에서 공헌이익률은 58%에서 72%로 올랐다. 공헌이익 기준 86% 증가였고, 고객의 90%는 청구서가 한 푼도 달라지지 않았다. 가격 수준에서는 매출과 이익이 서로 물리지만 가격 구조에서는 둘이 함께 좋아진다.

공헌이익률	공헌이익	청구서가 그대로인 고객
58% → 72%	+86%	90%
정가는 그대로, 구조만 전환	같은 고객 · 같은 사용 분포	한 푼도 달라지지 않았다

구조를 방치한 채 가격 수준으로 대응하면 선택지는 급격히 좁아진다. 공헌이익률 45%인 멀티스텝 에이전트 제품이 10% 인하를 하려면 판매량을 28.6% 늘려야 본전이다. 사실상 인하가 불가능하다. 구조로 풀 수밖에 없다.

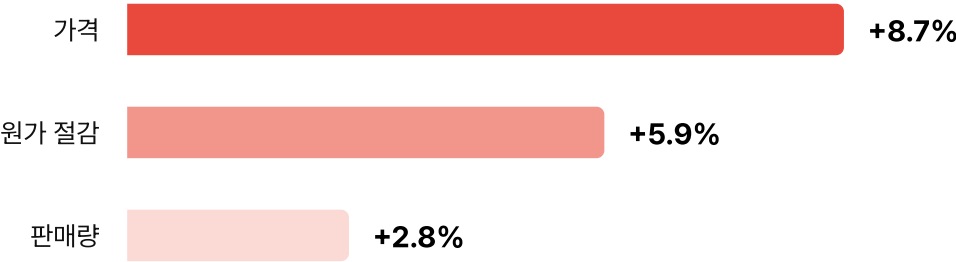
지금 가장 싼 시점이다

가격 구조는 나중에 고칠수록 비싸지는 항목이다. 무료 기간에 형성된 기대는 되돌릴 수 없고, 제품에 미터링을 심어 두지 않으면 나중에는 바꾸고 싶어도 근거가 없다. 이미 시장에 나간 요금제를 바꾸려면 최소 90일의 절차가 필요하다. 고객별 청구액을 시뮬레이션해 인하·동일·인상 그룹을 나누고, 석 달 전에 예고하고, 인상 그룹은 개별로 만나고, 기존 요금 유지 같은 선택지를 준다. 출시 전에 설계하면 이 비용이 전부 0이다.

가격 1%를 개선하면 영업이익은 평균 8.7% 움직인다. 같은 1%라도 판매량은 2.8%, 원가 절감은 5.9%다. 마케팅의 4P 가운데 제품·판촉·유통은 전부 비용이고 수익을 만드는 요소는 가격뿐이다. AX에 투입하는 자원의 일부만 PX에 돌려도 손익의 방향이 바뀐다.

[도표 5] 1%를 개선했을 때 영업이익이 움직이는 폭

같은 1%라도 어디를 건드리느냐에 따라 결과가 다르다



가로축은 영업이익 변화율. 로이젠트 정리.

AI 제품의 요금제는 원가 구조의 거울이다. 원가 곡선의 모양을 모르고 만든 요금제는 언젠가 그 곡선에 무너진다. 사업이 잘될수록 손해가 커지는 역설에서 벗어나는 길은 더 많이 파는 것이 아니라, 파는 방식을 다시 설계하는 것이다.

로이젠트 심층분석 보고서 · 2026년 9월 17일 발행 · 백민기 로이젠트(ROIGENT) 대표
원문 <https://www.roigent.com/report-ai-pricing> · 문의 contact@roigent.com
본 보고서의 수치는 2026년 9월 15일 공개 가격표를 기준으로 산출했으며, 사용량 값은 모형에 넣은 가정입니다.