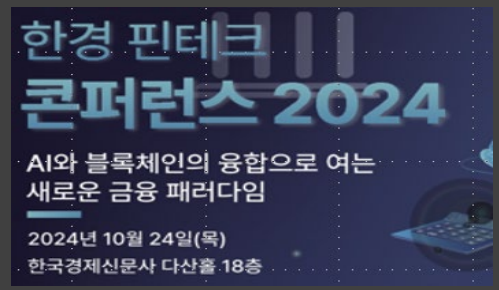
Web3 AI
지향점, 솔루션과 가능성

Heung-No Lee 이흥노
Professor GIST/CEO LiberVance



기조연설:
Web3 AI
지향점, 솔루션과 가능성

Agenda

1. LLM 혁신
2. Web3 AI
3. 국제적 Web3 AI 동향

LLM 혁신

Open models are available to make AI agents.

Open Source vs. Closed Source

- GPTs are not open source
- Transformer model is open source
- Meta's LLaMa models are OW
- LLaMa3.1 is as good as GPT4.

Meta

Open Source AI Is the Path Forward

July 23, 2024
By Mark Zuckerberg, Founder and CEO

Why? People want
- Their own model
- Control their own destiny
- Protect their data

Good for Meta?
- First mover advantage

Meta

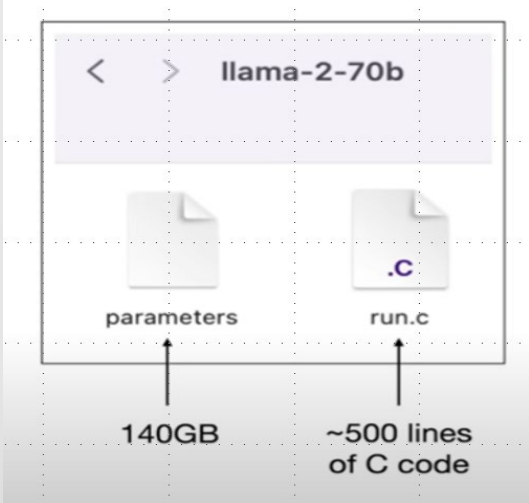
a BigScience initiative
BLOOM
176B params 59 languages Open-access

Technology Innovation Institute
Innovation for a better world.

Google Research

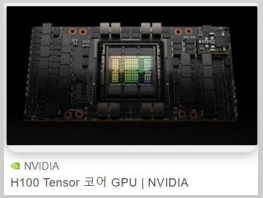
Model	Source Code	Training Data	Checkpoints
LLaMa 3	Open	Partially Open	Open
BLOOM	Open	Open	Open
Falcon 180B	Open	Not Open	Open
Flan-T	Open	Not Open	Open

LLaMa is a two file system.



Revolutionary processing power

Meet the fastest, most powerful Snapdragon for laptops, available first in Galaxy Book4 Edge.[4]



NVIDIA
H100 Tensor 코어 GPU | NVIDIA

80GB Memory

AMD EPYC 9334 듀얼 NVIDIA A100 80GB 2WAY 최대 8GPU GPU서버 AI 연구 학습 딥러닝 이중화 4U 랙타입

94,714,590원 **8 GPU**

8/24(토) 도착 예정

Small - hand phone
Medium - PC or Laptop
Large - Server
ULarge - Cluster

Llama Models

Model	Launch date	Model sizes	Context Length
Llama 2	7/18/2023	7B, 13B, 70B	4K
Llama 3	4/18/2024	8B, 70B	8K
Llama 3.1	7/23/2024	8B, 70B, 405B	128K

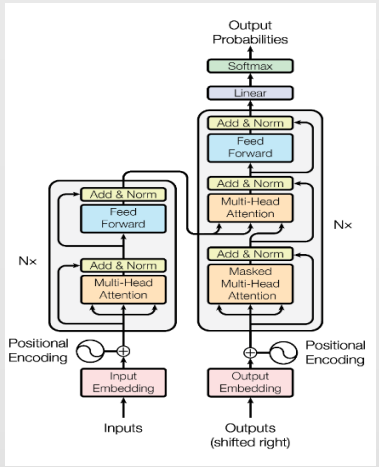
GPT3.5 175B
GPT4.0 1T?

Transformer is a next word prediction model.

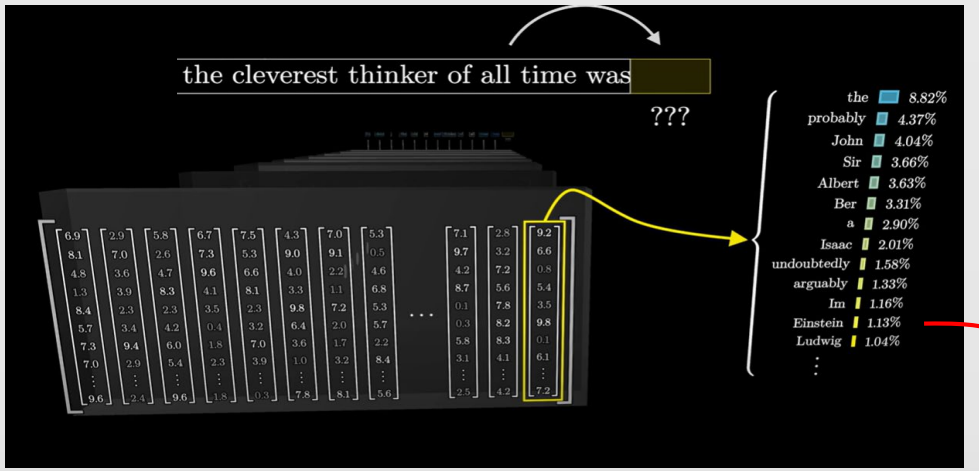
Input

the cleverest thinker of all time was MASK

X



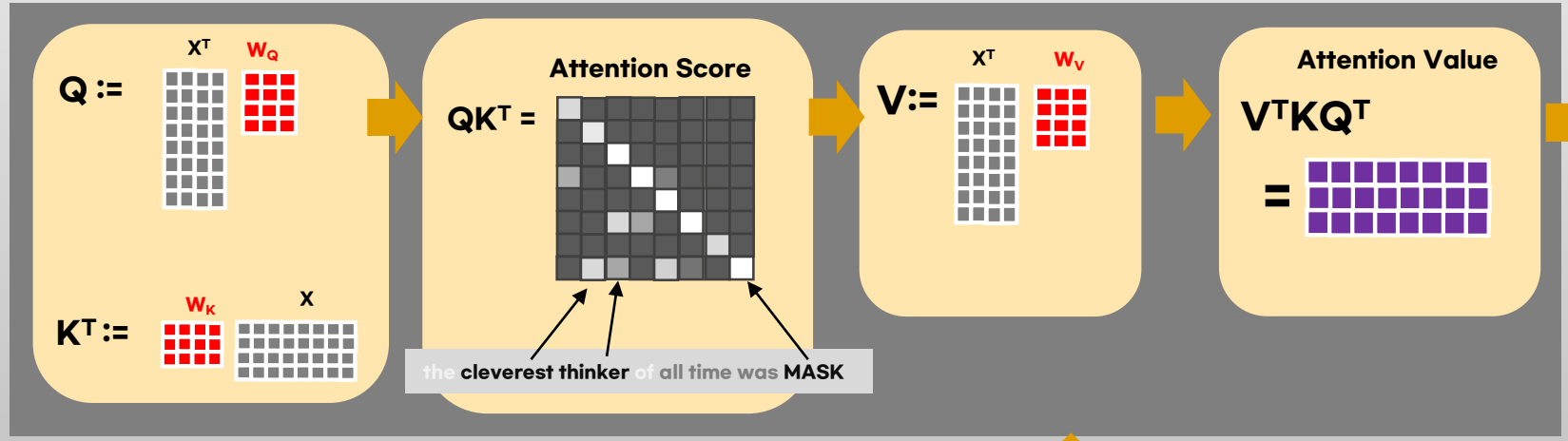
Context length = 8



Output

"the"

"Einstein"



Compute Loss

Back propagation

Parameter matrices W_Q, W_K, W_V are updated

How does it work?

We know
- How to iteratively adjust billions of parameters and make it better at predicting the next word

We don't know
- How these parameters work together achieving it

It builds and maintains a knowledge database, but it is imperfect.

- "Reversal Curse" example
- Q: Who is Tom Cruise's mother?
- A: Mary Lee Pfeiffer (correct)

- Q: Who is Mary Lee Pfeiffer's son?
- A: I don't know.

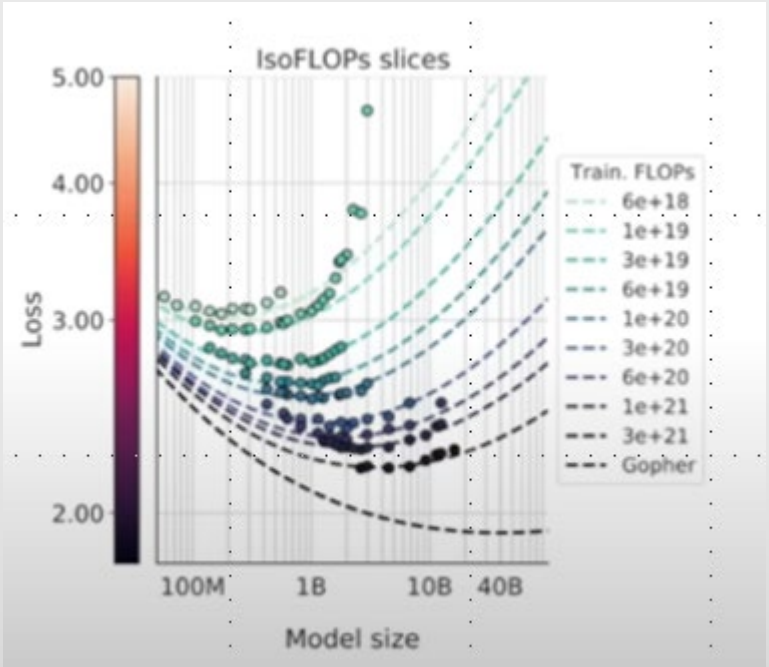
Aware that LLM outputs are probable artifacts!
- Need to develop sophisticated method to verify and correct them!

Retrieval Augmented Generation

We can expect better intelligence by scaling!

LLM Scaling Laws:

- Parameter Scaling: Larger models with more parameters tend to perform better.
- Data Scaling: Models trained on more data show improved performance.

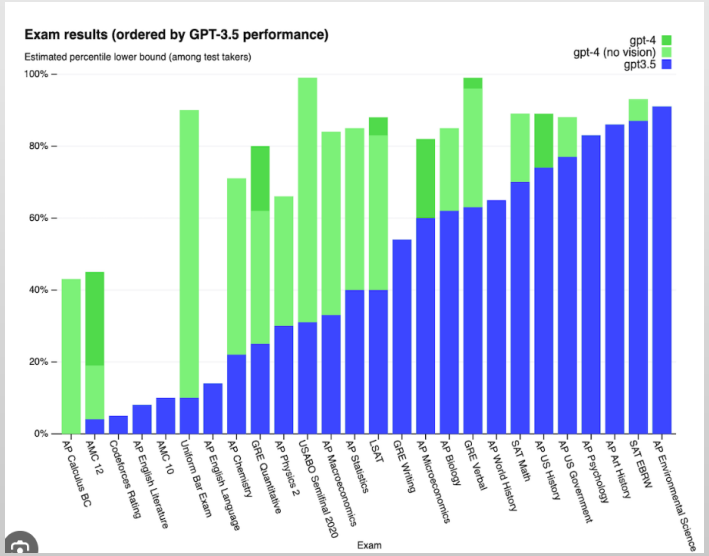


GPT3.5 175 billion parameters

GPT4.0 possibly 1 trillion parameters

Emergent Abilities

- Few shot
- Zero shot



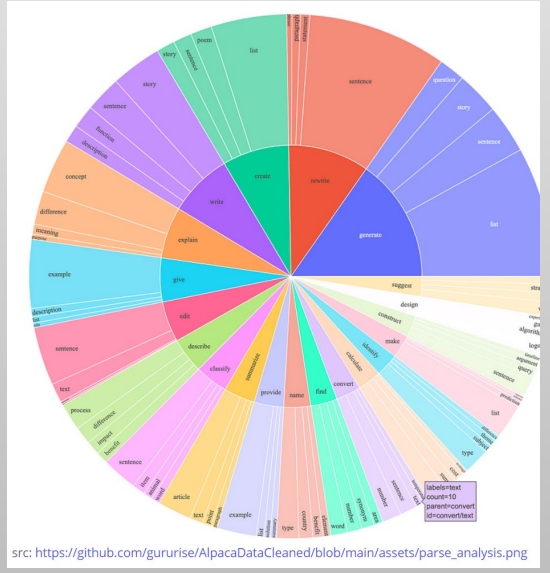
Two types of training

Pre-training (once/year)
- Collect 10 TB of text
- Get a cluster of 10,000 GPUs
- Feed the text into LLM ($2M, month)
- Obtain the pre-trained (PT) model

Next word prediction enables the LLM to learn a lot about the world!

