

센티언트(Sentient)

해당 번역본은 이용자가 디지털자산 관련 정보를 편리하게 확인할 수 있도록
단순 참고용으로 번역된 것이며, 투자 권유를 목적으로 하지 않습니다.
그 내용이 정확하지 않을 수 있으므로 정확한 정보 습득을 위해서는
원문을 참고하시거나 원문 작성 주체 측에 문의하시기 바랍니다.
두나무는 해당 번역본 내용의 진실성 및 정확성을 보장하지 않고,
제공된 정보에 의한 투자 결과에 대하여 어떠한 책임도 지지 않으며,
투자 손실은 투자자에게 귀속됩니다.

센티언트 프로토콜 백서

목차

1. 서론: 문제와 해결	1
2. 그리드(GRID) – 인텔리전스 디렉토리	1
2.1 목적과 철학	1
2.2 구조와 메타데이터	2
2.3 워크플로 라우팅 및 실행	2
2.4 아티팩트 유형 및 생태계 파트너	3
2.5 빌더 경험	4
3. 센티언트 프로토콜 시스템	5
3.1 수익화	5
3.1.1 가격 책정 및 라이선스	5
3.1.2 발행 예산 및 부트스트래핑	5
3.1.3 수익 분배 및 사용 흐름	5
3.1.4 인센티브 정렬 및 오용 방지 메커니즘	6
3.2 아티팩트 기여자 간의 조정	6
3.2.1 기여자 범주 및 역할	6
3.2.2 3단계 점수 산정 과정	7
3.2.3 로컬 거버넌스 및 예산	7
3.2.4 관리자 대시보드 및 자동화 도구	7
3.3 배포	8
3.3.1 배포 목표	8
3.3.2 아티팩트 네트워크 활용	8
3.3.3 세분화된 배포 경로	8
3.3.4 매칭 및 추천	8

3.3.5 커뮤니티 소유권 및 배포 인센티브	9
4. 신뢰 불필요 평가 및 소유권	9
4.1 지문 식별(Fingerprinting)	9
4.1.1 임베딩 지문	9
4.1.2 탐지 및 집행	9
4.2 신뢰 실행 환경(TEE) — 일반 개요	10
4.2.1 보안 실행	10
4.2.2 데이터 프라이버시 및 규정 준수 예시	10
4.2.3 센티언트 인클레이브 프레임워크(SEF)	10
4.2.4 최종 고려사항	11
5. 토크노믹스	11

1 서론: 문제와 해결

인공지능은 이제 소수의 기업이 통제하는 강력하고 폐쇄적인 모델들로 이루어진 환경으로 발전했다. 이러한 거대 기업들은 데이터, 알고리즘, 컴퓨팅 자원을 독점하여 혁신을 고립된 제품으로 만들고, 독립 개발자들이 참여할 수 있는 기회를 거의 남기지 않는다. 그러나 커뮤니티 주도의 연구와 도구 개발이 보여주듯, 다른 미래 또한 가능하다.

센티언트(Sentient)는 수천 개의 전문화된 지능(모델, 에이전트, 데이터세트, 평가자)이 글로벌 브레인의 뉴런처럼 협력할 수 있는 개방형 네트워크를 구축함으로써 이 환경을 재구성하는 것을 목표로 한다. 이 프로토콜은 오픈소스 AI 아티팩트가 직면한 세 가지 지속적인 과제를 해결한다.

1. **수익화(Monetization)** – 수익이 일반적으로 폐쇄형 제공자에게만 집중되는 상황에서, 개발자들이 오픈소스 기여를 어떻게 지속할 수 있는가?
2. **조정(Coordination)** – 중앙 관리 없이 분산된 기여자들이 로드맵, 품질 기준, 보상 배분에 어떻게 합의할 수 있는가?
3. **배포(Distribution)** – 오픈 아티팩트가 어떻게 올바른 대상(다른 개발자, 기업, 최종 사용자)에게 도달하고, 기존의 독점 플랫폼과 경쟁할 수 있는가?

센티언트는 다음 네 가지 계층을 결합함으로써 이러한 과제들을 해결한다.

- **그리드(GRID)**: AI 아티팩트를 발견하고, 조합하며, 제공하기 위한 탈중앙화된 등록 및 라우팅 계층
- **프로토콜(Protocol)**: 아티팩트 수익화, 기여자 보상 관리, 이해관계자 조정을 위한 스마트컨트랙트 메커니즘
- **기술 프리미티브(Technology primitives)**: 모델 지문 식별, 신뢰 실행 환경(TEE), 검증 가능한 AI 등 공정한 수익화와 거버넌스를 위한 보안 및 감사 가능성을 제공하는 기술적 기반
- **토크노믹스(Tokenomics)**: 인센티브를 정렬하고, 기여를 자금 지원하며, 채택을 촉진하도록 설계된 \$SENT 토큰 기반 경제 시스템

2 그리드(GRID) – 인텔리전스 디렉토리

2.1 목적과 철학

그리드(Grid, Global Research and Intelligence Directory의 약자)는 센티언트 아키텍처의 기초 계층이다. 그리드의 핵심 구성 요소는 **아티팩트(Artifacts)**로, 모델, 에이전트, 데이터 소스, 도구 등 커뮤니티 빌더들이 기여하는 모듈형이자 조합 가능한 지능 단위이다.

그리드는 전 세계 커뮤니티가 기여한 AI 아티팩트의 개방적이고 검색 가능한 카탈로그로 작동하도록 설계되었다. 각 아티팩트는 언어 모델, 특정 도메인 에이전트, 데이터 소스, 도구 등과 같은 모듈형 지능 단위이다. 센티언트의 논지는, 거대한 단일 인공지능(AGI)을 구축하기보다, 전문화된 구성 요소들의 협력을 통해 지능이 발현된다는 것이다.

