



# NEUROSEED

WHITEPAPER

Version 1.1

May 1, 2018

Contents of the white paper are subject  
to changes and improvements

# ABSTRACT

The Fourth Industrial revolution brings the implementation of such technologies like Big Data, Internet of Things, Virtual Reality, Augmented Reality, Machine Learning and Artificial Intelligence. Current market requirements are changing rapidly and use of machine learning algorithms based solutions can significantly increase business competitive advantages in the context of globalization. Back in 1995 on "Neural Information Processing Systems Workshops"<sup>[1]</sup> an idea of creating a platform that could provide a reliability of data sources and trained neural networks has been declared. Platform implementation would lead to the rapid development of the whole stack of machine learning based technologies and help to reduce the cost of its development, mass adoption and significant growth of derived systems efficiency. The concept of rationality of several machine learning models merging with their further transfer learning has been proposed and proved later. The common limiting factor of development and implementation of similar systems was the lack of reliable technology that could provide a decentralised digital trustworthiness for final machine learning models and data sources.

NeuroSeed platform solves this problem by means of combining cryptography and blockchain technologies and providing a reliable tool for creation, validation, trading and reusing of datasets and uniform reusable clusters of final machine learning models.

1 - NIPS\*95 Post-Meeting Workshops: Learning to Learn: Knowledge Consolidation and Transfer in Inductive Systems  
December 1-2, 1995; Marriott Hotel, Vail, Colorado  
<http://www.cs.cmu.edu/afs/cs/project/cnbc/nips/NIPS95/Workshops.html>

# DISCLAIMER

The attached white paper is meant to describe NeuroSeed's currently anticipated plans for developing its business and its NSD Tokens. Nothing in this document should be treated or read as a guarantee or promise of how NeuroSeed's business will develop or of the utility or value of the NSD; the document outlines our current plans, which could change at our discretion, and the success of which will depend on many factors outside our control, including market-based factors and factors within the cryptocurrency industries, among others. Any statements about future events are based solely on our analysis of the issues described in this document, and our analysis may prove to be incorrect. Purchasing NSD is subject to many potential risks. Purchasers of NSD could lose all or some of the value of the funds used to purchase NSD.

<b>ABSTACT</b>	<b>2</b>
<b>DISCLAIMER</b>	<b>3</b>
<b>1 INTRODUCTION</b>	<b>5</b>
1.1 Definitions, acronyms and abbreviations	5
1.2 Mission statement: why we created NeuroSeed	6
<b>2 OVERALL PLATFORM DESCRIPTION</b>	<b>7</b>
2.1 What is NeuroSeed	7
2.2 Benefits	7
2.3 Platform principles	7
2.4 Platform roles	8
2.5 Platform roles and components interaction	10
<b>3 TECHNOLOGY STACK OVERVIEW</b>	<b>12</b>
3.1 Blockchain	12
3.2 Final ML models merging	12
<b>4 PROOF OF CONCEPT</b>	<b>14</b>
4.1 Testing image dataset	14
4.2 Testing text dataset	15
4.3 Testing audio dataset	15
<b>5 INTERNAL ECONOMY</b>	<b>16</b>
5.1 What is Token	16
5.2 Internal micropayments and micro purchases	16
5.3 Payment distribution	16
5.4 Payment for updates	17
5.5 Motivation and validation	17
<b>6 BUSINESS MODEL</b>	<b>18</b>
<b>7 CHALLENGES</b>	<b>19</b>
7.1 Universal classifier of final ML models	19
7.2 Motivation system economical model	19
7.3 Abstraction vs reusability	20
<b>8 ROADMAP</b>	<b>21</b>



# 1 INTRODUCTION

## 1.1 Definitions, acronyms and abbreviations

**Artificial Intelligence (AI)** is an area of computer science that emphasizes the creation of intelligent machines that work and react like humans.

**Application programming interface (API)** is a set of defined methods of communication between various software components.

**Big Data** is a voluminous amount of structured, semistructured and unstructured data that has the potential to be mined for information.

**Blockchain** is a technology that describes the way of keeping the actual copies of uniform database between distributed network nodes, that includes description of principles of reaching an agreement between these nodes about updating the database information.