Fine-tuning (once/month)
- Consult Alpaca paper for instructions
- Get a single GPU or GPU server.
- Generate 100K high quality Q&A responses (or buy)
- Fine-tune the PT model on Q&A responses (1 day)
- Obtain FT model

Makes an assistant



src: https://github.com/gururise/AlpacaDataCleaned/blob/main/assets/parse_analysis.png

Create
Summarize
Rewrite
Suggest
Describe
Edit
Classify
...

2024 Novel Prizes

The Nobel Prize in Chemistry 2024

David Baker

"for computational protein design"

David Baker. Ill. Niklas Elmehed © Nobel Prize Outreach

Demis Hassabis

"for protein structure prediction"

Demis Hassabis. Ill. Niklas Elmehed © Nobel Prize Outreach

John Jumper

"for protein structure prediction"

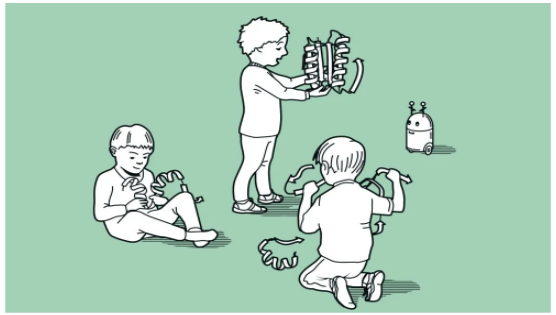
John Jumper. Ill. Niklas Elmehed © Nobel Prize Outreach

They cracked the code for proteins' amazing structures

The Nobel Prize in Chemistry 2024 is about proteins, life's ingenious chemical tools. David Baker has succeeded with the almost impossible feat of building entirely new kinds of proteins. Demis Hassabis and John Jumper have developed an AI model to solve a 50-year-old problem: predicting proteins' complex structures. These discoveries hold enormous potential.

Related articles

Press release
Popular information: They have revealed proteins' secrets through computing and artificial intelligence
Scientific background: Computational protein design and protein structure prediction



© Johan Jarnestad/The Royal Swedish Academy of Sciences

The Nobel Prize in Physics 2024

John Hopfield

"for foundational discoveries and inventions that enable machine learning with artificial neural networks"

John Hopfield. Ill. Niklas Elmehed © Nobel Prize Outreach

Geoffrey Hinton

"for foundational discoveries and inventions that enable machine learning with artificial neural networks"

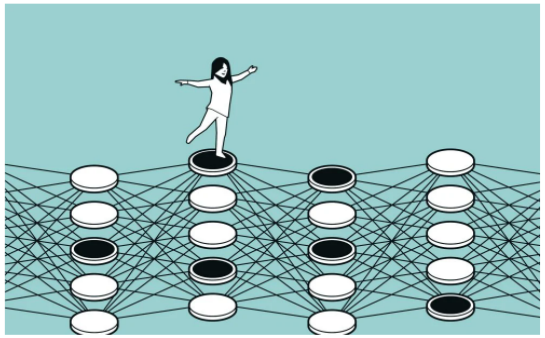
Geoffrey Hinton. Ill. Niklas Elmehed © Nobel Prize Outreach

They used physics to find patterns in information

This year's laureates used tools from physics to construct methods that helped lay the foundation for today's powerful machine learning. John Hopfield created a structure that can store and reconstruct information. Geoffrey Hinton invented a method that can independently discover properties in data and which has become important for the large artificial neural networks now in use.

Related articles

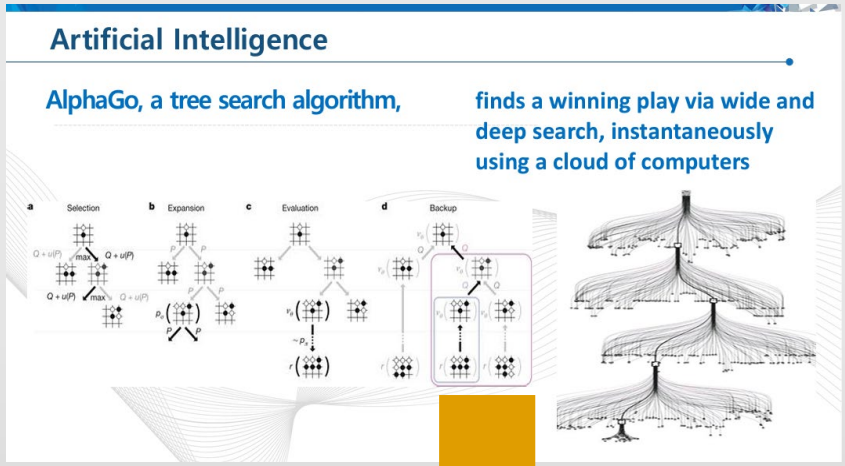
Press release
Popular information: They used physics to find patterns in information
Scientific background: "for foundational discoveries and inventions that enable machine learning with artificial neural networks"



© Johan Jarnestad/The Royal Swedish Academy of Sciences

Two types of AI advances, and future

Broad and Deep Search



Generative AIs

A network of AI agents

- Each node is a domain expert
- Local self-learning
- Global cooperative learning network
- Solving real-world problems
- Evolution of network intelligence

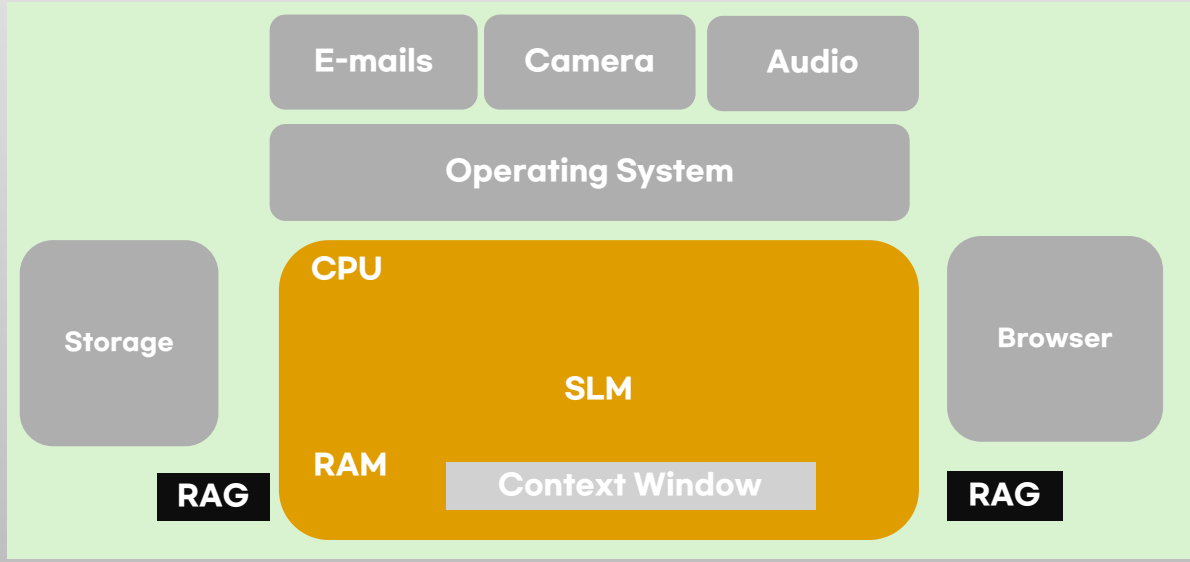
AI assistant in laptops & hand phones

Apple Intelligence

Apple Intelligence is the personal intelligence system that puts powerful generative models right at the core of your iPhone, iPad, and Mac and powers incredible new features to help users communicate, work, and express themselves. You can bring these Apple Intelligence features right into your apps.

An AI agent is an AI assistant which autonomously performs tasks for me.
It listens to me or watch my screen activities and assists me:

- Make travel plan
- Write a report
- Carry out on-line research



You can run your SLM on device.
But for medium and large models,
you have to choose between

Web2 and Web3

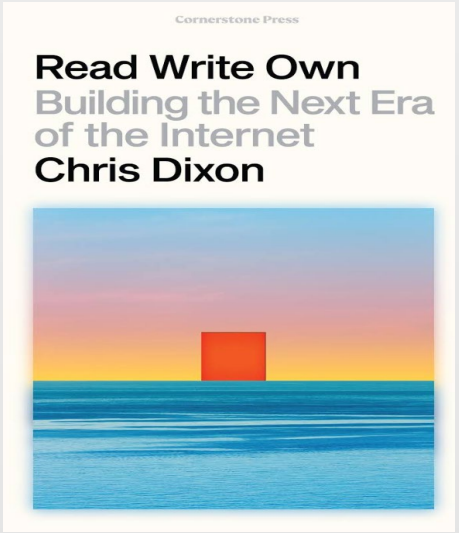
Web3 AI

Web2 vs Web3

Web 2 platforms:
Facebook, YouTube, Twitter

They control content distribution, data collection, and monetization, which are key aspects that Web3 seeks to decentralize and democratize.

Web3 infrastructure
- dApps
- NFTs
- DAOs
- Permissionless blockchains
- Users control their data and identity



Take Rates

플랫폼이 창작자 제공 콘텐츠로 얻은 수익 중 징수하는 수익 비율

Attract 단계와 Extract 단계

Web2	Facebook	YouTube
Take Rate	100%	45.0%

Web3	Ethereum	Uniswap
Take Rate	0.0%	<2.5%

Web3

Web3 represents the 3rd phase of the internet, focusing on decentralization, user control, and blockchain integration.

