CHAPTER 16

The Art of the Newly Possible: Transforming Health with Emerging Technology and Federated Learning

This chapter is derived from a presentation, “The Art of the Newly Possible: How 5 Emerging Technologies Will Transform Health,” delivered by the authors (Heather Flannery, Jonathon Passerat-Palmbah PhD, and Sean T. Manion, and edited by Vikram Dhillon) at the online ConsenSys Health COVID-19 Veterans Health Summit on August 26, 2020.

Introduction

Not since the dawn of the commercial internet have we experienced such a viable convergence of disruptive and transforming technologies as we do now in the world of healthcare. Our earliest attempts to revolutionize healthcare technology, unfortunately, were not fully inclusive, leaving providers, patients, and payers in an increasingly dire situation. We ended up with a massive digital divide with a worsening of disparities, a crisis in physician suicide due to the dehumanization of the art of medicine, and an array of other global problems. For the first time, we have innovations—with extraordinary positive potential for humanity—as significant as the obviously hypothetical,
simultaneous debut of the Gutenberg printing press, the steam engine, the integrated circuit, and the internet. Of course, as with any disruptive technology, extraordinary risks and reasons for circumspection and ethically intentional consideration are paramount as we approach these changes that can transform healthcare. The overture to a new era is playing, and the time is now.

This transformation of healthcare is an achievable reality through the implementation of three key areas of technology:

1. **Blockchain** – including optional *tokenization*
2. **Decentralized Artificial Intelligence (AI)** – primarily *machine learning*
3. **Privacy-in-Depth** – including *zero-knowledge cryptography* and *verifiable and confidential computing*

The largest healthcare enterprises in the world have been working on applied R&D, experimentation, planning, and business development regarding new models of blockchain technology in recent years. For example, the Synaptic Health Alliance was announced in April 2018 as a blockchain-focused healthcare industry consortium that includes industry leaders such as United Healthcare, Optum, Quest Diagnostics, Aetna, and other key stakeholders. The Synaptic Health Alliance has led the way not only in technology implementation but also in methods of governance, methods of capital formation, and methods of examining how different aspects of what we do in our industry can be removed from the basis of competition and mutualized.

### Blockchain and Tokenization

Blockchain is more than a technology. It is a technology that exists simultaneously at multiple layers of the engineering stack as well as an innovation in the arena of behavioral economics. It has created a new discipline: applied behavioral economics in real time. It is a brand new opportunity to express game-theory principles, form hypotheses, and test the way that different incentive designs inform stakeholder groups in healthcare along with every other industry it touches.
Structurally, it opens the door to new business models that have never been logistically possible before and, therefore, have never been meaningfully considered. And, it introduces radical new levels of personal identity or self-sovereignty, including the potential for credentials or verifications about who you are, your data, and data about you. All of these things are made possible by blockchain. Blockchain exists in the form of networks. Those networks can be public resources, public infrastructures, and neutral platforms like the Ethereum main net or the Bitcoin blockchain. The Ethereum main net can run software that can focus on any kind of application—not only digital money, but any sort of software application.

Compared to public blockchain networks, private blockchain networks have completely different characteristics, though some of the same benefits along with several trade-offs. Looking to the future of blockchain in the healthcare industry, it seems likely that our end state will be an array of hybrid networks—lots of private enterprise networks that are bridged, or tethered, through a single neutral public infrastructure.

Another defining characteristic of blockchain, particularly newer blockchain platforms like Ethereum and others, is the ability to execute and run smart contracts: cryptographically stored and cryptographically verified stored procedures expressed as microservices that run on a blockchain infrastructure. These smart contracts are key because they facilitate secure automation across organizational boundaries. For the last thirty years, most of the focus of enterprise computer engineering work has been looking inward into the four walls of any given enterprise. For the first time, this distributed smart contract infrastructure allows the use of the same kinds of principles for automation and efficiency gains at an ecosystem level, providing technology that can actually deliver the goods for the future of healthcare.

In addition, blockchain provides for digital scarcity from an economics and behavioral economics perspective. In our current legacy data environment, there is no scarcity because data can be infinitely copied. This endless copying continues to expand the disk space required for storage and creates governance difficulties. Through tokenization—either tokenization of single items like something with a serial number or tokenization of units that are replicable like a cryptocurrency, a certificate, or things with a value of any kind—scarcity can be created for actual transactions with established units of value.
One of the most appealing things about blockchain is that a user will not be aware that the technology is “running” in the background to exchange and manage data. Along with a human design–focused and streamlined user experience and interface, frontend Web 3.0 applications, and decentralized apps, a user would never know that the background is any different. Some of the things that a public blockchain delivers, particularly transparency and decentralization, are traded off when implemented as private blockchain networks. What is gained is extreme privacy.

