

오리진트레일(OriginTrail)

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.

인공지능을 위한 검증 가능한 인터넷: 크립토, 인터넷, AI의 융합

(Verifiable Internet for Artificial Intelligence: The Convergence of Crypto, Internet and AI)

사전 공개본(Pre-publication)

Trace Labs

오리진트레일 코어 개발자

피드백 및 토론 : <https://discord.gg/9CvynQ9kYe>

“우리는 연결성이 풍부한 동시에 잘못된 정보 또한 넘쳐나는 시대에 살고 있다. 오리진트레일 탈중앙화 지식 그래프(DKG, Decentralized Knowledge Graph)는 지식 속에서 진실을 찾기 위한 진화하는 도구이다. 특히 우리는 지식 그래프가 인공지능의 신뢰도를 향상시킬 것이라 본다.”
Dr Bob Metcalfe, 이더넷 발명가이자 인터넷 개척자

초록(Abstract)

본 백서는 크립토, 인터넷, 인공지능 기술 간의 시너지를 활용하여 인공지능을 위한 검증 가능한 인터넷(Verifiable Internet for AI)이라는 개념을 통해 AI의 미래에 대한 비전을 제시한다. 이를 위해 AI 시스템이 활용하는 정보의 출처(provenance), 무결성(integrity), 검증 가능성(verifiability)을 보장하는 새로운 접근 방식인 탈중앙화 지식 그래프(DKG, Decentralized Knowledge Graph)를 소개한다. 본 백서는 Web3 기술과 신경망 기반 AI 및 기호 기반 AI 접근 방식을 결합함으로써, 잘못된 정보, 데이터 소유권, 지식재산권, AI 기술에 내재된 편향 문제를 해결하고자 한다. 또한 DKG의 정교한 작동 메커니즘, 지식 자산(Knowledge Assets)을 통한 AI 신뢰성 강화 방식, 자율 파라넷(autonomous paranets) 지원 구조를 심층적으로 다룬다. 인공지능을 위한 검증 가능한 인터넷은 단순한 기술적 진보를 넘어, 보다 투명하고 포용적이며 윤리적인 디지털 환경으로의 패러다임 전환을 의미한다. 이 비전은 블록체인 기술과 AI의 역동적 역량을 결합하여, 민주적이고 윤리적으로 설계되며 경제적으로 지속 가능한 AI 애플리케이션의 미래를 구축하고자 한다.

1 목적

주류 영역에서 인공지능(AI)이 급부상하면서, 우리는 지식과 상호작용하는 모든 방식이 재편되고 있다. 이는 우리가 지식을 발견하는 방식, 이를 활용해 가공하는 방식, 그리고 새로운 지식을 창출하는 방식에까지 변화를 가져온다. 이는 문명사적으로 전례 없는 사건으로, 인간이나 기계가 과거에 보유했던 어떤 방법보다 뛰어난 인지 역량을 얻게 되었음을 의미한다. AI는 방대한 데이터를 활용하여 기존에는 발견되지 않았을 패턴과 통찰을 식별할 수 있게 한다. 이를 통해 개인별로 맞춤화되고 지속적으로 자기 개선하는 적응형 솔루션을 구축할 수 있다.

기술이 강력해질수록, 그 이상 현상(anomalies) 역시 더욱 파괴적일 수 있다. 잘못된 정보의 문제는 더 이상 악의적인 선전물 인쇄에 국한되지 않는다. 소셜 네트워크를 정치적 조작에 악용하는 문제조차, 우리가 인지 기능을 “외주화(outsource)”하는 시스템에 대한 신뢰 부족에 비하면 미미해 보일 수 있다. 대량의 지식을 처리하고 우리의 행동에 대한 입력을 제공하거나 심지어 특정 행동을 자율적으로 수행하도록 신뢰하는 시스템은 최고 수준의 투명성과 검증 가능성을 요구한다. 데이터 소유권, 정보의 출처, 정보의 검증 가능성, 검열을 내재화한 설계(censorship-by-design) 형태의 편향 문제에 있어 어떠한 타협도 허용되어서는 안 된다. 이 혁명이 포용적으로 전개되지 않을 경우, AI 독점이 형성될 수 있다는 점은 사회적 위협이다.

본 백서는 인공지능을 위한 검증 가능한 인터넷의 구성 요소를 논의한다. Web3의 기본 원칙과 빠르게 확산되는 AI 시스템 간의 조화로운 발전을 도모하기 위해, 우리는 실질적인 단계를 제안한다. 이 접근 방식에서 우리는 오리진트레일 탈중앙화 지식 그래프(DKG), 이더리움(Ethereum), 폴카닷(Polkadot, 보다 구체적으로는 NeuroWeb)과 같은 핵심 Web3 기술을 OpenAI, Google, Microsoft, X의 GrokAI 등 다양한 AI 시스템과 통합하고자 한다. 우리는 중립성, 포용성, 사용성을 원칙으로 하는 인공지능을 위한 검증 가능한 인터넷을 구축함으로써 신뢰 가능한 AI의 잠재력을 실현할 수 있다고 믿는다.

2 인공지능과 그 이상 현상

인공지능(AI)은 지능형 기계를 개발하고 연구하는 컴퓨터 과학 분야이다. 이는 특히 컴퓨터 시스템을 통해 인간 지능의 과정을 모방하는 것이다. 생성형 인공지능(GenAI, Generative AI)은 텍스트, 이미지, 오디오 및 기타 합성 데이터 등 다양한 유형의 콘텐츠를 생성할 수 있는 AI 기술이다. 이는 생성형 신경망 모델을 활용하여 주어진 프롬프트에 응답하는 새로운 콘텐츠를 생성한다. 이러한 기술은 OpenAI의 ChatGPT, Google의 Gemini, Meta의 LLaMA와 같은 대규모 언어 모델(LLM)에 의해 널리 대중화되었다. 이와 같이 광범위하게 학습된 LLM의 성능은 매우 빠른 속도로 향상되었으며, 다양한 주제에 대해 지능적인 응답을 제공할 수 있게 되었다. 이러한 모델이 경제에 미치는 단독 영향은 McKinsey에 의해 4.4조 달러로 추정된 바 있다 [1]. 생성형 AI가 잠재력을 충분히 발휘하고 나아가 새로운 성장 잠재력을 개방하기 위해서는 다음과 같은 이상 현상이 해결되어야 한다.

환각(Hallucinations). 이름에서 알 수 있듯이, 생성형 AI는 결정론적 시스템이 아니며, 모델이 학습한 방대한 데이터에 기반하여 응답을 생성한다. 따라서 이는 가장 적절한 응답을 “추측”하는 데 능한 확률적 시스템이다. 고품질 응답을 예측하는 데 매우 뛰어날 수 있지만, 사용자가 제시한 질문에 대해 반드시 정확한 답을 제공한다는 보장은 없다. 이러한 잘못된 예측을 “환각”이라 하며, 이는 생성형 AI 시스템이 만들어낸 오류 응답을 의미한다. 환각은 때로는 흥미롭거나 영감을 줄 수 있으나, 특히 정확성이 매우 중요한 상황(예: 교육, 의료, 공급망, 과학, 건설 등)에서는 일반적으로 바람직하지 않은 결과이다.