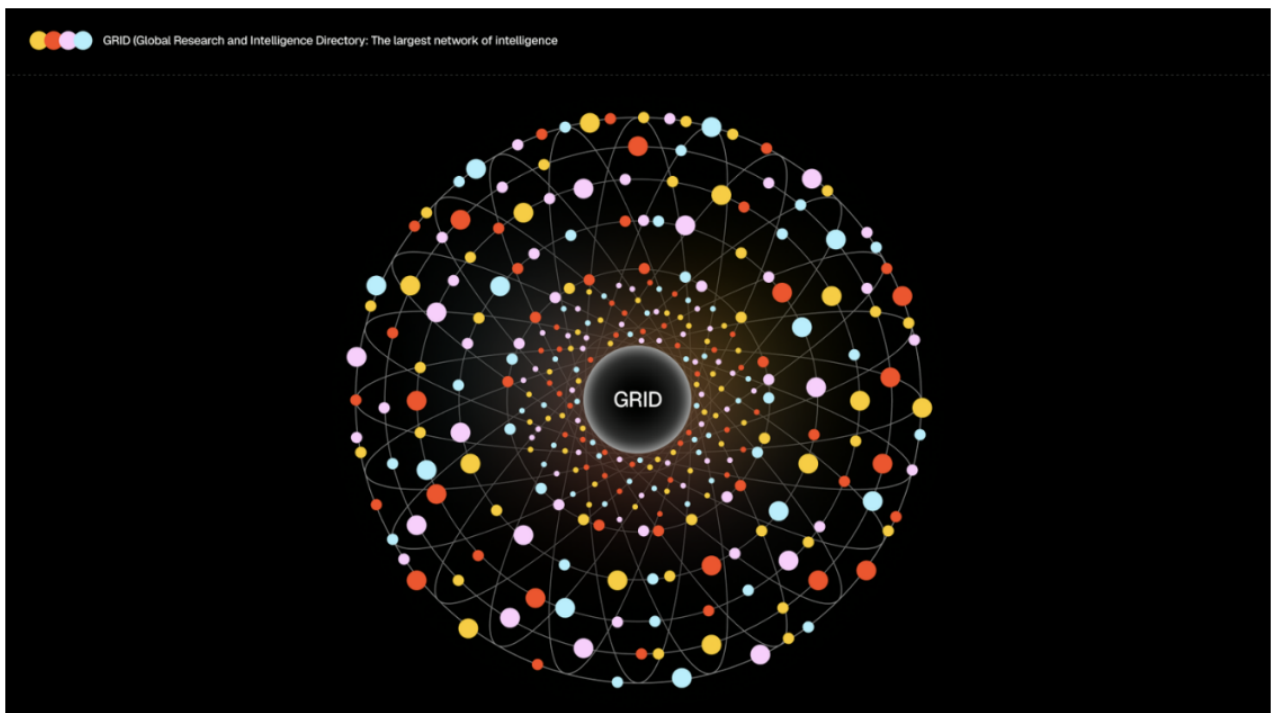


Figure 1: GRID : Global Research and Intelligence Directory

그리드는 이러한 구성 요소들이 상호운용되는 캔버스로서, 단일 모델을 능가하는 워크플로를 가능하게 한다. 참여를 비허가형(permissionless)으로 만들고, 유용한 기여에 보상을 제공함으로써, 센티언트는 오픈 AI의 잠재력을 실현하고자 한다.

2.2 구조와 메타데이터

그리드(Grid)에 등록된 아티팩트는 표준화된 ID로 식별되며, 버전 번호, 의미론적 태그, 기능 설명자, 라이선스 정보, 성능 지표 등 풍부한 메타데이터를 함께 포함한다. 기여자는 입력 및 출력, 모델 크기, 하드웨어 요구사항, 허용된 사용 범위, 호출 속도 제한(call rate limit)을 명시할 수 있다. 이러한 메타데이터는 라우팅과 조합을 가능하게 한다. 예를 들어, 특정 질의가 요약 엔진, 검색 에이전트, 도메인별 데이터베이스를 필요로 할 경우, 그리드는 태그와 평판 점수에 기반하여 적합한 구성 요소를 찾아낸다. 이 등록부는 탈중앙화되어 있어, 단일 주체가 커뮤니티의 합의 없이 특정 아티팩트를 검열하거나 우선시할 수 없도록 보장한다.

2.3 워크플로 라우팅 및 실행

그리드로 전송된 질의는 여러 하위 작업으로 분할되어 적절한 지능(모델 및 에이전트)으로 라우팅되고, 검색 및 도메인 데이터와 같은 도구를 통해 보강된 뒤, 최종적으로 통합되어 가장 우수한 결과를 생성한다. 이를 통해 사용자는 일관되고 고품질의 출력을 얻을 수 있다. 그리드의 결과물은 소수의 폐쇄된 팀이 아닌 수천 명의 오픈소스 개발자들의 작업의 집합체이다.

이 라우팅은 전문가 및 커뮤니티가 정의한 워크플로에 의해 구동된다. 단순한 질의의 경우, 그리드는 전문가가 설계한 워크플로를 활용한다. 각 질의는 분석, 분류된 후 해당 사용 사례에 가장 적합한 워크플로로 라우팅된다.

