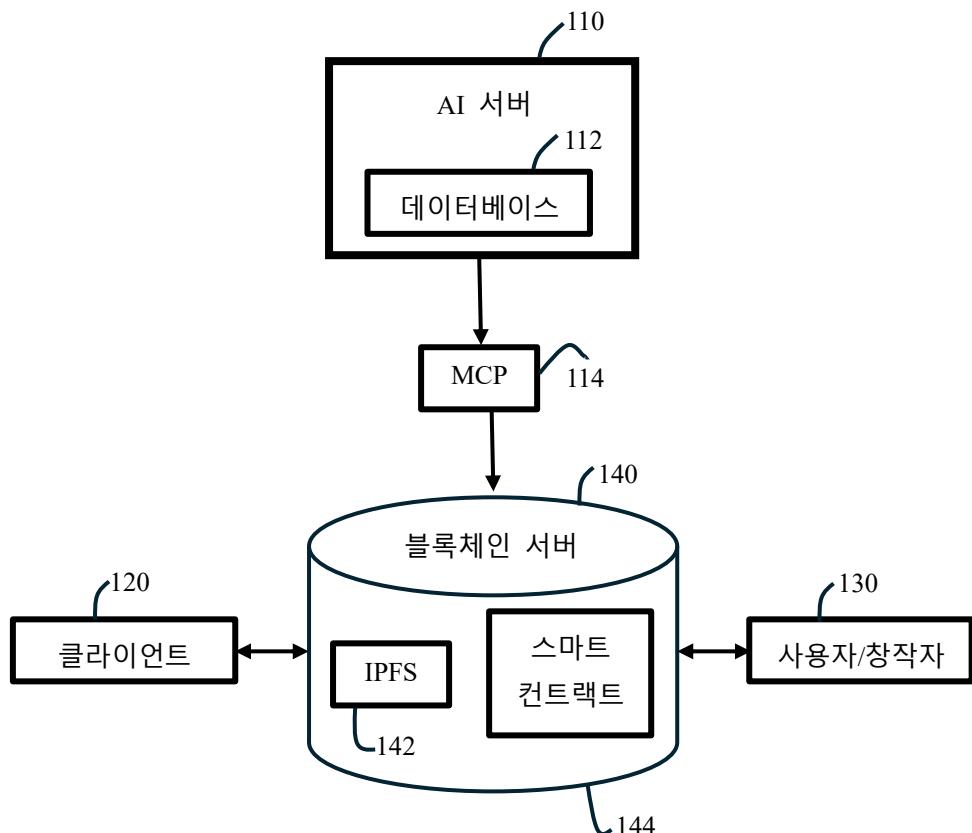

(54) 블록체인 기반 AI 기여도 보상 방법 및 시스템

(57) 요약

본 개시는 블록체인 기반의 AI 기여도 보상 방법 및 시스템을 제공한다. 생성형 AI 이용 과정에서 사용자 피드백 또는 창작자 콘텐츠가 AI 성능 향상에 기여하지만, 이에 대한 명확한 출처 표시나 보상은 존재하지 않는 문제를 해결하기 위해, 본 발명은 다음과 같은 단계로 구성된다: 사용자 또는 창작자로부터 AI 응답에 대한 정정 제안 또는 도메인 지식 기반 콘텐츠를 수신하는 단계; MCP(Model Context Protocol)를 통해 AI 서버와 블록체인 서버를 연동하여 제1 블록체인을 생성하는 단계; 웹 검색 결과 또는 복수 사용자 검증을 통해 기여 정보의 진위 여부를 확인하는 단계; 검증된 기여 정보 IPFS(InterPlanetary File System)에 업로드하고 해시값을 블록체인에 기록하는 단계; 샤플리 값(Shapley Value)을 이용하여 각 기여자의 기여도 비율을 산정하는 단계; 및 산정된 기여도에 따라 스마트 컨트랙트를 통해 토큰 기반 보상을 분배하는 단계를 포함한다. 이를 통해, AI 지식 생태계의 공정성 및 투명성을 확보하고, 콘텐츠 창작자와 사용자의 기여를 정당하게 보상할 수 있다.

