

planck.ai  
interpretable AI, democratized on the blockchain

Fangda Li, Ellis Roberts

## Contents

<b>1</b>	<b>Mission Statement</b>	<b>2</b>
<b>2</b>	<b>Background</b>	<b>2</b>
<b>3</b>	<b>Applications</b>	<b>3</b>
3.1	Employment in the Post-AI Economy . . . . .	3
3.2	An Interface to Empower the Citizen Data Scientist . . . . .	4
3.3	Auditing-As-A-Service on the Blockchain . . . . .	4
3.4	Liquidity for Data-As-A-Commodity . . . . .	4
3.5	Counterparty Risk Elimination . . . . .	4
<b>4</b>	<b>Technologies</b>	<b>4</b>
4.1	Interpretable Deep Learning . . . . .	4
4.2	Interactive Semantic Segmentation . . . . .	5
<b>5</b>	<b>Alpha Release</b>	<b>5</b>
<b>6</b>	<b>Crowdsale</b>	<b>5</b>
<b>7</b>	<b>Roadmap</b>	<b>6</b>
<b>8</b>	<b>Legal</b>	<b>6</b>

9	References	6
10	Why This Name?	6

## 1 Mission Statement

As more facets of everyday life become governed by AI, it is incumbent upon consumers, engineers, and researchers to insist that such systems be capable of not only making correct decisions, but also *justifying* them.

The automated institutions of the future will wield the social and political power currently held by a handful of humans who are, at least in principle in Western democracies, ultimately accountable to the citizens or shareholders they govern. Decentralized transparency into the inner workings of these institutions is necessary for the continuation of the accountability and avenues of recourse that underlie equitable, just, and prosperous societies.

Planck.ai empowers its democratic user base to enforce human-interpretable justifications on prospective machine learning models. Our decentralized deep learning visual collaboration platform builds the necessary trust with key stakeholders to accelerate AI adoption in critical-use cases across industry verticals.

## 2 Background

In the past decade, advances in deep learning have shattered machine learning benchmarks firmly out of reach of previous state-of-the-art algorithms, in diverse fields ranging from image recognition to natural language processing. The rapid pace of innovation in these areas suggests that the greatest barriers to their commercial adoption will soon not only be technical, but also sociological.

Beginning in their first math class, schoolchildren are indoctrinated to *show their work*. The pedagogical benefits are obvious - it is difficult to imagine how one could learn anything beyond rudimentary arithmetic if given only a set of question-and-answer pairs. Partial credit incentivizes partial progress, because the final answer to a particular question is secondary to the correct application of the underlying principles. It is not enough to memorize the textbook example questions; one must be able to generalize to new, unseen examples. It is not enough to know how wrong one's final answer is; one must also know *where* the error was made.

In practice, machine learning can be simplistically defined as function approximation via optimization of a specific loss function over a dataset.

Do I understand and accept the assumptions and rationale behind your predictions? Where should I expect your knowledge to be strongest and weakest? Why should I trust you? These are all reasonable questions a human being is capable of answering to varying degrees of satisfaction, yet the very complexity underlying the power of deep neural networks makes generating these justifications nontrivial.

In parallel, the past decade has also upended paradigms of electronic payment via the proliferation of cryptocurrencies - digital assets that leverage cryptographic protocols to secure a distributed public ledger of transactions - the blockchain. Bitcoin (BTC), the market capitalization leader as of 2017 (~70B USD), pioneered the concept of storing digital assets on a network of miners - computational nodes that preserve the integrity of the system by verifying existing transactions and broadcasting new transactions into the network. Miners are incentivized by BTC block rewards, distributed approximately proportionally to the computational work done.

In 2014, Ethereum (ETH) unified a fractured ecosystem of distributed applications by introducing *smart contract* functionality to the blockchain, allowing parties to enter into binding contracts whose internal logic is visible and verifiable. Planck (PLK) builds on top of the Ethereum platform by providing a UI on which individual contributors are immediately compensated for improving the state-of-the-art deep neural networks by gathering and labeling data from the wild, and auditing model predictions.