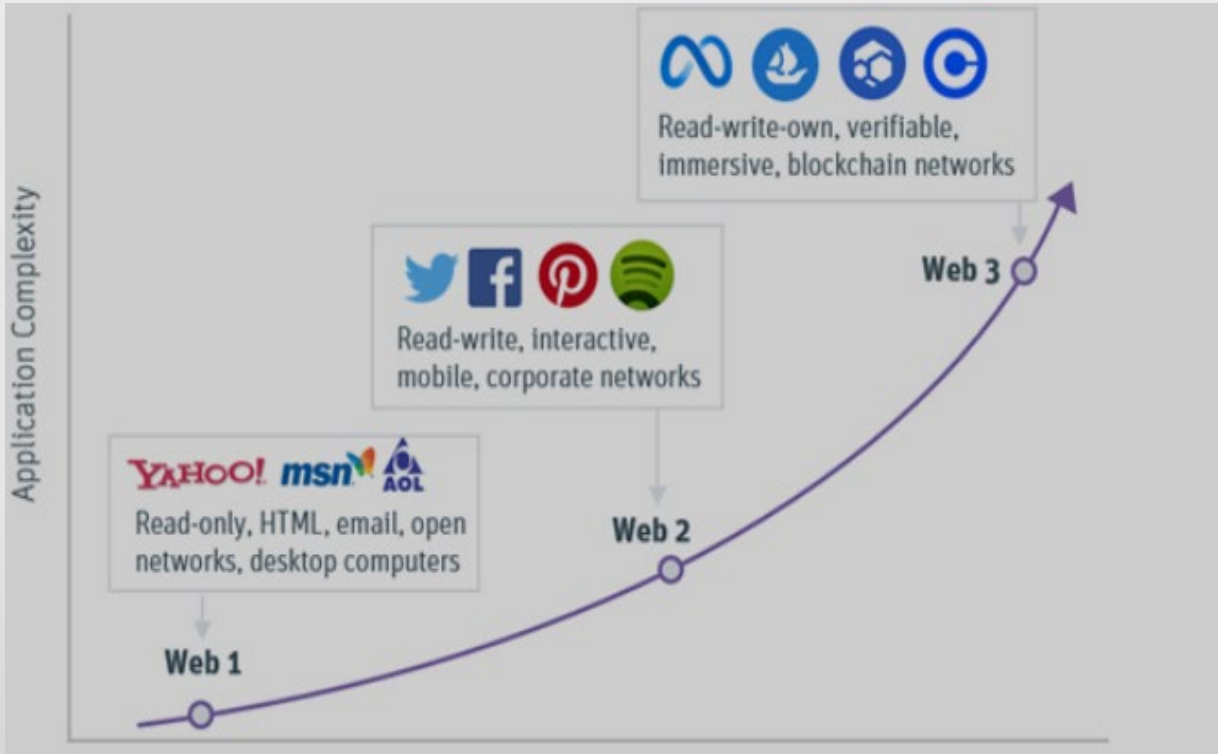
Origin:

- The term "Web3" was popularized by Gavin Wood, co-founder of Ethereum, around 2014.
- He envisioned a user-centric internet where applications would not rely on a central authority, and data would be owned by users.

Key Players:

- Gavin Wood (Polkadot), Vitalik Buterin (Ethereum), and companies like ConsenSys, Chainlink, and Protocol Labs.

Web3



Impact: Web3 could democratize the internet by enabling peer-to-peer transactions, reducing reliance on intermediaries, and creating new economic models based on tokens and decentralized governance.

Challenges:

- Scalability: Many blockchain networks face limitations in transaction speed and capacity.
- User Experience: Web3 applications are often less user-friendly than their Web2 counterparts.
- Interoperability: Connecting different blockchains and ensuring they work together smoothly is still a work in progress.

Overall, Web3 is rapidly evolving but still in its early stages.

The next few years will be critical in determining how widespread and impactful it becomes.

Web3 AI

Web3 AI refers to the fusion of Web3 and AI.

- The idea of Web3 AI is emerging.
- There are two types available today:
- (Type 1) Web3 AI aims to run and train AI models on decentralized, transparent, and permissionless Web3 infra such as blockchains, smart contracts, and tokens. For this, we need to create new systems where users can retain ownership of their data and AI models, benefit economically from AI models, and ensure privacy.
- (Type 2) This fusion allows AI to operate on decentralized platforms, enabling applications like decentralized autonomous organizations (DAOs) to utilize AI for governance, decision-making, and operations without relying on centralized entities.

Web3 AI

Prospect

- The near-term prospects for Web3 AI are promising but also challenging.
- The integration of AI into Web3 platforms could disrupt various industries, including finance, healthcare, and content creation.

Challenges

- Scalability of blockchain technologies
- Need for decentralized AI models

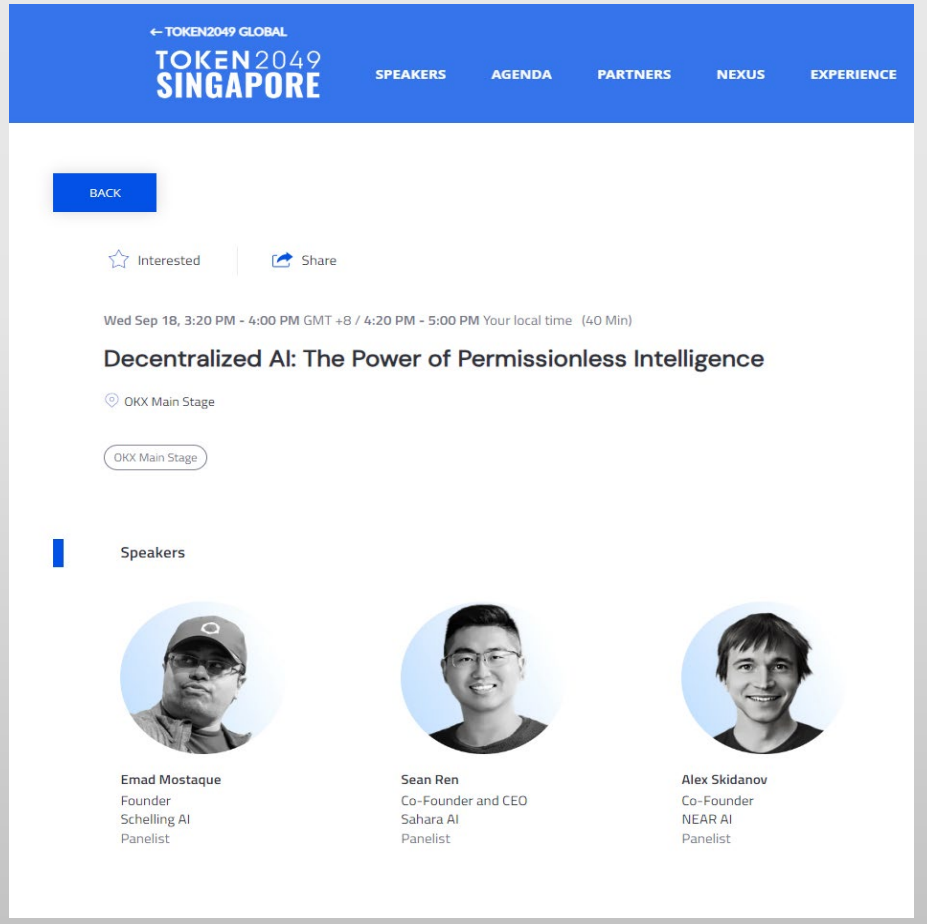
Key Takeaways

- Use not only ZKML, OPML, Federated Learning…
- But also layer-2 chains, sidechains, and off-chain computations/storages for a greater scalability.
- Track AI transactions
- Allow people monetize their AIs
- Encourage cooperation among AI agents
- Enable solving real world problems
- Empower people at the edge

Web3 AIs in Consensus 2024

1. AKASH Network: Tokenized decentralized GPU Network
2. Morpheus: Open Source Decentralized AI Network
3. GAIANET: Decentralized GenAI Agents Network
4. Oraichain: Multichain and off-chain AI compute
5. SingularityNET
4. Rise of the Machine Economy

Many more projects exist in the direction of Web3 AI.
➔ Will discuss this in My AI Network (competitive landscape)



DUBAI AI & WEB3 FESTIVAL

국제적 Web3AI동향

My AI Network

사용자가 프라이빗한 Web3 클라우드에서

AI 에이전트를 직접 생성하고, 소유 및 관리하며,

수익화할 수 있는 플랫폼

이흥노

GIST 교수 / CEO LiberVance

블록체인지능융합센터

LiberVance

비전

AI Agent 시대가 도래하고 있습니다.

인간 교사

AI 교사

인간 변호사

AI 변호사

인간 의사

AI 의사

문제점

플랫폼 중심형 Web2 AI 기업은 전 세계 사람들의 콘텐츠로 이익을 독점합니다.

비 오픈 소스!

독점적 수익 추구형!

Web2 AI

모델 및 데이터 비소유!

독점형 거버넌스!

문제점

AI 위협

"AI가 생계 위협"…노벨상 이시구로·배우 무어 등 예술인 1만명 성명

중앙일보 | 입력 2024.10.23 10:14 업데이트 2024.10.23 10:38

조문규 기자 구독

일본계 영국 국적의 2017년 노벨 문학상 수상자 가즈오 이시구로. AP=연합뉴스

서명은 온라인으로 받고 있다. 해당 성명 웹사이트에는 작가 이시구로를 비롯해 영국의 인기 록밴드 라디오헤드의 톰 요크, 전설적인 스웨덴 팝그룹 아바의 비에른 울바에우스, 할리우드 배우 케빈 베이컨, 멀리사 조앤 하트, 케이트 맥키넌, 코미디언 로지 오도넬, 미국 소설가 제임스 패터슨 등도 이름을 올렸다. 또 미국음악가연맹과 미국 배우노조(SAG-AFTRA), 유럽작가위원회 등 단체들도 서명에 참여했다. 현재까지 1만500명이 서명했다.

그는 "우리가 얘기하는 것은 글, 미술, 음악 등 사람들이 만든 창작물"이라며 "AI 회사가 이를 '학습 데이터'라고 부르는 것은 비인간적인 행위"라고 비판했다. 이어 "AI 기업의 창작물 사용을 제도적으로 막아야 한다"며 "예술인 개인이 직접 거부 의사를 표시할 경우에만 기업 측에서 해당 저작물을 제외하는 '옵트 아웃'(opt out) 방식으로는 문제가 해결되지 않는다"고 주장했다.

리버밴스 솔루션: My AI Network는 사용자 중심

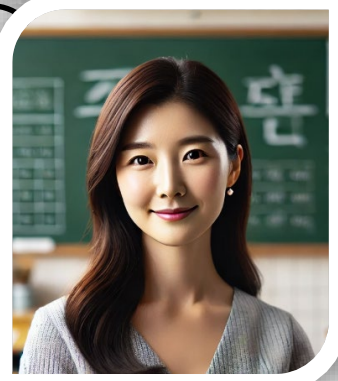
Local Data Learning
My AI agent

Mutual Cross Learning

Local Data Learning
My AI agent

Mutual Cross Learning

Local Data Learning
My AI agent



Local Data Learning
My AI agent

Mutual Cross Learning

Local Data Learning
My AI agent

My AI Network

오픈 소스 Web3 플랫폼으로서, 사용자에게 데이터와 AI 통제권을 제공하고 수익화할 수 있게함

리버밴스의 솔루션

My AI Network의 주요 기능

1 Drop & Run 에이전트 생성: 코딩 없이 누구나 에이전트를 만들 수 있습니다!

- 자신의 데이터를 Drop하여 프라이빗 공간에서 자신의 에이전트를 생성하고 실행합니다.

2 에이전트 수익화: 에이전트로 지식증류(distillation) 서비스를 제공하고 수익을 창출합니다!

- 나의 에이전트는 다른 사람과 그들의 에이전트에게 서비스를 제공하여 수익을 창출하고, 상호 교류와 학습을 통해 더욱 더 스마트 해집니다!

3 검증 가능한 AI 컴퓨트 클라우드 제공: 독창적인 프로토콜 스택 보유!

- 오픈 소스 모듈형 AI 컴퓨트 프로토콜을 기반으로 구축
- PQS 암호화 / ECCVCA / PT Model / FT-RAG / ZKP / Web3 등 주요 기술 보유

리버밴스의 솔루션

리버밴스는 Web3 AI 비전을 실현하는 MY AI Network를 구축합니다.

Web3 AI

AI Teacher

AI Lawyer

AI Doctor

My AI Network

AI + Blockchain

Law | Medical | Research | Education | Design | Coding

Web3 AI는 사용자가 개인 데이터와 AI모델을 온전하게 소유하는 방식을 의미합니다.

My AI Network의 주요 특징

- Web3 인프라와 토큰 인센티브를 통해 GPU등의 주요 Compute 자원이 확보 및 공유되며,
- 오픈 소스 프로토콜로 자신의 AI 에이전트를 지속적으로 학습 시킬수 있고 자주적인 AI 에이전트 관리가 가능한 탈중앙 환경을 제공하고,
- AI 사용자, 데이터제공자, 개발자를 포함한 전 세계인에게 참여가 개방된 플랫폼입니다.

강화 가능한 사용자의 역량, 왜 Web3를 선택해야할까?