대표도 - 도1



명세서

청구범위

청구항 1

블록체인 서버의 적어도 하나의 프로세서에 의해 수행되는, 블록체인 기반의 AI 기여도 보상 방법에 있어서,

클라이언트로부터 사용자의 피드백 또는 창작자의 콘텐츠를 수신하는 단계 – 상기 사용자 피드백은 AI 응답에 대한 정정 제안 및 근거를 포함하고, 상기 창작자 콘텐츠는 도메인 지식 기반 자료를 포함함 -;

MCP(Model Context Protocol)를 통해 AI 서버와 블록체인 서버를 연동하여 제1 블록체인을 생성하는 단계;

상기 사용자 피드백 또는 창작자 콘텐츠를 기반으로 AI 응답 변경 여부를 추적하는 단계;

웹 검색 결과 또는 복수 사용자 검증을 통해 기여 정보의 진위 여부를 확인하는 단계;

검증된 기여 정보를 IPFS(InterPlanetary File System)에 업로드하고 해시값을 블록체인에 기록하는 단계;

샤플리 값(Shapley Value)을 이용하여 각 기여자의 기여도 비율을 산정하는 단계; 및

상기 블록체인 서버에 연결된 적어도 하나의 노드에 의해 생성된 상기 기여도 산정 결과에 관한 검수 결과를 상기 적어도 하나 이상의 노드에 송신하는 단계 – 상기 검수 결과는 상기 기여도 산정 결과에 대하여 상기 적어도 하나 이상의 노드 중 일부에 의해 결정된 승인 또는 미승인 여부에 기초하여 생성됨 –

를 포함하고,

상기 스마트 컨트랙트를 통해 산정된 기여도에 따라 토큰 기반 보상을 분배하는 단계는,

상기 사용자 피드백 및 상기 창작자 콘텐츠를 기초로 샤플리 값 및 PageRank 알고리즘 사이의 가중치를 조정하는 단계; 및

상기 가중치를 반영한 최종 기여도를 산정하는 단계

를 포함하는, 블록체인 기반의 AI 기여도 보상 방법.

발명의 설명

기술분야

[0001] 본 개시는 블록체인 기반의 AI 기여도 보상 방법 및 시스템에 관한 것으로, 구체적으로 생성형 AI의 성능 향상에 기여한 사용자 피드백 및 창작자 콘텐츠에 대해 공정한 기여도를 산정하고 보상을 분배하는 방법 및 시스템에 관한 것이다.

배경기술

[0003] 최근 OpenAI가 공개한 GPT-4, Anthropic의 Claude, Google의 Gemini 등 대규모 언어 모델(LLM, Large Language Models)은 대화형 인터페이스를 사용하는 다양한 애플리케이션의 핵심 요소로 각광 받고 있다. 여기서, 언어 모델은 방대한 양의 텍스트 데이터를 기반으로 사전 학습된 딥 러닝 모델이다. 언어 모델은 텍스트를 생성, 번역, 요약, 질의 응답, 코드 생성 등 다양한 자연어 처리(NLP) 작업을 수행할 수 있다.

[0004] 그러나 생성형 AI 모델은 사용자의 피드백(정정, 질문, 평가 등)과 창작자의 도메인 지식 기반 콘텐츠를 통해 지속적으로 개선되지만, 이러한 기여자들의 출처가 명확히 추적되지 않으며, 기여도에 따른 보상 체계가 존재하지 않아 AI 지식 생태계의 불공정성이 심화되고 있다.

구체적으로, 약 1,500만 개의 웹 콘텐츠가 동의 없이 AI 훈련에 사용되었으며, 60%의 AI 응답이 올바른 출처 인용에 실패하고 있다. 또한 AI 도입 이후 상위 500개 퍼블리셔의 트래픽이 27% 감소하였으며, 61%의 크리에이터가 AI를 생계 위협으로 인식하고 있다. 더 나아가 80.2%의 아티스트가 자신의 작품이 AI 훈련에 사용되었는지 공개를 요구하고 있으며, 57.5%의 챗봇 응답이 부정확한 인용을 포함하고 있다. 이에 따라, 콘텐츠 창작자와 사용자의 기여를 정당하게 보상하고 AI 응답의 출처를 투명하게 추적할 수 있는 시스템의 필요성이 대두되고 있다.

발명의 내용

해결하려는 과제

[0006] 본 개시는 생성형 AI의 성능 향상에 기여한 사용자 피드백과 창작자 콘텐츠에 대해 정량적 기여도 측정 방법을 제공함으로써, AI 지식 생태계의 불공정성 문제를 해결하고자 한다. 이를 위해 본 개시는 샤플리 값(Shapley Value)을 활용한 공정한 기여도 산정 메커니즘을 구현하여, 각 기여자의 실제 기여 비율을 협력 게임 이론에 기반하여 수학적으로 산출한다. 또한, 본 개시는 IPFS(InterPlanetary File System)와 블록체인 기술을 결합하여 기여 이력을 투명하게 추적하고 변조 불가능하게 저장함으로써, 콘텐츠 창작자의 저작권을 보호하고 사용자 피드백의 출처를 명확히 한다. 나아가, 본 개시는 스마트 컨트랙트 기반의 자동 보상 분배 시스템을 구축하여, 중개자 없이 기여도에 비례한 토큰 보상이 즉시 이루어지도록 한다. 더 나아가, 본 개시는 MCP(Model Context Protocol)를 통한 AI 서버와 블록체인 서버의 실시간 연동을 구현하여, AI 응답 생성 과정에서 발생하는 기여 정보를 즉시 블록체인에 기록하고, 이를 기반으로 실시간 기여도 추적 및 보상 분배가 가능하도록 한다.

