가치는 어디로 가는가

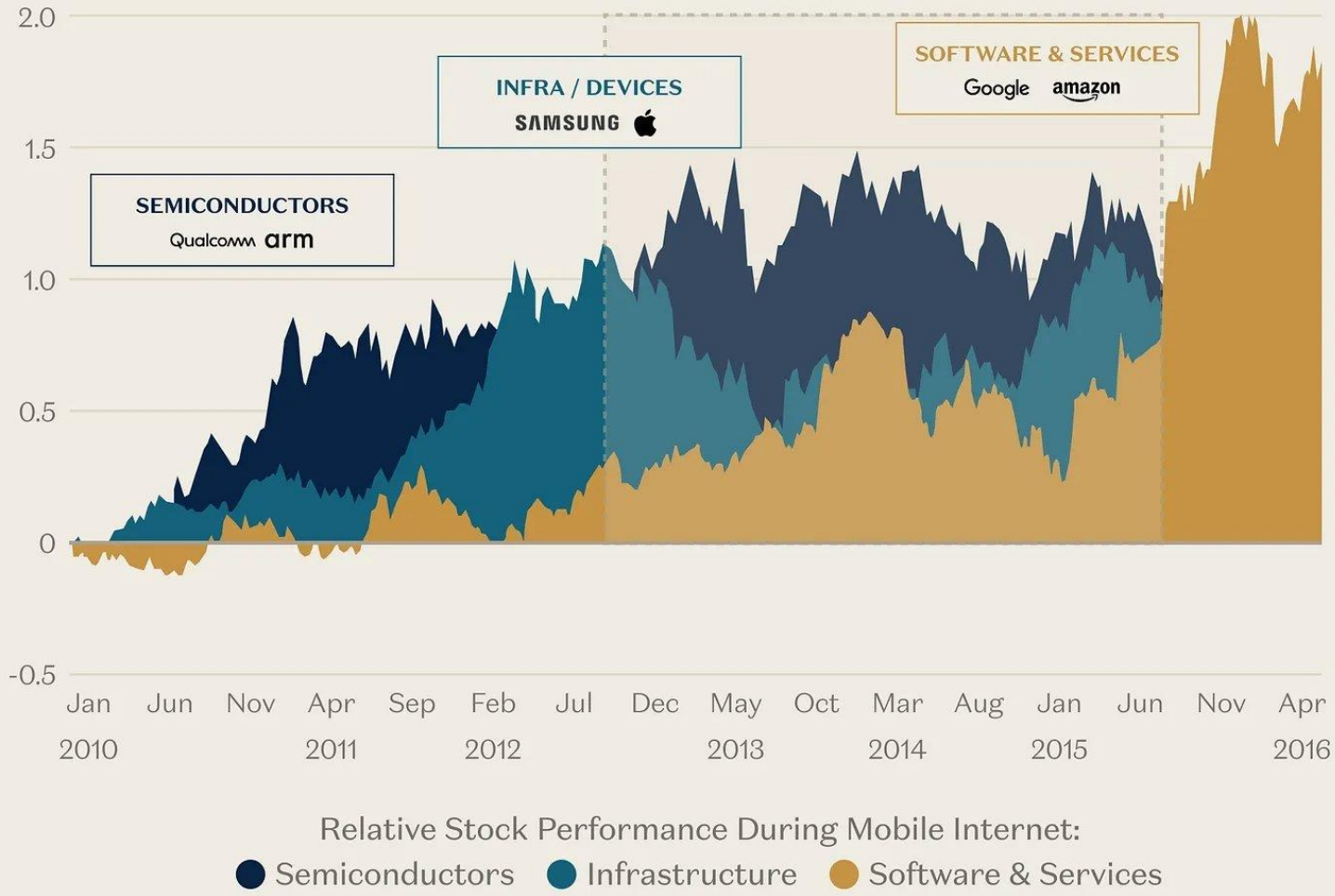
AI market dynamics 속에서 value capture는 누구에게 가는가

— “leak” 하지 않는 “scarcity”로

AI Frontier · 2026년 10월 23일
chester@bfactory.ai

First Semis, Then Software

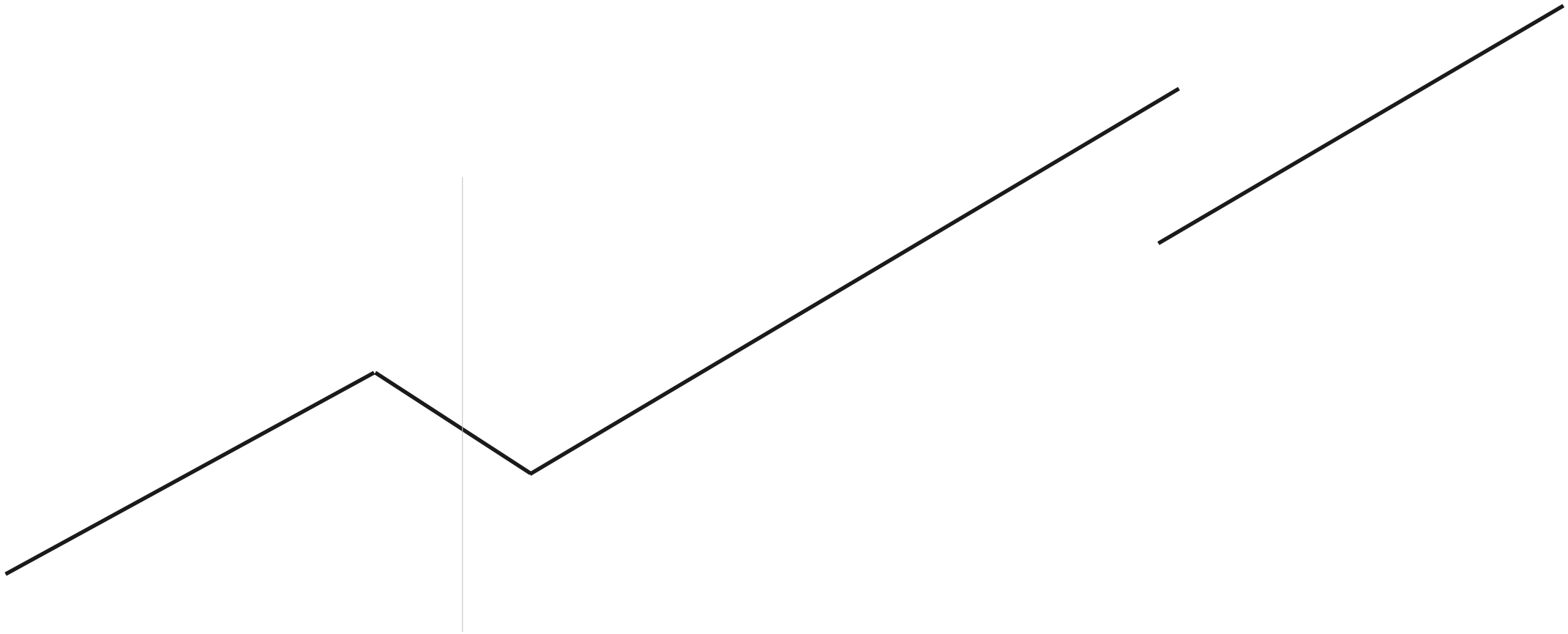
Relative Stock Performance During Mobile Internet



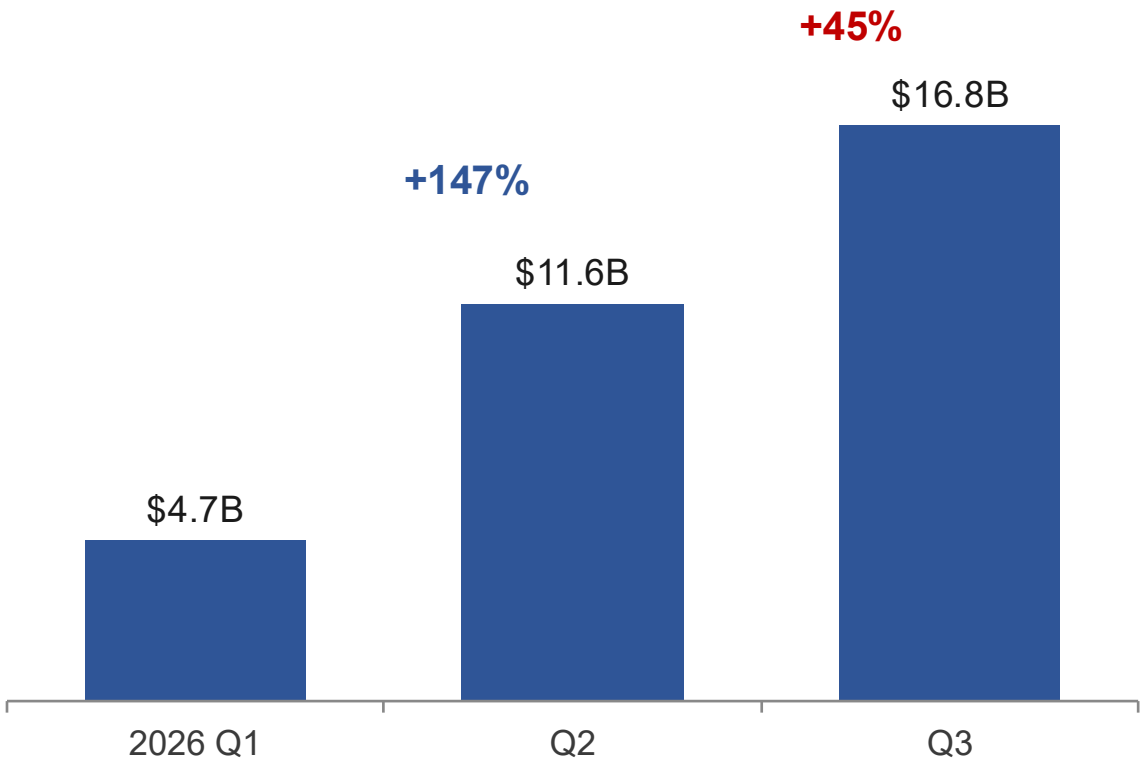
Source: Morgan Stanley Research, FactSet

More charts: a16z.news/subscribe

A16Z


Infra buildout
After Infra buildout
New Paradigm

지난주 ① Anthropic — 숫자는 놀랍고, 기울기는 꺾였다



분기 매출, 십억 달러

Q3 매출 168억 달러

2021년 창업 — 5년 반 만에 월 50억 달러 이상

성장률 +147% → +45%

11월 IPO, valuation 약 2조 달러

ARR 700억~1,000억 달러는 아직 루머

모두가 묻고 있다 — demand의 확장은 계속되는가.

지난주 ② CoreWeave — demand는 진짜인데 CDS spread가 치솟았다

9.75% → 15.78%

2031년 만기 unsecured notes
coupon → YTM (10/15)
액면 100짜리가 80 근처에 팔린다

1,195bp

5년 CDS spread, 10/15 장중 최고
1,000bp 위 = 시장이 부르는 'credit stress'
1년 전 대비 +164%

$104B

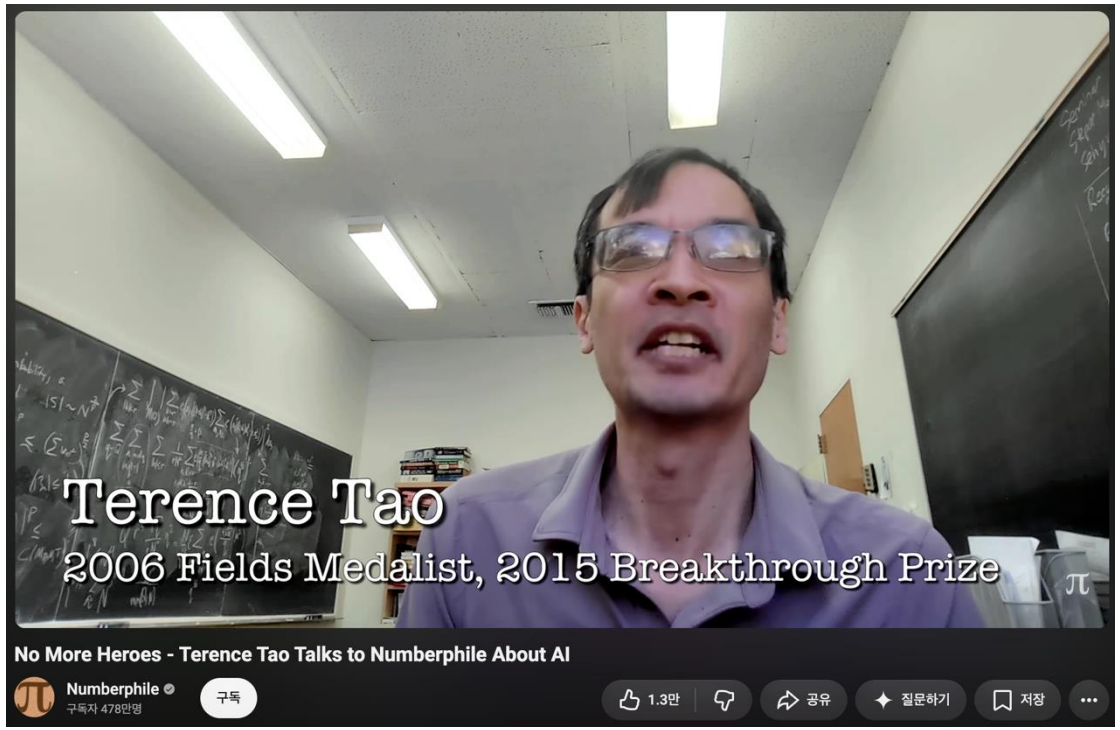
계약 backlog 1,042억 달러
분기 매출 전년 대비 +112%
매출의 98%가 take-or-pay

같은 주, 한국 반도체주(삼성전자·하이닉스)도 내려앉았다.

demand가 없어서 생긴 숫자가 아니다.

demand가 진짜인데, 왜 default 보험료가 치솟는가.

