돌핀(POD)

주요내용설명서(국문백서)

Korean White Paper

2026년 10월 6일

Disclaimer

본 번역본은 2026년 10월 6일 기준의 돌핀(Dolphin) 홈페이지의 관련 내용 위주로 번역됐습니다.

빗썸은 발행주체 또는 운영주체가 제공하는 가상자산의 총발행량, 유통량 계획, 사업 계획 등이 포함된 정보를 이용자들의 편의를 위해 참고용으로 제공하고 있습니다.

본 번역본은 그 내용이 정확하지 않을 수 있으며 원문의 내용이 일부 누락될 수 있으므로, 정확한 정보 습득을 위해서는 원문을 참고하시거나 원문 작성 측에 문의하시기를 바랍니다. 또한 본 번역본은 오픈 커뮤니티의 검토에 따라 내용이 변경될 수 있습니다.

빗썸

프로젝트 소개

돌핀(Dolphin)

돌핀은 AI 모델 개발 및 분산형 추론의 한계를 극복하고자 하는 프로젝트입니다.

현재 상황 및 성과

- 돌핀은 모든 베니스(100만 명 이상의 전체 사용자를 대상으로 시간당 약 60,000회의 프롬프트에 대응하고 있음) 사용자를 위한 무검열 모델을 개발했습니다. – 돌핀 미스트랄[1] 24B 베니스 버전
- 허깅 페이스[2] 내 돌핀 모델의 월간 다운로드 횟수가 500만 회를 돌파한 바 있습니다.
- 돌핀 추론 네트워크가 현재 서비스 중입니다. –https://x.com/dphnAI/status/1978158402431041657.

[1] Mistral

[2] Hugging face

비즈니스 모델

돌핀 AI 모델

돌핀의 모델은 사용자가 원하는 수준의 정렬[3]을 지원하기 위한 목적을 갖고 선구적인 오픈 소스 프런티어 모델[4]들을 토대로 학습되고 있습니다. 돌핀은 현재 일반에 공개된 최고 성능의 오픈 소스 AI 모델들의 무검열·무정렬·무편향 버전을 생성하기 위한 파이프라인을 구축하고 있습니다. 돌핀은 메타의 라마[5] 3, 미스트랄 및 믹스트랄[6] MoE[7], 구글의 젬마[8], 알리바바의 큐웬[9] 3 및 기타 유명 모델들의 자체 버전을 개발했습니다.

자체 버전 개발은 맞춤 데이터세트를 활용해 기반 모델을 수정하는 '파인 튜닝[10]' 과정을 통해 이루어집니다. 돌핀은 돌핀 모델을 위한 데이터세트를 구성할 때 응답 거절과 편향된 답변을 최대한 식별한 뒤 소거합니다. 그런 다음 필터링한 데이터세트를 활용해 기존 인스트럭트 모델[11]이 학습된 방식과 동일한 방식으로 모델을 학습시킴으로써 기반 모델의 더욱 '순응적인' 버전을 개발하여 사용자의 요청을 거부하지 않게 합니다.

