planck.ai

interpretable AI, democratized on the blockchain

Fangda Li, Ellis Roberts

Contents

1 Mission Statement

As more facets of everyday life become governed by AI, it is incumbent upon both end users and machine learning engineers to insist that such systems be capable of not only making correct autonomous decisions, but also *justifying* them. Planck is a decentralized deep learning visual collaboration platform that empowers a democratic user base to enforce appropriate, human-interpretable justifications on prospective machine learning models. In doing so, we build the necessary trust to accelerate their adoption in critical-use cases across industry verticals.

2 Background

In the past decade, deep learning has shattered benchmarks firmly out of reach to previous machine learning methods in fields as diverse as 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.

Do I understand 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 these types of justifications highly nontrivial.

In parallel with machine learning advances, 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 ($\sim$70B USD), pioneered the concept of storing digital assets on a network of miners - computational nodes that preserve the integrity of the ledger by verifying existing and broadcasting new transactions into the network. Miners are incentivized by BTC block rewards, distributed approximately according 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 between them.

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.

3.2 An Interface to Empower the Citizen Data Scientist

3.3 Auditing-As-A-Service on the Blockchain

One application of crowdsourcing that has really taken off in some segments of the machine learning community is the use of crowdsourcing to evaluate learned models. This is especially useful for unsupervised models for which there is no objective notion of ground truth. As an example, consider topic models. A topic model discovers thematic topics from a set of documents, for instance, New York Times articles from the past year. In this context, a topic is a distribution over words in a vocabulary. Every word in the vocabulary occurs in every topic, but with a different probability or weight. For example, a topic model might learn a food topic that places high weight on cheese, kale, and bread, or a politics topic that places high weight on election, senate, and bill. Topic models are often used for data exploration and summarization. In order to be useful in these contexts, the learned

model should be human-interpretable in the sense that the topics it discovers should make sense to people. However, human interpretability is hard to quantify, and as a result, topic models are often evaluated based on other criteria such as predictive power.

3.4 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 that provide justifications behind correct decisions to human users are more likely to be trusted. 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. 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 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 buyback structure allows users to seamlessly reserve GPU cloud 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

6 Roadmap

7 Legal

8 References

9 Alpha Release

The alpha release demonstrates the

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,

10 Why This Name?

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