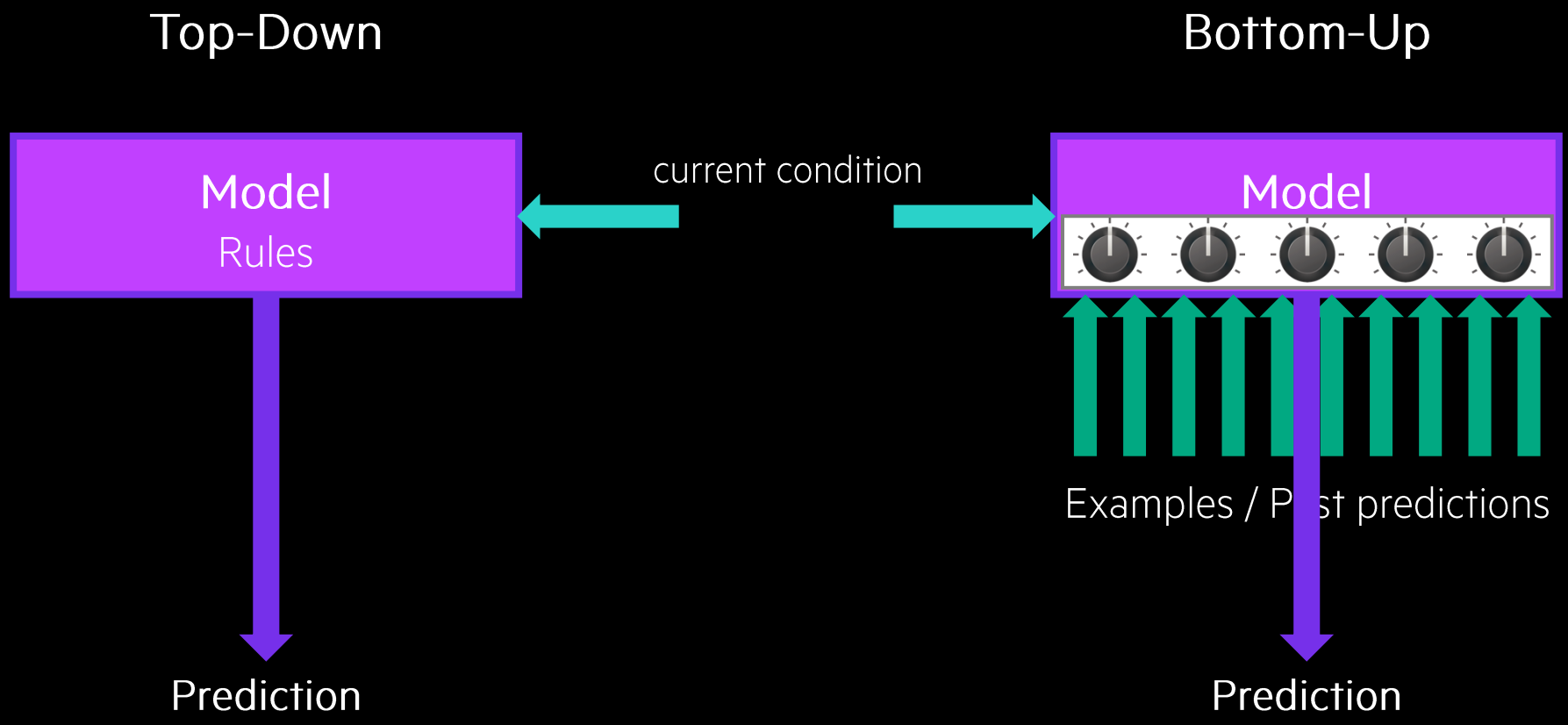
Jeff Winterich
jeff.winterich@hpe.com

June 2022

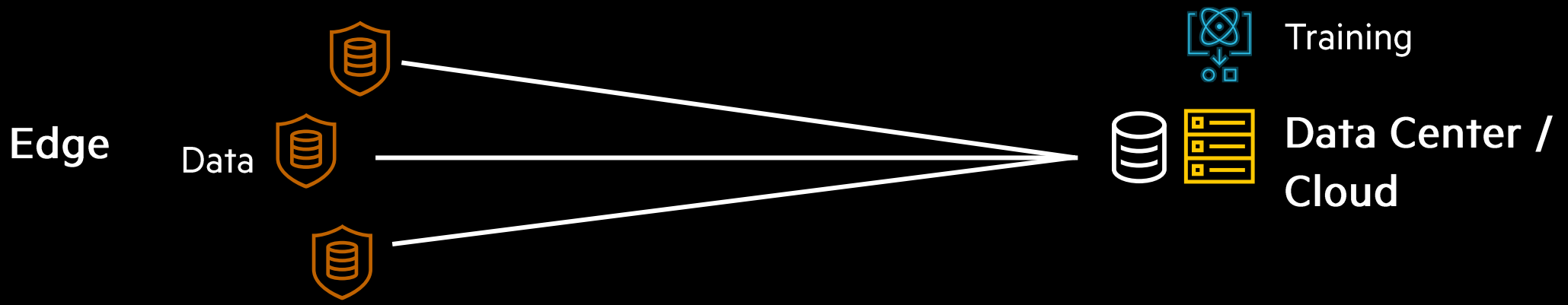
ARTIFICIAL INTELLIGENCE – MACHINE LEARNING

Replacing programming with training



MACHINE LEARNING—TODAY'S APPROACH

Centralized Learning



Process

- Data moved to central data center / cloud
- Model training performed at data center / cloud
- Trained model transferred back to Edge for inferencing

Data is distributed but model training happens at Data Center or Cloud

MACHINE LEARNING—TODAY'S CHALLENGES

 Low efficiency	Lots of data movement and duplication.
 Data Privacy/Security	Privacy acts. e.g. GDPR, prevent movement of data…data classifications, ITAR, etc
 Biased Data	Data biases due to demographic distribution
 Collaboration challenges	No framework to enable cross enterprise collaboration
 Insufficient Bandwidth	Low bandwidth, high latency, or DDIL environment

WHAT IS DRIVING THE NEED FOR SWARM LEARNING?

	Need	Need description
Primary	Improve accuracy of models	• Improve accuracy of models, and reduce bias by accessing a large data pool
	Protect Privacy	• Protect privacy of local data while satisfying the need for better performance
Secondary	Efficient use of resources	• Solve financial and operational pain points of transferring (and storing) large data volumes to a central location • Optimal leveraging of available computing resources
	Collaboration across organizations	• Collaboration across enterprises without central custodian, and appropriate incentive mechanism • Enable governance across multiple entities

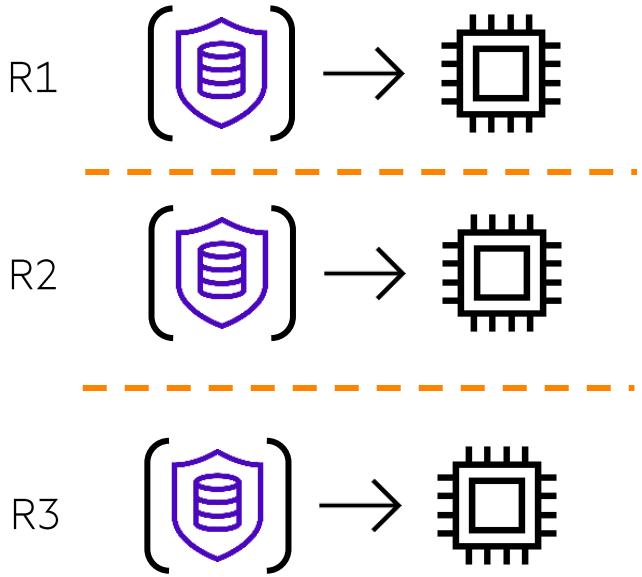
Tertiary	Ease of use	• Important given the nascent nature of technology, and lack of talent
		  • Ease of administration – deploy and manage solution
		  • Ease of development/ programmability – easily change ML model into a SL model

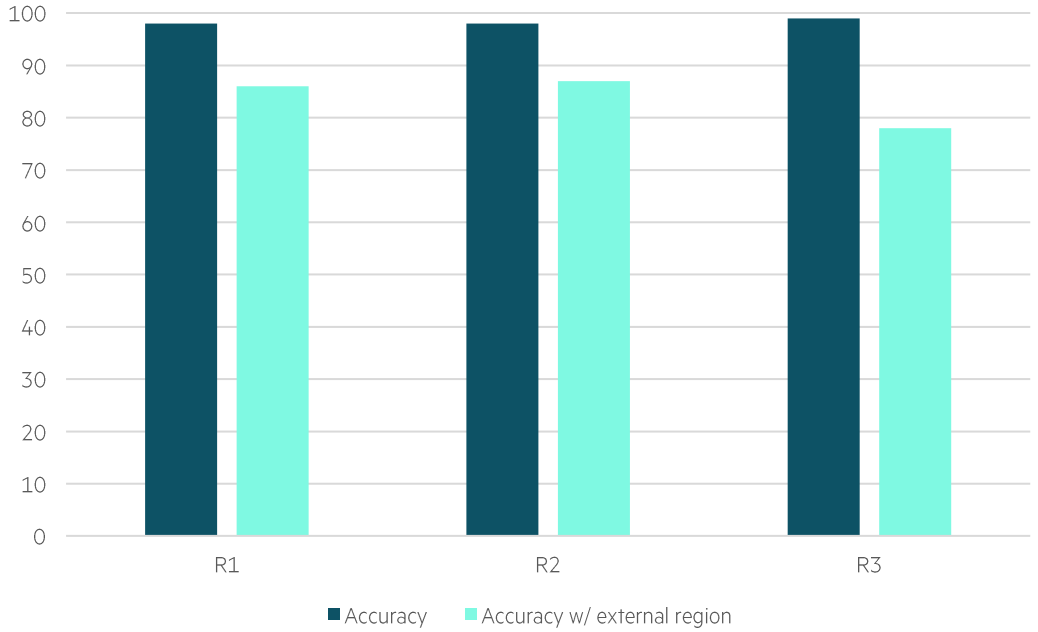
HOW DOES SWARM LEARNING SOLVE THESE NEEDS?

	Need	Feature	Differentiation (wrt alternatives)
Primary	Improve accuracy of models	• Enable models to run on decentralized data-sets, increasing training data pool	• Enhanced granularity of merge and adaptive adjustment → higher accuracy • Equal rights to all participants (full decentralization) makes collaboration easy, thus access more data/insights
	Protect Privacy	• Only model weights shared with other trusted nodes using secure communication	• Protected access to merge algorithm - blockchain smart contracts • Future - Secure containers (secure root of trust) • Future - homomorphic encryption
Secondary	Efficient use of resources	• Enable edge nodes to perform training, removing need for centralized data transfer	• Efficiently select coordinator based on resource usage • Adaptive merge algorithm balances accuracy vs performance trade off (resource usage and time)
	Collaboration across enterprises	• Enable collaboration by preventing raw data sharing between participants	• All participants equal – no central custodian • Enhanced traceability • Smart contract to prevent from insider attacks to poison the model • Future - Incentive mechanism
Tertiary	Ease of use	• NA (nascent technology)	• Administrative ease of use – a) containers, b) automated deployment • Programmability ease of use – Add few lines to 'swarmify' it ☺

HOW CAN REGIONAL DATA BIAS IMPACT MODEL ACCURACY?