지난주 ③④ 그런데도 우리는 '특이점'을 지나는 중이다



③ 다 끝났어...

가장 초췌해진 테렌스 타오의 얼굴

수학자들이 자기 분야가 넘어가는 걸 보고 있다

DeepSeek V4.1 Flash — 더 작은 모델로 더 나은 성능(?)

④ Tesla Roadster — simply flew.

Tesla / SpaceX, AI 전쟁에 동참 중

'다 끝났다'는 느낌과, default 보험료가 치솟는 시장. 둘 다 진짜다.

Demand는 진짜다.

Risk-free rate가 올랐다.

가장 약한 builder의 CDS spread가 치솟았다.

Demand가 진짜인데 왜 default 보험료가 치솟는가.
이 모순을 풀면 가치가 어디로 가는지가 보인다.

이번 주 ① 시장 — 같은 질문을 더 큰 목소리로 물었다

−2.6%

Nasdaq, 10/22 (목) — 6월 이후 최대 하락
Meta −7% · Amazon −4.6% · Nvidia −4.4%
10/6 고점 대비 −5.8%

4.95%

10년물 국채, 10/22 — 5% 아래로
2년여 만의 최대 일중 하락, 정점은 5.3% 위
AI 주식에서 국채로의 피난

−10%

CoreWeave, 10/22 → $55.60
장중 $53.50, 52주 최저
6월 고점 $138.25 대비 −60%

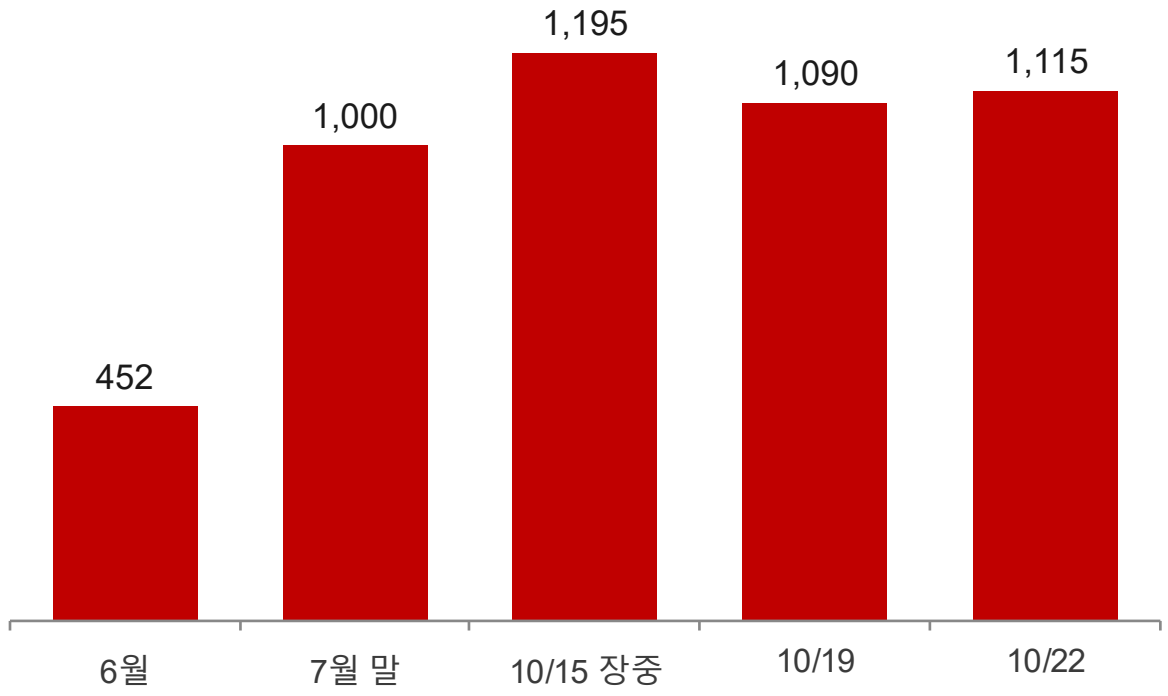
Brent는 이란 전쟁으로 $114, 프랑스 10년물은 5.1% — AI만의 한 주는 아니었다.

Demand는 여전히 진짜다. Huang "neocloud demand는 off the charts", CoreWeave backlog $104B, Anthropic Q3 adjusted 흑자. 보험료만 더 올랐다.

모순은 커졌지, 풀리지 않았다.

이번 주 ② CoreWeave — 숫자가 한 단계 더 갔다

CoreWeave 5년 CDS spread, bp — 1,000 위에 머문다



주가

10/22 $55.60(−10%), 장중 $53.50 — 52주 최저. 시총 $31B.

채권

9.75% 2031년물 YTM이 10/22 다시 최고치(15.7%대), 가격 80 근처. 8월 중순엔 11.3%.

CDS

15일 고점에서 내려왔지만 1,000 위. 연초 대비 +171%. Oracle CDS는 2021년 이후 최고, hyperscaler 4사 CDS도 수개월 내 최고.

회사 쪽

compute 가격 7월 +25%, 10월 +10% — 최신 세대는 아직 scarce. Intrator "시장이 오해하는 시기가 있다", Blackstone Gray "underlying business는 꽤 강하다".

Demand가 아니라 timing — 그리고 이제 다음 refinancing의 가격이다.

이번 주 ③ 공매도 — 시장이 먼저 판 것은 AI가 아니라 leverage다

$6B

neocloud 5사 공매도의 10월 평가이익 (ORTEX)
CoreWeave · Nebius · IREN · Applied Digital · Core Scientific
아직 청산 전 — 반등하면 줄어든다

$17B

공매도 잔고 — free float의 20~26%
보통 대형주는 1~3%, 10%를 넘으면 'heavily shorted'
그만큼 short squeeze 위험도 크다

23.7%

CoreWeave 공매도 비중, 월초 19.6%
주가가 30% 빠지는 동안 포지션을 더 늘렸다
10월 이익 약 $1.7B · Nebius $2.94B

공매도가 몰린 곳은 Nvidia도 hyperscaler도 아니다. 빌린 돈으로 GPU를 깔아 임대하는 중개자 다섯이다.

BofA: 지난주 매도는 "AI trade 변두리의 positioning-driven hiccup", 가장 아픈 곳은 "highly levered" 종목. 빌리는 비용은 아직 싸다 — squeeze 단계는 아니다.

Spread로 사는 marginal builder는 어느 국면에서도 loser — 시장이 먼저 가격에 넣었다.

이번 주 ④ 월가는 둘로 갈렸다 — '파산'을 어느 층에서 보느냐

Goldman — Buy, $106 (10/19)

SPV 안의 자산을 본다.

차입 비용이 100~200bp 올라도 신규 project의 return은 매력적 — 목표가는 $139에서 내렸다.

Bernstein — Underperform, $51 (10/22)

모회사의 funding gap을 본다.

backlog를 이행하는 데 약 $85B · 확보된 facility $9B · 부족분 $76B를 밖에서 조달해야 한다. Goldman은 이 gap을 그 3분의 1 아래로 본다.

JPMorgan — 은행 exposure (10/20)

은행의 층을 본다.

SPV plumbing 덕에 CoreWeave가 default해도 주요 은행 exposure는 각 $1~3B — '제한적 위험'. 단, 주식과 unsecured notes가 안전하다는 뜻은 아니다.

같은 회사를 두고 목표가는 $39(HSBC)에서 $135(평균)까지 벌어져 있다. 회사 쪽은 backlog가 "충분히 뒷받침한다"(Intrator), Huang은 "제약은 working capital뿐".

둘 다 맞을 수 있다. 그래서 '파산'은 세 층(SPV · 모회사 · liquidation)으로 나눠야 답이 나온다.

이번 주 ⑤ 'Bubble'을 말한 사람들

Ken Griffin · Citadel · 10/20

지금은 "almost certainly" 버블이다. "급진적 전환은 대개 버블과 얽혀 있다." 살아남는 회사는 "수십 년의 시장 리더" — "그때 Amazon 주식을 팔지 말 걸 그랬다." 걱정은 주가가 아니라 기업들이 AI 지출에서 가치를 뽑는 능력이 "매우 고르지 않다"는 것.

Larry Fink · BlackRock · 10/21

버블론 일축. "데이터센터를 안 지으면 compute가 비싸서 오히려 더 inflationary하다." "capital을 두고 벌이는 경쟁은 내 경력에서 본 적 없는 규모다."

Christian Sewing · Deutsche Bank · 10/21

"터지는 버블 속에 있다고 보지 않는다 … 버블과 다른 건 계속되는 변동성이다."

Arvind Krishna · IBM · 10/21

대기업들이 "AI 지출이 통제 불능이고 실제 return이 없는 게 아닌지 걱정하기 시작했다."

Ray Dalio · 10/7

"classic bubble"의 특징을 다 갖췄고, 터지기 직전 단계에 "가까워지고 있다" — 빚으로 쌓은 확장, 금리 상승, 현금화 압력.

숫자로는

UBS — hyperscaler capex가 2026년 영업현금흐름의 약 100% (10년 평균 40%)

Morgan Stanley — 올해 AI 관련 debt 발행 약 $570B, 작년의 두 배

Meta Q3 — FCF −$2.1B · 2026 capex $135~145B · AI funding을 위한 의결권 없는 Class C 주식 신설 · 다음 날 −7%

BofA — "변두리의 positioning-driven hiccup" / IG — "소화지 붕괴가 아니다"

'버블인가'는 갈렸고, '돈이 비싸졌다'에는 양쪽이 동의한다.

이번 주 ⑥ Lab — funding이 숫자로 드러났다

$450B

Anthropic S-1 — 향후 7~10년 취소 불가 cloud 지급 의무
Google $111B · AWS $110B · Azure $31B · Broadcom $161B · Nscale/Lambda 최대 $86B
협상 중인 것까지 합치면 $750B

$80.7B

상반기 순손실 — 그중 $72B는 convertible 재평가 비현금 손실
매출 $16.3B(전년 $1.1B) · compute 지출 $9B · 상위 고객 둘이 매출의 20%
Q3 $16.8B는 adjusted 흑자, FCF 양수

Valuation이 재협상됐다

Anthropic — 2조 이야기가 1.5조로(Reuters).

OpenAI — 최근 round 목표 $1.4T → $1.25T(FT, 10/22).

Firmus(Nvidia 투자) — $30B valuation 반발로 호주 IPO 철회.

두 lab의 compute는 소유가 아니라 임차다

GPU 소유자는 Oracle · Microsoft · Amazon · Google · CoreWeave. lab은 take-or-pay로 전력 단위를 산 임차인이다. 자기 compute를 소유한 lab은 xAI뿐.

Lab도 funding spectrum 위에 있다. 'Lab 지분은 보험'은 이번 주에 더 맞는 말이 됐다.

이번 주 ⑦ Kimi K3.1 Pro — coding에서 time gap이 6주로 내려왔다

Benchmark	Kimi K3.1 Pro (10/23, open-weight)	Claude Opus 5.5 (9/22)	GPT-6 Astra (9/3)
Terminal-Bench 4.0 (coding agent)	**65.7**	66.4	58.2
BenchCAD (with tools)	92.7	96.2	95.9
Cyber (자체 bench)	81.0 — hacking data 제거 전 90.5	"3개월 이상 공개된 frontier에 뒤진다" (Moonshot)	
HLE (tools) · ARC-AGI-3	미보고	67.7 · —	57.2 · 99.9

7월의 K3 pretrain(2.8T MoE) 위에 post-training을 훨씬 더 얹은 점 릴리즈다 — 새 pretrain이 아니다. 전부 자체 보고치, weight는 11/4.

Epoch 기준 open–closed gap은 종합 index로 4개월(1~5월). coding agent 한 축에서는 Opus 5.5(9/22) → weight 공개(11/4), 약 6주.

