TuringNet

World's First Open & Trustable Artificial Intelligence Platform

v0.9.9 08/20/2018

Table of Contents

Executive Summary

As we proudly pronounce that we live in an “AI-first age”, there are still problems industry-wise in regards to opaque data refinery, siloed and fragmented data sets, and repetitive model training that’s wasting massive computational resources, just to name a few.

TuringNet is a decentralized public chain, aiming to create world’s first open and trusted Artificial Intelligence platform. The platform sets future advancement and applications of Artificial Intelligence free from any controls of centralized organizations, thus serving the best interests of the AI community that reshapes humanity.

We are reinventing how the public participates in scalable and collaborative model training and prediction, significantly reduces repetitive model training and fully utilizes idle computing resources, at the same time ensuring data security. This is achieved by TuringNet’s new infrastructure design and incentivizing LBFT consensus mechanism which is tailor-made for and fully compatible with various AI model training needs.

Our next generation solution will foster a frictionless system as TuringNet’s DModels are accessible, verifiable and trusted by the public, and will maximize benefits and values to all ecosystem partners and participants.

Roadmap

2018 Q1

- Project kickoff
- Pre-research and conceptual design, with early validation

2018 Q2

- Complete system architecture design
- Release technical whitepaper

2018 Q3

- Launch product demo
- Implement consensus mechanism for subchains

2018 Q4

- Launch main chain test network
- Collaborate with initial partners for preliminary AI models training/inferencing
- Build and run communities with early developers and partners

2019Q1

- Launch subchain test network
- Build global community and expand ecosystem partners

2019Q2

- Launch version 1 main network with both main chain and subchains
- Continue to promote for global adoption with world-wide partners

2020 and after

- Launch version 2 main network world-wide
- Expand training tasks for consumer devices

Introductory Background

“Every day millions of people decide to grant their smartphone a bit more control over their lives or try a new and more effective antidepressant drug. In pursuit of health, happiness and power, humans will gradually change first one of their features and then another, and another, until they will no longer be human.”

— Yuval Noah Harari, Homo Deus: A Brief History of Tomorrow

AI has progressed exponentially in the past ten years, in large part thanks to cloud computing and machines having access to massive amounts of data generated on Internet, for example, the success of AlphaGo[1]. AI is not just changing the equation for online companies like Google, Amazon and Netflix with machine learning: it is also contributing to how we shop for tangible, everyday products like food and beverages with the use of big data. In every aspect of our lives, AI is creating powerful transformative innovations that are changing the way companies market to consumers, and changing the way consumers live their everyday lives. Life is becoming, more personalized, more convenient, and much, much smarter. We are officially living in the era of artificial intelligence, and the question is no longer whether or not artificial intelligence will infiltrate our society, the question is how we will succeed or fail at making use of it.

Challenges in AI platform

The common belief is that there are three future potential threats from AI: the first threat is that AI superpasses human control and effectively begins to control or destroy human; the second potential threat is that a small percentage of powerful people will have sole control of the artificial intelligence technology, leading the rest of society to be controlled by machines that a small population of powerful people control. This could lead to a monoLBFTy of resources, a huge threat to individual privacy, or at worst, world domination by this small group.The third possible threat from AI is that we develop the technology in a way that allows society as a whole to remain in control of it by programming it to think more like humans. For example, we train AI to prioritize obedience over problem solving, and to prioritize empathy and compassion over power.

In other words, the question is becoming: who will control the AI, once we enable it to run and manage the world? The question is urgent, and currently remains unanswered.

There are a lot of variables that will contribute to determining how the AI revolution turns out: and much of it will be playing out in the next 20 years. What is certain, is that if one player gets too much control over the intelligence, then the rest of humanity has potential to suffer greatly. On the other hand, if we make efforts now to democratize artificial intelligence, then humanity could see immense benefits ranging from less daily traffic, to better healthcare, to eradicating poverty.

Furthermore, a centralized AI system contains the following problems compared to our proposed decentralized, democrat, and trusted AI system:

Models are controlled by centralized owners

Models in a centralized AI system often owned by big companies like Google, Amazon, etc. The models and the processes of training and inferencing are only visible internally. This problem is one of the root cause for the fear of AI from general public. People can only rely on the goodwill of such companies. However, if some malicious owners or some ignorant owners misuse the models in their products, the consequence may be critical.

Waste and imbalance of computation powers

Typical AI system requires lots of computation powers in model training and inference. Multiple centralized systems have built their own data centers but not shared to each other, which lead to enormous resource waste. Users on different centralized systems have imbalance computation powers for their AI models. Moreover, general public cannot participate in global AI model training and inferencing; instead, such computation powers are wasted. There is a great opportunity to organize computation powers from public in a decentralized but trustable way.

Public is not incentivized to train and use AI

As we state above, general public cannot participate in AI. One of the major reasons is that they are not incentivized on centralized AI systems. Owners of the centralized AI systems are not obligated to pay people who want to contribute their computation powers, as they all possess large data centers with much lower cost per unit power. The ecosystem then separates the public from these big companies, who further strengthens their owned centralized AI systems and models.

High cost for operations

Centralized AI systems are operated separately by their owners. Even though the cost per unit power is lower, the scales of the systems often lead to enormous cost to run.

Such high cost may easily urge the owners to do harmful business to users while generating much higher revenue and profit.

Data access and privacy

Training and optimizing AI models require large amount of data. The primary sources of such training data are from user characteristics and interactions, such as genome, faces, user behaviors on clicking ads, etc. Such private data are sensitive to use in training. The black box of a centralized AI system has no audition on how the data are used. On the contrary, the high operation cost for running centralized AI systems drives the owner companies to further exploit private data in order to train highly targeted models. On the other hand, data owners, including ordinary users and businesses owners, are hesitant to use centralized platforms with concerns of data exfiltration.

Challenges in Blockchain

With the technical breakthrough from the last decade, blockchain offered a mechanism to build distributed consensus, and in turn, brought a network of shared values. One of the most important applications of blockchain is a distributed ledger based on consensus. Using blockchain to store transaction information allowed issuance of digital / cryptocurrencies. The most prominent being Bitcoin is known as blockchain 1.0[2]. Built on top of it, the industry went one step further and added Turing complete virtual machine which enabled programs to develop and run on blockchain network. Ethereum project is the most well known example of such implementation [3]. Thus, smart contract based blockchain 2.0 was born. Blockchain uses cryptography technologies to enable parties in a trustless environment to reach an agreement. At the same time, tokens act as a medium to carry real values similar to currencies issued by centralized banking agencies, allowed members of the ecosystem to receive rewards in participation of various activities. Given the fact that blockchain technology is decentralized, anonymous and immutable, it is well equipped to serve as an educational platforms, and has the ability to generate education credentials and contents.

Based on these properties of blockchain technology mentioned above, we believe that blockchain will provide a solid foundation for us to build an open and trusted AI platform. However, the current architecture of blockchain faces the following challenges in order to implement our goal.

Consensus and Incentivization

The earliest blockchain consensus protocol is Proof-of-Work (POW)[2]. When a new transaction is added to the blockchain, a community of miners, or computing nodes, compete to solve difficult cryptographic problems, these who reaches an answer first adds the transaction to the blockchain, and in return receives a fraction of cryptocurrencies as rewards. While PoW empowered world's first famous cryptocurrencies, such as Bitcoin and Ethereum, its limitations appear to be more obvious as the blockchain industry evolves. PoW consensus requires intensively concentrated power to operate and making it hard to scale. Furthermore, PoW does not properly incentivize as if someone tries to cheat when creating a new block, their inappropriate activities can be forgiven by other nodes in the network. Thus, more granular considerations are needed around consensus protocols in order to encourage the right incentives.

Virtual Machine
Today's virtual machines aim for Turing completeness by allowing anyone to develop any programs in one platform, however, virtual machines on blockchain really merely support transactions and programs with standard control flow, which is not ideal and cannot be optimized for AI models. Unlike traditional programs with their standard parameters and debuggable codes, AI models come in all sizes and shapes, and they demand huge amounts of data sets that can also come in with various types. The virtual machines are simply not a sufficient one-size fit all solution for AI models.

Transaction
Transaction capacity on blockchain has become a scarce resource, resulting in slow confirmations and high fees. While training and inferencing AI models require large amounts of data, time, and computing resources, the limitation of slow and costly transactions will significantly affect the efficiency to quickly iterate of AI models. Only with the right incentives, effective rewarding system and a fast enough transaction delivery infrastructure will catalyze healthy and collaborative contributions for AI to advance.

Block storage
AI models usually contain large scale parameters, and training and testing data set are also large. The current block storage model cannot support such large data from a variety of AI models and problems. The limited extensibility impedes the development of an open AI platform.

TuringNet Solution

To address the challenges of AI and blockchain, the time has come for us to come forward and offer world's first and most comprehensive solutions to ensure the future of AI technology remains accessible, democratized and trustable. This goal has been motivating TuringNet to create a decentralized artificial intelligence platform with trusted models that follows Bitcoin's blockchain path that can be widely implemented and eventually take the stage as the new norm for artificial intelligence in the years to come.

The platform sets future advancement and applications of Artificial Intelligence free from any controls of centralized organizations, thus serving the best interests of the AI community that reshapes humanity. TuringNet encourages the public to participate in scalable and collaborative model training and prediction, significantly reduces repetitive model training and fully utilizes idle computing resources, at the same ensuring data security. This is achieved by TuringNet’s novel infrastructure design and groundbreaking incentivizing LBFT consensus mechanism, which is tailor-made for and fully compatible with various AI model training needs. We believe our groundbreaking approach will be hugely disruptive as TuringNet’s DModels are accessible by the public, verifiable and trusted, and will maximize the benefits to all ecosystem partners and participants.

