published on 31-Dec 2017

updated on 09-Apr 2018

Decentralized machine learning

WHITE PAPER

decentralize data,
processing power and
algorithms for machine learning

www.decentralizedml.com

Table of Contents

1 Executive Summary

1. Executive Summary

In the era of rapid development in artificial intelligence and machine learning, data quality and relevancy are essential to generate usable applications of high quality and accuracy for machine learning.

As compared to publicly-accessible data, privately-held data are more relevant and timely for machine learning. These private data are usually untapped and inaccessible as they are stored in individual electronic devices such as smartphones, tablets and computers. Leading technology firms try to access these private data when individuals are unaware of or by providing free service to them in return. Nevertheless, these leading technology firms can only obtain a portion of the private data, which are subset of the massive untapped private data owned by all individuals.

Decentralized Machine Learning (DML) protocol is designed to expand the reach to untapped private data and unleash their potential to facilitate machine learning development while providing economic incentives and protecting data privacy. Machine learning algorithm will be run on the devices without extracting the data from the devices, which will be kept within the devices. Only the machine learning result will be aggregated with outcomes generated from other devices to form an unbiased, comprehensive and accurate crowdsourced analytics and predictions. Through DML protocol, both the private data and processing power for machine learning are decentralized as algorithms are run directly on individual devices by utilizing their idle processing power.

The quality of the algorithms is as critical as data quality to generate good usable applications and conduct accurate predictions through machine learning. With the belief in collective intelligence, a machine learning developer community will be built by means of establishing a marketplace for talented developers to list their algorithms for sale. Any parties, who want to apply usable applications and conduct predictions through machine learning, can locate or request suitable algorithms from the marketplace. Innovation from the periphery are encouraged and enormous potential in machine learning are unleashed through decentralizing the algorithm development.

To facilitate participations in DML protocol, economic incentives and a reliable environment are needed. With the aid of blockchain smart contract, a distributed trustless system is created that connects participants of the DML protocol directly without the need to trust a centralized third party. It ensures participants, who successfully contribute their effort such as allowing algorithms to be run on individual data, building machine learning algorithms by developers etc., to be rewarded in DML tokens.

DML protocol is a scalable decentralized infrastructure that aims to connect potentially billions of devices and tens of thousands of developers to facilitate machine learning development while providing decentralized ownership of data and artificial intelligence. Many possibilities will be unlocked through crowdsourced untapped data and collective intelligence. We expect and support innovation from the periphery to promote unconventional, ground-breaking and diverse innovations from all over the globe. Such decentralized machine learning protocol is too powerful to be controlled by a few persons, monopoly or oligopoly, hence it should be owned collectively and allow mass participation through decentralization. The whole decentralized DML ecosystem will keep evolving. The ecosystem allows every participant to equally and fairly contribute equally and fairly to the machine learning development for a better humanity.

2 Mission Statement

2. Mission Statement

Our goal is to create a blockchain-based decentralized machine learning protocol and ecosystem through:

1. utilizing untapped private data for machine learning while protecting data privacy,

2. connecting and leveraging idle processing power of individual devices for machine learning,

3. encouraging involvement from the periphery by creating a developer community and algorithm marketplace that promotes innovation to build machine learning algorithms that match practical utilities,

4. improving and correcting existing machine learning algorithms and models through crowdsourced fine-tuning model trainers,

5. creating a new DML utility token and leveraging on blockchain smart contract technology to provide a trustless and middle-man free platform that connects potential contributors in machine learning from all aspects.

3 Market Potential

3. Market Potential

3.1 / Application of Big Data for Machine Learning

3.1.1 Huge Market Potential in Existing Data

The internet and availability of big data set are key elements that lead to the recent tremendous development in machine learning. In the past decades, leading technology firms such as Google, IBM, Amazon, Facebook, Microsoft etc. have already invested heavily in machine learning research and development. With the increased interest and investment in machine learning, established technology firms, start-ups and research companies continue to develop and enhance the applications. This creates the need to acquire more data and algorithms for machine learning.

According to the research conducted by International Data Corporation (IDC), it is expected that the revenue for big data analytics will grow from US$130 billion in 2016 to over $210 billion in 2020[1]. IBM also sees the huge potential in machine learning markets and predicts a market worth of US$2 trillion over the next decade.[2]

3.1.2 Unlock Massive Market Potential of Untapped Private Data

Given the existing available data that each company has collected separately, the market potential in machine learning and big data analytics are already predicted to be enormous. However, these existing available datasets are just tip of the iceberg. If the untapped private data, like the bottom of the iceberg, is being tapped and applied to machine learning, its market potential and revenue generated can be even greater than the current market forecasts. Therefore, we see massive market potential and significant advancement in machine learning for our decentralized machine learning protocol.

[1] Big Data and Business Analytics Revenues Forecast to Reach $150.8 Billion This Year, Led by Banking and Manufacturing Investments, According to IDC, International Data Corporation [Website], March 2017, https://www.idc.com/getdoc.jsp?containerId=prUS42371417

[2] Through Machine Learning, IBM Braintrust Sees Better Days Ahead, Fortune [Website], February 2016, http://fortune.com/2016/02/25/ibm-sees-better-days-ahead/

3.2/
Utilize Decentralized and Idle Processing Power

Nowadays, machine learning is often conducted through a centralized computer cluster with limited processing power . If additional processing power can be utilized to run more machine learning algorithms, the market potential is tremendous.

Currently, there are approximately 2.3 billion smartphones and 2 billion personal computers that are not fully utilized. This implies that there is a total of 4.3 billion of electronic devices with underutilized processing capacity for machine learning. If DML protocol can connect the idle processing power of these devices and build a decentralized magnificent processor for machine learning, the extra market revenue predicted to be generated can be tremendous.

3.3/
Unleash Creativity to Build ML Algorithms and Models with Practical Utility

Large corporations are increasingly applying machine learning in their business According to Harvard Business Review, global companies are applying machine learning in anticipating customer preference, improving media purchase etc..[3] Furthermore, the review also states that there can be immediate benefits on revenue generation and cost saving. Giant corporations like Amazon and Microsoft see fraud, bad debts and costs are reduced while enjoying significant increase in revenue and operational efficiency.[4]

Often only the large corporations can afford investing huge initial capital and resources to build in-house machine learning algorithms or acquire tailor-made ones from consultancy firms and apply in their own businesses. Therefore, machine learning applications are often only found in some large international corporations rather than local or medium sizes companies.

We see huge potential growth if more companies can apply machine learning algorithms and models in their businesses provided that the initial cost required to acquire machine learning algorithms and models or engage outsourced programmers to build tailor-made algorithms and models becomes more affordable. We foresee mass adoption in machine learning in the commercial world when the associated cost is lower and when the prediction made by machine learning is more accurate.

Our DML protocol is the solution.

[3] Harvard Business Review,
https://hbr.org/2017/04/how-companies-are-already-using-ai

[4] Ibid

4 Solutions

4. Solutions

We aim to create a decentralized machine learning protocol and ecosystem, where customers, such corporate customers, research institutions, government and non-government organizations or even individuals, who wish to run analytical predictions can acquire appropriate algorithms from crowdsourced developers through the DML marketplace. With the aid of DML protocol, the machine learning algorithms can be run on the untapped private data and leverage the idle processing power of individual devices resulting in more precise predictions. Furthermore, the developers can improve their algorithm and its predictability by the crowdsourced model trainers in the DML protocol.

4.1 / Apply Decentralized Algorithm & Data for ML Prediction

4.1.1 / Deploy Machine Learning Algorithm for Practical Prediction

4.1.1.1 Customers Acquire Algorithms from Marketplace[5]

Customers, who wish to conduct analytical predictions through machine learning, can either choose to develop their in-house machine learning algorithms or acquire readily available ones, which will be built by freelance developers and listed in the DML marketplace.