**Blockchain ledger** is a shared, replicated and synchronized database that is distributed between blockchain network nodes.

**Dataset** is a collection of related sets of information that is composed of separate elements but can be manipulated as a unit by a computer.

**Composite Dataset** is a virtual dataset that consists of parts of several datasets.

**Machine Learning (ML)** is a field of computer science provides systems the ability to automatically learn and improve from experience without being explicitly programmed.

**Machine Learning Model** is a machine learning algorithms implementation with a certain configuration for solving specific task.

**Final Machine Learning Model** is a model that is saved for later use.

**Transfer Learning** is a method that focused on storing knowledge gained while solving one problem and applying it to a different but related problem.

**Neural Network (NN)** is a mathematical model of biological neural networks to identify complex dependencies between input datasets and output data, and perform generalization.



## 1.2 Mission statement: why we created NeuroSeed

Our vision is to allow anyone to participate in the global machine learning markets provided by NeuroSeed. We provide the empowerment, access and tradability so anyone can take part by creating a viable hub for Big Data suppliers, data storage suppliers, computing power suppliers, machine learning models developers, distributors, validators and customers.

We introduce a new alternative for machine learning market participants interaction by making an extremely simple tool for machine learning models design, training and use.



## 2 OVERALL PLATFORM DESCRIPTION

### 2.1 What is NeuroSeed

NeuroSeed is a platform that combines Big Data suppliers, ML algorithms developers, computing power suppliers, data storage suppliers, validators with domain expertise, distributors and customers. As a result of a synergy there is a creation of the unified ecosystem that provides incentive mechanisms for all participants in the machine learning market.

### 2.2 Benefits

**Reuse of final ML models.** Final ML models are reused for merging and transfer learning. This approach allows to reduce the cost of initial ML models training, and also reduce time for new ML models development and time for their implementation in customers' software.

**Datasets composing.** New virtual composite datasets are created on top of existing ones by means of necessary data ranges selection. This approach enables both storage expenses reduction and increase in available datasets diversity.

**Final ML models updating.** All interconnected final ML models are updated by transfer learning one of final ML models on a new dataset. This enables keeping final ML models in actual state.

**Use of blockchain for ML models and datasets reliability.** Use of blockchain makes it possible to check the authenticity of datasets and final ML models together with related metadata and history of related transactions. Thereby it is possible to check any of datasets and final ML models to be consistent and unmodified.

**New liquidity hub for ML market players.** NeuroSeed introduces the new liquidity hub for ML models developers, Big Data suppliers, computing power suppliers and data storage suppliers.

### 2.3 Platform principles

**Auditability.** Every transaction is cryptographically signed, and the history of all operations and transactions are stored on the blockchain, together with hashes and metadata of datasets, machine learning code and final ML models.



**Platform independence.** NeuroSeed platform acts as a complex marketplace and is independent from specific computing power suppliers, data storage suppliers, Big Data suppliers, ML models developers and validators.

**Separation of roles.** Platform concept assumes roles separation according to the defined list of operations, performed by an entity.

**Data and solutions reliability.** Big Data sets, machine learning code and final ML models are validated by companies and users with recognized reputation in domain expertise. Reliability problem will be solved by means of a complex motivation system, combined with non-repudiable authorship metadata stored on a distributed blockchain ledger. Incorrect validation service leads to a reputation loss outside of the NeuroSeed platform.

**Product delivery.** Platform charges customers for API calls to complex final ML models.

**Rewards system.** Suppliers are proportionally rewarded for any usage of values provided. Values can be represented by computing power services, data storage services, Big Data sets, ML models code, final ML models, validation services.

