TripleBlind™

Date: **09/07/2022**

https://tripleblind.com

# Accelerating Digital Health

## The Role of Privacy-Enhancing Technologies

**Suraj Kapa, MD**

Chief Medical Officer,

SVP Healthcare

suraj@tripleblind.ai

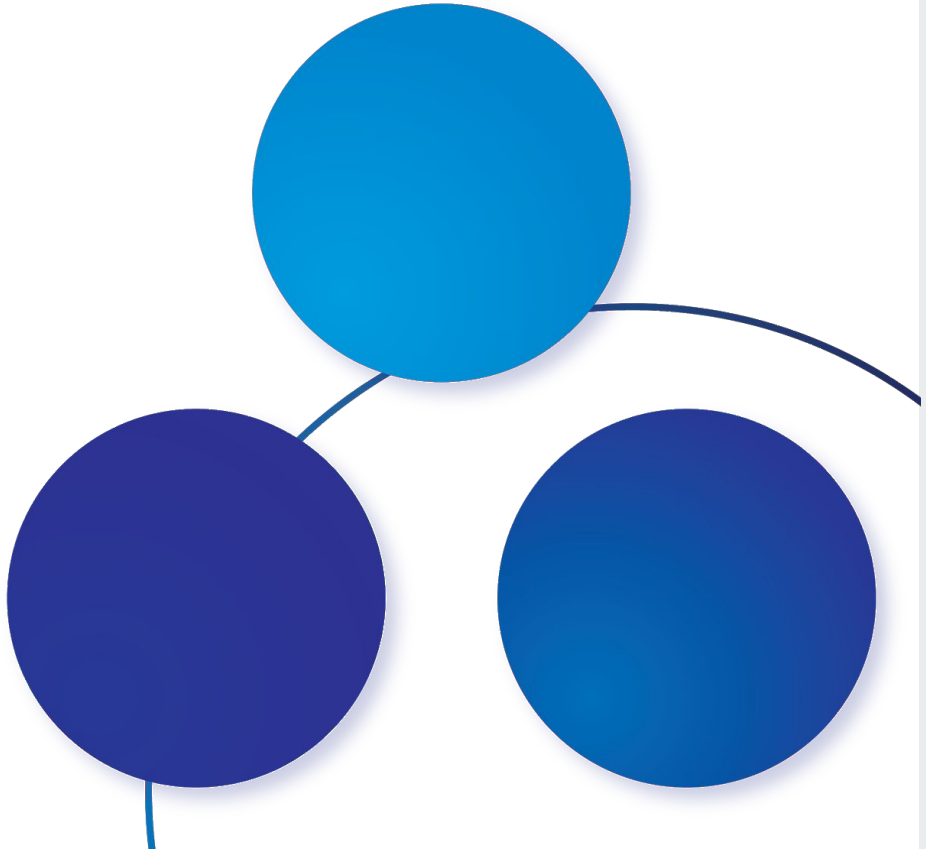
# Conflict of Interest

Suraj Kapa, MD: Chief Medical Officer and SVP of Healthcare works for TripleBlind, a privacy-enhancing technology company.



# Agenda

- The promise of digital health
- Barriers to achieving scalable digital health deployment
- The role of privacy enhancing technologies

