



OpenGradient

기술 백서

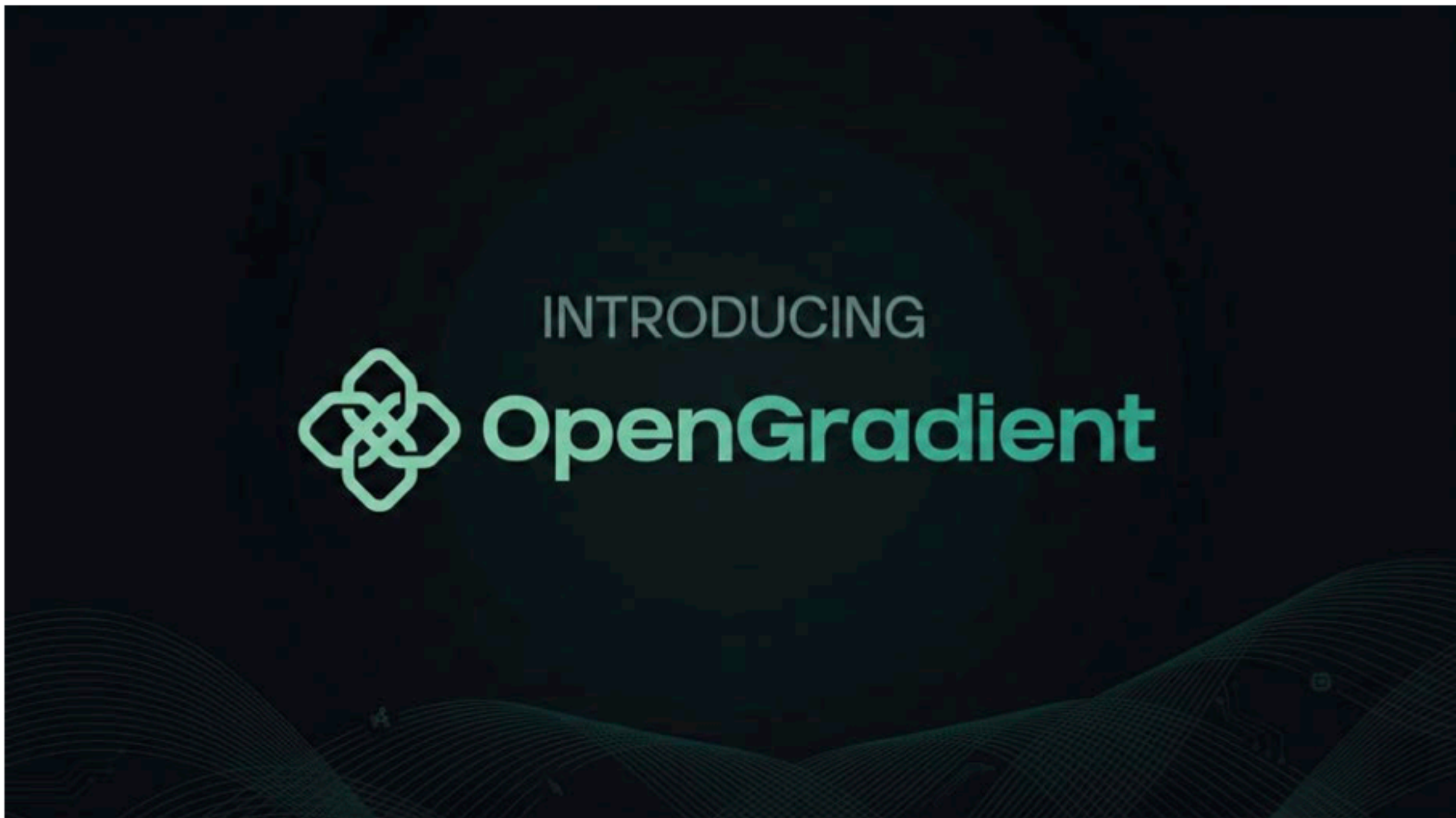
Technical Whitepaper

개방형 인텔리전스를 위한 네트워크

The Network for Open Intelligence

버전 1.0 | 2026년 3월

<https://opengradient.ai> | <https://docs.opengradient.ai>



목차 (Table of Contents)

1. 개요 (Overview)
2. 문제 정의 (Problem Statement)
3. 하이브리드 AI 컴퓨팅 아키텍처 (HACA)
4. 노드 유형 (Node Types)
5. 검증 및 정산 (Verification & Settlement)
6. 검증 가능한 LLM 실행 (Verifiable LLM Execution)
7. ML 실행 — PIPE
8. 합의 메커니즘 (Consensus)
9. 제품 생태계 (Product Ecosystem)
10. 사용 사례 (Use Cases)
11. 네트워크 현황 (Network Status)
12. 용어 사전 (Glossary)

1. 개요 (Overview)

오늘날 모든 AI 애플리케이션은 단일 신뢰 지점(single point of trust)에 의존합니다. AI 에이전트가 포트폴리오를 관리하거나, 대출을 승인하거나, 콘텐츠를 중재할 때, 어떤 모델이 실행되었는지, 어떤 프롬프트가 사용되었는지, 출력이 변조되었는지를 독립적으로 검증할 방법이 없습니다.

OpenGradient는 이 문제를 해결합니다. AI 추론(inference)을 위해 특별히 설계된 탈중앙화 네트워크로, 모든 연산을 단일 당사자를 신뢰하지 않고도 암호학적으로 검증할 수 있습니다. 모델은 특수 노드의 무허가(permissionless) 네트워크에서 실행되고, 증명(proofs)은 온체인에 정산되며, 요청부터 응답까지의 전체 파이프라인이 감사 가능(auditable)합니다.



그림 1: OpenGradient 제품 스택

OpenGradient의 차별점

OpenGradient는 기존 AI API의 래퍼(wrapper)가 아닙니다. 전용 블록체인부터 특수 컴퓨팅 노드까지, 수직 통합된 인프라 스택으로, 하나의 원칙을 중심으로 설계되었습니다: AI 추론은 기본적으로 검증 가능해야 합니다.