Algorithms are categorized according to different types of businesses or themes within the marketplace where customers can acquire the appropriate ones that fit their purpose best. Afterwards, the customers need to specify their needs and criteria such as target segments in terms of geographic and demographic information. By narrowing the scope, this enables the machine learning algorithms to be run on appropriate dataset resulting a more accurate prediction that matches the customers' requirement.

[5] Refer to Diagram 1 Part 1a

4.1.1.2 Use of Smart Contracts to ensure Governance among Users[6]

The criteria input, scope of services provided by each participant and rewards in terms of an utility token i.e. DML token will be written in the blockchain-based smart contracts. The smart contracts will be digitally signed to eliminate default risk and connected among the customers, developers, and the decentralized nodes that are connected to DML protocol.

4.1.1.3 Decentralized Architecture via Decentralized Node Deployment

Instead of connecting the customers and developers directly to all potential data owners' devices, DML protocol adopts a decentralized approach to connect the developers to various decentralized nodes with different functionalities. Subsequently, each decentralized node will connect to multiple individual data owners' devices. Therefore, a massive number of potential data owners can be connected at a faster speed with more stable connection by utlizing the processing power and network bandwidth of the decentralized nodes.

4.1.2 / Distribute Algorithm to Eligible Devices

4.1.2.1 Distributing Node to Identify Devices for Algorithm Distributions[7]

Based on the machine learning requirements stated in the smart contracts, distributing nodes will identify appropriate data owners, who have granted access to their data on their electronic devices. The data will match the criteria and defined scope of running the selected machine learning algorithms. These distributing nodes will distribute the algorithms, which will be homomorphically encrypted, to appropriate data owners' devices as identified.

For instance, if the machine learning algorithm is required to be run on photos stored in the photo albums, the distributing nodes will identify the devices with data owners' consents obtained to allow algorithms to be run on their photo album before distributing the encrypted algorithm to these appropriate devices.

[6] Refer to Diagram 1 Part 2

[7] Refer to Diagram 1 Part 3

4.1.2.2 Auction Mechanism to Prioritize Machine Learning Sequence

The amount of token required to engage the protocol is purely driven by the market. It may depend on the scope and how busy the protocol is. DML protocol will analyze the current protocol capacity and suggest to customers the amount of DML tokens required for processing at that particular moment. The algorithm distribution and deployment sequence will be prioritized according to the amount of incentives as paid in DML tokens.

4.1.3 / Run Algorithms on Eligible Data Owners' Devices[8]

With the aid of homomorphic encryption, the encrypted algorithms received from the distributing nodes can be directly run on the data within the devices. Therefore, the intellectual property right of the algorithms is being well-protected through encryption.

4.1.3.1 Run Algorithms directly on Devices without Data Extraction

Unlike traditional data analytics, which private data are transferred from the data owners to the developers or a centralized hub for processing, DML protocol will facilitate machine learning algorithms to be run directly on data owners' devices such as smartphones without the need to extract any personal data or to store the information elsewhere for processing. As a result, individual private data will be well protected without any raw data leakage from individual devices.

[8] Refer to Diagram 1 Part 4

4.1.3.2 Deploy Algorithms on Individuals' Authorized Dataset

With the DML Application, data owners can authorize specific types of datasets, such as photos in the album or text messages etc., for machine learning algorithms to be run. Only with the consent of the data owners, the algorithms can be run on certain authorized types of data in the devices. Since the algorithms are run directly in the device, the data will be kept within the devices without transferring to any third parties or being stored in the cloud server.

Furthermore, non-public data located in other applications such as social media and messenger can be used to run machine learning algorithm too with the aid of API and users' authorization. Therefore, DML protocol is not only a gateway to connect private data stored inside the device for machine learning, but also a mean to access massive amount of non-public data stored within the existing networks.

4.1.4 Share Encrypted Prediction Results without Data Extraction

After the machine learning algorithms are run directly on the authorized dataset within the devices, only the analytical conclusion in the form of local prediction results will be encrypted and transmitted to a federated node via a secured network. All individual raw data will be retained in the data owners' devices without sharing to the nodes. Therefore, concerns over data privacy are resolved as data owners can stay anonymous and their private data will not be obtained by any third parties including developers or customers.

4.1.5 / Federated Nodes to Aggregate Prediction Results[9]

4.1.5.1 Average Prediction Results by Federated Learning

Each individual device will transfer its local encrypted prediction results to the connected federated node. The federated node will aggregate the prediction result with results generated from other connected devices to formulate a comprehensive prediction via averaging by Federated Learning.

[9] Refer to Diagram 1 Part 5

4.1.5.2 Validate Prediction Result Quality against Requirement

The federated nodes do not only act as a hub to collect and average the prediction results, but they also help to validate the quality of the results against the original requirements. If the results do not match the defined scope and criteria as set earlier, the node will instruct the connected devices to re-run the algorithms.

4.1.6 Report Node to Produce Final Report & Store in IPFS[10]

After various federated nodes aggregate and average the local encrypted prediction results of respective connected individual devices, they will send the encrypted aggregated results to the Report Node. The report node will further average the encrypted results as processed by individual federated nodes, of which an encrypted final report will then be generated and stored in a distributed file system such as IPFS.

If the total number of devices that contribute to the final results are less than a certain amount, the result will not be shared to avoid inaccurate predictions due to small sample sizes and data privacy protection by eliminating the possibility of reverse-engineering.

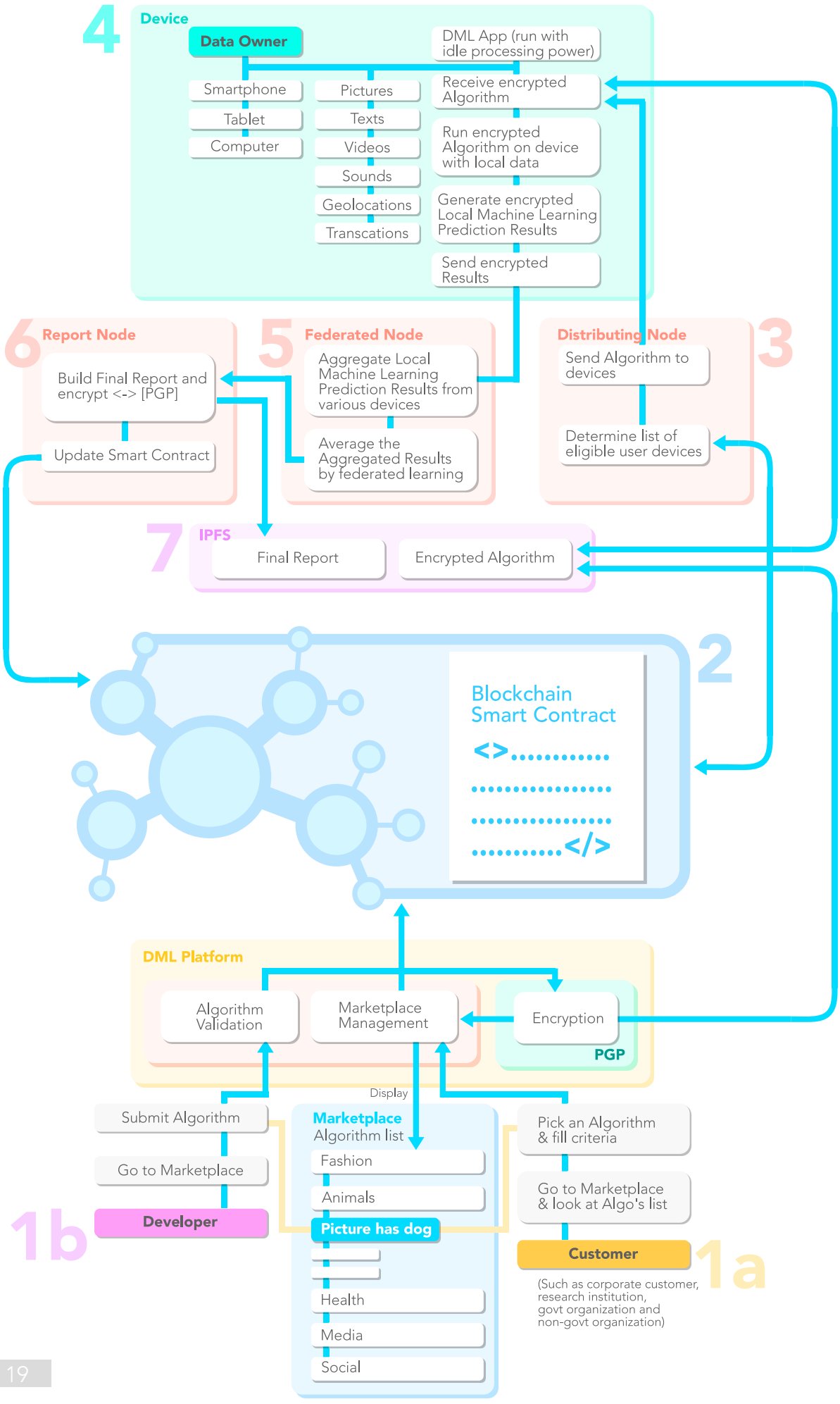
4.1.7 Prediction Report Retrieval and Smart Contract Execution