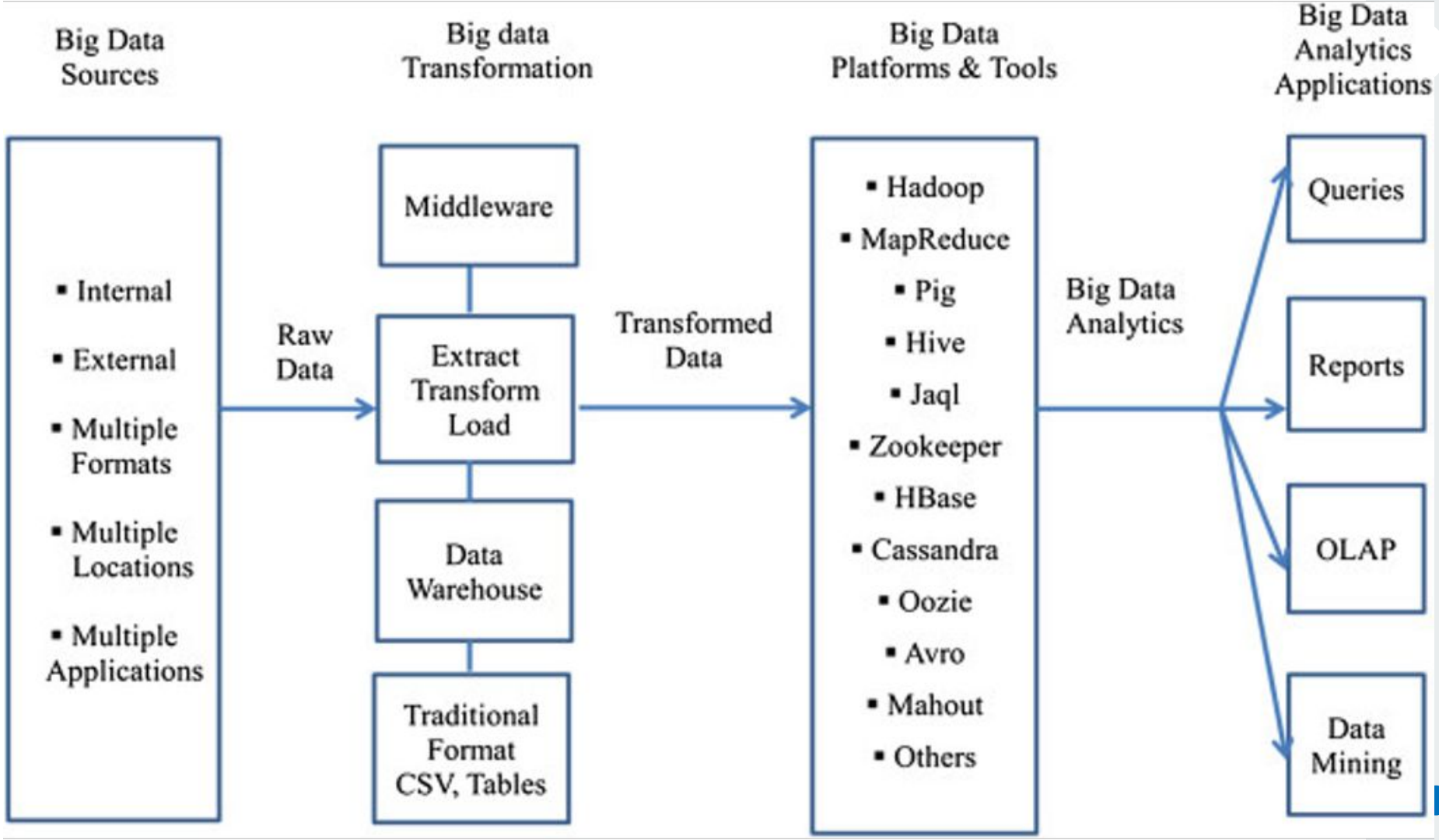


# The promise of digital health

- Enable scalable global health
- Create improved, efficient pathways for care
- Augmented insights into human disease




