Connecting global datasets that you never move or see

Decentralized Federated Artificial Intelligence

Presagen
A.I. Enhanced Healthcare

SAN FRANCISCO | ADELAIDE | LONDON

Dr Jonathan Hall
Co-Founder

e: Jonathan@Presagen.com
l: http://www.linkedin.com/in/dr-jonathan-hall-41164a90/
t: @DrJMMHall
w: Presagen.com

Scalable, unbiased AI requires a globally diverse dataset

Healthcare: global diversity means different patient demographics and different clinical settings

Privacy policies and data laws prevent movement or centralization of medical data beyond borders
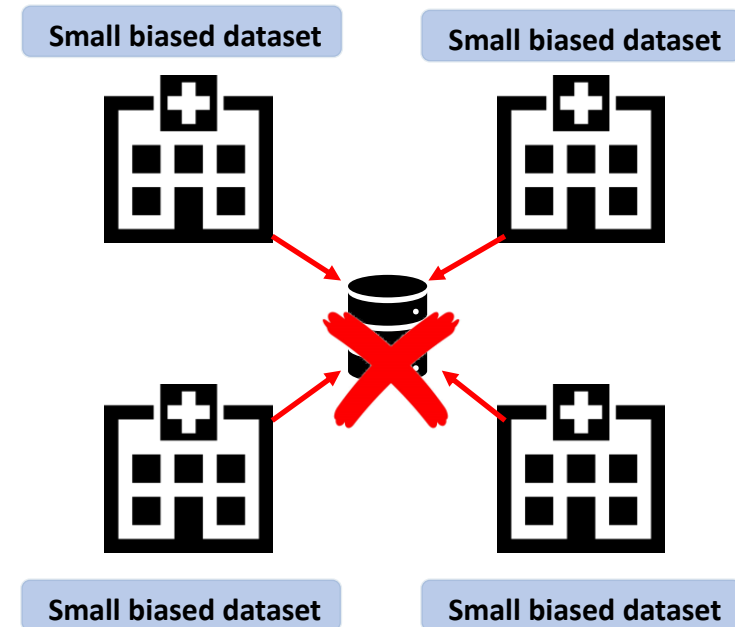
Centralizing global datasets for "traditional" AI training is not possible

Institutions want to leverage data but can't compromise on data privacy regulations

Silos typically don't have a large enough or diverse enough dataset to train scalable unbiased AI

Can't collect data in one place, Can't view each other's data

Federated Learning allows AI training on distributed data

AI does not require data to be moved, seen or owned

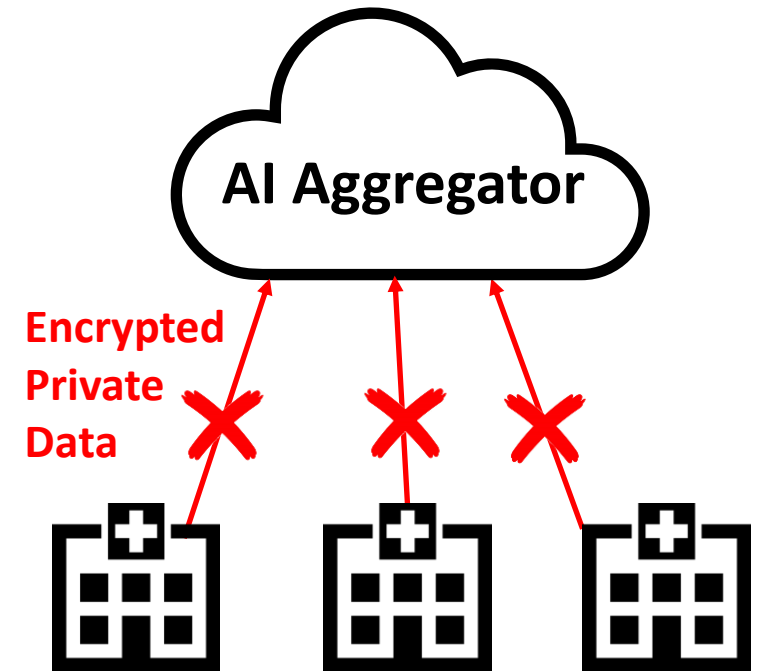AI just needs to **access** and **learn** from the data

Many federated learning algorithms are unsuitable

Still moves private data to a central server, by using heavy encryption

Very slow and inefficient

For pre-existing AI to tweak them on new data, not to create a new AI algorithms

AI Aggregator

Encrypted Private Data

Decentralized Federated AI Learning

Safely connects global datasets to AI which are siloed, private and distributed

Private data is not moved or shared, only the general learning from the data

Scalable and efficient, with great AI training performance

Patent: WO 2021/056043 A1

How does AI learn from data distributed around the world?