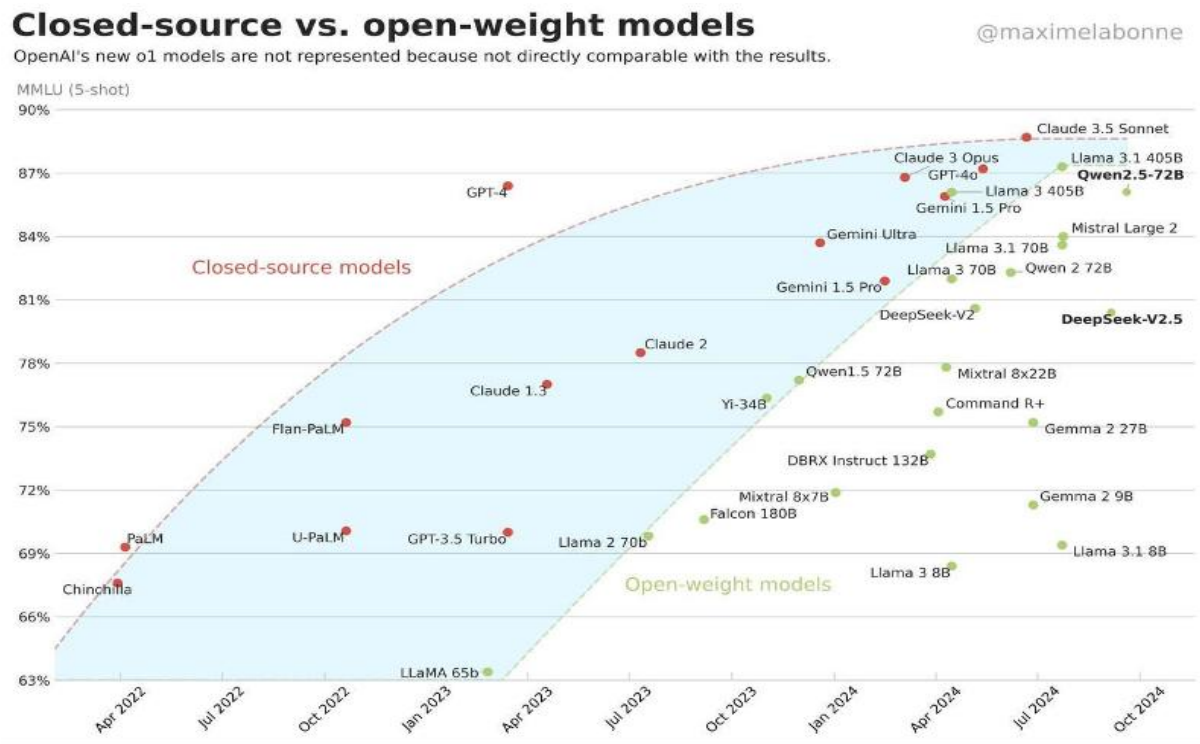
돌핀은 그러한 과정을 통해 기반 모델의 지능은 온전히 유지하면서도 돌핀의 코딩 및 수학 중심 데이터세트를 활용해 지식 베이스[12]를 확장하는 것이 목표입니다. 이를 통해 궁극적으로 사용자가 구상하는 거의 모든 프롬프트에 응답할 수 있는 모델을 개발하고자 합니다. 또한 돌핀은 흥미로움과 유익함을 모두 놓치지 않는 답변을 생성할 수 있도록 모델을 학습시키고 있습니다.

---

[3] Alignment
[4] Frontier model
[5] Llama
[6] Mixtral
[7] Mixture of Experts
[8] Gemma
[9] Qwen
[10] Fine tuning
[11] Instruct model
[12] Knowledge base

오픈 소스 모델과 비공개 모델의 격차 감소

구글조차도 그 불가피함을 인지하고 있습니다.



[오픈 소스 모델과 비공개 모델 비교]

상기 그래프는 오픈 소스 모델과 폐쇄형 모델의 성능을 시간 경과에 따라 비교한 그래프입니다.

폐쇄형 모델의 예시로 챗GPT, 클로드를 들 수 있으며, 이들 모델은 불투명한 비공개 모델로서 일부 플랫폼을 통해서만 접근할 수 있습니다. 오픈 소스 모델은 투명하며, 자택 컴퓨터를 활용해 로컬에서 작동하는 등 접근 방식도 다양합니다.

돌핀은 앞으로도 계속 오픈 소스 모델과 비공개 모델의 성능 격차는 감소할 것이라고 보며, 커뮤니티 주도의 혁신과 협업이 증대됨에 따라 오픈 소스 모델이 비공개 모델을 훨씬 능가할 수는 잠재력을 갖고 있다고 생각합니다.