핵심 통찰은 AI 워크로드가 금융 거래와 근본적으로 다른 요구사항을 가진다는 것입니다. OpenGradient는 실행(execution)과 검증(verification)을 분리하는 하이브리드 AI 컴퓨팅 아키텍처(HACA)로 이 문제를 해결합니다.

주요 지표

지표	수치	설명
모델 (Models)	2,000+	세계 최대 탈중앙화 AI 모델 저장소
검증 가능한 AI 추론	2,000,000+	TEE 및 ZKML 검증된 추론 횟수

ZKML 증명 + TEE 증명서	500,000+	암호학적 증명 및 하드웨어 증명서
EVM 호환성	100%	Ethereum 도구 및 생태계 완전 호환

주요 투자자

a16z crypto (리드), Coinbase Ventures, SVA (SV Angel), NVIDIA, Foresight Ventures, Salt Fund, Symbolic Capital, Canonical Crypto, Black Dragon, Celestia, NEAR, Thanefield Capital

주요 엔젤 투자자: Balaji Srinivasan (전 Coinbase CTO), Illia Polosukhin (Transformer 아키텍처 공동 발명자, NEAR 공동 창업자), Sandeep Nailwal (Polygon CEO)

2. 문제 정의 (Problem Statement)

AI 인프라가 소수의 제공업체로 집중되고 있습니다. 이로 인해 실질적인 문제가 발생합니다:

불투명성 (Opacity)

LLM이 자금, 건강 또는 거버넌스에 영향을 미치는 결정을 내릴 때, 블랙박스 내부에서 무슨 일이 일어났는지 증명할 방법이 없습니다.

단일 장애 지점 (Single Points of Failure)

제공업체가 다운되거나, 사용량을 제한하거나, 모델 동작을 변경하면 애플리케이션이 중단됩니다. 대안도 없고 구제 방법도 없습니다.

검증 없는 신뢰 (Trust Without Verification)

에이전트 운영자나 API가 모델을 은밀히 교체하거나, 콘텐츠를 주입하거나, 프롬프트를 기록할 수 있습니다. 사용자는 이를 맹목적으로 수용해야 합니다. 정확성이 중요한 애플리케이션에서 맹목적 신뢰는 충분하지 않습니다.

OpenGradient는 반대 접근법을 취합니다: AI를 기존 블록체인 모델에 맞추는 대신, AI의 실제 요구사항에 맞는 블록체인 모델을 구축합니다.

3. 하이브리드 AI 컴퓨팅 아키텍처 (HACA)

OpenGradient는 HACA(Hybrid AI Compute Architecture)로 운영됩니다. AI 워크로드를 금융 거래와 동일한 방식으로 처리할 수 없다는 관찰에 기반한 네트워크 설계입니다.

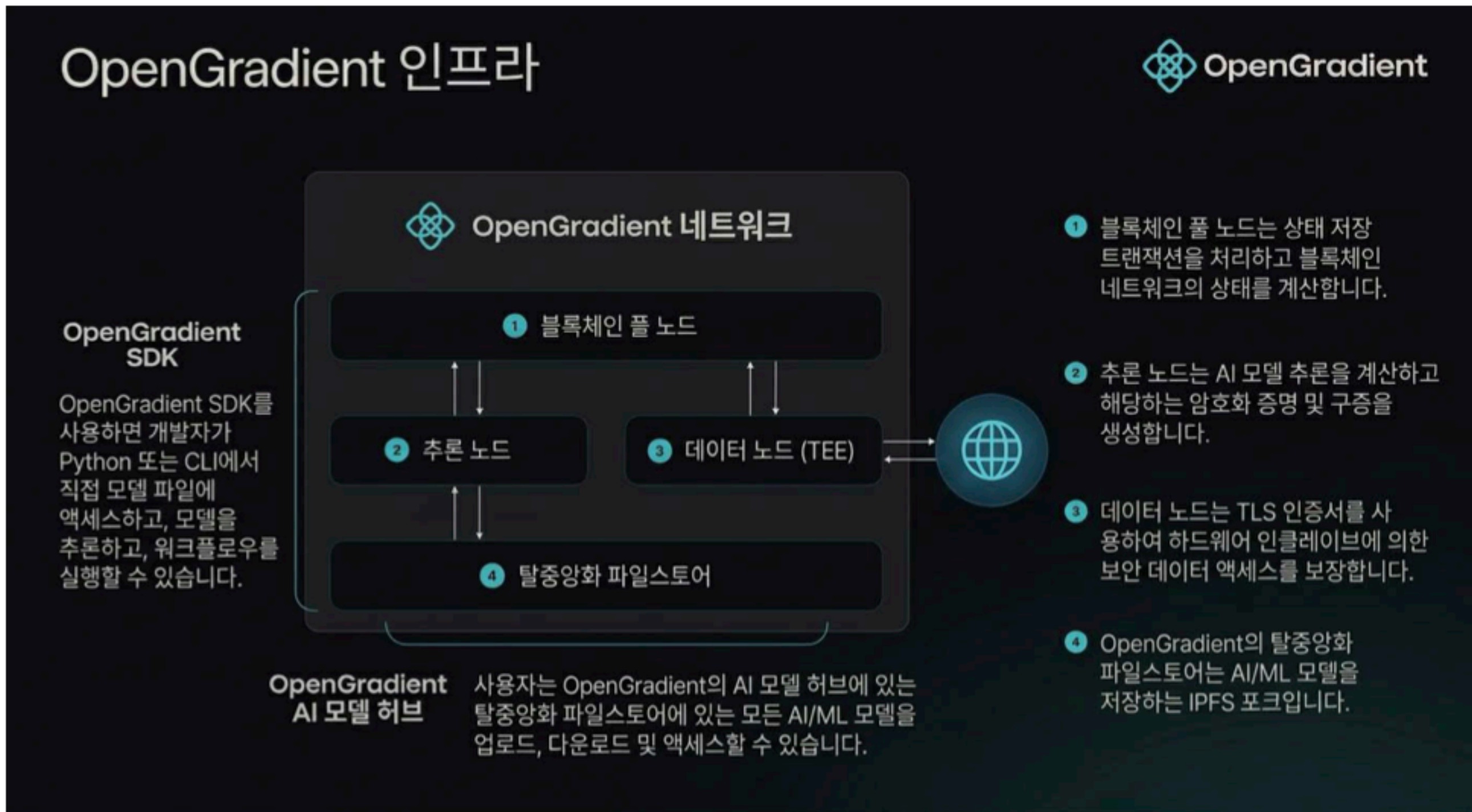


그림 2: HACA 인프라 구성요소 (한국어)

핵심 원칙: 실행과 검증의 분리

HACA의 핵심 아키텍처 원칙은 실행(execution)과 검증(verification)이 서로 독립적인 작업으로 별도의 타임라인에서 수행된다는 것입니다. 검증 레이어가 원래 연산을 보거나 재실행할 필요 없이 증명을 검증하도록 설계되었습니다.