AI trains on data in its local data in a region

AI at each center is transferred to the next center

The transferred AI, existing local AI, and local data at each center is re-trained using a technique called distillation
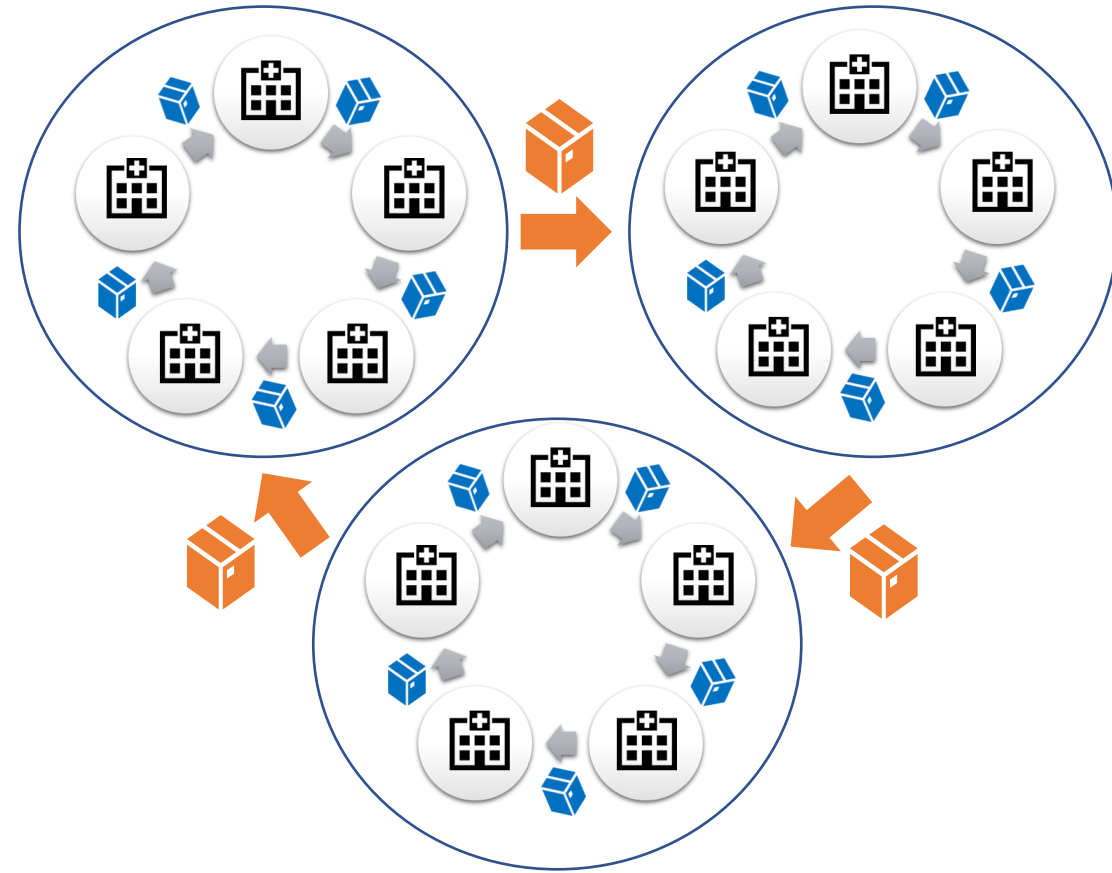
Repeat for all center, for *n* cycles, and then aggregate the final AI from each center