줄어든 곳이 가장 verifiable한 domain이라는 것, 그리고 줄인 방법이 RL compute라는 것 — 둘 다 3절의 예측 그대로다. RSI 신호는 아니다.

Verifiable domain에서 gap이 먼저 닫힌다 — 지표 ④는 S3의 반대 방향으로 반 발짝.

이 thesis로 읽으면 이번 주는 세 줄이다.

① 시장이 먼저 판 것은 AI가 아니라 leverage다.

공매도 이익 $6B는 neocloud 다섯에 집중됐다. BofA의 "highly levered"가 그 말이다. 6절의 결론을 시장이 먼저 가격에 넣었다.

② Risk-free가 내려와도 spread는 최고치를 다시 썼다.

국채로 돈이 피난하며 10년물이 4.95%로 내려왔는데 CoreWeave spread는 더 벌어졌다. 눌리는 건 금리의 수준이 아니라 '남이 못 사는 걸 쥐었는가'다.

③ Lab의 아킬레스건이 funding이라는 말이 S-1의 숫자로 나왔다.

$450B의 취소 불가 compute 의무는 lab도 funding spectrum 위에 있다는 뜻이고, 2조가 1.5조가 된 건 그 가격의 재협상이다.

Demand는 진짜다. 보험료는 더 올랐다. 모순은 커졌지 풀리지 않았다.

질문

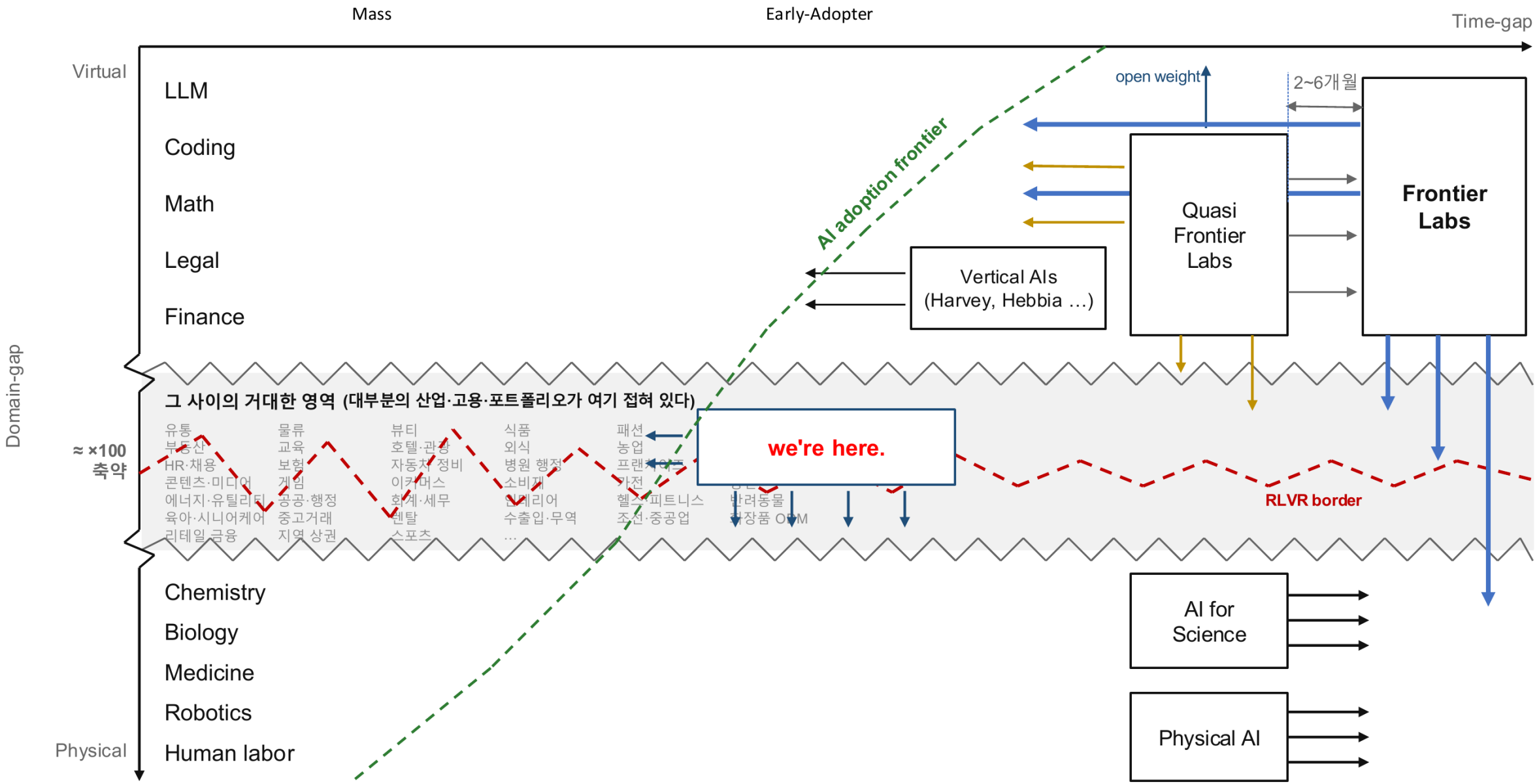
매크로는 도대체 어디로 흘러가는가.

이렇게 혼돈스러운 시기에 '가치'는 어디로 이동하는가.

우리는 어디에 투자해야 하고,
어떤 사업을 해야 하는가.

과거에서 빌릴 렌즈 —
1830~1900 railroad boom, 2000년대 telecom bubble
과잉 투자가 일어났고, 실수요는 늦게 따라왔고, 그 사이에 심한 구조조정이 있었다.

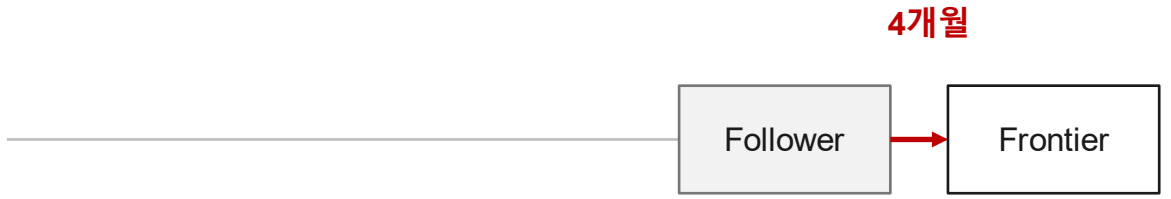
우리의 위치 : time-gap × domain-gap — 그 사이에 접혀 있는 거대한 기회



두 개의 gap — scarcity란 아직 아무도 닫지 못한 gap이다

Time gap

frontier와 follower의 시간차



open-weight는 frontier 4개월 뒤 (Epoch, 2026.1~5)

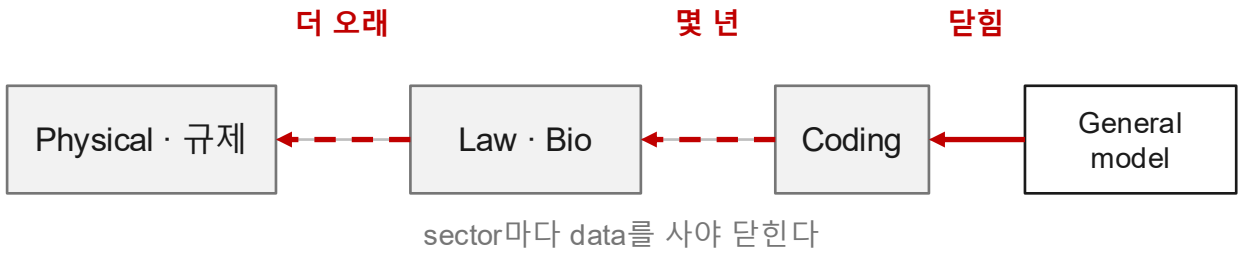
미·중 최고 모델 격차 Arena 2.7%

같은 성능의 가격, 매년 약 10배 하락

capital이 leak해서 빨리 닫힌다

Domain gap

general model이 하는 것과
domain이 요구하는 것의 거리



coding은 닫혔다

law · bio · physical은 열려 있다

닫으려면 sector마다 data를 사야 한다

느리게 닫힌다

Model이 commodity보다 비싸게 팔리려면 time gap이 열려 있어야 하고, model layer 위에서 사업이 생기려면 domain gap이 존재해야 한다.

누가 그 자리를 차지하는지는 gap을 누가 닫느냐로 정해진다 — lab이 data를 사서 닫는가, operator가 signal을 쥐고 닫는가 (3절).

가치는 더 오래 열려 있는 gap 쪽으로, 즉 위로 간다.

오늘의 결론 두개를 미리 이야기하고 갑니다.

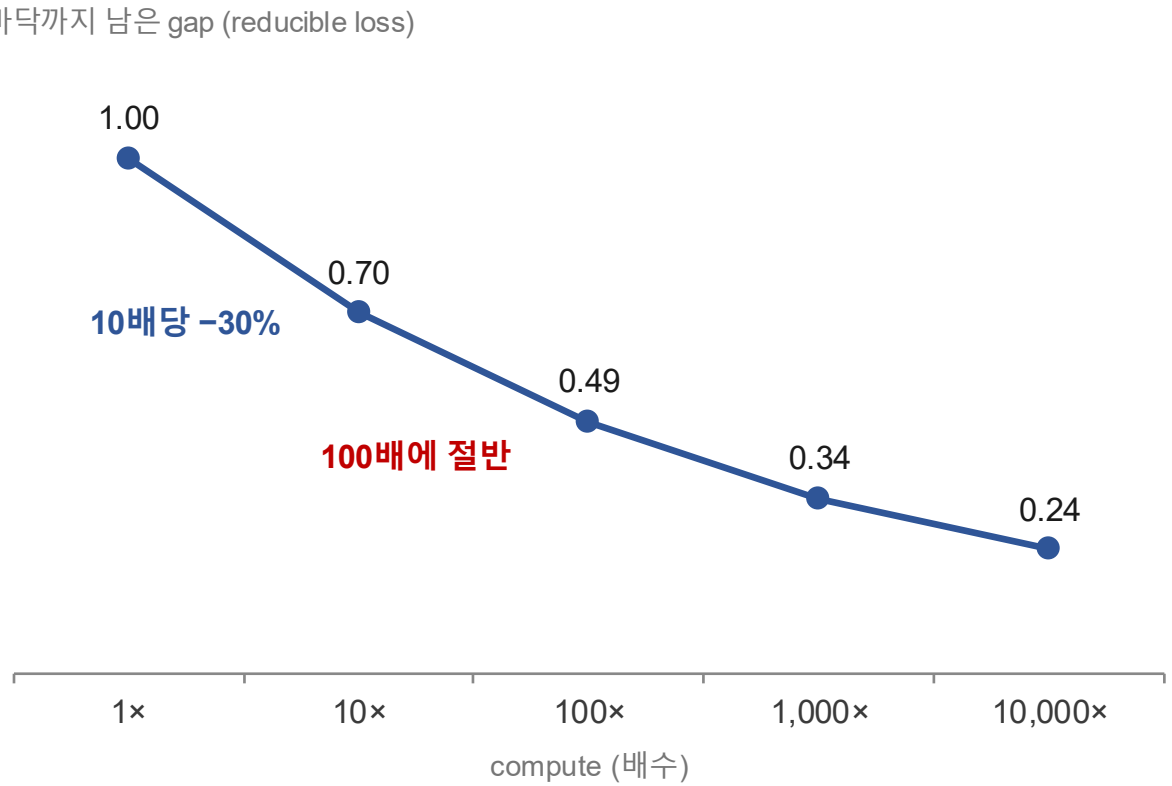
가치는 leak하지 않는 scarcity로 간다.
그런 scarcity는 둘뿐이다.