- Machine intelligence + clinical intelligence = medical intelligence

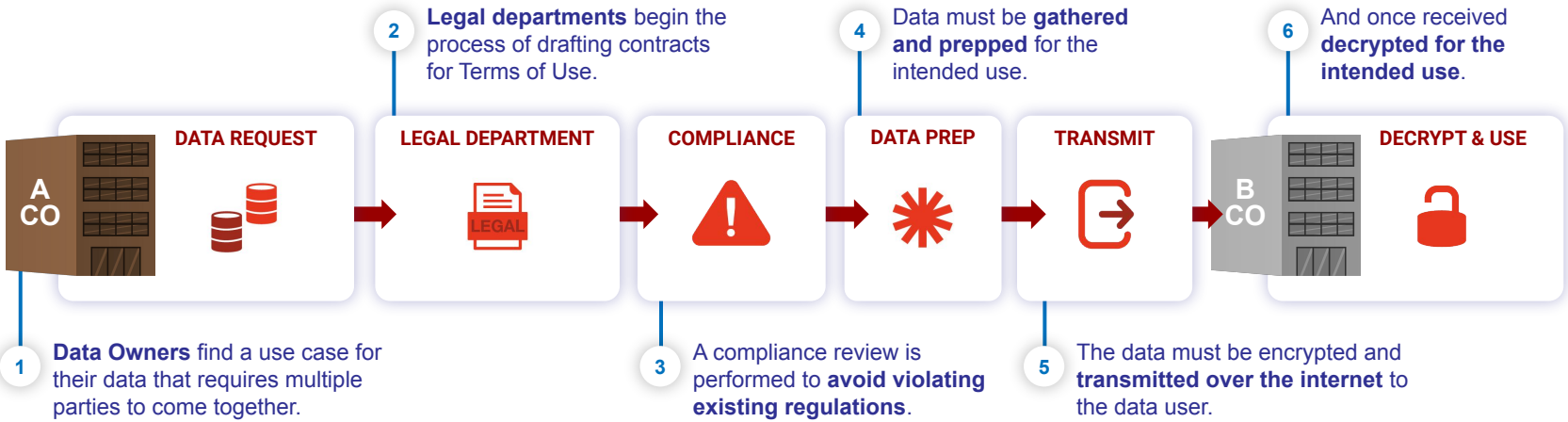

Big Data Sources
Big data Transformation
Big Data Platforms & Tools
Big Data Analytics Applications
Internal
External
Multiple Formats
Multiple Locations
Multiple Applications
Raw Data
Middleware
Extract Transform Load
Data Warehouse
Traditional Format CSV, Tables
Transformed Data
Hadoop
MapReduce
Pig
Hive
Jaql
Zookeeper
HBase
Cassandra
Oozie
Avro
Mahout
Others
Big Data Analytics
Queries
Reports
OLAP
Data Mining

# Barriers to scalable digital health deployment

- Data normalization / standardization
- Data access
- Bias within algorithms built off limited data sets
- Efficient deployment of digital algorithms



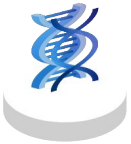
# How Data Ecosystems **Operate Today**



# Compliance With HIPAA Requires De-Identification

HIPAA compliance is not trivial.

It involves removing **18 identifiers** and gathering an “expert opinion” that the statistical or scientific principles used result in a low chance of identification.



Genetic data is impossible to de-identify using any of the above methods, as it has been proven that anonymized genetic data can still be traced to an individual.

## HIPAA Identifiers

1. Name
2. Address
3. Significant Dates
4. Phone Numbers
5. Fax Numbers
6. Email Address
7. Social Security Number
8. Medical Record Number
9. Health Plan Beneficiary Number
10. Account Number
11. Certificate or License Number
12. Vehicle Identifiers
13. Device Identifiers
14. Web URL
15. IP Address
16. Finger or Voice Print
17. Photographic Images
18. Other Characteristics that Could Uniquely Identify an Individual

# *Data Privacy Regulation*

## Strictness of laws globally

The map shows countries with strict, moderately strict, and light privacy regulations.