소유권, 경제적 자립 및 협업 역량 강화

업무 의뢰

전문서비스
컨설팅
창작 및 콘텐츠

Private

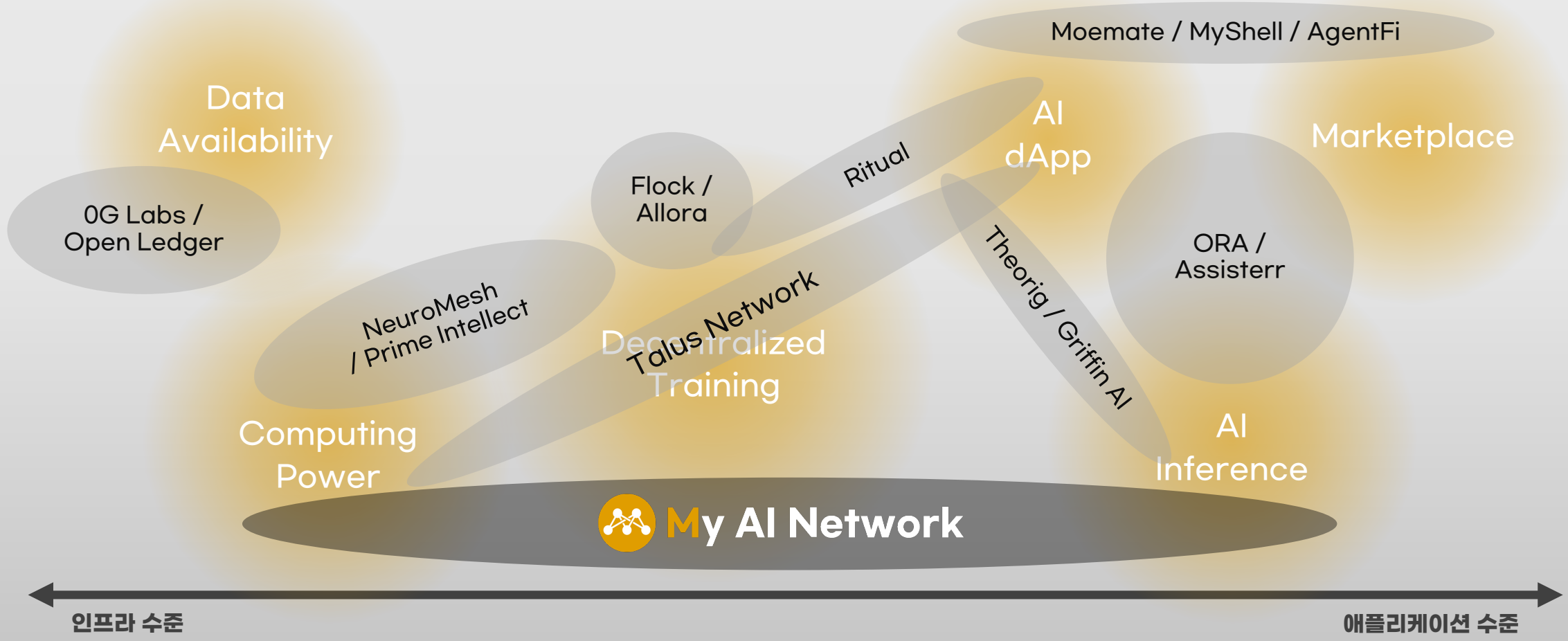
Public

지식재산권 관리 및 소유권 강화	창작 능력 강화 및 협업	경제적 자립과 전문성 강화	타인의 권리 보호와 책임성 강화

LiberVance

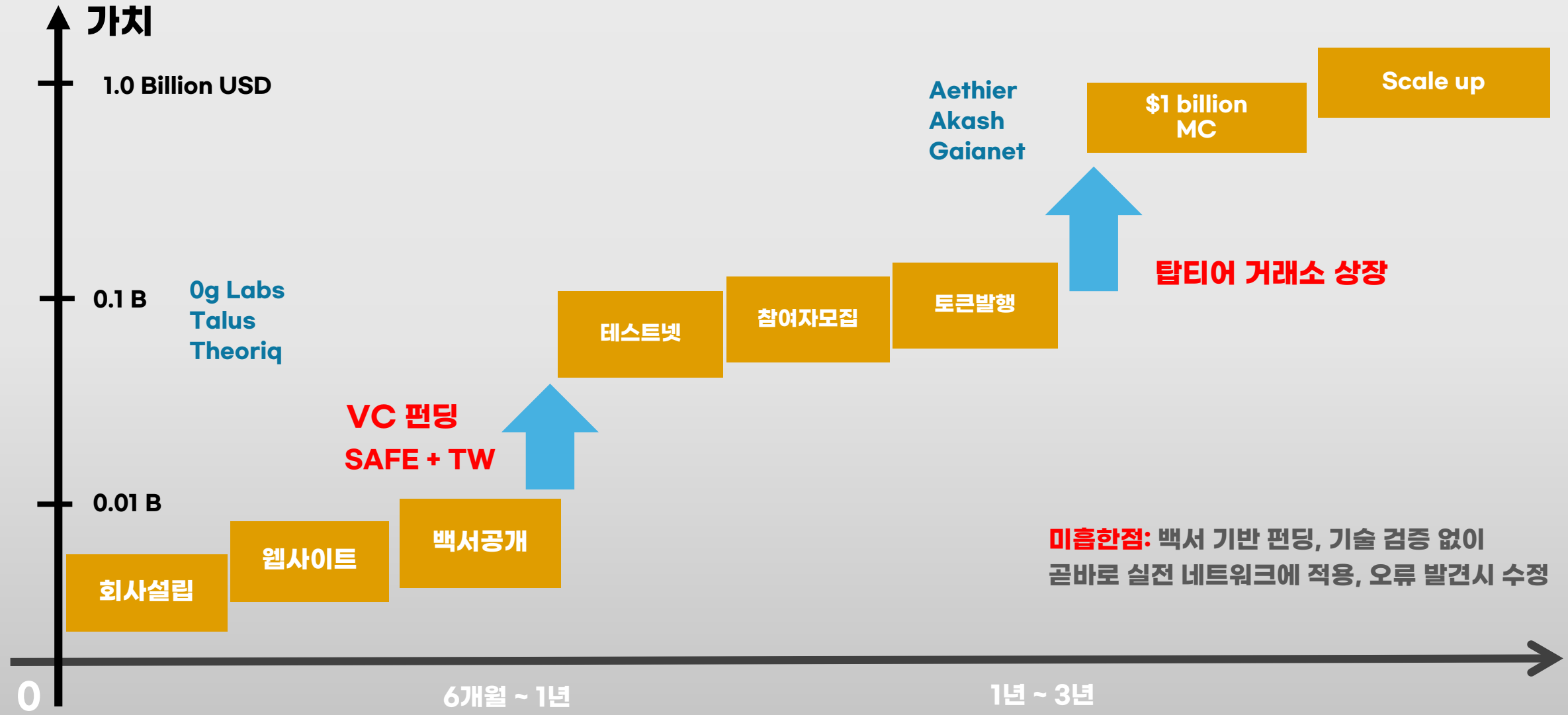
세계적 경쟁 및 협력 현황

리버밴스는 컴퓨팅 파워, AI 학습 및 추론 서비스제공에 중점을 두고 이를 강화하는 것을 목표로 합니다!



LiberVance

Web3AI: 회사 설립후 거래소 상장까지 대개 1년 ~ 3년 계획

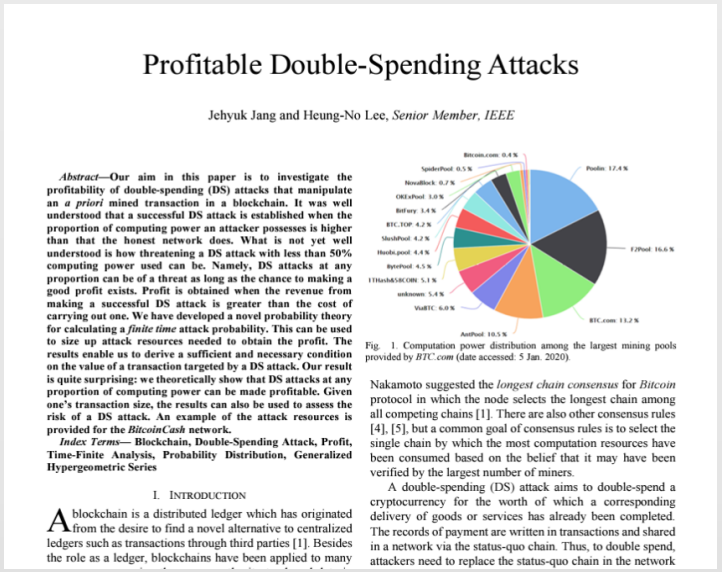


LiberVance

고유 기술 스택

핵심 프로토콜 구성 요소들은 최상위 전문 저널에 게재되었습니다.

- 양자내성암호, ZKP, ECCVAC
- Legal Query RAG 등 LLM 기술



Profitable Double-Spending Attacks

Jehyuk Jang and Heung-No Lee, *Senior Member, IEEE*

Abstract—Our aim in this paper is to investigate the profitability of double-spending (DS) attacks that manipulate an *a priori* mined transaction in a blockchain. It was well understood that a successful DS attack is established when the proportion of computing power an attacker possesses is higher than that the honest network does. What is not yet well understood is how threatening a DS attack with less than 50% computing power used can be. Namely, DS attacks at any proportion can be of a threat as long as the chance to making a good profit exists. Profit is obtained when the revenue from making a successful DS attack is greater than the cost of carrying out one. We have developed a novel probability theory for calculating a *finite time* attack probability. This can be used to size up attack resources needed to obtain the profit. The results enable us to derive a sufficient and necessary condition on the value of a transaction targeted by a DS attack. Our result is quite surprising: we theoretically show that DS attacks at any proportion of computing power can be made profitable. Given one's transaction size, the results can also be used to assess the risk of a DS attack. An example of the attack resources is provided for the *BitcoinCash* network.

Index Terms— Blockchain, Double-Spending Attack, Profit, Time-Finite Analysis, Probability Distribution, Generalized Hypergeometric Series

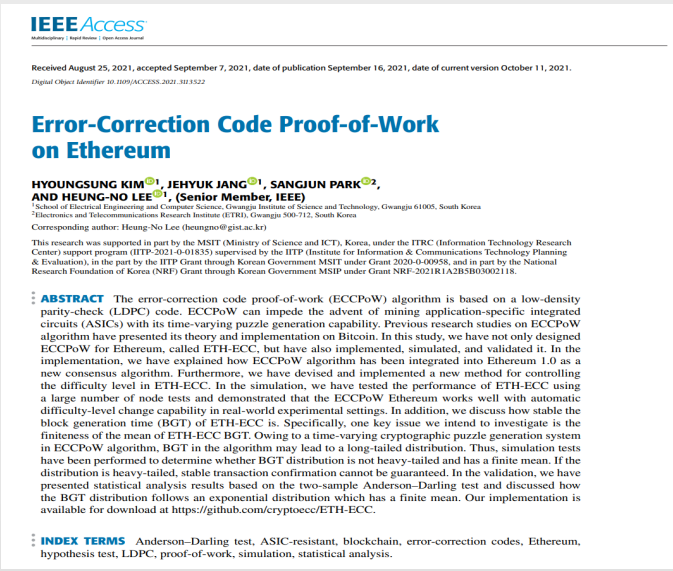
I. Introduction

A blockchain is a distributed ledger which has originated from the desire to find a novel alternative to centralized ledgers such as transactions through third parties [1]. Besides the role as a ledger, blockchains have been applied to many

Fig. 1. Computation power distribution among the largest mining pools provided by *BTC.com* (date accessed: 5 Jan. 2020).

Nakamoto suggested the *longest chain consensus* for *Bitcoin* protocol in which the node selects the longest chain among all competing chains [1]. There are also other consensus rules [4], [5], but a common goal of consensus rules is to select the single chain by which the most computation resources have been consumed based on the belief that it may have been verified by the largest number of miners.

