# Blockchain-Orchestrated Machine Learning for Privacy Preserving Federated Learning in Electronic Health Data

Jonathan Passerat-Palmbach, PhD[1,2*], Tyler Farnan[1], MS, Mike McCoy[1,3], Justin D. Harris[4*], Sean Manion, PhD[1*], Heather Flannery[1], Bill Gleim, MS[1]

1 - ConsenSys Health
2 - Imperial College London
3 - Thomas Jefferson University
4 - Microsoft Corporation
* - Presenting

# What is Decentralized AI?

CONSENSYS HEALTH

Integration of public and private **Blockchain**, **Privacy-in-Depth**, and **Decentralized AI**

**Decentralized Apps** (“dApps”)
*Web3 User Experience*

**Optional Tokenized Assets**
*Secured and Transferred*

**Smart Contracts**
*Secure Automation Across Organizational Boundaries*

**Blockchain Networks**
*Public, Private, and Hybrid*

## Privacy-in-Depth

- Zero-Knowledge Proofs (ZKPs)
- Trusted Execution Environments (TEEs)
- Secure Encrypted Virtualization (SEV)
- Blind Computation
- Verifiable Computation
- Secure Multi-Party Computation
- Differential Privacy
- Homomorphic Encryption
- Quantum-Resistant Encryption

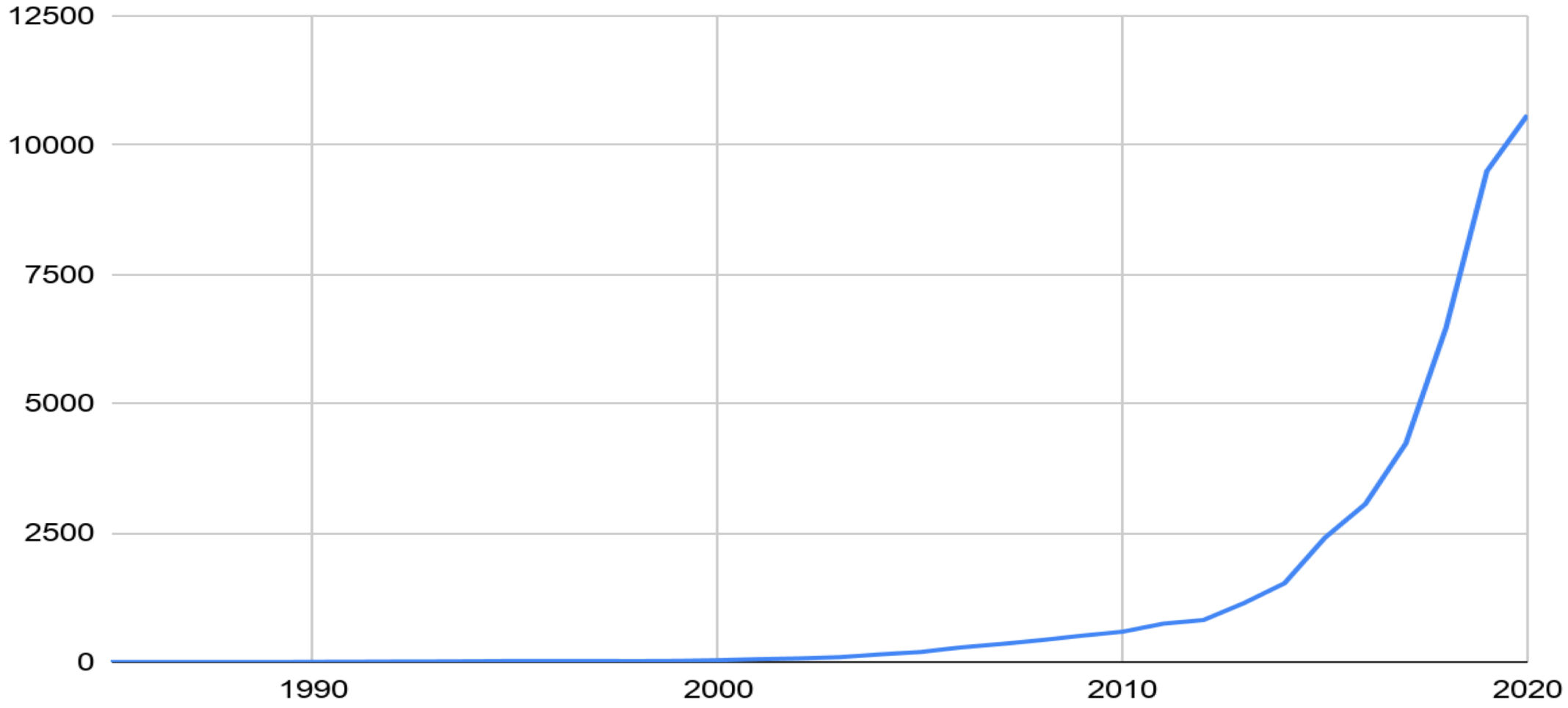
## Decentralized AI

- Federated learning in blockchain networks
- Intelligent agent-based automation
- Optimization w/agent-based simulation
- New paradigm in training data provenance

