

forget PRD v2.0 — 비즈니스 에디션

"에이전트 상태 인프라는 어떻게 돈이 되는가"

작성: 2026-07-19 (연구 포크, 사용자 지시). v1(제품 요구사항)과 쌍 — v1은 "무엇을 만들까", **v2는 "누가, 언제, 왜, 얼마를 내는가"**를 소유한다. 충돌 시 게이트 문서 (validation-criteria.md)가 우선.

0. 비즈니스 한 문장

forget은 오픈소스로 에이전트 빌더의 기본 부품이 된 뒤, 그 위에서 세 가지를 판다: 연속성(sync), 신뢰(증명), 운영(managed) — 전부 좌석이 아니라 사용량으로.

1. 시장 구조 판단 (왜 이 순서인가)

1.1 우리가 소거한 것

- **B2C 구독**: 개인은 md 파일·무료 기본 기능과 경쟁 — 지불 의사 낮고 CAC 회수 불가. 개인층은 **매출원이 아니라 유통층**으로 재정의 (MySQL의 무료 사용자 = 영업 부대).
- **탐다운 B2B SaaS**: 1인 회사가 엔터프라이즈 영업 사이클(6-18개월)을 버틸 산소 없음. 좌석 과금 자체가 에이전트 시대에 침식 중.

1.2 우리가 선택한 것 — 문 2 (개발자 기본값) + 고래 1마리

- **주 경로**: OSS + 60초 온보딩으로 "에이전트 세우는 사람의 반사신경"이 된다. 그 개발자들이 다니는 회사가 나중에 관리형·거버넌스 버전을 산다 (bottom-up).
- **부 경로**: 규제 산업 고래 1마리 — 매출보다 **요구사항 명세서**가 1차 가치 (Oracle-CIA 패턴). 삭제 증명·감사·격리가 법적 요건이라 하이퍼스케일러 기본 기억을 못 쓰는 팀.
- **타이밍 논거**: 에이전트 지출은 폭증 구간, 상태(기억)는 실행·신원·결제와 함께 에이전트 경제에서 죽지 않는 원시 자원. 파도를 만들 필요 없음 — 파도의 부품이 됨.

2. 수익 모델 — 3층 케이크 (시간순 활성화)

층	상품	과금	활성 시점	지불 논리
L0 무료	OSS 엔진 + MCP + 로컬 전부	\$0 영원히	지금	유통. 여기서 돈 받으면 문 2가 닫힘
L1 연속성	E2EE sync (개인 멀티디바이스 → 팀 공유 스코프)	개인 \$8/mo · 팀 \$15/석/mo	스파이크 1-2 통과 후 (~9월)	"기억을 잃기 싫다" — 검증된 심리, waitlist가 측정 중
L2 신뢰	삭제 증명 · 접근 감사 · dirty-score 리포트	팀 애드온 +\$10/석 또는 사용량	고래 명세서 확보 후	"규제가 지불을 강제" — 참는 게 불법인 시장
L3 운영	Managed forget (Atlas 모델)	사용량: per-op + per-GB	기본값 확보 후 (게이트 이후)	"네 대신 운영해줌" — OSS 인프라의 검증된 최종 수익화

과금 철학: 에이전트는 좌석을 안 산다 — L3는 순수 사용량(per-op/per-GB), L1·L2의 석 과금도 "사람 수"가 아니라 "스코프 수" 기준으로 전환 가능하게 설계 (P2-4 계층 혹은 배관).

결제 인프라: Paddle KYB 통과 완료 — L1 활성화일에 결제 링크는 하루 작업.

3. 경쟁 지형과 방어선 (돈의 관점)

경쟁자	그들이 파는 것	우리가 파는 것 (비대칭)
AWS AgentCore / Vertex Memory Bank	편의 (플랫폼 통합)	"우리도 못 본다"(E2EE) + "지웠음을 증명" — 하이퍼스케일러가 구조적으로 못 파는 문장
Mem0 / Zep (호스팅 메모리 SaaS)	클라우드 메모리 API	로컬 퍼스트 + 이식성(MCP) + 오염 측정(β) — 그리고 그들의 스토어도 우리 벤치마크의 피평가자
Letta (OSS 에이전트 프레임워크)	에이전트 런타임	우리는 런타임 종립 부품 — 경쟁 아닌 보완 포지션 유지
모델 내장 기억 (ChatGPT 메모리 등)	계정 = 기억 한 통	스코프 격리(복수 맥락) + 도구 간 이식 — 구조적으로 못 따라옴

레이어 흡수 해지 (최심층 리스크): 기억이 모델/플랫폼 안으로 흡수돼도 **상태의 거버넌스**(뭐가 썩었나, 뭘 지웠나, 누가 봤나)는 남는다. L2가 그 보험.

라이선스 (결정 필요, 지금이 공짜): 제안 — 엔진 Apache-2.0 영구 유지(신뢰·채택 극대화) + L2 신뢰 스위트·L3 관리형은 상용 코드로 분리(BSL 불요, 그냥 비공개). Elastic식 사후 변경의 평판 비용 회피. **사용자 승인 대기.**

4. 유통 기계 — 연구가 곧 영업

마케팅 예산 \$0. 대체 자산:

- β 프리프린트** (Dirty Stores): "우리는 진짜 오염된 스토어를 열어봤다" — 인프라 개발자는 광고를 안 믿고 측정을 믿는다. 발행 = Show HN v3 = 유통 이벤트. 부산물 상품화: **dirty-score 리포트**(L2)가 "당신 스토어는 몇 점?"이라는 흑.

