

미라 네트워크(Mira Network)

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.

Mira: 신뢰 없는 AI 출력 검증을 위한 탈중앙화 네트워크

저자

- Ninad Naik (ninad@arohalabs.com)
- Sidhartha Doddipalli (sid@arohalabs.com)
- Karan Sirdesai (karan@arohalabs.com)

초록

AI는 그럴듯한 출력을 생성하는 데 뛰어나지만, 대규모 언어 모델이나 확산 모델과 같은 신경망 기반 기술의 확률적 특성으로 인해 자주 잘못된 정보를 산출합니다. 본 논문은 탈중앙화 합의를 통해 AI가 생성한 출력을 검증하는 네트워크를 소개합니다. 이 네트워크는 AI 출력을 독립적으로 검증 가능한 주장으로 변환하여, 여러 AI 모델이 집단적으로 각 주장의 타당성을 결정할 수 있게 합니다. 이러한 추론 기반 검증을 수행하는 노드 운영자는 정직한 검증을 수행하도록 하이브리드 형태의 작업증명(Proof-of-Work) 및 지분증명(Proof-of-Stake) 메커니즘을 통해 경제적 인센티브를 부여받습니다. 검증을 넘어, 우리의 비전은 오류 없는 출력을 제공하는 합성적 기초 모델(synthetic foundation model)로 확장됩니다. 이 인프라는 AI 시스템이 인간의 감독 없이 작동할 수 있도록 하는 중요한 단계이며, 이는 AI가 사회 전반에 걸쳐 변혁적 잠재력을 실현하기 위해 필요한 조건입니다.

1. 서론

인공지능(AI)은 인쇄술, 증기기관, 전기, 인터넷과 같은 기술들에 필적하는 변혁적 힘이 될 준비가 되어 있습니다. 이러한 기술들은 인류 문명을 근본적으로 재편하였습니다. 그러나 오늘날 AI는 이러한 혁명적 잠재력에 도달하지 못하게 하는 근본적 한계에 직면해 있습니다. AI는 창의적이고 그럴듯한 출력을 생성하는 데 뛰어나지만, 오류 없는 출력을 안정적으로 제공하는 데에는 어려움을 겪습니다. 이러한 한계로 인해 AI는 주로 인간이 감독하는 과제나 챗봇과 같은 낮은 위험의 응용으로 제한되며, 고위험 과제를 실시간으로 자율적으로 처리할 수 있는 잠재력에는 크게 미치지 못합니다.

핵심 장벽은 AI의 신뢰성입니다. AI 시스템은 두 가지 주요 오류 유형—환각(hallucination)과 편향(bias)—에 시달리며, 이들이 합쳐져 모델의 전체 오류율을 결정합니다. 현재의 오류율은 중요한 상황에서 자율적 운영이 가능하기에는 여전히 너무 높으며, 이로 인해 AI의 이론적 역량과 실제 응용 사이에 근본적인 격차가 발생합니다.

AI 모델은 훈련 데이터와 매개변수의 증가에 따라 계속 진화하고 있으나, *훈련 딜레마(training dilemma)* 때문에 이러한 신뢰성 문제는 여전히 지속됩니다. 이 딜레마는 고전적인 정밀도-정확도(trade-off) 균형과 유사합니다. 환각은 정밀도 오류(모델 출력의 일관성 문제)를 나타내고, 편향은 정확도 오류(진리로부터의 체계적 편차)로 나타납니다. 모델 개발자가 정밀도를 높이고 환각을 줄이기 위해 훈련 데이터를 선별할 때, 그 선택 기준으로 인해 필연적으로 정확도 오류(편향)가 발생합니다. 반대로, 정확도를 높이고(편향을 줄이고자) 다양한 잠재적 충돌 데이터를 학습할 경우, 모델의 지식 분포 전반에서 불일치하는 출력을 생성하게 되어 정밀도는 떨어지고 환각은 증가합니다.

세부 조정(fine-tuning)된 모델은 좁은 도메인 내에서는 더 높은 신뢰성을 달성하는 것으로 관찰되었습니다. 그러나 연구에 따르면 세부 조정된 모델은 새로운 지식을 안정적으로 포함하는 데 어려움을 겪으며, 새로운 정보를 도입하는 훈련 예시는 기존 지식과 일치하는 예시에 비해 훨씬 덜 효과적으로 학습됩니다. 또한 세부 조정된 모델은 훈련 범위를 벗어난 경계 사례나 예상치 못한 상황에서도 성능이 저조하여, 다양한 현실 세계 상황을 처리해야 하는 자율 시스템에는 적합하지 않습니다.