## 3 Applications

### 3.1 Employment in the Post-AI Economy

One of the central tenets of deep learning is end-to-end feature extraction, but there is no intrinsic guarantee that these features learned on a training set *in the lab* reflect those required for business objectives *in the wild*. Symptoms of this misalignment include catastrophic predictions on small perturbations of image inputs (adversarial example ref) and

Whereas the industrial revolution created new wealth by leveraging advancements in mechanical and chemical engineering to multiply the productivity of manual labor, the AI revolution creates new wealth by leveraging advancements in data

ubiquity, machine learning algorithms, and parallelized hardware to multiply the productivity of *intellectual labor*.

### **3.2 An Interface to Empower the Citizen Data Scientist**

### **3.3 Auditing-As-A-Service on the Blockchain**

### **3.4 Liquidity for Data-As-A-Commodity**

### **3.5 Counterparty Risk Elimination**

## **4 Technologies**

### **4.1 Interpretable Deep Learning**

The greater the stakes of a decision, the greater the importance of the rationale behind it. The rapid pace of progress on various benchmarks in the machine learning research community suggests, Black-box models eventually reach a sociological barrier to adoption rather than a technical one. Studies in this area (refs here) unsurprisingly show that systems providing justifications behind correct decisions to human users are more likely to be trusted to those that are only simply correct. Take for example image segmentation models, which must balance two intertwined criteria: *class discrimination* and *accuracy*.

In 2016, the General Data Protection Regulation was passed by the European Parliament, one of the legislative bodies of the European Union. The act reaffirmed the directives of its predecessor, the Data Protection Directive, by specifying:

1. It is well known that popular convolutional neural network architectures do not generalize well on images far from their training distributions. They are almost universally highly susceptible to adversarial attacks - even small, pathological perturbations from canonical inputs can lead to wildly divergent, high-confidence predictions.

Insight from the network's human-interpretable justification of erroneous predictions can allow non-experts to meaningfully contribute to

2. Whereas a core value proposition of blockchain technology is its trust-free nature,

1. Right to Explanation: usual regression and classification algorithms are fundamentally function approximators that minimize prediction error on a unseen test set, with no intrinsic constraint on causality.

## 4.2 Interactive Semantic Segmentation

# 5 Alpha Release

The initial release of the Planck platform is a proof-of-concept demonstrating several core features, allowing users to:

- visualize salient regions in image classification, image captioning, and visual question answering (VQA) tasks.
- update the weights of convolutional neural networks pre-trained on ImageNet, after contributing new data & labels or correcting incorrect predictions on existing data.
- share datasets, neural network architectures, & corresponding weights with other contributors
- identify systematic biases in models by inspecting a visualization of the topological structure of a low-dimensional (i.e. t-SNE) embedding of its performance on a test dataset.

# 6 Crowdsale

Planck (PLK) is a utility token that incentivizes contributors on the Planck platform to gather data, provide labels, audit predictions, and improve models. A PLK buy-back structure allows users to seamlessly reserve GPU infrastructure for interactive neural network training and inference.

The immutable total supply of 1,000,000,000 PLK will be distributed in the following manner:

- 10% pre-sale on [[date1]]
- 20% stage 1 distribution on [[date2]]

- 20% stage 2 distribution on [[date3]]
- 15% founding team alignment
- 35% administration, composed of:
  - 15% project seed incentivization
  - 15% employee compensation
  - 5% contingency reserve

## 7 Roadmap

## 8 Legal

In *Securities and Exchange Commission v. W. J. Howey Co.*, 328 U.S. 293 (1946), Justice Murphy of Supreme Court of the United States formulated the so-called *Howey test* to determine whether a particular contract qualifies as an "investment contract" under the *Securities Act of 1933* 15 U.S.C. : 77b. PLK tokens do *not* qualify as such because they are utility tokens that allow

## 9 References

## 10 Why This Name?

Planck.ai quantizes "making the world a better place" into a series of singular, indivisible contributions.