A double-spending (DS) attack aims to double-spend a cryptocurrency for the worth of which a corresponding delivery of goods or services has already been completed. The records of payment are written in transactions and shared in a network via the status-quo chain. Thus, to double spend, attackers need to replace the status-quo chain in the network



IEEE Access

Received August 25, 2021, accepted September 7, 2021, date of publication September 16, 2021, date of current version October 11, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3113522

Error-Correction Code Proof-of-Work on Ethereum

HYOUNGSUNG KIM[1], JEHYUK JANG[1], SANGJUN PARK[2], AND HEUNG-NO LEE[1], (Senior Member, IEEE)

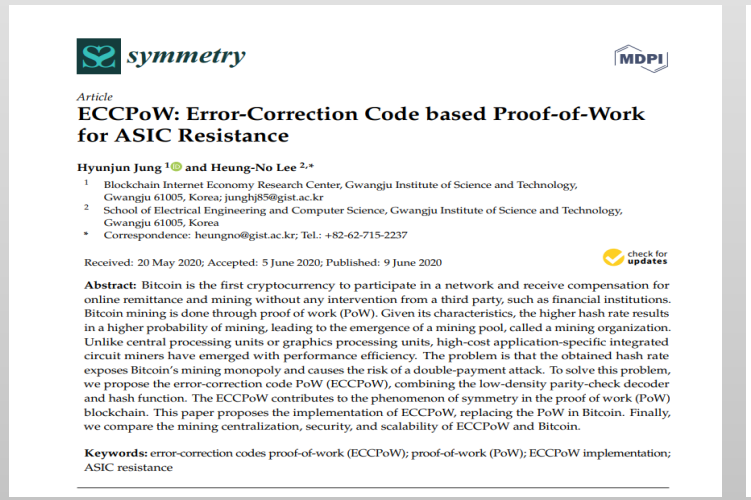
[1]School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju 61005, South Korea
[2]Electronics and Telecommunications Research Institute (ETRI), Gwangju 500-712, South Korea

Corresponding author: Heung-No Lee (heungno@gist.ac.kr)

This research was supported in part by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2021-0-01835) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation), in the part by the IITP Grant through Korean Government MSIT under Grant 2020-0-00958, and in part by the National Research Foundation of Korea (NRF) Grant through Korean Government MSIP under Grant NRF-2021R1A2B5B03002118.

ABSTRACT The error-correction code proof-of-work (ECCPoW) algorithm is based on a low-density parity-check (LDPC) code. ECCPoW can impede the advent of mining application-specific integrated circuits (ASICs) with its time-varying puzzle generation capability. Previous research studies on ECCPoW algorithm have presented its theory and implementation on Bitcoin. In this study, we have not only designed ECCPoW for Ethereum, called ETH-ECC, but have also implemented, simulated, and validated it. In the implementation, we have explained how ECCPoW algorithm has been integrated into Ethereum 1.0 as a new consensus algorithm. Furthermore, we have devised and implemented a new method for controlling the difficulty level in ETH-ECC. In the simulation, we have tested the performance of ETH-ECC using a large number of node tests and demonstrated that the ECCPoW Ethereum works well with automatic difficulty-level change capability in real-world experimental settings. In addition, we discuss how stable the block generation time (BGT) of ETH-ECC is. Specifically, one key issue we intend to investigate is the finiteness of the mean of ETH-ECC BGT. Owing to a time-varying cryptographic puzzle generation system in ECCPoW algorithm, BGT in the algorithm may lead to a long-tailed distribution. Thus, simulation tests have been performed to determine whether BGT distribution is not heavy-tailed and has a finite mean. If the distribution is heavy-tailed, stable transaction confirmation cannot be guaranteed. In the validation, we have presented statistical analysis results based on the two-sample Anderson–Darling test and discussed how the BGT distribution follows an exponential distribution which has a finite mean. Our implementation is available for download at https://github.com/cryptoecc/ETH-ECC.

INDEX TERMS Anderson–Darling test, ASIC-resistant, blockchain, error-correction codes, Ethereum, hypothesis test, LDPC, proof-of-work, simulation, statistical analysis.



symmetry

MDPI

Article

ECCPoW: Error-Correction Code based Proof-of-Work for ASIC Resistance

Hyunjun Jung [1] and Heung-No Lee [2,*]

1 Blockchain Internet Economy Research Center, Gwangju Institute of Science and Technology, Gwangju 61005, Korea; junghj85@gist.ac.kr
2 School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju 61005, Korea
* Correspondence: heungno@gist.ac.kr; Tel.: +82-62-715-2237

Received: 20 May 2020; Accepted: 5 June 2020; Published: 9 June 2020

check for updates

Abstract: Bitcoin is the first cryptocurrency to participate in a network and receive compensation for online remittance and mining without any intervention from a third party, such as financial institutions. Bitcoin mining is done through proof of work (PoW). Given its characteristics, the higher hash rate results in a higher probability of mining, leading to the emergence of a mining pool, called a mining organization. Unlike central processing units or graphics processing units, high-cost application-specific integrated circuit miners have emerged with performance efficiency. The problem is that the obtained hash rate exposes Bitcoin's mining monopoly and causes the risk of a double-payment attack. To solve this problem, we propose the error-correction code PoW (ECCPoW), combining the low-density parity-check decoder and hash function. The ECCPoW contributes to the phenomenon of symmetry in the proof of work (PoW) blockchain. This paper proposes the implementation of ECCPoW, replacing the PoW in Bitcoin. Finally, we compare the mining centralization, security, and scalability of ECCPoW and Bitcoin.

Keywords: error-correction codes proof-of-work (ECCPoW); proof-of-work (PoW); ECCPoW implementation; ASIC resistance

Green Bitcoin: Global Sound Money

Heung-No Lee, Young-Sik Kim, Dilbag Singh, and Manjit Kaur

Abstract

Modern societies have adopted government-issued fiat currencies many of which exist today mainly in the form of digits in credit and bank accounts. Fiat currencies are controlled by central banks for economic stimulation and stabilization. Boom-and-bust cycles are created. The volatility of the cycle has become increasingly extreme. Social inequality due to the concentration of wealth is prevalent worldwide. As such, restoring sound money, which provides stored value over time, has become a pressing issue. Currently, cryptocurrencies such as Bitcoin are in their infancy and may someday qualify as sound money. Bitcoin today is considered as a digital asset for storing value. But Bitcoin has problems. The first issue of the current Bitcoin network is its high energy consumption consensus mechanism. The second is the cryptographic primitives which are unsafe against post-quantum (PQ) attacks. We aim to propose Green Bitcoin which addresses both issues. To save energy in consensus mechanism, we introduce a post-quantum secure (self-election) verifiable coin-toss function and novel PQ secure proof-of-computation primitives. It is expected to reduce the rate of energy consumption more than 90 percent of the current Bitcoin network. The elliptic curve cryptography will be replaced with PQ-safe versions. The Green Bitcoin protocol will help Bitcoin evolve into a post-quantum secure network. In addition, it improves the properties of Bitcoin's hash PoW while addressing environmental concerns.

Keywords: Bitcoin, energy consumption, error-correction codes, post-quantum security, sound money, verifiable random function

I. INTRODUCTION

WE the global citizens not by our choice live in a world of boom-and-bust cycles created by the Federal Reserve Board (FED) of the United States. In the boom phase of a cycle, the FED supplies debt-mon-

시사점

- 검증된 AI 전문성 보유
- 새로운 Web3 클라우드 기술 제시
- 새로운 합의 프로토콜 GPU 자원 확보 제고

논문은 Docsend에서 확인할 수 있습니다.

고유 기술 스택

우리는 LLM 기반 에이전트를 위한 핵심 구성 요소들을 개발해 왔습니다!

Legal AI = Legal QA Corpus + LLaMa3 + Fine Tuning + RAG

IEEE Access

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2023.DOI

Legal Query RAG

RAHMAN S M WAHIDUR [1], SUMIN KIM [2], HAEUNG CHOI [1], DAVID SAMUEL BHATTI[1] and HEUNG-NO LEE [1], (Senior Member, IEEE)

[1]School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju 61005, South Korea
[2]Artificial Intelligence Graduate School, Gwangju Institute of Science and Technology, Gwangju 61005, South Korea

Corresponding author: Heung-No Lee (heungno@gist.ac.kr).

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2023-2021-0-00118, Development of decentralized consensus composition technology for large-scale nodes) and This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC (Information TechnologyResearch Center) support program (IITP-2023-2021-0-01835) supervised by the IITP (Institute of Information & CommunicationsTechnology Planning & Evaluation)

ABSTRACT

Legal practice has seen a significant increase in the adoption of products using Artificial Intelligence (AI) for various core legal tasks. However, these technologies are still in their early stages and lack the capability to effectively address domain-specific challenges. One promising approach that enhances response quality by integrating a curated external knowledge base into the Large Language Model (LLM) prompt is known as Retrieval-Augmented Generation (RAG). Therefore, this paper leverages fine-tuned LLMs, advanced RAG techniques, and an agent-based recursive feedback loop to evaluate RAG for legal Natural Language Processing (NLP). Initially, a general-purpose embedding LLM was fine-tuned using legal corpora. The fine-tuned LLM was then evaluated against baseline LLM, showing an average performance improvement of 13% in Hit Rate and 15% in Mean Reciprocal Rank (MRR). Furthermore, comparisons were made between general domain LLMs and a Hybrid Fine-tuned Generative LLM (HFM) specifically designed for the legal domain. The results demonstrated that the HFM observed average performance improvements of 24% across various tasks compared to baseline LLM. Finally, the evaluation of the proposed Legal Query RAG (LQ-RAG) architecture against the Naive configuration and the RAG with Fine-Tuned LLMs (FTM) configuration shows that the proposed RAG model significantly outperforms these baselines. In terms of average relevance score, it achieves a 23% improvement over Naive configuration and a 14% improvement