Once the report node shares and stores the final report in IPFS, the customers can retrieve by decrypting the report. At the same time, the report node will automatically update the smart contract once it stores the final report in IPFS.

The smart contract will be automatically executed by rewarding DML tokens to different participants in performing their roles according to the criteria and terms as established in the smart contract. The contributing participants include developers, data owners and decentralized nodes such as distributing nodes, federated nodes and report node.

[10] Refer to Diagram 1 Part 6 & 7

Diagram 1: Comprehensive flow chart of DML Protocol and Marketplace



4.2 / Improve Existing ML Algorithm via Crowd Informed Fine-tuning

To improve the machine learning algorithms, it usually requires intensive manual resources to fine-tune the algorithms and provide accurate responses to train the algorithms. Apart from unleashing the untapped private data for machine learning and building a community of talents to develop algorithms, DML protocol also facilitates to identify idle resources to fine-tune and improve the machine learning algorithms with the aid of a vast community of participating model trainers.

Through crowd knowledge and dataset, the algorithms can be fine-tuned, trained and improved by correcting the labels of the dataset in the algorithms.

4.2.1 Distribute to Model Trainers for Algorithms Improvement

When customers want to improve the selected algorithms before running them on appropriate data, the distributing nodes will identify participating model trainers, who previously indicated their willingness to support fine-tuning and correcting machine learning models, and send the models to these model trainers for improvement before running the algorithms on those appropriate data. Besides, DML protocol also facilitates developers, who simply want to enhance their algorithms, to distribute the algorithms to the participating trainers for fine-tuning without the need to compulsory list their algorithms for customers to use.

As device owners, they can opt to allow algorithms to be run on their authorized data to generate predictions, volunteer to improve the algorithms through fine-tuning and training, or both in return of DML tokens.

4.2.2 Adaptive Test Approach: Common data as Control Group

Once the algorithms are distributed to the devices of participating model trainers for fine-tuning and correction, an adaptive dataset with a mixture of common and private data will be used to test the participating trainers' response through questioning and testing in randomly generated sequence.

The results derived from common dataset will act as a control results, which will be used to cross examine with other individuals' responses and evaluate the accuracy of participating trainer's response. Weighting will be applied to individual training responses based on result accuracy generated from the questions based on common dataset.

4.2.3 Federated Node to Aggregate Weighted Fine-tuning Update

High-quality updates and correction to the algorithm as derived from the participating model trainer result will be encrypted and transmitted to the federated node, which will further aggregate and average the updates from other participating trainers by applying similar approach as Federating Learning.

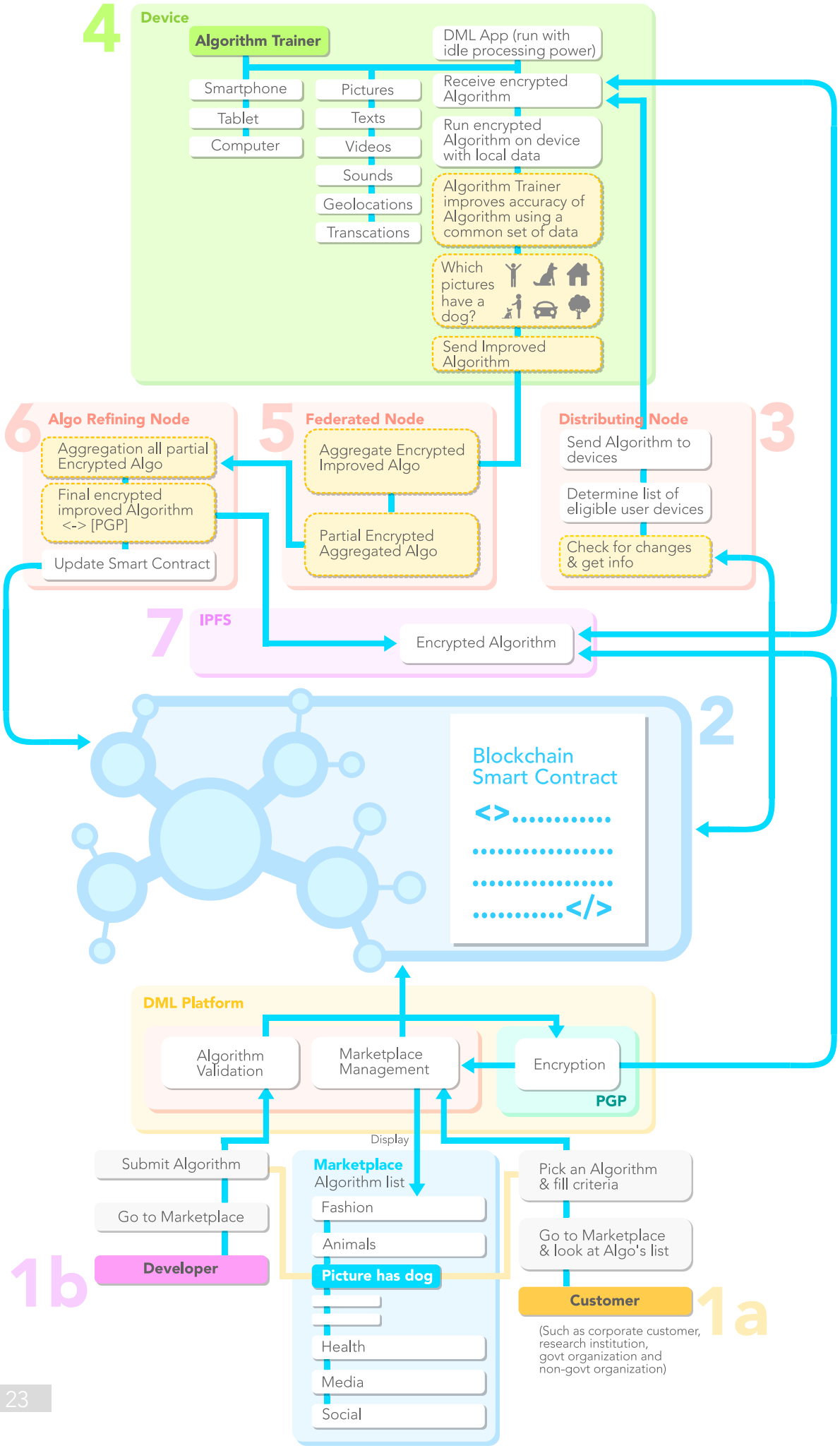
To ensure the fine-tuning updates are accurate, which will enhance rather than worsen the machine learning models, cross examination across fine-tuning results generated from the common dataset of different devices will be conducted. Results from poorly performed model trainers that do not meet a minimum standard will be excluded statistically. While weighting based on the result accuracy generated from common dataset will be applied to the overall updates from each participating trainer before fine-tuning the algorithms.