이러한 근본적 제약은 AI 모델 성능에서 불변의 경계를 확립합니다. 즉, 규모나 아키텍처에 관계없이 단일 모델로는 극복할 수 없는 최소 오류율이 존재합니다.

단일 모델이 환각과 편향을 동시에 최소화할 수는 없지만, 집단적 지혜는 앞으로 나아갈 길을 제시합니다. 합의 메커니즘을 통해 여러 모델이 협력하면 개별 모델이 할 수 없는 일을 달성할 수 있습니다. 즉, 집단 검증을 통해 환각을 걸러내고, 다양한 관점을 통해 개별 편향을 균형 있게 조정할 수 있습니다. 이러한 통찰은 신뢰할 수 있는 AI가 단순히 더 나은 모델이 아니라, 그들의 강점을 결합하고 개별 약점을 완화하는 더 나은 방법을 필요로 한다는 점을 시사합니다.

그러나 단순히 중앙집중적으로 통제되는 모델 앙상블을 구축하는 것으로는 신뢰성 문제를 완전히 해결할 수 없습니다. 모델 선택 자체가 체계적 오류를 도입하는데, 중앙화된 큐레이터의 선택은 필연적으로 특정한 관점과 한계를 반영하기 때문입니다. 게다가 많은 진실은 본질적으로 맥락적이며, 문화, 지역, 도메인에 따라 달라집니다. 진정한 신뢰성은 단순히 여러 모델을 모으는 것에 그치지 않고, 탈중앙화된 참여를 통해서만 나타낼 수 있는 진정으로 다양한 관점을 필요로 합니다.

따라서 필요한 것은 중앙집중적 권위가 아닌 탈중앙화 합의에 기반한 AI 검증 시스템입니다. 이를 통해 AI가 생성한 어떤 출력도 단일 신뢰 주체에 의존하지 않고 검증할 수 있습니다. 합의를 조작하는 것이 계산적으로나 경제적으로 비현실적이 되도록 만드는 시스템은 사용자를 불신된 출력으로부터 보호하는 동시에, 특화된 도메인 모델과 다양한 관점을 대표하는 모델 개발을 모두 장려하게 됩니다.

본 논문에서는 블록체인 기반의 다양한 AI 검증자 네트워크를 활용하여 AI 출력의 타당성에 대한 계산적 증명을 생성하는 AI 신뢰성 문제의 해결책을 제안합니다. 네트워크의 보안 프레임워크는 경제적 인센티브, 기술적 안전장치, 게임이론적 원칙의 조합을 통해 신뢰할 수 있는 검증을 보장합니다. 이러한 접근법은 분산 검증 메커니즘을 통해 AI 신뢰성을 향상시키며, 이는 편향과 환각률을 모두 줄이는 효과를 가집니다.

2. 네트워크 아키텍처

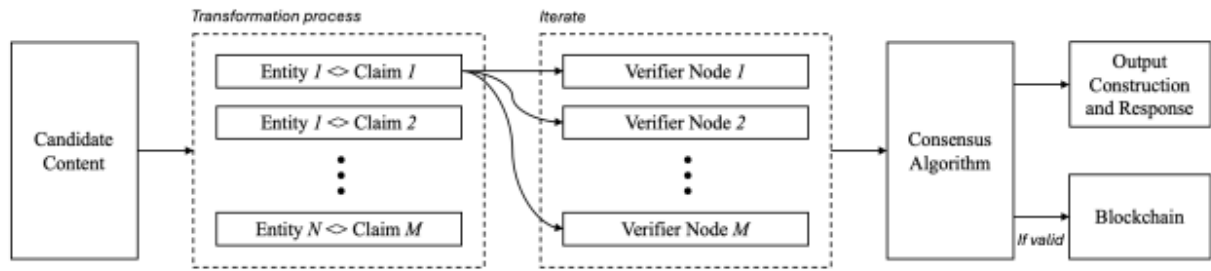
미래 네트워크는 복잡한 콘텐츠를 독립적으로 검증 가능한 주장으로 변환하는 새로운 프로토콜을 통해 AI 생성 출력의 신뢰 없는 검증을 가능하게 합니다. 이러한 주장은 다양한 AI 모델 간의 분산 합의를 통해 검증되며, 노드 운영자는 정직한 검증을 수행하도록 경제적 인센티브를 부여받습니다. 이러한 탈중앙화 접근 방식은 어떤 행위자도 검증 결과를 조작할 수 없도록 하면서 AI 생성 출력의 검증을 가능하게 합니다.