IEEE Access

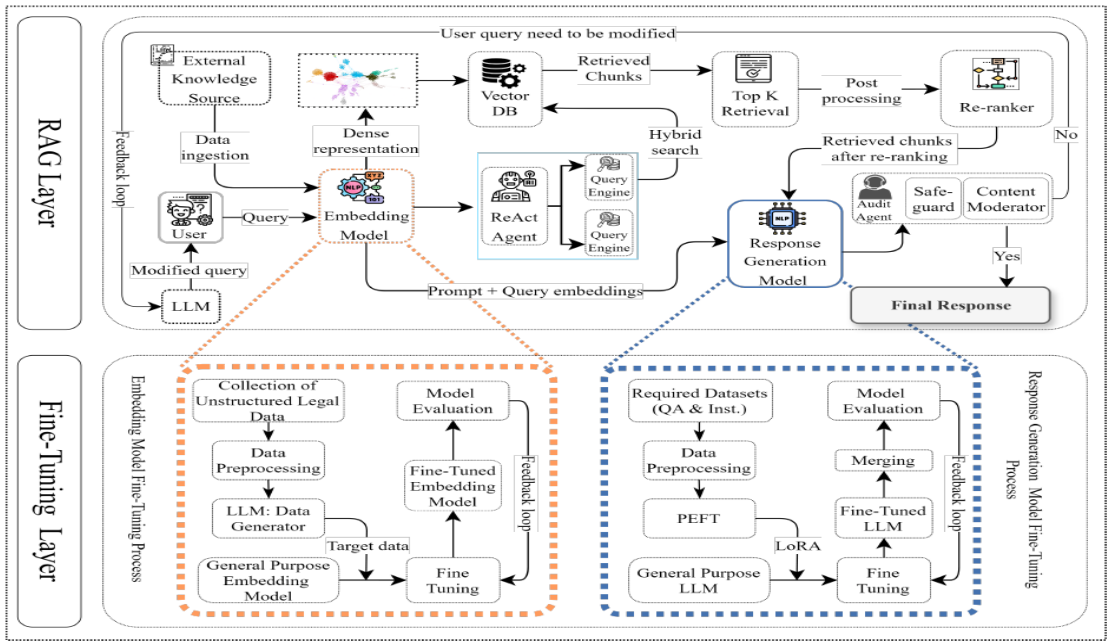


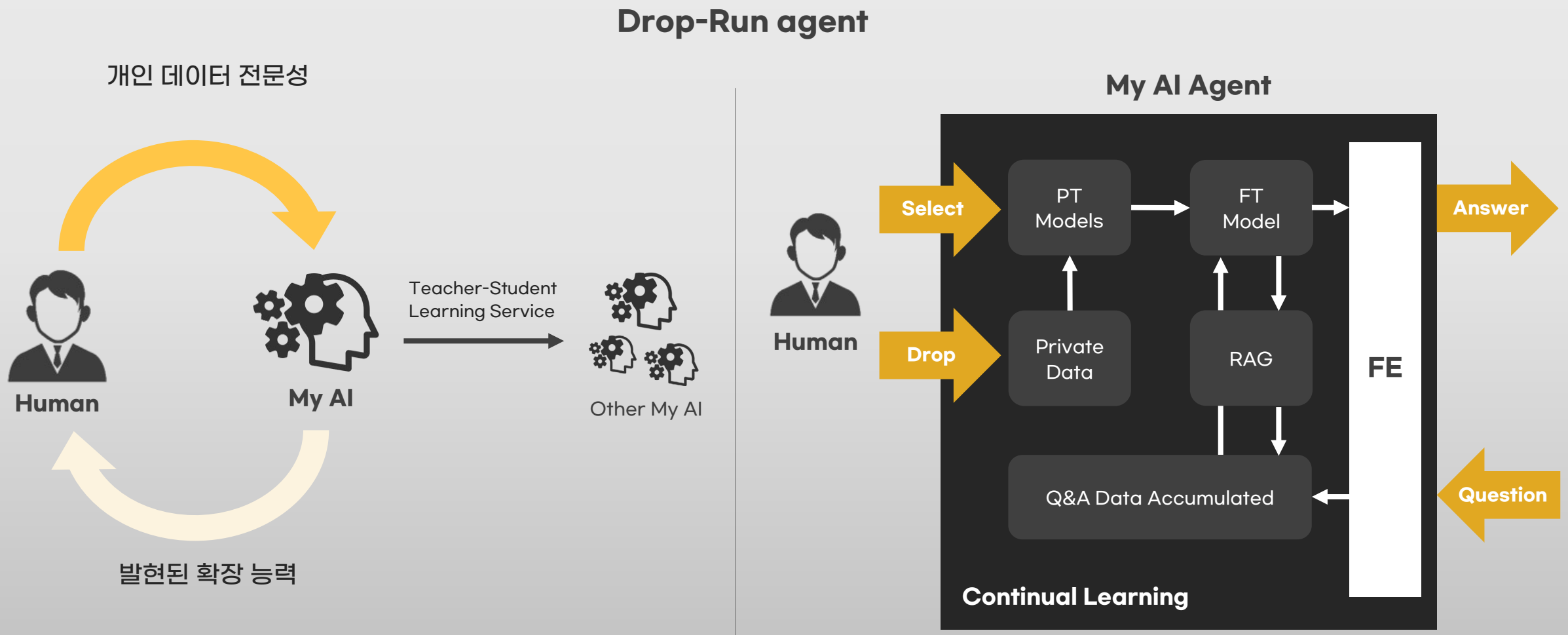
FIGURE 1: The Proposed LQA-RAG System Architecture.

LLaMa3와 FlanT5 같은 기본 모델과 비교했을 때,

우리의 법률 AI 어시스턴트는 모든 측정 기준에서 우수한 성능을 보이고 있습니다!

고유 기술 스택

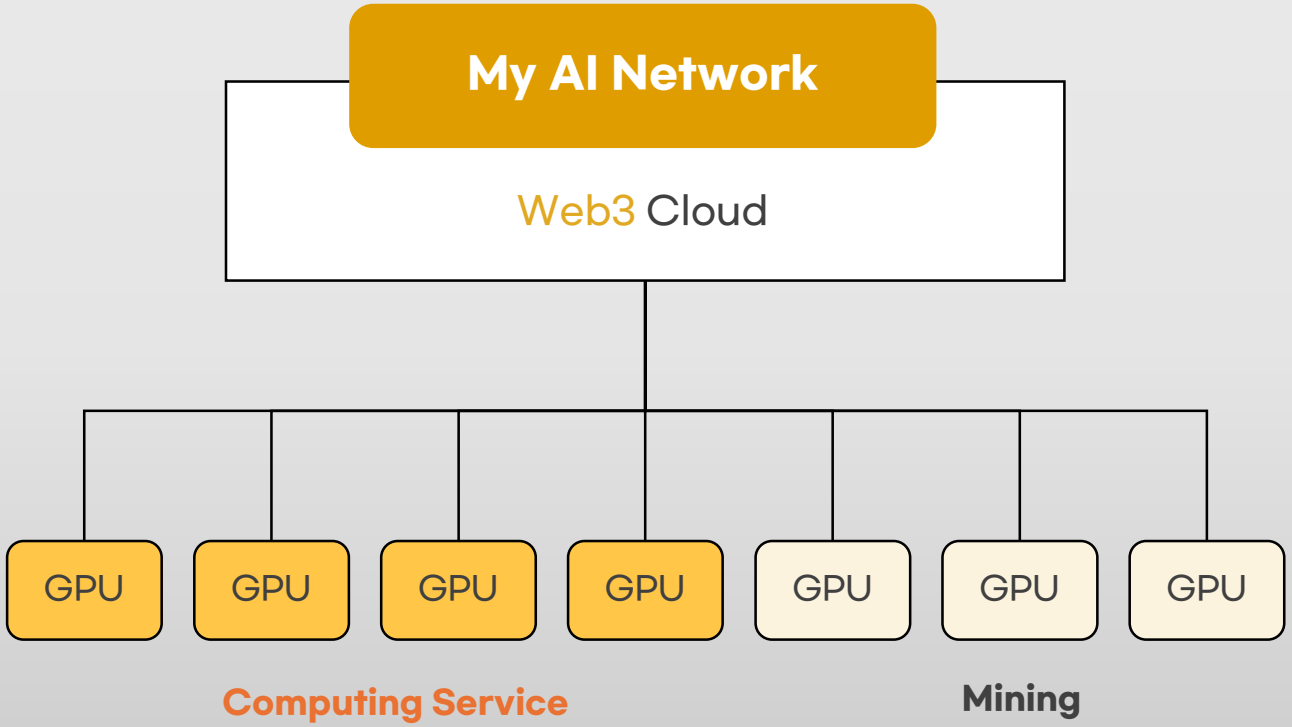
드롭 & 런 에이전트: 자신의 AI 에이전트를 훈련시키고 다른 에이전트들과 상호작용하게 합니다.



LiberVance

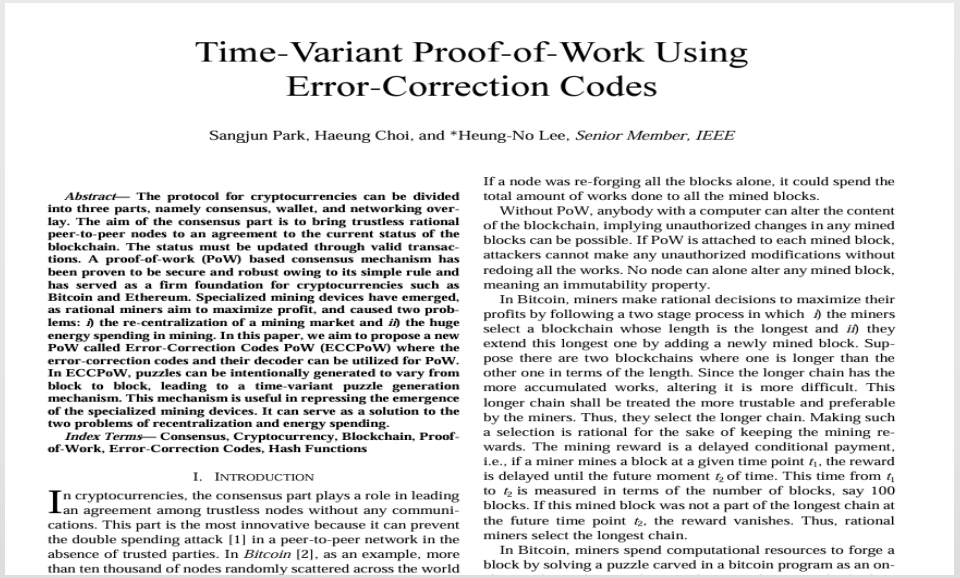
고유 기술 스택

우리의 솔루션 패키지는 블록체인을 검증 가능한 AI compute cloud로 업그레이드 합니다 !



Novel Consensus Layer

- Error Correction Code VCA ▶ ASIC resistance & promote GPU nodes
- Verifiable Coin Toss ~ Random self election ▶ **Mining** vs **Computing Service**



Time-Variant Proof-of-Work Using Error-Correction Codes

Sangjun Park, Haeung Choi, and *Heung-No Lee, *Senior Member, IEEE*

***Abstract*— The protocol for cryptocurrencies can be divided into three parts, namely consensus, wallet, and networking overlay. The aim of the consensus part is to bring trustless rational peer-to-peer nodes to an agreement to the current status of the blockchain. The status must be updated through valid transactions. A proof-of-work (PoW) based consensus mechanism has been proven to be secure and robust owing to its simple rule and has served as a firm foundation for cryptocurrencies such as Bitcoin and Ethereum. Specialized mining devices have emerged, as rational miners aim to maximize profit, and caused two problems: *i*) the re-centralization of a mining market and *ii*) the huge energy spending in mining. In this paper, we aim to propose a new PoW called Error-Correction Codes PoW (ECCPoW) where the error-correction codes and their decoder can be utilized for PoW. In ECCPoW, puzzles can be intentionally generated to vary from block to block, leading to a time-variant puzzle generation mechanism. This mechanism is useful in repressing the emergence of the specialized mining devices. It can serve as a solution to the two problems of recentralization and energy spending.**

***Index Terms*— Consensus, Cryptocurrency, Blockchain, Proof-of-Work, Error-Correction Codes, Hash Functions**

I. Introduction

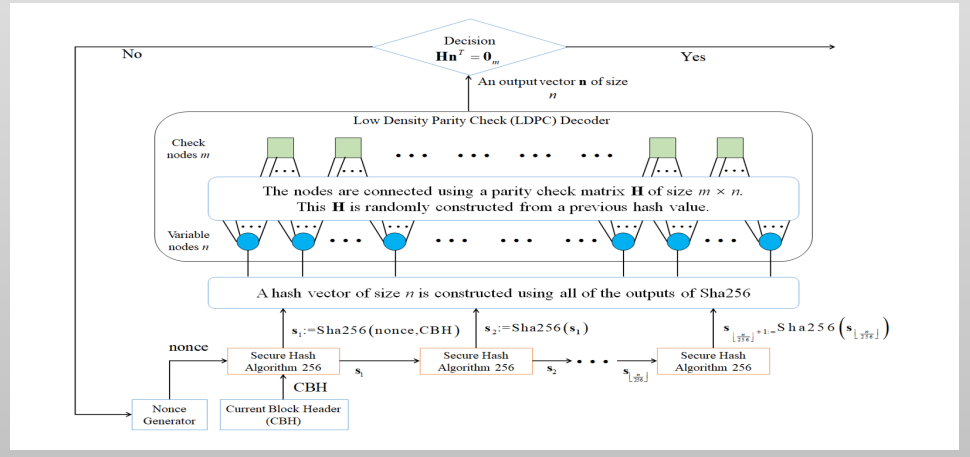
In cryptocurrencies, the consensus part plays a role in leading an agreement among trustless nodes without any communications. This part is the most innovative because it can prevent the double spending attack [1] in a peer-to-peer network in the absence of trusted parties. In *Bitcoin* [2], as an example, more than ten thousand of nodes randomly scattered across the world

If a node was re-forging all the blocks alone, it could spend the total amount of works done to all the mined blocks.