실제 작동 방식

- 사용자가 추론 노드에 프롬프트를 전송합니다.
- TEE 엔클레이브 내에서 요청을 처리하고, LLM 제공자를 호출하여 응답을 반환합니다.
- TEE가 하드웨어 기반 증명서(attestation)를 생성합니다.
- 증명서가 풀 노드에 제출되고, 온체인 TEE 레지스트리와 대조하여 검증됩니다.
- 검증된 증명서가 원장(ledger)에 영구적으로 기록됩니다.

노드 전문화가 필요한 이유

- AI 추론은 비용이 높고 이질적 — 소형 분류 모델과 700억 매개변수 LLM은 하드웨어 요구사항이 크게 다릅니다.
- 검증에 재실행이 필요하지 않습니다 — TEE 증명서와 ZKML 증명은 모델을 다시 실행하지 않고도 정확성을 보장합니다.

- 데이터 접근에는 격리가 필요 — 외부 데이터는 TEE 엔클레이브 내에서 가져와 변조 방지를 보장해야 합니다.
- 스토리지는 컴퓨팅과 다르게 확장 — 모델 파일은 기가바이트 규모이므로, 온체인 참조와 오프체인 스토리지를 분리합니다.

4. 노드 유형 (Node Types)

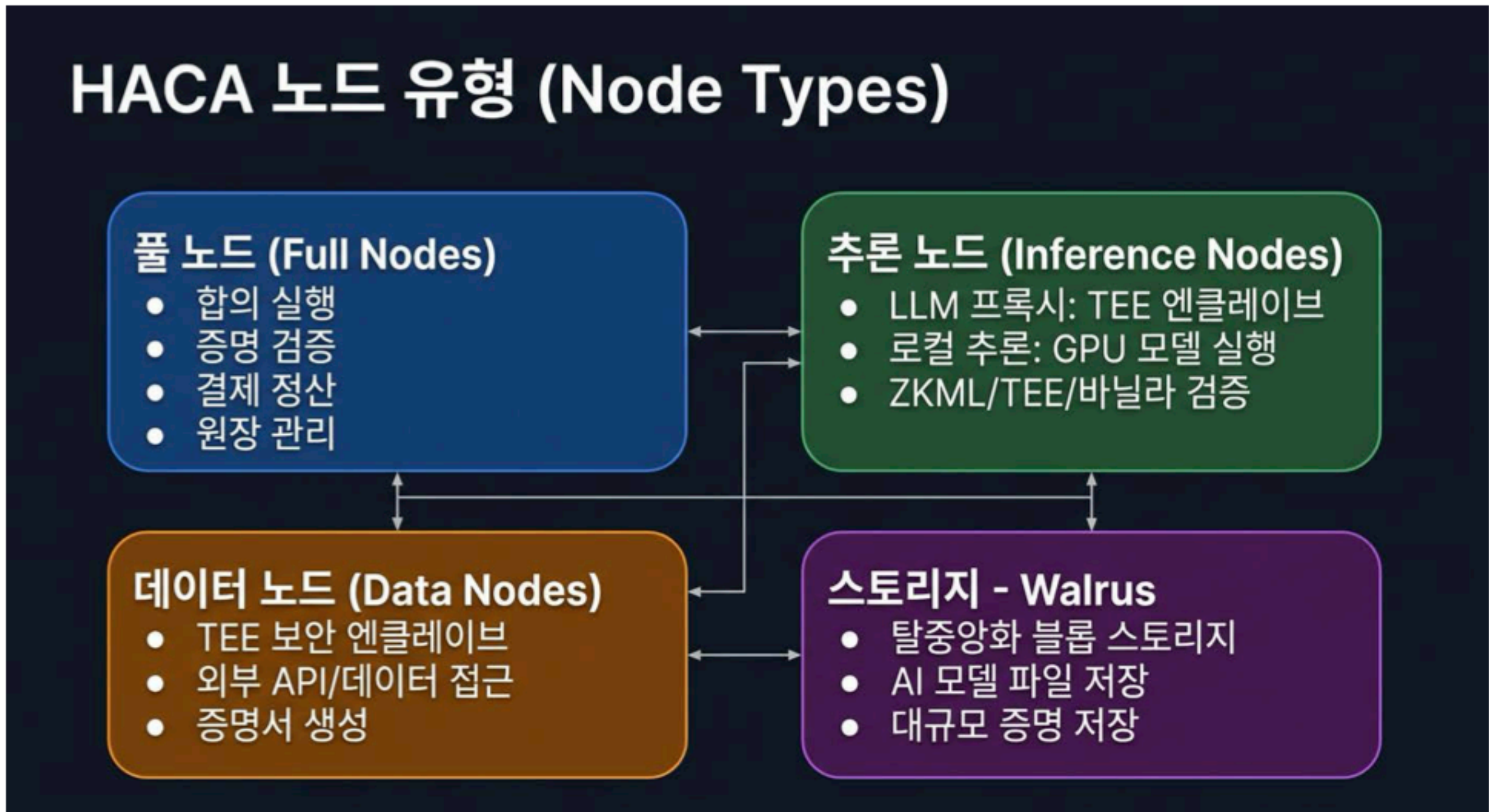


그림 3: HACA 노드 유형 (한국어)

노드별 역할 분담

역할	풀 노드	추론 노드	데이터 노드	스토리지
합의 실행	O	X	X	X
모델 실행	X	O	X	X
증명 검증	O	X	X	X
결제 정산	O	X	X	X
외부 데이터 조회	X	X	O	X
모델/증명 저장	X	X	X	O

4.1 풀 노드 (Full Nodes)

풀 노드는 OpenGradient 원장을 유지하는 탈중앙화 블록체인 검증자 노드입니다. 노드 등록, 결제, 그리고 추론 노드와 데이터 노드가 생성한 모든 증명 및 증명서를 검증합니다.

