Integration of OpenMined and Zero-Knowledge Proofs into Ocean Protocol

Version 1.00

March 11, 2021

Contents

1 Introduction

Zoracles is announcing the partnership with Ocean Protocol to collaborate on integration of Machine Learning and Zero-Knowledge Proofs (ZKPs) for the Ocean Ecosystem. Ocean Protocol is a decentralized data exchange protocol that lets people share and monetize data while guaranteeing control, auditability, transparency, and compliance to all actors involved. It supports DataTokens and Compute-To-Data which enables a robust marketplace for data. Their ecosystem is enabling a Web3 economy vital for democratizing the monetization of data.

Data is the new currency.

Humans are considered to be the most valuable resource on Earth. It is not directly related to our ability to influence an economy, but rather due to a result of the data we generate in the information age. Data is the new oil and humans generate a tremendous amount of data every day across the globe. Monetizing data is allowing corporations such as Google to amass a fortune in terms of big data.

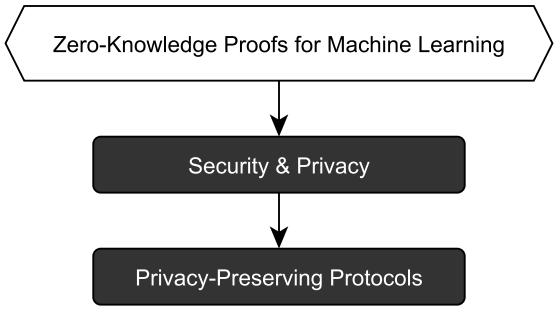
Data can be public data, data generated and owned by the organization developing the application, or private data acquired by 3rd parties. Public data is not an issue. When data that should be kept private gets in the wrong hands, adversity can ensue. A data breach at a government agency, for example, can put confidential information in the hands of an enemy state or malicious actor. A breach at a corporation can put proprietary data in the hands of a competitor.

Thus data privacy is too important.

Privacy is a concern not only related to Artificial Intelligence (AI) but any data-related field in general. It is about liberty - people having control over their personal data and the decisions taken based with such data.

Introducing OpenMined, an open-source community focused on building technology to facilitate the decentralized ownership of data and intelligence by providing a peer-to-peer network which allows any new company or person to train their AI models on user data, without the actual owner of data losing control over it. OpenMined aims to provide data sets which are very similar to those owned by big data companies and to provide compensation for the actual owners of that data. This will help reduce the monopoly of large companies by reducing the barrier of entry for newer players who wants access to large data sets, and to ensure the financial gains are equally distributed.