## 2.4 Platform roles

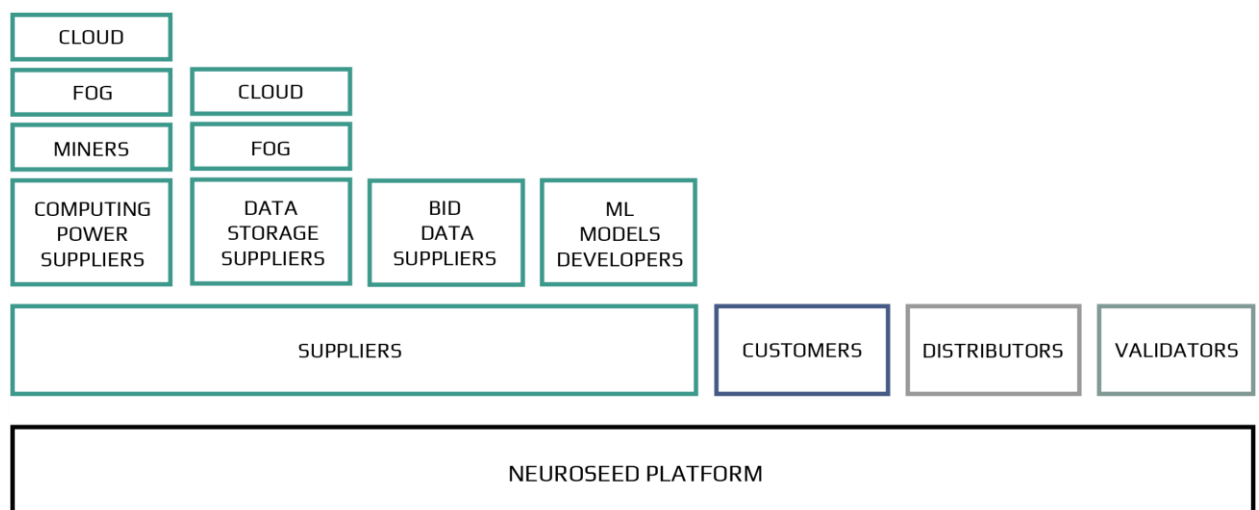


Figure 1. Platform roles

Platform actors are divided into roles. Actor's role prescribes the list of their allowed actions.





**Big Data Suppliers** upload datasets of Big Data that are used for ML models training.

**Data Storage Suppliers** connect their systems via the API and provide the services of storing data on their hardware. NeuroSeed platform can work with different types of data storage suppliers, such as centralized cloud storage providers and decentralized fog storage providers.

**Computing Power Suppliers** connect their systems via the API and provide the computational services on their hardware. NeuroSeed platform can work with different types of computing power suppliers, such as centralized cloud computing and decentralized fog computing power suppliers.

**ML Models Developers** are professional individual developers or companies that develop ML models and train them.

**Validators** are the establishments with significant proficiency in the domain field that perform quality, integrity and compliance assessments for the Big Data and ML models described in metadata specifications.

**Distributors** are wholesale resellers of API calls. Usually have a specific domain expertise and can participate in creation of uniform reusable clusters of merged final ML models.

**Customers** are any entities that buy API calls for NSD tokens in order to get ML services or embed in their own systems.