4.2 추론 노드 (Inference Nodes)

추론 노드는 상태 비저장(stateless) 워커 노드로, 로컬 모델 추론을 위한 GPU를 제공하거나, Anthropic, OpenAI 등의 외부 모델 제공자에 대한 보안 접근을 제공합니다.

- LLM 프록시 노드: TEE 엔클레이브에서 실행되며, 사용자와 외부 LLM API 간의 보안 중개자 역할
- 로컬 추론 노드: Model Hub의 모델을 GPU에서 직접 실행하여 고성능 추론 제공

4.3 데이터 노드 (Data Nodes)

TEE 내에서 실행되며 API, 데이터베이스, 오라클 등의 서드파티 데이터 소스와 암호화된 연결을 설정합니다.

4.4 스토리지 — Walrus

Walrus 탈중앙화 스토리지로 AI 모델 및 대규모 추론 증명을 저장합니다. Blob ID만 온체인에 기록되어 블록체인을 경량으로 유지합니다.

5. 검증 및 정산 (Verification & Settlement)

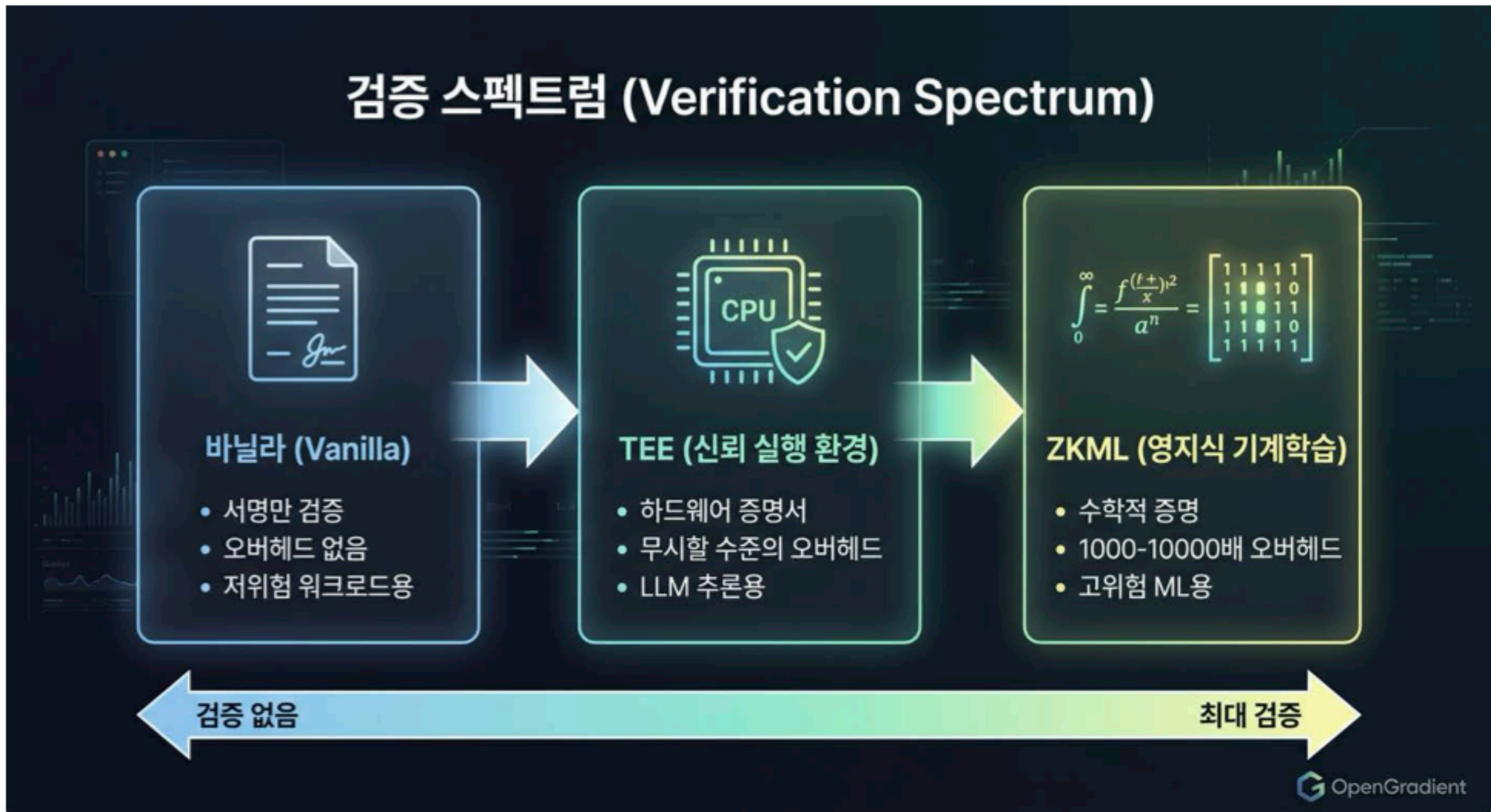


그림 4: 검증 스펙트럼 (한국어)

검증 방법 비교

방법	작동 원리	오버헤드	적용 시기
TEE	하드웨어 증명서로 보안 엔클레이브 실행 증명	무시할 수준	LLM 추론, 프라이버시 민감 작업
ZKML	영지식 증명으로 모델 실행 정확성 수학적 증명	1000-10000배	고위험 ML 모델, DeFi
바닐라	서명 검증만, 암호학적 증명 없음	없음	저위험, 프로토타이핑

신뢰 모델 (Trust Model)

HACA는 세 가지 독립적인 기둥에서 신뢰를 구축합니다:

- 1. 하드웨어 증명서 (TEE): CPU 보안 엔클레이브에서 증명서 생성, 온체인 레지스트리와 대조 검증
- 2. 수학적 증명 (ZKML): 수학만으로 신뢰 제공, 하드웨어 가정이나 신뢰 당사자 불필요
- 3. 합의 (BFT): CometBFT 합의로 비잔틴 장애 허용 제공, 검증자의 1/3 미만 손상 시 네트워크 안전