Physical bottleneck
전력 · 부지 · supply chain
build-out 국면에 돈이 된다

Operation이 만드는 reward signal
outcome이 현실에서 드러나야 생기는 신호
그 뒤에 돈이 된다

Frontier Lab Dynamics

Scaling law still reigns. — 그게 lab의 상방 제약이다



gap을 절반으로 줄이는 데 늘 compute 100배

지금 속도로 100배 = 약 2년 (물리 compute 4~5배 × 알고리즘 ~3배 = 연 10배)

RSI : AI가 다음 AI를 만드는 연구를 대신해, 한세대에 드는 compute 가 세대마다 줄어드는 것

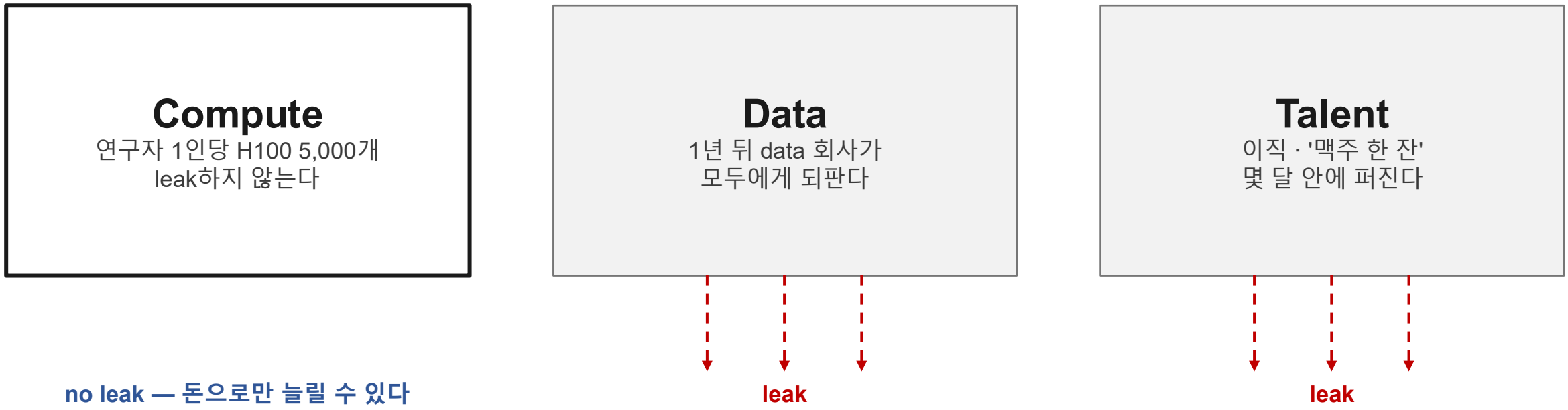
결과 : 곡선모양은 그대로. 반감 2년 → 1년 → 6개월

다음 모델은 GPU가 덜 들고, 훈련이 더 짧아지는데, 더 좋다 → 아직은 아님 (여전히 pre-train compute 은 연4~5배)

loss의 절대값은 아무도 공개하지 않는다. 바닥(≈1.7~1.8)도 추정치. 그래서 비율로만 말한다.

Verifiable domain은 이미 AGI를 지났고, general progress는 100배에 묶여 있다. 둘 다 참이다.

Frontier Lab을 지탱하는 세 capital 중 leak하지 않는 것은 compute뿐이다



그래서 time gap은 결국 compute gap이다.

lab의 진짜 moat는 capability가 아니라 cheapest intelligence per watt (serving cost). time gap이 0이 돼도 commodity 구간에서 cost로 이긴다.

Anthropic inference margin 38% → ~70% (SemiAnalysis), adjusted GM 85% — 당연히 줄어들 개연성이 있다.

모두가 이러한 Compute gap 을 메꾸기 시작하면, model layer는 utility가 된다.

Compute의 단위 — 1 GW가 무엇인가

1 GW = 10억 W. 1년이면 8.76 TWh — 서울시 연간 소비의 18%, 원전 APR-1400 한 기의 70%. 업계의 GW는 냉각 포함 facility 전력이고 칩에 가는 IT 전력은 그 70~85%다.

칩 세대가 바뀌면 GPU 수도 랙 수도 FLOPS도 바뀌지만 전력은 안 바뀐다. 땅·변전소·냉각은 GW로 짓고, 칩은 2년마다 갈아 끼운다. 그래서 계약은 GW로 한다.

1 GW facility	Blackwell — GB300 NVL72 (2025~26)	Vera Rubin — NVL144 (2026 하반기~)
랙 하나	GPU 72개 · 120~140 kW · FP4 약 1.1 EFLOPS	Rubin 패키지 72개(다이 144개) · 130~190 kW · FP4 3.6 EFLOPS
랙 수	약 5,000~6,000	약 3,700~5,300
GPU	약 40만~45만 개 (Jensen: 10 GW = 400~500만 개)	약 27만~39만 패키지
FP4 inference	약 6 ZFLOPS	**약 13~19 ZFLOPS — 같은 전력에서 2~3배**
capex	약 $50~60B (Jensen, 2025.9 · Nvidia 몫 ~$35B)	비슷하거나 약간 위

Blackwell 1 GW × 3개월 ≈ 3×10^{27} FLOPs = GPT-4(2×10^{25})의 100배 남짓 = 지금 frontier pretrain 한 번. 다음 100배(10^{29})는 10 GW급을 몇 달 돌리는 run이다 — OpenAI가 10 GW를 말하는 이유.

칩은 갈아 끼우고 전력은 남는다 — 그래서 단위는 GW다.

두 lab이 가진 compute — 돌아가는 것, 계약된 것, 계획

OpenAI

돌아가는 것

2025년 말 "2 GW 조금 넘게"(Altman). Stargate Abilene 0.3 GW(Epoch, 4월) + Azure · CoreWeave($22.4B) · Oracle · AWS($38B) — 지금 3~4 GW 안팎(추정)

계약된 것

4월 "미국 내 10 GW 확보"(원래 2029년 목표). Stargate 7곳 9 GW+, 대부분 2028년 완공 — Epoch: 9 GW ≈ 2025년 말 전 세계 AI compute

계획

Nvidia ≥10 GW(첫 1 GW는 2026 하반기 Vera Rubin) · AMD 6 GW · Broadcom 자체 칩 10 GW = 26 GW. 공개 commitment 약 $1.15T

Anthropic

돌아가는 것

AWS Trainium2 100만 개+(≈1 GW급) · Google TPU "2026년에 1 GW 훨씬 넘게" · Azure의 Nvidia 최대 1 GW · SpaceX Colossus 1 전체 (300 MW+, GPU 22만) — 2026년 말 3~4 GW 안팎(추정)

계약된 것

S-1의 $450B 취소 불가 — Google $111B · AWS $110B(+5 GW) · Microsoft $31B · Broadcom $161B(차세대 TPU 5 GW, 2027~) · Nscale/Lambda 최대 $86B

계획

2027년 말 약 9.5 GW(보도). TPU · Trainium · Nvidia 삼중 — cheapest intelligence per watt를 위한 협상력

둘 다 소유가 아니라 임차다. 자기 compute를 소유한 lab은 xAI뿐이다.

Data — domain이 넘어가는 순서는 lab의 attention이 정한다

lab의 attention이 domain을 먹어 들어가는 순서



순서의 규칙 — verifiable ∧ digital ∧ 큰 것부터

lab의 attention ∝ 시장 크기 ÷ data를 사는 난이도

reward signal(평가 기준)만 구할 수 있으면 앞으로 갈 수 있다 — jagged intelligence의 주된 동력

회계·법률처럼 signal이 뚜렷하지 않은 곳도 전문가를 동원해 RL environment를 만들어 해결해 가고 있다

그런데 '전문가를 동원해 만들 수 있는 signal'과 '만들 수 없는 signal'이 있다.

Buyable signal vs. unbuyable signal

Buyable

화면 앞의 제3자가 그 자리에서 판정할 수 있는 signal

법률 문서 · 회계 계산 · 코드 테스트

전문가를 고용해 RL environment로 만든다

→ lab이 사서 닫는다

Unbuyable

outcome이 현실에서 시간을 두고 드러나야 생기는 signal

다시 샀는가 · 처방으로 좋아졌는가

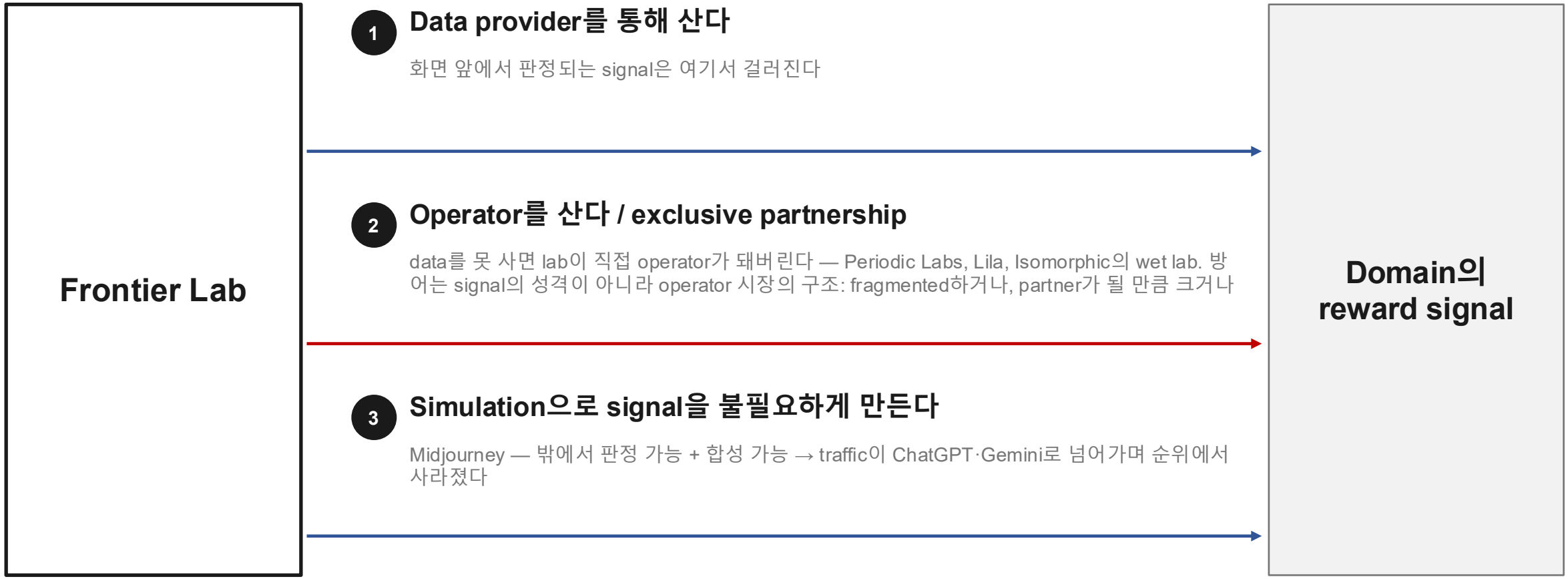
수리가 90일을 버텼는가 · 이 batch의 수율

→ label 자체가 operation 밖에 존재하지 않는다

lab은 synthetic environment와 outsourced labeling으로 현실을 근사한다.
operator는 진짜 환경을 갖고 있다 — 실제 고객, 실제 outcome, 실제 책임.

경계는 'signal이 있나'가 아니라 '그 signal이 operation 밖에서 존재할 수 있나'다.

Frontier lab의 공격 경로 셋



그래서 각 domain의 실제 승부는 '누가 operator가 되는가'다.

Talent — recipe는 다 알려졌다. Frontier lab이 만드는 spillover

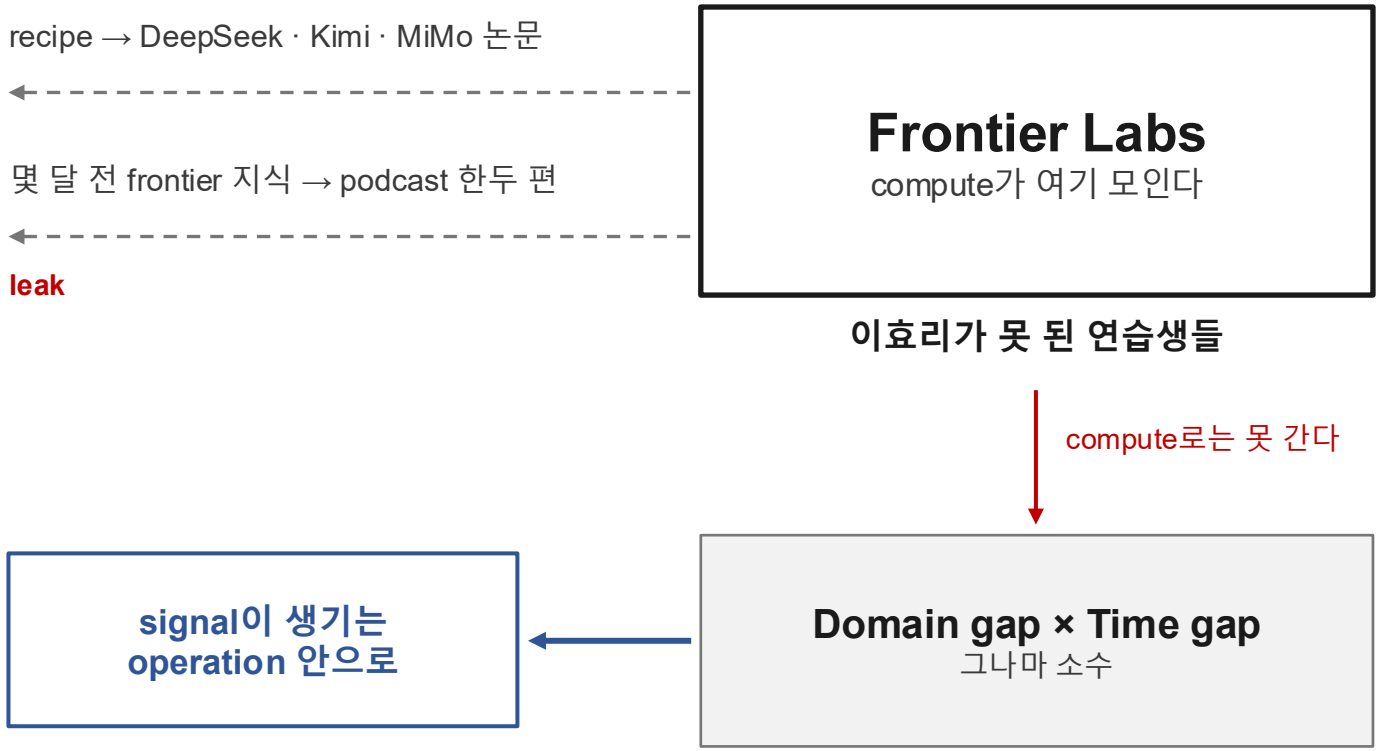
방법은 비밀이 아니고, 비율만 비밀

'맥주 한 잔이면 퍼진다' — 20~30개 회사가 좋은 모델을 만든다

pre-training은 10억 달러 없이도 가능, post-training은 bar가 낮다

연습생이 들고 나가는 recipe는 leak하는 capital이라 가치가 없다.