Localized and Disconnected



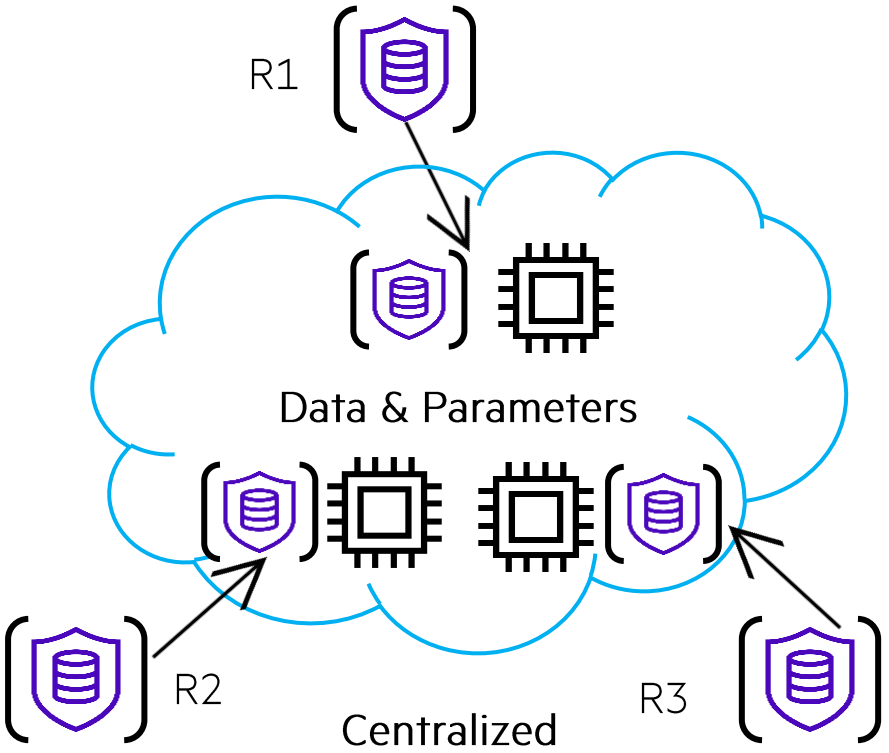


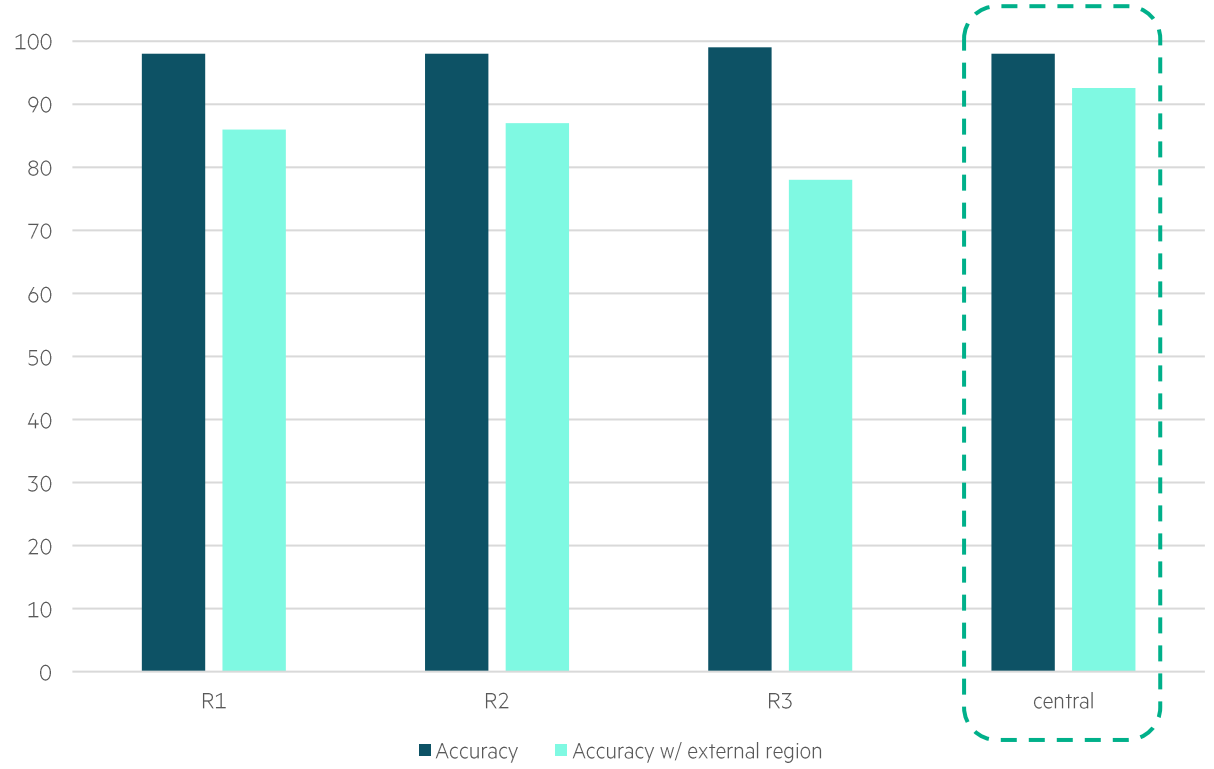
Biased Model due to localized data

A machine learning model is only as good as the data it is fed

MACHINE LEARNING JOURNEY

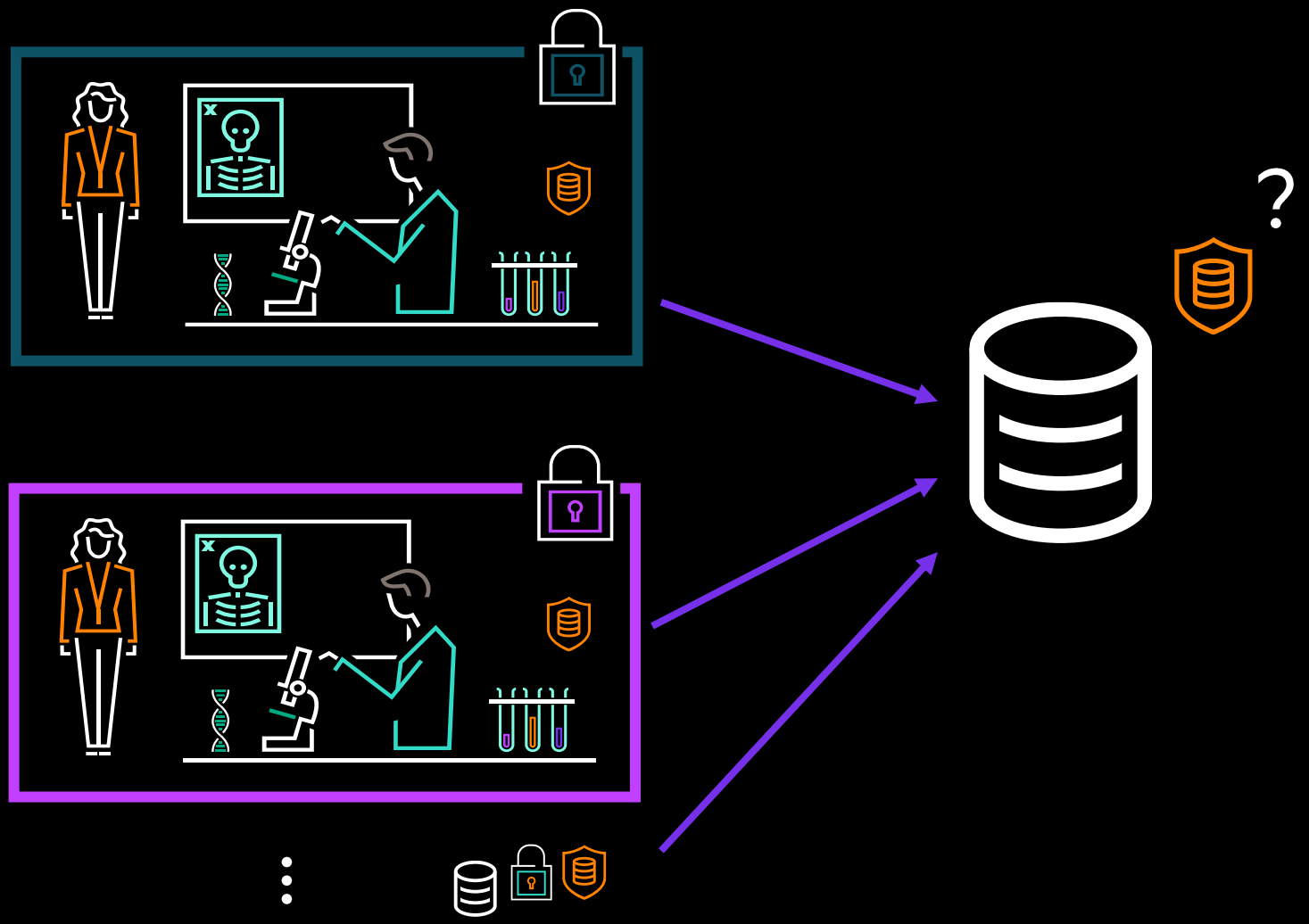
Centralized Learning





Centralized model is generic and demonstrates good accuracy across regions

HEALTHCARE—SHARE DATA?