Web2: Today’s Mainstream Modern Web

CONSENSYS HEALTH

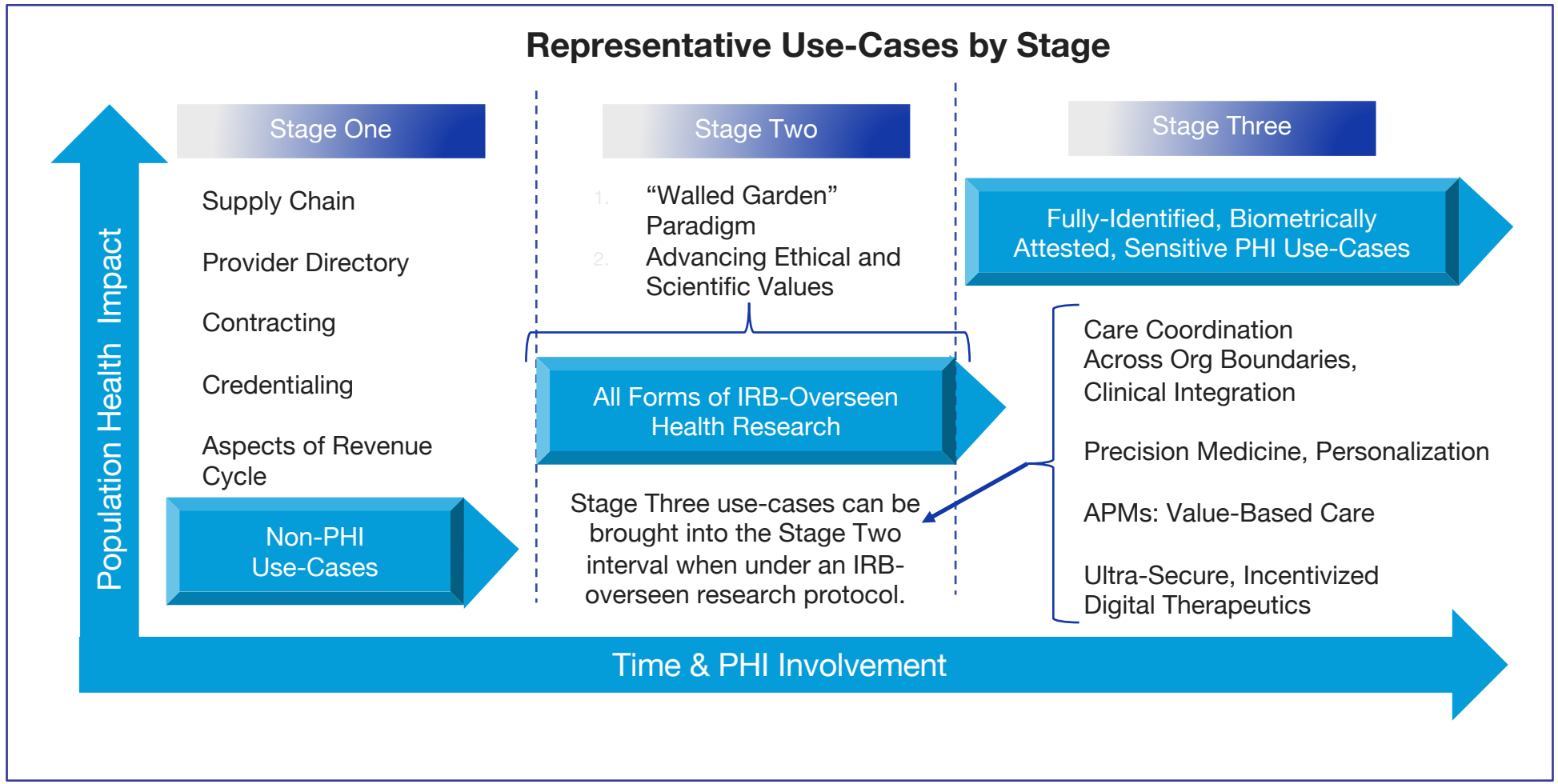
# Machine Learning for Medicine - “machine learning” results by year in PubMed (medical literature database)



# The PHI Readiness Framework

## PHI = Protected Health Information

CONSENSYS HEALTH



# Two Families of *Converging*, Emerging Technologies

CONSENSYS HEALTH

Integration of public and private **Blockchain**, **Privacy-in-Depth**, and **Decentralized AI**

**Decentralized Apps** (“dApps”)
*Web3 User Experience*

**Optional Tokenized Assets**
*Secured and Transferred*

**Smart Contracts**
*Secure Automation Across Organizational Boundaries*

**Blockchain Networks**
*Public, Private, and Hybrid*

?