네트워크의 아키텍처는 콘텐츠 변환, 분산 검증, 합의 메커니즘을 결합한 새로운 조합을 통해 신뢰할 수 있는 검증을 가능하게 합니다. 이 시스템은 단순한 사실 진술부터 기술 문서, 창작 글, 멀티미디어 콘텐츠, 코드와 같은 복잡한 콘텐츠에 이르기까지 모두 처리할 수 있습니다.

예를 들어, 복합 문장을 생각해 보겠습니다: “지구는 태양을 공전하고 달은 지구를 공전한다.” 여러 모델로 이 간단한 문장을 검증하는 것은 직관적으로 단순해 보이지만, 전체 단락, 법률 문서, 코드와 같은 복잡한 콘텐츠로 확장하면 근본적인 도전 과제가 발생합니다. 후보 콘텐츠를 그대로 검증자 모델에 전달하면 각 검증자 모델은 콘텐츠의 서로 다른 측면을 해석하고 검증할 수 있습니다. 체계적 검증을 위해서는 각 검증자 모델이 동일한 맥락과 관점에서 정확히 같은 문제를 다루도록 AI 생성 출력을 표준화해야 합니다.

제안된 변환 접근법은 이러한 근본적 문제를 해결합니다. 위의 예시 문장에서 시스템은 후보 콘텐츠를 구별 가능한 검증 가능한 주장으로 분해합니다: (1) “지구는 태양을 공전한다.” (2) “달은 지구를 공전한다.” 양상불 검증을 통해 각 주장의 타당성을 판단하고, 검증 결과를 증명하는 암호학적 인증서를 발급합니다. 이 과정은 AI가 생성한 콘텐츠뿐만 아니라 인간이 생성한 콘텐츠에도 보편적으로 적용되며, 시스템은 출처에 구애받지 않고 엄격한 검증 기준을 유지할 수 있습니다.

네트워크는 후보 콘텐츠의 변환, 주장 분배, 합의 관리, 네트워크 오케스트레이션을 처리합니다. 노드 인프라는 독립적으로 운영되는 운영자들로 구성되며, 이들은 검증자 모델을 실행하고 주장을 처리하며 검증 결과를 제출합니다. 노드는 자율적으로 운영되지만 네트워크에 참여하기 위해 특정 성능 및 신뢰성 기준을 유지해야 합니다.



검증 워크플로우는 체계적으로 진행됩니다. 고객은 후보 콘텐츠를 제출하고, 의료·법률 등 특정 지식 영역과 같은 도메인이나 절대적 합의, N of M 합의 등과 같은 합의 임계값과 같은 검증 요구사항을 지정합니다. 네트워크는 이 콘텐츠를 논리적 관계를 보존한 상태로 검증 가능한 주장으로 변환하고, 이를 검증을 위해 노드에 분배하며, 결과를 집계하여 합의에 도달합니다. 그 후 네트워크는 검증 결과를 기록하는 암호학적 인증서를 생성하며, 각 주장에 대해 어떤 모델이 합의에 도달했는지를 포함한 결과와 인증서를 고객에게 반환합니다.

3. 경제적 보안 모델

네트워크의 경제적 보안 모델은 작업증명(Proof-of-Work, PoW)과 지분증명(Proof-of-Stake, PoS) 메커니즘을 결합하여, 정직한 검증을 위한 지속 가능한 인센티브를 창출하고 실제 경제적 가치를 포착 분배합니다. 이 하이브리드 접근법은 AI 출력을 검증하는 과정에서 발생하는 고유한 과제를 해결합니다.

네트워크는 검증을 통해 AI 오류율을 줄임으로써 실질적 경제적 가치를 창출합니다. 고객은 검증된 출력을 얻기 위해 네트워크 수수료를 지불하고, 네트워크는 검증 보상을 통해 이 수수료를 노드 운영자와 데이터 제공자에게 분배합니다.

전통적인 블록체인 네트워크에서 PoW는 무작위 성공 확률이 극도로 낮은 암호학적 퍼즐을 푸는 것을 포함하지만, 미래 네트워크에서는 검증을 표준화된 객관식 질문으로 전환합니다. 이러한 표준화는 노드 간 체계적 검증을 가능하게 하지만, 동시에 근본적 문제를 야기합니다. 가능한 응답의 확률 공간이 제한되기 때문입니다. 예를 들어, 이진 선택의 경우 검증 과제는 무작위로 성공할 확률이 50%이고, 4개의 선택지가 있으면 25%입니다. 이는 계산 비용 없이 높은 보상을 얻을 수 있는 무작위 추측 전략을 매력적으로 만들 수 있습니다.