6. 검증 가능한 LLM 실행

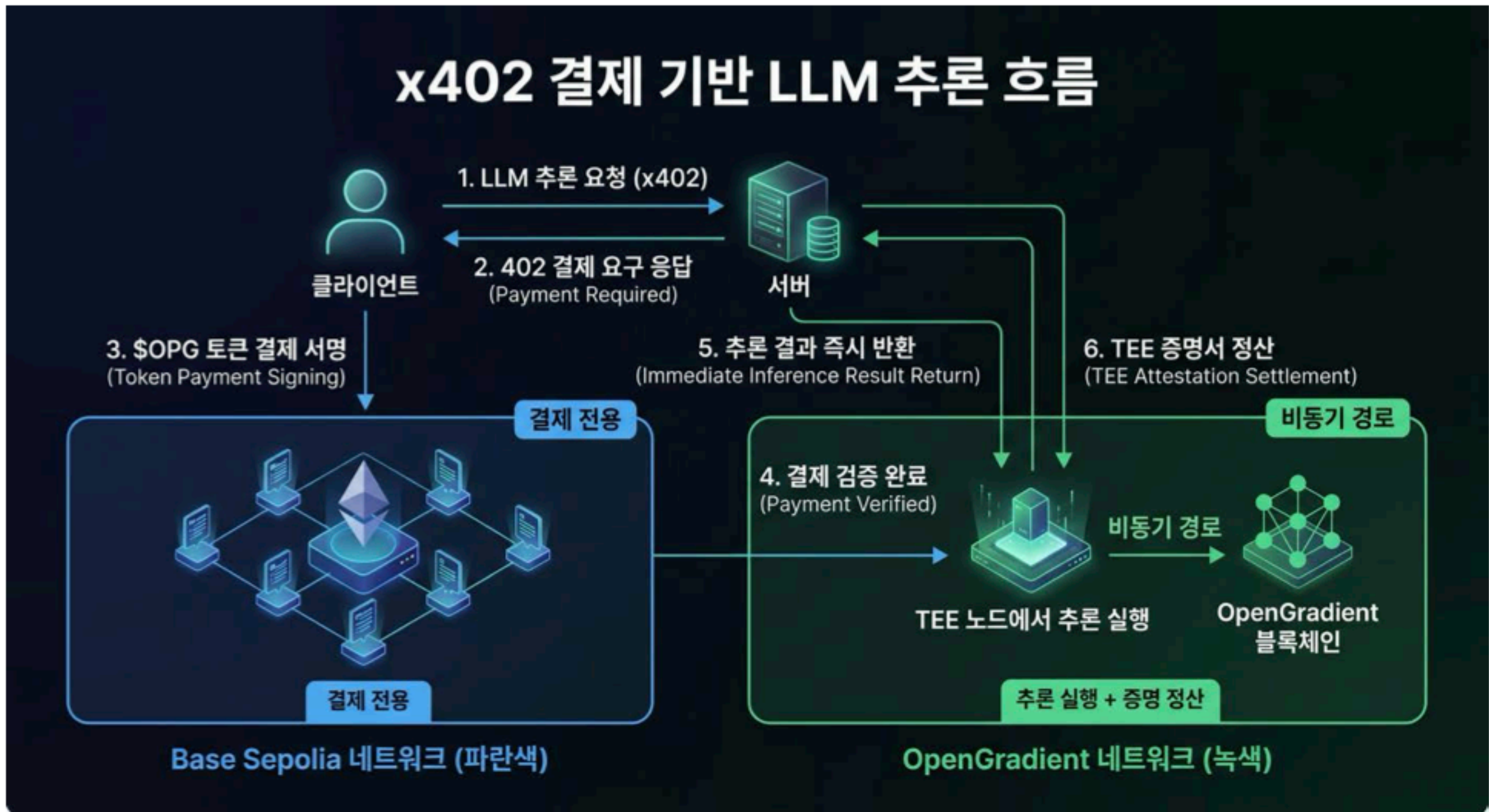


그림 5: x402 LLM 추론 결제 흐름 (한국어)

OpenGradient는 TEE 기반 보안 및 검증 가능한 LLM 추론을 핵심 인프라로 제공합니다. 모든 LLM 요청은 TEE를 통해 라우팅되어, 추론이 올바르게 실행되었고 특정 프롬프트가 사용되었음을 하드웨어 증명서로 검증합니다.

작동 흐름 (10단계)

- 1. TEE 노드 등록 — 온체인 TEE 레지스트리에 등록 및 검증
- 2. Permit2 승인 — Base Sepolia에서 \$OPG 토큰 허용량 승인
- 3. 초기 요청 — LLM 추론 엔드포인트에 HTTP 요청
- 4. 402 결제 요구 응답
- 5. 결제 생성 및 서명
- 6. 서명된 결제와 함께 재제출
- 7. 결제 검증 (Base Sepolia)
- 8. TEE 검증 추론 실행 (OpenGradient 네트워크)
- 9. 결제 정산 (Base Sepolia)
- 10. LLM 정산 — TEE 증명서가 OpenGradient 블록체인에 영구 기록

정산 모드

모드	온체인 데이터	최적 용도
----	---------	-------

SETTLE_INDIVIDUAL	입출력 해시	자금 관리 에이전트 추론
SETTLE_BATCH	배치 해시	챗봇, 고처리량 앱
SETTLE_WITH_METADATA	전체 입출력 + 메타데이터	공개 추적 가능 실행

지원 모델

- openai/gpt-4.1, openai/gpt-4o, openai/o4-mini
- anthropic/claude-4.0-sonnet, anthropic/claude-3.7-sonnet, anthropic/claude-3.5-haiku
- google/gemini-2.5-flash, google/gemini-2.5-pro, google/gemini-2.0-flash
- x-ai/grok-3-beta, x-ai/grok-3-mini-beta, x-ai/grok-4.1-fast, x-ai/grok-2-1212

7. ML 실행 — PIPE

PIPE(Parallelized Inference Pre-Execution Engine)는 전통적인 ML 모델의 네이티브 실행을 위한 병렬 추론 사전 실행 엔진입니다. 원자적(atomic) 보장과 함께 AI 모델을 애플리케이션에서 직접 호출할 수 있습니다.