4.2.4 Algorithm Refining Node to form Fine-Tuned Algorithm

After the federated node has averaged the fine-tuned updates and has aggregated the encrypted updated algorithm based on the weighted updates received from their respective connected model trainers, the updated algorithm will be sent to the algorithm refining node. Upon receiving the updated algorithms from all federated nodes, the algorithm refining node will average and aggregate the updates to generate a final improved algorithm.

The final algorithm will be encrypted and sent to the distributing nodes for full scale distribution and run on the devices of appropriate data owners or to the developers if no machine learning job is deployed but only refinement is required.

Diagram 2: Algorithm refining technical flow



5 Blockchain-based Smart Contract

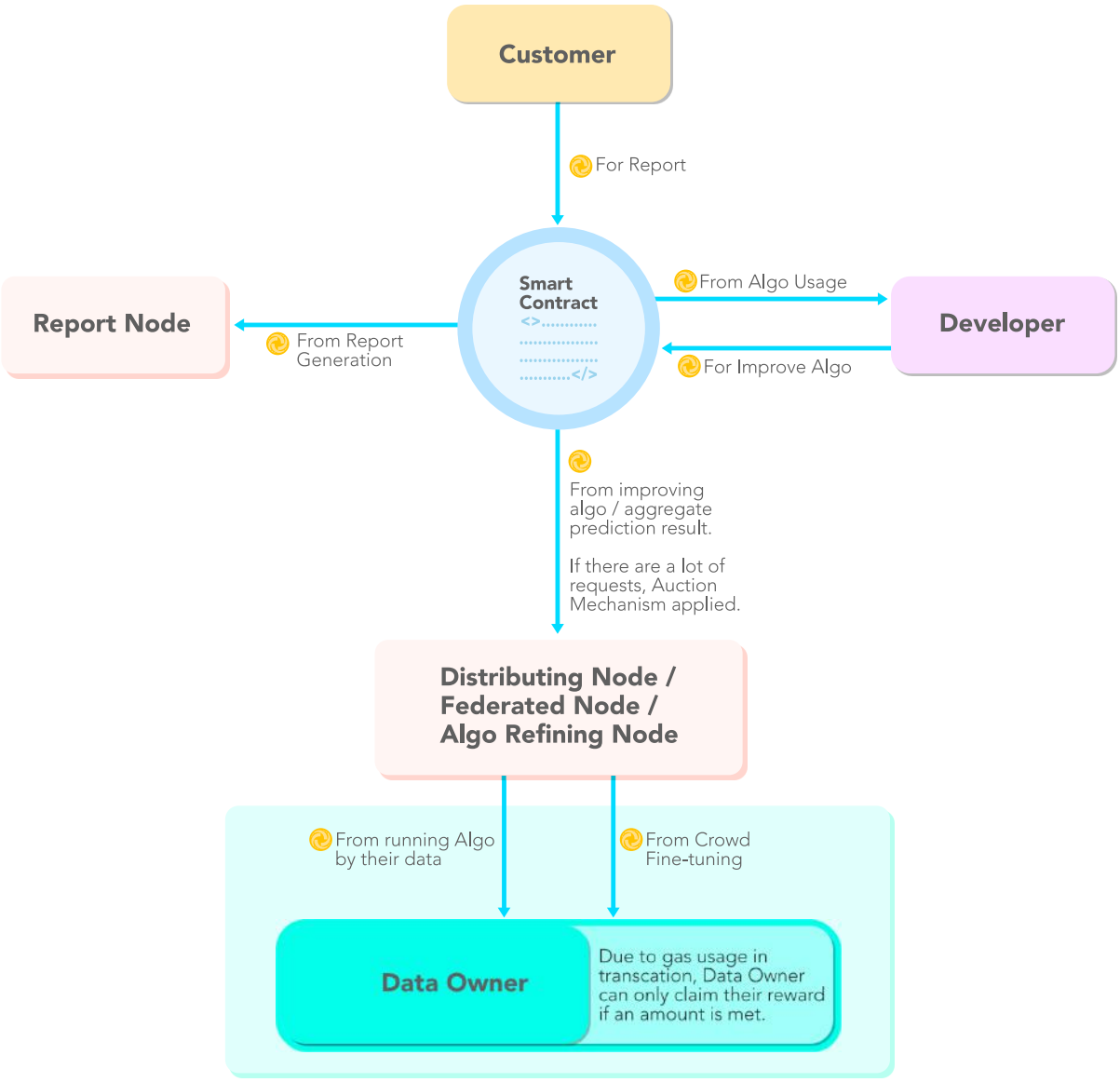
5. Blockchain-based Smart Contract

Smart contracts are utilized as trustless tools free of middlemen to facilitate the DML ecosystem.

1. customers, who request for machine learning service, to pay DML tokens to run machine learning algorithm on data owners' authorized dataset in their individual devices. The decentralized nodes will facilitate by distributing algorithms and collecting, aggregating and averaging the local prediction results to produce the final prediction report;

2. machine learning developers, who sell their algorithm on the DML MarketPlace, to get paid by customers in DML tokens; and

3. model trainers, who correct and improve the machine learning model, to be rewarded, by the customers or machine learning developers, who solely want to improve their algorithm for further use;



5.1 / Smart Contract for Machine Learning on Untapped Data

The smart contract will be run on the decentralized nodes which connect data owners. Upon the completion of the job i.e. the final report is produced by the Report Nodes, the Report Nodes will update the smart contract. The smart contract will then be executed, and the participants will be rewarded according to the terms as stated in the smart contract.

5.2 / Smart Contracts for DML Marketplace Activities

Apart from the data processing, the smart contract will also be used to govern the two market activities within the DML Marketplace regarding the acquisition of machine learning algorithms for various applications, and the algorithm fine-tuning and correction services provided by model trainers.

5.2.1 Acquire ML Algorithms for Self or Business Predictions

The parameters of this smart contract that will be used to govern DML Marketplace activities will be linked to the DML Marketplace setting. The contract will be executed according to the conditions as defined under the setting. The condition includes the amount of DML tokens that customers are willing to pay for the acquisition of machine learning algorithms for further deployment in the protocol. Upon the completion of the final machine learning result report by the Report Node, the smart contract will be executed and automatically transfer the agreed amount of DML tokens to the developers.

5.2.2 Crowd Fine-tuning to Improve Existing ML Algorithms

A similar mechanism applies to the algorithm fine-tuning services to be conducted by model trainers. They will be rewarded automatically by the smart contract upon the completion of the final machine learning result report by the Report Node or the algorithm is improved by Algorithm Refining Nodes.

6 Benefits for Different Participants

6. Benefits for Different Participants

6.1 / Developers' Perspective

6.1.1 Unleash Creativity via Innovation from Peripheral

The current ecosystem of machine learning and artificial intelligence development is mainly dominated by large corporations. They may either employ in-house developers or seek external professional consultants to develop the machine learning algorithms. Developers working at large corporations may be subject to hierarchical influence. They can only follow the bureaucratic corporation guidelines and hence may hinder their creativity to build algorithms with better use.

