MarkerDAO Decentralizes AI Data Crowdsourcing

Steven Wong, Zhen Li, Shuguang Cui

markerdao.io

Abstract

Artificial intelligence (AI) shows great potential to bring the 4th industrial revolution in human history. Deep learning is driving AI to improve its knowledge and recognition of the world, disrupting and advancing various fields. This includes enabling more accurate medical diagnoses, improving natural disaster response times, and creating more efficient transportation systems. However, to achieve all of this, deep learning AI applications require large-scale annotated datasets which are then utilized for pretraining. e.g., ChatGPT is trained on trillions of data points which cost hundreds of millions of dollars for annotation/raw data structuring. MarkerDAO is a decentralized crowdsourcing marketplace focusing on AI data annotation. In general, training AI models requires large datasets which need to be structured and labeled. Structuring and Labeling data is a labor-intensive manual process. Nowadays, AI researchers still rely on services provided by centralized platforms such as ScaleAI and MTurk. However, in the Web3 era, MarkerDAO leverages blockchain technology to deliver a trustless, permissionless, and cross-border labor market being strategically incentivized by crypto-economics and having instant cross-nation payment settlements. These features significantly reduce the cost of dataset annotations for AI companies and increase earnings for data annotators. The whitepaper illustrates the design principles and underlying mechanisms of implementation.

1. Introduction

In 1986, Geoffrey Hinton proposed a backpropagation approach [1] which became the foundational theory of deep learning. In general, deep learning approaches require structured data with manual annotations or labels to train the model. Once the deep learning model is trained using the labeled data, the trained model can recognize and classify totally new input information and make relevant predictions accordingly. Nowadays, all advanced AI models are based on deep learning theory which requires large-scale structured and labeled data to train the algorithms. Thus, the AI data labeling industry is booming while also showing great promise to bring AI models at par with and beyond human intelligence levels.

There are three steps necessary to create a deep learning based AI.

1. Create an AI algorithm.
2. Collect and label Data manually.
3. Train the Algorithm with the annotated data.

However, the second step poses the greatest barrier since it requires massive labor work to structure or annotate data. Not providing the model with high quality annotated data will affect its training which in turn will result in the AI producing low quality results. This highlights the importance of providing the model with high-quality labeled data. The image in Figure 1 illustrates the procedure of the AI data annotation.

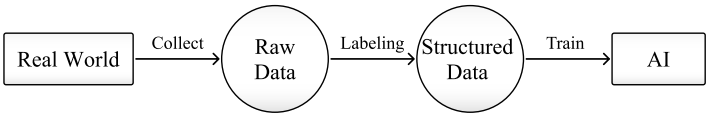

Figure 1: Procedure of AI Data Labeling.

1.1. Self-Supervised and Reinforcement Learning

In recent years, many AI scientists are trying to invent new approaches to replace fully supervised learning with self-supervised learning labels [2], and reinforcement learning learning [3] which could make AI learn from the raw data without labels. However, those methods suffer from some limitations and most of them only can

work in several special scenarios. For example, reinforcement learning is only effective to solve strategic puzzles such as chess and computer games and self-supervised learning only works for fundamental deep learning model pretraining [4] and shows no effectiveness in real-world applications.

1.2. Supervised Learning

In supervised learning field, researchers focus on significantly improving the labeled data scales and applying massive computations to train the AI algorithms, this has led to miracles and breakthroughs we could not imagine. For instance, the Alphafold2 which is proposed in [5] addresses the protein folding problem, the 50-year-old grand challenge in biology. Alphafold2 is trained with a large-scale dataset called the Big Fantastic Database (BFD) which contains billions of protein structures. ESM-1b which is proposed in [6] is a state-of-the-art protein language model which uses 0.25 billion i.e. 250,000,000 protein sequences to train the model and shows great performance in protein structure prediction. In 2022, OpenAI launched ChatGPT, achieving a milestone in the AI industry, and bringing confidence to humans in developing general AI, since the performance of ChatGPT is nearly on par with humans. However, the fact behind this performance is that ChatGPT utilizes the DaVinci-003 model, which is trained with trillions of words. The data annotation cost of ChatGPT is estimated to be around 300 million dollars. Therefore, the AI data annotation industry will be a key sector in the AI revolution.

Data Labeling Tasks are Labor Intensive

Most AI data labeling tasks can be implemented by generic human labor manually. For example, data for an image classification task can be easily labeled by ordinary people to classify dogs or cats in images, as shown in Figure 2. However, deep learning approaches usually require large-scale labeled datasets for training, which means there will be millions of different animal images to be labeled as either a cat or a dog by humans manually. This is a labor-intensive task.

Image annotation tasks are the easiest work among all supervised AI labelling tasks. Point cloud segmentation is an example of this where a radar collects millions of points from the real word e.g., a street view, as shown in Figure 3 but the category of points themselves remain unknown. To train the deep learning models [7] to recognize the points, the raw street view image data is to be labeled. The data labeling workers should give each point a label, e.g., point-cloud 1 belongs to the road, and that point cloud 2 belongs to the car. One 3D point cloud scan includes millions of points and one dataset contains millions of scans. Thus, one can get an

Figure 2: Labeled data for image classification which can be utilized to train the AI algorithms. the last column is labeled data manually.

idea of the humongous scale of the overall annotation work required to label these types of datasets. This work would require a large number of workers to spend thousands of hours on annotation. Once the dataset is labeled, it could be utilized to train deep learning algorithms for various applications such as self-driving cars, machine vision, object recognition, and detection.

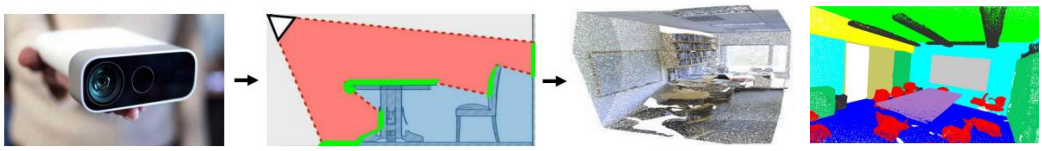

Figure 3: Labeled data for point cloud segmentation. The last image is the labeled data manually

1.3. The Middlemen in AI Data Annotation Market

Currently, there are two kinds of solutions to provide AI data labeling services.