TuringNet’s solution has the following key advantages:

- **Reaching Consensus on Useful Work:** To incentivizes computing devices for useful work, we innovates a novel and groundbreaking consensus protocol: **LBFT protocol (TuringNet patent pending)**.
- **Building Trust in Decentralization:** AI models trained on TuringNet’s decentralized platform are guaranteed trustworthy.
- **Unleashing Unlimited Computing Power for AI:** Computing power capacity available for model training and inferencing is far beyond any centralized platform could ever provide.

TuringNet Overview

With a world class pioneering engineering and research team, TuringNet is creating the world's first blockchain network with consensus, incentives, extensibility, and virtual machine features dedicated for an open and trusted AI platform. TuringNet's technology has the key innovations that are described below.

World's First True AI Consensus

The most cutting-edge technology used in TuringNet is the consensus mechanism named **LBFT (Learning Byzantine Fault Tolerance)**. The name itself is fairly self-explanatory, the consensus is part of the Proof-of-Useful-Work consensus family which has very few members in the blockchain world. Compared to other Proof-of-Useful-Work consensus mechanisms that are either trying to integrate off-chain or real world work with questionable feasibility, or solving practical but limited problems such as searching for prime numbers[4], **the LBFT consensus employs AI tasks themselves as useful work, both in digital world and on chain, which both guarantees its feasibility and solves practical problems.**

Although many blockchain projects are combining AI and blockchain technologies, most of them either use AI as a tool to reach consensus, or use traditional consensus while executing AI tasks. Communities of these projects are not incentivized by the performance of nodes in AI tasks. As a result, few participants are willing to contribute their computing power, while most power is still being wasted.

On the contrary, the innovation of LBFT in TuringNet lies in the integration of AI task execution within every iteration of consensus. In each iteration, there is a well-defined value regarding the performance, particularly in training AI models, i.e., the loss. By voting the minimum loss in each iteration, the consensus is able to select the best performing node and confirm it with token rewards. **The LBFT consensus mechanism is the hallmark of TuringNet and empowers it as world's first blockchain with seamless integration between AI tasks and the blockchain consensus. Therefore we call it the world's first true AI consensus.**

Building Trust in Decentralization

As we stated above, models trained in a centralized AI platform are controlled and only accessible for internal usage, and mainly serve the interest of centralized organizations.

To sustain a profitable business, it is hard to put the interest and privacy of users and communities first. TuringNet's decentralized models (DModels) are trained in a decentralized and transparent environment. The model architecture is open and goals of training are visible to everyone. As shown below, both model architecture, and parameter updates are written on the chain and accessible to everyone.

The underlying new consensus protocol invented by TuringNet is also designed for building trust. Any malicious nodes or data, under certain limits, will be excluded from consensus and block generation. In addition, the trained AI models on TuringNet can be verified, for example, what kind of features of user data are being used and what the goals are for optimization. **As a result, the models on TuringNet's decentralized platform are guaranteed trustworthy.**

Next-Generation Multi-Layer Blockchain with Extensibility for AI

TuringNet incorporates the next-generation multi-layer blockchain architecture, in order to meet the requirements of fast speed and large scale data for AI tasks, both in training and inference.

Performance, i.e. transaction speed, and extensibility are often identified as the key issues for Bitcoin[2] and Ethereum[3] as they are both of single-chain architecture. Multi-layer blockchain architecture is one of the most popular designs for solving extensibility and performance problems in many next generation blockchain projects[5–8]. The design idea is to keep the main chain secure and decentralized, while implementing fast transactions and extensibility to subchains or sidechains. However, all projects are competing over transaction performance, and none of them has the extensibility that supports a variety AI models with large amounts of data.

TuringNet also implements multi-layer architecture with more than 10 thousand transactions per second, and in addition, supports the extensibility for AI as another key innovation. The main chain in TuringNet is used for transactions, while each subchain is created for a particular AI model and problem. A secure deposit is frozen while the transaction is transferred from the main chain to a dedicated subchain. The huge training and testing datasets as well as model parameters for a particular model are stored in the corresponding subchain or off-chain. The actual execution of training and inference of an AI model is running on the nodes in the corresponding subchain. The LBFT consensus is able to run very fast as it only depends on the results in each iteration. As a result, without introducing any burden on the main chain, the performance of TuringNet is able to reach more than 10 thousand transactions per second across multiple subchains, and the extensibility is tailored for AI models, in which various AI models are executed and owned separately by subchains.

World-Class Virtual Machine for Deep Learning Network

As deep learning is becoming the mainstream and is capable of solving increasing numbers of AI problems[9–11], the success of TuringNet as an open and trusted AI platform should be supported by its intriguing and world-class deep learning virtual machine, named, Graph Virtual Machine (GVM). GVM adopts graphs as AI models, one of the key components in TensorFlow[12]. The abstraction of a graph can best describe a variety of deep learning network models, with nodes as operations and edges as data. The team is thriving to port all types of deep learning models to GVM, while optimizing the model storage and loading in a decentralized environment. Therefore, users and the community can easily migrate their TensorFlow models into TuringNet and grow TuringNet into one of the world's largest AI platforms.

TuringNet Architecture

Design Principles

TuringNet is creating the next generation open and trusted AI platform with the following architecture design principles:

High extensibility to run large scale AI models
The blockchain should be able to run large scale AI models with enormous training and testing data, as well as very high dimension model parameters. The extensibility is also designed for many different AI models that are trained to solve as many problems as possible.

Open protocols and virtual machine to support a variety of AI models
The protocols on the blockchain should be open so that the community can contribute to the AI models. The virtual machine of the blockchain is designed and optimized for AI models. A variety of AI models that solve different types of AI problems should be supported.

Ecosystem to incentivize the community and build trust
The blockchain as an open AI platform should create an ecosystem to encourage the community to participate, particularly with their computation resources. As a result, a carefully designed consensus and incentivization is required to support healthy growth for the community, while building trusted AI models that are beneficial to the community.

Guarantee security and decentralization
In a decentralized environment, the blockchain should be secure against attacks from malicious nodes and data.

System Architecture

TuringNet takes a completely new approach to create an AI blockchain that will reward each participant based on learning work done through data and computing resource contribution, thanks to a new LBFT protocol (TuringNet patent pending).

TuringNet employs two-level architecture with one main chain and multiple subchains. The underlying principle is that the main chain is used for the transactions and incentives ledger, while subchains are tied to particular AI problems in which the consensus of nodes is well defined for the same problem.

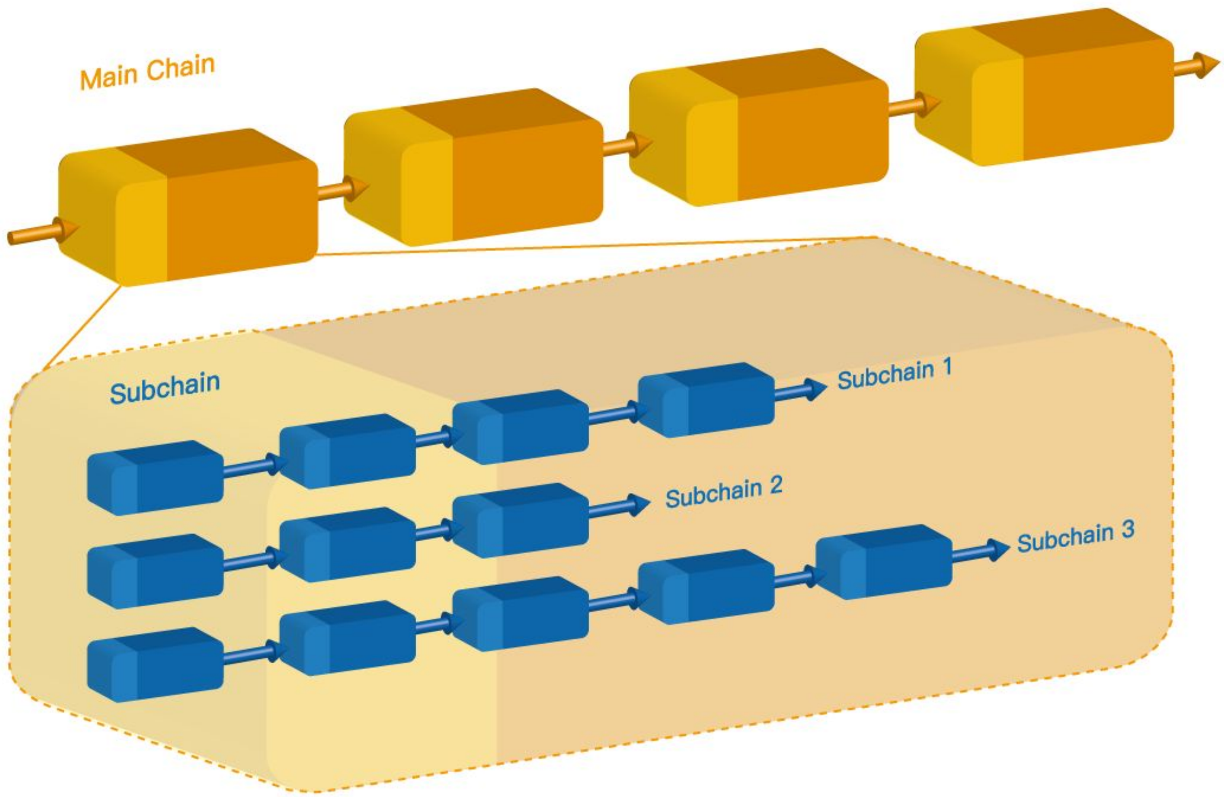

Figure 1. TuringNet Architecture

The advantages of the two-level architecture include:

- The separation of subchains from the main chain prevents the main chain from storing a large number of large scale AI models, and maintains high transaction speed of the main chain.
- The design with each model corresponding to a subchain, enables the extensibility for supporting various AI models, and provides a well-defined way to reach consensus and incentivize the community.
- The subchains are responsible for training and inferencing trusted AI models, while the main chain as a public ledger guarantees the security of transactions.