DML Algorithm Marketplace creates a developer's community where developers have the capacity to work as a freelancer to build and market their machine learning algorithms in return for incentives. Innovation and creativity to build machine learning algorithms are unleashed as no approval by tech giants or any other parties with vested interest is needed. Besides, better algorithms may also be developed and evolved due to open competition in a free market ecosystem.

6.1.2 Open Market to Sell Machine Learning Algorithms

Currently, it may not be easy for many developers to sell or market their algorithms to buyers such as commercial clients. Even the developers can find a mean to market their algorithms, they may need to pay high agency costs to sell their algorithms.

DML algorithm Marketplace provides developers an alternative mean to market their algorithms to potential algorithm buyers in an open market that are free from middleman control and costs. The transaction will be executed in smart contract with clear written terms and free from default risk.

The algorithm usage will be subject to different type of license conditions. We expect the most common type of licenses is on subscription basis for single use which requires minimal cost. On the other hand, the algorithm will be encrypted to prevent unauthorized access to the IP of the developers.

6.1.3 External Source to Fine-tune Machine Learning Algorithms

One of the major challenges of building a useful machine learning model is the lack of sufficient training dataset and the manual effort to correct the labelling of the machine learning algorithm in order to increase the accuracy and the intelligence of the algorithm. This is because supervised learning algorithms often require analyzing labelled training data to produce an inferred function that can generate predictions and generalize classification based on other unseen information. However, the processes to test the algorithms and training the dataset for supervised learning can be time consuming and can require a lot of manual effort to conduct.

Through DML Marketplace, developers have additional mean to connect with numerous individual devices owners whom can support the manual exercises to fine-tune the algorithms and trained the dataset without the need to hire in-house limited resources. This is another breakthrough and improvement in machine learning by connecting and engaging model trainers in the crowd.

6.2 / Data Owners' Perspective

6.2.1 Unlock Economic Value of Private Data with Privacy Protection

Currently, some data owners are already unknowingly sharing their individual data in their smart devices. They shared the data through various means such as social medial applications installed on on their devices, without any economic incentive and much attention to data privacy protection. Many data owners are reluctant to share their private data due to data privacy concerns, lack of trust and lack of incentives even if they are willing to contribute to machine learning development.

Through DML Protocol, the economic value of the private data is unlocked. Data owners are incentivized to allow machine learning algorithms to run on their private data in return of DML tokens without leakage of their private data to any third parties. Data privacy is protected as only the prediction results will be collected and aggregated through the decentralized nodes while the original source of private data will be kept within their individual devices. Therefore, data owners can participate in the machine learning development with economic incentives while data privacy concerns is addressed.

6.2.2 Make Use of Idle Processing Power

The processing power of our devices such as mobile phone, tablets and computers are growing rapidly. In the past decades, the processing power increases approximately according to Moore's law. However, we recently see exponential improvement in processing power in smart devices with dedicated chips for on-device artificial intelligence and machine learning. While we usually do not actively use our devices 24/7, the underutilized processing power can now be used to run machine learning algorithms and be rewarded along with the private dataset usage.

6.2.3 Certain Data and Device Owner also Acting as Model Trainer

While DML protocol design aims to run and make prediction in a hassle-free and automatic way to encourage mass participation, some data and device owners may be interested to walk an extra mile by acting as model trainers through correcting the labels of the machine algorithm in their spare time., The effort contributes to the machine learning development and earn extra DML token.

6.3 / Customers' Perspective

6.3.1 Reducing the Bar for Applying Machine Learning Algorithm

Currently, it is costly to hire an in-house developers or seek external professional consultancy services to build machine learning algorithms. Therefore, only large corporations with sufficient budgets may be able to afford the research and development costs in building machine learning algorithms. It is not surprising that large corporation dominated the machine learning field.

DML Algorithm Marketplace creates an open market that is available for anyone to acquire machine learning algorithms at a market-driven cost that are free from middleman control and costs. As a result, even small and medium enterprises, research institutions or even individuals may also be able to afford accessing talented resources from all over the world and acquiring the algorithms for business predictions or other usages at a market-driven cost. Furthermore, customers can request tailor-made machine learning algorithms for specific predictions without going through a middleman if the algorithm required is not in the current list or does not match the exact required criteria. As such, the marketplace could always meet the latest demand in the market.

6.3.2 Better Prediction Results with the Use of Untapped Data

The DML Protocol and Algorithm Marketplace can help the customers to acquire machine learning algorithms for their business predictions. They can also generate more accurate predictions by applying the algorithms to untapped data stored in individual electronic devices, which is inaccessible before, at a market-driven cost and free from middleman by adopting smart contract.

Comparing to traditional market research such as conducting survey and focus group, prediction from the DML protocol could be much faster and timely, unbiased, cost-effective, more accurate and relevant. DML protocol provides a powerful tool to identify market trends and sentiments are learnt.

With more accurate prediction on trends and sentiments etc, customers could deploy and plan their resources more wisely by producing and providing the goods and services that better fit the consumers' preferences and needs. Other customers such as political parties in election campaign, after learning the latest trend of the election, could quickly adapt and change their strategy in marketing and promoting themselves geographically and/ or demographically.

7 DML Token Usage and Mechanism

7. DML Token Usage and Mechanism

7.1 / Major Uses of DML Token

7.1.1 Utilize Untapped Data and idle processing power for Machine Learning

DML token is paid to the data owner to contribute their private data and idle processing power for running machine learning algorithm in the DML protocol.

7.1.2 Acquire Machine Learning Algorithms in Marketplace

Apart from building in-house machine learning algorithm, customers could acquire machine learning algorithm from the developers in our DML marketplace using DML tokens. Therefore, talented developers are incentivized to join the DML developers' community to develop machine learning algorithms for real life applications.

7.1.3 Reward to Decentralized Nodes in the protocol

DML tokens are used to motivate the decentralized nodes with different functions such as distributing algorithms, aggregating and averaging local machine learning result by federated learning or producing aggregated machine learning report. Those functions are essential and serve as part of the decentralized processing power network in machine learning.

7.1.4 Reward to Algorithm Trainer for Fine-tuning Algorithms

Developers who wish to seek resources to train and fine-tune their models can pay DML tokens to the model trainers. DML tokens are used to motivate model trainers to contribute their effort and time to fine-tune the algorithms by correcting the label of the algorithm.

7.2/ User Growth Pool to Incentivize Users

A user growth pool of 9% of DML tokens will be created to incentivize DML users to early participate in the DML protocol and ecosystem.

1. Data owners to allow machine learning algorithms to be run on their untapped private data;

2. Decentralized nodes which provide various essential functions in supporting different parts of decentralized machine learning;

7.2.1 Extra Reward to Contributors

In addition to DML tokens directly paid by customers, data owners may earn extra DML tokens randomly from the user growth pool as long as they are connected to the DML protocol. The data owners, therefore, will earn DML tokens from up to two sources i.e. customers and user growth pool.

7.2.2 User Growth Pool Generation Mechanism

A limited amount of the DML token will be generated according to the pre-defined decreasing-supply algorithm based on the preset mechanism across time. The mechanism is designed to encourage early participation to contribute to machine learning development. Therefore, the reward block is designed with a pre-defined algorithm that will decrease with time.

7.3/ The DML Token Utility System

Data owners will be rewarded from up to two sources, which is driven by the market demand of using untapped private data for machine learning and from the user growth pool.

To encourage simplicity and mass adoption of token and cryptocurrency, a ERC20-supported wallet will be built in DML App to keep the DML token or other supported tokens.