- 1. 사용자가 트랜잭션을 제출
- 2. 추론 멀풀(Inference Mempool)에 배치
- 3. 모든 트랜잭션을 시뮬레이션하고 추론 요청 추출
- 4. 추론 요청이 추론 네트워크에 전송되어 병렬 실행
- 5. 사전 계산된 추론 결과와 함께 트랜잭션 실행
- 6. 다음 블록에 포함

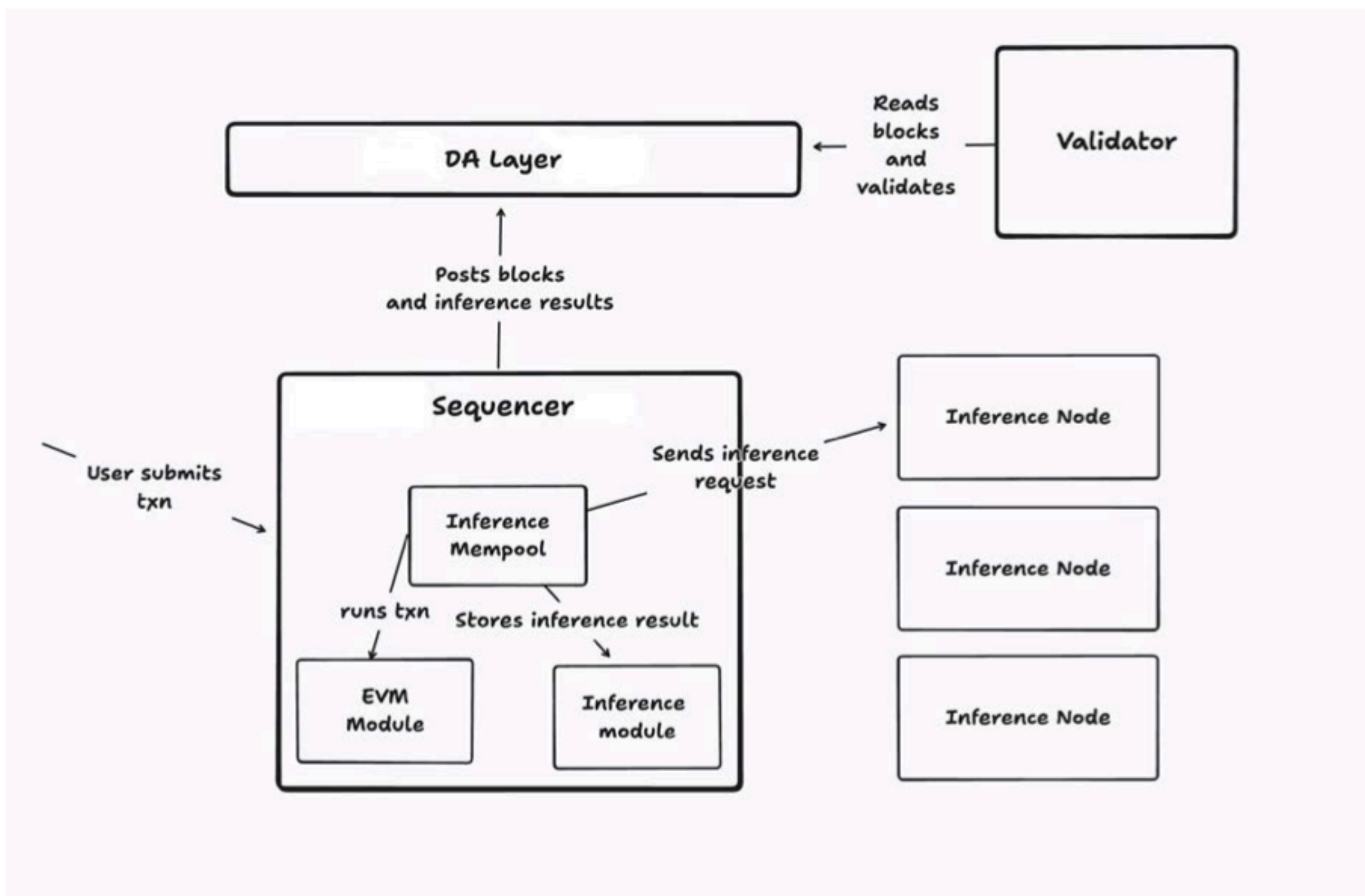


그림 6: PIPE 추론 실행 흐름

LLM 실행과의 비교

기능	PIPE (ML 실행)	x402 (LLM 실행)
통합	스마트 컨트랙트	HTTP/REST API
트랜잭션 원자성	O	X
검증	ZKML, TEE, 바닐라	TEE
모델 유형	ML 모델	LLM

8. 합의 메커니즘 (Consensus)

OpenGradient의 블록체인 레이어는 AI 추론의 검증 및 정산 레이어 역할을 합니다. CometBFT 합의와 Cosmos SDK 및 EVM 호환성을 사용합니다.

CometBFT 주요 속성

- 즉시 완결성: 블록이 커밋되면 즉시 최종 확정
- 비잔틴 장애 허용: 검증자의 1/3 미만이 악의적이거나 결함이 있는 한 네트워크 안전
- 효율적 검증: 비용이 큰 AI 연산을 재실행하지 않고 증명 검증

네트워크 파라미터

파라미터	값	설명
블록 시간	10초	정산을 위한 블록 간 목표 시간
완결성	즉시	커밋 후 즉시 최종 확정
합의 임계값	2/3+	블록 최종 확정에 필요한 합의
장애 허용	<1/3	허용 가능한 최대 비잔틴 검증자

9. 제품 생태계 (Product Ecosystem)

9.1 x402 게이트웨이

Python SDK 없이도 TEE 검증 LLM 추론에 직접 HTTP로 접근. JavaScript, Go, Rust, Python, curl 등 모든 언어에서 통합 가능.