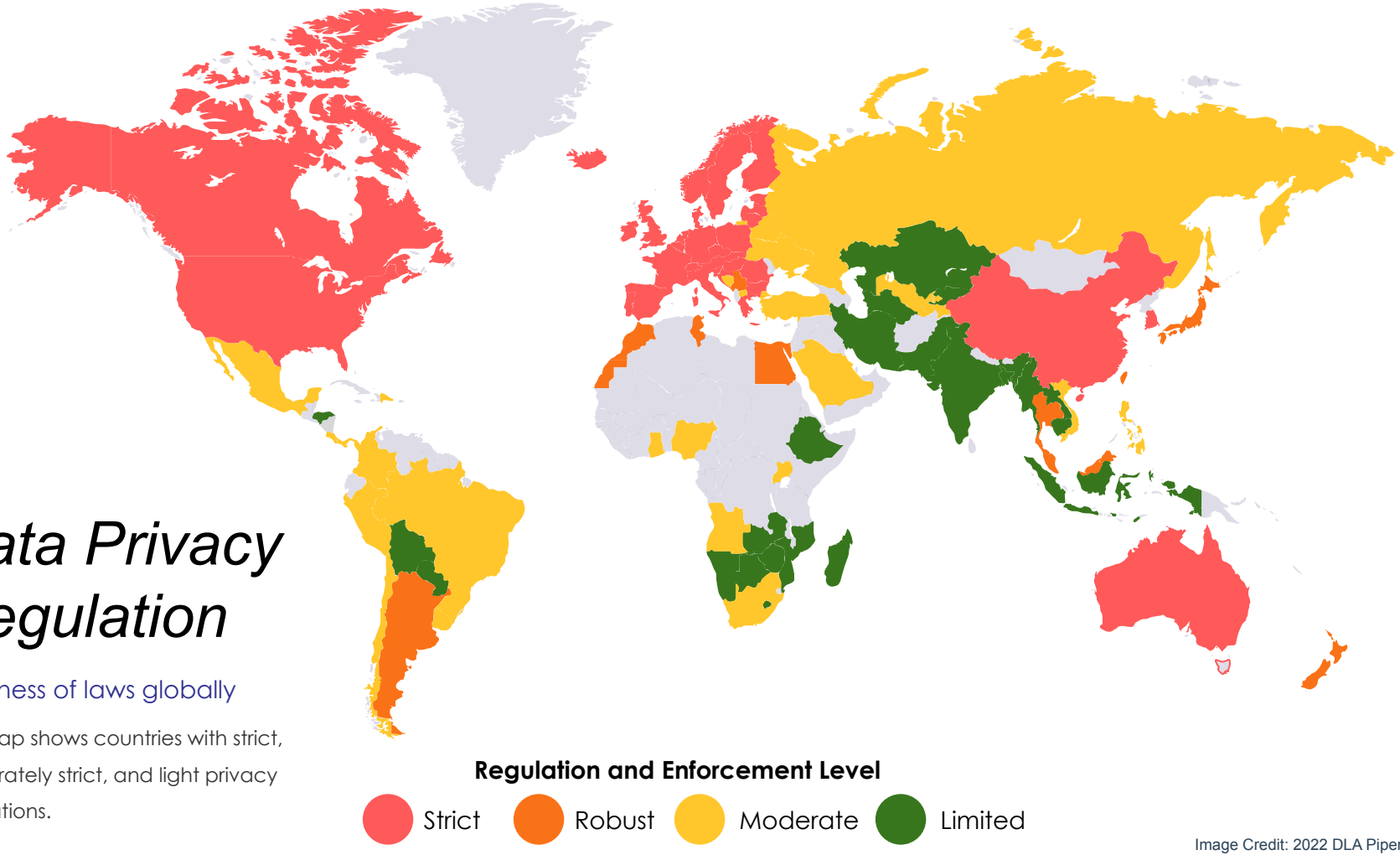



Image Credit: 2022 DLA Piper

# The role of privacy enhancing technologies

- Goal: Allow data to be operationalized while retaining individual privacy
- Minimize risk of use of data in a way that is not approved / acceptable
- Maintain regulatory compliance



## *Tokenization*

"Masks" sensitive data elements, replacing them with a token generated using an algorithm.