AI Algorithms (models) are shared, not the private data
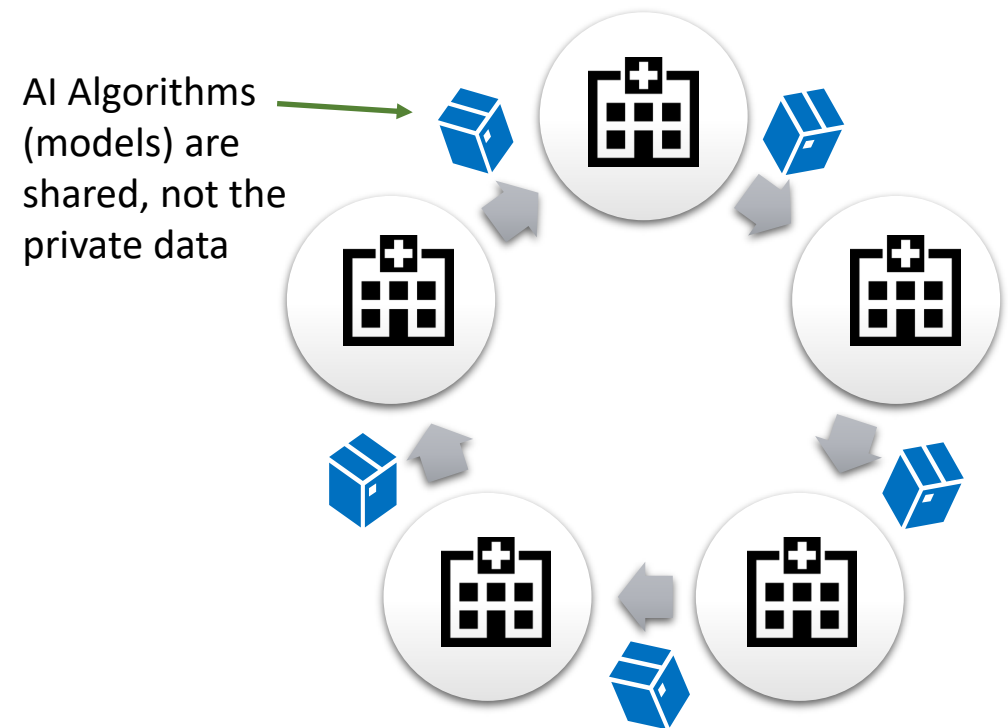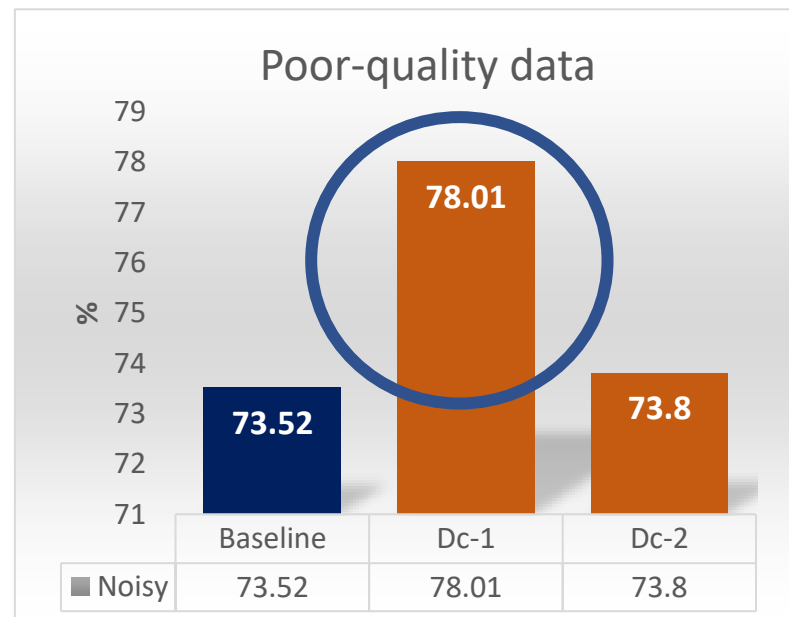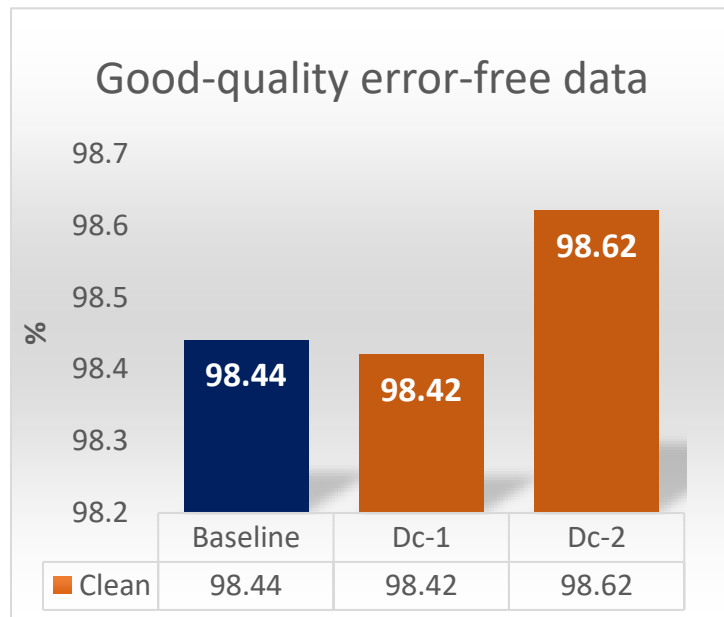
Scalability using clusters

Efficiency is configurable

Can optimize AI performance
versus time and cost of
training

Outstanding performance

Improved AI performance (accuracy) compared with centralized training

Good-quality error-free data

	Baseline	Dc-1	Dc-2
Clean	98.44	98.42	98.62

Poor-quality data

	Baseline	Dc-1	Dc-2
Noisy	73.52	78.01	73.8

- **Baseline**: Traditional Centralized AI training

- **DC-1**: Federated AI on 5 data centers, configuration 1

- **DC-2**: Federated AI on 5 data centers, configuration 2

Robust to real-world poor-quality data

AI performance improved when training on poor-quality data, even compared with traditional centralized training

This scenario is likely in most real-world situations, because:

- Healthcare data typically comprises inherent errors
- Each siloed data source (node) cannot be independently manually verified due to privacy

Applications

Safely connecting distributed private siloed data to AI

Applications where learnings from the data is more important or can be more efficiency shared than the data itself

Applications where collaboration and sharing of global data has greater value than local data

Finance, retail, cyber security, space/satellites, and defense.

Presagen
A.I. Enhanced Healthcare

Dr Jonathan Hall
Co-Founder

e: Jonathan@Presagen.com
l: http://www.linkedin.com/in/dr-jonathan-hall-41164a90/
t: @DrJMMHall
w: Presagen.com

info@Presagen.com
www.Presagen.com
www.linkedin.com/company/Presagen/
https://twitter.com/Presagen