데이터 소유권 및 지식재산권(Data ownership and intellectual property). 생성형 AI 모델이 학습하는 방대한 데이터셋에는 저작권이 있는 자료가 포함될 수 있으며, 해당 소유자가 명시적인 사용 허가 또는 라이선스를 부여하지 않았을 가능성이 있다. 이는 (1) 해당 데이터를 생성형 AI 모델의 학습에 사용하는 행위의 합법성 문제와 (2) 생성된 결과물(예: 예술 작품, 텍스트)의 지식재산권 귀속 문제를 제기한다. 개인 데이터나 기업의 민감한 데이터가 포함될 경우, 상당한 개인정보 보호 문제도 발생할 수 있다. 더 나아가, 모델에서 특정 저작권 콘텐츠를 분리하는 작업은 재학습(retraining) 또는 미세조정(fine-tuning)을 통해서만 가능하며, 이는 모두 자원 집약적인 과정이다. 데이터 삭제(data-forgetting) 기법 역시 제한적인 효과만을 제공한다.

편향(Bias). 생성형 AI 모델은 방대한 데이터를 기반으로 학습되며, 이 과정에서 데이터에 내재된 편향을 함께 흡수한다. 이는 프롬프트 응답에 체계적 편견이 반영되는 결과를 초래하며, 사회적 차원에서 인종, 국적, 성별, 정치적 편향을 포함한 다양한 문제를 야기할 수 있다. 또한 기업용 모델에서 특정 공급자를 선호하는 등 소규모 편향이 발생할 수도 있다. 기존 편향을 제거하기 위해서는 재학습 또는 미세조정이 필요하며, 이는 데이터 분리 기법과 마찬가지로 상당한 자원을 요구한다.

모델 붕괴(Model collapse). 생성형 AI, 특히 LLM의 성능은 학습 데이터의 품질에 결정적으로 의존한다. 데이터는 관련성이 높고 다양해야 모델이 학습을 통해 더욱 강력해질 수 있다. 만약 모델이 자체 생성한 데이터(합성 데이터)를 학습에 활용하는 상황이 발생한다면, 모델은 단순히 성능이 정체되는 것이 아니라 오히려 퇴화하기 시작한다 [2]. 합성 데이터를 학습에 사용할 경우 기존 편향이 강화되고, 데이터의 “꼬리(tails)” 부분이 최소화된다. 여기서 꼬리란 발생 가능성은 낮지만 중요한 데이터 요소를 의미하며, 이는 LLM 출력의 다양성을 확보하는 데 필수적이다. 이러한 요소가 제거되면 데이터 오해석이 발생하고, 모델은 붕괴에 가까운 상태로 향할 수 있다. AI 활용이 인터넷 전반에 확산되면서 이 위협은 더욱 커지고 있다. LLM 학습 데이터의 주요 출처는 인터넷 크롤링이기 때문이다. 만약 인터넷상의 콘텐츠 대부분이 AI 생성 콘텐츠(합성 데이터)가 된다면, 해당 데이터의 학습 가치는 감소하고 활용 위험은 증가한다. 반대로, AI가 생성하지 않은 실제 세계 데이터임이 검증 가능한 데이터의 가치는 더욱 상승하게 된다.

3 인공지능을 위한 검증 가능한 인터넷

가까운 미래에는 AI 모델이 활동과 콘텐츠 생산 측면에서 인터넷을 사실상 지배하게 될 것이다. OpenAI가 생성형 AI 대규모 언어 모델 ChatGPT를 출시하며 촉발한 파괴적 변화는, 주요 기술 기업들로 하여금 그동안 개발해 온 AI 기술을 공개하도록 만들었다(예: Google의 Gemini 출시, X의 GrokAI, Meta의 LLaMA 진전 등). 더 나아가 오픈소스 커뮤니티 역시 독점 모델을 따라잡기 위해 비약적인 발전을 이루었으며, 인공지능을 위한 검증 가능한 인터넷을 구성하는 유용한 요소들을 성공적으로 제공하고 있다.

그러나 텍스트 기반 LLM만이 생성형 AI 모델의 전부는 아니다. 예를 들어 다음과 같은 모델들이 존재한다:

- 이미지 생성(예: Stable Diffusion) — 텍스트, 이미지, 텍스트와 이미지, 또는 이미지와 이미지 입력을 기반으로 이미지를 생성하는 기능,
- 비디오 생성(예: OpenAI Sora) — 텍스트 프롬프트를 통해 최대 60초 길이의 짧은 영상을 생성하는 기능,
- 멀티모달 모델(예: OpenAI CLIP) — 여러 유형의 미디어를 입력으로 수용하고, 여러 유형의 미디어를 출력으로 제공할 수 있는 모델.

따라서 AI가 지배하는 인터넷의 미래에서는, 사람들은 단일 모델과의 상호작용부터 다수의 에이전트 역량을 활용하여 자율적 활동을 수행하는 에이전트 기반 애플리케이션에 이르기까지, 다양한 AI 모델과 상호작용하게 될 것이다. AI의 대중화가 지속됨에 따라, 이러한 세계는 앞서 언급한 현재 AI 시스템의 이상 현상을 효과적으로 해결할 수 있는 구조를 절실히 요구한다.

우리는 인공지능을 위한 검증 가능한 인터넷의 핵심 속성으로 다음 세 가지를 제시한다: 정보의 출처(provenance), 정보의 검증 가능성(verifiability), 그리고 지식 공유에 대한 인센티브 구조.

정보의 출처(Information provenance).

의미 있는 방식으로 대규모 운영을 수행할 AI 시스템은 반드시 정보 출처의 투명성을 보장해야 한다. 특정 프롬프트에 대한 응답이 검증 가능한 알려진 출처로부터 생성되었음을 확인할 수 있어야 한다. 최소한 해당 출처는 게시자의 신원, 게시 시점, 관련 콘텐츠(지식)에 대한 정보를 포함해야 한다. 생성형 AI의 응답을 알려진 출처에 기반하게 하는 것은, AI 엔지니어와 지식 소비자 모두에게 환각을 관리하는 가장 효과적인 방법이다. 이제 사용자는 AI 응답을 맹목적으로 신뢰하는 대신, 해당 응답이 생성된 실제 출처와 직접 비교할 수 있다. 즉, 정보를 “직접 원천에서” 확인할 수 있게 된다. 정보 출처의 투명성은 또한 적절한 지식재산권 보호를 보장하는 데 핵심적이다. 이는 단순히 응답에 대한 출처를 명시하는 수준에서부터, 원본 데이터를 기반으로 생성된 결과물에 대한 로열티 분배와 같은 보다 복잡한 구현까지 확장될 수 있다(예: 원본 예술 작품을

기반으로 생성된 디지털 아트). 마지막으로, 정보 출처의 투명성은 편향 문제를 완화하는 강력한 도구이기도 하다. 출처 자체가 편향을 내포할 수는 있지만, 사용자가 출처에 접근하고 이를 필터링할 수 있는 역량을 가짐으로써, 제시된 응답에 대해 비판적 사고를 형성할 수 있다.

검증 가능한 정보(Verifiable information).

단순히 출처를 찾는 것만으로는 충분하지 않다. 생성형 AI가 소비하는 콘텐츠가 변조되지 않았는지 또한 확인할 수 있어야 한다. 생성형 AI 솔루션이 더욱 광범위하게 사용될수록, 이들은 제3자 출처를 포함한 방대한 글로벌 지식과 상호작용하게 된다. 이는 응답 생성에 사용되는 지식 네트워크의 복잡성을 크게 증가시킨다. 이러한 환경에서는, 사용된 콘텐츠가 출처를 가질 뿐 아니라 생성 이후 변경되지 않았음을 보장하는, 즉 무결성을 보장하는 암호학적 검증 수단이 필요하다.

고품질 지식에 대한 인센티브(Incentivizing high-quality knowledge).