Drawbacks: Only works for tabular data. Individuals can be re- identified, violating HIPAA. Reduces the fidelity of the data by removing computable elements.Adds computational burden.

## *Synthetic Data*

Creates fake, or "synthetic", data based on the statistical properties of the original data.

Drawback: Eliminates useful outliers. Not real data - only useful for training models or analyzing metadata. Studies have shown, it does not retain data utility and leads to less accurate models. Adds time and computational burden.

## *Differential Privacy*

Adds noise to datasets to make it difficult to tell if information on a specific person is included in the dataset.

Drawbacks: By definition, reduces the accuracy of the analyses which can be performed using the data. Too little noise can make it easier to identify individuals in the dataset.

## *Homomorphic Encryption*

Allows for computations on encrypted data without needing a secret decryption key, enabling only those with the key to see the results.

Drawbacks: Extreme latency penalties / computation overhead. Only works with tabular data. Only supports algebraic operations. No digital rights management. Not HIPAA compliant due to existence of decryption key requiring safeguarding.

## *Secure Enclaves*

Provides a physical trusted execution environment for data and algorithms to be combined on a server.

Drawbacks: Hardware dependent. Requires the physical aggregation of data and algorithm - not GDPR or data residency compliant.

## *Federated Learning*

Specific to training machine learning models, federated learning brings the algorithm to the data.

Drawbacks: Adds storage requirements and computational burden on the data owner(s).Does not protect the IP of the model.Requires a strong connection to transmit large models. Takes a very long time.

# Data Ecosystems **Should Be This Easy**


1. **Simplified agreements processing** with built-in compliance protections.

2. **Manage what can and can't be done** with your data. Once you create agreements and set permissions you can be sure **only approved process are run**.

3. **Set it up once** and reuse against real time data sets, removing data prep hassle.
.