돌핀은 무검열 및 조정 가능한 AI 모델 생성의 선두 주자로서 입지를 다지고자 합니다. 돌핀의 궁극적인 목표는 챗GPT같은 범용 챗봇 및 API(애플리케이션 프로그래밍 인터페이스)[13]를 만들되 정렬 기능을 내장시키지 않고 최대한 편향성을 제거하는 것입니다.

AI 검열의 대두

오픈 소스 모델 개발사는 대부분 챗GPT나 클로드가 생성한 데이터를 학습시키는데, 이는 다운스트림 모델[14]의 편향성과 검열 수준을 더욱 높입니다. 또한 오픈 소스 모델은 '안정성'을 고려해 편향성이 의도되었습니다. 답변 거부가 포함된 인스트럭트 데이터를 통해 조정되기 때문에 소위 '나쁜' 주제가 언급되면 모델이 답변을 거부하는 쪽으로 기울게 됩니다.

메타의 라마 1과 같은 초기 모델에는 그러한 편향성이 학습되지 않았습니다. 라마 2부터 기본 모델 또한 특정 요청에 대한 답변을 거부하도록 학습되기 시작했습니다. 돌핀은 라마 3에서 라마 4로 업그레이드되는 과정에서 이미 볼 수 있듯이(이전 버전보다 답변 거절 빈도가 눈에 띄게 상승함) 오픈 소스 모델 검열은 갈수록 증가할 것이라고 생각합니다.

이 같은 경향성은 주요 비공개 모델에서 더욱 명확하게 확인할 수 있습니다. 오픈AI의 새로운 GPT-5 모델은 이미 답변 거부율이 높았던 GPT-4보다 거부율이 훨씬 높아졌습니다. 이는 AI 개발의 진행 방향을 단적으로 보여줍니다.

대부분의 사람들은 선택권이 주어진다면 자신들의 AI에 엄격한 검열을 적용하지 않을 것입니다. 이러한 경향이 지속됨에 따라 돌핀의 주장은 더욱 힘을 얻고, 돌핀이 도달할 수 있는 전체 시장 규모는 재미 삼아 AI 모델을 사용하는 이부터 믿을 수 있는 무편향 AI 어시스턴트를 원하는 사용자까지 포괄하며 점점 확장되게 될 것입니다.

---

[13] Application programming interface

[14] Downstream model

돌핀 네트워크(Dolphin Network) – 분산 AI 추론 및 학습 네트워크

돌핀 네트워크는 전 세계 유휴 GPU(그래픽 처리 장치)[15]의 잠재력을 발현시킵니다.

돌핀 네트워크를 간략히 소개하면 아래와 같습니다.

- AI 추론 및 학습을 위한 보안과 프라이버시, 검증 가능성이 보장되는 분산 네트워크입니다.
- 유휴 GPU를 활용하여 현재 시장 가격보다 저렴한 수준으로 AI 연산 서비스를 제공합니다.
- 검열 저항 설계를 통해 중앙화된 호스팅 서비스 제공자의 기준에 구애받지 않는 다양한 AI 모델을 호스팅하고 학습시킬 수 있습니다.

돌핀 네트워크는 전 세계에 배치된 기계를 통해 작동되는 이기종[16] AI 모델을 지원하고, 조작 방지를 위한 강력한 검증을 수단을 제공하며, 중앙화된 서비스 제공자보다 저렴한 수준으로 응답이 일정하게 생성될 수 있도록 합니다. 돌핀 네트워크의 목표는 게이머나 데이터 센터가 돌핀 추론 네트워크에 참여함으로써 GPU 미사용 기간 동안 유휴 장치를 활용할 수 있게 지원하는 것입니다.

참여자는 네트워크 요청 처리에 기여하면서 돌핀(POD) 토큰을 획득할 수 있으며, 이후 POD를 통해 추론을 실행하거나 토큰을 2차 시장에 매도하는 등 GPU 비용을 회수하는 데 이용할 수도 있습니다.