과제의 해결 수단

[0008] 본 개시는 방법, 장치(시스템) 및/또는 컴퓨터 판독 가능 저장 매체에 저장된 컴퓨터 프로그램, 컴퓨터 프로그램이 저장된 컴퓨터 판독 가능 저장 매체를 포함한 다양한 방식으로 구현될 수 있다.

[0009] 본 개시의 일 실시예에 따르면, 블록체인 서버의 적어도 하나의 프로세서에 의해 수행되는, 블록체인 기반의 AI 기여도 보상 방법이 제공될 수 있다. 상기 방법은 클라이언트로부터 사용자의 피드백 또는 창작자의 콘텐츠를 수신하는 단계 – 상기 사용자 피드백은 AI 응답에 대한 정정 제안 및 근거를 포함하고, 상기 창작자 콘텐츠는 도메인 지식 기반 자료를 포함함 -; MCP(Model Context Protocol)를 통해 AI 서버와 블록체인 서버를 연동하여 제1 블록체인을 생성하는 단계; 상기 사용자 피드백 또는 창작자 콘텐츠를 기반으로 AI 응답 변경 여부를 추적하는 단계; 웹 검색 결과 또는 복수 사용자 검증을 통해 기여 정보의 진위 여부를 확인하는 단계; 검증된 기여 정보를 IPFS(InterPlanetary File System)에 업로드하고 해시값을 블록체인에 기록하는 단계; 샤플리 값(Shapley Value)을 이용하여 각 기여자의 기여도 비율을 산정하는 단계; 및 상기 블록체인 서버에 연결된 적어도 하나의 노드에 의해 생성된 상기 기여도 산정 결과에 관한 검수 결과를 상기 적어도 하나 이상의 노드에 송신하는 단계를 포함할 수 있다.

[0010] 본 개시의 일 실시예에 따르면, 상기는 검수 결과에 기초하여 적어도 하나 이상의 상기 노드에 의해 제공되

는 보상을 상기 클라이언트에 송신하는 단계를 더 포함할 수 있다.

[0011] 본 개시의 일 실시예에 따르면, 상기 기여도 산정은 사용자 피드백 및 창작자 콘텐츠의 가중치 조정을 통해 수행되고, 상기 가중치 조정은 샤플리 값과 PageRank 알고리즘 사이의 가중치를 기초로 결정되며, 상기 가중치는 사용자 피드백, 창작자 콘텐츠, 대화 힌트 및 미리 결정된 지시사항을 기초로 산출될 수 있다.

[0012] 본 개시의 일 실시예에 따르면, 상기 샤플리 값 및 PageRank 알고리즘 사이의 가중치를 조정하는 단계는, 사용자 피드백을 기초로 미리 결정된 힌트 생성 기준을 이용하여 대화 힌트를 생성하는 단계 및 창작자 콘텐츠, 사용자 피드백, 대화 힌트 및 미리 결정된 지시사항을 기초로 샤플리 값과 PageRank 알고리즘 사이의 가중치를 조정하는 단계를 포함하고, 지시사항은 샤플리 값 알고리즘이 PageRank 알고리즘과 가중치를 조정함에 있어서 대화의 가이드라인이 될 수 있다.

[0013] 본 개시의 일 실시예에 따르면, 방법은 힌트 생성 기준, 지시사항 및 사용자 피드백을 해시 키(hashed key)로 산출하는 단계를 더 포함할 수 있다.

[0014] 본 개시의 일 실시예에 따르면, 제1 블록체인은, 해시 키를 키값(key)으로 포함하고, 창작자 콘텐츠, 대화 힌트 및 대화 기록을 내용값(value)으로 포함하는 key-value 쌍을 포함할 수 있다.

[0015] 본 개시의 일 실시예에 따르면, 평가자로 대화 기록의 평가를 위해 대화 기록을 송신하는 단계 후에, 평가자가 원요한 대화 기록에 대응되는 해시 키와 연관된 식별정보를 평가자로 송신하는 단계, 평가자로부터 미리 결정된 평가 기준 및 해시 키에 기초하여 실행된 대화 기록의 평가 결과를 수신하는 단계 및 평가 결과를 기초로 제2 블록체인을 생성하는 단계를 포함할 수 있다.