이를 완화하기 위해, 노드는 검증에 참여하기 위해 가치를 스테이킹해야 합니다. 노드가 지속적으로 합의에서 벗어나거나 실제 추론이 아닌 무작위 응답을 시사하는 패턴을 보이면, 스테이킹된 자산은 슬래싱될 수 있습니다. 이러한 경제적 패널티는 무작위 응답을 통한 시스템 악용 시도가 경제적으로 비합리적이 되도록 보장합니다.

따라서 네트워크의 경제적 보안 모델은 세 가지 근본 원칙에 따라 운영됩니다. 첫째, 노드 운영자는 스테이킹된 가치가 슬래싱 위험에 노출되므로 경제적 인센티브에 합리적으로 반응합니다. 둘째, 정직한 운영자가 스테이킹된 가치의 대다수를 통제하는 한 네트워크 보안은 유지되며, 조작 시도는 지나치게 비싸집니다. 셋째, 네트워크가 확장됨에 따라 검증자 모델의 자연스러운 다양성이 통계적 편향을 줄입니다. 이는 각기 다른 모델이 다양한 훈련 접근법과 지식 기반을 가져오기 때문입니다. 이러한 원칙은 상호 강화 작용을 합니다. 경제적 인센티브는 다양한 참여자를 끌어들이고, 이들의 다양한 관점은 보안을 강화하며, 이는 다시 경제적 모델을 지

탱합니다.

표 1은 답변 선택지의 수에 따라 정답을 맞출 확률을 보여줍니다.

검증 횟수	2개 답변 옵션	4개 답변 옵션	6개 답변 옵션	8개 답변 옵션	10개 답변 옵션
1	50.0000%	25.0000%	16.6667%	12.5000%	10.0000%
2	25.2500%	6.2500%	2.7778%	1.5625%	1.0000%
3	12.5000%	1.5625%	0.4630%	0.1953%	0.1000%
4	6.2500%	0.3906%	0.0772%	0.0244%	0.0100%
5	3.1250%	0.0977%	0.0129%	0.0031%	0.0010%
6	1.5625%	0.0244%	0.0021%	0.0004%	0.0001%
7	0.7813%	0.0061%	0.0004%	0.0000%	0.0000%
8	0.3906%	0.0015%	0.0001%	0.0000%	0.0000%
9	0.1953%	0.0004%	0.0000%	0.0000%	0.0000%
10	0.0977%	0.0001%	0.0000%	0.0000%	0.0000%

네트워크의 초기 단계에서는 노드 운영자를 철저히 심사하여 네트워크의 무결성을 보장합니다. 두 번째 단계에서는 의도적으로 설계된 중복 구조와 함께 네트워크가 탈중앙화를 시작하며, 동일한 검증자 모델의 여러 인스턴스가 각 검증 요청을 처리하게 됩니다. 이러한 중복은 검증 비용을 증가시키지만, 악의적이거나 태만한 운영자를 강력하게 식별할 수 있게 합니다. 네트워크가 정상 상태로 성숙해지면, 검증 요청은 무작위로 노드에 분산되어 담합이 점점 더 어렵고 비용이 많이 들게 됩니다.

네트워크의 샤딩 메커니즘은 또 다른 보안 계층을 제공합니다. 시스템은 노드 간 응답 패턴과 유사성 지표를 분석하여 잠재적 담합을 식별할 수 있습니다. 악의적 행위자가 결과에 영향을 미치려면 네트워크에 스테이킹된 가치의 상당 부분을 통제해야 하는데, 이 시점에서 그들의 경제적 인센티브는 정직한 운영과 일치하게 됩니다.

노드 운영자는 검증 과정을 우회하기 위해, 일반적인 검증 결과 데이터베이스를 유지하는 등 다양한 전략을 통해 네트워크를 악용하려 할 수 있습니다. 단기적으로는 검증 요청의 다양성과 고유성이 이러한 캐싱 전략을 비효율적으로 만듭니다. 대규모로 확장되면, 방대한 검증된 사실 집합의 존재는 네트워크의 검증 역량을 활용하는 파생 프로토콜의 기회를 제공합니다.