## 2.5 Platform roles and components interaction

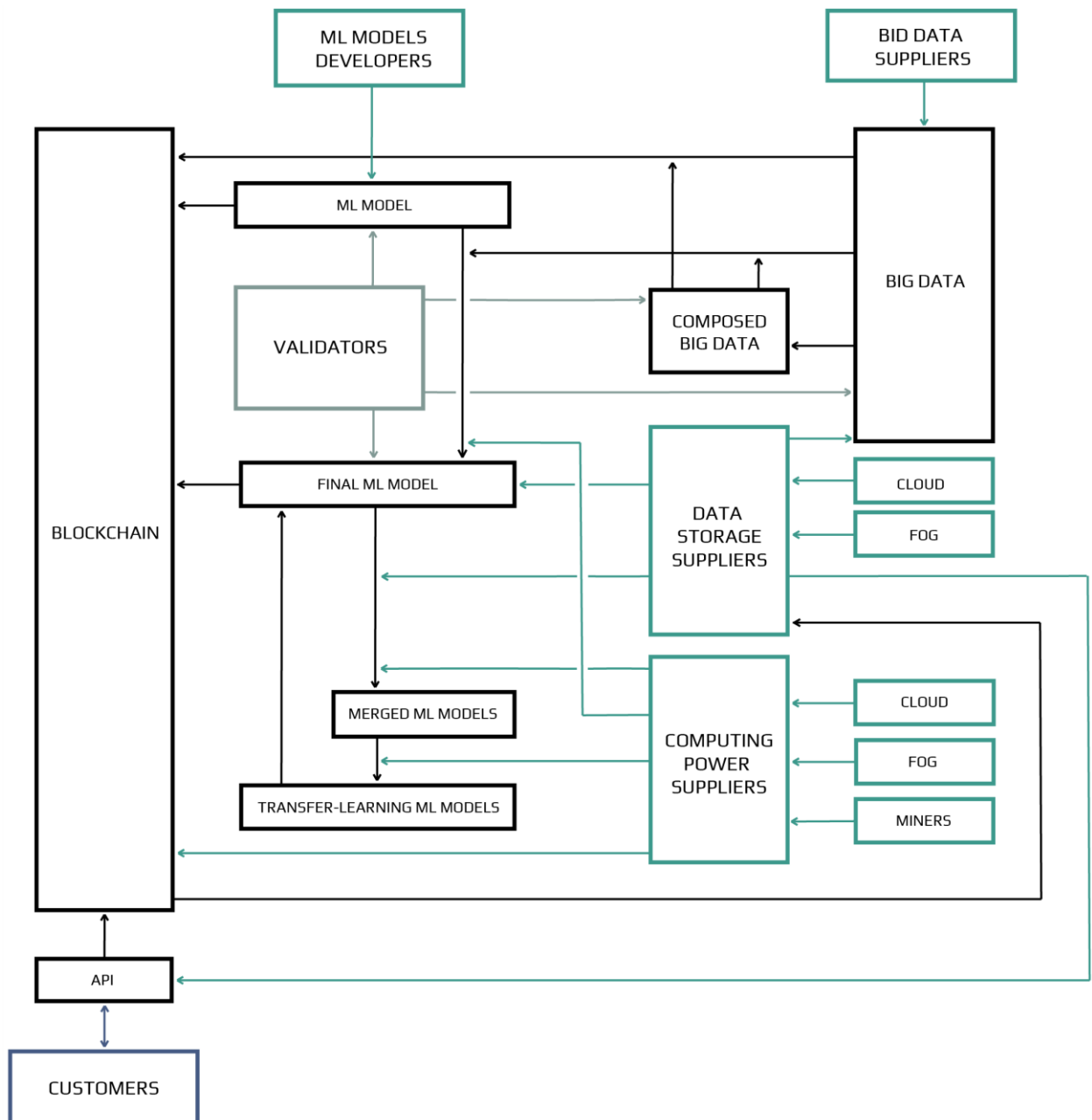


Figure 2. Platform roles and components interaction

Big Data suppliers upload datasets prepared for ML models training. New virtual composite datasets are created on top of existing ones.

The validators carry out the expertise of the data.



Data storage suppliers provide storage services for Big Data suppliers and ML model developers.

Computing power suppliers provide computing power required for ML model training.

ML model developers create and train ML model. The validators carry out the expertise of final ML models. ML model developers are able to use several models to merge and transfer learning processes in order to create the new one.

Distributors combine several final ML models to create clusters pack and sell API calls for customers.

Datasets, final ML models hash and metadata of all processing carried out by Big Data suppliers, ML model developers and Validators, are added to the blockchain.



## 3 TECHNOLOGY STACK OVERVIEW

### 3.1 Blockchain

Blockchain is used for storing the hashes of Big Datasets and final ML models together with their authorship metadata. The metadata is used by search algorithms when a Customer/Validator/Distributor request a dataset or a final ML model. The hash is used to identify and extract the target item from the distributed data storage.