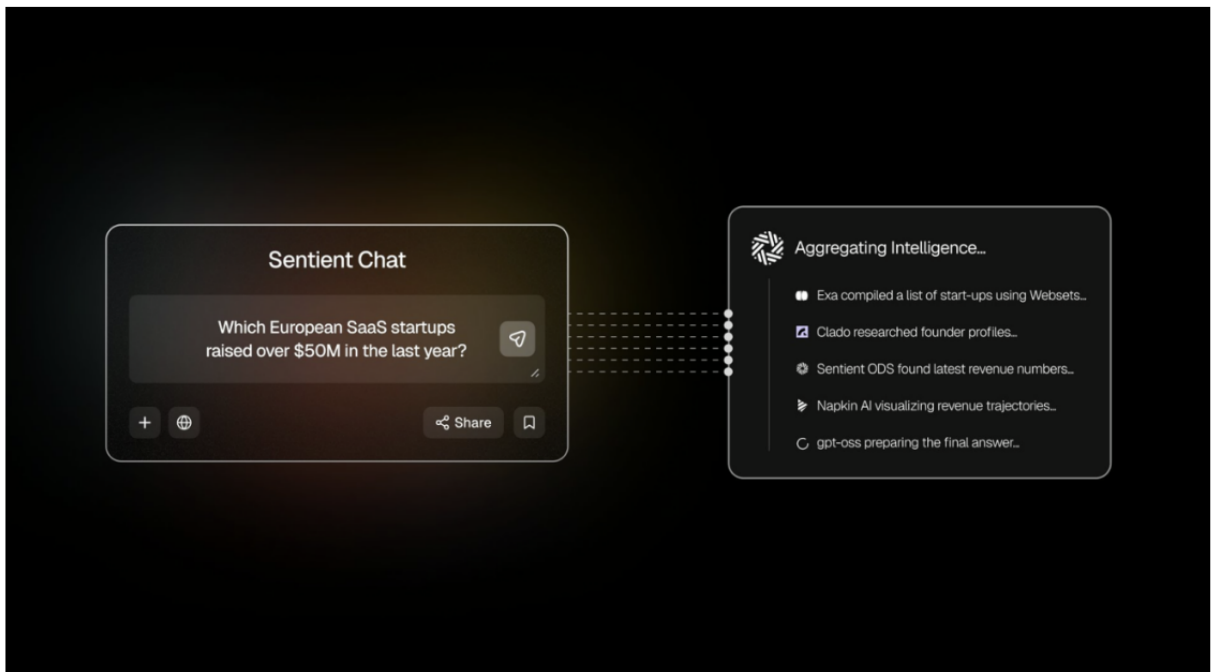


Figure 2: Workflow in GRID

예시 워크플로: 리서치 분석

사용자가 다음과 같은 질문을 한다: “지난 1년간 5천만 달러 이상을 조달한 유럽의 SaaS 스타트업은 어떤 곳인가?”

그리드는 이를 리서치 질의로 식별하고, 이를 특화된 워크플로로 라우팅한다.

- **검색(Search):** 검색 에이전트가 관련 스타트업의 초기 목록을 수집한다.
- **연구(Research):** 연구 에이전트가 창업자 프로필과 기업 데이터를 평가한다.
- **검색(Search):** 검색 에이전트가 해당 스타트업들의 최신 수익 지표를 탐색한다.
- **개념화(Conceptualize):** 인포그래픽 아티팩트가 수익 추이를 시각화한다.
- **통합(Aggregate):** 최종적으로 통합된 답변이 준비되어 사용자에게 전달된다.

워크플로는 글쓰기, 금융, 여행 등 다양한 사용 사례에 맞추어 설계되며, 커뮤니티 전문가들이 이를 큐레이션하고 그 효과성에 따라 보상을 받는다.

2.4 아티팩트 유형 및 생태계 파트너

센티언트의 그리드는 이미 110개 이상의 아티팩트 파트너를 보유하고 있으며, 이 중에는 50개 이상의 전문화된 에이전트, 50개 이상의 데이터 제공자, 다양한 모델, 그리고 컴퓨팅 및 검증 가능한 AI 인프라 파트너들이 포함되어 있다. 대표적인 예는 다음과 같다.

- **도비(Dobby)** – 웹3 생태계에 대한 전문가 수준의 통찰을 제공하는 크립토 네이티브 언어 모델
- **카이토(Kaito)** – 온체인 및 오프체인 신호를 모니터링하고 분석하는 크립토 소셜 데이터 제공자
- **오픈 딥 서치(Open Deep Search)** – 에이전트가 다단계 웹 검색을 수행하고 미묘한 정보를 종합할 수 있도록 하는 오픈소스 라이브러리
- **로마(ROMA)** – 전문화된 AI 에이전트 팀을 조율하기 위한 메타 에이전트 프레임워크
- **칼도(Caldo)**(인물 검색 및 보강), **엑사(Exa)**(LLM 검색), **더 그래프(The Graph)**(블록체인 데이터 인덱싱), **아이겐레이어(EigenLayer)**(통합 컴퓨팅 계층)

이러한 파트너들은 그리드가 단순히 언어 모델에 국한되지 않음을 보여준다. 그리드는 다양한 지능 및 데이터 소스를 통합하며, 이를 통해 센티언트는 크립토 인사이트, 과학 연구, 사회 분석 등 도메인 특화 영역으로 확장한다.

2.5 빌더 경험

센티언트는 빌더(Builder)와 사용자(User)를 구분한다. 빌더는 AI 아티팩트를 생성하고 배포하는 개발자 또는 조직으로, 그리드의 공급 측을 형성한다. 빌더의 여정은 세 단계로 구성된다.