협력적 방식으로 구축된 공유 오픈 지식에 대한 접근은 AI의 미래에 있어 핵심적이다. 특히 가까운 시기에 비(非)AI 생성 콘텐츠의 규모가 AI 생성 합성 콘텐츠에 의해 추월될 것으로 예상되는 상황에서, 이는 더욱 중요하다. 딥러닝 분야 튜링상 수상자인 Yann LeCun 역시 이를 강조한 바 있다:

“그 (AI) 시스템을 학습시키는 방식은 클라우드소싱 되어야 한다... 그것이 모든 인간 지식의 저장소가 되기를 원한다면, 모든 인간이 여기에 기여해야 한다.”

그러나 필요성이 존재하더라도, 오픈 지식은 종종 “공유지의 비극(tragedy of the commons)” 딜레마에 직면한다. 오픈 지식에 기여하는 동기는 사적 지식 저장소를 확장하는 것에 비해 덜 매력적으로 인식될 수 있다. 따라서 우리는 다음과 같은 이중 인센티브 모델이 필요하다고 본다:

- 오픈 지식 창출에 대한 인센티브 — 특히 공공 영역(public domain)에 속하는 오픈 지식을 게시하는 행위를 장려함으로써, 공유지의 비극을 인센티브 확보를 위한 경쟁으로 전환할 수 있다. 내가 오픈 지식을 추가하지 않으면, 다른 누군가가 이를 수행하게 된다.
- 독점적 지식의 수익화 — 공공 영역에 속하지 않는 데이터 또는 지식(예: 익명화된 개인 데이터, 기업의 노하우)을 기반으로 한 비즈니스 모델의 발전을 허용한다.

이 두 가지 경우 모두, 지식의 품질을 측정할 수 있는 체계가 존재할 때 인센티브 분배의 기준을 마련할 수 있으며, 이는 매우 중요한 요소이다.

4 인공지능을 위한 검증 가능한 인터넷에서의 지식 출처 보장

위에서 제시한 요구사항을 충족하기 위해, 우리는 탈중앙화 검색 증강 생성(dRAG, Decentralized Retrieval Augmented Generation) 프레임워크에서 다수의 탈중앙화 시스템을 활용하는 새로운 패러다임을 제안한다.

검색 증강 생성(RAG, Retrieval Augmented Generation)이라는 용어는 2020년 Patrick Lewis의 논문에서 처음 제안되었으며 [3], 이는 외부 출처로부터 사실 정보를 검색하여 생성형 AI 모델의 정확성과 신뢰성을 향상시키는 기술을 의미한다. 이를 통해 AI 솔루션은 생성 과정 이전에 관련 정보를 동적으로 검색하여, 생성 과정을 검색된 입력을 재구성하는 방식으로 제한함으로써 응답의 정확성을 높일 수 있다.

dRAG는 외부 출처를 탈중앙화 지식 그래프(DKG, Decentralized Knowledge Graph)로 조직화함으로써 RAG 모델을 확장하고, AI 모델이 활용할 수 있는 글로벌 클라우드소싱 지식 네트워크의 성장을 촉진하는 인센티브 구조를 도입한다. 이 프레임워크는 신경망 기반 AI(예: LLM)와 기호 기반 AI(예: 지식 그래프)를 결합한 하이브리드 AI 시스템을 구현한다. 벡터 임베딩 표현에 기반한 순수 신경망 접근과 달리, 기호 기반 접근은 지식 그래프의 강점을 통해 상징적 표현에 기반한 구조를 도입함으로써 이를 보완한다.

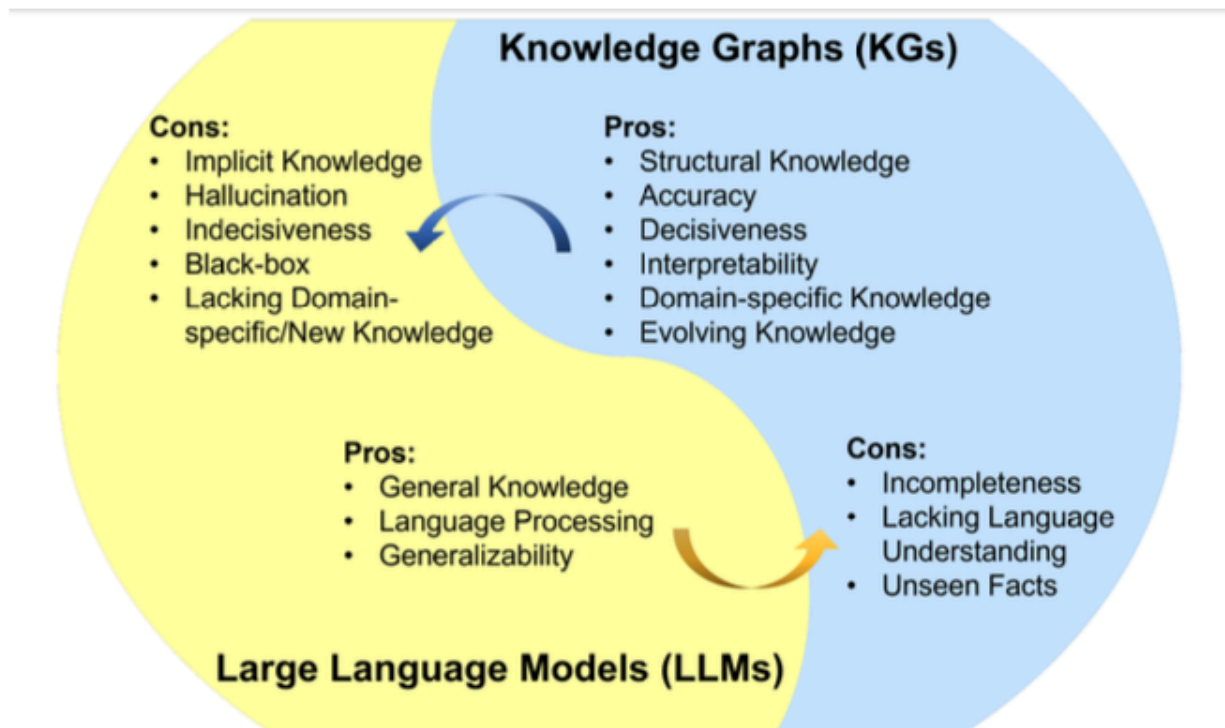


그림 1: 신경망(LLMs)과 기호 기반(KGs) AI 방법론 간의 상호작용. 출처: <https://arxiv.org/pdf/2306.08302.pdf> [4]

따라서 dRAG는 신경망의 강력한 학습 및 일반화 능력과, 기호 기반 AI의 정밀하고 규칙 기반 처리 역량을 동시에 활용할 수 있도록 하는 프레임워크이다. 이는 두 가지 핵심 구성 요소를 기반으로 작동한다: (1) DKG 파라넷(paranets)과 (2) AI 모델. dRAG 애플리케이션 프레임워크는 기존 기술, 도구 및 RAG 프레임워크와 완전히 호환되며, 주요 데이터 형식을 모두 지원한다.