2. **벤치마크 신용**: 81.8%/full-500 테크리포트 — 절대치 자량이 아니라 "독립 재현 + 정직 방법론"으로 신뢰 구축 (자기발표 수치 관행과의 차별화).
3. **카테고리 언어 선점**: Verified Takeover Rate — "인수인계가 검증되는가"를 에이전트 기억의 표준 질문으로 만들면, 질문의 저자가 기본값이 된다.
4. **정직 브랜드**: null도 공개($E1, \alpha$) — 연구자 신뢰가 곧 인프라 신뢰.

5. 지표 체계 (비즈니스 갈때기)

unknown-connects/week ← 유일 주간 지표 (무명 = 사인 1순위)
 ↓ (전환: 7일 후 스토어에 실기억 ≥ 10 개)
 활성 스토어 수
 ↓ (전환: 2대 이상 기기 or 팀 스코프 생성)
 L1 유료 전환 ← 게이트 지표: 유료 팀 3개 (Day-90)
 ↓
 L2/L3 확장 (게이트 이후)

- **선행 지표**: β 인용/포크, Show HN 반응, waitlist 성장 — 전부 unknown-connects의 원인
- **허수 경보**: 스타 수·방문자·데모 호평은 지표가 아님 (이미 한 번 속을 뻔한 항목들)

6. 산소 계획 (생존 재무)

- **번레이트**: 창업자 생활비 + 인프라 $\sim \$50/\text{mo}$ + API 실험비(월 $\$200\text{--}500$, 연구 페이지즈 한정). 4090은 소유 자산. **낮은 번레이트 = 파도 지각에 대한 유일한 보험.**
- **펀딩 트랙** (매출과 별도): YC는 W2027 본게임 (F2026 7/27은 스킵 권고 유지 — 벤치마크· β 서사 완성 후가 밸류에이션 유리). 시드는 L1 유료 신호 확보 후.
- **손익분기 산수**: 고정비 월 $\sim \$300$ 기준, L1 팀 20석 $\times \$15 = \$300/\text{mo}$ — **팀 1개가 인프라 손익분기.** 생활비 포함 현실 분기: 팀 ~ 15 개 또는 고래 1건. 이게 "몇 억 펀딩"이 아니라 "팀 몇 개"의 문제라는 것이 이 모델의 핵심 방어력.

7. 마일스톤 — 게이트 정렬 로드맵

시점	비즈니스 이벤트	성공 신호
~7/27	리포 push + Show HN v2 (81.8 동반)	unknown-connects 첫 주 ≥ 30
~8월 중	β 프리프린트 + Show HN v3	인용·포크 + L2 혹은(dirty-score) 문의
~9월	L1 결제 오픈 (스파이크 통과 시)	waitlist→유료 전환율 측정 시작
10/11	Day-90 게이트	유료 팀 ≥ 3 or 고래 명세서 1건 → 지속 / 미달 → 게이트 문서 분기 실행
11월	YC W2027 지원 (게이트 통과 시)	서사: 연구→기본값→매출의 실증

8. 고래 사냥 계획 (부 경로, 병렬)

- **프로파일**: 국내 우선 — 의료(진료 보조 에이전트), 법률(리걸테크), 금융(로보어드바이저 상담 로그). 공통 요건: 대화/기록의 보존·삭제·감사가 규제 사항, AWS 기본 기억 도입 불가.
- **접근**: 영업이 아니라 **연구 인터뷰로 진입** ("에이전트 기억 오염 연구 중인데 귀사 요건을 듣고 싶다") — β 가 명함. 판매 시도 금지, 명세서 수집이 목표.
- **산출물**: 요구사항 명세서 1건 = L2 로드맵의 외부 검증. 20곳 리스트(targets-2026-07.md) 재할용, 규제 산업 필터로 재정렬.
- **소유**: 실행 세션 + 사용자 (인터뷰는 사람이 해야 함 — 기존 병목과 동일하나, 이제 β 라는 진입 명분이 생김).

9. 비즈니스 리스크 (v1 리스크 레지스터의 돈 버전)

리스크	신호	대응
유통 실패 (무명)	unknown-connects 정체 3주	콘텐츠 케이던스 점검 → 채널 실험 (게이트 전 피벗 금지)
L1 지불 가설 오류	waitlist→유료 <2%	L1 가격 재설계 1회 → 실패 시 게이트 분기(L2 중심 전환)
파도 지각	에이전트 기억 지출 부재 지속	산소 최소화로 대기 — 번레이트가 전략
하이퍼스케일러 무료화	AgentCore 기억 기본 무료	예정된 미래 — L2 비대칭으로 선회 가속
창업자 정체성 (연구≫영업)	인터뷰 0 지속	연구 인터뷰 프레임(\$8)으로 재정의 완료 — 그래도 0이면 커머셜 코파운더 안전 상정

10. v1(제품)과의 접점 — 돈이 요구하는 빌드 순서

비즈니스가 제품에 거는 주문은 셋뿐: 1. **P2 (MCP 표면) 완성도** — 유통 기계의 전부. 최우선. 2. **P3-5 (dirty-score)** — β 의 상품화 혹은 L2의 첫 실물이자 가장 싼 것. 3. **P3-1/P3-4 (E2EE)** — L1 활성화의 전제. 스파이크 일정이 곧 매출 일정. 나머지 요구사항은 이 셋을 늦추지 않는 한도에서만.