The first category is labor-employ. These are centralized platforms that hire a massive number of workers who are trained in advance to label data, these workers aim to offer data annotation services to these companies and earn salaries. In this category, these workers usually have a long-term agreement with their employers

and receive fixed low salaries. e.g., Scale AI[1]. These labor-employ companies receive massive profits from AI research institutes and commercial companies, but distribute tiny incomes to the data labeling workers thus these companies act as a huge middleman.

The second category is on-demand crowdsourcing websites which allow AI companies to submit raw datasets with specific data labeling tasks, and funds and publish them to the marketplace. The individual workers could log in to the website, check the marketplace to annotate the raw dataset, and earn the division from the funds provided by the publishers. Conventionally, in such crowdsourcing platforms, workers are not required to sign long-term agreements with the website owners because is an open marketplace. As workers complete more tasks, their earning increases accordingly. Therefore, AI companies allocate larger funds to get their raw data annotated faster, since in these types of marketplaces workers prefer tasks that yield higher rewards. Amazon MTurk [2], Appen[3], and Fiverr[4] are the representative products in this category.

Exorbitant Commissions

However, whether labor-employ or on-demand crowdsourcing platforms, both are based on centralized entities and charge massive commissions to act as the middlemen which significantly increases the cost of AI data labeling and reduces the individual worker's income. Additionally, because of these huge commission fees large-scale datasets suffer from high costs for annotation making it impossible for AI companies to annotate them on such centralized platforms, hindering the AI research & Development of the industry.

Cross-Border Payment Issue

International payment issues limit the above-mentioned platform's growth to increase their workforce of data labellers worldwide which in turn limits their capacity for data annotation and often bounds them to limited geographical locations.

The fiat payment system usually only works well in particular national regions e.g., different countries use different payment systems and fiats which presents an inherent barrier for cross-board payments. The lack of flawless international payment

[1] https://scale.com
[2] https://www.mturk.com
[3] https://appen.com/
[4] https://www.fiverr.com/

settlement also hinders AI companies from accessing talent pools of data annotators present in different geographical locations of the world that are ready to do the annotation tasks at a fraction of what they are spending right now. This inconsistency of data annotation cost across various different regions of the world occurs because labor cost varies a lot among different countries.

More specifically, AI companies are mostly located in developed countries in which labor costs are usually expensive. Whereas, there is a hugely talented workforce of AI data workers present in third-world countries that usually require low costs, a result of being located in a poor economic environment. Thus, to access this huge existing talented labor pool without restrictions, an effective cross-border payment system is required. However, neither labor-employ nor crowdsourcing solutions cannot solve this problem efficiently since they only adopt fiats payment. While labor-employ companies are focusing on hiring workers with an agreement for offering data labeling services, this method is infeasible for crowd-sourcing platforms which involve micro-workers who work part-time with different fiat payment systems. The centralized crowdsourcing platform such as Amazon MTurk, because of these reasons at the time of writing, only allows workers from 25 countries to join their workforce and support fiat payments only to US bank cardholders while the rest of their workforce receives compensation in terms of amazon gift cards.

Centralized Trust Issue

The current solutions are built on trust in a centralized way between the AI companies and data labeling workers. But such centralized trust is fragile, especially in cross-border conditions with a distributed workforce. The platform can shut down at any time, and the terms of commission can be modified by the platform owner's will. The traditional law system fails cross-border workers and cannot guarantee the workers' rights and interests. Therefore, trust is hard to build between AI data labeling workers and the data annotation middlemen.

1.4. Cryptocurrency, Smart Contract, and DAO

In 2008, Satoshi realized that a fully decentralized system with miners would provide the first game-changing breakthrough for traditional finance, and purposed the Bitcoin whitepaper [8]. Utilizing blockchain technology it enabled the world's first decentralized cryptocurrency called Bitcoin. In 2014, Ethereum was proposed [9], introducing a new technology called Smart Contracts. Bitcoin only facilitates transactions of the Bitcoin cryptocurrency. Smart contracts are intended to build upon this technology to enable cryptocurrency to facilitate logic-based applications. With

blockchain decentralized EVM (Ethereum Virtual Machine), developers can build decentralized applications (Dapps) that could form the basis of the next-generation web (Web3). These Smart contracts allow for the development of self-executing applications that are transparent and decentralized. This means that various Web2 companies that charge huge fees for providing a platform could be just developed as a self-executing Dapp on the blockchain. Since the Dapp code is present and executed on the blockchain, it makes decentralized governance possible, further reducing the necessity of Web2 companies. These centralized companies and institutes can now be replaced by decentralized autonomous organizations (DAOs) which are a set of decentralized smart contracts deployed on the blockchain and gives web service participants rights to share incomes from the treasury of the DAO and becomes involved DAO's governance by voting. The Web3 protocol functions with the same effect as DAO but with more generic options.

2. MarkerDAO

We deliver MarkerDAO in this section to leverage blockchain and smart contracts to address all the above issues. MarkerDAO aims to build a Mark-to-Earn (M2E) platform that leverages the global workforce for crowdsourcing AI data labeling tasks to advance model training and R&D of AI companies.

2.1. Participant Coordination

There are four roles in the DAO workflow for AI data labeling which are **Markers, Validators, Fishers, Publishers** as shown in Figure4.