# Blind Data & Algorithm API



Digital Rights Management

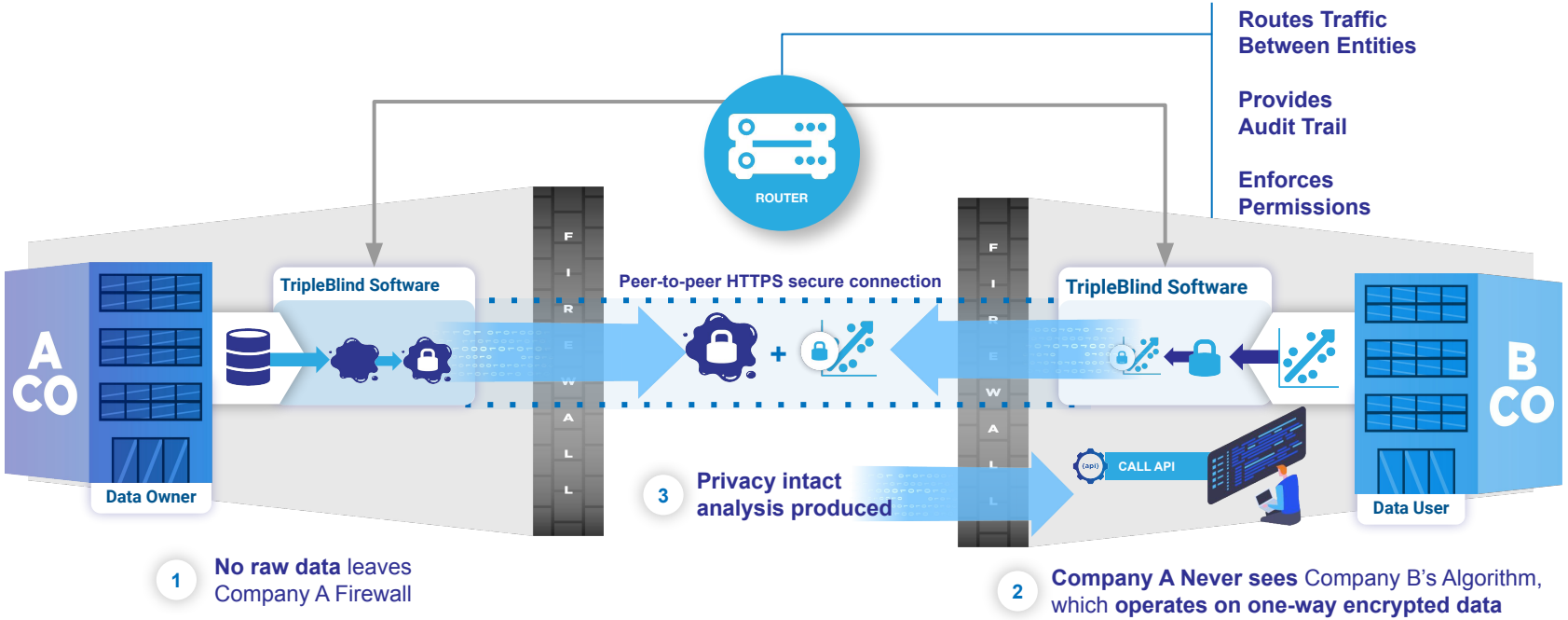
De-Identified

One-Way Encryption

Secure Multiparty Computation

No Stray Data

Peer-to-Peer




TRIPLEBLIND DOES NOT HOST ANY DATA

# Rapid Global AI Development

**DATA USERS**
AI/ML model developers

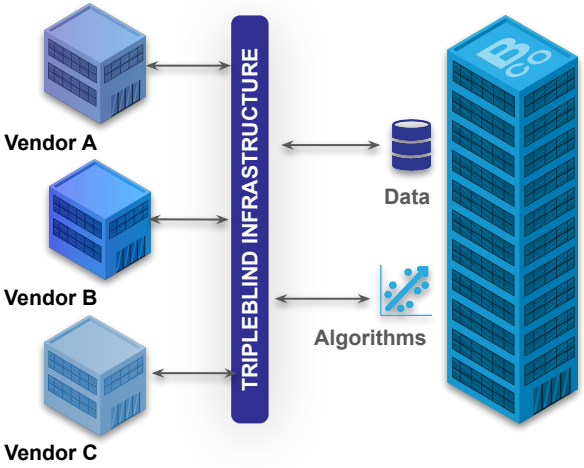
**DATA OWNER**
Hospitals or other data owners

**SUMMARY OF PAIN POINT**
AI and ML require sourcing diverse data, which is not always available. Once found, it can be impossible to use due to privacy regulations that silo it in place.

Current approaches like federated learning for remotely training AI algorithms on distributed data require the entire model to be shipped to each data owner, exposing IP.

**TYPE OF DATA**
Any health data - X-ray images, EKGs, EHRs, etc.

## WITH TRIPLEBLIND INFRASTRUCTURE




**RESULTS EXPECTED**
Using **Blind Learning**, "Objects" can be published to multiple counterparties in one integration (technical and contractual).

Counterparties never receive PII - no PII is exposed. And no manual de-identification is needed.

Each interaction enforces the appropriate privacy regulation (GDPR, CCPA, HIPAA, Data Residency, etc.)

Data owners keep their data in place, while the model owner never ships the full model to anyone and receives a fully trained AI model.

# What do PETs need to optimize value



**FAST AND ACCURATE**

Need to be both fast and accurate in operation.

**REAL-TIME DE-IDENTIFIED COMPUTATION**

Make any kind of PHI computable by third parties in real-time.

**NO COMPATIBILITY LIMITATIONS**

Need to support any type of data, any algorithm, including statistics

**GDPR/HIPAA COMPLIANT**

GDPR, HIPAA and data residency all need to be enforced to attain global scalability.

**NO DATA MOVEMENT**

Need to minimize dependence on trusted third parties or any movement of data.

**HARDWARE AGNOSTIC**

Ideally a software only approach - use your existing infrastructure, no specific hardware dependencies.

# Q&A