돌핀은 AI와 디핀[17]의 이상적인 결합 사례는 바로 분산형 추론이라고 생각하며, 그 이유는 아래와 같습니다.

1 – AI 수요

AI 추론 사용량은 매우 높으며 해마다 우상향하고 있습니다.

글로벌 AI 추론 시장의 규모는 생성형 AI 및 거대 언어 모델(LLM)[18] 사용에 힘입어 2025년 약 1,060억 달러에서 2030년 약 2,550억 달러까지 증가할 것으로 예상됩니다(연평균 성장률 약 19%). 또한 AI 학습량이 큰 영향을 미치는

---

[15] Graphics processing unit

[16] Heterogeneous

[17] DePIN (decentralized physical infrastructure network)

[18] Large language model

데이터 센터 GPIU 시장의 규모는 하이퍼스케일러[19] 및 다수의 기업이 모델 학습 증대를 위해 GPU클러스터를 점차 확장함에 따라 2025년 약 1,200억 달러에서 2030년 약 2,280억 달러까지 증가할 것으로 예상됩니다(연평균 성장률 약 14%).

(데이터 출처: AI 추론 시장 – 마켓츠앤마켓츠[20] | 데이터 센터 GPU 시장 – 마켓츠앤마켓츠)

2 – 소비자용 GPU 공급, 확장 및 공유 리소스화

설치된 게이밍 PC의 수는 전 세계 약 20억 개로 추정되는데, 이는 디핀 프로젝트가 접근할 수 있는 대규모 하드웨어 풀 중 하나입니다. 게이밍 GPU는 AI 추론과 학습을 처리하기에 꽤 적합하지만 게임을 실행할 때만 사용되기 때문에 설치 이후 95%가 넘는 시간 동안 유휴 상태로 남아 있습니다.

돌핀 네트워크는 피어 투 피어[21] 프로토콜을 통해 소비자용 GPU를 추론 및 학습을 위한 공유 하드웨어 풀로 통합함으로써 유휴 자원을 활용할 수 있도록 설계되었습니다. AI와 결합된 대부분의 디핀 네트워크가 노드[22]와 개별 사용자를 1:1로 연결하는 것과는 달리(엄격한 가동 시간 및 장기적인 서비스 제공 요건 또한 부과됨) 돌핀의 노드는 참여한 뒤 언제든 불이익 없이 그만둘 수 있습니다. 게이머는 게임을 하지 않는 동안 돌핀의 노드 소프트웨어를 실행한 다음 원할 때 중지할 수 있으며, 소프트웨어를 가동하더라도 운영체제는 계속 사용할 수 있습니다. 돌핀은 일부 GPU만을 사용하며 리눅스[23]까지 요구하지 않습니다.

핵심은 이러한 유휴 GPU가 중소형 모델에 중앙화 서버와 다름 없는 추론 속도와 품질을 제공할 수 있다는 점입니다. 이를 테면 RTX 4090으로 가동되는 라마 3.1 8B가 vLLM을 기반으로 363개의 요청을 동시에 배치[24] 처리했을 때 초당 토큰(데이터 단위) 생성 속도는 약 6,200입니다. 기업용과 비교해보자면 H100 하나로는 약 2배의 동시 요청을 배치 처리할 수 있었고 608개의 동시 요청을 처리하는 동안 약 11,300의 초당 토큰 생성 속도를 기록했습니다.

---

19 Hyperscaler

20 Markets and Markets

21 Peer to peer

22 Node

23 Linux

24 Batch

GPU 구매 가격을 고려했을 때 소비자용 GPU의 가격 대비 성능은 약 10배 높다고 할 수 있으며, 이는 소비자용 GPU의 전력 소비량이 낮다는 점을 아직 반영하지 않은 수치입니다. 대규모 데이터 센터의 토큰 생성 비용이 가장 낮을 것이라는 기존의 '규모의 경제' 관점과는 상반되는 결과입니다.