### 3.2 Final ML models merging

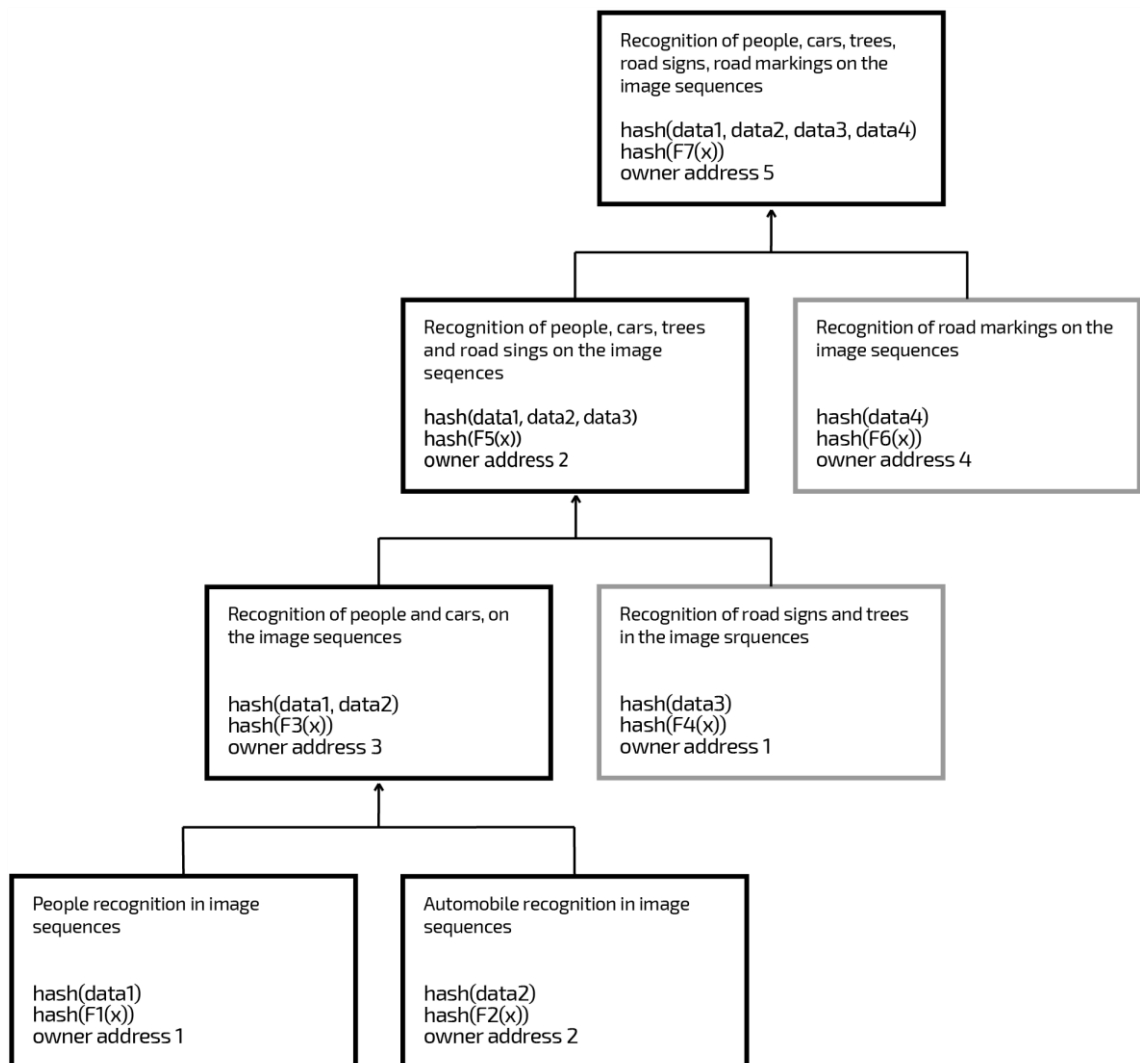


Figure 3. Neural Networks Merging



System automatically searches appropriate models that can be used through metadata in the blockchain. In order to maintain high level performance it is efficient to merge limited quantity of final ML models by each step. It creates inverted final ML models tree structure. Final ML models merging creates new model that include features of all parents models. All related tree nodes takes profit by child use in merging process.