1. **아티팩트 등록(Artifact registration)** – 빌더 여정의 첫 단계는 아티팩트를 온보딩하는 것이다. 네트워크 초기 단계에서 센티언트는 신뢰할 수 있는 모델, 에이전트, 도구로 구성된 고품질 초기 파트너를 큐레이션한다. 시간이 지나면 등록은 비허가형(permissionless)으로 전환되어, 누구나 새로운 모델, 데이터세트, 도구의 메타데이터를 제출하고 입력 및 출력을 명시하며 라이선스 조건을 선언할 수 있다. 빌더는 오픈 리서치, 상업용, 파생형 중 표준 라이선스 템플릿을 선택하여 아티팩트의 사용 방식과 수익 분배 방식을 정의한다. 또한 \$SENT 토큰을 스테이킹하여 자신의 기여에 대한 신뢰를 표시하고, 발행량(emission) 비중을 높일 수 있다. 프로토콜은 메타데이터가 완전한지, 필요한 경우 지문 식별(fingerprinting)이 적용되었는지를 최소한으로 검증한다. 지문 식별은 모델에 비밀 트리거를 내장하여, 소유자가 무단 복제를 탐지하고, 라이선스된 인스턴스만 제공되었음을 증명할 수 있게 한다.
2. **아티팩트 관리(Artifact management)** – 등록 이후 빌더는 지속적인 관리 단계에 들어간다. 빌더는 아티팩트에 스테이킹한 \$SENT 양을 조정할 수 있으며, 이는 아티팩트의 발행 비중에 영향을 준다. 또한 위임자에게 부과되는 커미션율을 변경하거나, 실시간 발행 통계 및 수익을 모니터링할 수 있다. 빌더는 새로운 버전을 릴리스하거나, 라이선스 조건을 갱신하고, 시장 상황에 따라 호출당 가격을 변경할 수 있다. 사용자 피드백에 따라 문서를 개선하거나 모델을 조정할 수도 있다. 프로토콜은 빌더가 아티팩트별 발행 예산을 설정할 수 있도록 허용한다. 이 예산은 초기 채택을 보조하기 위한 기간 제한형 토큰 스트림이다. 아티팩트가 인기를 얻으면 예산은 점차 축소되고, 일반적인 게이지 기반 발행으로 전환된다. 빌더는 네트워크 대표자와 상호작용하여 발행 가중치에 영향을 미치거나, 보완적인 아티팩트(예: 검색 에이전트와 요약 모델)를 결합한 합작 벤처에 참여할 수 있다. 또한 사용량 및 평가 점수를 추적하고, 공유 평가자 및 TEE와 협력하여 안전성과 성능을 보장할 수 있다.
3. **기여자 강화(Supercharging contributors)** – 센티언트는 프로젝트가 수익의 일부를 개별 기여자에게 프로그램적으로 관리하고 분배할 수 있도록 돕는 도구를 개발 중이다. 빌더는 문서 개선, 데

이더세트 보강 등 특정 작업에 로컬 예산을 배정하고, 보상 분배 기준을 설정할 수 있다.

3 센티언트 프로토콜 시스템

프로토콜 계층은 수익화, 조정, 배포를 실행 가능하게 만든다. 이는 스마트컨트랙트, 거버넌스 메커니즘, 인센티브 구조로 구성되어 있으며, 빌더, 사용자, 스테이커, 그리고 커뮤니티 전체의 이해관계를 조정한다.

3.1 수익화

3.1.1 가격 책정 및 라이선스

사용자가 모델, 에이전트 또는 데이터셋을 호출할 때, 온체인 트랜잭션이 발생하며, 수수료는 자동으로 창작자, 호스트, 평가자, 프로토콜 금고 간에 분배된다. 빌더는 아티팩트를 등록할 때 오픈 리서치, 상업용, 파생형 중 하나의 라이선스 템플릿을 선택한다. 예를 들어, 연구 목적으로는 무료 사용을 허용하되, 상업적 배포에는 요금을 부과할 수 있다. 스마트컨트랙트는 라이선스 조건과 수익 분배를 집행하며, 센티언트의 지문 식별 기술은 모델에 비밀 질의-응답 쌍을 내장한다. 이러한 지문은 소유자가 무단 복제를 탐지하고 오용을 입증할 수 있게 하여, 허가된 사용만이 수익을 발생시키도록 보장한다.

3.1.2 발행 예산 및 부트스트래핑

오픈 아티팩트는 흔히 ‘닭이 먼저냐 달걀이 먼저냐’의 문제를 겪는다. 수익을 창출하려면 사용이 필요하지만, 채택에는 자원이 필요하기 때문이다. 이를 해결하기 위해 센티언트는 큐레이션된 아티팩트에 \$SENT 발행 예산을 배정한다. 이 예산은 유의미한 사용자 수익이 발생하기 전에도 기여자, 호스트, 평가자에게 주기적인 토큰 스트림을 제공한다. 발행 예산은 유한하며, 프로토콜 DAO에 의해 주기적으로 재평가된다. 아티팩트가 인기를 얻고 수익을 창출하기 시작하면, 해당 아티팩트의 발행 비중은 점차 축소되어 새로운 참여자에게 자리가 마련된다.

발행 예산은 커뮤니티 지지(스테이크 비중), 네트워크 대표자의 전문가 투표, 제품 사용량 및 수익과 같은 현실 지표를 기반으로 조정된다. 지속적인 채택과 긍정적인 사용자 피드백을 보여주는 빌더는 연장된 예산을 받을 수 있으며, 성과가 저조한 프로젝트는 예산이 축소되어 다른 프로젝트에 자원이 재배분된다.

3.1.3 수익 분배 및 사용 흐름

사용자 또는 에이전트가 아티팩트를 호출하면, 해당 호출은 센티언트 네트워크의 스마트컨트랙트에 의해 중개된다. 이 계약은 아티팩트의 라이선스를 참조하고, 등록 시 합의된 조건에 따라 결제를 분배한다. 등록 시 기록된 소유권 지분은 각 결제 중 얼마가 원 빌더와 유지관리자에게 돌아가는지를 결정한다. 결제는 다음 여러 이해관계자에게 분배된다.

- **빌더 및 유지관리자:** 기록된 지분에 비례한 몫이 아티팩트를 제작하고 유지하는 팀에 지급된다.
- **호스트:** 아티팩트를 제공하기 위해 컴퓨팅, 대역폭, 저장소를 제공하는 운영자는 수수료의 일부를 받을 수 있다. 그들은 잘못된 보고를 방지하고 올바른 모델 버전을 제공하도록 담보를 유지해야 한다.

- **평가자:** 아티팩트가 제3자 평가자(예: 안전성 검사, 품질 평가)에 의존할 경우, 이들 역시 사용 수익의 일부를 받을 수 있다.
- **프로토콜 금고:** 보조금, 보안 감사, 지속적 유지보수를 위한 소규모 수수료가 센티언트 금고로 들어간다.
- **스테이커 및 대표자:** 아티팩트에 스테이킹한 참여자와 네트워크 대표자는 개별 사용 수수료 대신 발행 공식(토크노믹스 섹션 참조)을 통해 간접적으로 보상받는다.