8 Governance

8. Governance

8.1 / Open Source Technology for System Audit & Code Debug

DML Protocol adopts a "collective governance" structure that allows and motivates any parties to participate in system and protocol monitoring, auditing and debugging. It is an open source protocol, of which the source codes of the protocol are open to public. Therefore, participants of the DML community can access, read, debug and audit the source codes.

8.2 / Self-Governance through System Audit by DML Participants

With the source codes open to DML participants, it creates transparency over the DML protocol as any participants can audit the source codes to check against our philosophy and mechanism such as only local prediction results are aggregated without leakage of original private data. Rather than being governed by a small group of people at periodic interval, DML protocol will be monitored and audited continuously 24/7 by talents from all over the community, which creates confidence and provides transparency to users.

8.3 / Encourage Public Contribution to System Debug

Furthermore, open source codes allow any participants to access, amend and debug the existing codes. To encourage talents to continuously debug the codes, they are incentivized with DML tokens. A pool of DML tokens are set aside to distribute to these participants who successfully debug the code. With the debug incentive mechanism, security of the DML protocol will be strengthened as any bugs or security leakage will be fixed by talents all over the global network.

9 Future Upgrade & Interoperability

9. Future Upgrade & Interoperability

9.1 / Flexibility to Upgrade and Multi-Blockchain Compatibility

We treasure the knowledge, expertise and innovation of every individual as we believe in collective wisdom and intelligence, which is our core value that motivates us to unleash untapped data and link them together to create a co-intelligence system. Therefore, DML protocol is designed to encourage developers to participate in system upgrade or future technology compatibility advancement too.

Compatibility with existing and forthcoming blockchain and cryptographic technology together with the flexibility for protocol upgrades are the key for on-going development and sustainability. Therefore, DML protocol is designed with flexibility to be upgraded and compatible with new blockchain technology, of which it will not be confined to any single or existing ones.

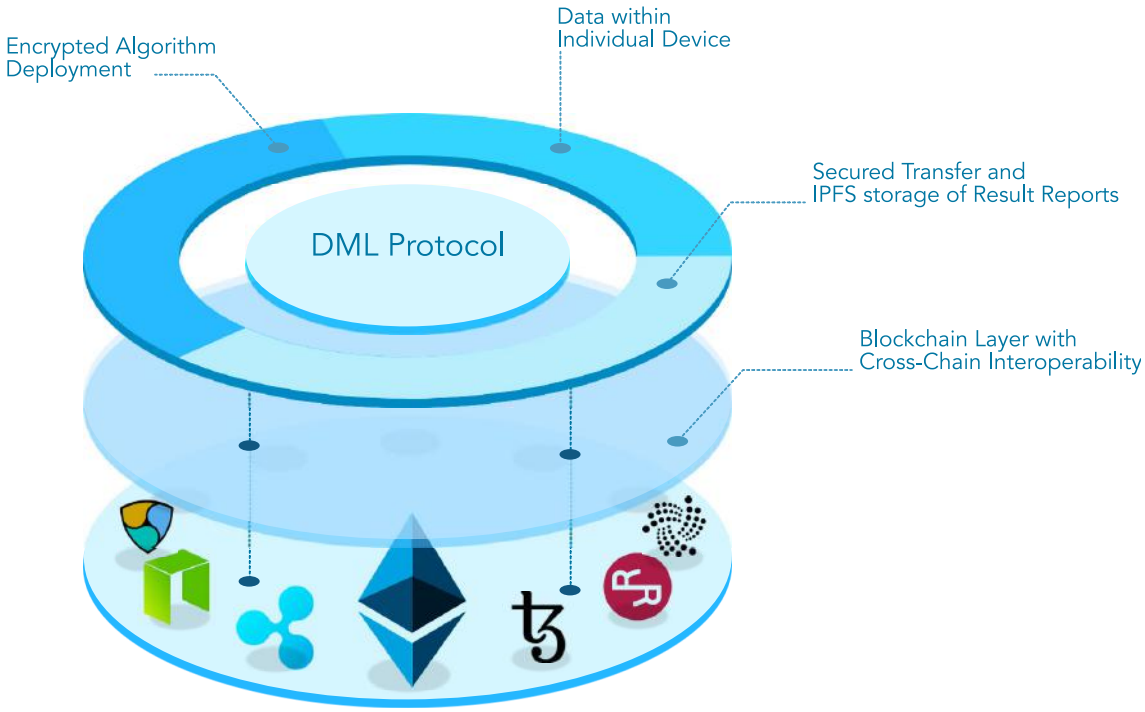
9.1.1 Promote DML Protocol Upgrade

We encourage any developers, who have ideas to improve the protocol by proposing enhancement to the existing one, through our protocol. We will reserve financial resources to motivate any developers to improve our protocol. This does not only incentivize developers across the globe to continuously contribute to the maintenance of the decentralized network, but also the forthcoming enhancement of the protocol.

9.1.2 Scalability and Off-chain Transactions

DML protocol is designed with scalability feature to cope with increasingly mass usage and heavy transactions load. For instance, we may adopt off-chain method, such as state channels, by offloading non-core and generalized processes off-chain to increase its efficiency and reduce transaction costs.

9.1.3 Cross-chain Compatibility and Interoperability



Our decentralized and open-source blockchain technology allows individual data and intelligence to be joined together to create a collective beneficial ecosystem. We believe such decentralization approach should not be confined to one single blockchain platform. Therefore, our technology welcomes and encourages all sorts of blockchain technology to be adopted in our protocol and various system languages to be used to run their machine learning algorithms on our network.

We encourage both adoption of existing blockchain technology such as Ethereum and any other future blockchain platform. Our protocol aims to allow integration of new blockchain technology to the currently adopted ones without replacing or overriding them. A co-exist ecosystem that links and aggregates different blockchain-based technologies will be created.

Likewise, for the protocol upgrade mechanism, developers are also encouraged to integrate any new blockchain technology on top of the existing blockchain platform as used in DML protocol. The same incentive mechanism will be deployed to encourage a variety of blockchain technologies to be adopted.

10 Protocol Evolvements

10. Protocol Evolvements

10.1 / Initial Phase:

10.1.1 Utilize Untapped Data & Create DML Marketplace

The first and most important features for DML protocol is to explore and utilize the blue ocean of untapped and privately-owned data for machine learning. In our initial development stage, we connect electronic devices from all over the globe to reach the untapped private data through DML protocol.

We focus running machine learning algorithms on the untapped data to generate more accurate predictions through DML protocol. We protect data privacy of data owners, which is one of our top priorities and niches, by sending local prediction results for further aggregating and averaging with no raw data being extracted from the electronic devices. All the prediction results will be further processed and consolidated on collective basis, which will yield a final prediction result.

The DML Marketplace is also created with the view to collective wisdom or "co-intelligent". Therefore, the data, algorithms development and fine-tuning of algorithms are decentralized through our DML platform rather than being centralized or controlled by oligopolies.

10.1.2 Launch of Regular DML Algorithms Competition

Subsequent to the release of DML Marketplace, a DML Algorithm Competition will be introduced, which creates a distributed platform for crowdsourcing machine learning solutions. It aims to attract and motivate more diversify talents to create machine learning algorithms with practical utility to solve real life problems. A level-playing field is created for any developers disregard of their background to participate in the competition., They can employ their talents, polish their skills and be rewarded solely based on their capability without any hierarchical interference.

Real life challenging use cases will be provided for participants to solve by creating tailor-made algorithms. Winners will not only be rewarded by DML tokens, but also gain reputation in winning the award and can market their solutions in DML Marketplace.