Graph Virtual Machine and Decentralized Models

To support a variety of AI models in TuringNet, we design the virtual machine by incorporating the graph models in TensorFlow[12] or similar machine learning platforms, while keeping the non-AI transactions running on EVM[13] (Ethereum Virtual Machine). We name AI related virtual machine and smart contract on TuringNet, as Graph Virtual Machine (GVM) and Decentralized Models (DModels).

Similarly to TensorFlow, a DModel in TuringNet is a computational program and smart contract specialized for an AI model. The DModel is described by a directed graph in GVM. The graph contains a set of nodes, each representing an operation respectively, and a set of edges connecting the nodes, each representing the data flowing between two nodes respectively.

Operations in DModel: Each node, as an operation, has zero or more inputs and zero or more outputs. The operations describe individual calculations of the data. For example, a constant operation has zero input and a single output to generate a constant value. A variable generation operation also has zero input and a single output to generate a writeable variable. A multiplication operation describing mult(x, y) = z, has two inputs x and y, and a single output z.

Tensors in DModel: Each edge, as a piece of data flowing in the graph, is described by a tensor. Tensors are of zero or more dimensions, and consist of a set of primitive values such as floats, strings, etc. The inputs and outputs of each node, i.e. operation, in DModel, are tensors.

API in DModel: After a DModel finishes training, the users of TuringNet can invoke the DModel to perform inferencing. In TuringNet GVM, we provide the description of the input and output of the DModel, both in tensors, and the address to the interface of the DModel. By providing the input tensors and invoking the DModel interface, the inference output will then be calculated and returned to the to the user.

On top of the low-level graph description of DModels, TuringNet GVM will also provide high level APIs for generating useful DModel templates for some typical deep learning networks, such as VGG[14], GoogleNet[15], ResNet[16], etc. As a result, instead of manually coding the graph for a given DModel, the users of TuringNet can use these API to create DModels by specifying some parameters like number of hidden layers, the number of nodes in each layer, and activation functions, etc.

Under the hood, TuringNet GVM provides various implementations on different devices, such as CPU or GPU acceleration, to execute the DModel, which is the abstract of an AI model. Our GVM aims to be compatible with TensorFlow models, so that many existing models running on TensorFlow can be ported seamlessly to TuringNet.

AI Subchain

The AI subchains in TuringNet are the core components for the extensible AI platform. Every DModel is supported by a subchain in TuringNet, which is connected to the main chain for transactions as a ledger. Each subchain is running LBFT (Learning Byzantine Fault Tolerance, TuringNet patent pending) consensus for block generation and node incentivization. The following figures demonstrate the structure of subchains in TuringNet.

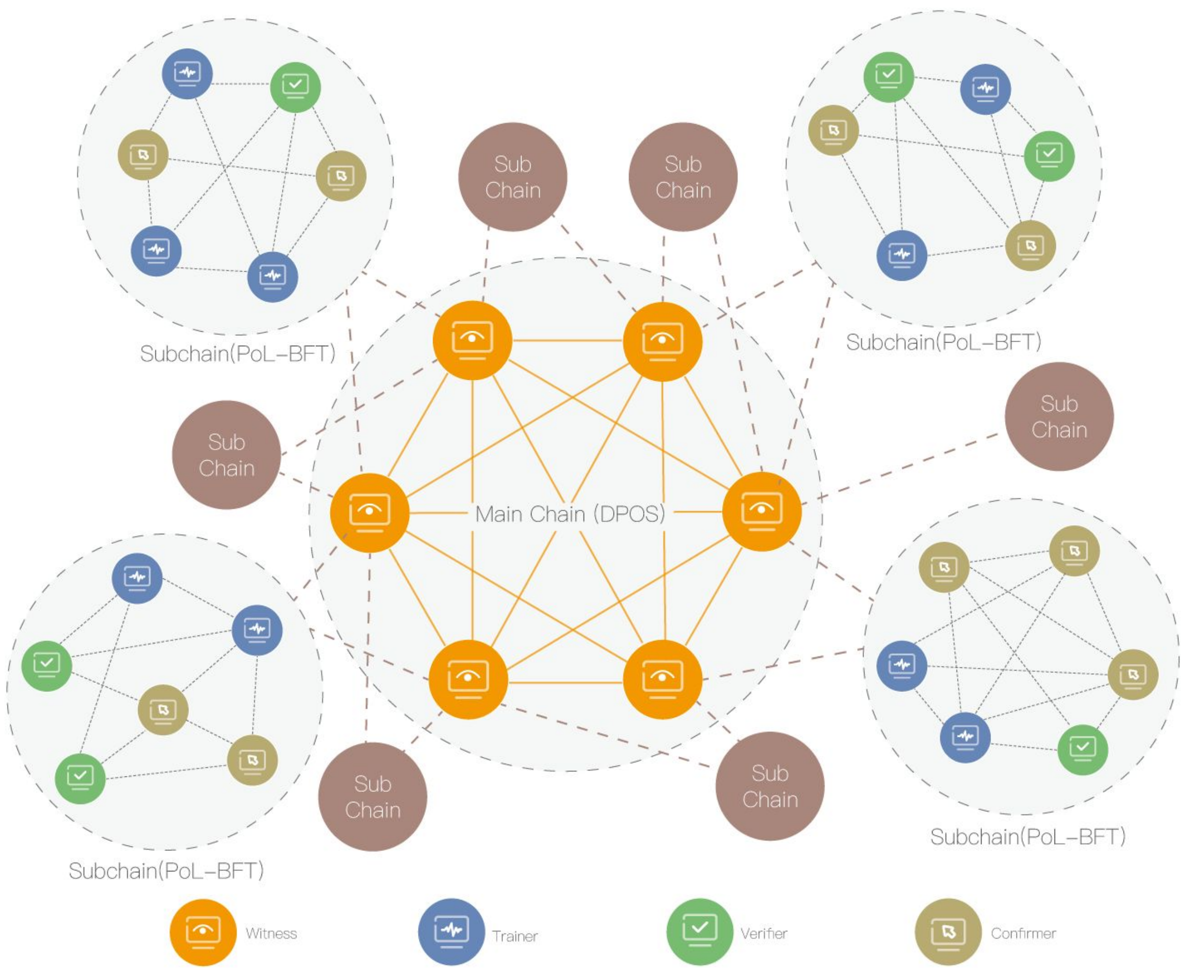

Figure 2. TuringNet AI Subchains

Structure: Each subchain consists of three types of nodes for the life-cycle of a DModel.

- Trainers are the nodes participating in the training of the DModel. When generating the next block, each trainer starts with the parameters in the previous block and searches for proper delta for the parameters. Then the trainer generates a new block, signs with its private key, and broadcasts it over the P2P subchain.
- Verifiers are the nodes participating in calculating and voting for the block with minimum losses. When generating the next block, each verifier receives blocks generated from the trainers for a certain period of time and calculates the loss from the new parameters in each block. Then all verifiers vote for the block with the minimum loss using BFT, sign with their private keys, and broadcast it over the P2P subchain.
- Confirmers are the nodes participating in confirming the newly generated block. Upon receiving the signed block with the minimum loss from the verifiers, the confirmers verify the calculated loss and the signatures, and also confirm the block by signing with their private keys.

Block producing: Each subchain produces and maintains a chain of blocks for the parameters of a DModel and the transactions for invoking the inference of the DModel. Upon creating a new subchain, the genesis block in the subchain includes the amount of initial deposit of TN tokens, the computational graph for the DModel, the initial parameters, and the testing data or their hash if stored elsewhere. The genesis block is signed by the private keys from the witness of the main chain, and can be verified by their public keys.

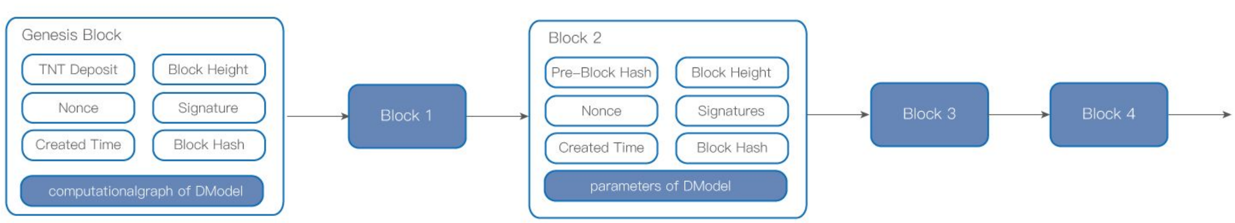

Figure 3. Blocks in the subchains

The following blocks in the chain consist of parameters with the minimum loss for each iteration on top of the previous block. The components in each block contains:

- The hashes of the previous block and the current block.
- The signatures of the block producer, verifiers and confirmers.

- The parameters of the DModel in the corresponding iteration, or the hashes of the parameters if stored elsewhere.

