Master Capstone Project Proposal
Blockchain AI Marketplace

Chenhan Ma
823289

1. Executive Summary

Machine Learning and Artificial Intelligence algorithms are the hottest topic around the world, thousands of researchers and developers are developing and researching new solutions to improve the training result of various datasets and use them in different application scenarios. This proposal will briefly introduce a Blockchain protocol that can able to connect the resource of the data all around the world to enormous group of students and researchers.

This protocol will solve two problems using Blockchain technology: one is by gathering more models from the crowd to improve the output result, one is to enable anyone to upload their labelled data to the net therefore improve the accuracy of Machine Learning process with larger amount of dataset.

2. Project Description

This project is aiming to improve the current Machine Learning community mode using Blockchain technology. The application will be deployed on the Ethereum network using Smart Contract, as a decentralized application. Here I won't detail the Blockchain too much. And for the storage of the data, the Blockchain will store a metadata IPFS record pointing to the actual location of the data.

As described in the previous part, two modes of will be introduced below:

- **Data Crowd Sourcing**

    Take a real world as an example, as described in [1], researchers are

working to build a system that can allow people to share their medical data. With a larger amount of data, researchers are able to improve the diagnose with the help of Machina Learning, which can also solve the security issue by gaining control to the user of their data.

- **Machine Learning Competition**

  This mode acts like a decentralized version of Kaggle. That anyone is able to upload their dataset and expected reward to the Blockchain, and anyone is able to download the data and train them, then upload their model to win the reward. The smart contract will be able to evaluate the result of the model and add the one with the highest accuracy to the chain and pay the uploader the reward.

3. Use Cases

Those are simple use cases that I can think of, which have limited usage with some drawbacks. Further improvements will be made later.
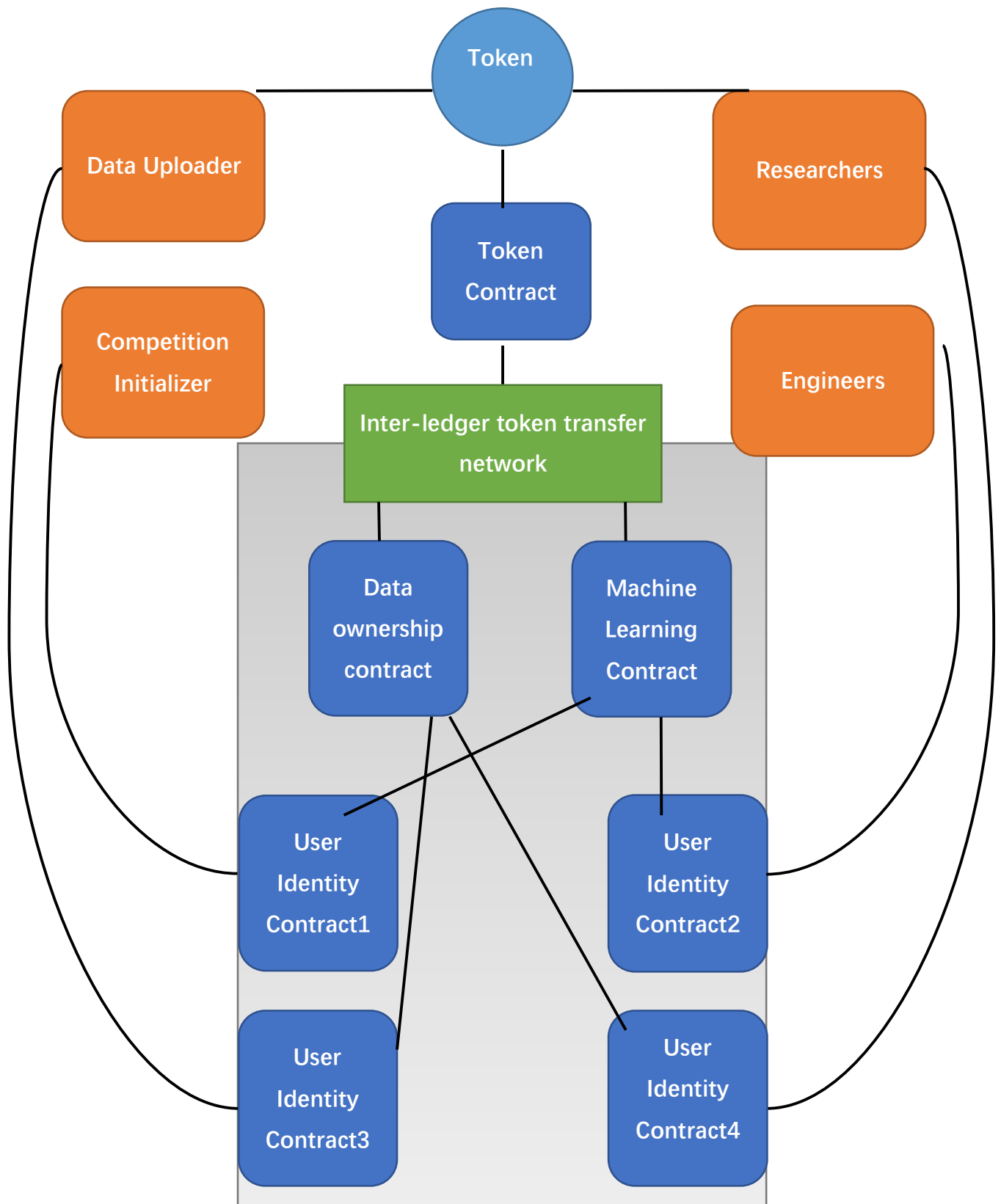
- **Data Crowd Sourcing**