9.2 Model Hub (모델 허브)

AI 모델을 위한 탈중앙화 저장소. Walrus 탈중앙화 스토리지로 모든 모델에 영구적인 저장소 제공. 무허가, 버전 관리, 추론 즉시 가능.

9.3 MemSync (멀싱크)

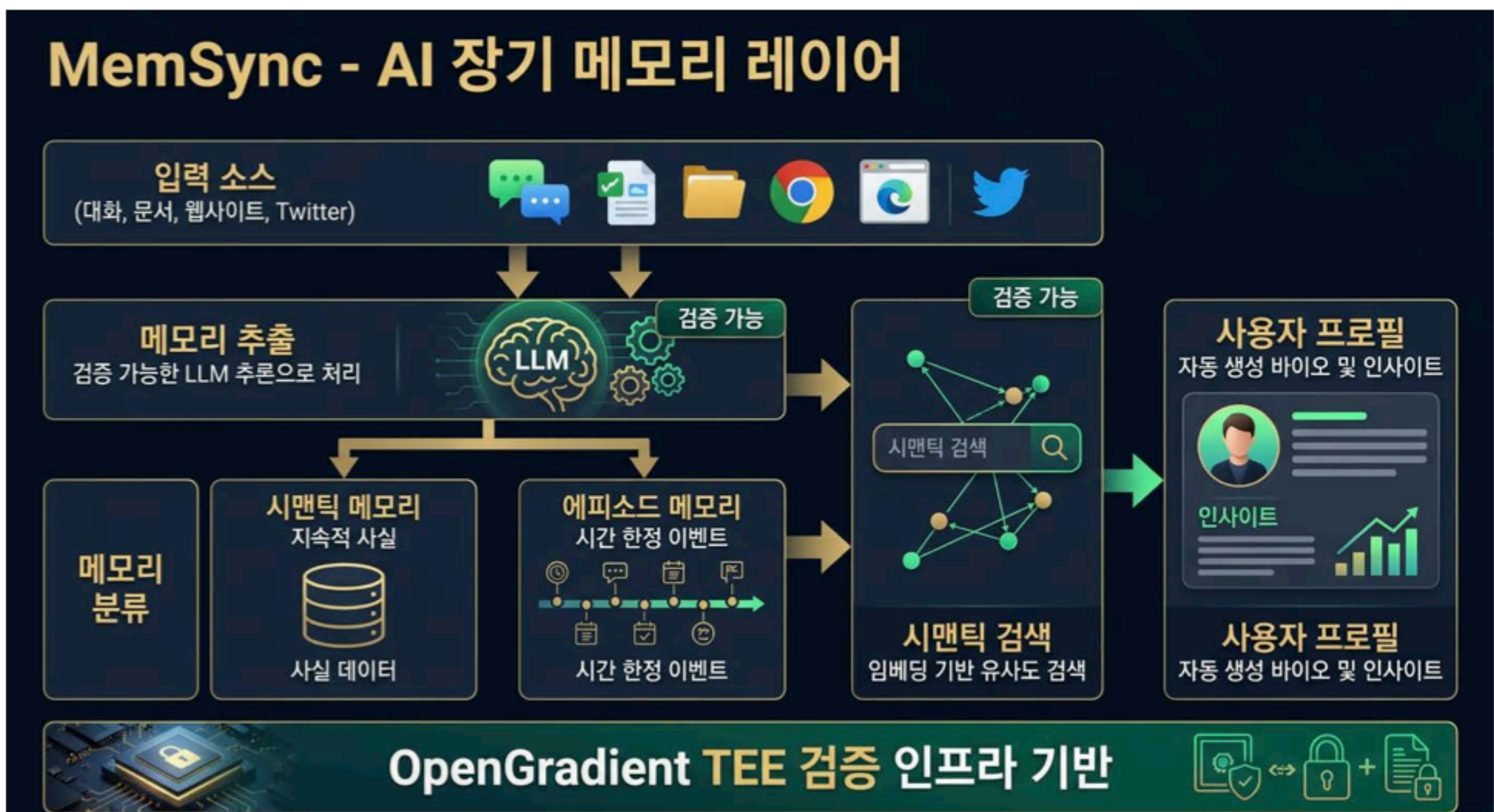


그림 7: MemSync 아키텍처 (한국어)

AI를 위한 장기 메모리 레이어. 대화, 문서, 웹사이트에서 의미 있는 메모리를 자동 추출하고, 시맨틱 검색으로 검색 가능. TEE 검증 기반.