Web2: Today’s Mainstream Modern Web

# What is Privacy-in-Depth?

CONSENSYS HEALTH

Integration of public and private **Blockchain**, **Privacy-in-Depth**, and **Decentralized AI**

**Decentralized Apps** (“dApps”)
*Web3 User Experience*

**Optional Tokenized Assets**
*Secured and Transferred*

**Smart Contracts**
*Secure Automation Across Organizational Boundaries*

**Blockchain Networks**
*Public, Private, and Hybrid*

## Privacy-in-Depth

- Zero-Knowledge Proofs (ZKPs)
- Trusted Execution Environments (TEEs)
- Secure Encrypted Virtualization (SEV)
- Blind Computation
- Verifiable Computation
- Secure Multi-Party Computation
- Differential Privacy
- Homomorphic Encryption
- Quantum-Resistant Encryption

Web2: Today’s Mainstream Modern Web

# What is Decentralized AI?

CONSENSYS HEALTH

Integration of public and private **Blockchain**, **Privacy-in-Depth**, and **Decentralized AI**

**Decentralized Apps** (“dApps”)
*Web3 User Experience*

**Optional Tokenized Assets**
*Secured and Transferred*

**Smart Contracts**
*Secure Automation Across Organizational Boundaries*

**Blockchain Networks**
*Public, Private, and Hybrid*

## Privacy-in-Depth

- Zero-Knowledge Proofs (ZKPs)
- Trusted Execution Environments (TEEs)
- Secure Encrypted Virtualization (SEV)
- Blind Computation
- Verifiable Computation
- Secure Multi-Party Computation
- Differential Privacy
- Homomorphic Encryption
- Quantum-Resistant Encryption

## Decentralized AI

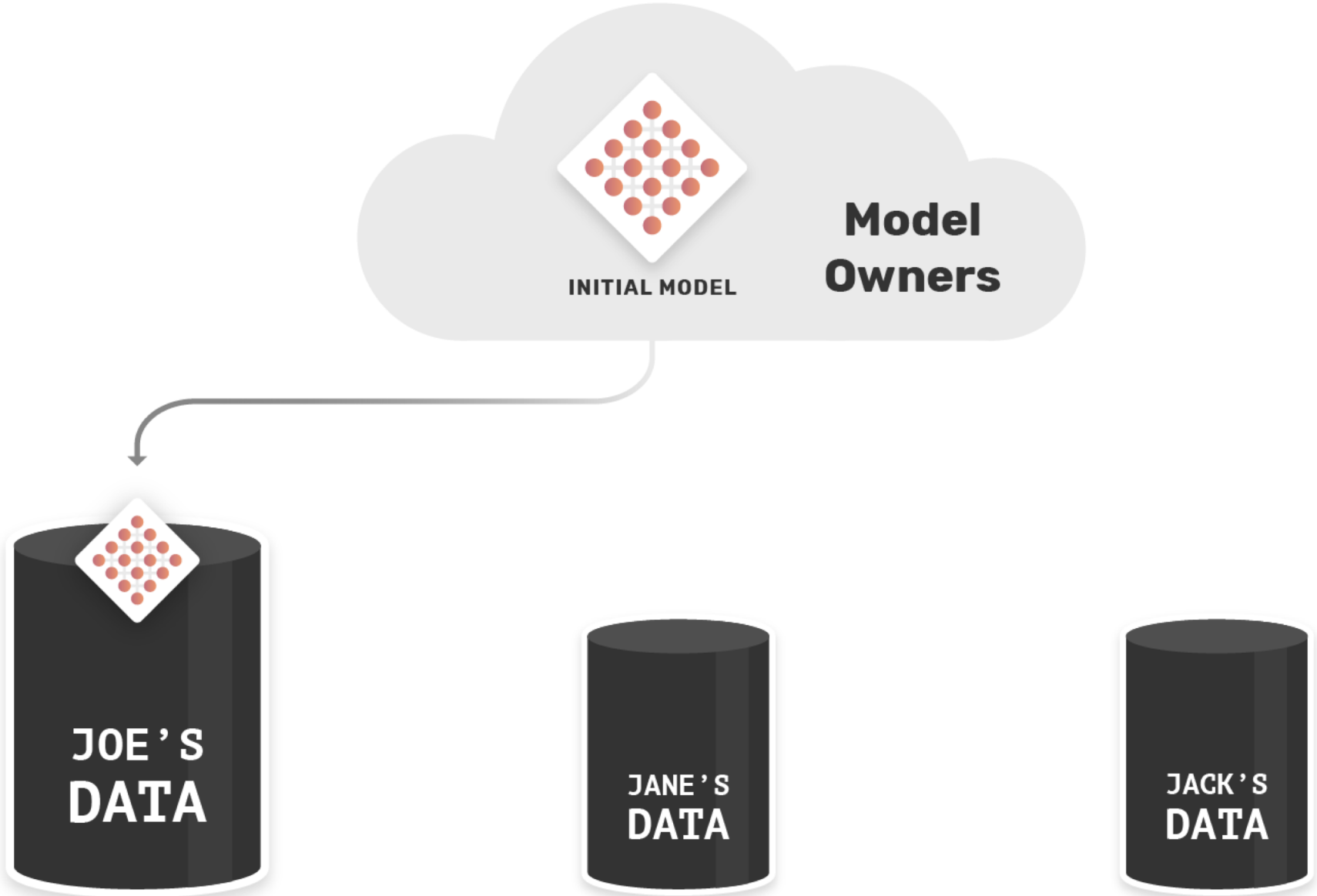
- Federated learning in blockchain networks
- Intelligent agent-based automation
- Optimization w/agent-based simulation
- New paradigm in training data provenance