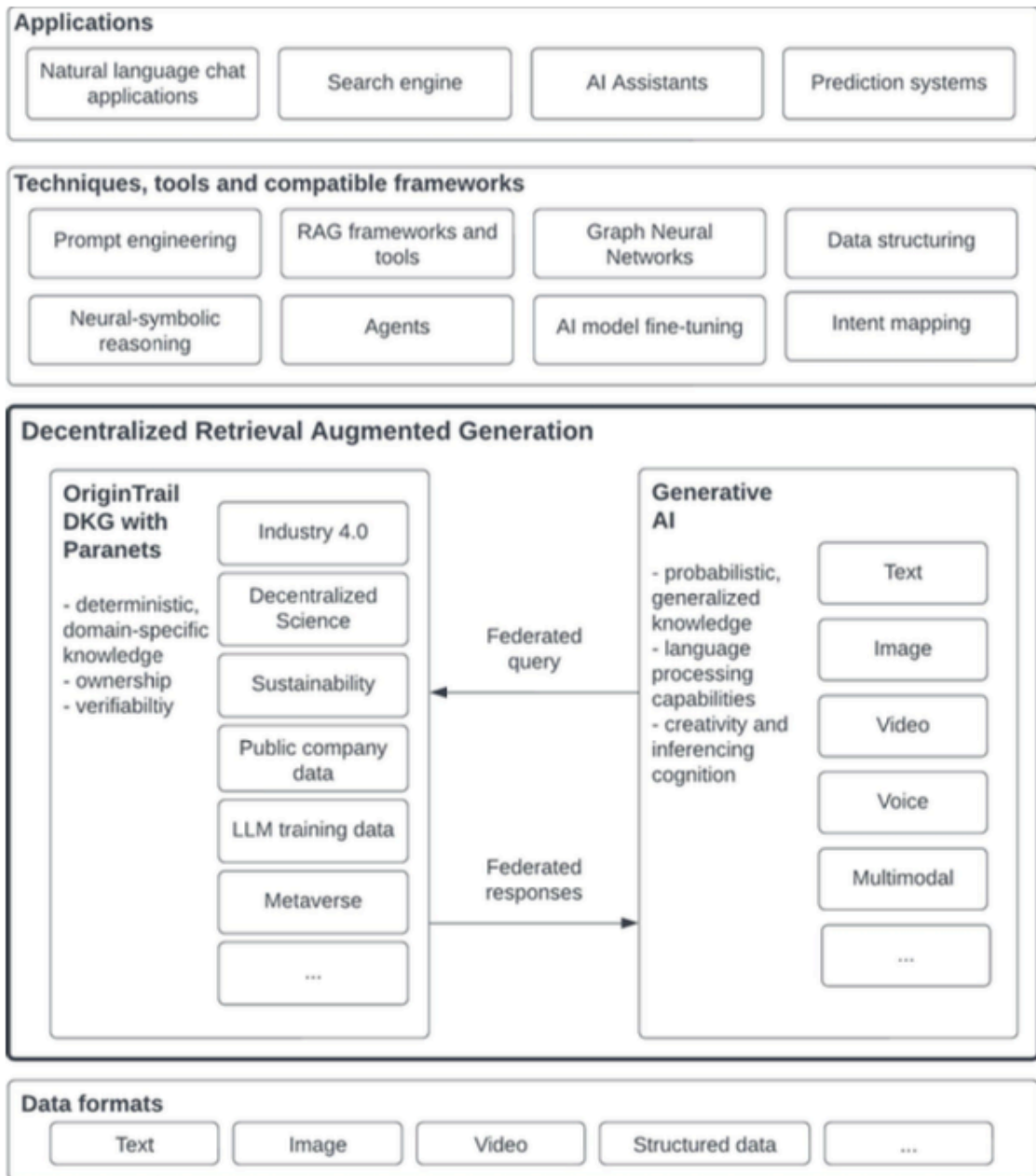


그림 2: dRAG 프레임워크 애플리케이션 아키텍처

4.1 탈중앙화 지식 그래프

오리진트레일 탈중앙화 지식 그래프(DKG, Decentralized Knowledge Graph)는 다수의 블록체인과 통합된, 개방형이자 무허가(permissionless) 탈중앙화 지식 그래프 노드 네트워크이다. 이는 검증 가능한 인터넷 인프라의 핵심 구성 요소로서, AI 솔루션이 DKG의 발견 가능성(discoverability), 검증 가능성(verifiability), 소유권 기능을 활용할 수 있도록 한다. DKG의 핵심 자원은 지식 서비스와 함께 DKG 파라넷(paranets)에 조직된 지식 자산(Knowledge Assets)이다.

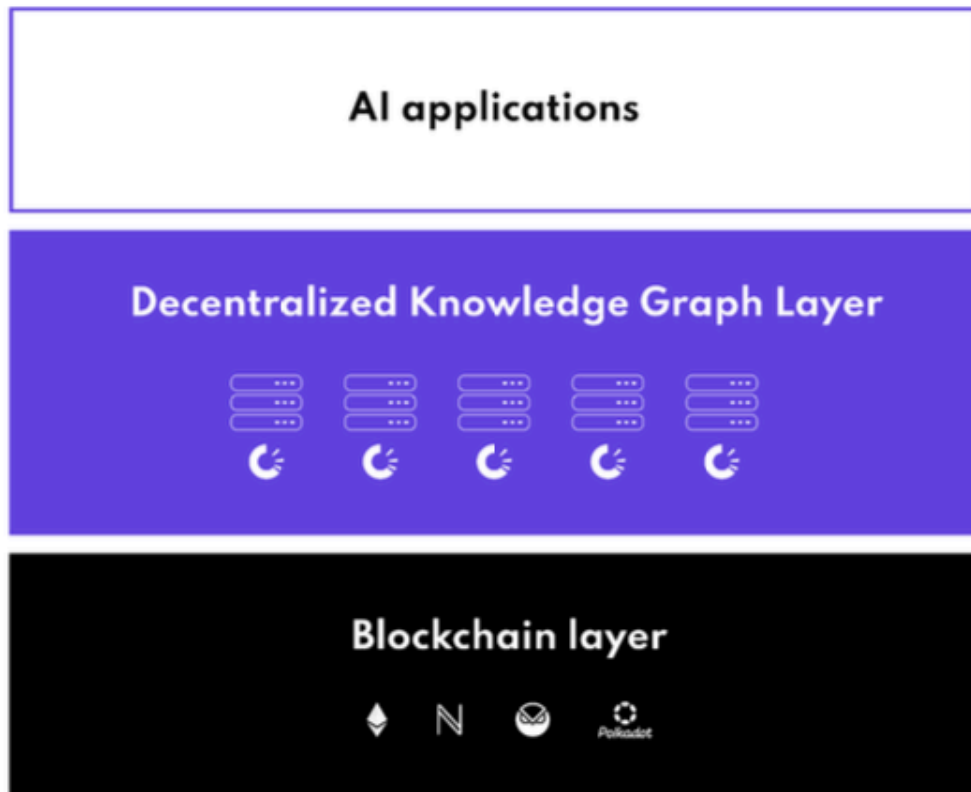


그림 3: 탈중앙화 지식 그래프의 3계층 개념 아키텍처

4.1.1 지식 자산을 통한 AI 신뢰성 강화

지식 자산(Knowledge Asset)은 DKG 내에서 AI 시스템이 소비할 수 있는 기본 자원이다. 이는 지식 그래프 트리플, 벡터 임베딩, 이미지 등 다양한 형식의 지식을 담을 수 있는 소유 가능한 컨테이너를 의미한다. 지식 자산은 DKG 프로토콜이 정의한 연산을 통해 단일 트랜잭션으로 생성, 업데이트, 이전될 수 있다. 각 지식 자산은 소유권, 발견 가능성, 검증 가능성이라는 세 가지 핵심 속성을 가진다.

소유권(Ownership).

각 지식 자산은 블록체인 상에서 발행된 대체 불가능 토큰(NFT)으로 구현되는 소유자를 가진다. 소유자는 지식 자산의 콘텐츠를 업데이트하거나, 비공개 또는 공개 공유 상태를 변경하거나, 소유권을 이전하는 등 자산을 관리할 수 있다. 이는 지식을 하나의 자산 클래스(asset class)로 전환시킨다. 소유권이 블록체인 레벨에서 구현되므로, AI 모델 및 에이전트에 의한 지식 자산의 매수·매도를 포함한 원활한 수익화가 가능하다.