핵심적인 원인은 시장 분리입니다. 엔비디아는 게이머를 비롯한 최종 사용자가 소비자 가격으로 GPU를 계속 구매할 수 있도록 데이터 센터가 소비자용 GPU를 대량으로 구매하는 것을 허용하지 않으며, 또한 소비자용 GPU를 다중 GPU 서버 설정에 통합하는 것을 까다롭게 만들고 있습니다. 돌핀 네트워크는 전 세계에 분산된 유휴 소비자용 GPU를 신뢰할 수 있는 무허가[25] AI 연산 클러스터로 합산하여 개발자와 사용자가 낮은 비용으로 인기 있는 오픈 소스 모델을 활용할 수 있도록 지원함으로써 이 같은 공백을 해소하고자 합니다.

3 – 인접성

기존 디핀 네트워크에서 인접성은 가치를 결정하는 요인입니다. 노드가 인접한 사용자에 한해 서비스를 제공할 수 있기 때문입니다.

그와 대조적으로 AI 추론은 지역에 크게 국한되지 않습니다. 몇 백 밀리초의 지연 시간이 발생하더라도 사용성에는 큰 차이가 없습니다. 따라서 돌핀 네트워크는 사용자에게 전 세계의 연산 리소스를 연결함으로써 확장성과 개별 노드의 사용률을 대폭 증가시킬 수 있습니다.

토크노믹스

돌핀의 네이티브 토큰[26] POD는 분산 추론 및 학습 네트워크에 적합하다고 판단한 여러 프로젝트의 우수한 요소를 참고하여 설계되었습니다.

- 이더리움(ETH) 방식: 노드 운영자 및 검증인이 예치한 보증금에 삭감[27] 적용

---

[25] Permissionless

[26] Native token

[27] Slashing

- 커브(CRV) 방식: 노드 운영자에게 보상 부스트 제공(스테이킹된 POD(xPOD) 본딩(예치)[28] 시 최대 1.3배 부스트 적용, 3개월 본딩 시 최대 1.2배까지 선형적으로 증가, 1.2배에서 1.3배 구간은 다른 노드 운영자의 본딩 현황에 따라 경쟁 방식으로 결정, 최대 부스트를 위해서는 6개월 본딩 필요)
- 스테이킹된 스시스왑(xSUSHI), 연 파이낸스 볼트[29]에 예치된 CRV(yCRV) 및 YFI(yYFI) 방식: 자동 복리 스테이킹(보상을 별도로 청구할 필요가 없어 xPOD를 노드 보상 강화를 위한 보증금으로 예치 가능, 보상 청구가 필요치 않고 모든 가치 상승이 xPOD 포지션에 반영되므로 수수료 측면에서 효율적)
- 스테이킹된 에이브(stkAAVE) 방식: 스테이킹 및 본딩에 3개월 대기 기간 적용 + stkAAVE 및 투표 락업[30] 컨벡스 파이낸스[31](vlCVX) 방식의 일정한 보상 제공

POD는 xSUSHI, yCRV 및 yYFI와 유사한 전송 가능한 자동 복리 볼트에 스테이킹할 수 있습니다.

보상이 분배되더라도 지갑 내 xPOD 잔고는 그대로 유지됩니다. POD 보상이 컨트랙트로 전송되면 상환 비율이 상승하게 됩니다.

예시와 함께 살펴보겠습니다.

- 첫날에 10 POD를 스테이킹합니다.
- 10%의 보상이 추가됩니다.
- 상환 비율이 1.1로 상승합니다.
- 스테이킹된 포지션을 인출하면 11 POD를 수령하게 됩니다.

투자자 또는 추론 서비스 사용자는 xPOD 보유 시 아래와 같은 혜택을 누릴 수 있습니다.