Mining rewards: Mining rewards are in the form of TN tokens in the subchain. During the training phase, an initial deposit of TN tokens is reserved from the main chain to the subchain, with an additional percentage of minted TN tokens. In each iteration, the voted block producer, one of the winning trainer, all verifiers and confirmers will receive the mining rewards.

During the inferencing phase, the verifiers will agree on a certain node in the subchain to perform the inference with the DModel, and the confirmers confirm the result and sign the block. Then the calculating node, all verifiers and confirmers will receive the mining rewards for the inference.

Data and parameters: Each subchain with the corresponding DModel requires training data, testing data, and the storage of parameters. TuringNet supports various storage form of the data and parameters.

- Training data are often stored off-chain, such as a separate data storage platform (IPFS[17] for example), or even locally in all or some of the trainer nodes.
- Testing data are the key for the consensus on the subchain, so either the entire testing data for small or medium size problems or the hash of the testing data are stored in the block. For public testing data, cross validation technique is employed for each iteration to calculate the loss. For testing data stored off-chain, such as IPFS, it is possible to use certain encryption technique to ensure only verifiers can access the testing data.
- Parameters of the DModel can be stored either on-chain or off-chain depending on the size of the DModel. In the former, the parameters are written in the block, while in the latter, the parameters are stored elsewhere, such as IPFS, and only the hash is written in the block.

Voting: Each trainer package the parameters get from training into a temperate block and sign the block with private key. Verifiers collect the blocks and check if the parameters in the block fit the DModel, then sign the block and broadcast it after verification. Every confirmer in a AI subchain gets to vote for the trainer with the lowest loss by certain loss-function of the DModel. The block from trainer gain more than ⅔ votes from all confirmers in the subchain will get signed by the first confirming confirmer. This block will add to the end of the subchian. Trainer will get rewards according to the blocks signed during the DModel training period.

Main chain

The consensus of the main chain uses DPoS[18,19] to reach fast transactions, high network security and maximum decentralization after initial launch of the network. It maintains the token ecosystem, and does bookkeeping for transactions, particularly the machine learning jobs in different subchains and incentives to the nodes of the subchains.

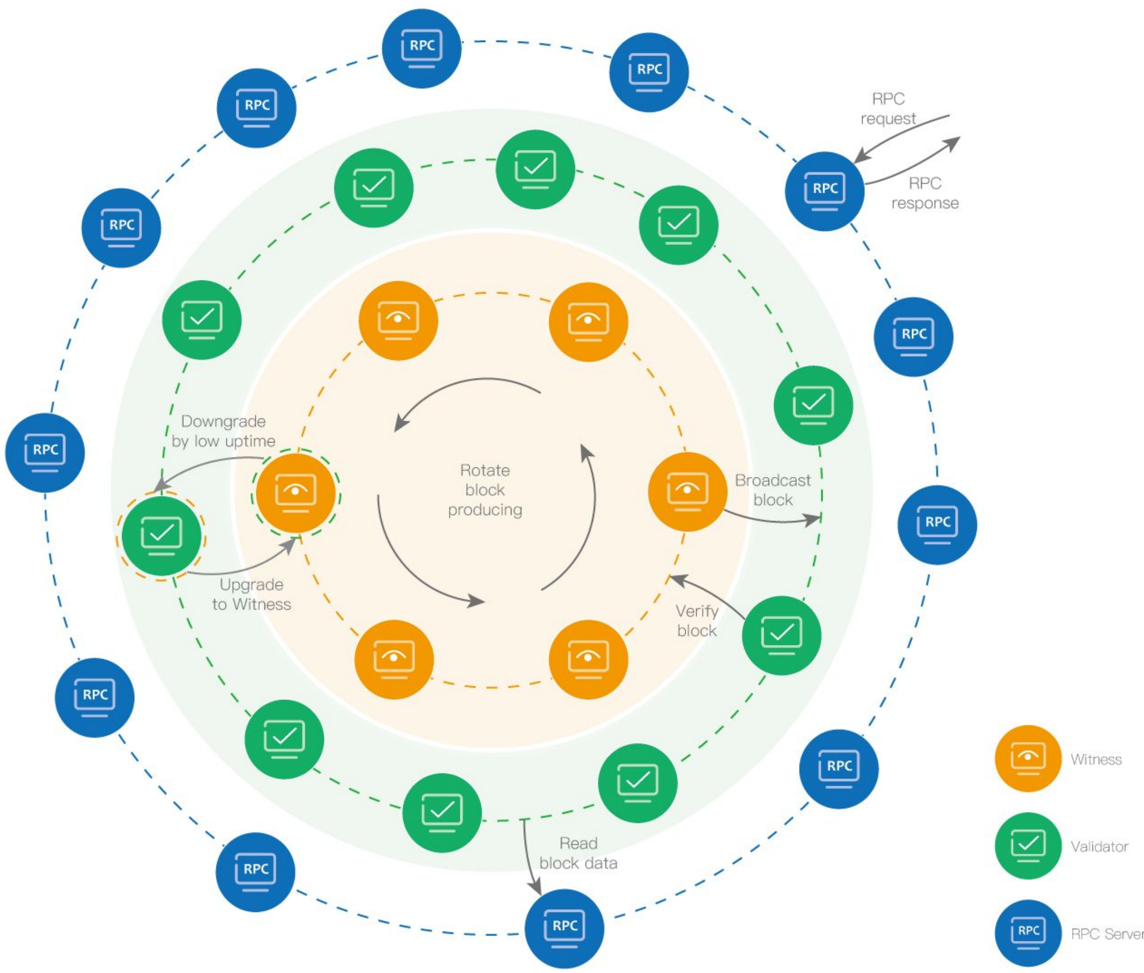

Figure 3. TuringNet Main Chain

Structure: The main DPoS network is composed and maintained by three tiers of decentralized nodes.

The top tier nodes called witness, who produce, sign and broadcast blocks.

The second tier nodes is validators, who provide the same service as witness except producing the blocks but verifying the blocks. A verifier could become witness when one of the witness failed to maintain the 99.9% uptime. Both witnesses and verifiers are voted by token holders.

The third tier nodes are RPC (remote process call) services, these node have entire data of main chain and response RPC requests from subchain and client.

Witness: A witness in main chain is an authority that is allowed to produce, sign and broadcast blocks. Producing a block consists of collecting transactions of the P2P network and signing it with the witness' signing private key. A witness' spot in the round is assigned randomly at the end of the previous block. A witness is approved by gaining enough approval voting from token-holders and required to have a 99.9% uptime to keep the approvement.

Block producing: A block of main chain will contain TN token transactions and necessary data to index subchian only to minimize the size of block. The producing rate will be 1-5 block per second for a maximum size of 32MB. A randomly selected witness will collect all transactions and new subchain requests during the processing period, pack the data into one block and sign the block, and broadcast the block on P2P network. All other witness and validators will verify the block, the block will be committed to the chain after more than ⅔ validators upvoted, followed by awarding the producing witness.

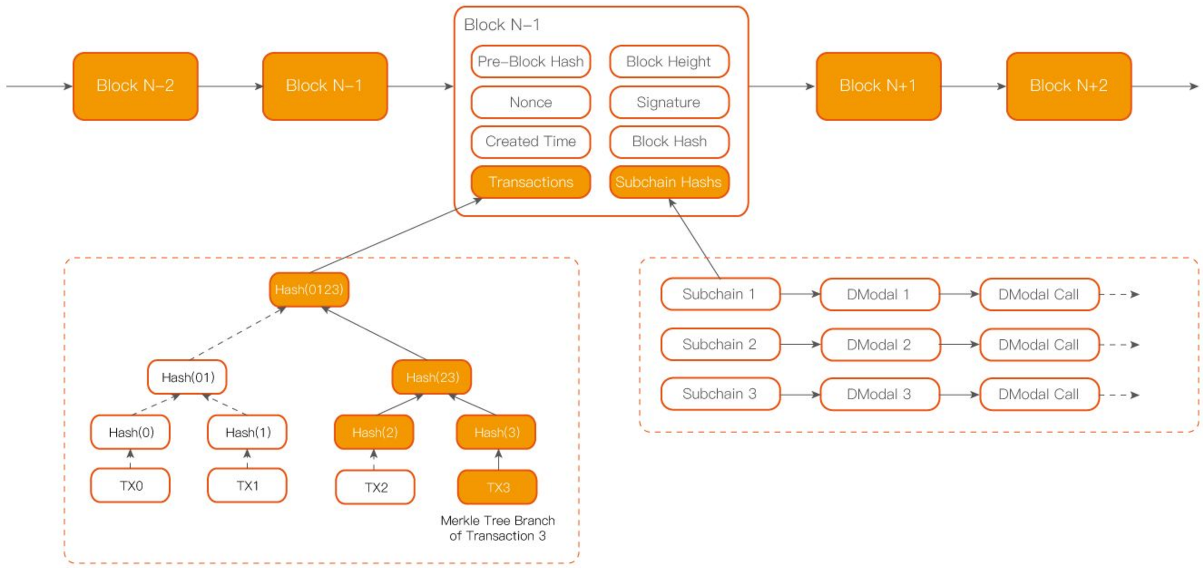

Figure 4. Blocks in Main Chain

Voting: Every token-holder gets to vote for someone to sign blocks in their stead (a representative if you will). Anyone who can gain 1% or more of the votes can join the board. The representatives become a “board of directors” which take turns in a round-robin manner, signing blocks. If one of the directors misses their turn, clients will automatically switch their vote away from them. Eventually these directors will be voted off the board and someone else will join. Board members are rewarded a small TN tokens to make it worth their time ensuring uptime and an incentive to campaign. They also post a small deposit equal to 1000x the average pay they rewarded for producing a single block. To make a profit a director must have greater than 99.9% uptime.