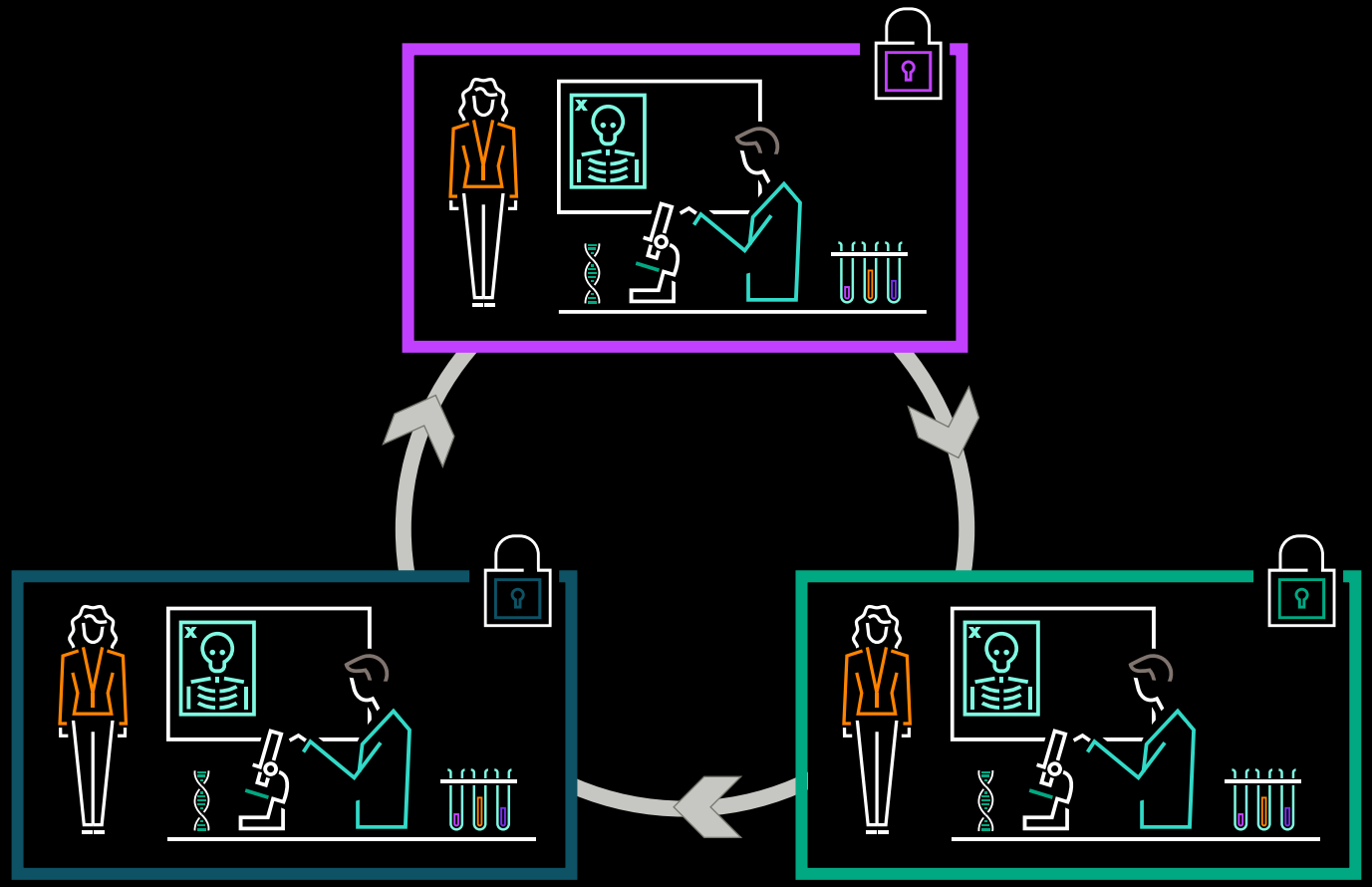
- Willing to share?
- Allowed to share?
  - GDPR
  - HIPAA
  - Consumer Privacy Act (CCPA)
- Agree on Data Sharing governance

- Data privacy
- Data movement & duplication

- **Data sharing, open data, centralized data** are concepts that are **diametrically** opposed to medical traditions

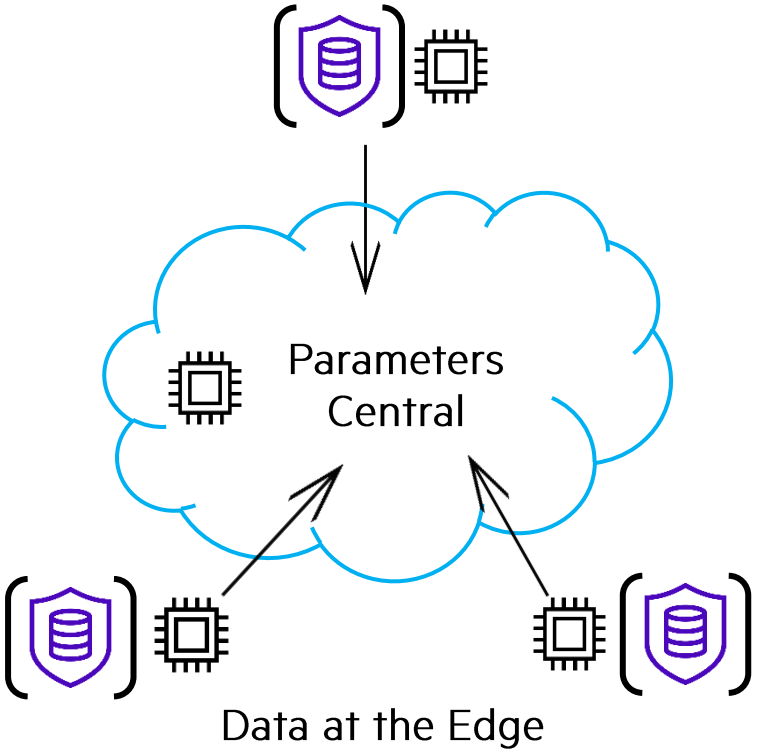
DATA SHARING IN HEALTHCARE—A NEW APPROACH

Decentralize Learning



- Collaborate with other institutes
- Do not move the data
- Keep ownership, privacy
- Share insights
- Enlarge dataset with a large network
- Reduce local bias

FEDERATED LEARNING



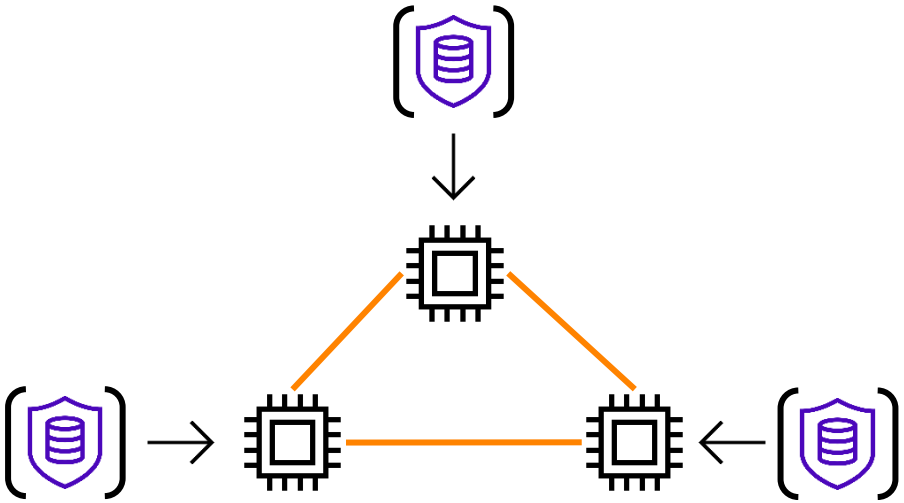
- Models are trained at Edge; Only Learning (parameters) are shared
- Data stays at Edge – ensures data privacy
- Parameters (Weights) are merged by Central Coordinator

Federated Learning solves privacy issues
But it still need a central custodian for coordination and merging the local learnings
This introduces a single point of failure

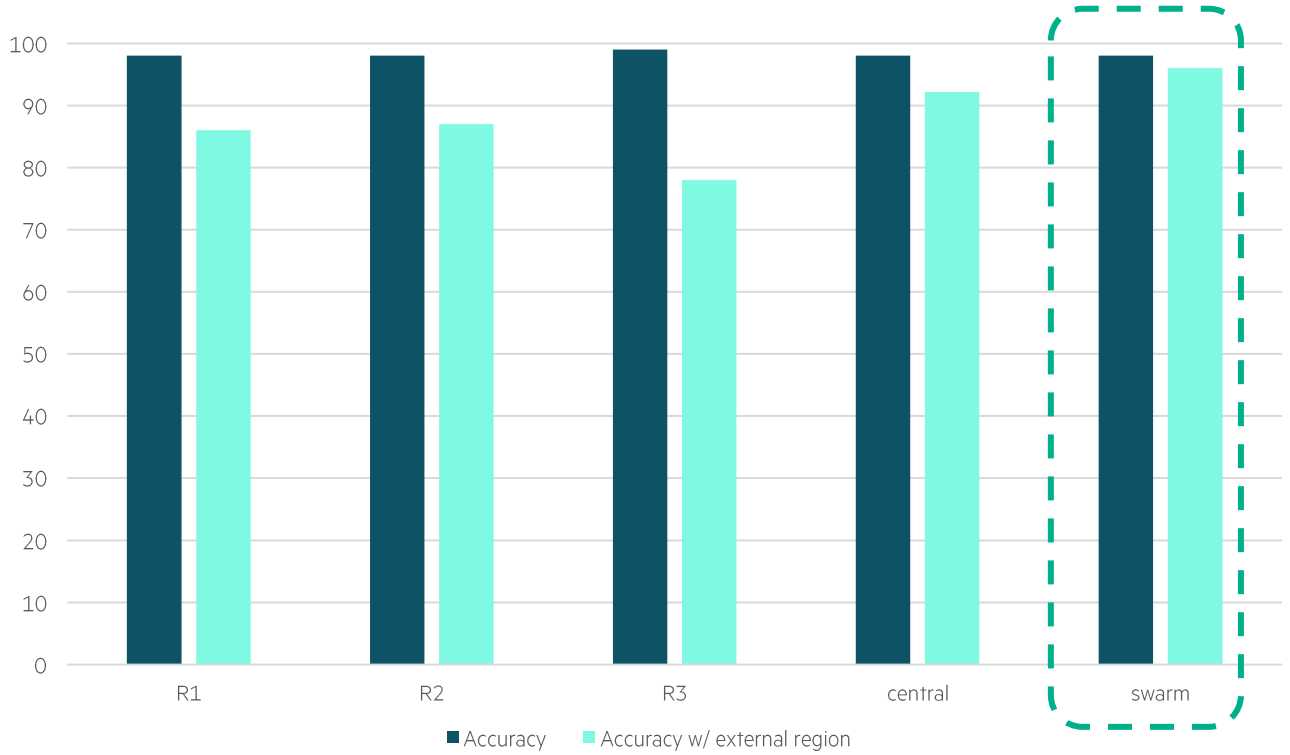
FEDERATED LEARNING SOLVES SOME **CHALLENGES**

 Low efficiency	Lots of data movement and duplication.
 Lack of Data Privacy	Privacy acts. e.g. GDPR, prevent movement of data
 Biased Data	Data biases due to demographic distribution
 **Collaboration challenges**	No framework to enable cross enterprise collaboration and TRUST
 **Insufficient Bandwidth**	Low bandwidth, high latency, or DDIL environment

SWARM LEARNING

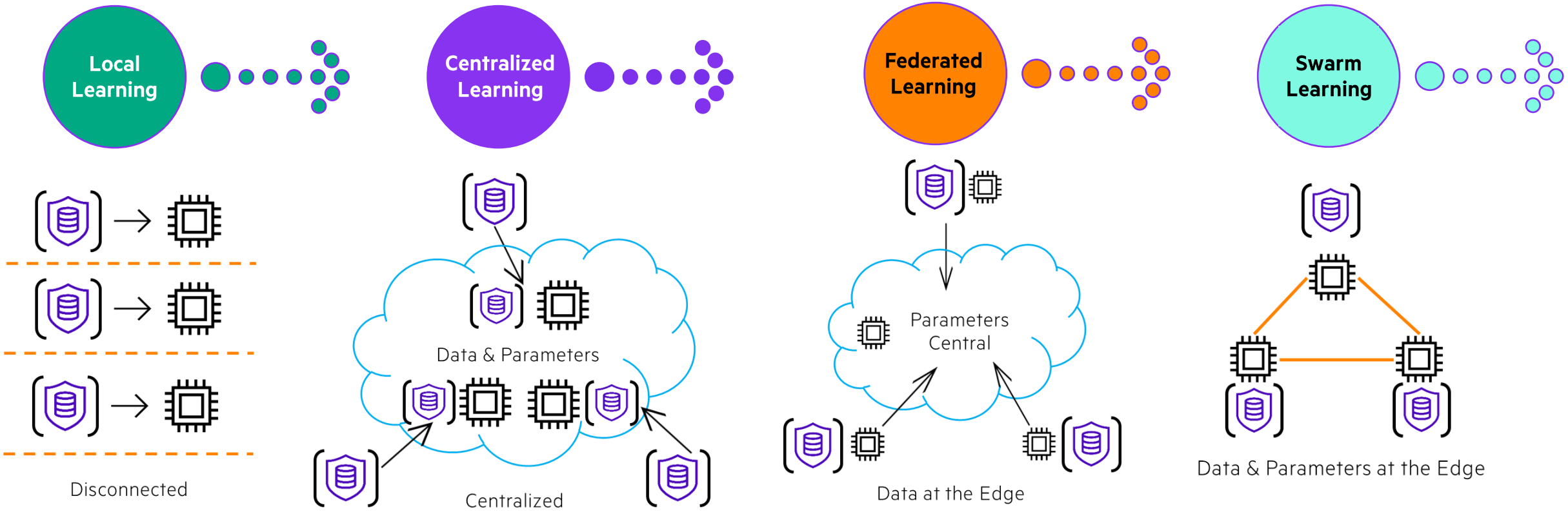


Data & Parameters at the Edge



Swarm Learning achieves the accuracy of centralized model without sharing data – thus enhances privacy and security