New technologies facilitate off-chain computation and storage to enable aspects of both transparency and privacy in a common infrastructure. When selecting from hundreds of different architectural patterns, dozens of blockchain protocols—public, private, and hybrid—there is always a trade-off between three criteria: 1) scalability of the system; 2) its cybersecurity posture; and 3) the degree to which it is decentralized or centrally controlled. It is possible to correct for any two as architectural patterns are formed.

Blockchain technology is necessary but not sufficient for the healthcare industry. There are additional important areas where the unique requirements of health can be greatly enhanced with this technology, including compliance, the unique cybersecurity environment, the unique privacy environment, and the challenges we have with identity of patients, providers, and devices. All of this also requires a clear eye toward bioethics of health. So, more than blockchain is required for the desired outcome.

**Decentralized Artificial Intelligence (AI)**

Currently, machine learning, which makes up the bulk of the current artificial intelligence applications in healthcare, is accomplished by 1) putting a lot of data together in a central place to curate it; 2) running analytics on the data sets; and 3) eventually training extremely complicated machine learning models. As an example of when this is a reasonable proposition and scalable, consider data from your phone. Your phone data is pooled together with others’ data in a single place to learn, for example, how to better predict next words on the keyboard of your smartphone. Then, once a new predictive model is ready, this improved model is pushed back to your phone. Inherently, the assumption is made either that the data pulled together in a central data lake is not too sensitive or that you are okay with sharing your data with the central entity. Obviously, this is not the case with health data, so the centralized model does not work as well in healthcare. With the decentralized model, the difference is that instead of bringing the data to the model, the model is delivered to the data, so the data does not travel anymore—it remains at rest.
In the decentralized approach, we send models to each of the locations where the data lives. It can be a mobile phone or larger data lakes owned by institutions. For example, if data is owned by different clinical institutions, a unique learning model can be developed at each of these institutions; they will basically each train their own model. Then, the key to machine learning is to combine the knowledge that has been learned at each of these institutions (or sites or devices) into a single new joint model. So, in the future, you will not need to consider how to gather data into a central location for storage and access; you will instead deploy a machine learning model to the network and receive back derived data.

So, why is this important? For starters, if it’s my personal data we are considering, a lot can happen with my data once it’s copied and whisked off out of my control. If it’s at rest within my system, or within my phone, I have more control. That peace-of-mind aspect is important to the consumer and the general public when you consider the constant stream of data breaches that show no sign of abating. Further, by leaving the data where it is, you suppress copies and introduce digital scarcity. Combined with blockchain, that’s all you need to create a decentralized marketplace. If data threats could be significantly decreased with a decentralized system that is still able to deliver the value of advanced knowledge that can only come from the extensive data analytics of many data sets, health and healthcare as we know it will be transformed.

An additional benefit of combining blockchain with machine learning, even in centralized machine learning models, involves sourcing and training data to have the cryptographically verified provenance of every data element that is ultimately included in a training set. That can go a long way toward resolving explainability problems and other long-standing issues with being able to source training data, understand where it came from, and determine whether the data itself was corrupted. Decentralizing the learning function and decentralizing the sourcing of training data is low-hanging fruit and represents a huge incremental benefit when conducting learning on decentralized data. Therefore, the convergence of blockchain infrastructures with decentralized machine learning approaches is absolutely necessary.

This may be news, but the ability to conduct decentralized machine learning has existed for quite some time. However, it has yet to be embraced by the healthcare industry because it has lacked the appropriate incentives. By integrating blockchain with these decentralized AI tactics for training models, we are able to connect incentives and to know conclusively exactly what permissions are granted, by whom, to do exactly what, for how long, etc. The combination of these technologies will be transformative for advancing healthcare.
Privacy in Depth

“Any sufficiently advanced technology is indistinguishable from magic.”

—Arthur C. Clarke’s Third Law

How do we ensure privacy and security? It’s the question on everyone’s mind when we are considering health data. It turns out that there is a family of advanced cryptographic privacy-preserving techniques that are being utilized in new ways that will have a profound impact on personal health information and regulatory compliance in healthcare and life sciences. A prime example of this is zero-knowledge proofs (ZKPs). It would take multiple chapters of another book to go over ZKPs and these other privacy-in-depth techniques in detail. So, from a high level, here is a brief look at ZKPs and how they are substantially different from standard legacy cryptographic models or methods that are generally used today.

ZKP cryptography has been around for thirty years, but it is getting extra traction in recent years, especially as it relates to blockchain. What ZKPs do is allow you to prove that a statement is correct and true without revealing the statement that is at the heart of the confirmation. Its value is to prove that you know something that you want to keep secret, and to convince another party or the rest of the network that you do actually know the secret, but without actually revealing the secret.