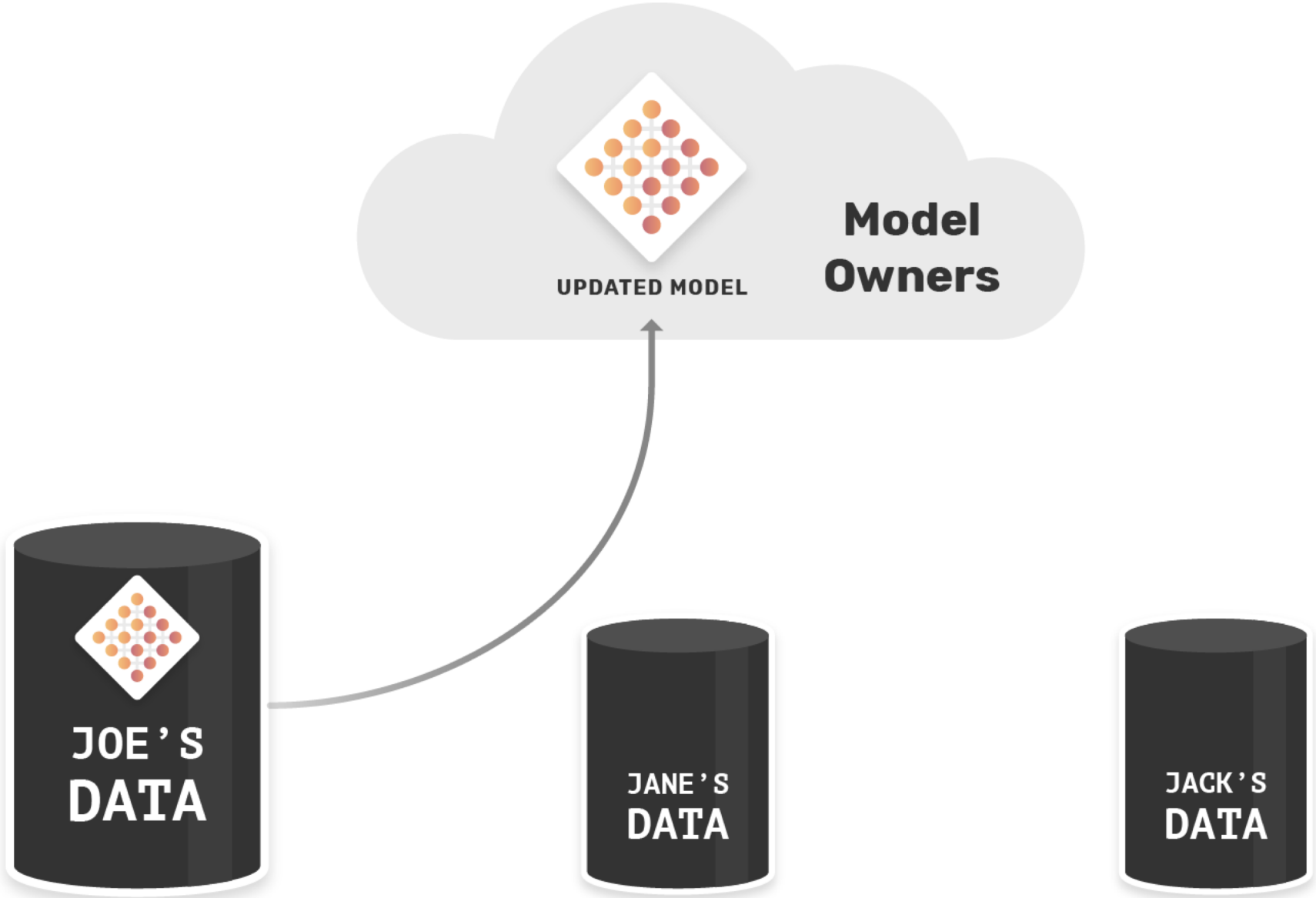
Web2: Today’s Mainstream Modern Web

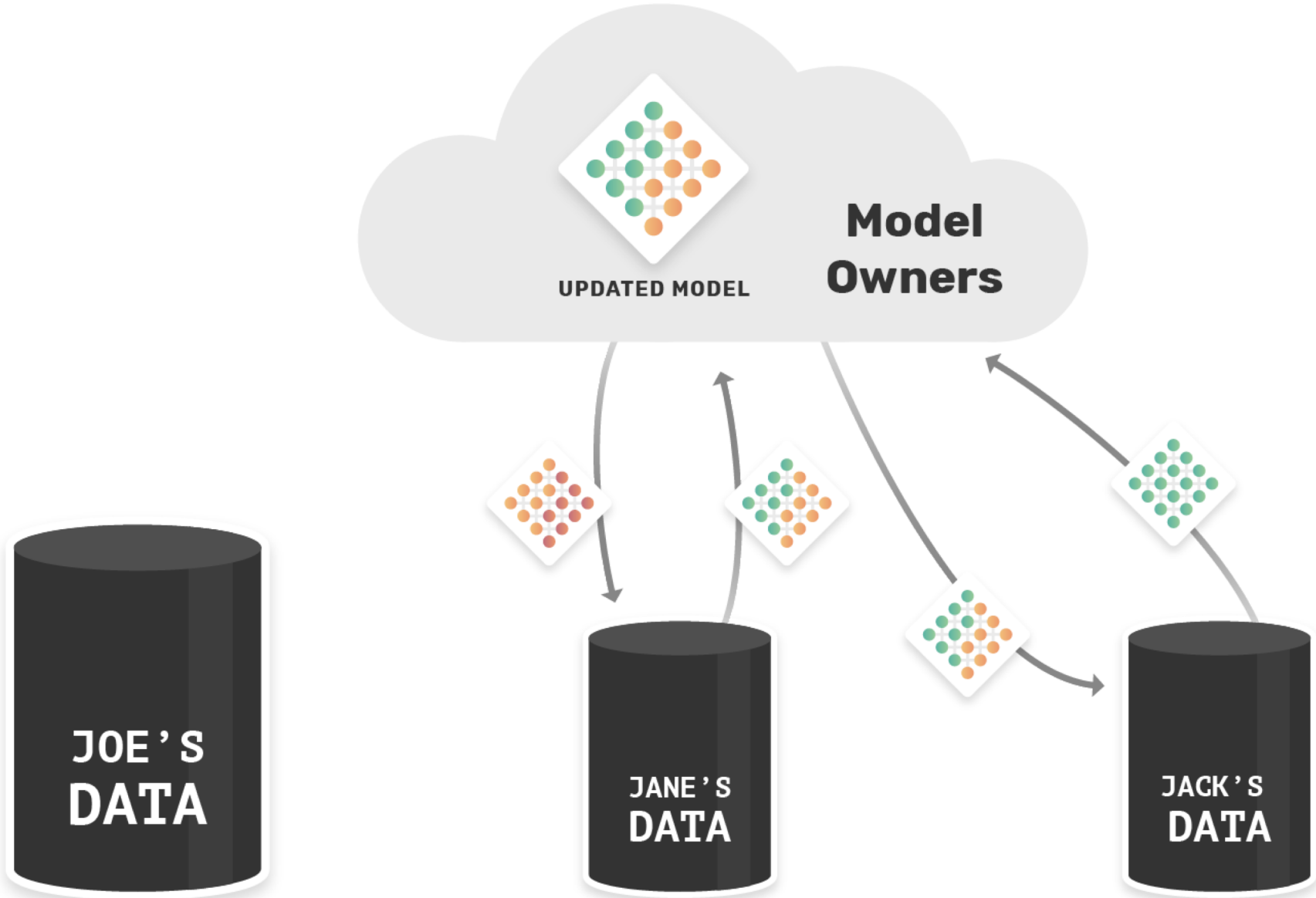
CONSENSYS
HEALTH

# Federated Learning

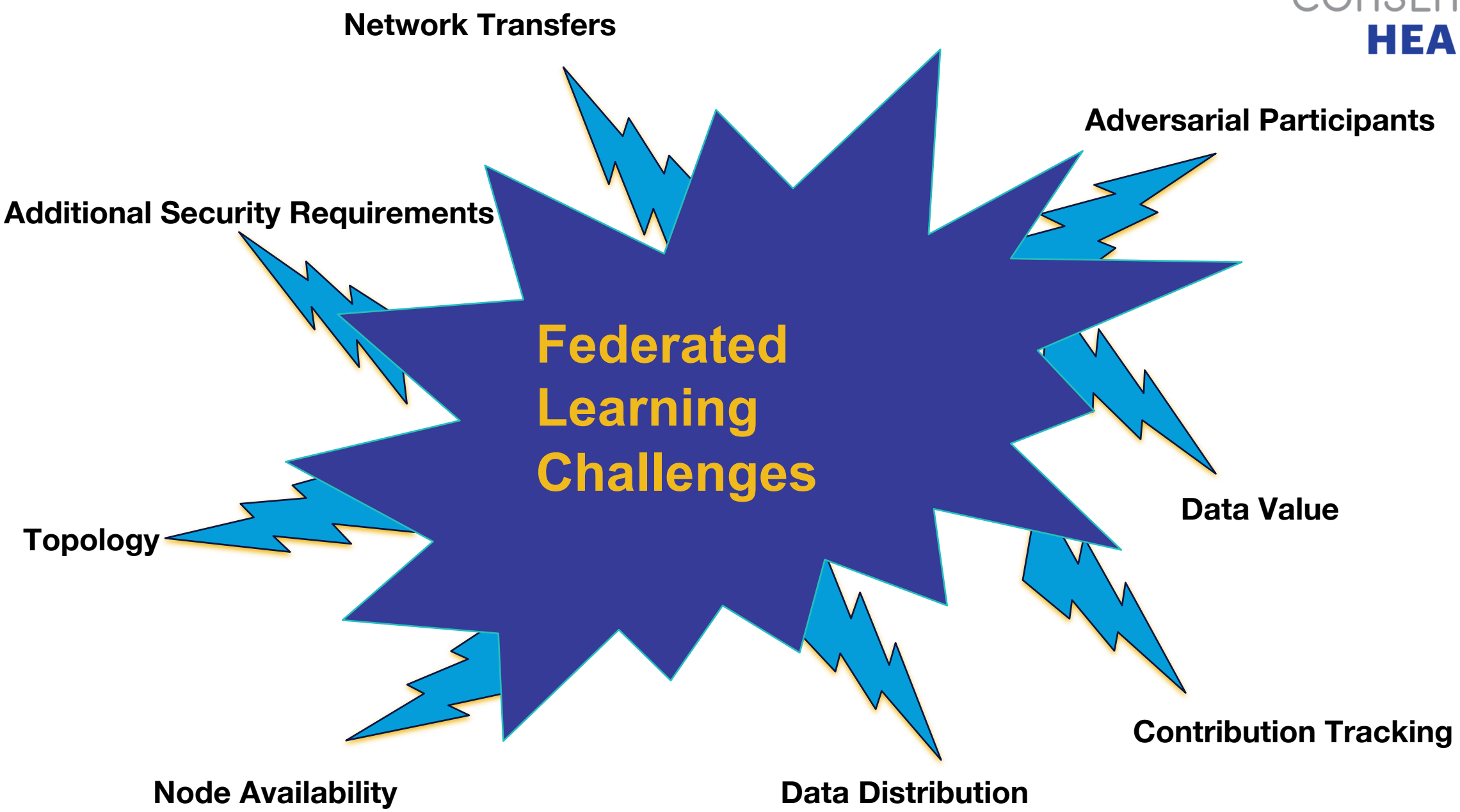
## Send the model, not the data.


INITIAL MODEL
Model Owners
JOE'S DATA
JANE'S DATA
JACK'S DATA


UPDATED MODEL
Model Owners
JOE'S DATA
JANE'S DATA
JACK'S DATA


UPDATED MODEL
Model Owners
JOE'S DATA
JANE'S DATA
JACK'S DATA

CONSENSYS HEALTH



# Federated Learning

- Academic references
  - **Bonawitz et al.,** *Practical Secure Aggregation for Privacy-Preserving Machine Learning*, 2017
  - **Ryffel et al.,** *A generic framework for privacy preserving deep learning*, 2018
  - **Kairouz et al.,** *Advances and Open Problems in Federated Learning*, 2019
- Software implementations
  - PySyft https://github.com/OpenMined/PySyft/
  - TF Federated https://github.com/tensorflow/federated

CONSENSYS HEALTH

# Distributed Systems / Blockchains Enable:

**AUDITABILITY**

Add trust and traceability to the learning process / data access

+

**ETHICS / GOVERNANCE**

Combined with PETs - provide a transparent, tamper-proof way to implement policies on AI models