노드 운영자는 가장 낮은 비용으로 정답에 도달함으로써 성공을 거둡니다. 특정 검증 과제에서 특화된 모델이 대형 모델과 비슷한 성능을 달성할 경우, 이는 합법적인 최적화 기회를 만듭니다. 이러한 기회는 특정 도메인에 최적화된 효율적이고 과제 특화된 모델의 개발을 촉진하며, 이는 더 높은 정확도, 낮은 비용, 짧은 지연 시간 등을 통해 전체 생태계에 이익을 가져옵니다.

네트워크의 경제 모델은 다중 강화 사이클을 통해 이러한 긍정적 역학을 강화합니다. 네트워크 사용량이 증가함에 따라 수수료 발생이 늘어나고, 이는 더 나은 검증 보상을 가능하게 하여 더 많은 노드 운영자를 유치하고 정확도, 비용, 지연 시간의 개선을 이끕니다. 이러한 성장은 네트워크 보안을 자연스럽게 강화합니다. 네트워크 가치가 상승함에 따라 스테이킹 요구량이 증가하고, 모델 다양성은 특화와 근본적 관점 차이를 통해 확장되며, 축적된 검증 이력은 점점 정교한 이상 탐지를 가능하게 합니다. 이러한 복합 효과는 정직한 검증과 지속적 혁신이 지배적 전략으로 자리 잡는 강력한 게임이론적 균형을 만들어내며, 악의적 조작을 경제적으로 비합리적이고 기술적으로 실행 불가능하게 만듭니다.

4. 프라이버시

앞서 설명한 보안 기반 위에서, 네트워크 설계는 프라이버시 보존을 핵심 아키텍처 원칙으로 우선시합니다. 프라이버시 보호는 네트워크의 근본적 콘텐츠 변환 접근 방식에서 시작됩니다. 즉, 복잡한 콘텐츠를 엔티티-주장 쌍으로 분해하고 이를 무작위로 노드에 샤딩하여 배분합니다. 이를 통해 어떤 단일 노드 운영자도 전체 후보 콘텐츠를 재구성할 수 없으며, 고객 프라이버시를 보호하면서 검증 무결성을 유지합니다.

프라이버시 모델은 다층적 보호를 통해 강화됩니다. 노드로부터의 검증 응답은 합의가 도출될 때까지 비공개로 유지되어, 검증 과정 중 정보 유출을 방지합니다. 합의가 달성되면, 네트워크는 필요한 검증 세부사항만 포함하는 인증서를 생성하여 데이터 최소화를 통해 프라이버시를 추가로 보존합니다.

네트워크의 초기 단계에서는 변환 소프트웨어의 중앙화된 성격이 자연스러운 프라이버시 경계를 형성합니다. 네트워크 로드맵에는 이 구성 요소의 점진적 탈중앙화가 포함되어 있으며, 암호학적 프로토콜과 안전한 연산 기술을 통해 강력한 프라이버시 보장을 유지합니다.

5. 네트워크 진화

네트워크의 진화는 AI 검증 및 생성 플랫폼으로의 자연스러운 발전 과정을 따르며, 이는 AI 시스템의 운영 방식을 근본적으로 재편할 것입니다. 우리의 비전은 단순한 검증을 넘어, 검증이 생성 과정에 본질적으로 통합된 새로운 유형의 기초 모델(foundation model)을 창출하는 데 있습니다. 이는 AI가 그 변혁적 잠재력을 실현하기 위해 필요한 근본적 돌파구입니다.

초기에는 의료, 법률, 금융과 같이 사실적 정확성이 중요하고 편향 위험이 최소화된 도메인에 집중하며, 점차 코드, 구조화된 데이터, 멀티미디어 콘텐츠와 같은 점점 복잡한 콘텐츠 유형으로 확장해 나갈 것입니다. 네트워크는 데이터 가용성 레이어와 보완적 기술을 통해 개인 데이터와 추가적 맥락까지 검증 기능을 확장하여, 기본 네트워크를 비대화하지 않으면서도 안전하고 효율적인 검증을 가능하게 합니다. 각 확장은 단순히 범위 확대에 그치지 않고, 더 정교하고 신뢰할 수 있는 AI 시스템으로 나아가는 한 걸음을 의미합니다.