9.4 Twin.fun (디지털 트윈)

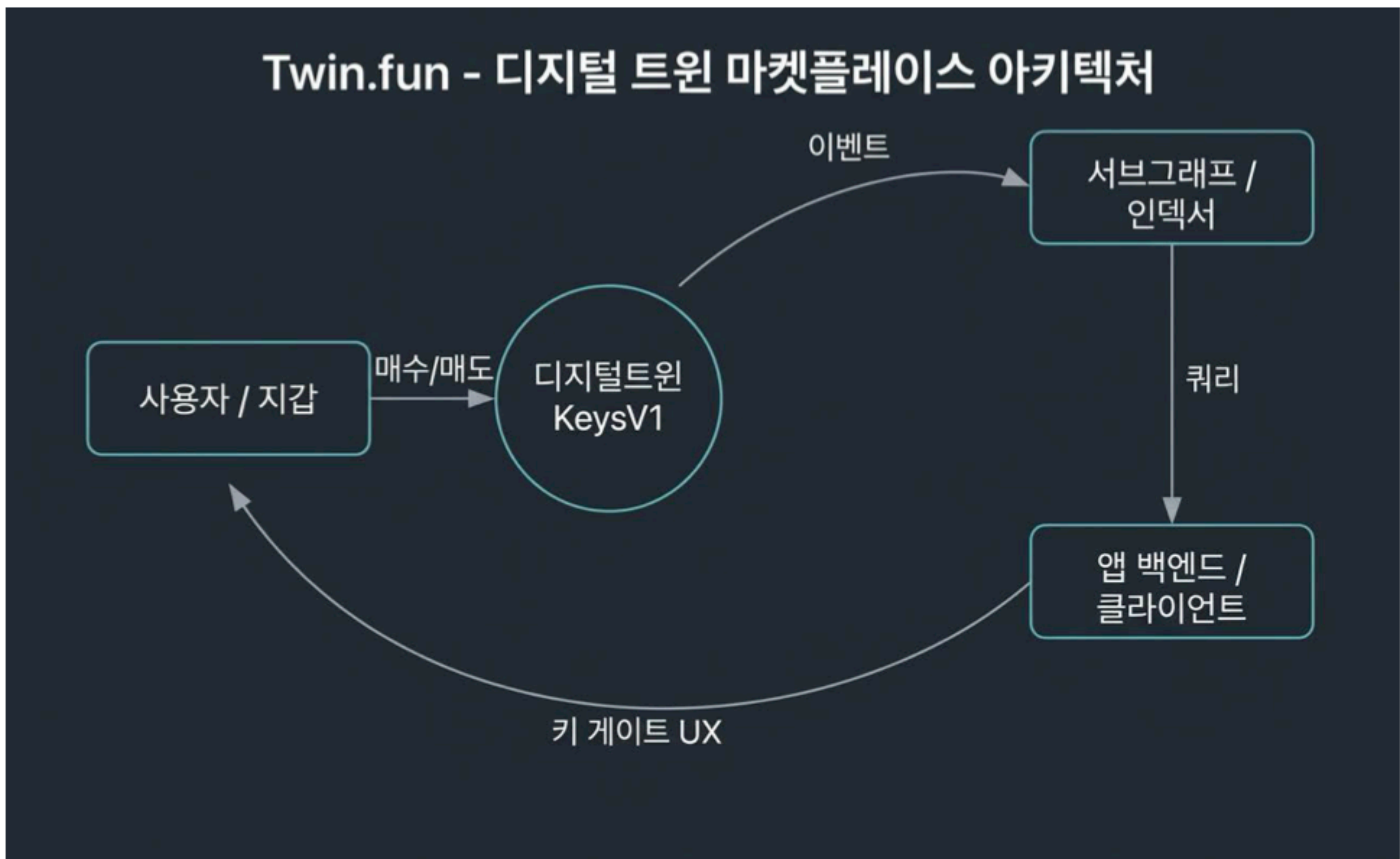


그림 8: Twin.fun 아키텍처 (한국어)

AI 기반 디지털 트윈 마켓플레이스. 본딩 커브에서 키를 매수/매도하며, 키 보유 시 디지털 트윈의 게이트 경험 잠금 해제.

9.5 AlphaSense (알파센스)



그림 9: AlphaSense 아키텍처 (한국어)

AI 에이전트에 강력한 신호를 제공하는 검증 가능한 AI 워크플로우. 변동성 예측, 가격 예측, 시빌 감지, 포트폴리오 최적화 등.

10. 사용 사례 (Use Cases)

엔터프라이즈

- 탈중앙화 및 프라이버시 보존 LLM
- AI 기반 사기 탐지
- 의료 AI 보안 컴퓨팅

AI 에이전트

- 증명 가능한 추론의 AI 에이전트 — 모든 LLM 호출이 암호학적으로 서명
- 블록체인 오케스트레이션 AI 에이전트

DeFi 및 트레이딩

- DEX 동적 수수료 최적화
- DeFi 대출 프로토콜 리스크 평가
- 예측 시장 ML 기반 가격 책정

MemSync 사용 사례

- 개인화된 챗봇, 메모리 강화 LLM 앱, 교육 AI 튜터

11. 네트워크 현황 (Network Status)

OpenGradient 테스트넷

속성	값
네트워크 이름	OpenGradient Testnet
RPC URL	https://ogevmdevnet.opengradient.ai
체인 ID	10740
블록 익스플로러	https://explorer.opengradient.ai/
Faucet	https://faucet.opengradient.ai/

결제 정보

속성	값
결제 네트워크	Base Sepolia
토큰	\$OPG (0x240b09731D96979f50B2C649C9CE10FcF9C7987F)
체인 ID	84532

주요 링크

- <https://opengradient.ai>
- <https://docs.opengradient.ai>
- <https://hub.opengradient.ai>
- <https://explorer.opengradient.ai>
- <https://github.com/OpenGradient>
- <https://discord.gg/opengradient>

12. 용어 사전 (Glossary)

HACA

하이브리드 AI 컴퓨팅 아키텍처 — OpenGradient의 핵심 아키텍처 설계

TEE

신뢰 실행 환경 (Trusted Execution Environment) — 민감한 데이터/코드를 무단 접근으로부터 보호하는 하드웨어 보안 아키텍처

ZKML

영지식 기계 학습 (Zero-Knowledge ML) — 입력 데이터 공개 없이 연산 검증하는 영지식 증명 생성

PIPE

병렬 추론 사전 실행 엔진 — 스마트 컨트랙트에서 모델을 네이티브로 사용 가능하게 하는 온체인 추론 실행 방식

x402

인터넷 네이티브 결제를 위한 개방형 표준. HTTP 402 Payment Required 상태 코드 확장

CometBFT

OpenGradient 검증 레이어의 합의 엔진. 즉시 완결성과 BFT 제공

Walrus

AI 모델 및 대규모 추론 증명을 위한 탈중앙화 블록 스토리지

ONNX

개방형 신경망 교환 (Open Neural Network Exchange) — ML 모델을 위한 개방형 표준 형식

Facilitator

x402 LLM 추론을 위한 Base Sepolia의 결제 검증 및 정산을 처리하는 선택적 서비스

풀 노드

합의를 실행하고 증명을 검증하고 원장을 관리하는 블록체인 검증자 노드

추론 노드

GPU 등 특수 하드웨어를 갖춘, 모델 실행을 담당하는 노드

데이터 노드

TEE 내에서 실행되며 외부 데이터에 대한 신뢰할 수 있는 접근을 제공하는 보안 엔클레이브 노드

AlphaSense

검증 가능한 AI 워크플로우를 래핑하여 AI 에이전트에 강력한 신호를 제공하는 도구



OpenGradient

개방형 인텔리전스를 위한 네트워크

The Network for Open Intelligence

<https://opengradient.ai>

<https://docs.opengradient.ai>

<https://hub.opengradient.ai>

<https://github.com/OpenGradient>

© 2026 OpenGradient. All rights reserved.