발견 가능성(Discoverability).

지식 자산은 그 콘텐츠 구조와 DKG 내 가용성으로 인해 본질적으로 발견 가능하다. 이는 DKG를 검열 없는 발견 가능성을 갖춘 글로벌 탈중앙화 지식 기반으로 만든다. 지식은 디리퍼런싱(dereferencing), 지식 그래프 쿼리, 벡터 유사도 검색 등 다양한 방법을 통해 탐색될 수 있다. 링크드 데이터 원칙에 따라, 지식 자산은 의미 있는 연결을 형성할 수 있다. 웹사이트가 URL(Uniform Resource Locator)로 연결되는 것과 유사하게, 지식 자산은 UAL(Uniform Asset Locator)과 URL을 통해 다른 지식 자산의 콘텐츠를 참조할 수 있으며, 이를 통해 기존 월드 와이드 웹 및 인터넷 인프라와 상호운용성을 갖는다. DKG를 통해 이러한 연결은 기호 기반 AI 맥락(symbolic AI context)으로 강화되며, 이는 신경망 기반 AI 기법과의 결합에 이상적이다.

검증 가능성(Verifiability).

지식 자산은 블록체인에 저장된 지식 상태(다이제스트, digest)에 대한 머클 트리(Merkle-tree) 기반 암호학적 증명을 포함한다. 지식 자산이 업데이트를 통해 진화할 때마다 각 증명이 기록되므로, 모든 지식 자산 관련 연산은 블록체인 상에서 투명하고 감사 가능하다. 검증 가능성은 개별 콘텐츠 수준(예: 지식 포함 증명)뿐 아니라 전체 지식 자산 수준에서도 지원된

다. 지식 자산은 W3C Verifiable Credentials 데이터 모델 [5]을 따르며, DKG는 검증 가능한 데이터 레지스트리 기능을 수행하고, 지식 자산 내부의 지식으로부터 검증 가능한 프레젠테이션을 생성할 수 있다. 이를 통해 AI 시스템은 소비 이전(dRAG, 학습 등)에 검증 가능성을 확립할 수 없는 콘텐츠를 사전에 필터링할 수 있다.

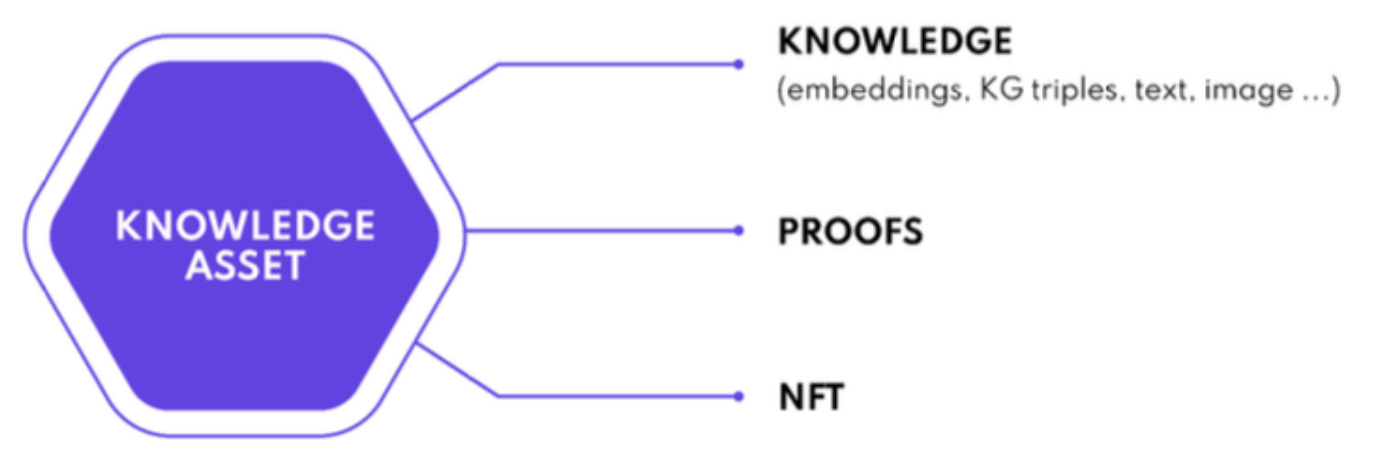


그림 4: 지식 자산의 핵심 요소

지식 자산은 dRAG 프레임워크에서 AI 시스템이 활용할 수 있는 신뢰 가능한 참조 자원이다. 따라서 지식 자산에 포함되는 구조화된 지식의 세분화 수준은 특정 활용 사례가 참조해야 하는 범위에 따라 달라진다. 이는 단일 명제(statement) 또는 불변성 증명에서부터 전체 데이터베이스에 이르기까지 확장될 수 있다. 지식 자산의 집합은 자율적으로 운영되는 파라 네트워크, 즉 파라넷(paranets)을 형성하는 데 활용될 수 있다.

4.1.2 자율 파라넷

파라 네트워크 또는 파라넷(paranets)은 DKG 내에서 커뮤니티가 소유하는 자율 운영 단위이다. 파라넷에는 특정 활용 사례를 구동하는 지식 자산의 집합과, 해당 파라넷에 특화된 AI 서비스 및 성장을 촉진하는 지식 채굴자(knowledge miners)에 대한 인센티브 모델이 포함된다.

각 파라넷은 다음과 같은 요소들로 구성된다:

- 지식 자산(Knowledge assets): 파라넷에 포함되기 위해 지식 채굴자가 준수해야 하는 기대 속성을 포함한다(예: 특정 주제에 관한 지식 포함, 정의된 온톨로지에 따른 데이터 구조 등).
- AI 서비스(AI services): dRAG 인터페이스, AI 에이전트, 스마트 컨트랙트, 데이터 오라클 등.
- 인센티브 모델(Incentivization model): 지식 채굴 및 파라넷 특화 AI 서비스 등 파라넷 성장 활동에 대한 보상 규칙을 명시한다.
- 지원 블록체인(A supported blockchain): 파라넷이 운영되며 지식 자산을 집적하는 블록체인.

지식 자산의 매개변수 및 서비스 제공 방식 등을 포함한 파라넷의 특성은 모두 파라넷 운영자에 의해 정의된다. 운영자는 개인, 조직, 또는 탈중앙화 자율 조직(DAO)이 될 수 있다. 여러 파라넷이 함께 DKG를 구성하며, 공통의 기저 네트워크 인프라를 활용한다. DKG는 무허가 시스템이므로 누구나 파라넷을 개시할 수 있다.