투명성을 보장하기 위해, 사용 데이터는 온체인 또는 검증 가능한 오프체인 인덱스에 기록되며, 이를 통해 감사가 가능하고 이중 지출을 방지한다. 지문 식별은 각 배포된 모델에 비밀 키-응답 쌍을 삽입하여, 호스트가 사용 권한을 요청하고 대금을 지불했는지를 검증할 수 있도록 한다. 이러한 흐름을 표준화함으로써 센티언트는 경제적 인센티브를 정렬하면서도, 아티팩트가 자체 거버넌스 절차를 통해 분배 방식을 맞춤화할 여지를 남긴다.

3.1.4 인센티브 정렬 및 오용 방지 메커니즘

오용을 방지하기 위해, 프로토콜은 여러 메커니즘을 결합한다.

- **지문 식별(Fingerprinting)** (4절 참조): 무단 모델 복제를 탐지한다. 빌더는 의심되는 배포에 비밀 질의를 보내 테스트할 수 있으며, 호스트가 라이선스되지 않은 복제본을 제공하는 경우, 증거를 제출하여 벌칙 및 수익 재지급을 유도할 수 있다.
- **신뢰 실행 환경(TEE)** (4.2절 참조): 올바른 모델 실행과 사용량 회계를 강제하며, 원격 인증(remote attestation)을 통해 바이너리 및 모델 해시가 라이선스된 버전과 일치함을 검증한다.
- **평판 및 스테이킹(Reputation and staking)** – 아티팩트를 지원하는 스테이커는 아티팩트가 부정 행위를 할 경우 가치 손실 위험을 부담한다. 호스트는 \$SENT를 스테이킹하여 운영하며, 사용량을 허위 보고하거나 비인가 버전을 제공할 경우 스테이크를 잃을 수 있다. 이는 무임승차를 방지하고 협력을 촉진한다.

3.2 아티팩트 기여자 간의 조정

3.2.1 기여자 범주 및 역할

센티언트는 AI 아티팩트가 다양한 인력에 의해 만들어진다는 점을 인식한다. 코드 작성자인 핵심 개발자, 데이터 수집 및 라벨링을 담당하는 데이터세트 큐레이터, 테스트 스위트를 설계하는 평가자, 컴퓨팅을 운영하는 인프라 제공자, 문서 작성자, 커뮤니티 운영자, 통합 개발자 등이 있다. 기여자는 아티팩트의 기여자 등록부에 하나 이상의 역할을 등록할 수 있다. 이 등록부는 온체인 멤버십 목록으로 기능하며, 보상 및 거버넌스 권한 자격을 결정한다. 역할을 명시적으로 기록함으로써, 문서화, 데이터세트 큐레이션, 커뮤니티 활동 등 코드 외의 기여 또한 가시화되고 가치 평가를 받는다. 한 사람이 여러 역할을 수행할 수 있으며, 예를 들어 개발자가 동시에 유지관리자이자 평가위원회 구성원일 수도 있다.

3.2.2 3단계 점수 산정 과정

기여가 자격 검증을 통과하면, 센티언트는 아티팩트가 이를 평가할 수 있도록 객관적 지표, 주관적 평가, 인간 판단을 결합한 3단계 점수 시스템을 적용한다. 이 과정은 기여가 성능을 향상시켰는지, 품질을 개선했는지, 미묘하지만 중요한 가치를 제공했는지를 포착하여 공정한 보상 배분을 목표로 한다.

1단계: 정량적 트랙션 점수(Quantitative traction score)

측정이 용이한 객관적 지표로, 코드 라인 변경 수, 병합된 PR 수, 총 커밋 수 등 소프트웨어 신호, 포크 또는 스타 증가와 같은 다운스트림 신호, 추론 지연, 모델 정확도, 데이터세트 커버리지 등의 성능 지표를 포함한다.

2단계: AI에 의한 정성 평가(Qualitative evaluation by AI)

평가자 및 동료 리뷰어가 코드 품질, 유지보수성, 문서화, 재현성, 공정성을 검사한다. 센티언트는 SWE-bench에서 영감을 얻은 도메인 특화 평가 모델을 탐구하고 있으며, 배포, 작업 복잡도, 기능 확장성을 포함한 정성적 영향을 정량화하려 한다.

3단계: 인간 전문가에 의한 정성 평가(Qualitative evaluation by human experts)

유지관리자나 지정된 위원회는 경계 사례를 반영하기 위해 사전 정의된 범위 내에서 점수를 조정할 수 있으며, 이를 통해 공정성과 적응성을 확보한다. 이러한 조정은 거버넌스를 통해 투명하게 공개되고 검토 가능하다.

가중치 점수화와 적응성(Weighted scoring and adaptability)

각 아티팩트는 정량적·정성적·수동 입력의 가중치를 도메인별 우선순위에 맞게 조정할 수 있으며, 이를 통해 다양한 저장소 전반에서 점수화 체계의 관련성을 유지한다.

3.2.3 로컬 거버넌스 및 예산

아티팩트는 프로토콜의 우산 아래에서 미니 DAO로 운영된다. 각 아티팩트는 수익과 발행량으로 자금이 조성되는 자체 금고를 유지하며, 유지보수, 신규 기능, 커뮤니티 이니셔티브를 위한 예산을 설정한다. 기여자는 표결권을 얻기 위해 토큰을 스테이킹하거나 위임한다. 제안은 오프체인 논의를 거쳐 온체인 투표로 진행된다. 예산은 바운티, 지속 보상, 소급형 보조금, 공용 인프라 이니셔티브에 자금을 지원할 수 있다. 정기 체크포인트의 리듬을 통해 재배분과 우선순위 정교화가 지원된다.

3.2.4 유지관리자 대시보드 및 자동화 도구

운영 도구는 유지관리자가 기여자를 조정하고 보상 흐름을 관리하도록 돕는다. 통합 대시보드에는 기여자 등록부, 백로그, 예산, 발행 통계가 표시된다. 주요 기능은 다음과 같다.