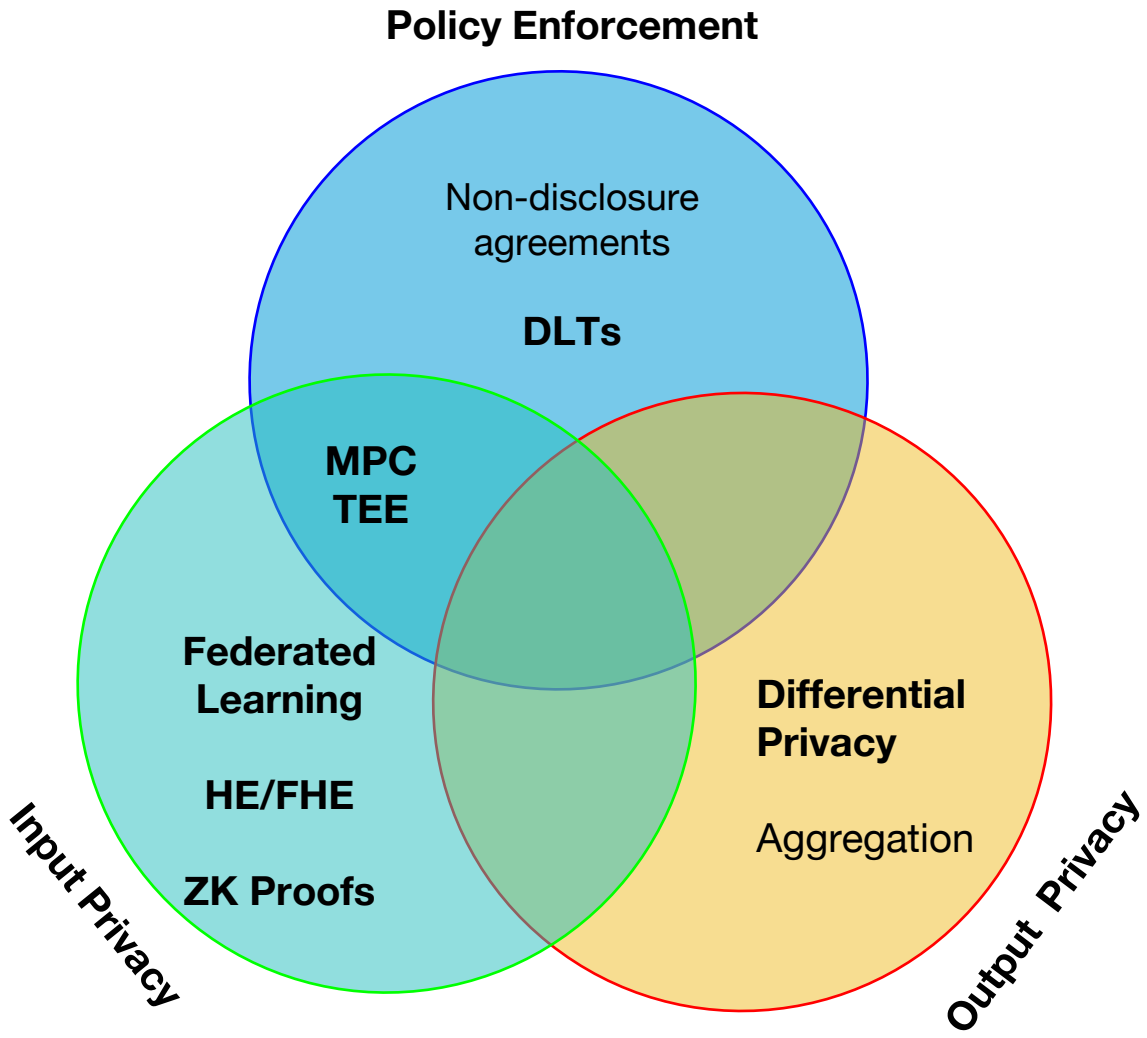
+

**MONETISATION**

From data sharing to data lending.