Mining pool: The witness node will be rewarded *n* TN tokens after producing, signing and broadcasting one block successfully. The rewarding amount *n* could be changed in a proposal during annual TN tokens holder conference.

TuringNet Core Functionalities

LBFT Consensus

One of the most core creation in TurningNet is the new consensus protocol for AI subchain: LBFT (Learning Byzantine Fault Tolerance). Compared to the current Proof-of-Work (PoW)[2] consensus which calculates useless hashes, LBFT is a next-generation type of consensus, or proof of useful work. The useful work is measured by the contribution to AI DModel optimization in TuringNet. The main consensus diagram is illustrated as follows.

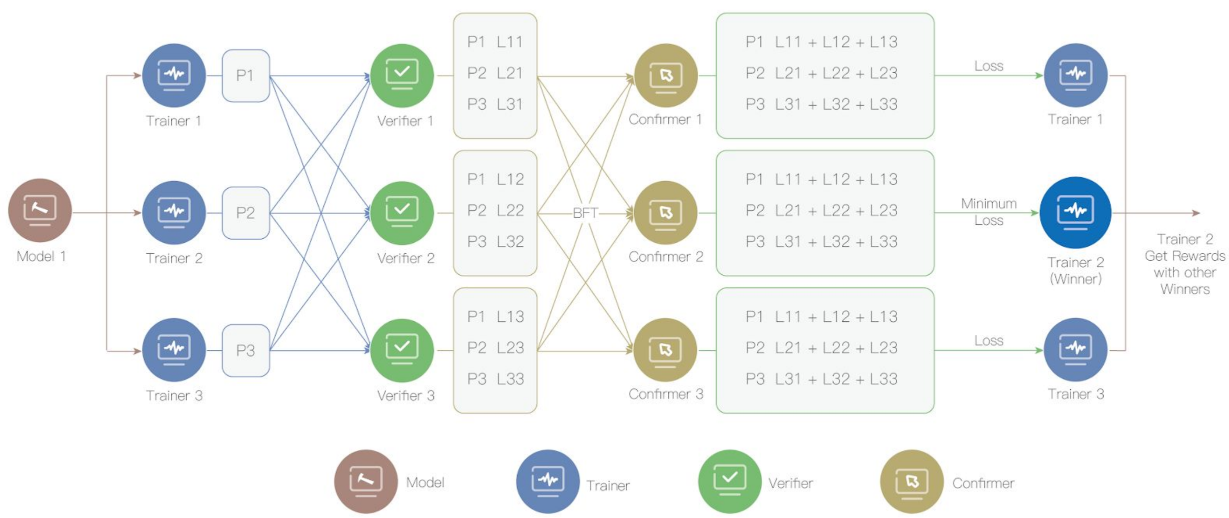

Figure 5. TuringNet LBFT Consensus

The consensus involves the three types of nodes in the subchain of TuringNet: trainers, verifiers, and confirmers. Each type is explained in details in the previous chapter.

Here we present the algorithms for LBFT:

$Algorithm\ 1: Block\ production\ for\ a\ DModel\ in\ TuringNet\ subchain$

$Initialize\ parameters\ w_0\ in\ B_0,\ the\ genesis\ block$

$for\ each\ iteration\ r = 1, 2, ...\ do$

$\quad w_r \leftarrow (Parameters\ retrieved\ from\ block\ B_r)$

$\quad T_r \leftarrow (Set\ of\ m\ nodes\ for\ training)$

$V_r \leftarrow$ (*Set of n nodes for verification*)

$C_r \leftarrow$ (*Set of t nodes for confirmation*)

for each training node $i \in T_r$ *in parallel do*

$w^i_{r+1} \leftarrow Train(i,\ w_r)$

generate B^i_{r+1} *with* w^i_{r+1} *and signs with private key* SK^i

broadcast B^i_{r+1}

for each verification node $j \in V_r$ *in parallel do*

receives a set of B^i_{r+1}

choose $x^j_{r+1} \leftarrow LBFT(\text{set of } B^i_{r+1})$

for all verification node $j \in V_r$ *do*

vote for $x_{r+1} \leftarrow BFT(x^1_{r+1}, x^2_{r+1}, ..., x^n_{r+1})$

sign $B^{x_{r+1}}_{r+1}$ *with private keys* SK^j

for all confirmation node $k \in C_r$ *do*

Confirm $B^{x_{r+1}}_{r+1}$ *and sign with private key* SK^k

$B_{r+1} = B^{x_{r+1}}_{r+1}$ *with all signatures*

$Algorithm\ 2:\ Train(k,\ w)$

$B \leftarrow$ (*split* P_k *into batches of size* B)

for each local epoch i from 1 *to* E *do*

for batch $b \in B$ *do*

$w \leftarrow w - \eta * \Delta l(w;\ b)$

return w

(B is the local minibatch size, E is the number of local epochs, and η is the learning rate.)

$Algorithm\ 3:\ LBFT(B)$

for each block B_k *from set B do*

L_k = *calculate loss with parameters* w^k *in* B_k

return $\underset{k}{argmin}(L_k)$

$Algorithm\ 4:\ BFT(x^1_{r+1}, x^2_{r+1}, ..., x^n_{r+1})$

each x^j_{r+1} *is the index of block* $B^{x^j_{r+1}}_{r+1}$

where the calculated loss L_j *is minimum from a given verification node* $j \in V_r$

use BFT algorithm over all verifications nodes in V_r *to reach consensus for*

$s = \underset{j \in V_r}{argmin}(L_j)$

$return\ x_{r+1} = x^s_{r+1}$

With the above algorithms, the nodes in a subchain can reach consensus on the blocks containing the parameters during the training phase. The consensus can generate trusted AI DModels against malicious nodes. First, the voting and confirming steps on the verifiers and confirmers can be secure against attacks with less than ⅓ of nodes respectively using BFT algorithm. Second, any malicious training node or malicious training data on the node may yield to results either failed to verify or with much higher loss. With less than ⅓ of malicious training nodes, the following voting and confirming steps will also eliminate the results from the malicious nodes and guarantee that the DModel is trusted in a decentralized environment.

Training as Mining

The major incentivization for the community in TuringNet is to mine TN tokens when training DModels in subchains. The figure illustrates the block generation between the main chain and subchains in TuringNet, from the mining perspective.

In the figure, the main chain and the subchains interact with each other when creating and finishing machine learning jobs. While each block in the main chain is rewarded with DPoS consensus, each block in a subchain is rewarded with LBFT consensus, during the training phases.

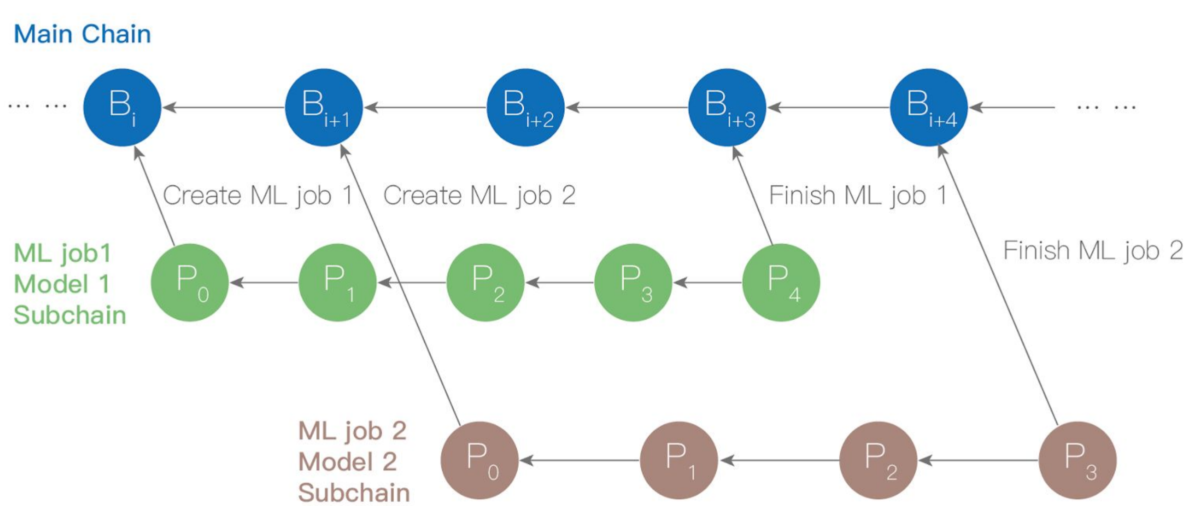

Figure 6: Block production during training phases

The following diagram depicts a workflow for a machine learning job:

1. An AI developer who wants to train a model, invokes the main chain with a machine learning job start transaction. Certain amount of TN tokens are submitted as a deposit for incentives for the miners.
2. The main chain freezes such amount of TN tokens submitted, as well as a percentage of newly minted TN tokens, as the deposit, and creates a subchain for the requested machine learning job.
3. The subchain starts for training, and generate sub blocks for each iteration, with the consensus to pay subchain's training tokens for the leader nodes.
4. After certain number of iterations, the subchain reaches a good model, exceeds the limit of iterations, or run out of TN tokens deposit. In any case, the subchain and invoke the main chain with finishing the machine learning job.
5. The main chain pays TN token deposits according to the portions of subchain's training tokens to the nodes in the subchain.
6. The main chain writes a transaction stating that the machine learning job is finished, and notify the developer.