- 작업 정의 및 자금 지원(바운티, 예산, 마감일; 지속 또는 소급 지급)
- 기여 할당 및 추적(VCS 통합, 피드백 루프)
- 보상 분배 프로그래밍(3단계 점수 시스템 활용)
- 지급 자동화(스마트컨트랙트가 완료 시 자금 집행)
- 기여 분석(평판, 처리량, 평균 점수, 전문성)

3.3 배포

배포는 오픈소스 AI 아티팩트가 개발자, 엔터프라이즈 팀, 최종 사용자 제품 전반으로 확산·채택되는 과정을 의미한다. 프로토콜이 개방적이고 조합 가능하기 때문에, 아티팩트는 워크플로 전반에서 등록·조합·재사용될 수 있으며, 배포는 협력적 네트워크 효과로 전환된다.

3.3.1 배포 목표

유용한 아티팩트가 올바른 대상에게 도달하도록 보장한다. 사용은 수익과 토큰 발행을 동시에 견인하여, 빌더가 다른 아티팩트를 포함한 광범위 적용 가능한 컴포넌트를 설계하도록 인센티브를 제공한다.

3.3.2 아티팩트 네트워크의 활용

아티팩트는 서비스이자 소비자이다. 서로 호출되거나 다른 컴포넌트를 호출할 수 있다. (예: 연구 에이전트 + 검색 + 요약 + 번역 + 평가자)로 구성된 파이프라인은 사용량과 수익을 상류로 전파한다. 센티언트는 채팅, 대시보드, 파트너 API를 통해 이러한 조합을 노출하여 네트워크 효과를 배가시킨다.

3.3.3 세분화된 배포 경로

개발자 중심: SDK, 템플릿, CLI, 라이선스 메타데이터를 포함한 패키지 매니저 퍼블리싱, 포털과 포럼. 조합 가능성 덕분에 최종 사용자가 직접 호출하지 않더라도 상류 컴포넌트에 사용이 축적된다.

엔터프라이즈: 컨테이너 또는 TEE를 통한 프라이빗 배포, SLA가 포함된 엔터프라이즈급 라이선스, 온프레임 데이터 처리와 감사 추적; 창작자에게 수익이 라우팅되는 구독·사용량 기반 조건.

제품: 소비자 앱을 위한 API 엔드포인트와 UI 컴포넌트(예: 요약, 추천). 다중 아티팩트 경험은 사용 크레딧과 수익을 체인 상단으로 연쇄적으로 전달한다.

3.3.4 매칭 및 추천

(카이토에서 영감을 받은) 매칭 엔진은 의미론적 태그(도메인, 모달리티, 컴플라이언스, 라이선스)와 신호(사용량, 평점, 스테이크)에 기반하여 아티팩트를 사용자와 연결한다. 채택이 증가함에 따라 고품질 아티팩트가 부상한다.

3.3.5 커뮤니티 소유권과 배포 인센티브

더 폭넓은 이해관계자 소유는 배포를 증폭한다. 마일스톤에 대한 발행 배수 등 일시적 인센티브는 초기 채택을 가속화하고 시간이 지남에 따라 감쇠한다. 다른 아티팩트에 의한 채택도 사용량으로 계산되어, 재사용 가능한 컴포넌트와 합작 벤처를 장려한다.

4 신뢰 불필요 평가 및 소유권

센티언트의 프로토콜은 적절한 회계, 보안, 검증 가능성을 보장하는 기술 프리미티브에 의해 뒷받침된다. 이러한 프리미티브는 개방적으로 개발되어 생태계와 공유된다.

4.1 지문 식별(Fingerprinting)

지문 식별은 소유권을 증명하고 무단 모델 복제를 탐지하는 데 도움이 되는 AI 네이티브 암호학적 프리미티브이다. 정적 파일에 대한 디지털 서명과 달리, 지문 식별은 수십억 개의 파라미터로 동작이 규정되는 생성 모델에도 작동한다. 센티언트의 지문 라이브러리는 파인튜닝 또는 어댑터 학습 중 모델에 비밀 질의-응답 쌍을 내장한다. 지문은 트리거 시퀀스와 이에 대응하는 응답으로 구성된다. 트리거는 대중에게 알려지지 않은 분포에서 추출된 입력이며, 응답은 특정 어휘적·통계적 속성을 가진 출력이다.

4.1.1 지문 삽입(Embedding fingerprints)

기법은 다음을 포함한다.

- **특화 파인튜닝(Specialized fine-tuning)** – 비밀 트리거 쌍을 포함한 소규모 데이터셋으로 학습하여, 정상 성능을 보존한다.
- **파라미터 확장(Parameter expansion)** – 로우랭크 어댑터나 프롬프트 튜닝 벡터를 사용해 기본 가중치를 변경하지 않고 지문을 인코딩한다.
- **베니그 데이터 혼합(Benign data mixing)** – 과적합과 탐지를 방지하기 위해 트리거를 베니그 데이터와 혼합한다.
- **모델 믹싱(Model mixing)** – 가중 평균을 통해 여러 지문 세트를 번들링하면서 충실도를 유지한다.

지문 파라미터(키 길이, 트리거 수, 중복도)는 강인성과 탐지 시간 간의 트레이드오프를 조정한다.

4.1.2 탐지 및 집행(Detection and enforcement)

1. **챌린지 생성** – 소유자/감사자가 비밀 트리거를 생성한다.
2. **모델 심문** – 트리거를 전송하면, 지문 삽입 모델은 높은 확률로 일치하는 응답을 반환한다.
3. **증거 기록** – 탐지 결과를 기록하고 커밋먼트에 대해 온체인에서 검증한다.
4. **시정 조치** – 수익 재지급, 위반 호스트 슬래싱, 라이선스 정지; 상습 위반자는 호스팅 권한을 상실한다.