- **Marker** or labeler is the annotation worker who takes labeling tasks and provides annotation services for unlabeled raw AI data which is the most important role in the ecosystem. The marker should read the annotation guide carefully and perform the labeling work whose reputation score will be slashed once the incorrect annotations are verified by validators.
- **Validator** can be the same group of people of Marker who perform verification works one by one for marker's outputs to determine their qualities by classifying them as acceptance or rejection. Both Marker and Validator are supervised by the third role - Fisher.
- **Fisher** is the DAO senior member and the expert in AI Data annotation works and knows well on the data annotation principles. Once the verification

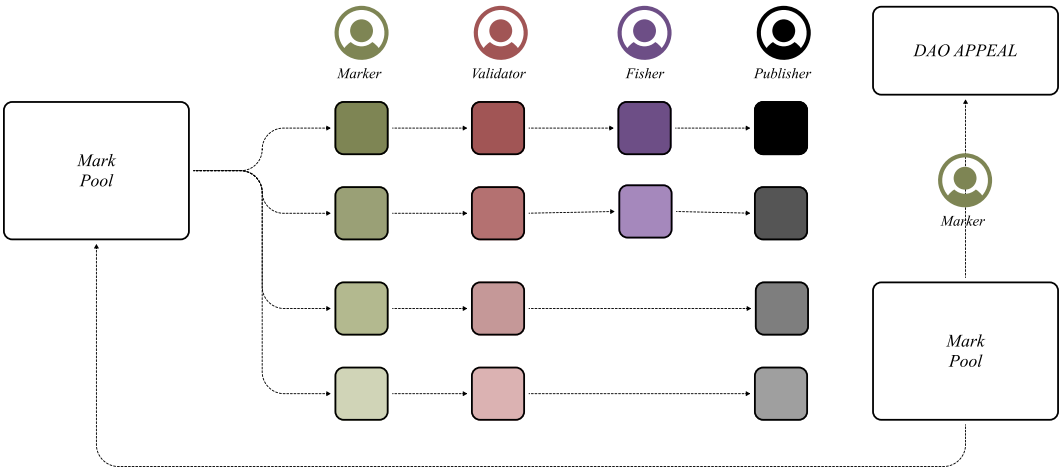

Figure 4: Participants behaviors with 4 roles: Marker, Validator, Fisher, and Publisher.

works are implemented by validators. An AI model will be used to select the inspection candidates among all the markers for Fisher verifications. Once verification is rejected by Fisher, the corresponding Marker, and Validator's reputations will be slashed. Otherwise, the annotations will be delivered to Publisher for the final check.

- **Publisher** is the data labeling jobs creator who usually is the AI R&D companies and research institutes which provide the raw data to MarkerDAO for annotation or labeling. The labeled data will be utilized to train the AI algorithms like ChatGPT to inject knowledge from data into the model and improve the performance.

2.1.1. Reputation Score System

The reputation score is adopted in MarkerDAO to evaluate the performance of Marker, Validator, and Fisher. The reputation score is increased, once the data labeling delivery is accepted by Publisher, whereas slashed when the outputs of the Marker and Validator are rejected by Fisher and Publisher. The Fisher reputation score only can be slashed by Publisher when the delivery is rejected. The overall reputation score is 1000, once less than 0, the user cannot take any tasks from Publisher. The initial reputation score is 500 when the user passes the LST (Labeling Standard Test).

2.1.2. Decentralized Identity and LST

The user joins MarkerDAO and wants to be Marker and Validator, which requires LST (Labeling Standard Test) for AI data annotations. The test will examine the user's fundamental skills for various data annotation jobs as follows.

- **Languages Annotation Test** includes Sequence Labeling, Text Summary, Question Answer Matching, .etc.
- **Image Annotation Test** includes Object Detection, Segmentation, Caption, .etc.
- **Medical Annotation Test** includes Medical Knowledge Understanding, Lesion Classification and Detection, Medical Data Tools, .etc.
- **3D Data Annotation Test** includes Point Cloud Detection, Segmentation, Caption. Point Cloud Tools, .etc.

Once the user gets acceptance in a Test, the system will transfer an SBT (soul-bound token) to that user's address as a certificate or DID (decentralized identity) which grants the user to take the related tasks for performing annotations and verification.

Particularly, the user gains acceptance by all the above tests and could apply to become Fisher on the DAO panel, and plays a crucial role in the quality assurance of annotations.

2.2. Payment Mechanisms

Publisher could create the AI data annotation jobs with payment to MarkerDAO smart contract. MarkerDAO accepts two kinds of payments which are crypto and fiat. The crypto payment can be one of USDT, USDC, DAI, and MAR. MAR is the utility token of MarkerDAO which is detailed in Section 4. If MAR is selected as payment by Publisher, the commission will be reduced against other payment currencies.

Nowadays, there are still AI companies that play the role of Publisher but do not recognize crypto payment and still adopt fiat money in their business. Thus, we adopt Marker Agent to transfer the fiat to cryptos.

Marker Agent

The Marker Agents are the centralized companies who act as AI data labeling service providers. Marker Agent takes tasks from the AI companies and research institutes with fiat payment to create corresponding labeling jobs on MarkerDAO with crypto payment. Therefore, the Marker Agent acts as the middleman between fiat payment Publishers and MarkerDAO to transfer their fiats to cryptos and interact with smart contracts.

3. Technical Implementation

In this section, we introduce technical implementation which consists of three parts: decentralized workflow engine (DWE), Upgradeable Dapp structure, and AI Boosts Data Labeling. Figure 5 illustrates the architecture of the overview of MarkerDAO technical design.