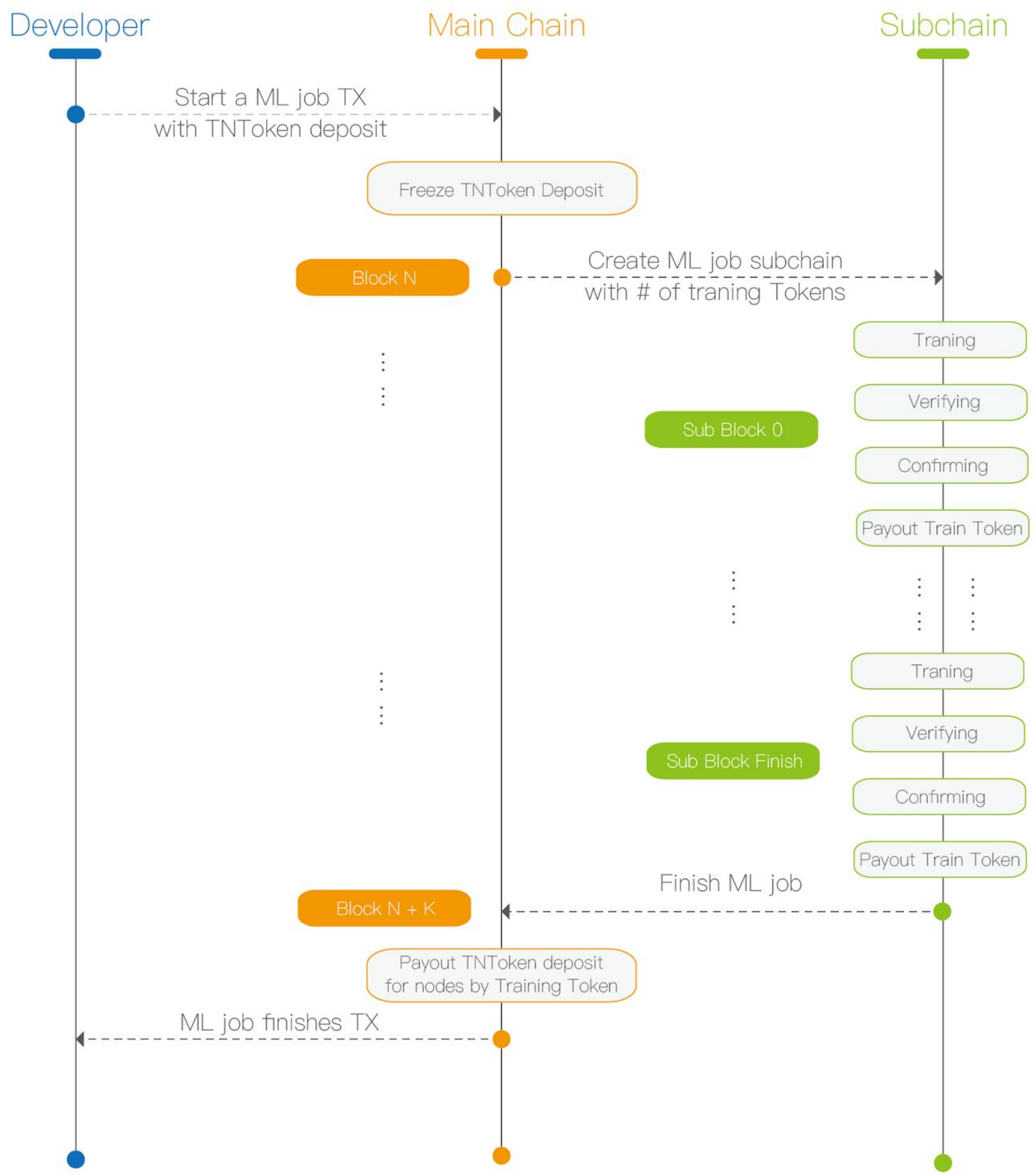

Figure 7: Training Workflow

Inferencing as Mining

Another important incentivization for the community in TuringNet is to mine TN tokens when inferencing DModels in subchains.

The following diagram depicts a workflow for a inferencing task:

1. An AI user who wants to inference a model, invokes the main chain with a inferencing task start transaction. Certain amount of TN tokens are submitted as a deposit for incentives for the miners.
2. The main chain freezes such amount of TN token deposit, and creates a inferencing block in the subchain of related DModel for the requested inferencing task.
3. The subchain starts for inferencing, miners execute the task in the order of higher fee paid and package the result of inferencing in to the block. The verifier check the result data in this block and sign and broadcast the block.
4. The main chain pays TN token deposits to the nodes in the subchain.
5. The main chain writes a transaction stating that the inferencing task is finished, and notify the user, the user could get the result from the subchain block.

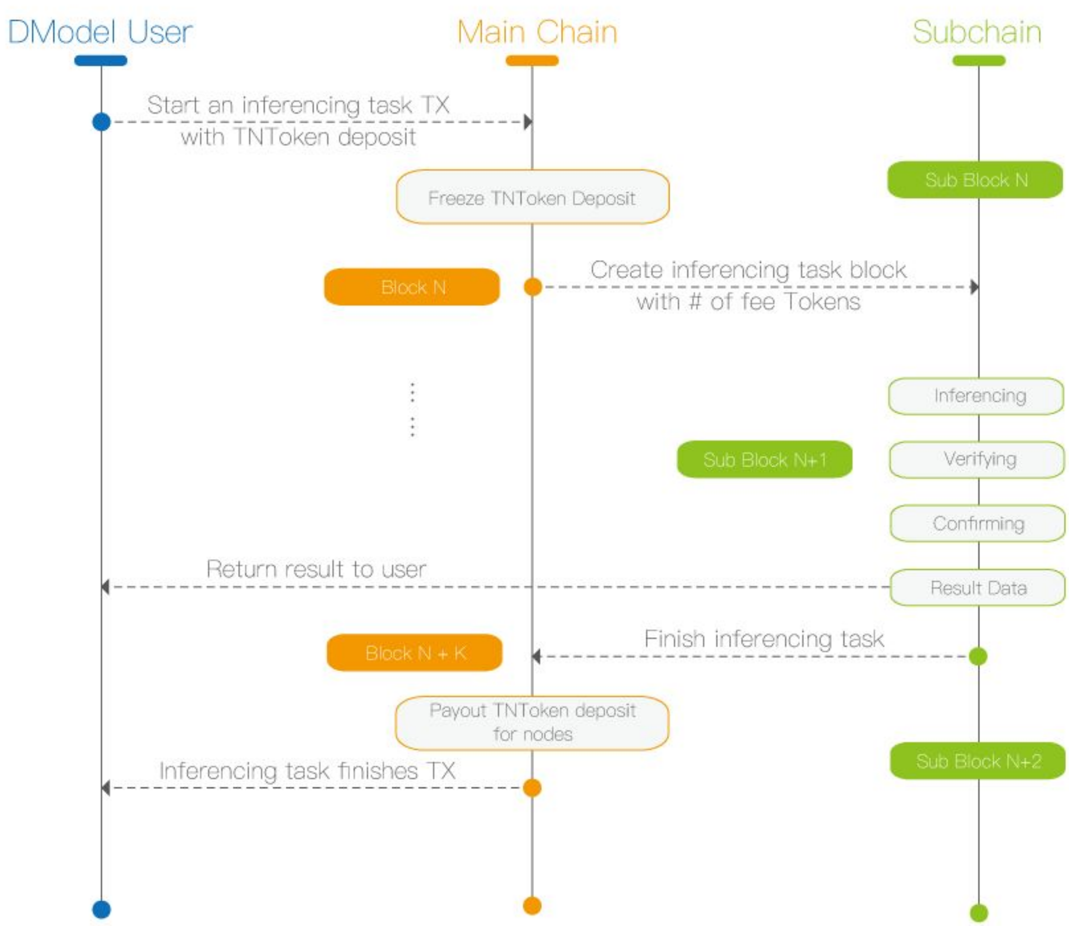

Fig.8 Inferencing Workflow

AI Marketplace

With the above core functionalities, TuringNet will create a marketplace for open and trusted AI DModels. As shown in the following figure, the marketplace is built on top of two key layers:

- The first layer is decentralized machine learning for training models in decentralized nodes;
- The second layers is LBFT Consensus for training trusted models in a decentralized environment and incentivizing the community.

On top of the above layers, the marketplace with TuringNet tokens creates a token economy for the open and trusted AI platform.