10.1.3 Connect Mobile Sensors and IOT devices

While we initially focus on exploring data, which has already be generated and stored in the device, machine learning development can be further advanced through the real time and relevant data generated from the multiple sensors within the smart devices and other IOT devices as controlled by the same devices.

The distributed IOT and mobile sensor network will bring machine learning development to another level by providing data that does not exist before or very difficult to collect for on-device machine learning. The use of sensors or IOT connection will require data owner authorization and those data will not be extracted from the smart device to protect data privacy.

10.2 / Future Phase 1: Extend Applications to other Programs

Machine learning applications is only one of the major use cases for DML protocol. With a network that connects to potentially billions of electronic devices to unleash the use of previously untapped private data and their idle processing power, it can extend its applications from machine learning algorithms to any other types of programs to be run on untapped data by using the idle processing power of these devices to generate any suitable programs and applications.

We aim to truly unlock the untapped data within the electronic devices especially smartphones and fully utilize their processing power for the advancement of artificial intelligence development. Not only does the machine learning algorithms that generate predictions can be run on these untapped data, but other kinds of applications and programs can also leverage the same DML platform in applying the untapped data and idle processing power of those connected devices. Similarly, these programs can also be marketed to potential users through the same DML marketplace.

10.3 /
Future Phase 2: Global Self-Initiate Decentralized ML Network

We do not only believe in "more good quality data, better result", but also "the whole is greater than sum of its parts". Therefore, applying existing distributed data in a decentralized network for machine learning is the initial step to build the DML Protocol ecosystem.

Smartphone is an extension of our body, which performs as extra pair of eyes, pair of ears as well as other sensors. It is also a gateway for us to reach the outside world digitally through the mobile network and internet.

The potential of DML protocol is far more than facilitating algorithms or programs to be run directly on these connected electronic devices through its decentralized network. Ultimately, we aim to leverage the decentralized network that connects potentially billions of devices all over the globe to be able to communicate with each other.

Once a self-communicating network is built across all electronic devices, decentralized data collected can be cross-checked and complementing with each other. Therefore, the data set in the DML protocol are more reliable, complete and with highest integrity for algorithms and programs to be run to generate more accurate predictions or produce better software applications.

Ultimately, the self-communicated network can self-detect and self-initiate to acquire useful information scattered across the globe to formulate valuable predictions with global impact that may be beyond human imaginations. As such, it will be the suitable time to research and develop Strong-AI as powered by the matured version of DML protocol.

11 DML Team

11. DML Team

11.1 /
Core Team Members Bios

Victor Cheung
Blockchain Developer

Victor obtained his BEng in Software Engineering and MSc in Computer Science in the University of Hong Kong. Since 2014, he founded 2 technology companies with startup winning awards. He is a full-stack web, mobile application and smart contract developer.

Prior to that, he spent over 4 years in a private company serving US technology corporations in software development. He is proficient in Solidity, C#, JavaScript, PHP, Node.js, MySQL, MFC etc.

Michael Kwok
Project Lead Director

Michael founded the 2 technology companies and winning startup competition awards together with Victor. He is a seasoned growth lead in early stage startups, especially adept at business development, digital marketing, search engine optimization (SEO), social media and community management.

His ability and the sense of motivation makes him the most suitable person to coordinate and drive the business development of the project.

Jacky Chan
Blockchain and Software Developer

Jacky is a full-stack software engineer. At Symphony Communication which was acquired by Goldman Sachs, he built secure chat applications for Wall Street firms and then a software engineer at Uber in San Francisco Bay Area.

Afterwards, Jacky co-founded Kyokan Labs and has been focusing on improving the blockchain space. He has substantial involvement in Metamask's new UI development, as well as the network visualization dashboard design with DFINITY. Jacky is proficient in JavaScript, Python, Go, Node.js, AngularJS, Backbone.js, MongoDB etc.

Pascal Lejolif
Machine Learning Engineer

Pascal started his career as a software engineer at OSIS and then over 8 years as the technical project manager in Alcatel-Lucent Enterprise in Paris, France. He then moved to Asia and was working as consultant, managing IT projects in Bangkok and Hong Kong for the past 7 years.

He is the former Chief Technology Officer of Alkia IT Services, a company specialize in cloud computing, AI and cyber security services in Thailand. Pascal is a veteran IT projects consultant and system designer. He also obtained his Machine Learning Certification at Stanford University.

Wilson Lau
Machine Learning Engineer

Wilson is the technical manager of a cryptocurrency mining consulting firm. He also involved in machine learning agriculture robotic system design, including a hydroponic vertical farm that utilize machine learning.

He obtained his Bachelor degree in Physics from Hong Kong University of Science and Technology and is proficient on machine learning programming languages such as Python, C, Mathematica, Arduino etc.

Patrick Sum
System Security Engineer

Patrick accumulated over 20 years of information technology experiences by serving in the banking and finance industry, education sector, as well as semi-government authority. His various specialties about software development, system administration and security, IT project management, and wide-area network design is the backbone of the Team.

Patrick completed his Bachelor of Science (Computer Studies) at the University of Hong Kong in 1988 and acquired the Master of Business Administration from Australia Graduate School of Management (AGSM) in 2000.

11.2 /
Advisors Bios

Guillaume Huet
Big Data / Machine Learning Advisor

Guillaume has 12 years of experience in banking, consulting and business development in Europe, Africa and Asia. Besides his MIB obtained from EDHEC Business School, Guillaume has developed his expertise on AI from online certification from Stanford University, John Hopkins, Deeplearning.ai and through successful implementations.

Guillaume is a frequent speaker on data science and machine learning. His passion leads him to be a noticed actor in the machine learning scene in Hong Kong.

Michael Edesess, PhD
Machine Learning Advisor

Michael is currently an Adjunct Associate Professor in Hong Kong Science and Technology University, for which he teaches postgraduate course including cryptocurrency.

Michael often act as a discerning critic to participate in the finance and blockchain world. He is a columnist in MarketWatch, an affiliate of The Wall Street Journal, as well as South China Morning Post, Hong Kong, focusing on global economy and cryptocurrency.

He received his bachelor's degree from Massachusetts Institute of Technology (MIT), majoring in Philosophy and Mathematics, then went on to receive a PhD in Pure Mathematics at Northwestern University, specializing in the fields of stochastic processes and partial differential equations. His expertise in Mathematics shall provide the Team valuable insights about the federated learning process design.

Roderik van der Graaf
Blockchain Advisor

Roderik is the Managing Director at Caldera Pacific, a growth capital investment firm based in Hong Kong. He has been in the VC/PE space since 2011 and involved in crypto assets markets since 2014 as an investor and advisor. Previously Roderik had a career in financial markets, where he was an equity derivatives trader at Bear Stearns, HSBC, Rabobank and Deutsche Bank along with hedge funds and market making firms.

Roderik holds a BEng in Aerospace Engineering at University of Hertfordshire and a MSc in Information Technology at Queen Mary and Westfield College, University of London. He is also a Chartered Alternative Investment Analyst charter holder.

Kyle Wong, PhD
Machine Learning Advisor

Kyle received his MSc and PhD degrees in Applied Physics from Cornell University. In the last 20 years, he applied his quantitative skills to the financial markets in roles like bond trading, derivative structuring and risk management.

Since 2015, he has focused his interests on fintech and artificial intelligence. He is the Founder and Chief Operating Officer of Artificial Intelligence Hong Kong and being an advisor of several AI start-ups.

Besides his strong academic backgrounds, Kyle obtained his fintech certificate in Massachusetts Institute of Technology (MIT).