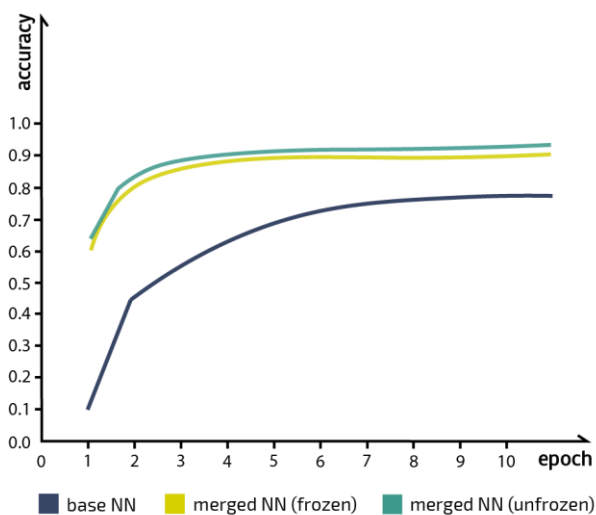
## 4 PROOF OF CONCEPT

Proof of Concept purpose is to prove economic benefits of final ML models merging. Neural network was chosen for several tests with text, audio and image datasets. The required time for training and merging the neural networks was compared. Neural networks merging allows to get a network with the necessary features 45-200 times faster than training a new neural network.

Visit our project page on [Github](#) to get full results description.

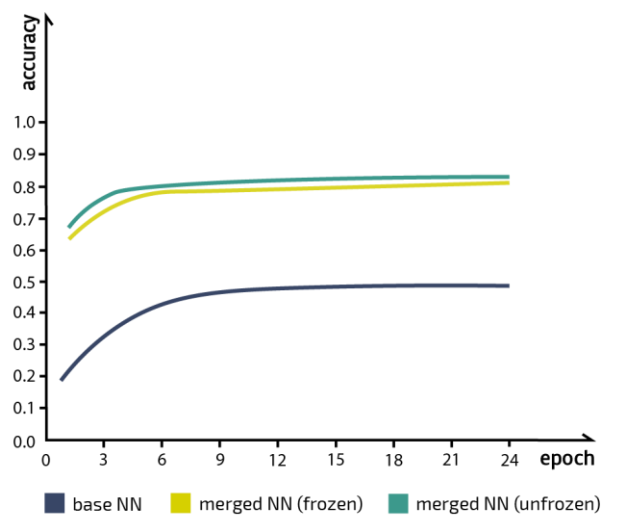
### 4.1 Testing image dataset

The greatest increase 200 times was obtained by training the neural network on image data (figure 4, 5).



Training neural networks on images from dataset cifar 10

Figure 4 - Image dataset learning rate



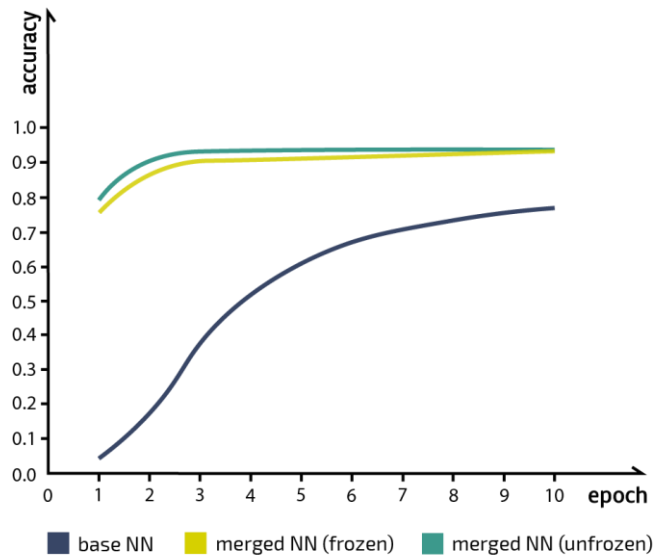
Training neural networks on images from dataset cifar 100

Figure 5 - Image dataset learning rate



## 4.2 Testing text dataset

It was possible to obtain an acceleration of 45 times on the text datasets (figure 6).

