

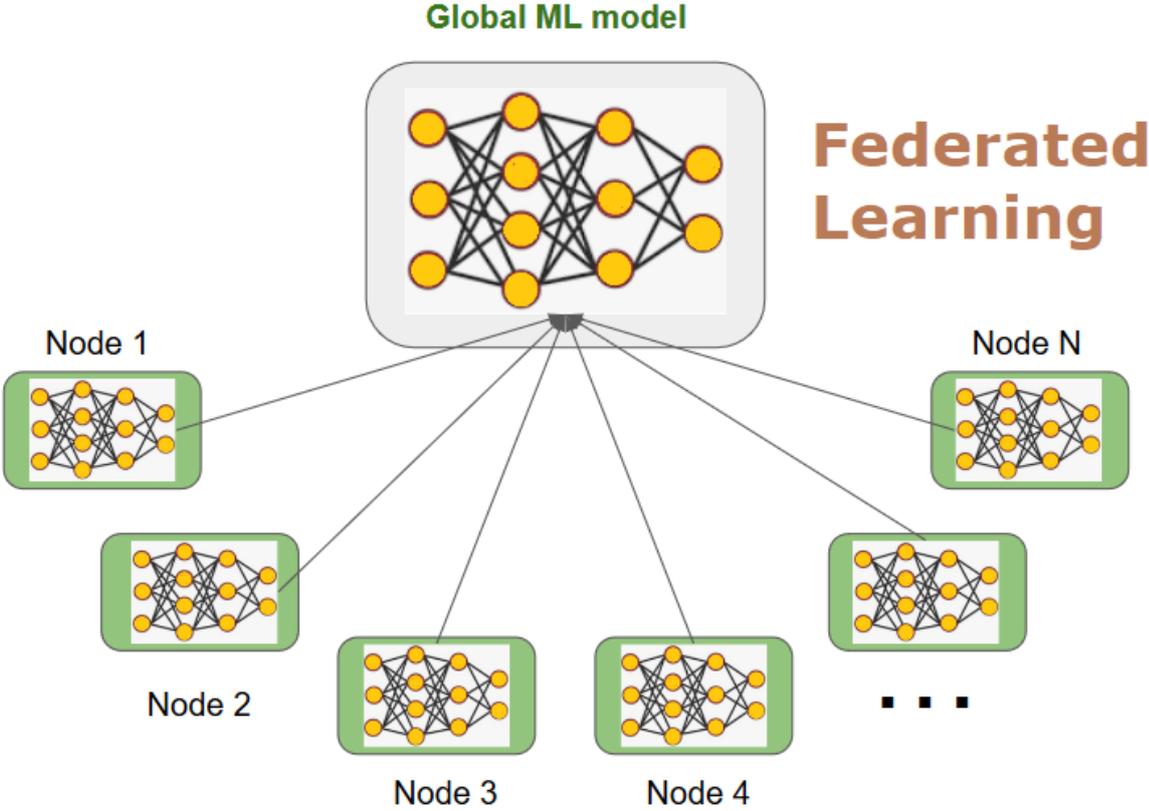
Pharma Companies Join Forces to Train AI for Drug Discovery Using Blockchain

June 5, 2019 by Andrii Buvailo

The newly organized research project “MELLODDY” (**M**achine **L**earning **L**edger **O**rchestration for **D**rug **D**iscovery), involving ten large pharma companies and seven technology providers, is that kind of deals which can catalyze a transition of the pharmaceutical industry to a new level -- a “paradigm shift”, as one might refer to it in terms of Thomas Kuhn’s “The Structure of Scientific Revolutions”.

The project aims at developing a state-of-the-art platform for collaboration, based on Owkin’s blockchain architecture technology, which would allow collective training of artificial intelligence (AI) algorithms using data from multiple direct pharmaceutical competitors, without exposing their internal know-hows and compromising their intellectual property -- for the collective benefit of everyone involved.

While artificial intelligence (AI) already proved to be a groundbreaking thing in many industries (robotics, finance, surveillance, cyber security, self-driving cars to name just a few), drug discovery still seems like a hard case for machine learning practitioners. A major reason for that is the lack of quality data to train models properly.



It might seem surprising, as pharmaceutical research generates enormous amounts of data daily. But when you consider a degree of secrecy and protectionism that competing pharmaceutical giants put on their research, it becomes clear that majority of data is actually not available for training, it is dispersed across hundreds of organizations, hidden behind their firewalls. Decades of screening, testing and validation research -- combined data from largest pharma companies might well be enough to make AI really smart at predicting next drug candidates. But how to deal with secrecy and competition, how to share data for the collective benefit without compromising own know how?

Two fundamental technologies that make it possible are “federated learning” and blockchain.

Federated learning, a new collaborative form of machine learning introduced in 2017 by Google AI, is a form of model training where the training process is distributed among many users. Instead of gathering all data from all users in one centralized location to train models, federated learning trains AI models on local devices in large batches, then transfers those learnings back to a global model without the need for data to leave any particular device.

Blockchain is a growing immutable list of records of transactional data, a sort of distributed ledger for maintaining a permanent and tamper-proof information. It can serve as a decentralized database

managed by computers belonging to a peer-to-peer (P2P) network. Each of the computers in the network stores a copy of the ledger to prevent a single point of failure (SPOF). All copies are updated and validated simultaneously.

The MELLODDY project aims, for the first time, to use machine learning methods on the aggregated chemical libraries data of ten pharmaceutical companies to create more accurate models of predicting next promising hit molecules. The platform uses Amazon Web Services as an infrastructure to execute machine learning algorithms from academic partners on a large scale. A central server allows each pharma company to share a common model to be consolidated collectively. Each partner trains model locally on their own data, and then shares only the updated weights of the model (without the actual dataset) with the central server, so that the whole model is updated altogether. In this way, the aggregate model eventually contains data features peculiar to each partner's dataset -- it learns in a federated way.

In the same time, the Owkin's blockchain architecture technology allows the whole platform to keep perfect transparency and control over each transaction by all partners, while preventing any leakage of proprietary datasets. "The MELLODDY consortium will use Owkin's blockchain architecture technology to extract insight from multiple datasets without having to first pool the data," said Mathieu Galtier, Project Coordinator, Owkin.

The MELLODDY platform allows for creating a more generalizable AI engine, boosting its predictive power and expanding applicability domain of the models by leveraging all available data.

The consortium of the project is funded from the Innovative Medicines Initiative (IMI) as a public-private partnership and comprises of the following partners:

Pharmaceutical companies: **Amgen, Astellas, AstraZeneca, Bayer, Boehringer Ingelheim, GSK, Janssen Pharmaceutica NV, Merck KgaA, Novartis, and Institut de Recherches Servier**

Academic institutions: **KU Leuven, Budapesti Muszaki es Gazdasagtudományi Egyetem**

Subject matter experts: **Owkin, Substra Foundation, Loodse, Iktos**

Machine learning chipset manufacturer: **NVIDIA**

It seems that MELLODDY might become a sort of illustrative example for the pharmaceutical industry as to how a mutually beneficial collaboration might look like in the era of artificial intelligence and blockchain.