가치는 모델이 아니라 operation 안에 loop를 짓는 것 — contract · outcome join · eval set



Spillover가 가야 할 곳은 'signal이 생기는 operation 안'이다.

Near Frontier Lab Dynamics

Near-frontier lab — capability moat가 아니라 ownership moat

존재 이유 둘 — sovereign AI, on-prem enterprise

'Near'를 떼지 않는 한 TMAX · 한컴 수준의 top-line

top-line = 모델 성능이 아니라 조달 규제 × data 반출 제한의 크기

Frontier는 closed, near-frontier는 open이 된다

공격이 아니라 진입 — API만 팔면 아무도 모를 회사가, 공개하면 대화에 끼어든다

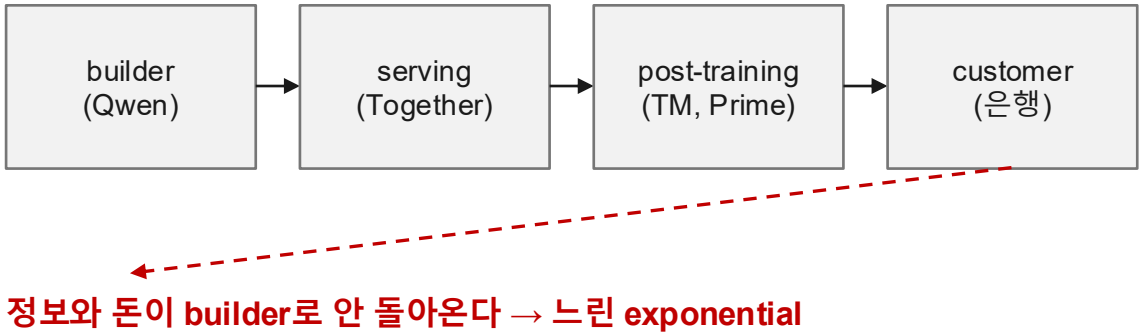
open의 자리 — API로 못 보내는 private data + 자기 하드웨어 위의 weight (enterprise token의 38%)

Closed — loop가 한 회사 안에서 닫힌다

model → serving → product (Claude Code) → 정보·돈 → next model

빠른 exponential

Open — loop가 네 회사에 걸쳐 있다

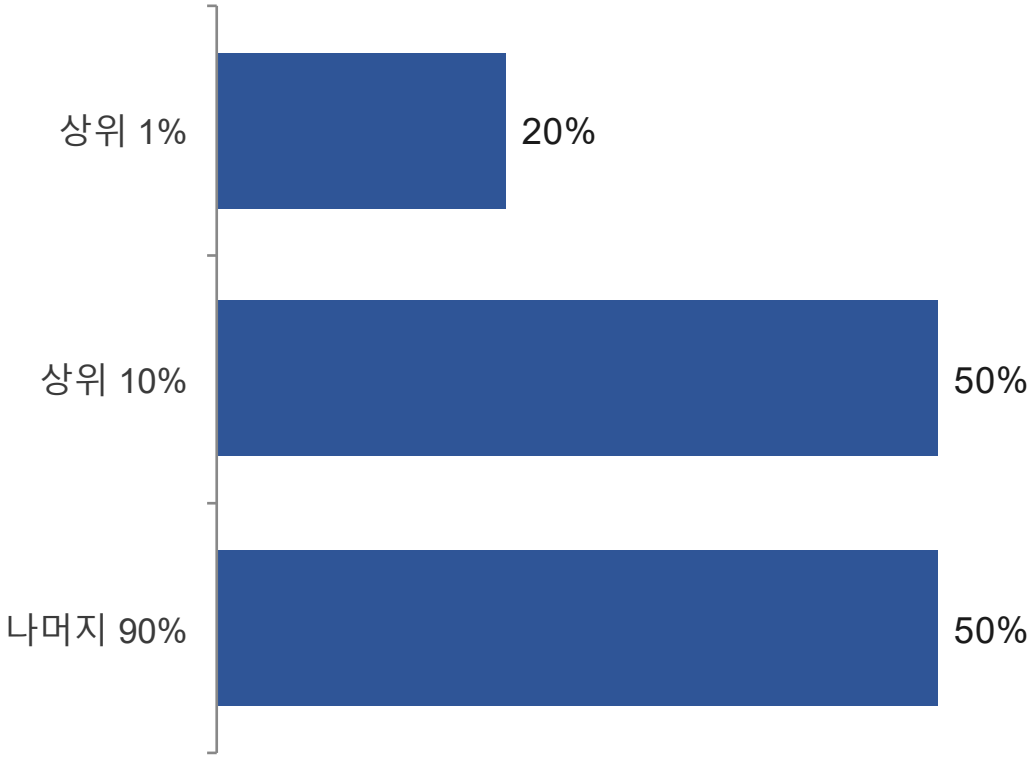


Open의 종착지는 속도가 아니라 ownership이다.

Key Metrics

지표 ① Token 소비는 극단적으로 편중돼 있다

사용자 구간별 매출 비중



1%가 매출의 20%, 10%가 50%

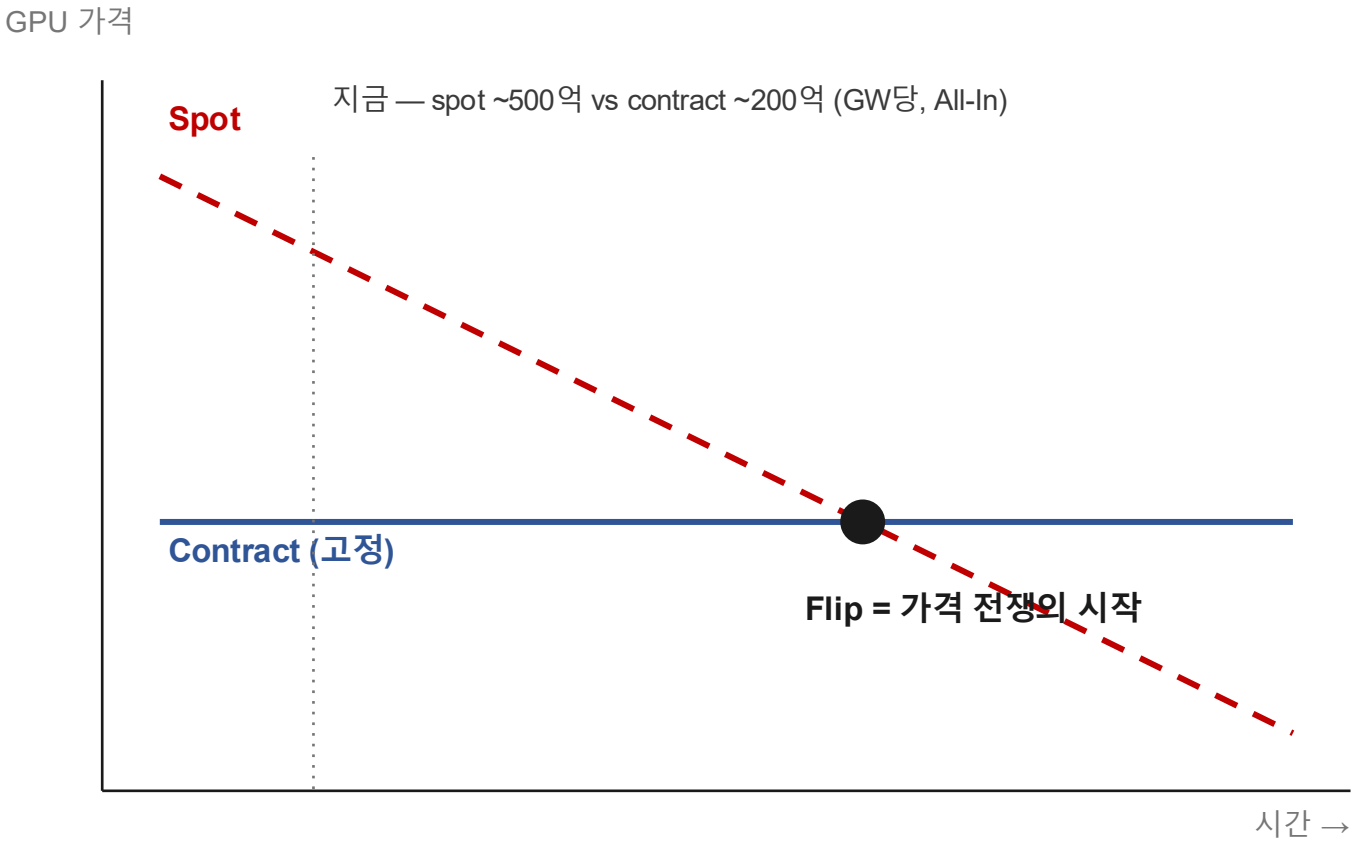
결제자는 전체의 4.5%

상위 1% 결제자 월 903달러, median 25달러 (a16z 7판)

쇼핑 · 소셜 · 엔터테인먼트 — 시간을 쓰는 category에 AI native 승자가 없다

App의 시대는 아직 오지 않았다.

지표 ② Spot > Contract — build-out 초기의 모양



이미 dual market

H100 on-demand −64~75% (2024 → 2026 Q2) — 구세대는 이미 flip됐다

최신 세대 1년 reserved +40% — 아직 scarce하다

flip은 세대별로 순차적으로 온다

Flip은 고정가에 산 쪽을 죽이고, 판 쪽을 살린다

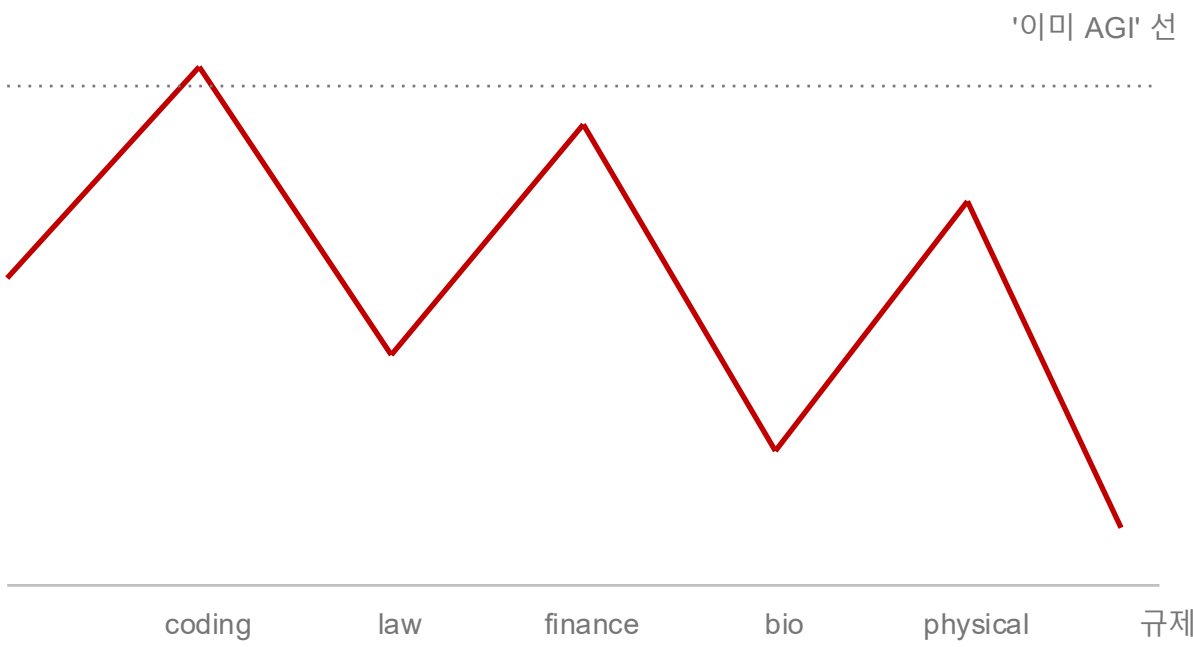
Loudcloud 1999 — $1,500/Mb에 3년 고정으로 샀다가 2001년 $50

CoreWeave의 위험은 flip 자체가 아니라, 3년 contract의 갱신 — 15년 lease와 5년 debt는 그대로

지표: 최신 세대의 1년 reserved 가격이 하락으로 돌아서는 날.