MACHINE LEARNING FRAMEWORK JOURNEY

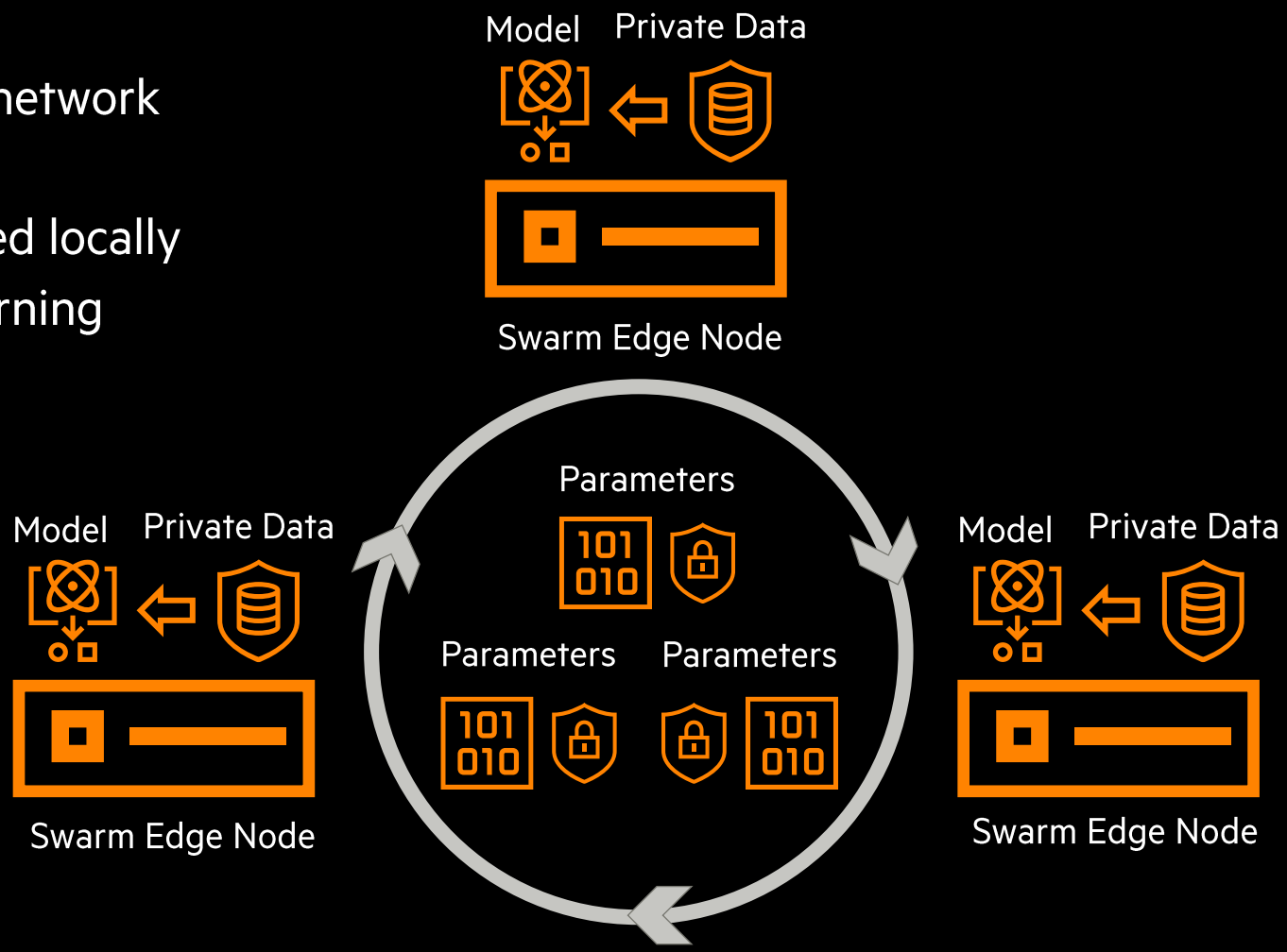


Swarm Learning enables privacy-preserving, secure and collaborative machine learning by treating all participants equally

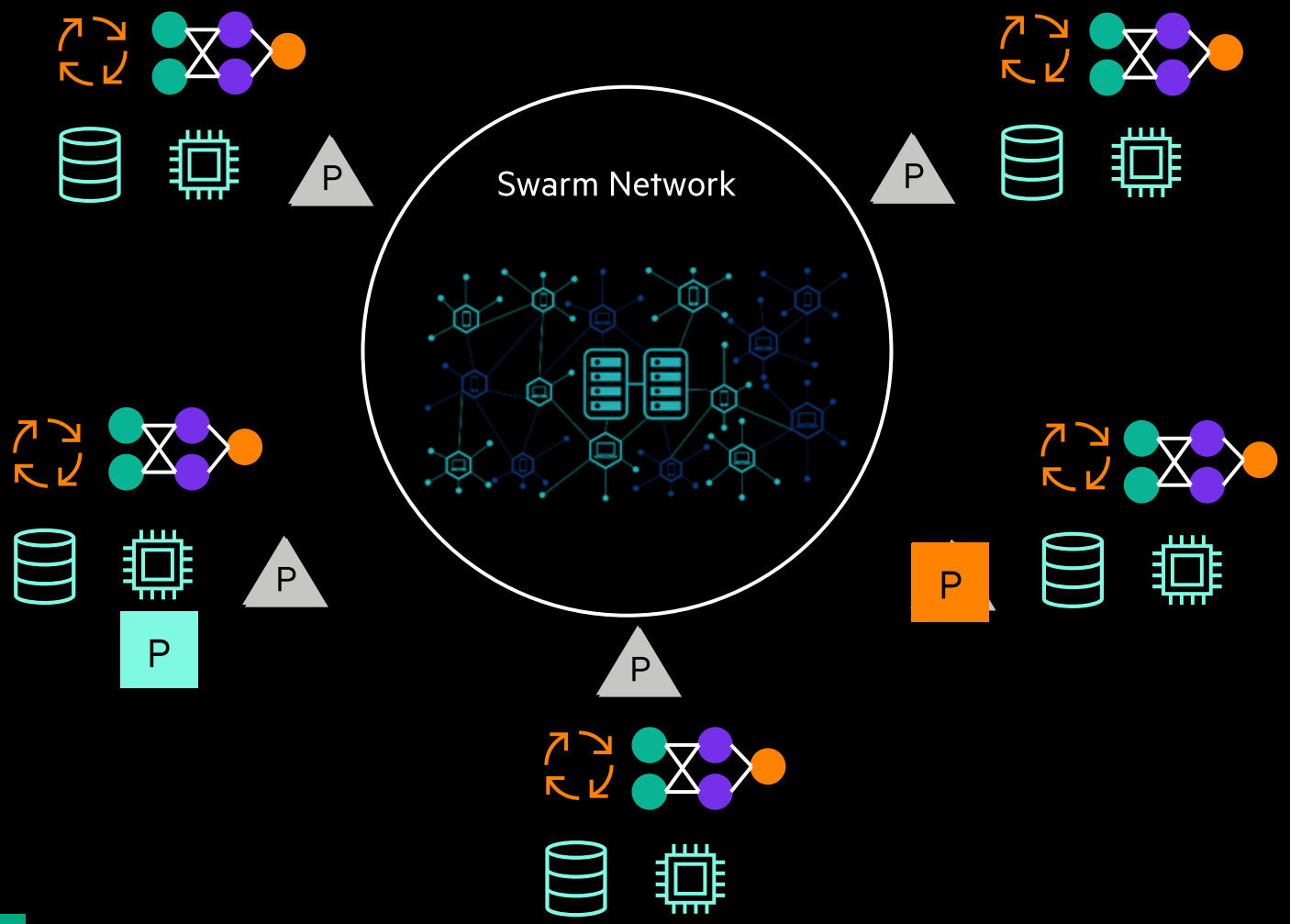
SWARM LEARNING PRINCIPLES

Democratic Machine Learning

- Equal and like-minded partners in the network
- Ownership of the data remains local
- Data protection and data security solved locally
- Less susceptible to bias in machine learning
- No control node
  - No single point of failure
  - Zero trust
  - Secure Ledger



SWARM LEARNING—PROCESS FLOW



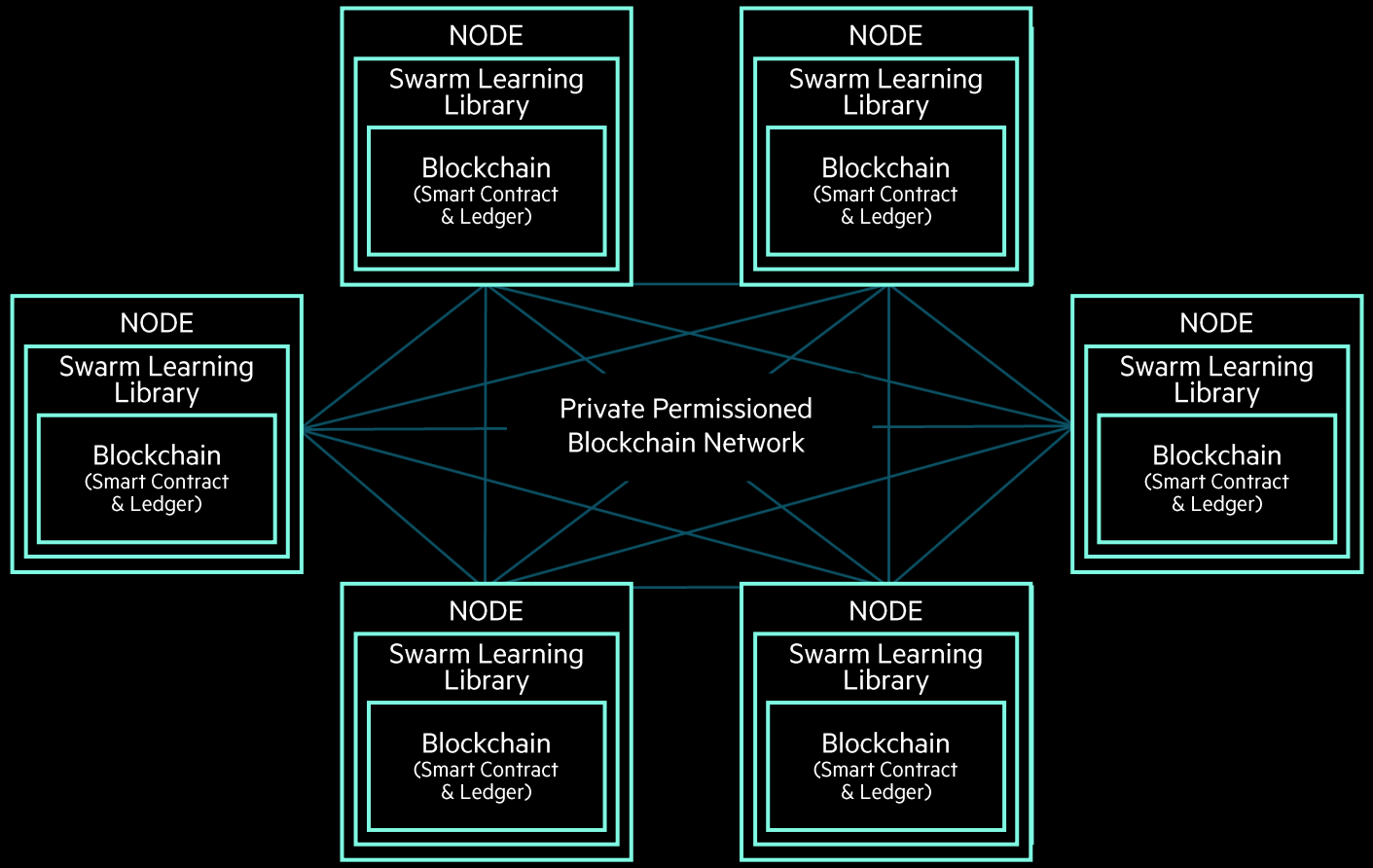
1. **Register**
   Nodes register to Swarm Network and receive ML model