As a real-world example, think about your birthdate as a data element and all the times you have to enter your birthdate in order to demonstrate your age, for whatever reason. In a world where ZKP cryptography was brought to global industrial scale, the data elements of your birthday would almost never need to be transferred. Instead, a cryptographic proof that your birthday was higher than or lower than or within the decade of the birthday requirement—yes or no—would be all that the other system would store. The month, day, and year of your birthday would remain at rest. That is just one simple example. If you imagine all the data that is involved in clinical, scientific, and business workflows in our industry, ZKP cryptography, along with the rest of the family of privacy-in-depth cryptography, has endless applications.

Why and how are these advanced cryptographic privacy-preserving techniques used in the context of this whole stack we are describing? First, they add more authenticity to the shared derived data. For example, imagine that a patient, Alice, must use a remote private inference service that screens patients prior to scheduling a consultation with a very busy specialist named Bob. Bob’s policy implies that the secure inference service returns a prediction with a probability greater than 80 percent in order to book an
appointment with him. When Alice gets the prediction back, it’s been encrypted such that she’s the only person who can decrypt it. Now Alice wants to present her diagnosis to Bob to request an appointment. How can Bob be certain that the prediction Alice is presenting has indeed been obtained from the secure inference prediction service? There is no way to do that unless Alice can provide a cryptographic proof that the prediction value actually came from the specified inference service. In ZKP terms, this means proving that the plaintext prediction sent by Alice to Bob corresponds to the encrypted content she received from the service. Not only does Alice send the prediction to Bob, but she also adds a ZKP of the correct decryption alongside the prediction number. Note that the whole process never revealed Alice’s input healthcare data in plaintext to the prediction service or to Bob.

Along with privacy-in-depth, there’s one other important aspect, and that’s the notion of secure inference. Inference is the function of a machine learning model making a prediction. After a model has been trained and is doing what it’s meant to do, it is theoretically making a prediction that might guide clinical decision making. It might guide patient behavior. It might guide access to care. It might guide anything where these technologies are implemented in our industry. It is critical that the resulting data can not be reverse engineered or inappropriately used to derive information about the data’s origin or source. Only the intended inference is produced by the model.

For those who wish to research these three technologies further, consider the following:

- **Blockchain**
  - Decentralized Apps (“DApps”) – *Web3 User Experience*
  - Optional Tokenization – *Creating Digital Scarcity*
  - Smart Contracts – *Secure Automation Across Organizational Boundaries*
  - Blockchain Networks – *Public, Private, and Hybrid*

- **Decentralized AI**
  - Federated learning in blockchain networks
  - New paradigms in training data provenance, incentivization, and secure inference
  - Intelligent agent-based simulation and automation
Source and Derived Health Data

Now that we've had a high-level discussion of these three game-changing technologies, let's consider a paradigm shift in value creation—moving copies of data is no longer the only way to create value from data.

In health, source data is enterprise data that lives behind firewalls, or the patient’s self-sovereign data, the volume of which has been exploding in scale. Source data in health has been expanding exponentially for the last fifteen years, and this will continue with the widespread availability of 5G wireless connectivity, the internet of medical things, and the growing understanding of social determinant data as health data. Currently, most of our strategies have to do with copying and moving source data from place to place: trying to centralize larger and larger bodies of source data.

Let’s define source data:

1. **Enterprise** data is behind firewalls – *at rest, not in motion*

2. **Patient** self-sovereign data – *within cloud storage infrastructures or at the edge on devices*
Source data has been expanding exponentially for the last fifteen years. As data grows, we are reaching the logistic extremes of data lakes and other centralization strategies. The statistics demonstrate the magnitude:

- 1100 Terabytes (TB) per lifetime of exogenous data (behavior, socioeconomic, environmental, etc.) representing roughly 60 percent of the determinants of an individual’s health
- 6 TB per lifetime of genomics/epigenomics data representing roughly 30 percent of the determinants of an individual’s health
- 0.4 TB per lifetime of clinical data representing roughly 10 percent of the determinants of an individual’s health

Another consideration is associated metadata, which are descriptions of the source data. Finally, there is the product of the attempts to create value out of source data, which we call derived data, composed of analytical findings and insights. As such, the current centralized system is a sunsetting paradigm. It is slow, labor intensive, error prone, and untrusted. Decentralization with a blockchain framework is the emerging paradigm. It will be rapid, automated, quality controlled, and trusted.