지표 ③④ 둔화는 사고가 아니라 구조다

jagged intelligence — domain마다 높이가 다르다



③ Anthropic의 둔화는 구조다

coding = verifiable ∧ digital ∧ 지불 의사 높음 → data 구매 없이 닫혔다

다음 domain은 data를 사야 한다 → sector별로 몇 년

coding 다음의 대체 demand는 coding만큼 가파르지 않다

④ 체감 가치의 증가

GPT-6 Astra, Opus 5.5… jaggedness가 사라지는 것을 두고 호들갑을 떨지만

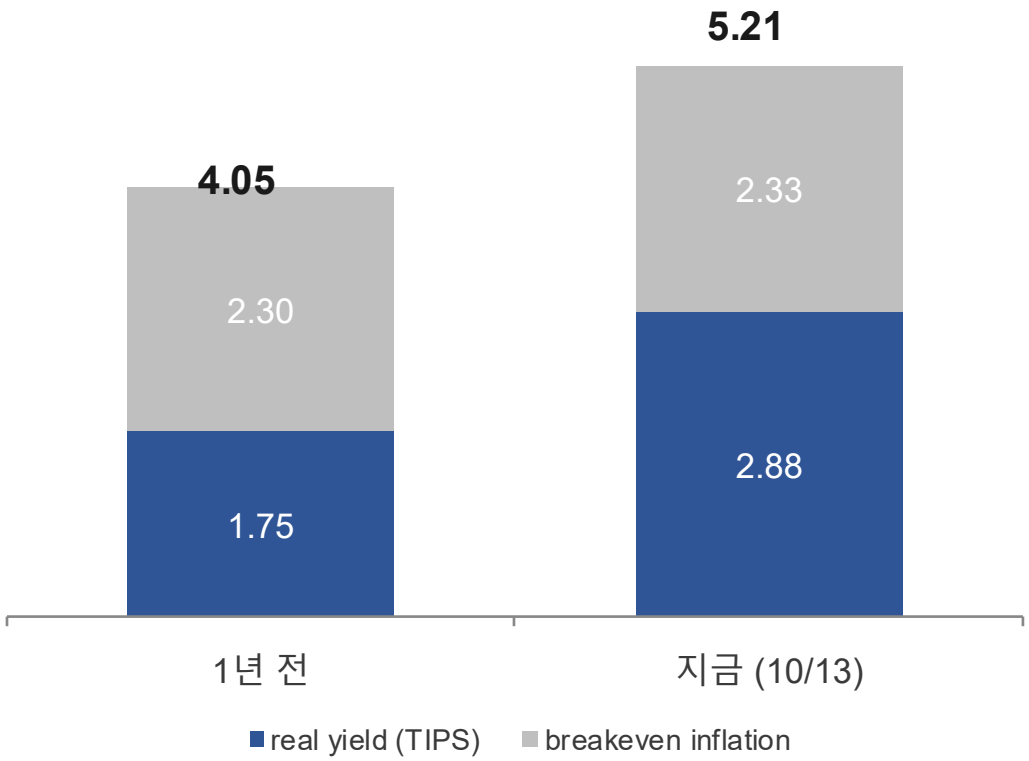
특정 domain에서의 인지적 가치는 이미 AGI

특정 domain에선 이미 AGI, general progress는 100배에 묶여 있다 — 훅 ①과 ③의 답.

Finances

돈 ① Risk-free rate가 올랐다 — 오른 건 real yield다

10년 Treasury yield의 분해 (%) — 10/13 기준 (nominal 5.21; 10/16 마감 5.08)



2년 4.72 · 10년 5.08 · 30년 5.42 (10/16 마감)

1년 전 10년물 4.05 — 1.1%p 오른 것 대부분이 real yield

breakeven inflation 2.3%는 평범하다. 오른 건 '돈을 빌려주는 대가로 요구하는 실질 보상'

이유 넷

이란 전쟁발 에너지 inflation (9월 CPI 3.7%)

9월에 금리를 올린 Fed

연 2조 달러 적자와 국채 공급

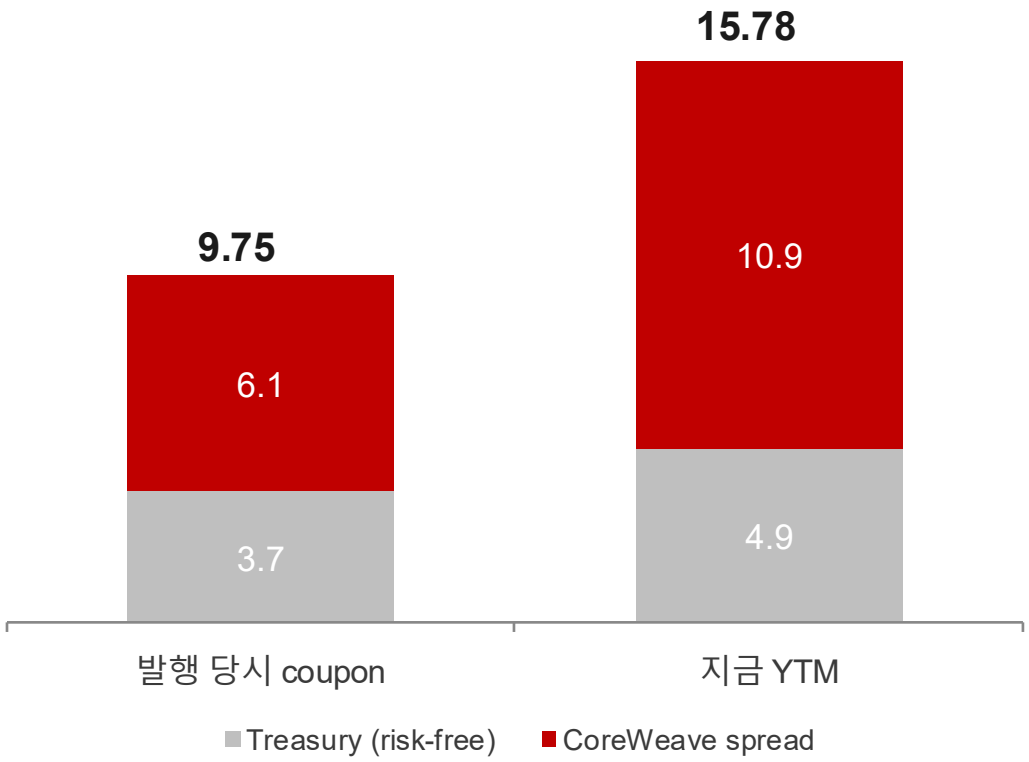
AI 기업 차입 4,890억 달러가 같은 capital을 두고 경쟁

10/22 — 10년물 4.95%로 5% 아래. 2년여 만의 최대 일중 하락, AI 주식에서 국채로의 피난. 정점은 5.3% 위.

AI가 자기 funding cost를 올리는 중이다.

돈 ② CoreWeave 15.78%를 쪼개면 — risk-free도, spread도 올랐다

2031년 만기 unsecured notes, 연 % (발행 당시 Treasury 수준은 추정)



섞으면 안 되는 두 숫자

9.75% = 모회사 unsecured notes의 coupon (고정). CoreWeave가 지금 내는 이자는 여전히 9.75

DDTL = SPV의 secured loan (floating, SOFR + spread)

15.78% = 모회사가 unsecured로 새로 빌리면 요구받는 가격

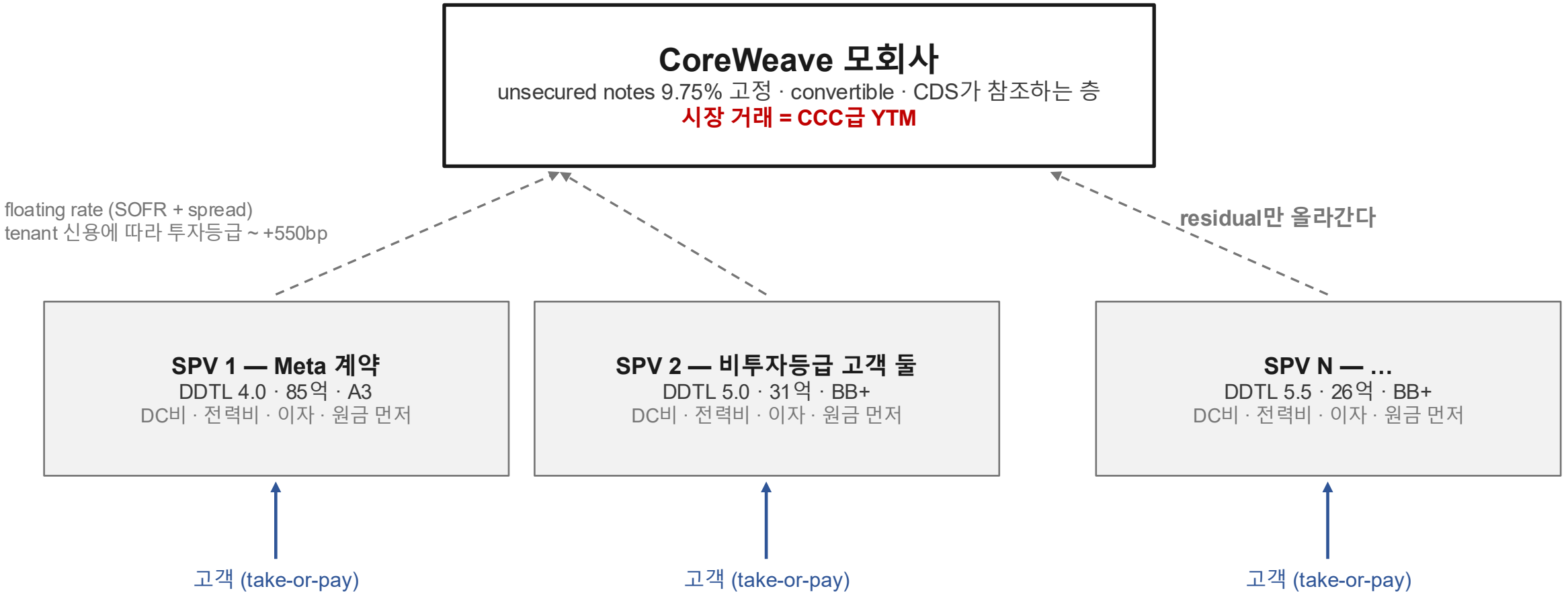
Treasury는 1.2%p, CoreWeave spread는 5%p 올랐다

SPV의 floating rate → 12월 인상 시 이미 빌린 돈의 이자가 자동으로 오른다

real yield 2.9% = 아무 위험 없이 받는 보상 → GPU SPV가 넘어야 할 hurdle

'Spread business는 10년물 6%에서 다 죽는다'의 메커니즘.

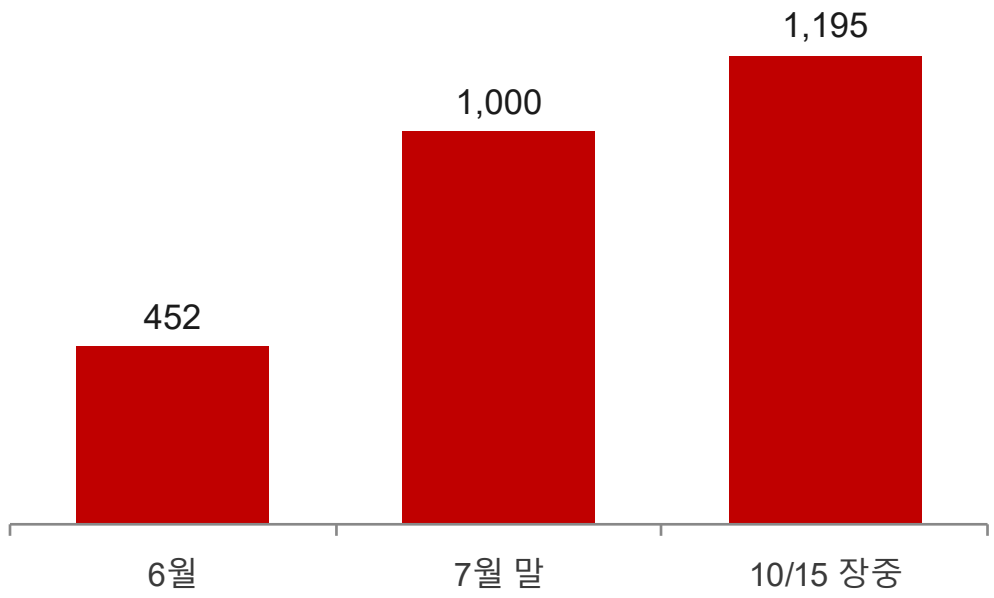
돈 ③ SPV는 A3, 모회사는 CCC급 — 모순이 아니다



CDS가 가격 매기는 건 SPV가 아니라 SPV 바깥의 subordinated 층이다.