2 Machine Learning with ZKPs

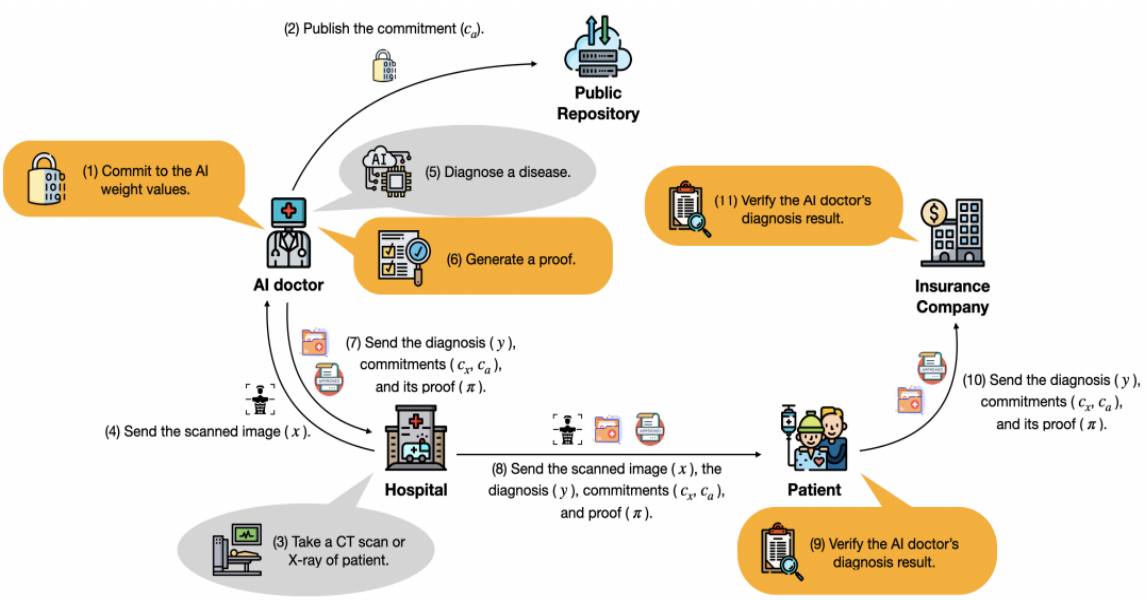


Let's consider a clinical decision support service application via AI. In this application, a hospital takes a patient's CT scan or X-ray, and sends the scanned image to the AI doctor. Then the AI doctor diagnoses the disease based on the image and returns the diagnosis result to the hospital and the patient. The integrity check of the AI results is required since incorrect results may endanger the life of the patient.

The most straightforward approach to verify the result is to re-execute the same AI program. However, it is impossible in most cases since the AI weight parameters are important Ips and are not available publicly. In addition, the privacy of input data is another issue to consider. In our scenario, we allow the AI doctor to know the user's input for diagnosis but it may be desirable to hide the user's private information when the diagnosis result is transferred to the third party - such as an insurance company. In this situation it should be possible for the insurance company to verify that the diagnosis result is correct without the private information of input data as well as AI weights.

Fortunately, the advanced cryptographic tool called zero-knowledge succinct non-interactive arguments of knowledge (zk-SNARKs) can solve the problem to verify the correctness of results without revealing private information.

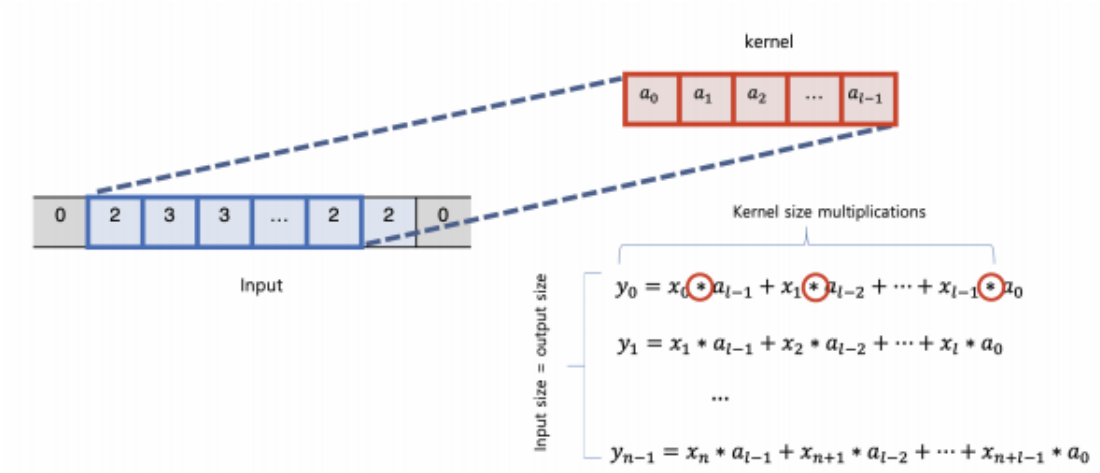
In zk-SNARKs, a prover generates a proof π using public input/output data (or statement ϕ) and secret input data (or witness ω) for a given function. A verifier can check the validity of the statement ϕ with the proof π without the secret input data ω. zk-SNARKs can also be used to protect the privacy of user's input data from the verifier when used together with a commitment scheme. The commitment scheme is a cryptographic primitive that allows one to commit to his choice while keeping it hidden to others (hiding) so that he can no longer change his choice (binding). Since the zk-SNARKs proof can include the correct computation of the commitment scheme, it can be verified with proof and commitment alone that the result is computed correctly.