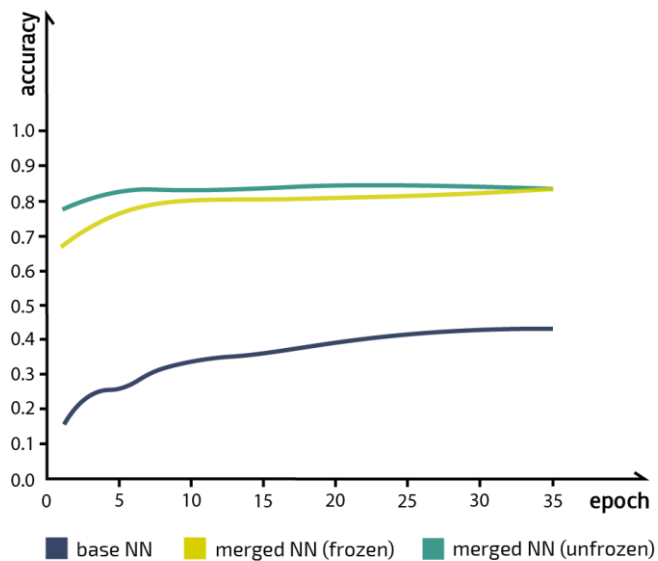


Training neural networks on text data

Figure 6 - Text dataset learning rate

## 4.3 Testing audio dataset

It was possible to obtain an acceleration of 150 times on the audio datasets (figure 7).



Training neural networks on audio data

Figure 7 - Audio dataset learning rate



# 5 INTERNAL ECONOMY

## 5.1 What is Token

Utility tokens represent unique opportunities to obtain all NeuroSeed features. NSD tokens are used to pay for access to datasets, final ML models, computing power and storage supply. NSD token cost is based on the platform usefulness and the combined fiat cost of Big Data and final ML models.

## 5.2 Internal micropayments and micro purchases

All internal micro purchases, such as final ML model, dataset or computing power are controlled by the platform. Micropayments technology are used to decrease additional expenses.

## 5.3 Payment distribution

ML training price includes dataset, computing power and final ML model price.

Income by using final ML models divides fairly between interconnected models due to accuracy, required computing power, related model count etc.