2. **Train**
   Nodes train the model on local data for a time-window (epoch)

3. **Merge**
   Nodes share and merge the trained models

4. **Repeat**
   Repeat 2 & 3 until desired accuracy is achieved

SWARM LEARNING—THE NETWORK WITH BLOCKCHAIN



- Private permissioned blockchain network
- Dynamic onboarding of new nodes
- Prevents unauthorized access to network
- Prevents insider attack from semi-honest or dishonest participants
- Reduces potential for reconstruction attacks

BLOCKCHAIN PROPERTIES

- Distributed: All network participants have a copy of the ledger for complete transparency
- Programmable: “Smart Contracts” – run when certain conditions are met
- Secure: All records are individually encrypted
- Anonymous: The identity of the participants is anonymous or pseudonymous
- Immutable: Any validated records are irreversible (permanent) and can not be changed
- Time-Stamped: A recorded time stamp is written on each block
- Unanimous: All network participants agree to the validity of each of the records

HOW TO EASILY TURN ML ALGORITHMS INTO SWARM LEARNING

Integrate using Swarm callback API

```
# 1. IMPORT SWARM CALLBACK #
from swarm import SwarmCallback
…
# 2. DEFINE SWARM CALLBACK #
swarmCallback =
SwarmCallback(rv_interval=1,
min_peers=2, num_epochs=500,
val_batch_size=VALID_BATCH_SIZE)
…
# 3. ADD SWARM CALLBACK IN THE
CALLBACK LIST FOR TRAINING #
callbacks_list = [swarmCallback]
```

Actual code difference for disease diagnosis model (in Keras)

```
$ diff nih_keras.py nih_swarm.py
29a30,31
> from swarm import SwarmCallback
>
359c361,367
<     callbacks_list = [checkpoint, early]
---
>     swarmCallback = SwarmCallback(rv_interval=1,
>             min_peers=3,
>             num_epochs=500,
>             val_data=valid_gen,
>             val_batch_size=32)
>
>     callbacks_list = [swarmCallback, checkpoint, early]
```

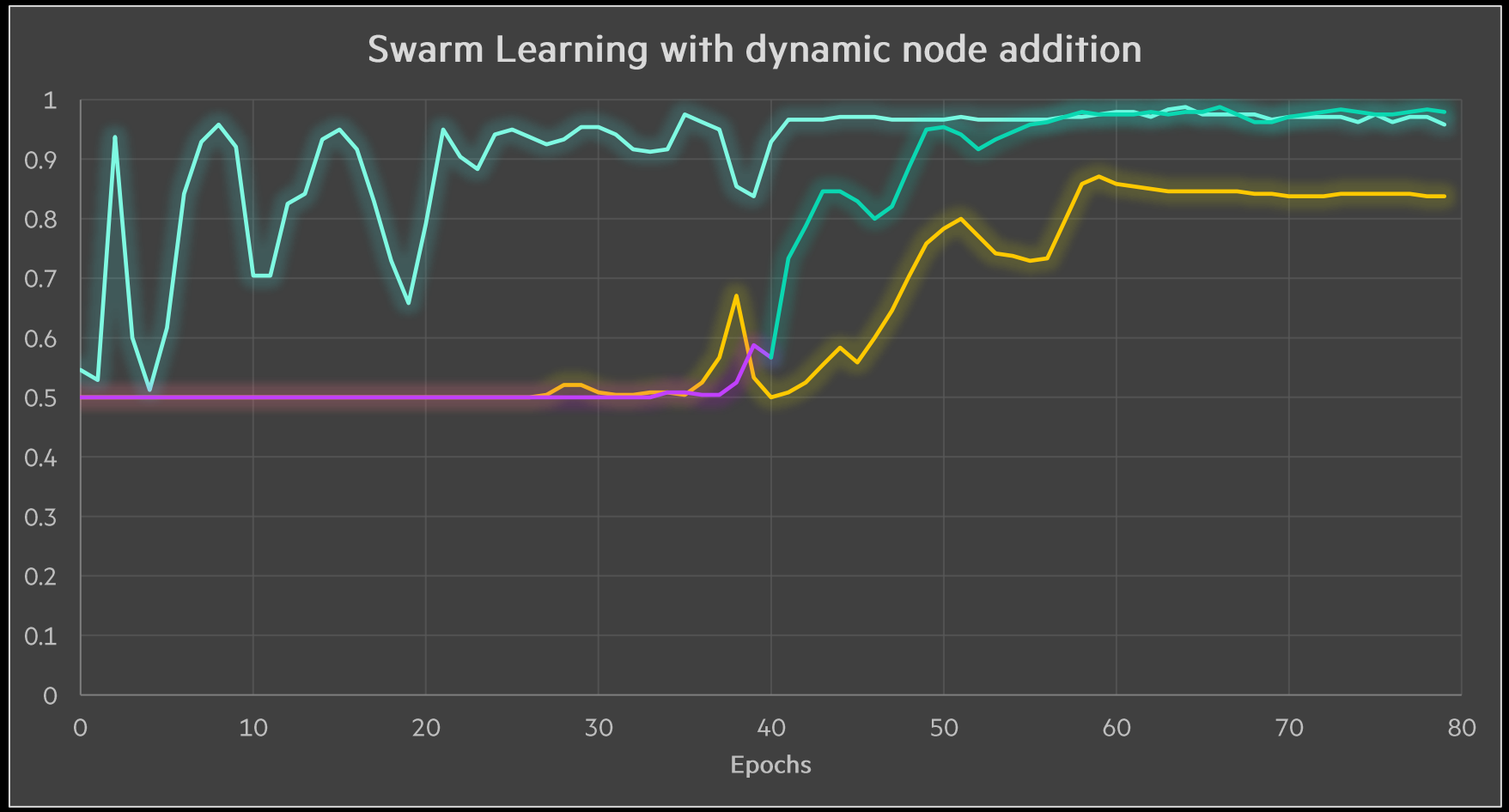
Swarm Learning Software is provided as Containers with APIs for easy integration

SWARM LEARNING SAFEGUARDS

How does Swarm Learning handle adversity?

- Dynamic swarm size
- Privacy
- Secure AI
- Secure Framework

DYNAMIC NODE ADDITION



- (a) train with 3 low prevalence nodes and hence achieves lower accuracy
- (b) train with 3 high prevalence nodes and achieves higher accuracy
- (c) 6 node scenario, starts with 3 low prevalence nodes, 3 additional nodes added around 40th epoch. Learning improves at this point and achieves higher accuracy

SECURE AND PRIVATE AI

- Privacy deals with protecting data sets from re-identification, reconstruction, tracing attacks
- Security deals with protecting models from adversarial attack and reconstruction attack
- The term secure AI is used for methods concerned with safeguarding algorithms, and the term privacy-preserving AI for systems allowing data processing without revealing the data itself. Their combination aims to guarantee sovereignty over the input data and the algorithms, integrity of the computational process and its results, and to offer trustworthy and transparently auditable technical implementations (structured transparency).
- The approach of retaining the global statistical distribution of a dataset while reducing individually recognizable information is termed differential privacy(DP)

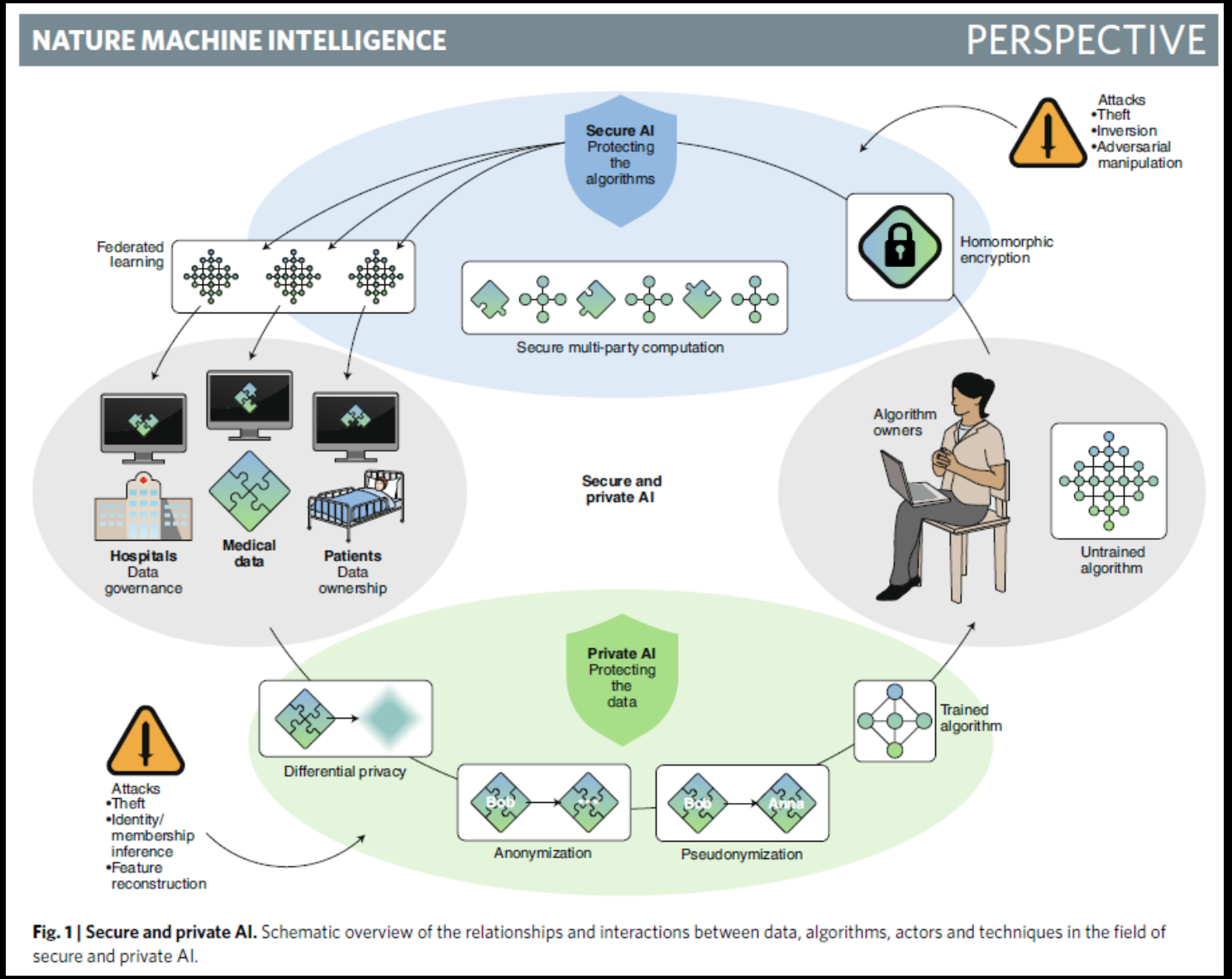


Fig. 1 | Secure and private AI. Schematic overview of the relationships and interactions between data, algorithms, actors and techniques in the field of secure and private AI.