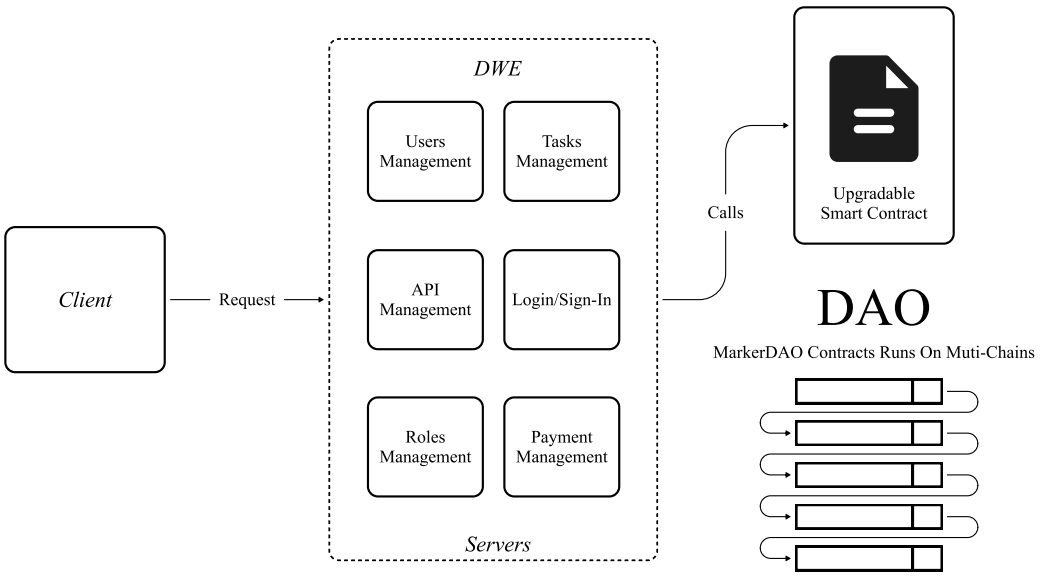

Figure 5: Architecture Overview

3.1. Decentralized Workflow Engine (DWE)

The platform is a decentralized AI dataset labor crowdsourcing platform based on blockchain, which can be implemented with a decentralized smart contract platform

for task distribution. These tasks are the discrete units of composition, and the units are transferred between different roles for approval and are finally completed and settled.

The implementation will focus on the design of a task-based decentralized workflow. We deliver a decentralized workflow engine (DWE) in the whitepaper which consists of smart contracts with related DAO governance functionalities.

As shown in Figure 6, there are three layers: Network Layer, Workflow Layer, and Blockchain Layer.

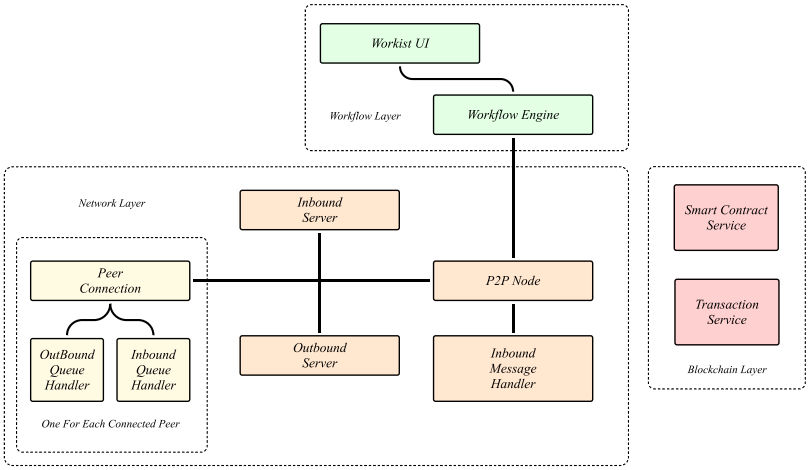

Figure 6: 3 layers of Decentralized Workflow Engine.

- **Network Layer** is constructed with a P2P infrastructure with a certificate authority that issues keys to participating actors.
- **Blockchain Layer** includes transaction services, block services, and mining services, and is implemented on top of the P2P layer. These services appear on all proof-of-work blockchains. The transaction service manages the pool of pending transactions that are created by the local workflow service or received from inbound message handlers and verified upon receipt.
- **Workflow Layer** consists of a workflow engine and a worklist handler with a user interface, implemented on top of the blockchain layer. When a transition is enabled, the workflow engine on the node to which it is assigned creates a work item and populates its input values from the current values of the workflow instance. The work item is then added to the local worklist. After a work item is executed (manually or through an external application

call), the output values are written back to the workflow instance. Data constraints can be specified for each activity, and these constraints are checked as part of the transaction validation performed by the transaction and block services. After the work item is completed, the local workflow engine submits the corresponding workflow transaction to the blockchain.

CovenantSQL is partially adopted to store the user information which is a decentralized blockchain SQL database. DWE receives HTTP RESTful API requests from users on the front end. The requests can be classified as login, registration, user information management, task creation, task submission, role permissions, payment, .etc. Particularly, there is corresponding permission management for the APIs which acts as the security assurance for the whole system. DWE interacts with the decentralized database CovenantSQL and the corresponding blockchain contracts.

Segmented Tasks Distribution Mechanism

Publisher could launch the whole dataset as one task in MarkerDAO and determine the split size which could be used to split the whole task into multiple small assignments with equal size. The related users such as Marker, Valiadtor, and Fisher cloud work in an appropriate size instead of the whole dataset. Assignments are the smallest indivisible units that can be processed by the DWE. The assignments will be distributed to the mark pool and reviewed by quality inspectors and spot checkers assignment-wisely.

Role Based Access Control

In the workflow, there is a corresponding role processing along with the assignment state. The role management is based on the RBAC (Role-based access control) permission model, as shown in the Figure 7.

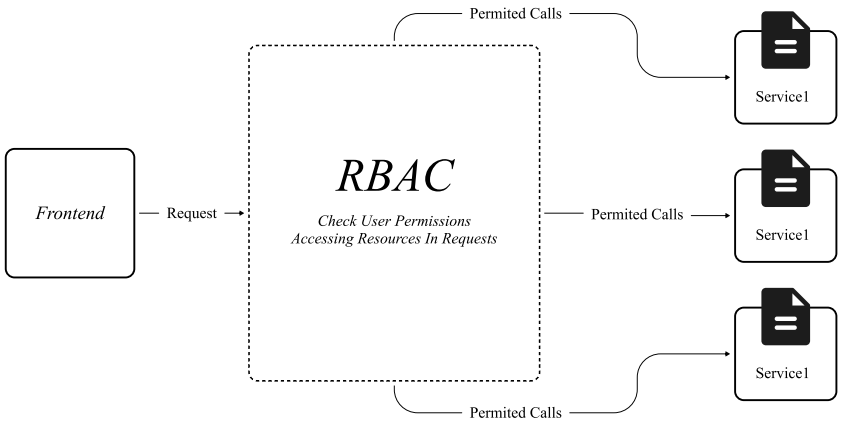

Figure 7: Overview of RBAC Model.