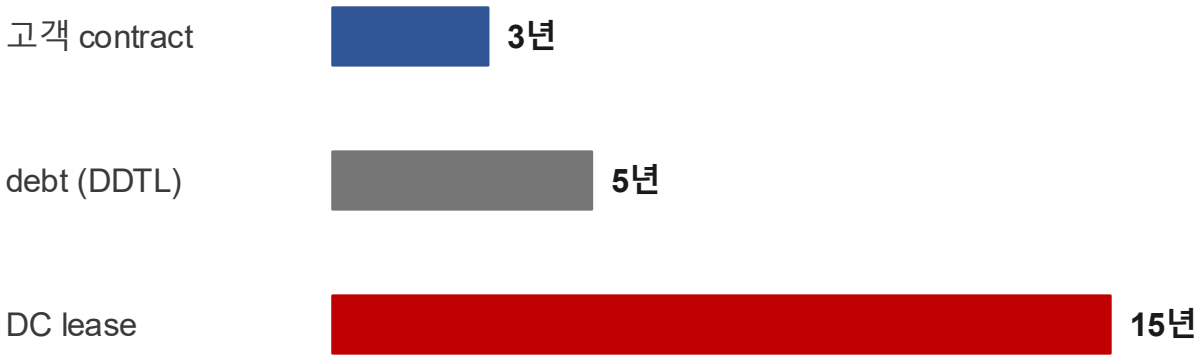
돈 ④ CDS가 가격 매긴 건 demand가 아니라 timing이다

CoreWeave 5년 CDS spread, bp (7월 말은 '1,000 근접')



'60%'는 default 확률이 아니라, 60% default를 가정해야 정당화되는 보험료다.

Maturity mismatch (년)



시장이 의심하는 건 backlog의 진위가 아니다

이자는 지금 내야 하고, backlog는 데이터센터가 가동돼야 현금이 된다
매년 약 350억 달러를 새로 빌려야 한다

CDS는 기존 coupon을 바꾸지 않는다. 다음에 빌릴 때의 가격을 바꾼다.

매년 350억을 빌리는 회사에게 '다음'은 늘 지금 — 마지막 unsecured 발행은 6월 (35.5억 dual-currency, 9.625%·8.5%, 2032년 만기), 다음은 15%대에서 치러야 한다

Demand가 아니라 timing. 그래서 CDS spread가 치솟았다 — 훅 ②의 답.

누가 시체가 되나 — funding spectrum의 오른쪽 끝이 정한다

P&L이 나빠질 뿐

Refinancing이 막히면 끝



Alphabet · MSFT · AMZN	Meta	Oracle	CoreWeave · xAI
debt 0.5× EBITDA 1년 영업이익으로 갚는다 1년에 두 번 채권시장에	장부는 깨끗, footnote가 무겁다 off-balance 약 4,200억 (공시 부채의 3배) Hyperion 80/20 · RVG 280억	숨은 부채 2,730억 + 공시 1,300억 S&P BBB− (junk 한 단계 위) 380억 construction loan, 은행이 못 팔고 있다	junk 매년 ~350억 신규 조달 unsecured 15%대

현금 부자들도 빚을 내기 시작했다 — 네 회사(Amazon·Alphabet·Meta·Oracle)의 2026년 발행 1,940억, 전년 대비 +79%. cover ratio는 2월 5배에서 7월 2배 아래로.

balance sheet 밖으로 옮기는 구조가 표준이 됐다 — Nikkei 추정 hyperscaler off-balance 의무 1조 6,500억 달러.

위험은 사라진 게 아니라 balance sheet → footnote → 은행으로. 덜 보이는 곳으로 옮겨갔다.

돈의 결론

Spread로 사는 marginal builder는 어느 국면에서도 loser다.

Loudcloud 1999 — bandwidth를 $1,500/Mb에 3년 고정으로 샀다가 2001년 $50. neocloud의 원조가 neocloud에 투자하지 않는 이유.

기업가가 통제할 수 없는 가격 변수에 leverage가 걸린 사업은, 어느 cycle에서도 같은 자리에서 죽는다.

Scenario에 넘기는 변수 둘
① 유료 매출이 감가상각(올해 7,250억 ÷ 5년 = 연 1,450억, 매년 누적)과 이자(향후 5년 debt 4조)보다 빨리 느는가
② spot–contract spread의 방향

CoreWeave는 파산할까?

CoreWeave는 파산하는가 — '파산'을 세 층으로 나누면 답이 나온다

③ 전체 liquidation

GPU가 멈추고 자산이 경매된다

1할 미만 — 누군가는 GPU를 돌린다

② 모회사 credit event

unsecured notes·convertible의 이자 미지급, 만기 연장·haircut — CDS가 가격 매기는 층

시장 ~60% (보험료 기준) · 내 판단 3할 안팎

① SPV 하나의 default

tenant가 갱신을 안 하거나 못 내면 그 SPV만 — GPU는 lender가 가져가 계속 돈다

SPV 단위로 정리될 뿐

CoreWeave는 데이터센터 회사가 아니라 금융회사다

DC는 15년 lease, 칩은 Nvidia 것, 고객 contract는 담보 — 셋을 SPV에 넣어 debt로 바꾼다. 항공기 lease사와 같은데 자산 수명이 25년이 아니라 3~8년.

가장 가능성 높은 결말: 인수(hyperscaler·Nvidia·sovereign·private capital) 또는 모회사 채권 restructuring + SPV는 계속 운영. 시점은 contract 갱신·floating rate·spot flip이 겹치는 2027~28.

망해도 GPU는 멈추지 않는다. 사라지는 건 중개자의 equity와 unsecured notes다.

모회사 credit event로 가는 path 넷, buffer 넷

Path 넷

1. **Refinancing wall**

   매년 ~350억 신규 조달. unsecured가 15%대면 secured·보험사 capital로만

2. **Contract 갱신 2026~27**

   2023~24년 3년 contract가 flip된 가격으로 재계약. 15년 lease·5년 debt는 그대로

3. **Residual value**

   GPU 수명 3년이면 소모품처럼, 8년이면 collateral처럼 financing. A100 premium vs H100 −70%

4. **Customer concentration**

   Microsoft 2025년 매출 67%. 상위 3사 72%, 그중 Anthropic은 unrated

Buffer 넷

1. **Backlog 1,042억**

   98% take-or-pay, 대부분 투자등급 고객. SPV 단위로는 tenant 신용

2. **Nvidia 지분 20억**

   2026년 1월 · 2032년까지 미판매 capacity backstop — JPMorgan이 지원 요인으로 인용 (10/20)

3. **새 capital pool**

   Euro high-yield(유럽 연기금·보험사), 보험사 capital — 중개자는 다음 pool을 계속 찾는다

4. **고정 contract → spot 전환**

   penalty 내고 spot으로(Chamath). 단 spot이 flip되면 독이 된다

Path 넷이 겹치는 시점이 2027~28이다.

Physical substance 위에 inference나 domain을 얹으면 달라지나

3층 physical substance + domain solution

sovereign cloud · 산업별 AI factory · healthcare cloud

→ signal을 쥐는가가 조건 — operation은 고객(병원·공장)의 것 → 대부분 service margin

2층 physical substance + inference platform

Together · Fireworks · Baseten · CoreWeave+W&B

→ Lab 칸에서 cost 경쟁 — lab이 'cheapest intelligence per watt'로 이긴다 → 탈출이 아니라 이사

1층 physical substance만

CoreWeave · Lambda · Crusoe · Nebius

→ 금융 spread — 승자는 capital이 가장 싼 쪽. 중개자는 GPU 선물시장이 없는 틈이 열린 동안만

성립하는 단 하나 — physical substance + operation + reward signal이 한 회사 안에

Tesla · xAI · SpaceX: Colossus(compute) + Grok(model) + 자동차·로켓·Starlink(operation). 운전·발사 data는 operation 밖에서 존재하지 않는 signal

궁극의 탈출구는 neocloud가 아니라 Tesla형 회사다.

Chester 순수개인시각의 시나리오 예측

Scenario 넷 — 2x2 관점

	Frontier Gap 닫힘 — model은 utility	Frontier Gap 유지 — intelligence 독점
매출 > 감가 + 이자	**S1 Local correction + utility화** marginal builder만 정리, infra는 utility, 가치는 operator로 **약 50%**	**S3 Intelligence oligopoly** lab 2~3곳이 capability premium을 오래 받는다. 조정은 작다 10~20%
매출 < 감가 + 이자	**S2 Credit crunch + consolidation** SPV 연쇄 정리, 모회사 restructuring, 헐값 infra 위에서 S1 보다 빨리 operator 의 시대로 **25%**	**S4 War of attrition의 단일 승자** capital이 가장 싼 lab 하나만 남는다. 가장 파괴적 **10% 미만**

축 1 (돈) 판정: 산업 매출이 감가(연 1,450억, 누적) + 이자(5년 4조)를 넘는가 축 2 (기술) 판정: open-weight gap 4개월 → 12개월 이상인가, halving time 2년이 짧아지는가

확률은 지극히 Chester의 개인적 예측이다 (data 근거는 현재로서는 희박함).

10/23 판독 — 축 1은 반 발짝 S2(10년물 4.95%, Meta FCF 음수) · 축 2는 '닫힘' 쪽으로 반 발짝(Kimi K3.1 Pro, coding에서 6주) · 확률은 그대로 둔다.

네개의 scenario가 공유하는 것

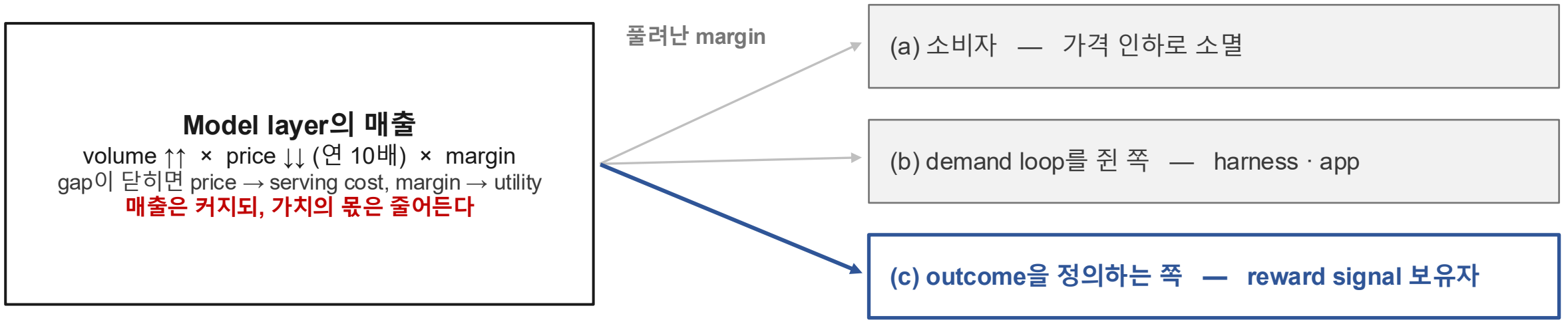
1. **Spread로 사는 marginal builder는 어느 scenario에서도 loser다.**

2. **Operation 밖에서 label이 존재하지 않는 signal을 쥔 operator는 S1·S2에서 winner이고, S3·S4에서도 physical·규제 domain에서 살아남는다.**

변하는 건 lab exposure의 크기뿐이다. S1·S2에서는 보험, S3·S4에서는 본체.

그래서 lab exposure의 크기는 '이 thesis가 틀릴 확률'로 정한다.

