Sponsored article

Distributed AI – A new way of industry cooperation in receivables finance

Artificial intelligence and blockchain technology will have a profound impact on many industries, and receivables finance is no exception. AI and blockchain each address a fundamental problem: how to make high quality data-driven predictions in a scalable way and how to build trust within an ecosystem respectively.

Igor Zaks, CFA
President and CEO
TenzorAI

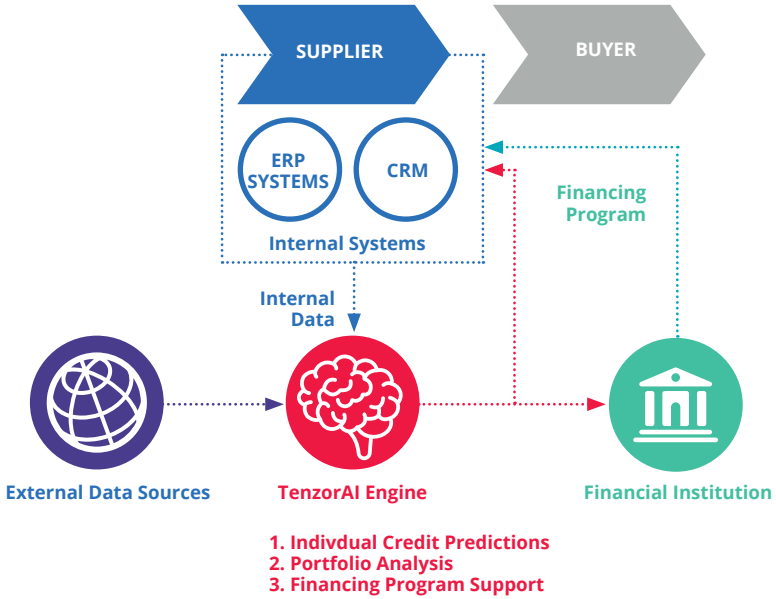
Alexei Lapouchnian, PhD
CTO
TenzorAI

Vlad Skorokhod, PhD
Chief Data Scientist
TenzorAI

In supply chain ecosystems, large companies and invoice financiers possess significant amounts of data, which can be used to inform and improve the assessment of both the buyers' credit and performance risks. However, up till now, most of the information available in supply chain ecosystems has not been used in a regular, data-driven way to assess receivables risk. Instead the companies rely on a combination of financial information (analysed typically in spreadsheet format) and ad-hoc analyses by a credit analyst/manager.

TenzorAI is leveraging AI and blockchain technology to build credit models that provide enhanced insight into individual and portfolio credit and performance risk. Utilising AI tools on a combination of internal data (such as from ERP and CRM systems) and external information (such as financials, third-party reports, etc.) can lead to major improvements in credit decision making, compared to simply relying on external financial information, which has traditionally been the case in the industry. (e.g., see Figure 1).

Figure 1. A typical single-client deployment of a TenzorAI engine in a supply chain



However, while a supplier knows a lot about their buyers (particularly in highly integrated value chains), they do not know everything, and neither do their financial institutions (FIs). Generally, the buyer wants suppliers and financiers to know about their credit risk (as in the absence of such knowledge, assumptions made by FIs/suppliers can lead to less favourable financing terms for the buyers), but they are very sensitive about exposing commercial information (e.g., the information on the end/next tier customers that the supplier can sell to directly or through competitors, the disclosure of margins that can trigger price increases for the buyer, etc.) and about legal issues (GDPR, client confidentiality, etc.).

Technically, this is a classical trust problem. It may be resolved through a trusted third-party (for example, a third-party auditor). However, doing this in real time is very expensive and impractical. This is where blockchain technology, designed to create trust without a central counterparty, may play a critical role.