The role indicates a collection of permissions, and the set of permissions of the role can have an intersection or not. Permissions can be classified into read-only, write-only, and other resource operation permissions. Once a user is assigned to a particular role, the operating privileges are limited to a specific range of operating resources. In the RBAC model, a user could be assigned multiple roles, but the principle is the same - essentially a concatenation of sets of role permissions.

RBAC model grants the users the exact access rights required by their roles. The governance panel of DAO can easily update permissions for specific roles.

3.2. Upgradeable Dapp Implementation

We design an upgradable Dapp architecture to make the whole decentralized system highly scalable. The proxy contract is the key to implementing it. Specifically, the transactions sent by users can be forwarded to different versions of MarkerDAO contracts through the proxy contract. Since the smart contracts on the blockchain are immutable, the requests could use the proxy contract to redirect to the specific implementation version address for the smart contracts. Therefore, A flexible contract upgrade mechanism can be achieved. Through the operations of the proxy contract, the context of the data state is shared between different versions, so there is no need to do data migration, which is equivalent to achieving a seamless and smooth smart contract upgrade.

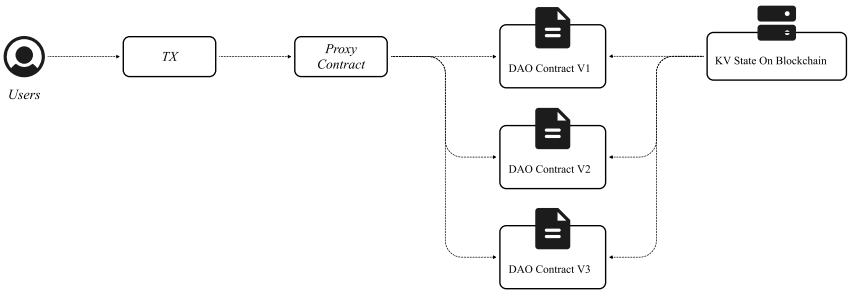

Figure 8: Upgradable Decentralized Application Architecture.

Multi-Chain Structure

The smart contract of MarkerDAO will be implemented in various languages. e.g., Solidity, Golong, Rust, !ink, Move, and deployed on multiple public blockchains. The marketplace data is shared among multiple blockchains, and the payment and transactions will be conducted on different chains with cross-chain bridges. For instance, the Publisher and Marker could make the deal in the marketplace of MarkerDAO, but their wallets are on different blockchains. Therefore, the cross-chain bridge will be applied to support transactions on different blockchains.

The first version of MarkerDAO will be implemented with Solidity and only support ETH mainnet and other layer2 evm-compatible blockchains.

Task Manager

Task Manager will be implemented by smart contracts. Our core design is to design two core smart contracts, probably a TaskManager and a Task contract, where the Manager is responsible for managing the life cycle of the Task contract, including the creation of the Task, the destruction of the Task, and the state tracking of the Task. The Task contract itself is a relatively small state machine that only represents the most critical states of the task to avoid unnecessary gas overhead.

The task contract also stores the addresses of task-related Markers. This way, payroll is issued when the task is finally settled.

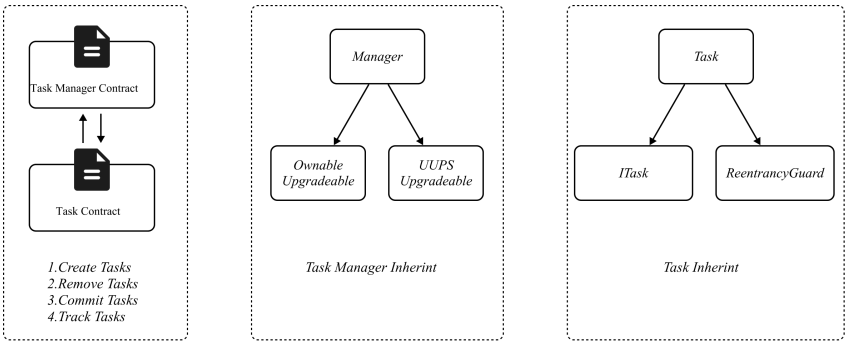

Figure 9: Contract Minimal Core Architecture

As you can see, the Task Manager in figure 9, the first parent contract ensures that Manager's owner can transfer ownership. The second parent contract ensures that the Manager contract is upgradeable. Additionally, the Task contract inherits a ReentrancyGuard parent contract for added security, as the Task contract will eventually be involved in the on-chain settlement of Markers. The ReentrancyGuard contract helps prevent reentrant attacks on the contract.

So, what is the relationship between the Manager and the Task contract? Actually, it can be understood like this, if we consider the Task contract as a process in the operating system, then the Task Manager acts as a process manager. The Manager is responsible for the life cycle of all Tasks, including the creation, deletion, submission, and tracking of certain metadata. This metadata includes the Task ID, the wallet address of the associated Marker, and the hash-based URL address of the Task's associated dataset packet storage.

3.3. AI Boosts Data Labeling

Although AI learning requires labeled data, similar to AI empowering other industries, we also could adopt AI to facilitate and accelerate the data labeling processing which will be illustrated by the next two sub-sections.

AI-assisted Spot Check

As we mentioned, Fisher is the role to do a spot check to inspect the low quality annotations which are labeled by the bad-performance Markers. However, the spot check not means that each assignment will be verified by Fisher, whereas, only a small proportion of assignments of Markers will be selected and verified by Fisher.

The traditional sampling approaches for spot check mainly focus on probability models e.g., a normal distribution that is straightforward and rough with low efficiency. In MarkerDAO, we finely design a deep learning models to learn from paired data which is formed by a triple array (Marker, Assignment, Quality). The paired data can be collected from the historical records of MarkerDAO operations. Once the paired data forms a large scale, the designed deep learning model could be trained by the collected paired data and automatically grade the assignments with quality scores for Fisher Spot Check. Particularly, M_i denotes the $i-th$ Marker, and A_i denotes the assignment annotated by Marker M_i. The scoring model is E. We have the quality score S_i for assignment A_i which can be calculated in Equation. 3.3.

$$S_i = E(A_i)$$

Once all the quality scores are calculated, the bad-score assignments will be selected for Fisher Spot Check. Therefore, we could largely improve the hit rate for the spot check by adopt deep learning models.