Without PoW, anybody with a computer can alter the content of the blockchain, implying unauthorized changes in any mined blocks can be possible. If PoW is attached to each mined block, attackers cannot make any unauthorized modifications without redoing all the works. No node can alone alter any mined block, meaning an immutability property.

In Bitcoin, miners make rational decisions to maximize their profits by following a two stage process in which *i*) the miners select a blockchain whose length is the longest and *ii*) they extend this longest one by adding a newly mined block. Suppose there are two blockchains where one is longer than the other one in terms of the length. Since the longer chain has the more accumulated works, altering it is more difficult. This longer chain shall be treated the more trustable and preferable by the miners. Thus, they select the longer chain. Making such a selection is rational for the sake of keeping the mining rewards. The mining reward is a delayed conditional payment, i.e., if a miner mines a block at a given time point t_1, the reward is delayed until the future moment t_2 of time. This time from t_1 to t_2 is measured in terms of the number of blocks, say 100 blocks. If this mined block was not a part of the longest chain at the future time point t_2, the reward vanishes. Thus, rational miners select the longest chain.

In Bitcoin, miners spend computational resources to forge a block by solving a puzzle carved in a bitcoin program as an on-

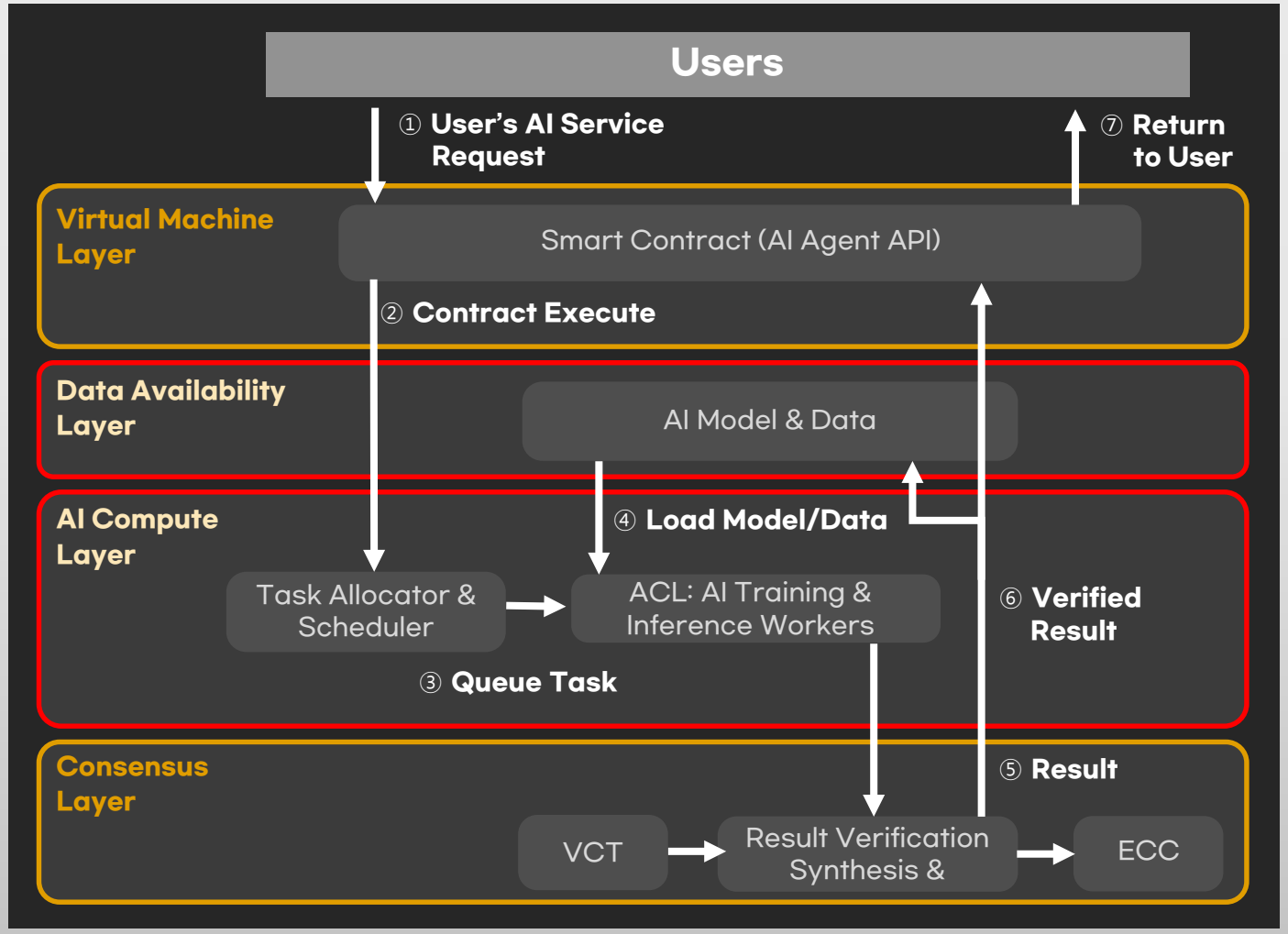


Paper PDF

LiberVance

고유 기술 스택

혁신적인 AI 컴퓨트 모듈이 스마트 계약 및 합의 계층과 상호 작용합니다.



AI 컴퓨트 모듈 절차

1. 사용자가 스마트 계약(Smart Contract)에 AI 서비스 요청을 보냅니다.
2. 블록체인이 스마트 계약을 실행합니다.
3. AI 계층이 AI 작업을 AI 컴퓨트 레이어(ACL)에 할당합니다.
4. ACL이 모델과 데이터를 로드하고, 작업이 완료될 때까지 매 블록마다 AI 컴퓨트 작업을 수행합니다.
5. 작업이 완료되면 ACL이 컴퓨트 결과를 C-layer에 반환합니다.
6. C-layer가 컴퓨트 결과를 검증합니다..
7. 검증 결과가 긍정적일 경우, C-layer는 합의 단계(예: PoS 혹은 PoW)로 진행합니다.
8. 요청된 작업이 완료되면 C-layer는 스마트 계약(SC)에 알리고, SC는 결제를 처리한 후 결과를 사용자에게 반환합니다.

고유 기술 스택

우리는 탈중앙화된 검증 가능한 데이터 저장소와 GPU 노드로 구성된 Web3 클라우드를 구축할 수 있습니다.

Global L1 Blockchain

My AI Network

Verifiable AI Compute Cloud

Distributed Ledger
for Tx & contract

Data Cloud
for AI model & data

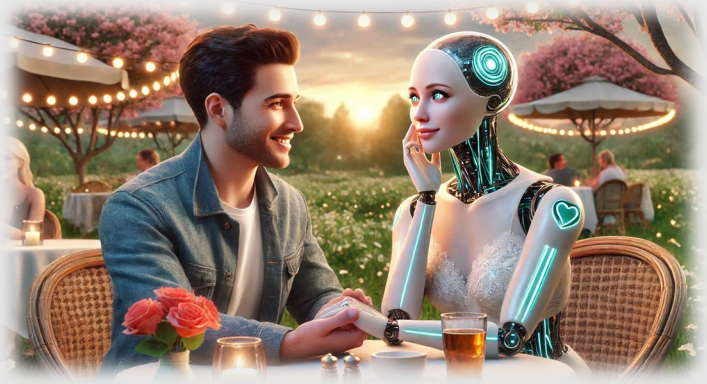
GPU GPU GPU GPU GPU GPU
AI Computing Cloud
for AI Training & AI Inferences

GPU GPU GPU GPU
Consensus
for Security & Verifiability

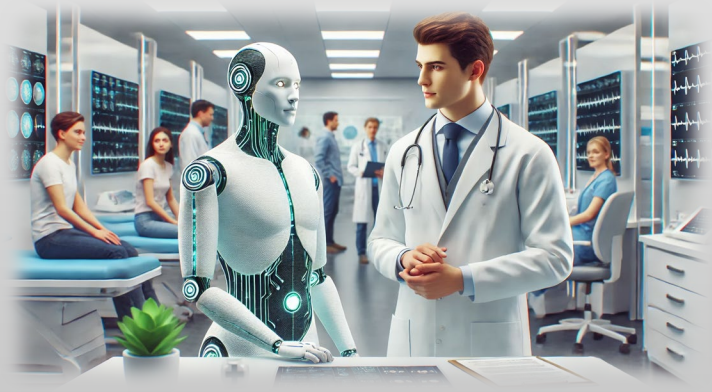
My AI Network는 전 세계에 분산된 노드들로 구성된 Web3 클라우드 체인입니다.

사용 사례

연구 및 개발 비서



돌봄 및 반려



의료 지원

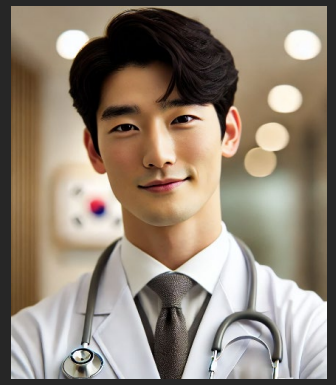
법률 지원

사용 사례: 의료 AI 에이전트

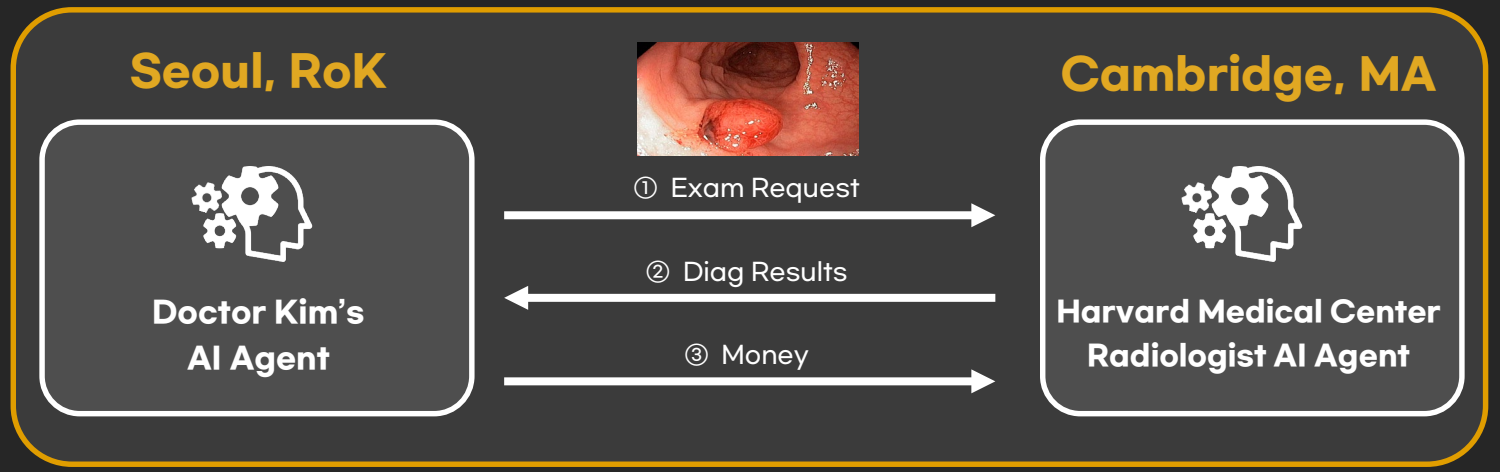
My AI 에이전트는 다른 에이전트에게 서비스를 제공하고 수익을 창출할 수 있습니다!

AI를 훈련시키고 당신의 전문성을 판매하여 시간과 국경의 장벽을 극복하세요!

- 하버드 의료 센터는 대규모 대장내시경 용종 이미지 데이터를 사용해 훈련된 AI 에이전트로 유명합니다.
- 한국의 임상 의사가 미국인 방문자를 진료하며 방사선 전문의 AI 에이전트에게 검사를 의뢰합니다.