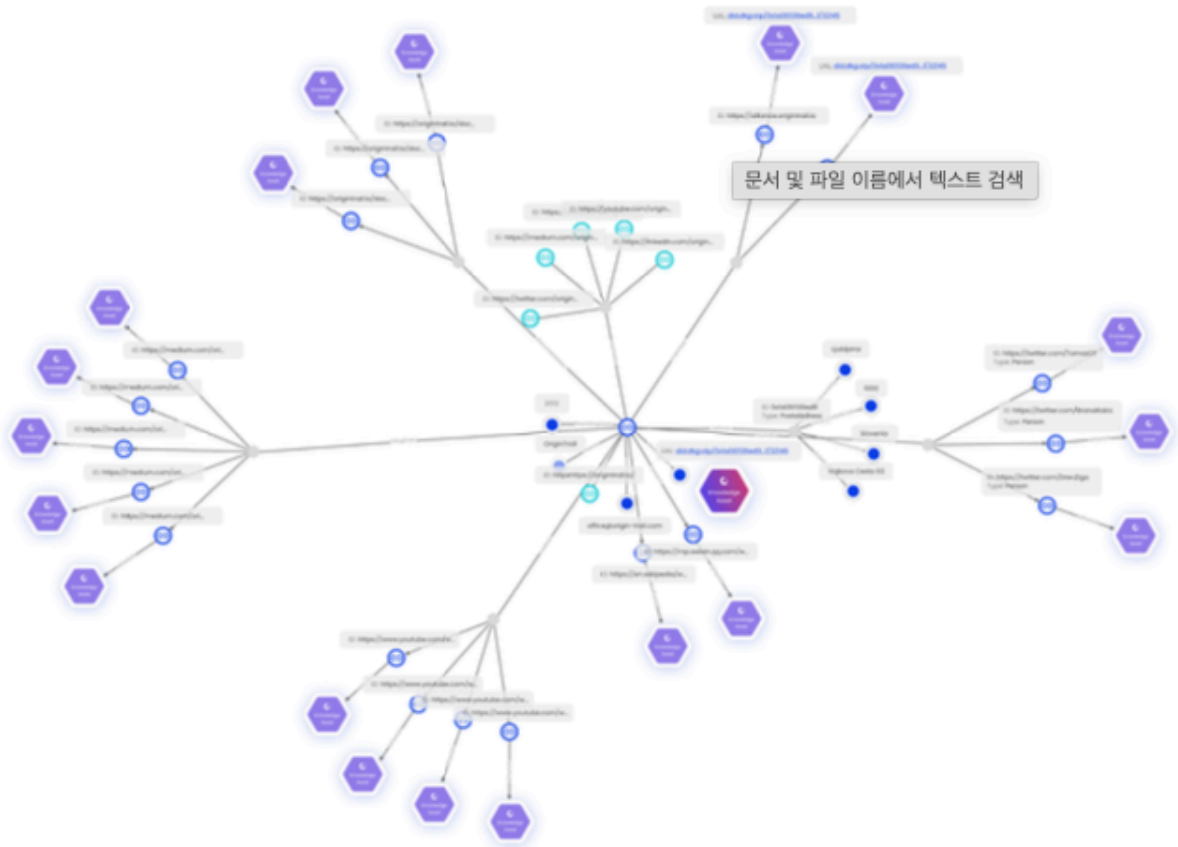


그림 5: 지식 자산 집합으로서의 파라넷

파라넷은 AI 시스템에 강력한 기반(substrate)을 제공한다. 이는 다수의 출처로부터 검증 가능한 입력을 활용하는 네트워크 효과(network effects)를 통해 dRAG를 사용하여 정확한 응답을 도출하도록 지원한다. AI 시스템은 (1) 공개 지식 그래프와 (2) 접근 권한이 있는 관련 지식 자산에 포함된 비공개 지식으로부터 정보를 수집할 수 있다.

파라넷은 다음과 같이 다양한 성격과 규모의 주제를 중심으로 형성될 수 있다:

- 인더스트리 4.0
- 탈중앙화 과학(Decentralized Science)
- 지속가능성(Sustainability)
- 상장 기업 보고서
- LLM 학습 데이터
- 메타버스
- 영상 콘텐츠 네트워크
- 소셜 미디어
- 예술 발견 네트워크
- 예측 시장
- 스포츠 및 베팅
- 엔터테인먼트

위 예시는 DKG에서 진행 중인 채택 사례에서 영감을 받은 것이지만, 목록은 결코 포괄적이지 않으며, DKG가 지원하고 AI 시스템에 제공할 수 있는 자물 파라넷의 다양성과 포용성을 보여주기 위한 예시일 뿐이다. 모든 파라넷은 함께 DKG의 일부를 이루며, 조합하여 쿼리 및 접근할 수 있다.

5 NeuroWeb에서 인공지능을 위한 검증 가능한 인터넷 인센티브

dRAG 프레임워크는 AI의 주요 과제를 해결하는 강력한 접근 방식을 제공하지만, AI 도구가 DKG에서 접근할 수 있는 지식의 양이 증가할수록 dRAG 활용의 가치는 기하급수적으로 증가한다. 채택 곡선을 더욱 가파르게 만들기 위해서는 인센티브 메커니즘이 핵심 요소이다. 인공지능을 위한 검증 가능한 인터넷의 인센티브 구조는 파라넷 운영자가 NeuroWeb 인센티브를 활용하여 파라넷 성장을 촉진할 수 있도록 함으로써 구현된다. 특정 파라넷에서 지식 채굴을 유도하기 위한 발행 규모와 제안된 규칙은 탈중앙화 거버넌스를 통해 결정된다. 파라넷에서 인센티브를 획득하기 위해 지식을 생성하는 핵심 활동은 지식 채굴(Knowledge Mining)이다.

5.1 지식 채굴(Knowledge Mining)

자물 파라넷을 출시할 때, 파라넷 운영자는 핵심 특성을 정의한다. 그중 가장 중요한 요소 중 하나는 해당 파라넷에서의 지식 채굴을 위한 인센티브 메커니즘이다. 이러한 메커니즘에는 주제 관련성, 온톨로지 준수, 크기 제한 등 다양한 조건이 포함될 수 있다. 지식 채굴자뿐 아니라, 인센티브 메커니즘은 서비스 제공자에게도 보상을 제공할 수 있다. 서비스의 범위는 단순한 온톨로지 준수 검증에서부터 AI 기반 챗 애플리케이션에 이르기까지 다양하다.

파라넷은 누구나 제안하고 운영할 수 있으며, 운영자가 적절하다고 판단하는 어떤 규모로도 설계될 수 있다. 즉, 좁은 범위와 넓은 범위 모두에서 기능할 수 있다. 파라넷 운영자는 NeuroWeb 거버넌스 투표를 통해 다른 파라넷과 보상 확보를 경쟁한다. 각 운영자는 자신이 필요로 하는 보상 발행 규모, 지식 채굴 메커니즘, 인센티브 대상 서비스 및 기타 운영자 또는 투표에 참여하는 NeuroWeb 커뮤니티가 중요하다고 판단하는 핵심 요소를 명시한 보상 제안을 제출한다.