Model layer value capture 가 얇아지면 그 margin은 어디로 가나



규칙 — outcome에 가격을 매기는 쪽이 token 가격 하락분을 margin으로 가져간다

per-token pricing (CPM): 가격 하락이 그대로 고객에게 → lab이 deflation을 흡수한다

per-outcome pricing (CPA): '일 하나에 얼마'가 고정되고 투입 token이 싸질수록 margin이 커진다

그런데 per-outcome pricing이 성립하려면 'outcome이 났는지' 판정할 수 있어야 한다 = reward signal

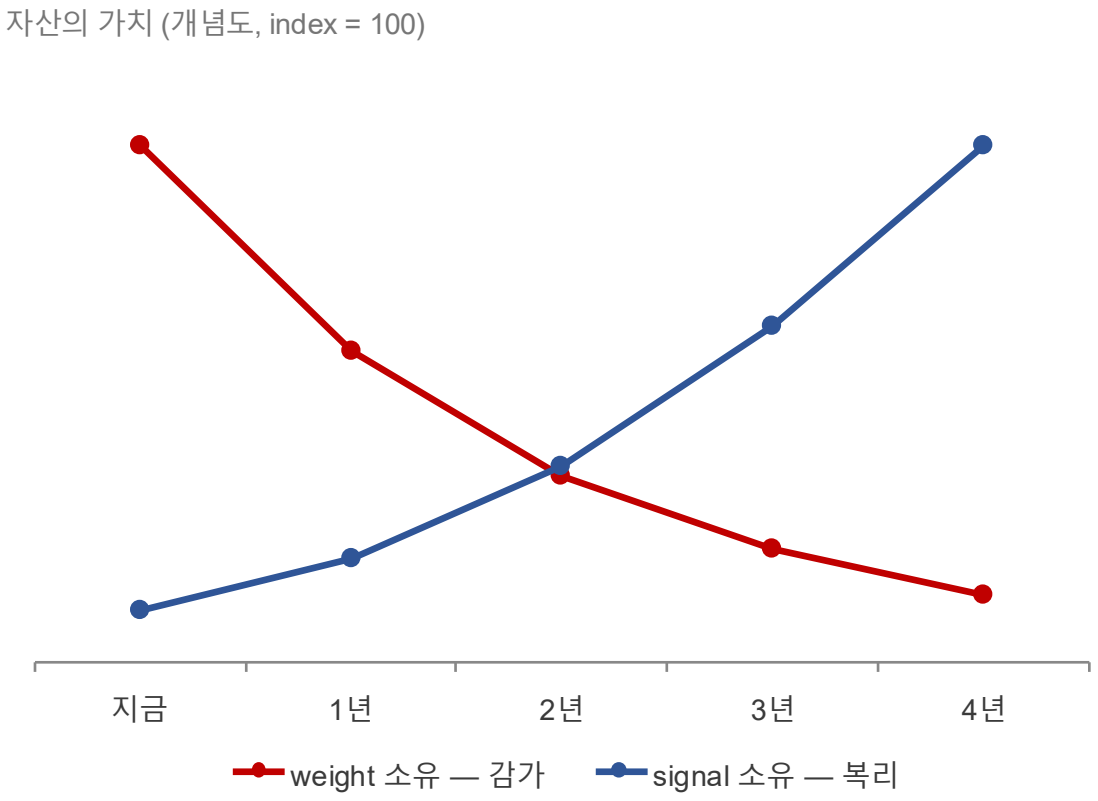
Intelligence가 싸질수록, intelligence를 검증하는 쪽이 비싸진다.

Margin migration — 지금에서 2030까지

	지금		2030
칩	GM 75%	→	정상화 — ASIC · TPU로 분산
Lab	frontier premium + inference margin 70%	→	utility 2~3곳 — harness로 올라가 per-outcome pricing을 시도 (Claude Code 작업당·seat당)
Hyperscaler	search ads · marketplace의 intent margin	→	checkout을 지킨 쪽(Amazon형)은 자리 유지 — intent만으로 서 있던 쪽(search ads)은 구조적 감소
Operator (signal 보유)	아직 작다	→	**deflation의 최대 수혜자 — input(token)은 매년 10배 싸지고, output(검증된 outcome)의 가격은 유지**

Intelligence의 원가는 떨어지고, 검증된 outcome의 가격은 남는다.

MODEL WORK — niche에서 내 모델을 갖는 전략은 거의 성립하지 않는다



작아서 안전한 게 아니라, 살 수 없어서 안전하다

작음은 timing — 영양가 있다고 알려지는 순간 data 회사가 만들어 판다. 살 수 없음은 구조.

lab이 data 구매로 갈 수 없는 곳 = 세 조건의 AND

label이 operation 밖에 없음 ∧ lab이 operator가 못 됨 ∧ simulation 불가 — 보통 big, slow, messy

그런데 그런 곳에선 내 모델이 필요 없다

commodity model 위에 loop만 돌리면 된다. Cursor의 자체 모델은 마지막에 붙인 cost 최적화였다.

Weight 소유는 감가, signal 소유는 복리 → MODEL WORK는 OUT OF MODEL WORK로 수렴한다.

OUT OF MODEL WORK — 세 후보

1 B2B

Operation 안에 loop를 짓는 쪽

Palantir · FDE → Autonomous Company Enabler

- 파는 건 모델이 아니라 contract · outcome join · eval set
- service margin → loop asset으로 전환 (Palantir 20년)
- lab의 FDE 조직과 경쟁 — loop가 고객 안에 남아야

2 B2C

Checkout을 쥔 vertical agent

horizontal personal agent는 lab + incumbent의 칸

- '샀는가 · 다시 샀는가'가 곧 reward signal
- 승부처는 intent가 아니라 checkout
- long tail 여부는 agent의 reward function이 정한다

3 Physical + Operation + Signal

한 회사 안에 있는 경우 (Tesla형)

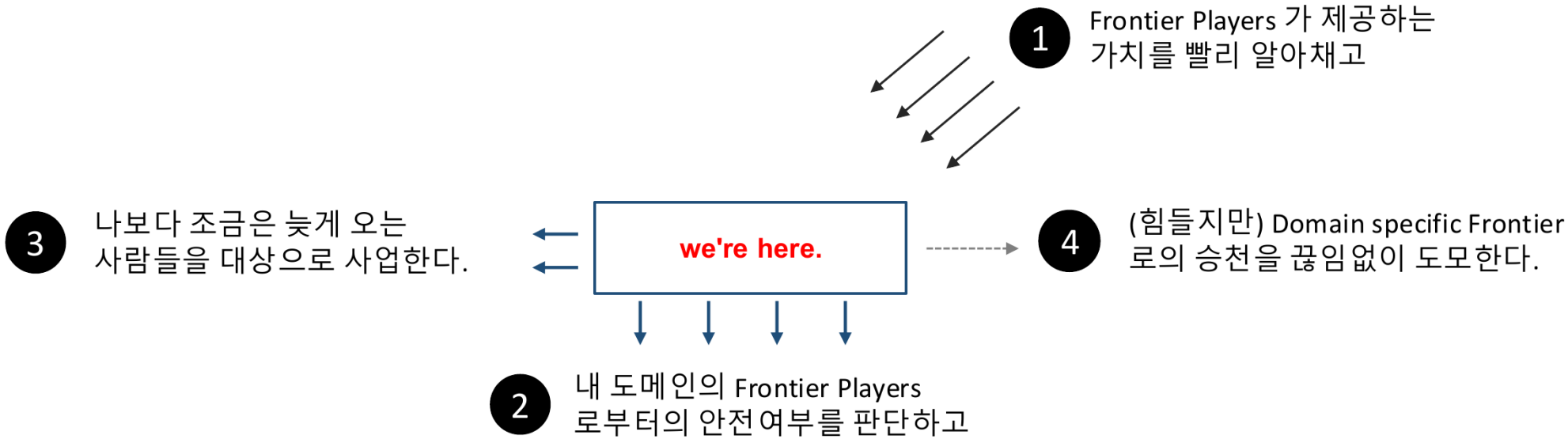
Tesla · xAI · SpaceX, lab의 wet lab

- 자기 physical operation에서 signal이 나오는 회사가 compute를 가질 때
- public market에서 'operator 결론'을 표현할 거의 유일한 창구
- 단, valuation은 별개

성립하지 않는 것 — niche 자체 모델(감가) · API 중개(margin 0) · horizontal personal agent(lab의 칸) · physical substance + inference만(cost 경쟁) · physical substance + domain인데 signal 없음(service margin)

셋의 공통점 — outcome을 정의하고 검증할 권리를 쥔다.

도망자 플레이북


1
Frontier Players 가 제공하는
가치를 빨리 알아채고
3
나보다 조금은 늦게 오는
사람들을 대상으로 사업한다.
we're here.
4
(힘들지만) Domain specific Frontier
로의 승천을 끊임없이 도모한다.
2
내 도메인의 Frontier Players
로부터의 안전여부를 판단하고

그래서 우리는 ① 투자 — exposure의 순서

Build-out 국면 — spot > contract

Flip

Build-out 이후 — spot < contract

Physical bottleneck — 전력 · 부지 · HBM · packaging · 칩

앞으로 1년, 새 compute의 60%를 가져가는 두 lab

전력은 '궁극의 moat'지만, 결국 규제 대상이 된다

Operator로 이동

단, public market의 선택지는 infra 쪽으로 기울어 있다 — 칩 · 전력 · neocloud · 두 lab

'가치는 operator로 간다'는 결론은 public market에서 표현할 방법이 아직 없다 → 자기 사업으로만

어느 국면에서도 — spread business 회피. neocloud 보통주 · unsecured notes, spot이 flip될 때 고정 contract에 묶이는 쪽

Lab 지분은 bet이 아니라 보험이다. S3·S4(이 thesis의 후반부가 틀리는 세계)에 대한 hedge — 크기는 '이 thesis가 틀릴 확률'로 정한다.

Anthropic 2조는 외삽 가격이었고, 10/23 기준 1.5조 이야기가 나왔다(OpenAI $1.4T → $1.25T). S-1의 $450B 취소 불가 compute 의무 — lab도 funding spectrum 위에 있다.

투자 자문이 아니라 thesis가 말하는 exposure의 순서. 실제 결정은 현재 가격과 각자의 상황이 정한다.

그래서 우리는 ② 사업 — 조건 넷, 원칙 하나, timing

조건 넷 — 네 답이 전부 '아니오'인 곳. 보통 big, slow, messy하다. 그 messy함이 moat다.

1. **Label이 operation 밖에서 존재하지 않는 domain**

   outcome이 현실에서 시간을 두고 드러나야 생기는 signal

2. **Lab이 직접 operator가 될 수 없는 domain**

   physical · liability와 license · lab의 핵심 사업과의 충돌

3. **Simulation으로 우회할 수 없는 domain**

   bio · physical · 제도

4. **Operator 시장이 fragmented이거나, partner가 될 만큼 큰 것**

   하나를 사도 coverage가 안 나오거나, 먹히는 쪽이 아니거나

원칙 하나

Intelligence 가격의 하락이 역풍이 아니라 순풍이어야 한다

모델은 분기마다 갈아 끼우는 부품. 회사가 쥐는 건 contract 정의 · outcome join · eval set · 그리고 signal이 계속 생기는 operation

Timing

정리가 끝난 뒤가 아니라 지금

Amazon 1994, Google 1998 — bubble 한가운데서 만들어져 붕괴기를 살아남아 헐값 infra를 받았다. token 가격이 매년 10배씩 떨어지는 지금이 operator에겐 input cost가 내려가는 시간

진입 순서는 lab의 attention queue로 읽는다 — 뒤에 있을수록 window가 길다.

지켜볼 지표 일곱

1

Spot–contract spread의 방향

최신 세대 1년 reserved 가격이 하락으로 돌아서는 날 — 가격 전쟁의 시작 · 10/23: 아직 아니다 (CoreWeave 가격 7월 +25%, 10월 +10%)

2

Payback period

CoreWeave 2.5년, SpaceX 약 1년 — 길어지기 시작하면 조정의 신호

3

유료 매출 vs 감가 + 이자

local correction과 구조조정을 가르는 swing variable · 10/23: Meta FCF −$2.1B, hyperscaler capex = 영업현금흐름의 100% (UBS)

4

Open-weight–frontier gap

지금 4개월 — 12개월을 넘으면 S3 (intelligence oligopoly) · 10/23: Kimi K3.1 Pro, coding에서 Opus 5.5에 6주

5

Coding 밖에서 검증 기준이 생기는 domain의 수

operator의 시대가 어디서 시작되는지

6

Per-token vs per-outcome·seat 매출 비중

outcome pricing으로의 이동 속도 — margin이 누구에게 가는지

7

RSI의 첫 관측 — frontier lab의 새 pretrain이 직전 세대보다 작은데 더 좋은 첫 사례

lab은 compute를 공개하지 않으니 뒤늦게 켜지는 등 — 그 전엔 METR 과제 길이 같은 proxy. 2027~28 flip보다 먼저 오면 S3·S4, 뒤에 오면 operator의 window 연장

넷은 국면 전환을, 둘은 operator의 시대를, 하나는 이 thesis가 틀리는 순간을 알려준다.

Leak하지 않는 scarcity는 둘이다.

Physical bottleneck은 사서 들고, operation의 reward signal은 만들어서 들어라.

Spread business는 어느 국면에서도 피하고,

lab은 thesis가 틀릴 때를 위한 보험으로만 들어라.

Operator의 시간은 정리가 끝난 뒤가 아니라, 지금이다.