- POD로 지급되는 네트워크 수익의 일부 또는 전부(자동 복리 방식)
- 비용을 직접 지불할 필요 없이 다양한 단계의 제품 구독 서비스 이용
- 일일 추론 크레딧 무료 제공

---

[28] Bonding

[29] Vault

[30] Vote-locked (vl)

[31] Convex Finance (CVX)

xPOD에는 거버넌스 권한과 일일 추론 할당량을 모두 위임할 수 있는 기능이 있습니다. 따라서 CRV 및 CVX 방식의 투표권 거래 시장이 형성되어 일일 추론 할당량을 사용하지 않는 스테이킹 사용자로부터 위임을 받기 위한 입찰이 진행될 수 있습니다. 할당량을 사용하지 않는 사용자는 추가 수익을 얻고, 입찰을 통해 할당량을 위임받은 사용자는 시장 가격보다 저렴하게 추론 서비스를 이용할 수 있습니다.

vlCVX의 동일 권한[32] 메커니즘은 CVX를 16주 동안 락업했을 때(vlCVX) 부여되는 투표권 및 스테이킹 APY(연간 복리 수익률)[33]이 해당 락업 기간 내내 동일하게 유지되는 컨벡스 파이낸스의 메커니즘입니다. 이는 시간 경과에 따라 투표권이 감소하여 보유자가 계속해서 락업 기간을 연장해야 하고 매몰 비용이 발생할 수 있는 커브 파이낸스 및 기타 투표 에스크로 토큰[34]과는 다릅니다.

이러한 방식은 자산 매수와 스테이킹에 대한 매력도가 더 크기 때문에 기존의 ve토큰 기반 토크노믹스보다 낫다고 많은 이들에 의해 평가됩니다. 돌핀은 vlCVX의 장점을 채택하면서도(반복적인 락업 없이도 APY가 온전하게 유지됨) 모든 스테이킹이 16주마다 언락[35]되게 하는 대신 사용자의 스테이킹 상태를 기본적으로 유지시킵니다.

토큰 할당

총발행량: 500,000,000 POD

항목	수량	비율
트레저리[36]	214,634,419개	42.93%
팀	125,000,000 개	25.00%
시드 라운드	117,600,000 개	23.52%
유니스왑 v4 풀	42,765,581 개	8.55%

[32] Constant power
[33] Annual percentage yield
[34] Vote-escrowed (ve) token
[35] Unlock
[36] Treasury

팀 할당분 베스팅[37]

팀 할당분 125,000,000 POD(총발행량의 25%)는 기할당된 물량과 향후 팀원을 위한 예비 물량으로 구분됩니다.

- 기할당된 87,500,000 xPOD: 팀 할당 볼트로 이동되는 85,000,000 xPOD 스트림[38]과, 각각 250,000 xPOD씩 이동되는 개별 스트림 10개로 구성됩니다.
- 향후 팀원 예비용 37,500,000 POD: 동일한 볼트로 전송되는 별도의 베스팅 스트림에 보관되며, 아직 개별 직원에게 할당되지 않았습니다.

두 개의 대규모 스트림으로 개별 직원에 대한 토큰 할당 및 분배를 관리하는 돌핀 팀 할당 볼트에 토큰이 공급됩니다. 볼트를 통해 직원 할당분이 xPOD로 지급되며, 예비 물량 스트림에서 언락된 POD를 스테이킹하여 지급 재원으로 사용될 수 있습니다.