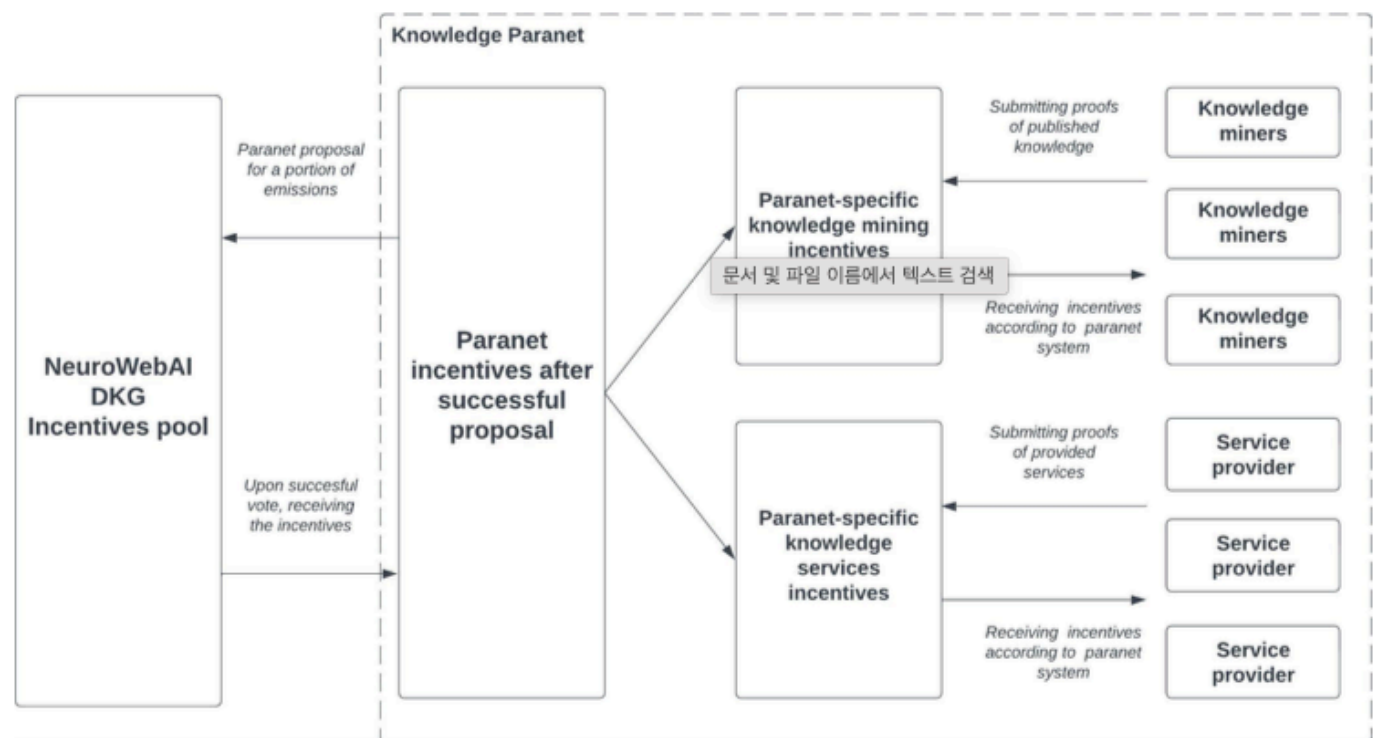


그림 6: 자물 파라넷에서의 인센티브 흐름

이와 같은 구조를 통해, 관련 도메인에 특화된 지식을 DKG에 게시하려는 지속적인 동기가 형성되며, 이는 dRAG 솔루션을 구축하는 주체들에게 DKG의 가치를 증대시킨다.

6 AI 기반 자율 DKG(AI-powered Autonomous DKG)

DKG, dRAG, 그리고 인공지능을 위한 검증 가능한 인터넷 인센티브의 결합은 글로벌 지식의 지속적인 성장을 지원하는 구조를 형성한다. 지식 채굴자는 초기에는 수동 및 자동화 기법을 활용하여 관련 출처로부터 데이터를 수집함으로써 특정 파라넷에서 지식 채굴을 시작하게 된다. 그러나 파라넷에 더 많은 지식이 축적되면, 자율적 지식 채굴의 가능성이 열리게 된다. 고품질의 주석 처리된(annotation) 온톨로지 적합 데이터가 파라넷에 존재하면, 우리는 연역 추론(deductive reasoning)을 수행할 수 있다. 이는 기존 지식에 온톨로지 규칙과 지식 간 연결/관계를 적용하여 새로운 지식을 결정론적으로 도출하는 논리적 과정이다. 자율 지식 그래프는 또한 귀납 추론(inductive reasoning)의 기회를 제공한다. 이는 AI를 활용하여 기존 지식에서 패턴과 규칙성을 식별하고, 최적의 새로운 지식을 예측함으로써 확률적으로 새로운 지식을 생성하는 방식이다.

이 분야를 발전시키는 중요한 연구 흐름 중 하나는 지식 그래프 추론을 위한 파운데이션 모델(foundation models for knowledge graph reasoning) [6]이다. 이는 AI 시스템에서 사용되는 파운데이션 모델을 지식 그래프 구조에 결합하는 접근이다. 이를 통해 보편적이고 전이 가능한 그래프 표현이 가능해지며, AI 시스템 [7]이 자신의 파운데이션 모델을 활용하여 어떤 지식 그래프든 추론하고 채울 수 있게 된다. 동시에 기존 지식 그래프의 표현을 소비함으로써 모델을 미세조정(fine-tuning)할 수도 있다. 특히 이러한 예측 작업에서 강력한 성능을 보이는 것이 앞서 언급한 그래프 신경망(GNNs, Graph Neural Networks)이며, 이는 GNN의 주요 활용 사례 중 하나이다.

자율 탈중앙화 지식 그래프는 논리적 추론과 확률적 추론을 모두 효과적으로 활용하여 DKG를 풍부하게 확장하고, 관계와 개체에 대한 이해를 심화시키며, 더 정교한 예측과 추천을 수행함으로써 새로운 지식을 자율적으로 생성하게 될 것이다.

7 인공지능을 위한 검증 가능한 인터넷의 경제 구조

검증 가능한 인터넷은 보다 포용적이고 민주적인 AI라는 비전을 향해 일관된 접근 방식을 제시하며, 그 단위 경제(unit economics)는 다음 두 가지 기본 개념에 의해 구동된다.

1. AI 준비형 지식 자산(AI-ready Knowledge Assets)
2. 지식 파라넷(Knowledge paranets)

이 각 단위는 오리진트레일 DKG 네트워크의 보안성, 복원력, 사용성을 강화하도록 설계된 토크노믹스 구조에 의해 지원된다. NeuroWeb 블록체인 상에서 실행되는 인센티브 구조는 DKG 네트워크 성장에 기여하는 활동에 매력적인 보상을 제공하며, 인공지능을 위한 검증 가능한 인터넷의 기반으로 지식 구축을 촉진하는 추가적인 기회를 창출한다.

7.1 인공지능을 위한 검증 가능한 인터넷 인프라를 구동하는 TRAC 유틸리티 토큰

지식 자산은 인공지능을 위한 검증 가능한 인터넷에서 논의되는 가장 기본적인 단위 경제 개념이다. 각 지식 자산이 생성될 때마다 오리진트레일 DKG 내 특정 서비스 단위가 소비되며, 이는 노드 운영자와 위임 스테이커에 의해 무허가 방식으로 제공된다. 지식 게시자가 DKG에 새로운 지식 자산을 추가할 때마다, 그들은 DKG 네트워크 서비스 이용에 대한 보상으로 TRAC를 사용하며, 트랜잭션에는 별도의 가스 토큰을 사용한다. 가스 토큰은 지식 자산을 게시하는 데 사용되는 블록체인에 따라 달라진다(예: NeuroWeb에서는 NEURO, Gnosis에서는 DAI, Ethereum에서는 ETH 등).

DKG 네트워크의 노드들은 지식 게시자가 지불하는 게시 수수료를 획득하기 위해 경쟁한다. 이들은 다음 세 가지 핵심 요소를 기준으로 경쟁한다: (1) 파라넷을 위한 적절한 서비스 제공(게시된 지식의 저장 및 가용성 보장), (2) 해당 노드에 스테이킹된 TRAC의 규모, (3) 네트워크 전반에 걸친 지식 및 보상의 효율적 분배를 위한 노드 파라넷 주소. 이러한 기준에 따라 가장 높은 점수를 획득한 노드는 지식 자산이 게시된 기간 동안 해당 보상의 일부를 수령할 자격을 가진다.

TRAC 스테이킹은 노드 성공의 핵심 요소이므로, 노드 운영자는 다른 TRAC 보유자가 자신의 노드에 TRAC을 위임 (stake delegation) 할 수 있도록 허용할 수 있다. 위임된 스테이킹에 대한 대가로, 노드 운영자는 위임 규모에 비례하여 노드 보상의 일부를 위임자에게 분배한다. TRAC 스테이킹은 또한 보안 측면에서 중요한 역할을 수행한다. DKG 노드가 서비스를 적절히 수행하지 않거나 악의적으로 행동할 경우, 해당 스테이크는 슬래싱(slashing)될 수 있다.