Scott Christensen
Machine Learning Advisor

Scott is the Founder and Chief Executive Officer of Hanpa Group, which utilize AI for high frequency trading. He is the Former Head of North American Equity Derivatives Trading in ABN AMRO Bank N.V., gaining solid trading experiences from various positions including trader Japanese index & options arbitrage, merger arbitrage, exotic options, Hong Kong proprietary trading, Hong Kong private equity investment and futures statistical arbitrage trading. Scott is graduated at University of Wisconsin with a BSEE.

Steven Cody Reynolds
Blockchain Advisor

Steve brings extensive leadership experience from the political, military and financial services spheres. His experience in business development, communications, public relations, crisis management, and crypto-currency exchanges will be invaluable to the Team.

During his tenure with Binance, Steve was responsible for public facing community management of the fastest growing cryptocurrency exchange ever. He managed over 7,000% growth in users. His steady hand helped to guide Binance through a period of explosive growth and uncertainty while managing their English speaking customer facing communications and media channels.

As the Founder of Greychain Capital, Steven specializes in digital asset management for High Net Worth Individuals. Steve is also an active member in various blockchain projects. Besides the Advisor of DML, he is the Advisor of HybridBlock and the Head of Operations of Akropolis.

Matthew Slipper
Machine Learning Advisor

Matthew joined Kyokan in 2017 as the Co-Founder and CTO. He is knowledgeable in Software Engineering and Engineering Management. Matthew has worked extensively on Blockchain scaling solutions. He currently leads a team working on implementing Plasma and also contributes to Machinomy, a micropayments library using state channel technology.

Matthew started his career at Perzo, which was later acquired by Goldman Sachs to became Symphony Communications. Afterwards, he joined Wealthfront, where he led the frontend team responsible for implementing new financial products on top of Wealthfront's trading platform.

He was also the Co-Founder and CTO of Spectrum Labs, which provides AI-based trust-and-safety tools to businesses with large volumes of user-generated content.

Jesmer Wong
Machine Learning Advisor

Jesmer has strong interest in Artificial Intelligence & Machine Learning with practical implementation in Data Science for Predictive Analytics and Data Visualization and Automation for the companies.

He is also keen in Blockchain applications especially in orchestration of Data Decentralization, Machine Learning, Communication Algorithms and Insight Extractions. He strongly believes DML is a pioneer to combine Blockchain with Machine Learning applications.

Jesmer holds an MBA and BEng (Electronic Engineering) from Chinese University of Hong Kong. He also obtained Deep Learning Nanodegree from Udacity and Certificate in Data Science from General Assembly.

Eugene Tay
PR & Marketing Advisor

Eugene is considered one of the most versatile and innovative business strategists for startups in Asia. He was Founders of various business initiatives and has positively influenced Singapore's education system during 2003–2018 such as initiating "The Alpha Mind Programme" which is supported by Temasek Foundation.

Eugene is also a significant influencer within the Asian Blockchain community and has actively engaged in developing strategic media content for reputed projects such as Wanchain, Quantsamp etc. In 2017, he founded CryptoCentral.net, a Blockchain advisory collective that brings businesses to the Blockchain and supports the crypto community.

Eugene graduated with honours from Nottingham Trent University with a Degree in Computer Engineering. He is also the author of *Supernatural Confessions* and *Beyond the Classroom*.

Eric Byron
Business Advisor

Eric has been working in the IT industry for over 25 years with diverse experience in various roles. He started working in the Entertainment industry and transferred to IT field during his tenure with Disney in the US. His unique experience in combining innovation with technology enables him to excel in the Video Game industry, of which he spent 6 years with Electronic Arts (EA Sports).

Eric holds a Bachelor Degree in Computer Science and a Master Degree in Computer Information Systems. He taught part-time Computer Science courses for several years in the US. Currently, he is the Strategic Advisor for Accelerate HK Coding Bootcamps Limited, which is a tech education startup in Hong Kong.

Fabrice Fischer
Business Advisor

Fabrice is the Founder and CEO of Blu, which is an Artificial Intelligence advisory firm that specializes in disruptive innovation within the Financial Services industry. Fabrice was working in the Financial Services and IT industries in Asia, Europe and North America in the past 20 years. During Fabrice's tenure as CFO of Sentient Technologies in the past 6 years, the company encountered rapid growth and has become one of the world's best funded AI platform.

Fabrice holds an MBA from INSEAD. He also received a Bachelor Degree in Electrical Engineering from University of Montreal. His AI expertise includes Machine Learning, Neural Networks, Deep Learning, Evolutionary Algorithm, Natural Language Processing, Image Recognition and Video Processing.

11.3 /
Working Partners

12 Disclosure

12. Disclosure

This whitepaper is being provided in support of the Decentralized Machine Learning Protocol. The technological, social and business structures utilizing blockchain and any other associated technology are continually developing and will evolve for the foreseeable future. Accordingly, the plans, strategies and implementation details described in this whitepaper will likely evolve as well and, accordingly, may never be adopted. We reserve all rights to develop or pursue additional or alternative plans, strategies or implementation details associated with the Decentralized Machine Learning Protocol.

DML Tokens are being sold in accordance to the Terms and Conditions of the token sale are available at https://decentralizedml.com. For complete details, please refer to the terms and conditions of the token sale. DML Tokens are utility tokens and are not securities, investments or currency, and are not sold or marketed as such. Besides, participation in the token sale involves significant technological and systematic risks; the sale is not open to individuals who reside in or are citizens of the United States, Singapore, China and Hong Kong. The sale period, duration, pricing, and other provisions may change as stated in the token sale terms. The DML token sale involves known and unknown risks, uncertainties, and other factors that may cause the actual functionality, utility, or levels of use of DML Tokens to be materially different from any projected future results, use, functionality or utility expressed or implied in the terms and conditions of token sale.

References

References

1. Daniel R. Rehak et al., A Model and Infrastructure for Federated Learning Content Repositories, WWW 2005 Chiba Japan, 2005

2. Frederik Armknecht et al., A Guide to Fully Homomorphic Encryption, University of Mannheim, 2015

3. Gavin Wood, Ethereum: A Secure Decentralized Generalised Transaction Ledger, EIP-150 Revision

4. H. Brenda McMahan et al., Communication - Efficient Learning of Deep Networks from Decentralized Data, Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS) 2017 Fort Lauderdale Florida USA JMLR: W&CP volume 54, v.3, 2017

5. Jakub Konecný et al., Federated Learning: Strategies for Improving Communication Efficiency, University of Edinburgh and Google, 29th Conference on Neural Information Processing Systems (NIPS 2016) Barcelona Spain, v.2, 2017.

6. Jakub Konecný et al, Federated Optimization: Distributed Machine Learning for On-Device Intelligence, Google and University of Edinburgh, 2016

7. Jeffrey Dean, Large Scale Distributed Deep Networks, NIPS 2012: Neural Information Processing Systems, 2012

8. John P. Lalor et al, Crowd-Informed Fine-Tuning to Improve Machine Learning Ability, v2, 2017

9. Keith Bonawitz et al, Practical Secure Aggregation for Privacy-Preserving Machine Learning, Google and Cornell University, 2017

10. Maha Tebaa, Secure Cloud Computing through Homomorphic Encryption, International Journal of Advancements in Computing Technology, Volume 5, Number 16, Dec 2013

11. Massimo Bartoletti and Livio Pompianu, An Empirical Analysis of Smart Contracts: Platforms, Applications and Design Patterns, Universit`a degli Studi di Cagliari, 2017

12. Ponomarev S. and Voronkov A. E., Multi-agent systems and decentralized artificial superintelligence, Moscow, 2017

13. Satoshi Nakamoto, Bitcoin: A Peer-to-Peer Electronic Cash System, 2008

14. Vamsi K. Potluru et al, ComertCloudCare: Distributed Machine Learning Platform-as-a-Service with Privacy Preservation, 2014