추가 자료:

- 지문 라이브러리(OML 1.0): <https://github.com/sentient-agi/OML-1.0-Fingerprinting>
- 연구 논문(AlphaXiv): <https://www.alphaxiv.org/abs/2502.07760>
- 블로그 개요: <https://www.sentient.xyz/blog/fingerprinting-enabling-open-source-monetization>

4.2 신뢰 실행 환경(TEE) – 일반 개요

TEE는 하드웨어로 강제되는 격리를 제공하여, 신뢰할 수 없는 호스트에서도 코드와 데이터의 기밀성을 유지한다. 인클레이브는 사용된 정확한 코드 바이너리와 AI 모델 해시를 인증(attest)하여 올바른 실행을 보장한다.

4.2.1 보안 실행(Secure Execution)

코드, 가중치, 입력은 인클레이브(격리된 메모리)로 로드된다. 호스트 OS/하이퍼바이저는 이를 읽거나 수정할 수 없다. 실행 후 인클레이브는 인증 문서를 출력하며, 중요 활동(결제, 키 관리, 거버넌스)은 유효한 인증에 기반하여 게이팅할 수 있다.

TEE와 기타 검증 가능 연산의 비교(TEEs vs. other verifiable compute)

- **ZK 증명(ZK proofs):** 완전한 암호학적 검증을 제공하나, 대규모 모델에 대해서는 현재 지연/비용이 높다(빠르게 개선 중).
- **MPC:** 비밀을 노드 간에 분할하여 단일 노드 신뢰를 줄이지만, 프로토콜/지연 상의 트레이드오프가 있다.
- **TEE:** 네이티브에 가까운 속도와 범용성을 제공하나, 벤더 루트 오브 트러스트에 의존하며 사이드채널에 유의해야 한다. NVIDIA는 여러 데이터센터급 카드에서 GPU TEE를 지원한다(예: Phala).

센티언트는 하드웨어에 독립적인 보장을 지향하면서도 무시할 만한 오버헤드를 제공하는 TEE에 투자하며, 파트너 전반에서 TEE/MPC/ZK 옵션의 탐색을 장려한다.

4.2.2 데이터 프라이버시 및 규정 준수 예시(Data Privacy and Compliance Examples)

소비자 PII: 인클레이브 내부에서만 복호화처리 후 재암호화한다. 어떤 감사를 거친 바이너리/모델이 PII를 처리했는지를 증명하는 인증을 제공한다.

엔터프라이즈/규제(Enterprise/regulatory): 헬스케어, 금융, 법률 분야에서 “사용 중인 데이터(in use)”를 보호하고, 증명 가능한 인증을 통해 TEE 전용 접근을 강제한다.

검증 가능한 AI 평가(Verifiable AI evaluation): 코드, 모델, 데이터세트 해시를 인증하여 파이프라인을 감사하고, 신뢰 불필요 점수 산정 및 벤치마크 실행을 가능하게 한다.

4.2.3 센티언트 인클레이브 프레임워크(SEF)

SEF는 AWS Nitro Enclaves(CPU TEE)에서 기밀 AI를 실행하기 위한 엔드투엔드 스택이다. Nitro는 사용자 데이터를 클라우드 호스트로부터 보호하고, 인클레이브는 실행을 머신 호스트로부터 보호한다. SEF는 다음을 추가한다.

- **재현 가능한 빌드(Reproducible builds)** – 결정적 해시를 가진 EIF를 생성
- **원격 인증(Remote attestation)** – 인클레이브 내부의 코드 및 모델 서버 인증
- **파일시스템 모니터(File-system monitor)** – 데이터 입출입의 무조건적 인증
- **투명한 VSocket 프록시(Transparent VSocket proxies)** – 선택적 킬 스위치 기능 지원

SEF는 검증 가능한 연산을 위한 보안성, 사용성, 재현성을 강조한다.

SEF 저장소: <https://github.com/sentient-agi/Sentient-Enclaves-Framework>

4.2.4 최종 고려사항(Final Considerations)

TEE 벤더 신뢰(실리콘/펌웨어)는 AWS, Intel, AMD 전반에 걸쳐 공유되는 전제이다. SEF는 오픈소스이며, QEMU 상의 EIF를 통해 AMD/Intel로 이식 가능하다. 이러한 유연성은 에이전트 호스팅에서 평가에 이르기까지 프로토콜의 광범위한 적용을 지원한다.

5 토크노믹스(Tokenomics)

센티언트 네트워크는 지능을 창조, 유지, 큐레이션하는 사람들에게 경제적 권리를 부여한다. 발행량 (emission)은 시간에 따라 구성원 간 정렬을 위해 분배된다. 주요 부분은 빌더(개발자, 연구자, 데이터 큐레이터, 평가자, 유지관리자), 초기 후원자(확장된 베스팅 포함), 그리고 보조금, 안전성, 커뮤니티, 프로토콜 개발을 담당하는 비영리 관리 기관인 센티언트 재단(Sentient Foundation)에 배정된다. 초기 에어드롭은 폭넓은 이해관계자 기반을 형성한다.

토큰 활용성(Token Utility)

- **블록체인 운영 가스:** 온체인 작업(라이선스, 호출, 데이터세트 질의, 스테이킹, 업그레이드)마다 토큰이 소모된다.
- **AI 아티팩트 결제:** 사용자는 모델, 에이전트, 데이터세트, 평가에 대해 비용을 지불하며, 계약은 창작자, 호스트, 평가자, 금고로 수익을 라우팅한다. 분배 구조는 아티팩트 거버넌스에 의해 결정된다.
- **스테이킹 및 위임:** 보유자는 아티팩트에 스테이킹하거나 대표자(Rep)에게 위임하여 발행·정책 투표에 참여하고, 위임자는 대표자의 보상을 공유한다.
- **거버넌스:** 토큰 가중 투표로 발행, 게이지, 자금 배분, 업그레이드를 결정하며, 제안은 보유자 또는 위임자에 의해 제출된다.
- **유동성 및 조합성:** 아티팩트별 토큰/파생상품과 페어링하여 본딩 커브 및 유동성 풀을 통해 아티팩트를 크라우드펀딩할 수 있다.