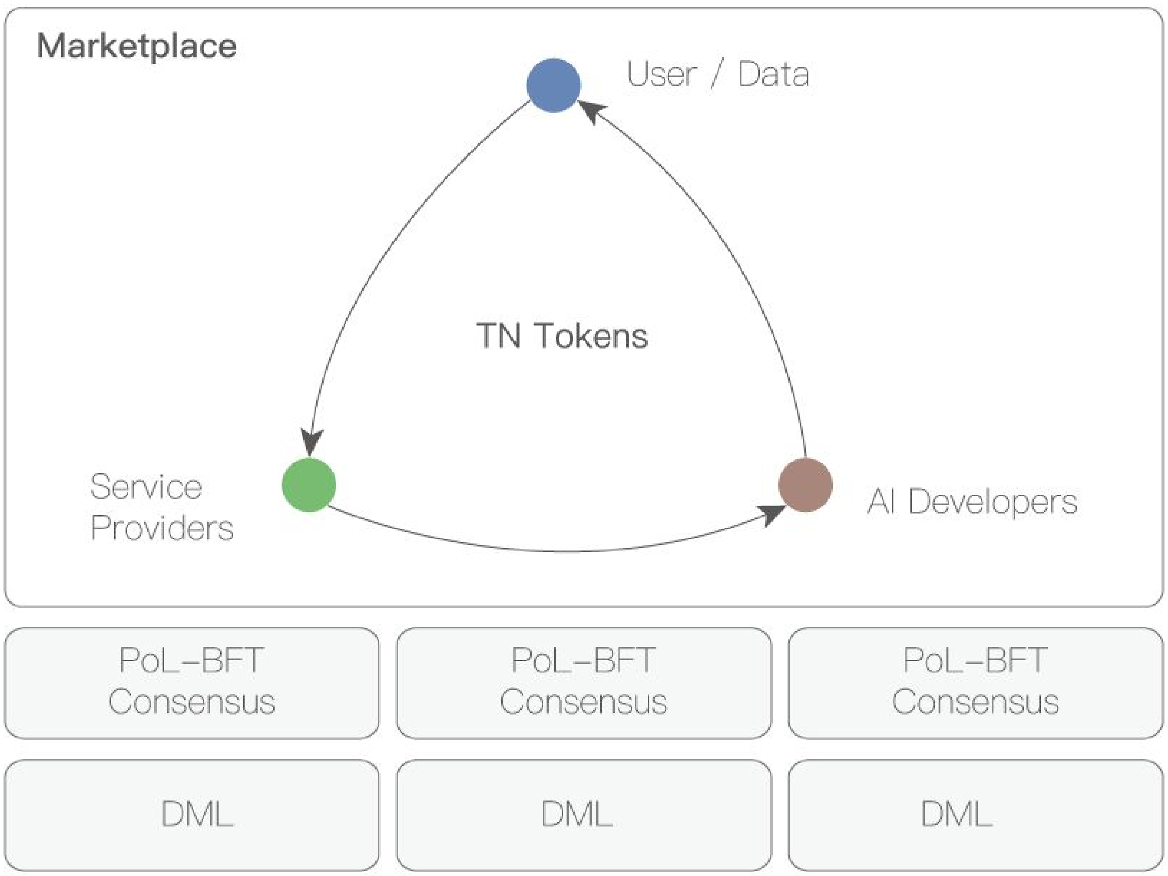

Figure 9: TuringNet Marketplace

Applicable Scenarios

We aim to establish a sustainable, safe, and highly efficient distributed AI ecosystem with endless possibilities that brings benefits to all of its participants, including ordinary users, AI companies, App Developers and service providers from all industries. By revolutionizing LBFT consensus mechanism, maximizing the value of computation power that sits on the network, allowing complete ownership data to its sources, TuringNet will dramatically broaden the usage of AI to everyday life.

Our Machine Learning models are designed to be collaborative and decentralized. This approach enables mobile devices to keep all the training data local on device, while working concurrently to learn a shared model, truly getting rid of the need to store data in a centralized platform.

On the other hand, the fact that the decentralized AI has become customizable means that individual users will be able to tailor the AI technology to their specific needs, for instance, sending personalized messages to your self-driving car, creating a software program that can translate and aggregate customer information into meaningful and actionable insights for a company, or recommending a course of action for a decision maker that is seeking intelligent and objective advice. All of these features of the future model will contribute to make the system becoming extremely high-efficient with low-latency, which is a drastic breakthrough of the today's blockchain systems and platforms.

Fast-forward to the future, and imagine multiple ways of how ordinary individuals and businesses will breath on AI applications. All of these will become possible and feasible based on a decentralized system on blockchain. Here are a few examples:

Smart City

Enhance cities, conserve energy, reduce congestion, improve traffic and safety, and empower IoT - The concept of a 'smart' city[20] has been around for years, and finally there's a reliable solution to harmonize data to provide city planners, residents, businesses, and applications real intelligence. This includes data collected from its residents or citizens, different kinds of devices, and personal assets that is analyzed or processed to monitor and manage traffic and transportation systems, power supply, water supply, waste management, law enforcement, schools, libraries, government systems, hospitals, and other community services.

In order to achieve this, data must work collaboratively together to train AI models[21]. However, the problem in today's world is that most data is still being held by tech giants, and isolated. For example, facial recognition technology is a well-known application for video processing and image detection that could be used to monitor and potentially reduce crime activities. But facial recognition alone won't be able to tell the whole story, if we have a safe and robust platform that helps to train the machine on one's images, along with their financial status, recent purchases, frequently visited locations and etc., that for sure will help to paint the picture. On the other hand, how would one feel about allowing one company to hold a zillion images of oneself? What if they are collecting one's facial images for identity verification and storing these images in one centralized database, what could happen if their system is hacked someday and all of the identity images get stolen for malicious uses? This may be just the tip of the iceberg.

Now with TuringNet decentralized AI model (DModel) training ecosystem, we can overturn how data is handled and harmonize it to maximize the impact. It all comes down to a safe community where users have the control over their images and all other personal data, at the same time still being able to contribute to a smart AI system that would benefit other participants in the ecosystem. Users would download the current AI DModel to their local devices, enlighten and improve the DModel by authorizing local Machine Learning training on their mobile devices, and then summarizes the update and upload for sharing. AI companies and service providers will be able to send tailor-made requests for data needs, and offer to reward TN tokens at a market price. Not only that TuringNet will be able to unlock massive untapped private user data, App Developers and service providers be able to adopt AI in a more meaningful way by rewarding and promoting for the good use of these user data. TuringNet's platform establishes a healthy and active ecosystem with fairness and transparency based on contributions.

Image and video processing is just an example of how Machine Learning can be applied in our daily lives. Today's existing data available (aka Big Data) is already capable of helping machines to gain deeper understanding of humans to better serve and increase productivities of human lives.

TuringNet is the disruptor in this space that brings everyone together in a collaborative, trustable, and fair environment that mitigate concerns around data privacy and protection. Decentralizing AI training is the only way to unleash the potential of data altogether. Participants within the Big Data ecosystem will get a fair share of reward based on their contributions. Whether it is sensor data, image related data, user behavior data, community data, traffic data, healthcare data and etc., we are essentially opening markets for useful private data, maximizing the true value of AI, promoting interpretability and supplementation of data training, and allowing all the resources to become more accessible. Ultimately, blending all these data together for the machine to make sense and gain a deeper understanding of how to serve its users is what it takes to build up a truly ‘Smart’ city.

In addition to that, TN Tokens play a critical role to promote a vibrant and sustainable Smart City ecosystem. TuingNet creates a healthy market that is available for anyone to acquire data, train models, and apply AI to their products and services at market driven prices. By eliminating all the middle men and layers of transactions, TuringNet ensures that its participants receive the largest benefits.

Genome

Would you like to leverage today’s technologies to decode your genome, but at the same have the full control and ownership of your genomic related data? Based on TuringNet’s technical implementation of the decentralized AI training platform, it is finally possible to bring this idea to life.

How does TuringNet’s platform looks like when incorporated and injected with genomic data? We will finally be able to eliminate all the middlemen, and solve bottlenecks in the current genomic data market such as low data quality, data fragmentation, lack of data standardization for effective algorithm training, and limited data acquisition from users. The blockchain based genomic data sharing and analysis platform combined with an effective LBFT consensus mechanism, which is all-in-one place of TuringNet platform, will empower scientific researches and unlock the power of genomic applications by training AI DModels on user’s devices. There exists tremendous opportunities to unleash the decoding of the human genome sequence:

- **Individual Opportunities**: Genome information is critical for making best choices in regard to all health-related questions, including medical treatments, family planning, personal diet, and exercise regimen. We are now entering the age of more precise and personalized medicine and healthcare, with an increasing focus on prevention.

- **Industry Opportunities**: Genomic data sets can be used to identify associations between genetic variants and diseases. There is tremendous value in this alone.

To prove the feasibility of this approach, TuringNet has locked down several genomic companies, such as Somur Technology Ltd.[22], for partnerships and defined the scope of the joint force.

This model empowers personal genomics companies by connecting **data owners** and **data buyers** directly. Data owners will have access to acquire their personal genomic data from sequencing facilities or other sources, join the TuringNet LBFT, peer-to-peer network and directly deal with data buyers using TN tokens. As detailed in the following sections, this model **reduces** effective sequencing **costs** and **enhances protection** of personal genomic data. It also satisfies the needs of data buyers in regards to data availability, data acquisition logistics and resources needed for genomic big data.

- Lower sequencing costs
- Local AI training
- Efficient data acquisition
- Big data ready

Smart Assistant

With TuringNet, smart assistants can be built on a decentralized network. In this system, all smart assistant devices work together to train and improve models with local data based on a consensus protocol. Users contribute their data, and get rewarded with tokens for their contributions for verifiable data and added value for the training model. Privacy issues are mitigated since user data was never taken out of user devices for training purposes, but stored in local. And users will be rewarded with a fair share as every time their data helps to improve a model, they will be compensated accordingly.

Once the system launches, third-party applications can be integrated to magnify the benefits to all participants. This self-forming, sustainable ecosystem will be appealing to all application developers as they will be able to gain easy access to:

- All **users** across the entire ecosystem: users can find and utilize your application data from the open marketplace. The large user base would be the perfect place to pilot, launch, or optimize new applications, and thus catalyze quicker and more successful product adoption.
- All **user data** across the entire ecosystem: the platform will empower application developers with accumulative user generated data over time. To illustrate an example, when using ride-sharing applications, users can disclose their location data. From there, an Ecommerce application can benefit from this set of location data to build a more robust user recommendation system (since they were able to gain a deeper understanding of user preferences, routine, and behaviors).