Source: https://www.nature.com/articles/s42256-020-0186-1

"Secure, privacy-preserving and federated machine learning in medical imaging", Kaissis et. al.

PRIVACY PRESERVING AI WITH SWARM LEARNING

Category	Attack Vectors	Counter measures	Swarm Learning (Today)	Swarm Learning (Future)
Securing Datasets (raw data, data stored in model)	Re-identification (Determining an individual's identity by linkage to other known datasets)	• Access control dataset • Anonymization (removal of private data), pseudonymization (replacement of private data)	• Data is not shared outside the node. • Identity and access control with SPIFFE, OPA (https://www.openpolicyagent.org/)	• Will work with anonymization techniques (orthogonal to decentralized learning)
	Reconstruction (Determining an individual's identity from computed values without dataset access)	• Add Noise to dataset • Encrypt model weights	• Secure containers prevent unauthorized access to data and model on local node	• Support differential privacy on local data • Homomorphic encryption (HE) of weights
	Tracing (Determining Membership of an individual to a class)	• Add noise to data (data perturbation, Differential privacy)	• Secure containers prevent unauthorized access to data on local node	• Support differential privacy on local data (orthogonal to decentralized learning)

Homomorphic encryption (HE) is an encryption scheme that allows computation on encrypted data as if it was unencrypted (plain text)

SECURE AI WITH SWARM LEARNING

Category	Attack Vectors	Counter measures	Swarm Learning (Today)	Swarm Learning (Future)
Securing Model	Adversarial (systematic poisoning of inputs during model Training)	• Secure model inputs	• Private Swarm network prevents unauthorized inputs to model • Smart contracts for protection from dishonest members	
	Model-Inversion / Reconstruction (Deriving information from model weights and behavior)	• Encryption of model parameters (weights) • Prevent white-box access	• Model is accessible only within secure containers • No central / global model	• Homomorphic encryption of model weights • Compiling models to binary and other obfuscation techniques (orthogonal to decentralized learning)

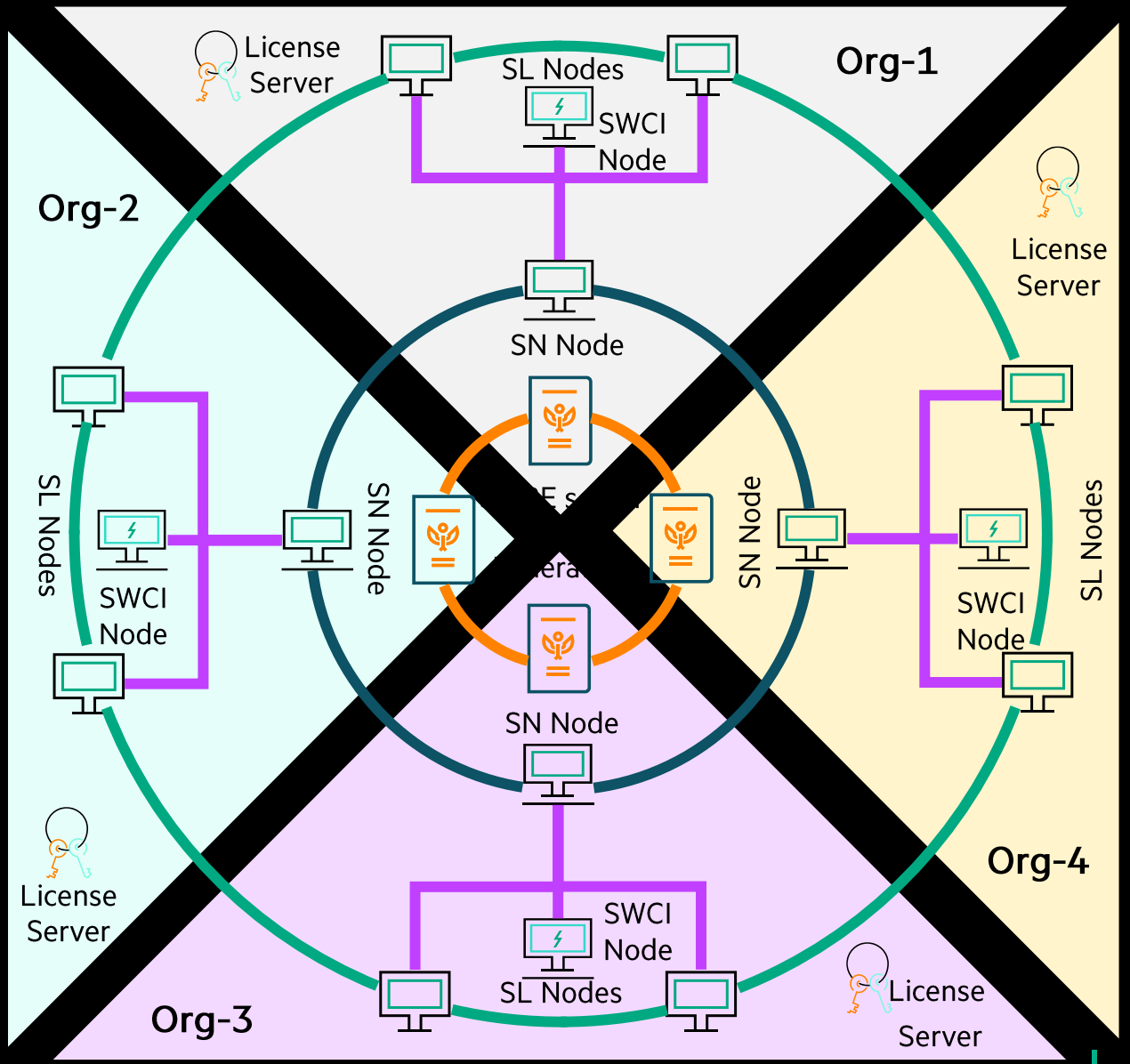
SECURE FRAMEWORK WITH SWARM LEARNING

Category	Attack Vectors	Counter measures	Swarm Learning (Today)	Swarm Learning (Future)
Secure framework (Overall Security)	Data on Move	• Authentication and Encryption	• P2P communication and no permanent sessions • mTLS for all communications	
	Containers	• Secure containers	• Hardened container – malware, vulnerability scans, code signing	• Secure route of trust • Leverage hardware-based solutions (SGX , TZ)
	Control Structure	• Access control	• private and permissioned network • Fully decentralized no central controller • Smart contract-based operations	

Mutual TLS, or **mTLS** for short, is a method for mutual authentication. **mTLS** ensures that the parties at each end of a network connection are who they claim ...

SWARM LEARNING DEPLOYMENT VIEW

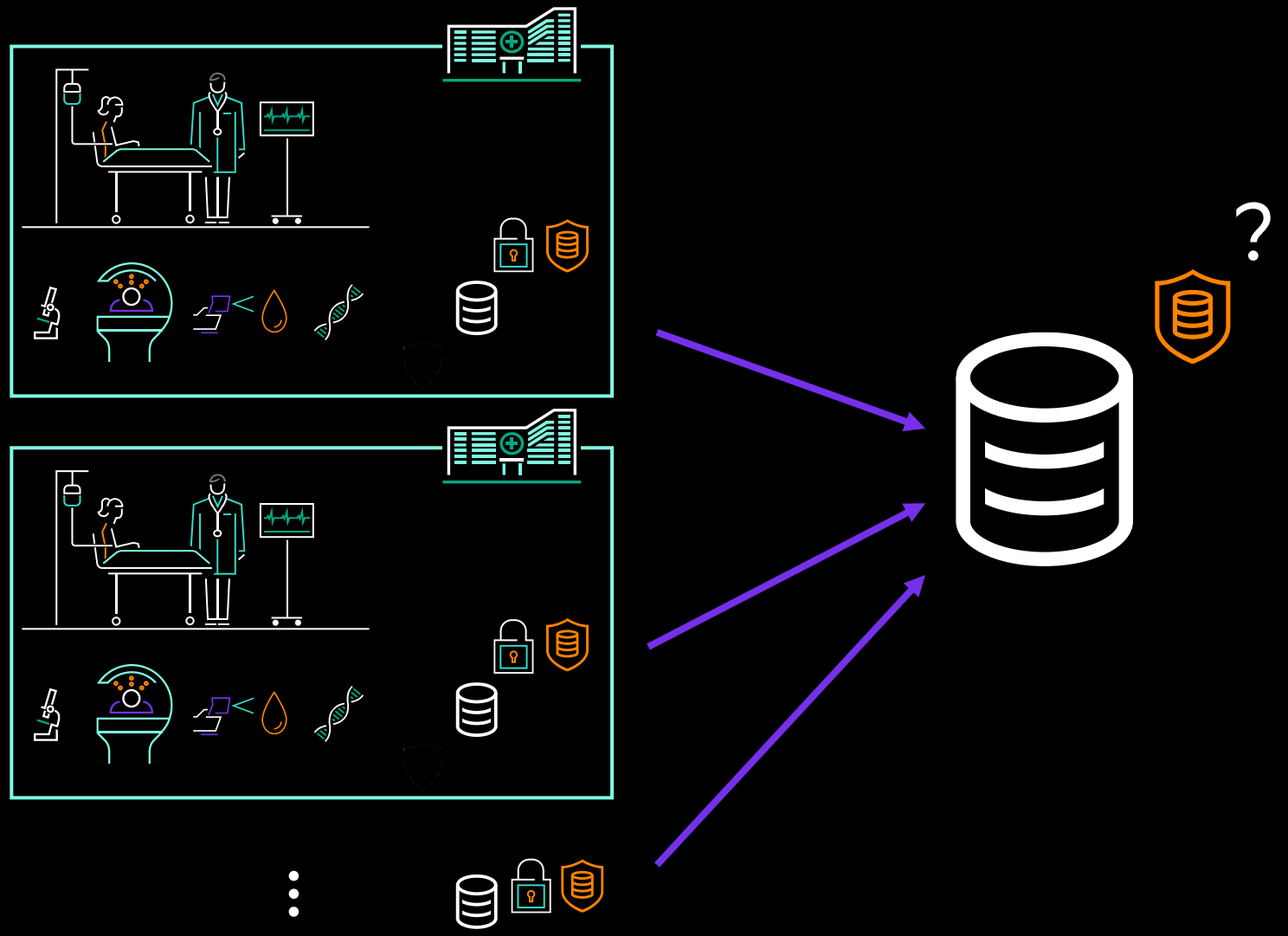
- Five components / containers
  - **Swarm Learning (SL)** – for running user-defined Machine Learning algorithm
  - **Swarm Network (SN)** - the blockchain network
  - **Spire** – security, identity integration
  - **License server** – Autopass license server
  - **SWCI** – command interface
- Support starting small Swarm Networks and then growing big by combining them