Multiple applications of blockchain within supply chains (by finance-based consortia such as Voltron* or TradeIX* or logistics-based ones like TradeLens)

are built on the premise that all supply chain transactions are moved to the blockchain. This approach enhances the ability to verify the flow of goods and money between parties and allows well-defined rule-based decisions through smart contracts to be made, which are programmes executed on blockchains when certain triggering conditions are met. However, due to what is commonly known as the network effect (whereby the value of a good/service is growing as more users use it, until after a critical mass of supply chain participants have joined these consortia and a reasonably large number of their transactions are accumulated on these blockchains) the efforts do not enable high-quality data-driven predictions. This is unlikely to happen in the immediate future.

At the same time, there are massive amounts of operational data available to each supplier and buyer within supply chains that can provide immediate input to AI-driven predictive models.

This information exists within the internal systems of supply chain participants, such as ERP, CRM, and other enterprise planning tools. The best way to analyse the risk profiles of supply chain partners is to directly access such data. This, however, requires a robust data governance framework, that needs to perform three key functions:

- Clear and verifiable rules and records on which information shall be provided to whom. This is similar to the notion of the 'clean room' in a M&A transaction. It is likely that supply chain participants are happy to share model outcomes, some summary data, but not individual customer/transaction data.

- Audit trail. The system needs to record the source of data and the timestamp for when the data was inputted into the model. This is essential for any future disputes, etc.

- Model cross training. In AI, more data means better models, which is beneficial for the whole ecosystem. There are technologies to do so without sharing the underlying data. This is a concept of distributed AI.

We will now discuss the above functions in more detail.

Blockchain-based access control:

Blockchains are distributed ledgers used for storing information/transactions about a business domain. Of interest to enterprises are mostly permissioned blockchains, where a central authority or existing participants are able to vet any new member, thus increasing security and stability of the chain. In blockchains, each block of records is linked to all the blocks before and after it. Information can be stored either on blockchain or off-chain (e.g., in a traditional database or an enterprise system), with only its hash – a digital fingerprint – residing on the chain. A combination of on-chain and off-chain storage is typically used in modern enterprise blockchains. Records on blockchains are cryptographically encrypted, with blockchain participants assigning their own private keys to their transactions to serve as digital signatures. While information is distributed across the blockchain network – i.e., each participant typically has a copy of the complete blockchain – due to the encryption, the information is secure. Only those parties given the right to access the information are provided with the keys to decrypt it. This assignment of access rights can be made by executing smart contracts, with the owner of information (e.g., a buyer) able to grant/revoke any other participants' (e.g., a supplier or FI) access. For information residing off-chain, having access to it may include the right to execute a specific query (which is itself recorded on the chain) on the information owner's system (ERP, accounting system, etc.) to retrieve the data.

Blockchain-based audit:

A number of blockchain features make it an excellent tool for maintaining audit trails of any kind. Blockchain is append-only and thus immutable. An attempt to change information previously recorded on blockchain will be immediately detected. For off-chain data, the fact that it has not been retroactively changed can be verified by reading it and comparing its newly generated hash with the one originally stored on the blockchain. The same hashes imply no data tampering. Moreover, all blockchain additions are timestamped. For instance, in many systems, the relevant parties are able to see which data was made available to the AI engines and other authorised parties and when, the predictions produced by the AI tools based on that data, etc. – all in a transparent and secure way.

Model cross training:

"Distributed AI" commonly refers to learning predictive models from data sets located remotely from the central server. In many use cases, scalability is the key motivation of distributed machine learning as it allows avoiding transmission of large data volumes into a central location. Additionally, it is useful in situations where raw data sharing is restricted due to data privacy concerns.