Today, we have to copy and move data everywhere. Analysis can only be performed on that of which we can secure custody. Often, we make compromises by de-identifying data, which introduces a huge spectrum of new challenges and risks. In this emerging paradigm source, data will be able to remain at rest in enterprise infrastructures or in patient self-sovereign infrastructures. Most of the time, it will become the exception instead of the rule that the data has to be exchanged for a workflow to occur. Analysis will frequently be performed in federated analytics models across network participants and the source data. Further, source data will be able to remain fully identified and clinically actionable without ever exposing it, sharing it, or transferring ownership.

In this new paradigm, normalization and analysis will change substantially. Normalization will occur by distributing rule sets and automating that normalization: operating on the data where it resides at its source. The analysis will be also performed where it resides at its source, and only that derived data will be shared. The aggregated findings will be shareable and, importantly, also cryptographically verified.

The new paradigm is a shift to a new person-centered health data architecture as part of this healthcare industry transformation. This includes five architectural components, all of which are interdependent and necessary:

1. The individual’s self-sovereign identity
2. The individual’s verifiable credentials

3. The diametric workflows that connect the individual represented to the physical person it represents

4. A self-sovereign data lake where all of an individual’s data can be brought together—and where $N$ equals one, for that individual alone, so that it is not a giant target for hackers.

5. Finally, the ability to transact with a tokenized value, to earn, to trade, and to hold all forms of a tokenized value

All of this will require privacy-preserving technologies, primarily for health-focused, decentralized machine learning. This is what we call **federated learning**.

**Federated Learning**

What does it look like to combine federated or decentralized machine learning with a blockchain infrastructure and privacy-in-depth? What is the heart of the value proposition for doing that technically? The synergy between blockchain and decentralized AI comes into play as we implement access control and content management of data, including data sharing. This is merely the low-hanging fruit for initial implementation.

If we just manage consent and access on the blockchain, but still rely on the old primitives of moving data around, we come back to the paradigm of copying and creating multiple copies of data. And, therein lies the problem of not being able to create digital scarcity. As such, we are unable to increase the value of the data, but we do increase its vulnerability. For example, we become the victims of data breaches that are totally out of our control—even though we have recorded our access—and have been denied specific access to our own data.

If you could access data in a decentralized model—as opposed to our strategies today, where sourcing training data is the single greatest inhibitor to performing machine learning functions—greater use of the data would be possible. The reality is that it is very difficult for data scientists and machine learning practitioners to obtain the data they need to create models. In this decentralized approach, we can turn the scarcity and the behavioral economics of that dynamic completely upside-down and gain/allow access...
to vast datasets without ever having to centralize them. Could this make a difference in the amount of total data available and the degree to which we could trust that data and its provenance in this new blockchain-facilitated decentralized AI?

If you look at most of the studies that are being run on healthcare data, it is rare to reach very large scales in terms of data, unless we are speaking of very carefully crafted datasets, and that is a complex and costly endeavor to put into place. Whereas, if you flip the paradigm around, everyone is able to contribute their personal data, healthcare data, or data that is only secondarily related to their health. You could contribute data from, for example, smartphone applications that receive requests from a clinical trial project. In this example, the project would be just using your data and not transferring your data, or owning your data, or being aware of its content while it’s being used. It is only giving the project/party permission to use it. That party cannot see it, doesn’t know what it is, and is never going to know what it is. This will create an incentive to come back and request more data to be used—and only that which is needed—instead of accumulating and stashing all of the data somewhere on an infrastructure—maybe securely and maybe not—with the hope of potentially using it someday.

We are shifting away from the old logic, which goes something like this:

- You can rarely monetize your data.
- Monetizing your data requires that you literally sell your data once and for all.
- Use of your data involves creating a whole new copy.

So, what we are discussing here with all this decentralization business is switching from data selling to data lending/leasing, and only providing the key elements.

Along with privacy-in-depth, there’s one other important aspect, and that’s the notion of secure inference. Inference is the function of a machine learning model making a prediction. After a model has been trained and is doing what it’s meant to do, it is theoretically making a prediction that might guide clinical decision making. It might guide patient behavior. It might guide access to care. It might guide anything where these technologies are implemented in our industry.
Conclusion

The combination of blockchain, decentralized AI, and privacy-in-depth is the data holy grail that has been missing from the fragmented, firewalled, and siloed systems that are currently in place in healthcare. We can free up more data for research and evidence-based medicine, get more expert eyes on all aspects of health, and allow more access to data without sharing and copying that data endlessly. This will allow more medical discovery and improved health outcomes without the exposure of people’s personal data and the risk of massive hacks and data breaches. When you layer these three areas of emerging technology together, you now have the complete technology stack that is both necessary and sufficient for the needs of the healthcare industry.