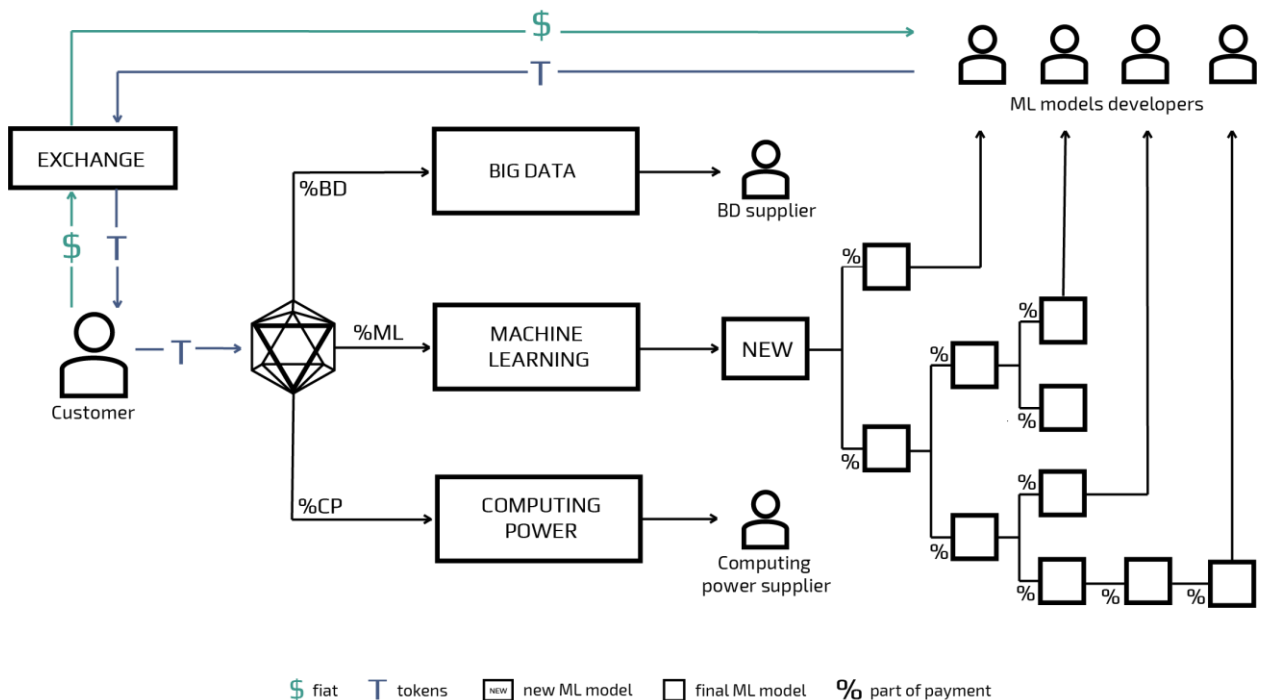


Figure 8 - Flowchart of cash flows





## 5.4 Payment for updates

All related final ML models can be updated as one of the models was updated. Each final ML model owner will be able to accept or decline updates. Updating price depends on model's accuracy that would be reached after update.

## 5.5 Motivation and validation

There is a special system of motivation for validators. The reputation and payment for verification within the system grows according to the quality of their work. Incorrect data validation affects the internal and external reputation of the validator.



## 6 BUSINESS MODEL

Revenue streams includes direct sales, service sales and asset sales.

### **Direct sales:**

- API calls to the NeuroSeed platform.

### **Assets sale:**

- Final machine learning models;
- Datasets or composed datasets.

### **Service sales:**

- Data Storage Supply;
- Computing Power Supply.



## 7 CHALLENGES

### 7.1 Universal classifier of final ML models

Universal signature analysis based classifier is designed specially for searching final ML models that have the same configuration. It requires information about ML model, layers and its configurations, verbal description of the problem being solved.

### 7.2 Motivation system economical model

The economic model describes a system of incentives as well as a system of checks and balances between all the roles on the platform:

- The formula for reward distribution for API calls between data providers, ML model developers and validators;
- Development of a complex motivation system combined with non-repudiable authorship metadata stored on a distributed blockchain ledger. Incorrect validation service leads to a reputation loss outside of the NeuroSeed platform.



## 7.3 Abstraction vs reusability

Reuse of final ML model is possible only when training ML model on generalized data that is not specified for a single task.

To get the most benefit, the correlation between the reuse of final ML model and the level of abstraction of the problem being solved should be optimal for a wide range of users.

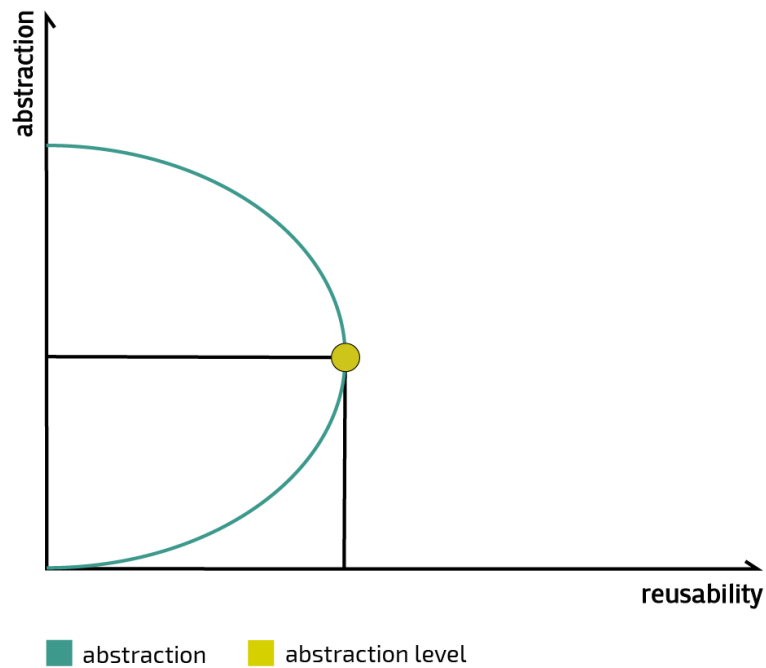


Figure 9 - Abstraction vs reusability diagram



# 8 ROADMAP