The AI doctor first generates a commitment c_a by committing to AI weight values a and publishes the commitment c_a in a public repository so that the correctness computation of the model can be transparently checked against the published c_a. The AI doctor computes a diagnosis result, y, on received input data x from the hospital and AI weights a. In addition, the AI doctor makes a commitment c_x from the input data x to hide the input data, and generates a proof π for a statement including the input commitment, the weight commitment and the output result $(\phi = (c_x, c_a, y))$ with a witness comprising $(\omega = (x, a))$. The proof π and the statement ϕ is provided to the hospital and the patient. Then they can check the correctness of the statement ϕ with the proof π. Moreover, the hospital and the patient can transfer the statement ϕ to any third party like an insurance company. The insurance company can also check the statement ϕ with the proof π without input data x and weight values a.

The zk-SNARKs require significant amounts of computations on the prover's side. In zk-SNARKs, a function is translated to an arithmetic circuit comprising addition and multiplication gates to be represented as quadratic arithmetic programs (QAPs). The proving time is proportional to the number of multiplications in QAPs. In addition, the size of public parameters called common reference string (CRS) linearly increases according to the number of multiplications.

Thus, the main hurdle to apply zk-SNARKs to real applications is how to minimize the proving time.

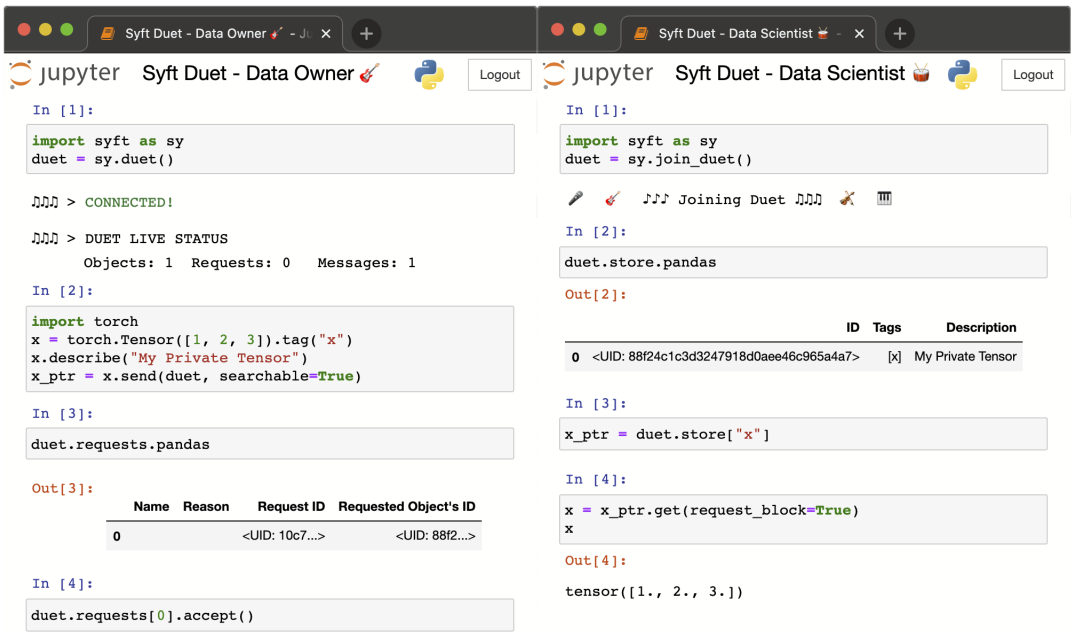


3 Bridging OpenMined to Ocean Protocol

The next challenging part is building Bridge to Ocean Protocol. Fortunately, OpenMined provides a Python library (PySyft) for computing on data. PySyft might be really helpful for us to connect to Ocean's data marketplace.

`https://github.com/OpenMined/PySyft/tree/master/examples/duet/`

`https://github.com/OpenMined/PySyft`



4 Conclusion

Zoracles partners with DeFi projects using zero-knowledge proofs to provide confidential data to smart contracts. Our product lines includes confidential credit, price feeds, and a "Snarks-As-A-Service" governance platform.

As we have the same vision with Ocean, our goal is to combine OpenMined and ZKPs on the top of Ocean Protocol.

Overall, we have devised a clear roadmap for an end-to-end solution for cutting releases.