SWARM LEARNING USE CASE

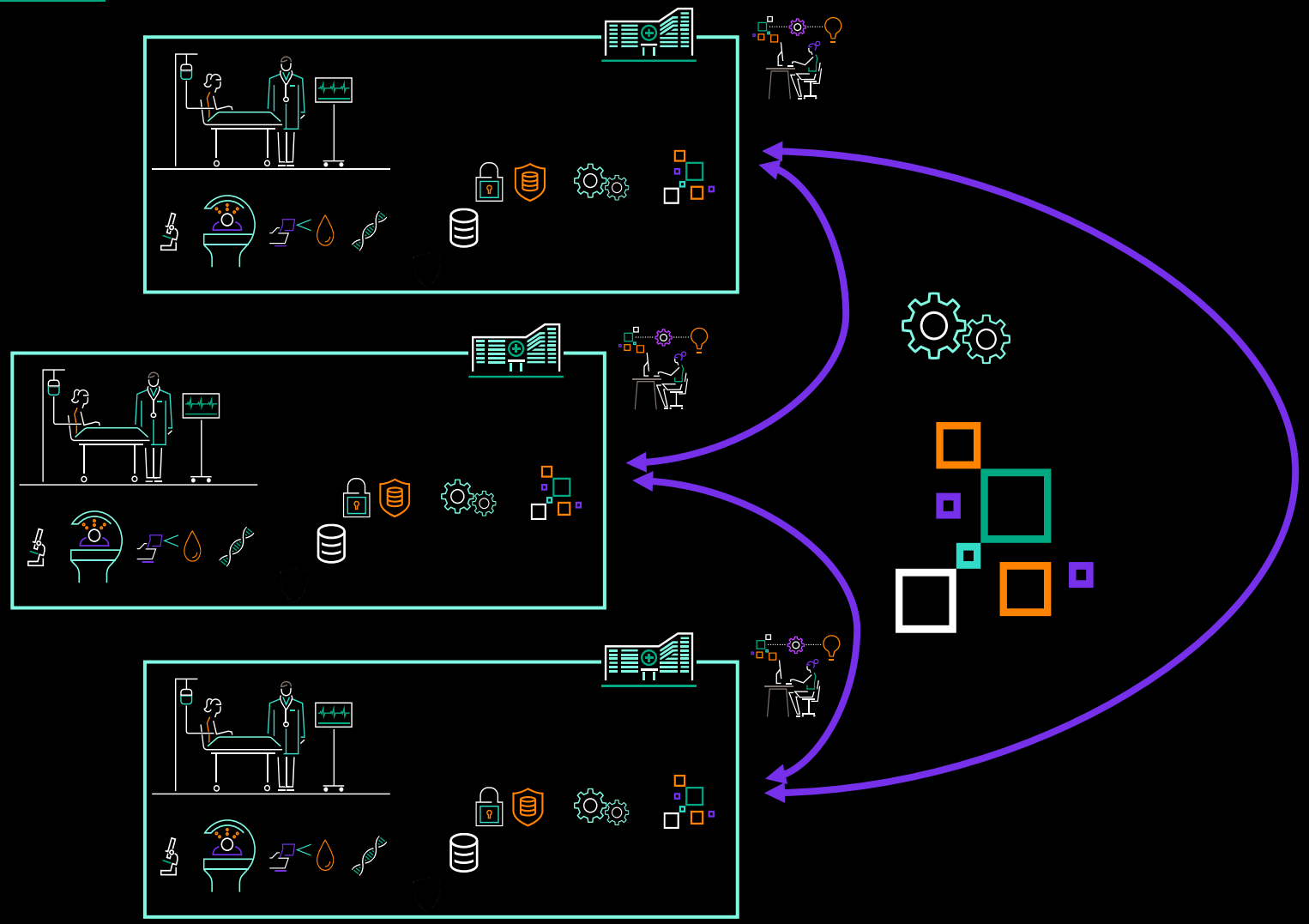
Medical

HEALTHCARE – SHARE DATA?



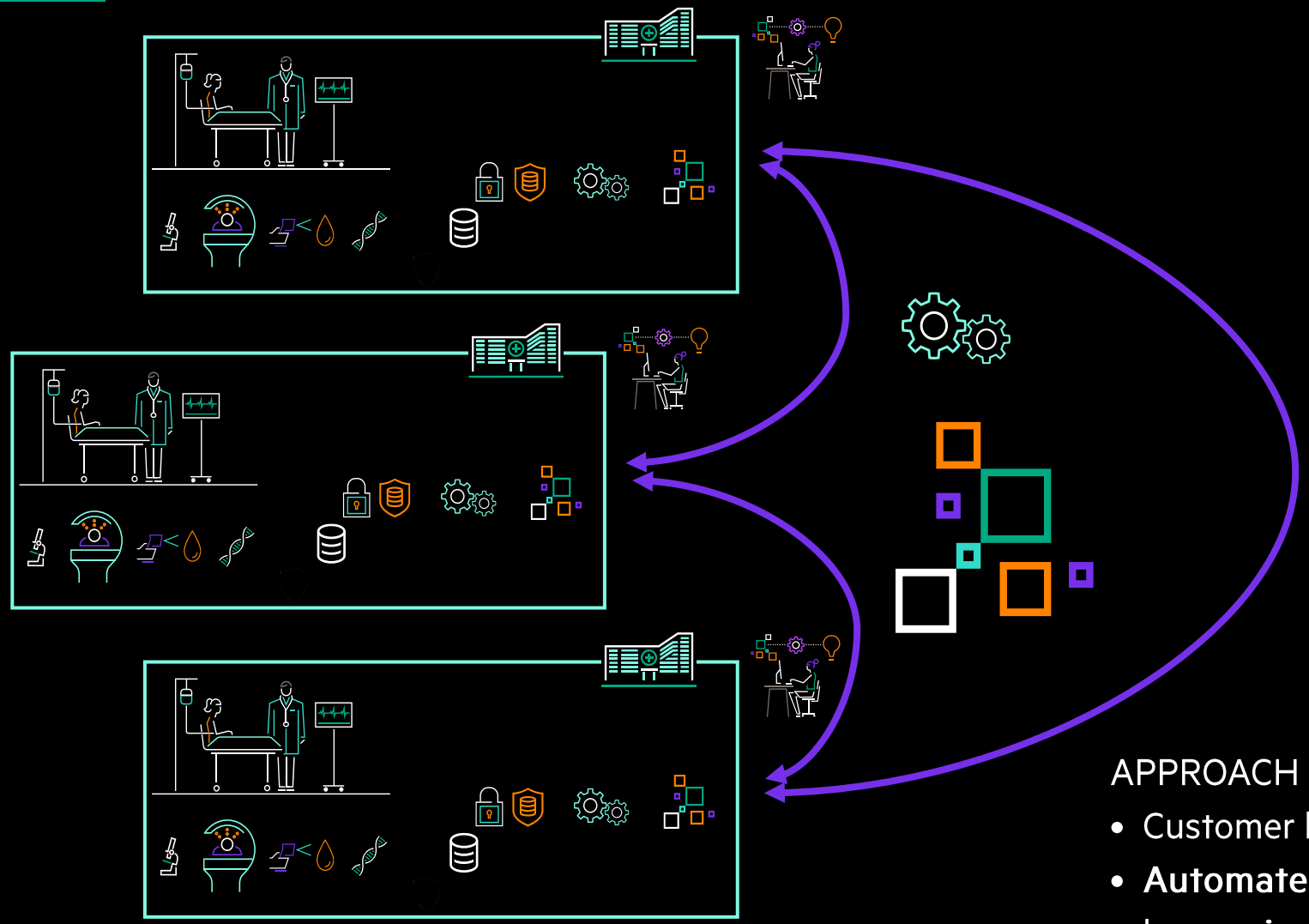
- **Data sharing, open data, centralized data** are concepts that are **diametrically** opposed to medical traditions
- Willing to share?
- Allowed to share?
  - GDPR
  - HIPAA
  - Consumer Privacy Act (CCPA)
- Agree on Data Sharing governance
- Data privacy
- Data movement & duplication

USING SWARM LEARNING



- Enables Decentralized learning
- Ideal to cope with large amount of sensor data

SWARM LEARNING IN HEALTHCARE - CHANGE APPROACH



- share insight / results
- do not move data
- do not duplicate data
- keep privacy & sovereignity
- enlarge dataset

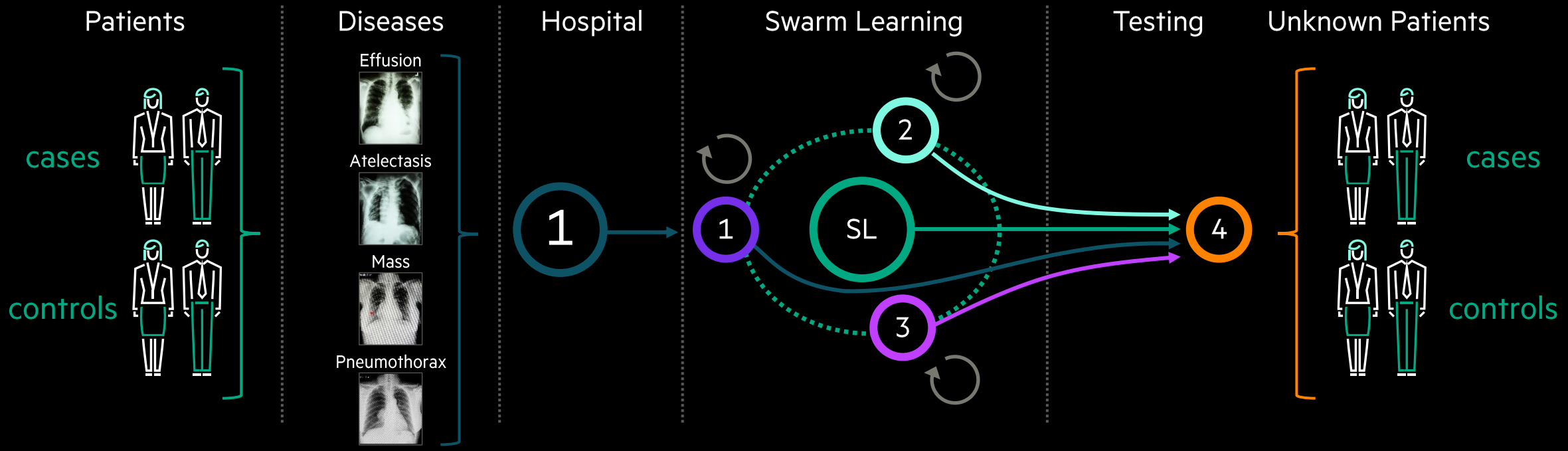
- decentralized learning

- AI will be the basis for precision medicine, in line with medical traditions and habits
- ideal to cope with medical traditions and habits
- make private data clinically usable

APPROACH

- Customer has (one or many) AI Model(s)
- **Automated decentralization** with Swarm Learning aaS
- keep **privacy** and **sovereignty** on a world wide swarm

DISEASE DIAGNOSIS WITH SWARM LEARNING—DOES IT WORK?