[0016] 본 개시의 일 실시예에 따르면, 제2 블록체인은, 평가자의 퍼블릭 키(public key)를 키값으로 포함하고, 식별 정보 및 평가 결과를 내용값으로 포함하는 key-value 쌍을 포함할 수 있다.

[0017] 본 개시의 일 실시예에 따른 방법을 컴퓨터에서 실행하기 위해 컴퓨터 판독 가능한 기록 매체에 저장된 컴퓨터 프로그램이 제공될 수 있다.

[0018] 본 개시의 일 실시예에 따르면, 블록체인 기반의 AI 기여도 보상 시스템은 클라이언트, 평가자, 노드 또는 블록체인 서버 중 적어도 하나 이상으로부터 트랜잭션을 수신하도록 구성된 풀 노드; 메이터베이스를 포함하는 메모리 및 메모리와 연결된 프로세서; 및 메모리와 연결된 프로세서를 포함하고, 적어도 하나의 프로세서는, 클라이언트로부터 사용자의 피드백 또는 창작자의 콘텐츠를 수신하고, MCP를 통해 AI 서버와 블록체인 서버를 연동하여 제1 블록체인을 생성하고, 평가자로 상기 대화 기록의 평가를 위해 대화 기록을 송신하고, 대화 기록에 응답하여 평가자에 의해 생성된 대화 기록에 대한 평가 결과를 수신하고, 블록체인 서버에 연결된 적어도 하나 이상의 노드에 의해 생성된 평가 결과에 관한 검수 결과를 적어도 하나 이상의 노드에 송신하거 위한 명령어들을 포함할 수 있다.

발명의 효과

[0020] 본 개시의 일부 실시예에 따른 방법은, AI 지식 생태계에서 사용자 피드백과 창작자 콘텐츠의 기여를 투명하게 추적하고 공정하게 보상할 수 있다.

구체적으로, 샤플리 값을 통한 기여도 산정으로 각 기여자의 실제 기여 비율을 정량적으로 측정할 수 있으며, IPFS 와 블록체인을 통해 기여 이력을 변조 불가능하게 저장할 수 있다. 또한, 스마트 컨트랙트를 통한 자동 보상 분배로 중개자 없이 즉시 보상이 이루어질 수 있다.

이에 더해, MCP(Model Context Protocol)를 통한 AI-블록체인 실시간 연동으로 기여도 추적이 즉시 이루어지며, AI 응답 생성 과정에서 발생하는 모든 기여 정보가 실시간으로 블록체인에 기록되어, 블록체인의 컨센서스(consensus) 구조를 활용하여 기여도 산정의 신뢰성이 보장되며, AI 응답의 출처 투명성 확보로 정보 소비자의 신뢰도가 향상될

수 있다.

[0021] 본 개시의 일부 실시예에 따른 방법은, 클라이언트에게 보상을 지급함으로써, 다양한 프롬프트와 모델을 이용한 챗봇의 평가를 최적화으로 이루어지도록 할 수 있다. 또한, 복수의 노드의 검수를 통해 챗봇의 평가 결과에 대한 신뢰성을 확보할 수 있다.

나아가, 본 개시는 콘텐츠 창작자의 저작권 보호와 사용자의 참여 동기 부여를 동시에 달성함으로써, AI 지식 생태계의 선순환 구조를 구축할 수 있다. 구체적으로, 창작자는 자신의 콘텐츠가 AI 훈련에 사용될 때마다 정당한 보상을 받을 수 있으며, 사용자는 AI 응답의 정정 및 개선에 기여함으로써 톤튼 보상을 획득할 수 있다.

또한, 본 개시는 기여도 산정의 투명성과 공정성을 확보함으로써, 기존 AI 생태계에서 발생하던 출처 불명확 문제(60%의 부정확한 인용), 창작자 보상 부재 문제(80.2%의 아티스트가 작품 공개 요구), 퍼블리셔 트래픽 감소 문제(27% 감소) 등을 근본적으로 해결할 수 있다.

더 나아가, 본 개시는 블록체인의 분산 원장 기술을 활용하여 중앙 집중식 관리 없이도 기여 이력을 영구적으로 보존할 수 있으며, 스마트 컨트랙트의 자동 실행 메커니즘을 통해 인간 개입 없이 보상이 즉시 분배되므로, 관리 비용을 절감하고 보상 지급의 신속성을 보장할 수 있다.