볼트 토큰 공급 스트림

세부 사항	기할당분	예비 물량
세이블리어[39] 스트림	85,000,000 xPOD — #1186	37,500,000 POD — #1185
수신 주소	팀 할당 볼트	동일 볼트
시작일	2026년 9월 20일	2026년 9월 20일
클리프[40] 종료일	2028년 5월 1일	2028년 5월 1일
클리프 종료 시 언락 물량(1%)	850,000 xPOD	375,000 POD
선형 베스팅 물량(99%)	84,150,000 xPOD	37,125,000 POD
베스팅 완료일	2038년 5월 1일	2038년 5월 1일
취소 가능 여부	취소 불가	취소 불가
스트림 NFT[41] 전송 가능 여부	전송 불가	전송 불가

---

[37] Vesting

[38] Stream

[39] Sablier

[40] Cliff

[41] Non-fungible token

2028년 5월 1일 이전까지 두 스트림의 어떠한 물량도 언락되지 않습니다. 클리프가 종료되면 1%가 언락되며, 나머지 99%는 이후 10년에 걸쳐 선형 베스팅되어 2038년 5월 1일에 베스팅이 완료됩니다. 계획된 일정은 UTC 기준 해당일의 01:00(AEST 기준 오전 11:00)에 진행됩니다.

개별 스트림

각 250,000 xPOD로 구성된 스트림 10개의 총수량은 2,500,000 xPOD이며, 해당 물량은 개별 수신자에게 기배정된 상태를 유지합니다. 기존 스트림에는 최초 일정이 그대로 적용됩니다. 스트림은 2026년 7월 4일에 시작되며, 2028년 5월 4일 클리프가 종료될 때 각 스트림에서 2,500 xPOD(1%)가 언락됩니다. 나머지 247,500 xPOD(99%)는 2038년 5월 4일까지 선형 베스팅됩니다. 계획된 일정은 UTC 기준 해당일의 12:03:51에 진행됩니다. 개별 스트림은 취소할 수 없으며, 스트림 NFT의 경우 전송 불가능합니다.

직원 지급분 및 보호 장치

상기 언급된 세이블리어 스트림 12개는 모두 취소할 수 없습니다. 토큰 공급을 담당하는 세이프 지갑은 스트림을 취소하여 베스팅되지 않은 자금을 회수할 수 없습니다. 전송 불가 설정은 인출되는 토큰이 아니라 스트림 NFT의 소유권에 적용됩니다.

볼트 내 개별 직원 할당분에는 규칙이 개별적으로 적용됩니다. 특정 팀원이 더 이상 프로젝트에 기여하지 않는 경우 볼트를 관리하는 트레저리 멀티시그[42] 7명 중 4명 이상이 찬성하면 해당 직원에 대한 토큰 할당을 취소할 수 있습니다. 다만 이미 획득한 권리는 유지되며, 아직 베스팅되지 않은 수량은 향후 팀원을 위한 할당분으로 사용됩니다. 이러한 조치로 인해 기본적인 세이블리어 스트림이 취소되거나 해당 스트림에 락업된 토큰 수량이 트레저리로 반환되지는 않습니다. 볼트는 볼트 관리자에 의해 업그레이드될 수 있으며, 직원에 대한 토큰 할당 정책은 세이블리어의 스트림 제한 규칙과 별개입니다.

기배정분과 직원 지급분은 스테이킹된 토큰 형태(xPOD)로 제공됩니다. 스테이킹 보상은 자동 복리 방식의 xPOD 볼트 영수증 토큰에 계속 반영되며, 베스팅 기간에는 별도로 청구할 수 없습니다. 아직 할당되지 않은 예비 물량 스트림은 POD로 표시됩니다.

일부 계약직 팀원에게는 더 짧은 비스통 일정이 적용되어 급여 보상의 일부로 지급된 바 있습니다. 이와 관련된 개별 스트림은 링크1, 링크2, 링크3에서 확인할 수 있습니다.

---

[42] Multisig

시드 라운드 및 트레저리

돌핀은 2024년 6월 20일 117,600,000 POD에 대해서 886,000달러 규모의 시드 라운드를 진행했으며, 해당 물량에 1년의 클리프 기간을 적용했습니다. 모든 시드 라운드 물량은 2025년 6월 20일부로 전량 언락된 상태입니다.