  User or institution upload their medical data or dataset (labelled or unlabeled separately) to earn a certain amount of token as reward. Researchers can pay the token to get the right using the dataset.
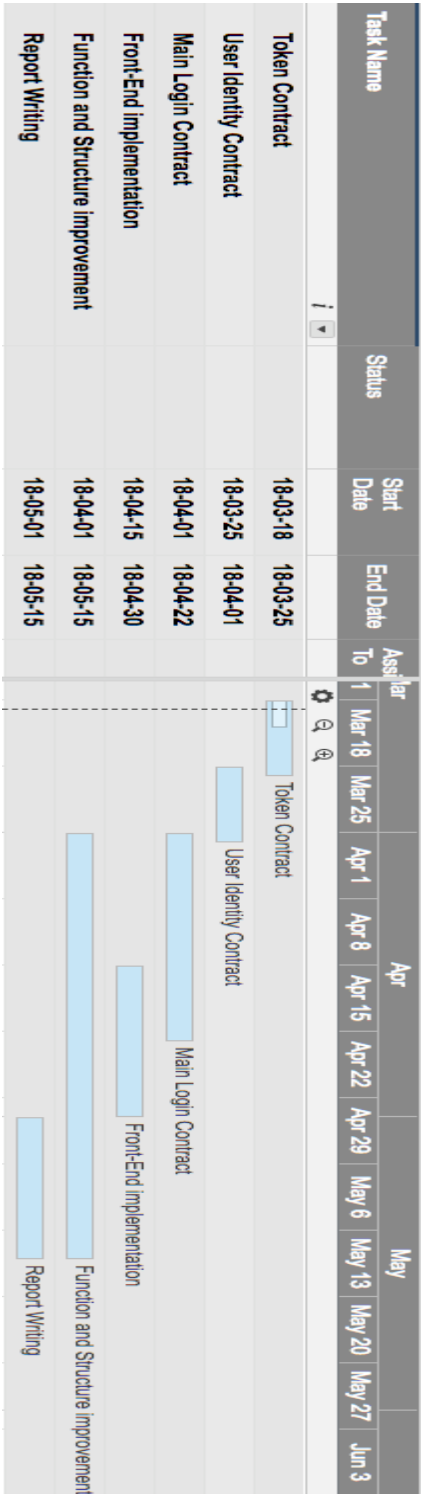
- **Machine Learning Competition**

  User creates a new contract with dataset and reward, then upload to the Blockchain consuming some gas. Then engineers download the contract and train with their model. After uploading their model to the Blockchain, the smart contract picks the best model as winner to pay with the reward. And the initial user will get the right to use the model.

4. Entity Relationship Diagram

5. Project Timeline GANTT Chart

Task Name	Status	Start Date	End Date	Assign To
Token Contract		18-03-18	18-03-25	
User Identity Contract		18-03-25	18-04-01	
Main Login Contract		18-04-01	18-04-22	
Front-End Implementation		18-04-15	18-04-30	
Function and Structure improvement		18-04-01	18-05-15	
Report Writing		18-05-01	18-05-15	

(Gantt chart timeline columns: Mar 18, Mar 25, Apr 1, Apr 8, Apr 15, Apr 22, Apr 29, May 6, May 13, May 20, May 27, Jun 3)

Bars: Token Contract, User Identity Contract, Main Login Contract, Front-End Implementation, Function and Structure improvement, Report Writing

6. Reference

[1] AI researchers embrace Bitcoin technology to share medical data, https://www.nature.com/articles/d41586-018-02641-7