AI-assisted Data Labeling Tools

Data labeling is a non-trivial task, and one of the critical components of the supervised learning pipeline. It is one such task that requires a lot of manual effort. So, then can we get the bulk of such mundane, labor-intensive & time-consuming effort to be autonomously driven by machines aiming to minimize the bulk of human tasks. We focus on this generic universal problem with our intuitive approach to largely alleviate the bottlenecks of having limited labels or the need of labeling tons of instances all by yourself from scratch.

Several multi-task models of CV, and NLP will be trained to facilitate and accelerate the labeling process. For example, in the object labeling task, the AI-assisted data labeling tool could automatically recognize most of the cars in Figure **??**. Only a few misses recognized cars should be labeled manually by laborers.

Figure 10: Pre-Labeling results of AI-assisted data labeling tool.

Therefore, the AI-assisted Labeling Tool could significantly improve the efficiency of AI Data Labeling which further could largely reduce the cost for Publisher since fewer Workers can perform more annotations.

4. Tokenomics

As shown in Section 2.1, four roles participate in the AI data labeling works by applying the DAO protocol and are incentivized by strategic tokenomics. For Marker, the Marker takes out the data package from the Mark pool, after labeling, and sends it to the quality inspector Validator. If Validator finds that there is a problem with the quality and calls back to Marker, Marker's Token will be calculated to do a pre-deduction according to the number of problematic data, but this pre-deduction is not a deduction. It is just similar to a record. When the final project is completed, the final deduction is made and the reputation value is reduced, which is designed to avoid excessive power of the Validator. On the contrary, Marker gets paid for his work and a certain amount of platform tokens and reputation value.

The Validator role, its main responsibility is to conduct a comprehensive quality check on each marked packet, so its responsibility is also relatively heavy, if the packet it passes the test is found to be faulty by Fisher, the random checker behind, then the packet is sent back to the Marker and the Validator is punished, but this time, the Marker is not punished. This time, the Marker is not punished, but also to reduce the disadvantages to the Marker, for maximum fairness, because imagine, if the Marker is punished, then with the quality check the more hurdles, in fact, the more unfair to the Marker, which is a chain of responsibility model.

The role of Fisher, its main responsibility is to selectively sample packets that have been vetted by Validator, and its reward and punishment mechanism is the same as Validators. If Fisher finds a faulty packet, he will punish Validator, but if the packet passes the sampling, he will not reward Validator, which is to prevent Fisher and Validator from colluding with each other to gather wealth. This chain of responsibility rewards, all only the final acceptance through, rewards and penalties will be executed, the latter role checks out the problem, will only punish the previous role.

For the Publisher, to ensure as fair as possible, there is only a corresponding token punishment mechanism, no reward mechanism, after all, the Publisher is equivalent to the role of an employer, it has the final decision and the employed personnel is a strong group. Once Publisher's acceptance is passed, then all the roles in the chain of responsibilities involved in this one task will be rewarded, but once Publisher refuses to pass and puts the task into Fix Pool, Marker raises objections to the community platform, and if the ruling is passed, then Publisher is punished, and if it is passed, Publisher's If the ruling is passed, the Publisher's reputation value will increase, but there will be no corresponding token reward. If the ruling is passed, the publisher's reputation will increase, but there will be no token reward.

For the rewards, we use the principle of proof of fraud in Optimistic Rollup, leaving

a window of time for locking the reward funds, during which any character can object to the result of the task. This is to prevent fraud with the highest probability possible. After the window has expired, the reward will be paid. The final settlement is completed.

Token Fundamentals

The native digital cryptographically-secured utility token of the Marker DAO (MAR Token) is a major component of the ecosystem and is designed to be adopted for use as the primary token on the network. MAR Token will be issued as EVM standard-compliant digital tokens on multiple blockchains.

MAR Token is designed to be a utility token that functions as the unit of payment and settlement between participants who interact within the ecosystem on the Marker DAO platform. While staked, it also represents rights and interests in the DAO governing system. MAR stakeholders are also entitled to fees generated by the Marker DAO platform.

Initially, MAR is distributed over 5 years. Beyond that, there will be no more new MAR tokens introduced to the supply.

MAR for MarkerDAO

There are two ways to pay for tasks on Marker Dao. Publishers can choose to either pay by fiats or by cryptos (USDT or MAR). As Marker DAO is a global platform, fiats will be converted to cryptos for cross-border transactions. The normal payment flow is listed below:

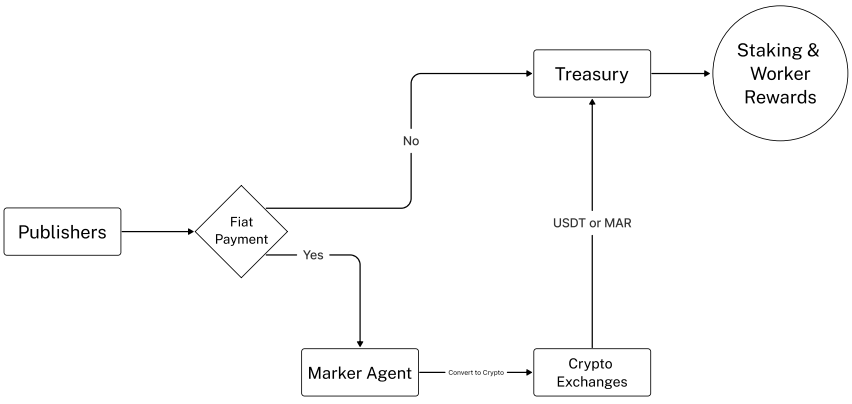

Figure 11: MAR Releasing Schedule

For continuing operations, Marker DAO platform charges a 20% publishing fees for tasks. However, MAR is designed to offer discounts and other benefits for publishers and workers on the platform:

- Reduced Platform Fees for Publishers: MAR offers a significant benefit to publishers on the Marker DAO platform. By using Mar as a payment method, publishers will receive a 25% discount on the platform fees charged. This reduction in fees will lead to significant cost savings for publishers.
- Increased Earnings for Workers: Workers on the Marker DAO platform can also benefit from using MAR. By choosing to accept MAR as payment, they will receive a 6.25% increase in their earnings. This increase in earnings will encourage more workers to use Mar and contribute to the platform.
- Faster Transaction Times: MAR is designed to offer faster transaction times than traditional payment methods. By using MAR, publishers and workers on the Marker DAO platform can enjoy quicker transactions, reduced waiting times, and increased efficiency.
- Enhanced Security: MAR is a secure cryptocurrency built on blockchain technology, offering participants on the Marker DAO platform enhanced security. By using Mar, participants can benefit from the platform's secure and decentralized nature, reducing the risk of fraud and hacking.

4.1. Staking

Yields

By staking MAR, users can enjoy multiple benefits. First, MAR block rewards are available for the first 5 years for stakers. The MAR release schedule is determined based on the formula, where y is the amount of MAR rewards to be released.

$$y = 500,000,000 * (1 - x/5)^2 \tag{1}$$

In addition to MAR block rewards, stakers will also earn a portion ($> 50\%$) of the platform's total revenue by backing great workers. Workers who have lots of backers will also likely earn more. The exact earnings will depend on the staking duration and the total staking amount, as well as the RG PoS system (Reputation Guided Proof of Stake).

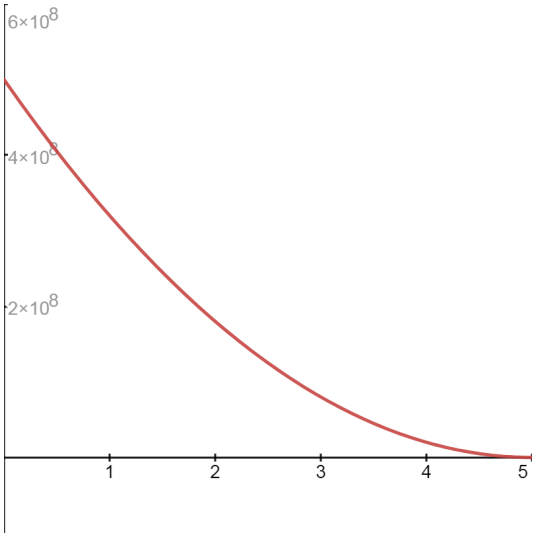

Figure 12: MAR Releasing Schedule

To join the staking system, stakers back markers using MAR. Workers have reputation scores, which are a reference for the quality and amount of their work. Workers gain a reputation for completed data labeling tasks but lose it if their work is rejected by publishers. Markers with more backers and completed work earn more rewards, while backers with the most backed workers have a higher probability of earning rewards. Dilution occurs if too many backers are backing the same workers. Rewards are calculated based on RG POS, and if workers fail jobs, backers' profits are reduced.

In the RG PoS system, when stakers back the same workers, the earnings will depend on staking duration. For instance, if the base annual percentage yield (APY) is x, stakers who stake for 1 month, 3 months, 6 months, and 12 months will earn 1x, 1.2x, 1.3x, and 1.5x of APY, respectively. However, it is important to note that stakers who unstake early will forfeit their earned stakes.

Ever-Protected Stakes

Ever-Protected Stakes

MAR prides itself on providing one of the most innovative and robust staking protection mechanisms in the market. Unlike other tokens where stakers do not get any compensation when the token's value declines in USD, MAR uses loss adjustment algorithms to give back to the stakers until they break even.

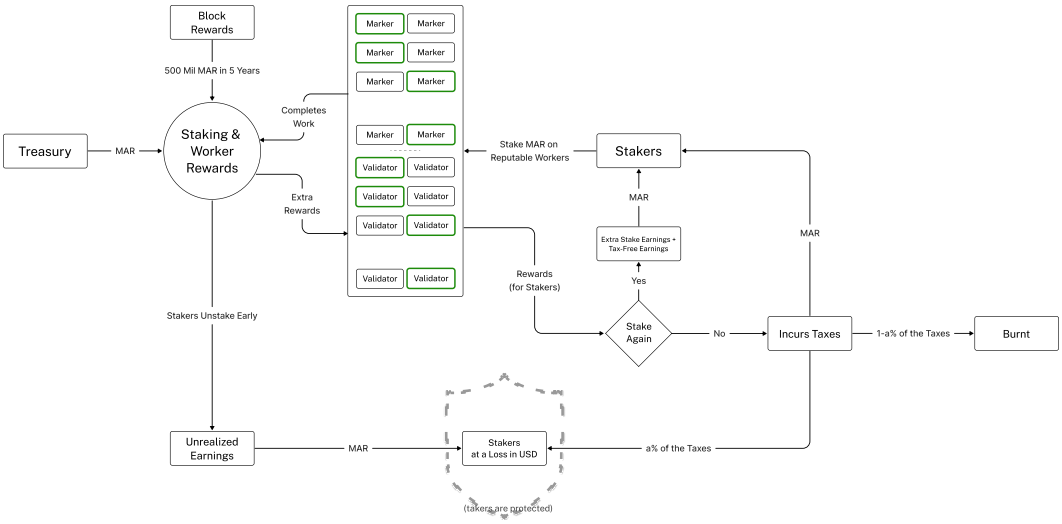

Figure 13: Ever-Protected Staking

For example, when a user chooses to unstake their MAR early, the system would utilize the dynamic price oracle to determine if any other stakers are currently in a loss in USD. If there are such stakers, the forfeited MAR tokens will be allocated to them. If no stakers are in a net negative position in USD, the forfeited earnings will be burnt to benefit the entire ecosystem.