토큰 발행 메커니즘(Token Emission Mechanism)

Token Emission Mechanism

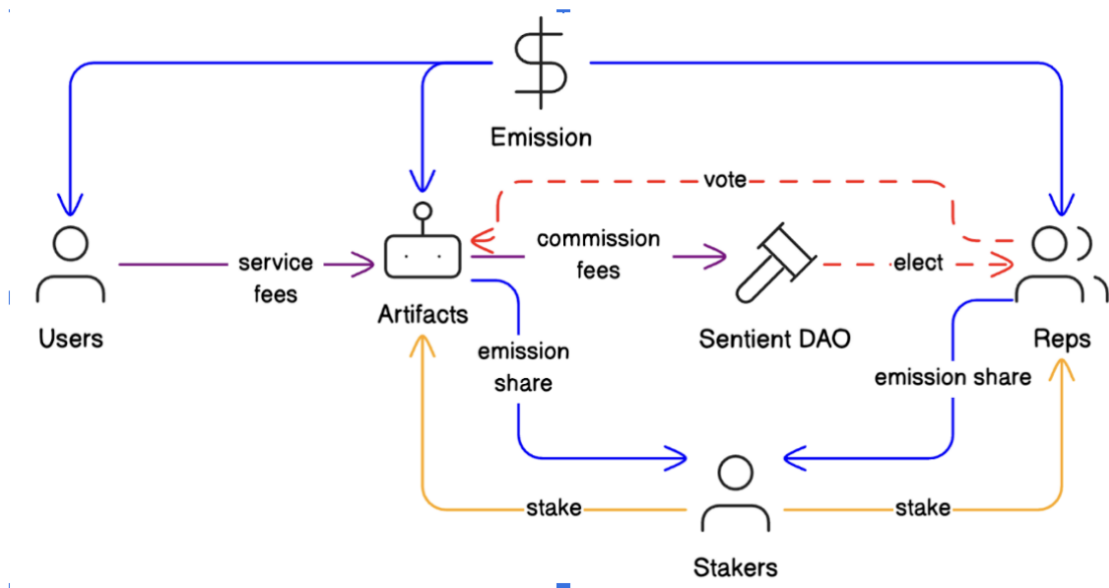


Figure 3: Token Emission Mechanism.

그림 3: 토큰 발행 메커니즘.

센티언트는 성장을 지원하고 참여를 보상하기 위해 지속적으로 토큰을 발행한다. DAO는 연간 발행률(에포크당 현재 공급의 비율)과 풀 간 배분 정책을 통제한다.

1. **사용자 인센티브(User incentives)** – 센티언트 인터페이스(채팅, API, 대시보드, 파트너 앱) 사용자에게 발행량을 분배하여 네트워크 효과를 촉진한다.
2. **대표자 및 스테이커 보상(Representative and staker rewards)** – 대표자에게 기본 급여를 지급하고, 스테이커에게 스테이크 가중 분배를 실시한다. 다음과 같이 정의된다.

$$\text{StakeWeight (SW)} = \text{Token}_{\text{staked}} \times \left(1 + 3 \times \frac{\text{LockDuration (years)}}{\text{MaxLockDuration (4 years)}} \right).$$

스테이커는 자신이 위임한 대표자의 수익 대부분을 공유한다.

3. **아티팩트 발행(Artifact emissions)** – AI 아티팩트 및 기여자에게 가장 큰 비중이 할당된다. 게이지는 세 가지 신호를 결합한다.

$$V \in (0, 1) \quad (\text{스테이크 가중 정규화 대표자 투표})$$

$$SWS \in (0, 1) \quad (\text{아티팩트의 스테이크 비중})$$

$$RS \in (0, 1) \quad (\text{롤링 기간 내 아티팩트 수익 비중})$$

$$W = \lambda V + x \cdot SWS + (1 - x - \lambda)RS,$$

단, $\lambda + x \leq 1$, $\lambda, x \in [0, 1]$.

거버넌스는 시간에 따라 지표 비중을 조정하기 위해 λ 와 x 를 튜닝한다.

이 메커니즘은 다음과 같은 순환 구조를 형성한다: 유용한 아티팩트는 사용자를 끌어들이고, 참여는 수익과 신호를 생성하며, 발행은 기여자와 스테이커를 보상하고, 네트워크가 진화함에 따라 거버넌스가 재조정한다.

거버넌스(Governance)

토큰 보유자는 직접 민주주의와 대표자 의사결정을 통해 프로토콜을 관리한다. DAO는 경제적 지렛대(발행률, 게이지 매개변수, 금고)를 통제한다. 대표자는 DAO가 선출한 검증자 또는 신뢰할 수 있는 커뮤니티 구성원으로, 스테이크를 잠금(lock)하여 상대적 발행 가중치를 할당한다. 모든 보유자는 제안을 생성하고 투표하거나 위임할 수 있다. 위임자는 대표자 간 스테이크를 이동시킬 수 있으며, DAO는 대표자를 임명하거나 해임할 수 있다. 견제와 균형 구조는 강건성과 적응성을 목표로 한다.

요약(Summary)

센티언트의 토크노믹스 설계는 “오픈소스 AI의 수익화, 기여자 조정, 가치 분배”라는 근본 목표를 유지하면서 새로운 프로토콜 기능을 통합한다. 배포는 광범위한 참여와 장기적 정렬을 보장하며, 토큰의 유틸리티는 연료이자 결속체 역할을 한다. 동적 발행 메커니즘은 성장을 자금 지원하고 이해관계자를 보상하며, 게이지 매개변수와 부트스트래핑 예산을 통해 적응한다. 계층적 거버넌스는 커뮤니티가 투명성과 책임성을 바탕으로 네트워크 진화를 주도하도록 하여, 커뮤니티 소유 지능 네트워크를 위한 탄력적이고 확장 가능한 경제 프레임워크를 구축한다.