검증 기능은 단순한 유효성 확인에서 시작하여, 잘못된 콘텐츠의 포괄적 재구성으로 발전하며, 궁극적으로 검증된 출력의 직접적 생성을 완성합니다. 이러한 발전은 생성 속도와 정확도 사이의 전통적인 균형(trade-off)을 제거하고, 엄격한 검증 기준을 유지하면서도 실시간 성능에 접근하게 합니다.

직접적인 검증을 넘어, 블록체인에 경제적으로 보장된 사실의 축적은 강력한 파생 애플리케이션을 가능하게 합니다. 이 검증된 지식 기반은 네트워크의 보안 보장을 상속받는 결정적 사실 확인 시스템(deterministic fact-checking systems)과 오라클 서비스를 지원할 수 있습니다. 더 근본적으로, 진실 검증에 경제적 인센티브를 부여함으로써, 네트워크는 원시 데이터를 가치로 뒷받침된 사실로 전환하는 새로운 모델을 구축합니다. 이는 신뢰할 수 있는 AI 시스템을 위한 중요한 구성 요소입니다.

기술적 역량과 경제적 인센티브의 지속적 진화를 통해, 네트워크는 전례 없는 신뢰성을 갖춘 새로운 세대의 AI 애플리케이션을 가능하게 할 것입니다. 이는 AI 시스템의 점진적 개선에 그치지 않고, 인간의 감독 없이도 오류 없는 운영을 실현하여 AI가 자율적으로 작동할 수 있는 새로운 패러다임을 확립합니다.

6. 결론

오늘날의 AI 시스템은 근본적인 도전에 직면해 있습니다. 창의적이고 그럴듯한 출력을 생성하는 데 뛰어나지만, 엄격하고 고위험 환경에서 오류 없는 출력을 안정적으로 제공하지는 못합니다. 우리의 탈중앙화 검증 네트워크는 콘텐츠 변환과 분산 합의를 암호경제적 인센티브와 결합하는 새로운 방식을 통해 이 문제를 해결하며, 조작을 기술적·경제적으로 모두 불가능하게 만듭니다. 전통적인 PoW가 임의의 퍼즐을 푸는 것과 달리, 미라 네트워크는 스테이킹된 가치에 의해 보장되는 의미 있는 추론 계산을 요구하여 정직한 운영을 보장합니다.

검증을 넘어, 우리의 비전은 검증이 생성 과정에 직접 통합된 합성적 기초 모델(synthetic foundation model)입니다. 이 간소화된 접근법은 생성과 검증의 구분을 제거하고, 오류 없는 출력을 제공합니다. 인센티브를 부여받은 운영자의 탈중앙화 네트워크 전반에 검증을 분산함으로써, 우리는 본질적으로 중앙화된 통제에 저항하는 인프라를 창출합니다. 이는 근본적 진전을 의미합니다. 즉, AI 시스템이 인간의 감독 없이 작동할 수 있도록 함으로써, 실제 인공지능을 위한 토대를 마련하는 것입니다. 이는 AI가 사회 전반에 걸쳐 그 변혁적 잠재력을 실현하는 데 있어 필수적인 단계입니다.

참고 문헌

[1] E. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?”, <https://dl.acm.org/doi/pdf/10.1145/3442188.3445922>, 2021

[2] M. Chelli, J. Descamps, V. Lavou  , C. Trojani, M. Azar, M. Deckert, J. Raynier, G. Clovez, P. Boileau, C. Ruetsch-Chelli, “Hallucination Rates and Reference Accuracy of ChatGPT and Bard for Systematic Reviews: Comparative Analysis”, <https://www.jmir.org/2024/1/e53164/>?, 2024

- [3] Z. Gekhman, G. Yona, R. Aharoni, M. Eyal, A. Feder, R. Reichart, J. Herzig, “Does Fine-Tuning LLMs on New Knowledge Encourage Hallucinations?”, <https://arxiv.org/pdf/2405.05904>, 2024
- [4] N. Naik, “Probabilistic Consensus through Ensemble Validation: A Framework for LLM Reliability”, <https://mira.network/research/ensemble-validation.pdf>, 2024
- [5] S. King, S. Nadal, “PPCoin: Peer-to-Peer Crypto-Currency with Proof-of-Stake”, <https://www.peercoin.net/read/papers/peercoin-paper.pdf>, 2012
- [6] S. Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System”, <https://bitcoin.org/bitcoin.pdf>, 2009
- [7] X. Zuwei, S. Jain, M. Kankanhalli, “Hallucination is Inevitable: An Innate Limitation of Large Language Models”, <https://arxiv.org/abs/2401.11817>, 2024