2025년에는 장외 거래[43] 3건을 통해 4,250,000 POD가 판매되었습니다. 장외 거래 물량에는 1년의 클리프와 2년의 선형 베스팅이 적용되었으며, 스테이킹된 토큰 형태로 발행되었습니다.

트레저리 할당분 및 유동성 포지션은 4/7 노시스 세이프 멀티시그 지갑에 보관 중입니다 서명자 7명은 팀원 6명과 돌핀 재단으로 구성되어 있으며, 트랜잭션 승인에는 7명 중 4명의 서명이 필요합니다. 트레저리 할당분은 노드 보상 지급 또는 향후 자금 조달에 한해 사용됩니다. 돌핀은 트레저리 토큰 판매를 총발행량의 10%로 제한하고 최소 1년의 클리프 기간을 적용하는 정책을 도입했습니다. 트레저리 토큰 판매로 조달된 투자금은 4/7 세이프 멀티시그 지갑에서 관리됩니다.

노드 보상은 노드 운영자가 처리한 추론 작업량만을 기준으로 지급됩니다. 이를 통해 추론 수요가 없는 상황에서 보상이 지나치게 지급되는 것을 방지합니다.

출금 대기

현재 스테이킹과 본딩에는 모두 stkAAVE 방식의 대기 기간 3개월과 출금 기간 1주가 적용됩니다. 본딩 컨트랙트는 대기 기간이 이중 적용되지 않도록 화이트리스트 처리되어 있기 때문에, 본딩된 자산의 경우 POD로 바로 출금할 수 있습니다.

출금 대기 중인 컨트랙트는 https://exitqueue.dphn.ai에서 실시간으로 확인할 수 있습니다.

---

[43] Over-the-counter (OTC)

노드 보상 분배

노드 보상은 POD 보상 분배 2/6 멀티시그 지갑에서 지급됩니다. 서명자 6명은 팀원들로 구성되어 있으며, 트랜잭션 승인에 2명의 서명이 필요합니다. 노드 보상 분배를 위한 멀티시그 지갑은 노드 운영자에게 지급되는 모든 POD 보상을 분배하는 데 사용되며, 트레저리를 통해 정기적으로 POD가 보충됩니다.

보안을 위해 해당 멀티시그 지갑에는 1,000,000 POD를 초과하는 수량을 보관하지 않습니다.

네트워크 대시보드에서는 노드 운영자에 대한 보상 지급 현황 및 과거 기록, 일평균 및 주평균 보상 분배량을 확인할 수 있습니다.

보상 분배량은 네트워크의 일간 전체 추론 작업량으로 결정됩니다.

로드맵

돌핀 네트워크 로드맵

- **추론 네트워크 V1**: 출시, RTX 3090/4090/5090 등의 소비자용 GPU 대상
- **추론 네트워크 V2**: 출시, 기업용 GPU 처리량을 위한 배치 처리 기능, 사용자가 돌핀 노드를 기반으로 실행되는 모델을 자신의 앱에 통합할 수 있도록 공개 API 제공
- **이미지 및 동영상 생성**: 출시 예정, 인기 모델을 저렴한 비용으로 이용할 수 있도록 지원
- **오디오 생성**: 출시 예정, 일레븐랩스[44] 대비 비용 대폭 절감, 다양한 최첨단 오디오 모델 제공
- **샤딩[45] 추론**: 출시 예정, 하나의 대규모 모델을 여러 소비자용 GPU 노드로 분할하여 실행

---

[44] ElevenLabs

[45] Sharding

- **분산형 LoRA[46] 학습**: 출시 예정, 소비자용 및 기업용 GPU에 적용, LoRA는 통신 효율이 매우 높은 모델 학습 방식으로 돌핀의 분산 학습 구현을 위한 합리적인 첫 단계임, 최근 블로그 게시물에서 단일 노드를 기반으로 메타의 최대 AI 모델인 라마 3.1 405B를 LoRA 학습을 통해 파인 튜닝한 방법을 설명한 바 있음

[46] Low rank adaptation