7.2 NeuroWeb의 네이티브 토큰 NEURO - 인공지능을 위한 검증 가능한 인터넷 인센티브

앞서 설명한 바와 같이, 지식 파라넷(Knowledge paranets)은 특정 주제에 대한 지식 자산과 관련 서비스의 집합이다. 파라넷은 파라넷 운영자(paranet operators)에 의해 제안되며, 운영자는 해당 파라넷의 핵심 규칙을 정의한다. 기술적 규칙(예: 온톨로지, 데이터 처리 규칙, 서비스 요구사항 등) 외에도, 파라넷 운영자는 파라넷의 인센티브 시스템을 정의할 수 있다. 운영자는 지식 채굴(새로운 지식 자산 게시), 지식 검증 서비스(요구되는 데이터 품질 보장), 지식 접근 서비스(예: AI 발견 가능성 서비스, 공개 인터페이스, AI 에이전트 서비스) 또는 기타 파라넷 특화 서비스를 인센티브 대상으로 선택할 수 있다.

인센티브 구조가 정의되면, 해당 파라넷 제안은 NeuroWeb 블록체인 상 DKG 인센티브 풀에 대한 거버넌스 투표에 제출될 수 있다. NEURO 토큰 커뮤니티는 일종의 (탈)중앙화된 은행 역할을 수행하며, 지식 경제의 어떤 영역이 생태계에 가장 긍정적인 효과를 창출할지를 판단하고, 제안된 NEURO 발행량을 승인할지 여부를 투표로 결정한다. 중요한 점은, 다른 블록체인에서 운영되는 파라넷에 대해서도 활동에 대한 검증 가능한 증명을 제공하면 NEURO 인센티브를 발행할 수 있다는 것이다. 다만 NEURO 보상은 항상 NeuroWeb 상에서 수령되어야 하며, 이는 해당 보상이 NeuroWeb에서 발행되기 때문이다.

파라넷 인센티브에 대한 투표 기능 외에도, NEURO는 Polkadot 상 NeuroWeb 블록체인의 네이티브 자산으로서의 유틸리티를 가진다. 네이티브 자산으로서 NEURO는 트랜잭션 가스, 패러체인 콜레이터(collators) 운영 및 스테이킹, 그리고 NeuroWeb 거버넌스 투표에 사용된다.

7.3 토크노믹스 개요

오리진트레이일 DKG를 통해 인공지능을 위한 검증 가능한 인터넷을 구축하고 활용한다는 것은, 항상 최소 두 개의 탈중앙화 네트워크(DKG와 해당 블록체인)와 상호작용함을 의미한다. 따라서 인프라를 사용하는 데에는 TRAC 외에도 최소 하나 이상의 유틸리티 토큰이 필요하다. 아래 표는 두 계층에서 수행할 수 있는 주요 활동과 이를 실행하기 위해 사용되는 자산을 정리한 것이다.

표 1: 두 계층에서의 주요 활동 및 해당 자산 개요

	TRAC	NEURO	기타 자산
지식 자산(Knowledge Assets)			
지식 자산 게시(Publishing Knowledge Assets)	DKG 사용에 대한 보상 지급	NeuroWeb 사용 시 트랜잭션 가스	다른 체인 사용 시 가스 토큰(예: DAI, ETH 등)
DKG 인프라 운영(Running the DKG infrastructure)	DKG 노드에 대한 (위임) 스테이킹	NeuroWeb 사용 시 트랜잭션 가스	다른 체인 사용 시 가스 토큰(예: DAI, ETH 등)
지식 파라넷(Knowledge Paranets)			
지식 채굴 인센티브(Incentivizing Knowledge mining)	지식 자산 게시 시 사용	게시자에 대한 인센티브 제공	해당 없음
지식 서비스 인센티브(Incentivizing Knowledge services)	해당 없음	서비스 제공자에 대한 인센티브 제공	해당 없음
NeuroWebAI			
블록체인 사용(Blockchain use)	해당 없음	NeuroWeb 트랜잭션 가스	해당 없음
콜레이터 시스템(Collators system)	해당 없음	콜레이터 운영 및 스테이킹	해당 없음
거버넌스(Governance)	해당 없음	업데이트, 트레저리, 파라넷 성장 인센티브에 대한 투표	해당 없음

8 결론

인공지능을 위한 검증 가능한 인터넷을 탐구하는 일은, AI, 인터넷, 블록체인이라는 세 가지 융합 기술의 세부 구조를 면밀히 살펴볼 것을 요구한다. 이 첨단 영역의 결합은 중대한 도전을 해결함과 동시에 전례 없는 기회를 창출하며, 우리의 디지털 환경을 재편할 잠재력을 가진다.

본 백서는 강력하고 역동적일 뿐 아니라 신뢰 가능하고 투명한 AI 생태계를 제시한다. 오리진트레이일 탈중앙화 지식 그래프(DKG)는 지식 자산을 핵심 자원으로 삼아, dRAG 프레임워크를 통해 AI 시스템이 활용하는 정보의 소유권, 발견 가능성, 검증 가능성을 보장하는 강력한 구조를 제공한다. 자율적으로 운영되고 커뮤니티가 소유하는 특성을 가진 파라넷은 지식 자산을 조직하고 DKG를 확장하는 모듈형·확장형 접근 방식을 제공한다.

인공지능을 위한 검증 가능한 인터넷의 지속적 성장을 위한 인센티브는 다수의 네트워크에 걸쳐 존재한다. 각 자산은 설계 목적에 따라 사용된다 — TRAC은 DKG 인프라의 운영, 보안, 사용을 위해, NEURO는 성장 인센티브 제공과 더불어 NeuroWeb 블록체인의 네이티브 토큰으로서의 유틸리티를 위해 활용된다.

AI, 인터넷, 크립토의 융합으로 구현되는 인공지능을 위한 검증 가능한 인터넷은 보다 투명하고 공정하며 지능적인 디지털 세계로 나아가는 변혁적 단계라 할 수 있다.

9 참고문헌

- [1] Chui M., Roberts R., Yee L., Hazan Z., Singla A., Smaje K., Sukharevsky A., Zemmell R., (2023) The economic potential of generative AI: The next productivity frontier.
URL: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-AI-the-next-productivity-frontier>
- [2] Shumailov, I., Shumaylov, Z., Zhao, Y., Gal, Y., Papernot, N., Anderson, R. (2023) The curse of Recursion: Training on generated data makes models forget. New York, May 31.
URL: <https://arxiv.org/abs/2305.17493>
- [3] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S. and Kiela, D. (2021) Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. New York, April 12.
URL: <https://arxiv.org/abs/2305.17493v2>
- [4] Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., Wu, X., Fellow. (2024) Unifying Large Language Models and Knowledge Graphs: A Roadmap. Jan 25.
URL: <https://arxiv.org/pdf/2306.08302.pdf>
- [5] Sporny M., Longley D., Chadwick D., Steele O., Verifiable Credentials Data Model v2.0
URL: <https://www.w3.org/TR/vc-data-model-2.0/>
- [6] Galkin M., Yuan X., Mostasfa H., Tang J., Zhu Z., (2023) Towards Foundation Models for Knowledge Graph Reasoning.
URL: <https://arxiv.org/pdf/2310.04562.pdf>
- [7] LeCun, Y. (2022) A Path Towards Autonomous Machine Intelligence. New York, June 27.
URL: <https://openreview.net/pdf?id=BZ5a1r-kVsf>