CONSENSYS HEALTH

# Privacy Enhancing Technologies - **3 Categories**

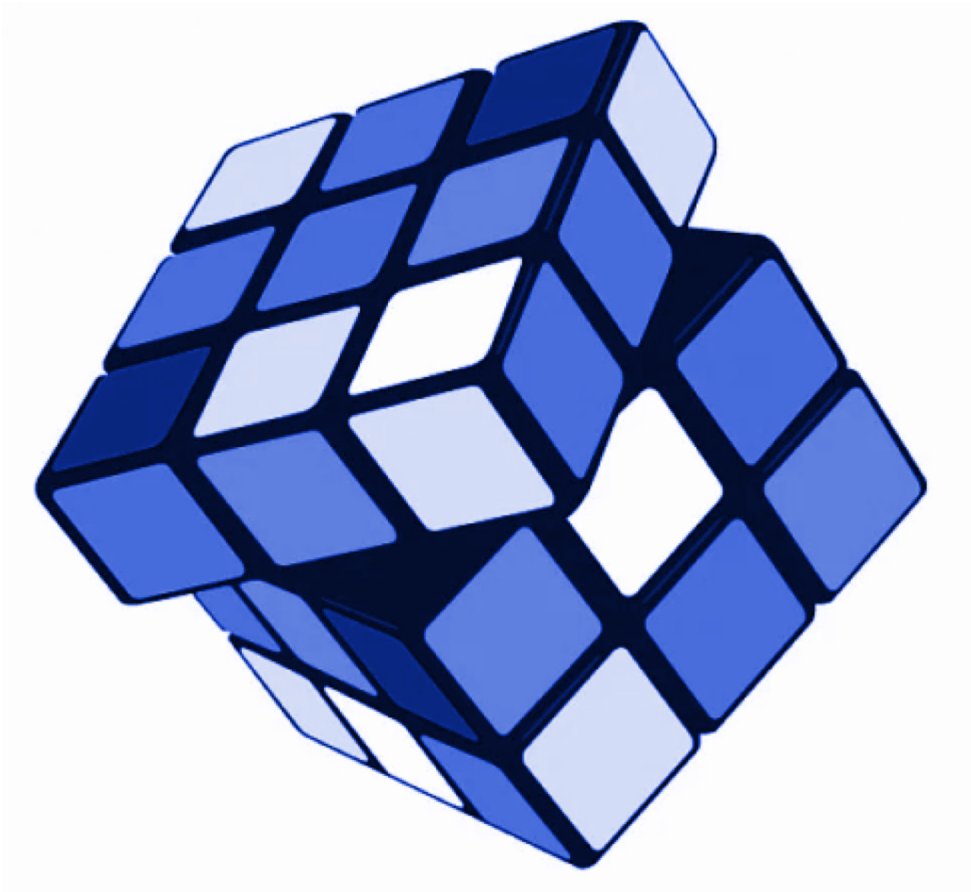



*[Credits] UN Handbook on Privacy-Preserving Computation Techniques*

CONSENSYS HEALTH

# Encrypted Computation - Challenges

- Computational requirements
- Numeric stability and precision
- Trusted third parties / hardware vendors
- TRAINING
- Adversaries and dishonest participants
- Leverage GPUs



# Secure Computing and Blockchain?

CONSENSYS HEALTH

**Elon Musk** @elonmusk · 12h

What should be developed on Ethereum?

960 764 4.3K

**Vitalik Non-giver of Ether**

@VitalikButerin

Follow

Replying to @elonmusk

## My top picks (1):

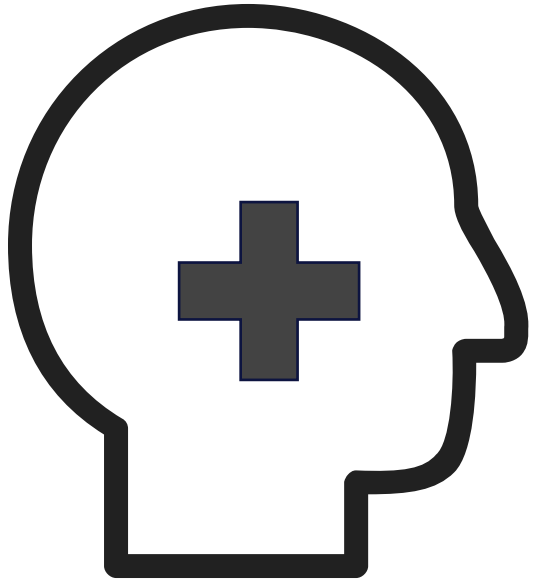
**Vitalik Non-giver of Ether** @VitalikButerin · 11h

(3)

* Markets for personal data for privacy preserving machine learning (you pay me X, I let you homomorphically execute function Y on my data that's been attestated to by Z...)

CONSENSYS HEALTH

# Incentives Introduced



**ENCOURAGE**

**Good quality data**

**Honest computation**

**e.g. model update or inference**

CONSENSYS HEALTH

# Solve this by answering some questions:

If data is private, then how do we determine its quality?

- **Trusted Execution Environments?**
- **Model Improvement with respect to a test set?**

How do we avoid biasing models?

- **Avoid accepting only consistent data.**
- **SOMETIMES OUTLIERS ARE GOOD.**

# Conclusion - Critical Elements

CONSENSYS HEALTH

| | |
|---|---|
| A | Data and analytic processes discoverable on a secure public blockchain while retaining privacy of the data and analytic processes |
| B | Value fabricated by generating data/compute matches that were previously illegal, unethical, and infeasible |
| C | Compute guarantees provided by federated learning and advanced cryptography |
| D | Privacy guarantees provided by software (e.g., Homomorphic Encryption, Secure Multi-Party Computation, ...) and hardware cryptography (e.g., Intel SGX and AMD SEV-SNP) |
| E | Data quality incentivized via tokenized reputation-based rewards |
| F | Discarding of poor data accomplished via model poisoning attack prevention techniques |

CONSENSYS
HEALTH

# Discussion

CONSENSYS
HEALTH

# Thank you!

CONSENSYS
HEALTH

# Thank you!