Another key component of the loss adjustment mechanism comes from the unstakers profits, which will be described in the following section, which would provide an additional layer of protection to stakers, ensuring that they are not subjected to significant losses due to market volatility. This mechanism, coupled with the staking yields, makes MAR a compelling investment opportunity for those seeking to participate in the decentralized microtask platform ecosystem.

Perpetual Earnings

Long-term supporters are the core of the MarkerDAO community. To encourage stakers to stay with MarkerDAO, anyone who withdraws their rewards immediately after they are incurred will face a 40% tax on their total equivalent USD profits. Half of the taxed amount will be burned, while the other half will be distributed to remaining stakers who still lost money in USD, providing them with strong protection against any losses and rewarding their loyalty.

However, stakers who stake their earned rewards again will enjoy tax-free earnings after 6 months and be rewarded with additional rewards depending on their staking

duration. This mechanism is designed to encourage stakers to stay invested for the long term and further strengthen the stability and security of the MAR ecosystem.

4.2. Service Mode

Two service modes are designed in MarkerDAO for Publishers to release the AI data labeling jobs which are Marketplace and MarkerDAO Pro.

Marketplace

In Marketplace mode, the Publisher releases the jobs with payment on the smart contract but MarkerDAO cannot guarantee the time costs or how many days and months that are required since it is full market behavior. For instance, the Publisher raises the rewards higher which could attract more Markers comes to take the tasks, otherwise, fewer Markers are interested. Thus, it is a fully open market operation and the platform does not have a specific date of delivery which is only determined by the market.

MarkerDAO Pro

In the pro mode, the DAO committee will arrange a dedicated service team to work for the Publishers with the AI data labeling jobs and provide guidance for them. In this mode, the Publisher could 'create jobs and forget'. The service team will help Publisher to supervise the data labeling progress to guarantee the delivery on time. The corresponding premium will be paid by Publisher to incentive the service team, thus this mode could be more expensive than the Marketplace mode.

4.3. Governance

The MAR token is also adopted as a governance token as the certificate to join the DAO events on the governance panel. The DAO events can be classified into the following categories.

- **Appeal Proposal** can be raised by any role in MarkerDAO to claim the disputes between the others. For instance, Marker works for the data labeling loyally whose outputs are also verified by Validator and Fisher, but still rejected by Publisher. The Marker could raise an Appeal Proposal to claim the

facts for judgment by the community. Once the Appeal Proposal got accepted, the Publisher payment in the smart contract will be automatically transferred to the corresponding users.

- **Upgrade Proposal** is raised for improving and upgrading the platform from different points such as security, multi-chain supports, UI/UX improvement.
- **Maintenance Proposal** aims to fix the bugs and defects of the DAO implementation. It could be technical or non-technical such as document improvement and correction.
- **Generic Proposal** fits other motivations which not belong to the above categories.

The treasury will be applied to support DAO events in terms of any economic costs.

5. Conclusion

The AI industry's requirement for data labelling service is ever-growing with high demand and fewer options to meet the requirement. In this white paper MarkerDAO illustrated its plan to systematically drive innovation in the AI data labelling sector to eventually become the forefront platform in AI data labelling in the industry. Firstly, MarkerDAO proposes to leverage crypto payments which enables it to expand its data labelling workforce to a global level thus increasing its capacity for labelling data on a global scale, something which current data labelling platforms are struggling with. By implementing the Reputation score system, DiD and LST builds a robust system of an effective workforce that guarantees a high quality labeled data standard for AI companies. The Decentralised Workflow Engine (DWE), Upgradeable App infrastructure coupled with the MarkerDAO governance ensures that MarkerDAO is a platform that is always ready to adapt according to industry requirements. All of this makes MarkerDAO the number one platform for AI companies to label their data.

The unique AI-assisted sampling mechanism is a novel mechanism that trains itself on its own previously labelled data to pre-mark new data sets, this is intended to reduce the time for MarkerDAO markers to label data, validators, and fishers to validate the labels and publishers to receive their order, thus decreasing the time taken and increasing economic efficiency for each order on the MarkerDAO platform. Lastly, by proposing the protocol native utility token $MAR, MarkerDAO intends to revolutionize the entire data labeling industry. $MAR acts as the unit of payment and settlement for parties in the ecosystem while also acting as the governance token

when staked. $MAR enables 25% reductions in cost borne by AI companies to label their data and 6.25% increase in revenue by the MarkerDAO workforce. $MAR 5-year staking offers a lucrative opportunity for 3rd party participants to buy the token and stake it for high returns. With its one of a kind "Loss adjustment algorithms" it penalizes early un-stakers, provides additional rewards to long-term stakers to break even in case of loss suffered by market fluctuations and lastly provides stakers with 50% revenues generated by the protocol depending upon the stake duration. To conclude, by leveraging blockchain, decentralization, and novel token economics MarkerDAO provides the AI industry with an extremely efficient, highly scalable, and economically lucrative platform for AI data labelling services. A platform that is designed to take a lead, direct the majority, and achieve a conglomerate monopoly in the AI data labeling industry.

References

[1] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[2] K. Kavukcuoglu, P. Sermanet, Y.-L. Boureau, K. Gregor, M. Mathieu, Y. Cun, *et al.*, "Learning convolutional feature hierarchies for visual recognition," *Advances in neural information processing systems*, vol. 23, 2010.

[3] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26–38, 2017.

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[5] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, *et al.*, "Highly accurate protein structure prediction with alphafold," *Nature*, vol. 596, no. 7873, pp. 583–589, 2021.

[6] A. Rives, J. Meier, T. Sercu, S. Goyal, Z. Lin, J. Liu, D. Guo, M. Ott, C. L. Zitnick, J. Ma, *et al.*, "Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences," *Proceedings of the National Academy of Sciences*, vol. 118, no. 15, p. e2016239118, 2021.

[7] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 652–660, 2017.

[8] S. Nakamoto, "Bitcoin whitepaper," *URL: https://bitcoin. org/bitcoin. pdf-(: 17.07. 2019)*, 2008.

[9] V. Buterin *et al.*, "A next-generation smart contract and decentralized application platform," *white paper*, vol. 3, no. 37, pp. 2–1, 2014.