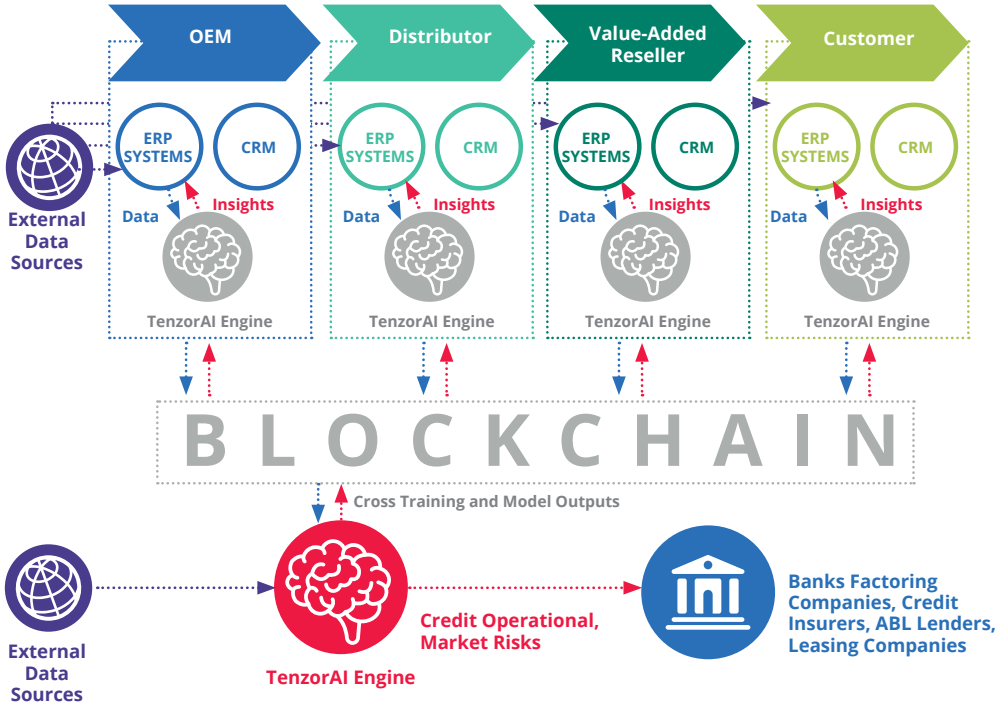
The fundamental mathematical principles and general strategy for algorithm design are based on the concept of sufficient statistics, i.e., a set of parameters learned on a small data sample to describe the entire population. This concept can be generalised such that a predictive model learned on a remote data set can be combined with similar models learned on similar data sets into a master model. Thus, only a limited set of parameters representing the model, for example; the coefficients and RSS for linear regression models, the equation for decision boundary for SVM models, split rules for tree-based models, centroids and within-cluster error for K-means clustering etc., needs to be transmitted to the central server where they are aggregated into a master model via ensemble learning.

The past few years witnessed significant growth in practical applications of distributed AI. While most of these applications are motivated by the developments in IoT and distributed sensor networks, their key principles are applicable to enterprise data repositories as "nodes" in a broad industrial network. In both cases, distributed learning becomes an attractive alternative to centralised model learning which requires costly or otherwise undesirable transmission of large data volumes.

One of the key technical challenges of remote learning is that our understanding of the remote data model may be inaccurate, which may potentially invalidate the findings of the predictive model or simply produce too many errors to be useful. This needs to be taken into account in model design to ensure the robustness of the distributed model. With all these challenges in mind, distributed AI is an efficient method of creating a rich knowledge base under severe data access restrictions.

A TenzorAI distributed AI blockchain-based architecture is illustrated in Figure 2.

Figure 2. Blockchain-based multi-client deployment.



When assessing risk in supply chains based on internal participant data, given the commercial and legal concerns related to data sharing, we are unable to combine data from many supply chain participants into one large dataset to train predictive models on. Instead, we use distributed AI as a way to achieve predictions that are as close to that ideal as possible while respecting the privacy and legal concerns described above.

The decentralised, distributed AI-based design has the advantage of supporting the pre-existing privacy/legal constraints in supply chains and not requiring the fundamental changes in how supply chains operate (e.g., by requiring all relevant transactions to be stored on a blockchain), whilst taking advantage of both the recent technological advances and the very large data sets available from multiple supply chain players.

** Voltron and TradeIX will present at **Consortia 2019**.*
www.consortia2019.com