Combining the data intelligence generated from the Ecommerce application and ride-sharing application, AI technology will be able to further improve the AI models that were trained locally, and ultimately improve personalized smart assistant. And the smart assistant AI model will keep evolving towards a true intelligent assistant as users have the incentives to contribute their data for local model training, at the same time, allowing the assistant to collect more user behavior data over time.

- AI **models**: TuringNet is committed to nurture application developers with the well-trained model that's shared on the blockchain, and enable developers to build more applications in an effective and scalable way.

Digital Advertising

Nowadays digital advertising ecosystem has been heavily monoLBFTized by a few giants, who also possesses huge amounts of centralized user data. User privacy and data protection have been constantly an ongoing battlefield for many that are involved in the digital marketing ecosystem.

In a perfect world, without compromising privacy issues, advertisements would still create tremendous values to both end users and businesses by leveraging existing technologies to match relevant information between customers and products or services, and help customers to make more informed decisions at the same time increase business exposure. TuringNet aims to bring this perfect world to live.

In TuringNet's ecosystem, users and businesses are rewarded by contributing useful data for AI model training purposes. Data providers are motivated to share these data since they will not have to worry about privacy issues in this environment, and they are financially rewarded for enabling a smarter system that ultimately used to serve for their good. AI models will become ever more intelligent overtime as they accumulate learnings from data, for example, the system will understand what matters the most for a user, what the user have been looking for, what would the user's comfortable spending level, or what would be an ideal to show certain ads to the user. And thus these AI models will become more capable of helping Marketing gurus to accurately target their highest potential audience and increase ads ROI.

Decentralization of AI models builds a solid foundation of user trust, and at the same time maximizes the power of AI to better serve targeted ads.

Self-driving Vehicles

Self-driving vehicles are cars that can drive themselves without human control. Ideally, self-driving cars can drive safely in an orderly manner together to increase road capacity, and also increase the value of the time that human spends while travelling by freeing up human and allowing them to perform other activities while on the road. However, one of the biggest bottleneck or challenges for self-driving vehicles is how to mitigate risks for potential accidents to 100%. In order to achieve that, the self-driving vehicles industry will benefit from collecting more user behavior and previous accidents data to uncover 1) user preferences and routines,, 2) drivers' cognitive system, 3) drivers and pedestrians' decision making processes, and gain a much deeper learning from previous accidents to further advance the predictive system.

Most of the self-driving car companies are currently isolated, in terms of collecting useful data to train their algorithms, and they live with the fear that their competitors might get ahead of them once gained access to additional relevant data. And this will linger around for as long as these companies decide not to collaborate for data sharing, which ultimately hammer the safety problem for customers.

TuringNet's solution helps to remove the roadblock by lowering the high bar of data collection to improve inter-predictability in this cases. App Developers and AI companies will be able to reach millions of user data specifically tailor to the needs of their models with the on device model training approach. With more accurate predictions over road trends and sentiments, the self-driving vehicles industry will eventually be able to ensure a reliable system and reduce accident factors to the minimum.

Fintech

Financial Technology is a new form of technology that aims to revolutionize the traditional financial services, by leveraging existing technologies to empower financial activities. Some of the examples include the use of mobile banking, online insurance platforms, P2P lending companies, or the use of cryptocurrencies. As much as the Fintech companies would like to disrupt the traditional financial services sector, they still need to overcome problems such as reliability and security, and how to avoid one centralized platform to mitigate risks of unauthorized attack.

With TuringNet solutions, these problems can be avoided, and furthermore, by allowing the cross training of user data around behaviors, preferences, predictions, historical transactions, investment choices and portfolios, and etc., we will be able to further expand the landscape of what Fintech companies could offer. Knowing that the data is only being accessed when allowed, and not being taken to somewhere else for the use of other purposes, plus the incentives of potential financial rewards, users can share sensitive personal financial data with peace of mind.

Conclusion

The world is ever changing, and as it does, the powers will shift as they always do. Artificial intelligence is becoming a new and very integral part of our daily lives, and it is absolutely essential for us to hold on to the control in order to guide the power dynamics that result from this shift. Establishing the world’s first and most comprehensive decentralized AI system that is reliable, designed for self-sustaining with endless growth potential, and offers AI technology and financial benefits to individual users is promising to ensure that the future rests in our collective hands. The evolution of blockchain, from Bitcoin to the decentralized computers that we have today, has opened the door to creating a truly decentralized system that can harness the power of artificial intelligence. Our next generation system will be able to a) maximize the usage of resources and energy and b) allow the general public to profit from the value of their computer resources and data to artificial intelligence, and c) further advance industries for providing smarter and more trustable products or services to their users. Not only TuringNet will bring great technical and financial benefits to the general public, it is also the ultimate groundbreaking solution to defend against the threat of a monoLBFTized artificial intelligence system in the future.

Team

Jun Hong - Jun was previously a Senior Engineer from Google, Microsoft and Amazon, where he filed multiple patents in the area of Machine Learning, and built a ML system from scratch for ads recommendation and generated hundreds of millions of revenue for Youtube. He is a crypto entrepreneur, and one of the earliest founders of Alphabet Blockchain Community (ABC). He is also an ex-Chairman of the Open Source Software Associations of Chinese Academy of Sciences. Jun specializes in Distributed Systems, Artificial Intelligence, Machine Learning, Deep Learning, Video / Image recognition, and GAN.

Kai Wen - Kai received his Ph.D from Stanford whose research fields include Quantum Computing, Cryptography, Information Theory, and Optimization Algorithm. He has hands on experience in AI and Machine Learning from Google before he co-founded multiple startups in the areas of Internet Services, Mobile Internet, Artificial Intelligence for Cloud and Ads, and provided AI solutions serving government sectors and Fortune 500 companies.

Dany Houde - Dany brings expertise in Distributed Systems and Encryption Algorithms. He had drastically increased the robustness of Amazon's EDI platform by implementing key security enhancements using encryption as well as secure communication and storage mechanisms. He graduated with distinction from University of Waterloo majoring in Mathematics and Computer Science.

Bella Wang - Bella is an innovative business strategist and crypto enthusiast. She has a diverse background of over 10 years in business operations, business development, partnership, and people operations in the high tech industry from Google and Tencent. She had designed and launched multiple strategic initiatives to engage users at multiple levels, and served as a trusted advisor for engineering teams in forming joint development to talent and resource management. Bella joins TuringNet to oversee operations, ecosystem partnerships and community growth.

Bibliography

1. Silver, D. *et al.* Mastering the game of Go with deep neural networks and tree search. *Nature* **529,** 484 (2016).
2. Nakamoto, S. Bitcoin: A Peer-to-Peer Electronic Cash System. (2008).
3. White Paper· ethereum/wiki Wiki · GitHub. *GitHub* (2015). Available at: https://github.com/ethereum/wiki/wiki/White-Paper.
4. PrimeCoin. Available at: http://primecoin.io/. (Accessed: 15th July 2018)
5. Poon, J. & Dryja, T. The Bitcoin Lightning Network: Scalable Off-Chain Instant Payments. (2016). Available at: https://lightning.network/lightning-network-paper.pdf. (Accessed: 15th July 2018)
6. Poon, J. & Buterin, V. Plasma: Scalable Autonomous Smart Contracts. (2017). Available at: https://www.plasma.io/plasma.pdf. (Accessed: 15th July 2018)
7. Loom Network - The Next-Generation Blockchain Application Platform for Ethereum. Available at: https://loomx.io/. (Accessed: 15th July 2018)
8. QuarkChain - A High-Capacity P2P Transactional System. Available at: https://quarkchain.io/. (Accessed: 15th July 2018)
9. Bengio, Y., Courville, A. & Vincent, P. Representation Learning: A Review and New Perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **35,** 1798 (2013).
10. Bengio, Y., LeCun, Y. & Hinton, G. Deep Learning. *Nature* **521,** 436 (2015).
11. Schmidhuber, J. Deep Learning in Neural Networks: An Overview. *Neural Netw.* **61,** 85 (2015).
12. Abadi, M. *et al.* TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.

(2015). Available at: https://www.tensorflow.org/.

13. Ethereum Yellow Paper. *GitHub* (2018). Available at: https://ethereum.github.io/yellowpaper/paper.pdf.
14. Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR* **abs/1409.1556,** (2014).
15. Szegedy, C. *et al.* Going Deeper with Convolutions. **arXi:1409.4842 [cs.CV],** (2014).
16. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. **arXi:1512.03385 [cs.CV],** (2015).
17. IPFS is the Distributed Web. Available at: https://ipfs.io/. (Accessed: 2018)
18. King, S. & Nadal, S. PPCoin: Peer-to-peer crypto-currency with proof-of-stake. (2012). Available at: peercoin.net. (Accessed: 23rd December 2013)
19. Tendermint: Byzantine-fault tolerant state machine replication. Or blockchain, for short. Available at: https://tendermint.com/. (Accessed: 2018)
20. What is smart city. Available at: https://internetofthingsagenda.techtarget.com/definition/smart-city. (Accessed: 15th July 2018)
21. Deloitte. Smart Cities, Big Data. (2015). Available at: https://www2.deloitte.com/content/dam/Deloitte/fpc/Documents/services/systemes-dinformation-et-technologie/deloitte_smart-cities-big-data_en_0115.pdf. (Accessed: 15th July 2018)
22. Somur Technology Ltd. Available at: http://www.somur.com/. (Accessed: 15th July 2018)