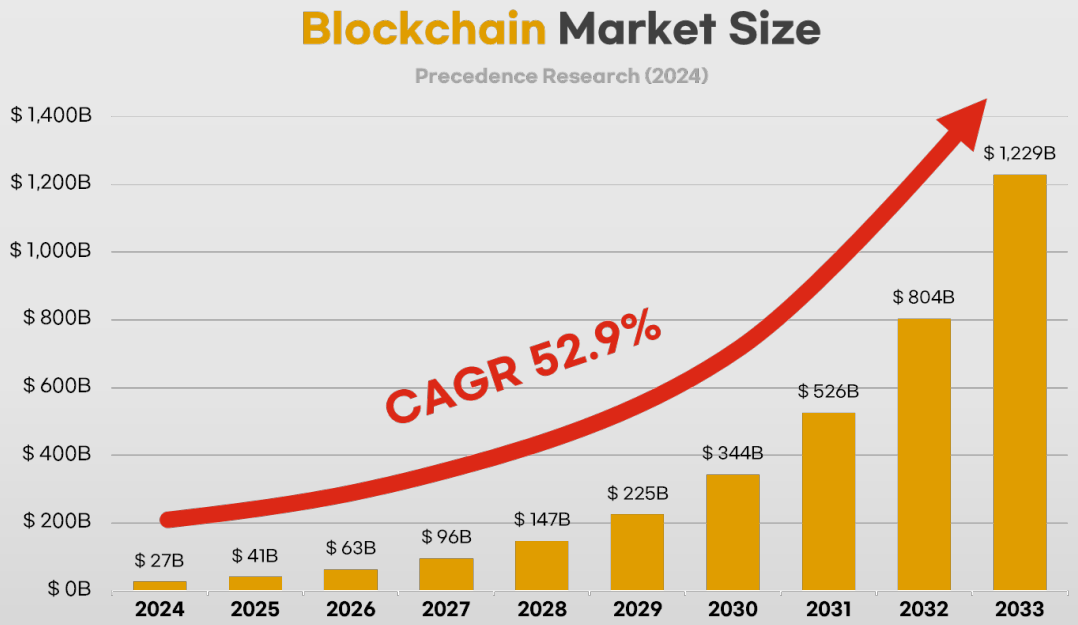


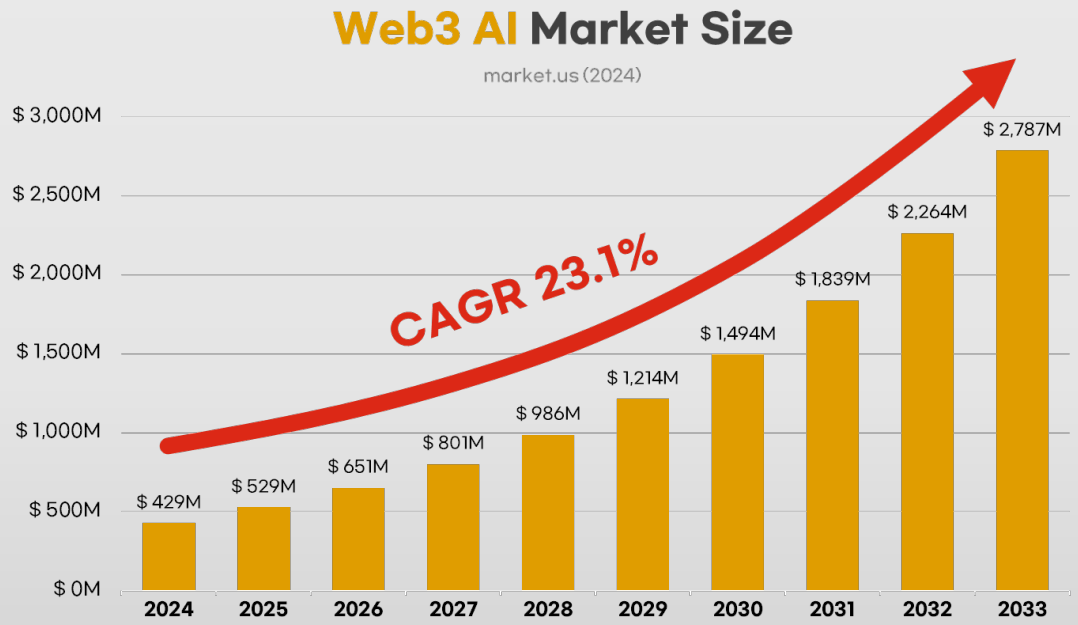
Doctor Kim, Seoul
Seeing American patient



Doctor Jones
Radiologist at Harvard

시장 기회





Blockchain 과 **Web3 AI** 시장 규모는 **급격히 성장하고 있습니다!**

LiberVance

매출 모델

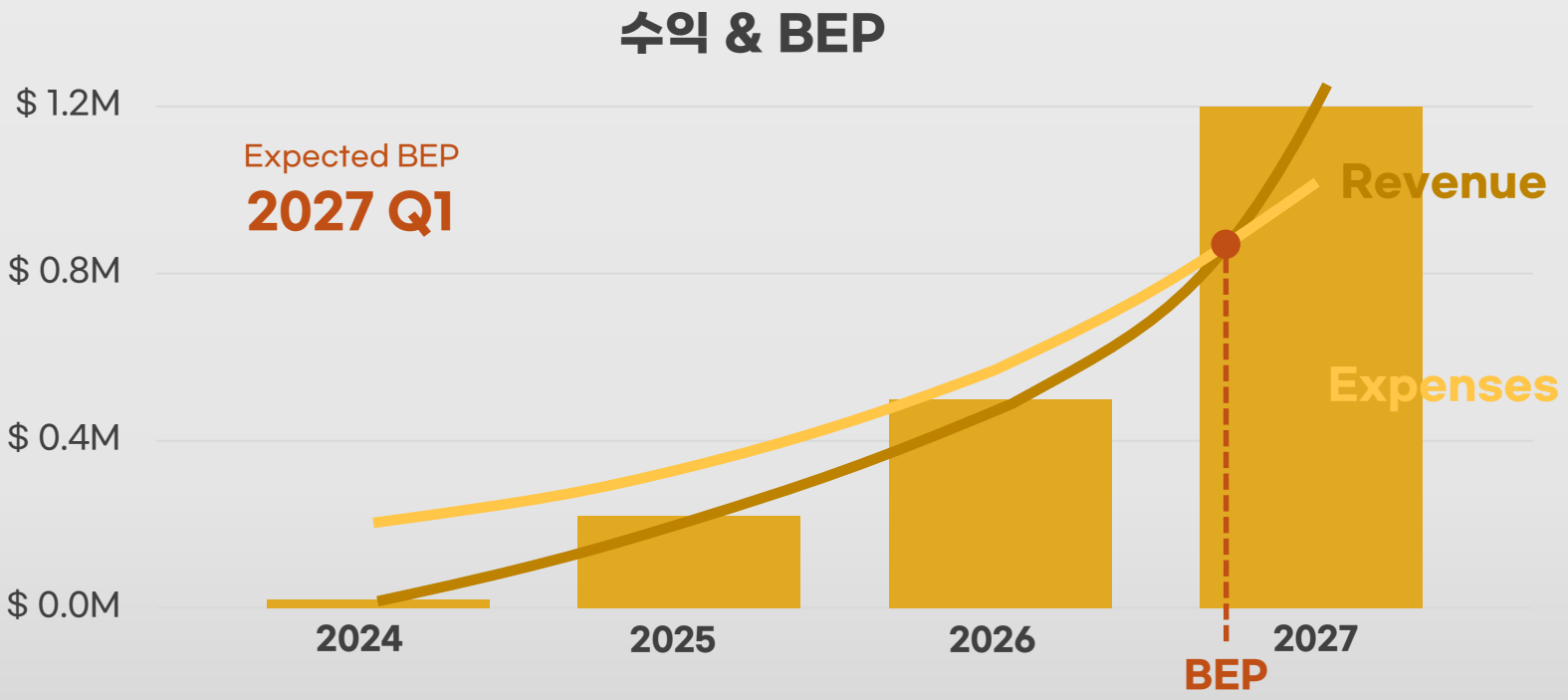
사용자는 이용 수준에 따라 요금 지불

	Free	Basic	Premium
Persona AI와 대화하기	20 Messages / 2hours	Unlimited	Unlimited
MY AI agent 만들기	1 Agent	Unlimited	Unlimited
MY AI agent와 대화하기	20 Messages / 2hours	100 Messages / 2hours	Unlimited
Agents 수익화	-	-	O
요금	$ 0	$ 5 /month	$ 10 /month

네트워크 기여자에게 수익 배분

Drop&Run 서비스 제공자	AI 컴퓨트 프로토콜 제공자	GPU 제공자	스토리지 제공자
40%	10%	30%	20%

출시 이후 3년간 매출 성장 목표



	2025	2026	2027
유료 사용자	**2K**	**4K**	**10K**
수익	$ 0.24M	$ 0.48M	$ 1.2M
운영 비용	$ 0.3M	$ 0.6M	$ 1.0M
순 이익	$ -0.05M	$ -0.1M	$ 0.2M

LV 단독 서비스 제공 기준

질문과 답변: My AI Network

11. 전세계적으로 볼 때 Web3 AI 사업을 진행하는 기업들이 있나요? 만약 있다면, What are their main focus areas and products?

1. **0g Labs:**
   - **중점 분야:** 0g Labs 는 분산형 AI 애플리케이션을 위한 모듈형 AI 블록체인을 구축합니다.
   - **혁신적 제품:** 이들은 데이터 가용성 계층(DA Layer)과 데이터 저장 계층(Data Storage Layer)을 결합하여 온체인 AI 모델 훈련과 신뢰할 수 있는 데이터 검증을 효율적으로 지원합니다. 목표는 블록체인 상의 AI 모델의 확장성과 보안을 강화하는 것입니다.
2. **Talus Network:**
   - **중점 분야:** Talus Network 는 분산형 자동화를 위한 스마트 에이전트를 개발하는 데 집중하고 있습니다.
   - **혁신적 제품:** 스마트 에이전트 기술을 블록체인과 통합하여 자율적인 스마트 계약을 생성하고, Web3 애플리케이션을 효율적이고 협력적으로 자동화할 수 있도록 합니다.
3. **Flock.io:**
   - **중점 분야:** Flock.io 는 분산형 연합 학습 플랫폼을 개발하고 있습니다.
   - **혁신적 제품:** 이 플랫폼은 원시 데이터를 공유하지 않고도 AI 모델을 공동으로 훈련시킬 수 있어 데이터 프라이버시와 보안을 보장합니다. Flock 의 AI Arena 플랫폼은 개발자가 모델을 훈련, 검증하고 보상을 받을 수 있는 탈중앙화된 Kaggle 과 유사한 환경을 제공합니다.
4. **dFusionAI:**
   - **중점 분야:** dFusionAI 는 분산형 AI 를 사용해 커뮤니티 기반 지식 그래프를 생성합니다.
   - **혁신적 제품:** 지식 융합 프로토콜을 사용해 커뮤니티 참여를 통한 데이터 검증을 수행하며, 참여자들은 토큰 보상을 받습니다. 주로 Web3 와 암호화폐 관련 주제에 중점을 두고 있습니다.

5. ChatGPT GPTs 에 데이터를 올려 학습시키고, 타인에게 서비스를 제공해서 소득을 만들면 되지않나요?

7. OpenAI 사는 GPT-4 서비스 공개로 전 세계인의 신뢰를 얻고 있어요. Web3 서비스는 뭔지 복잡해 보이고 경쟁할 수 있을지 모르겠어요.

13. Open-Source 기업이 성공한 사례가 있는가요?

15. My AI Network 가 만들어지면 사용자들은 누가 될까요?

16. My AI Network 를 통해 가능한 협업 시나리오로는 어떤 것들이 있을까요?

18. My AI Network 를 활용한 글로벌 협력과 그에 따른 투자 의사결정이 가능하여 개미들의 연대가 가능할 것으로 보이는데, 이 가능성에 대해 어떻게 보고 계신가요?

My AI Network

전 세계적 AI 에이전트 네트워크를 구축하여 사람들의 역량을 강화합시다!

이흥노

GIST교수 / CEO LiberVance