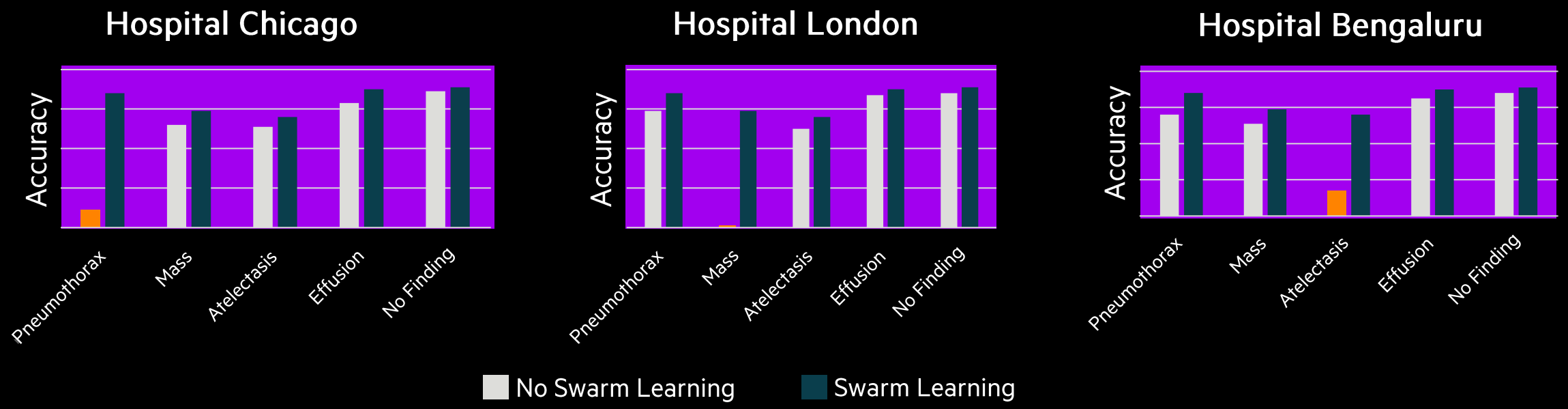


	Pneumothorax	Mass	Atelectasis	Effusion
Chicago Hospital (1)	●	●	●	●
London Hospital (2)	●	●	●	●
Bengaluru Hospital (3)	●	●	●	●

● Sufficient Images
● Limited Images

SWARM LEARNING TURNS THE DISTRIBUTED DATA INTO A COMPETITIVE EDGE

Use Case: Disease Identification Based on Chest X-rays



Models at each hospital fail to detect infrequently observed diseases

Swarm Learning model can detect those diseases where the hospital has limited data for that category

Even with sufficiently available data, Swarm Learning model is either better or at par with any individual model

SWARM LEARNING USE CASES

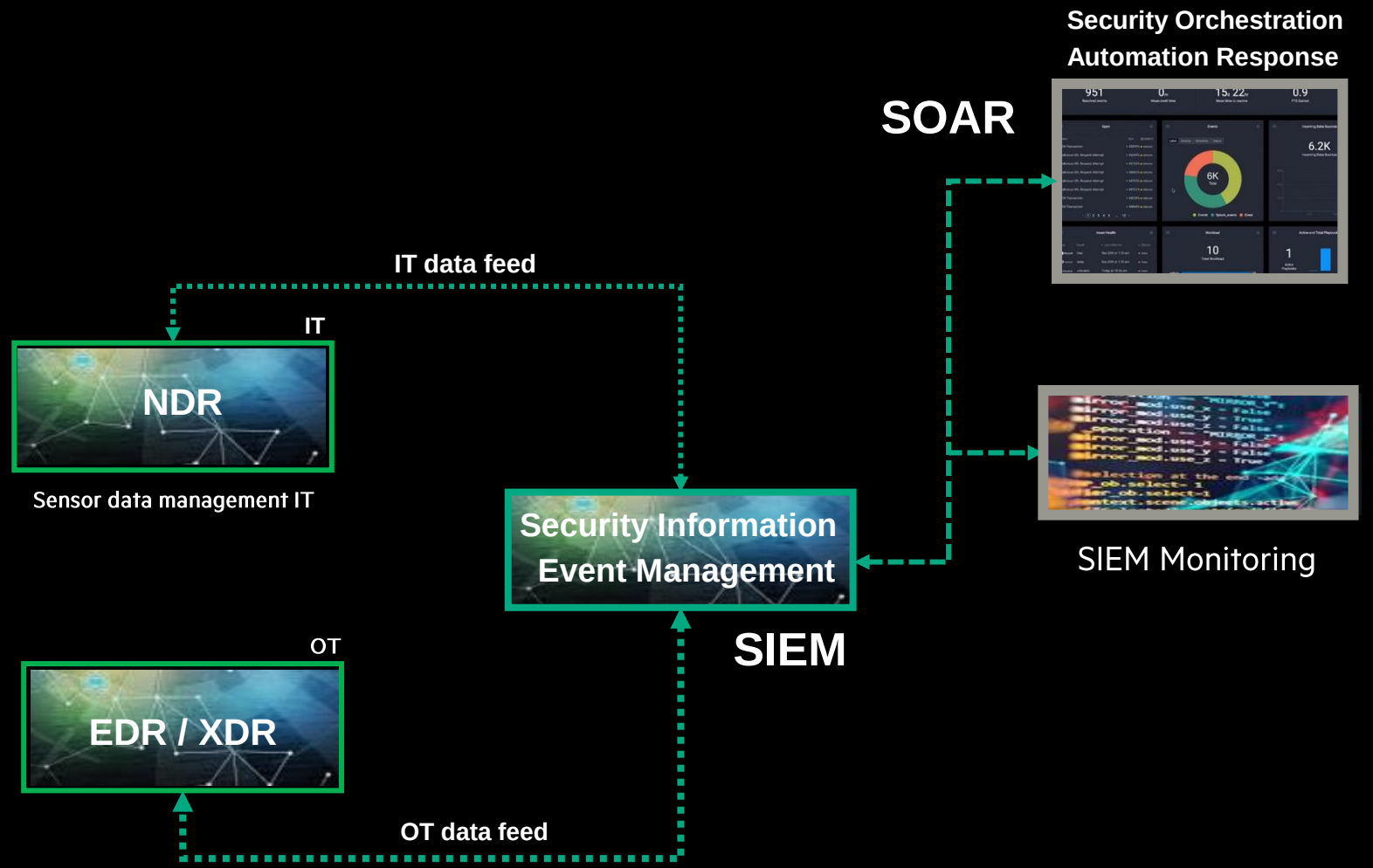
Satellite & Drones

CYBER ANALYTICS GROUND STATION FORWARDING DATA

Security Information Event Management Software

Network, Detection & Response

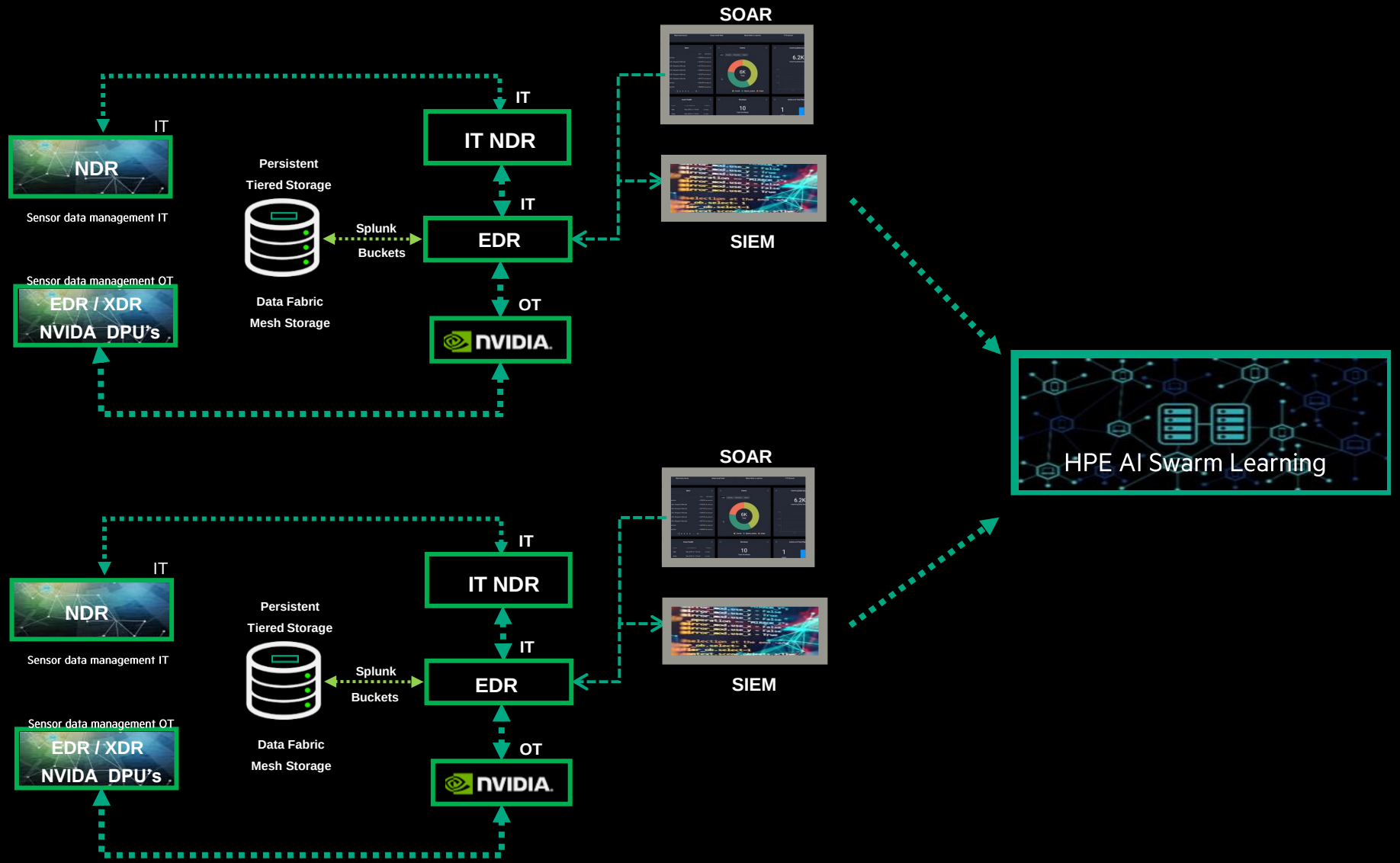
End Point, Detection & Response
+ eXtended Detection & Response



Add Swarm Learning....

Network, Detection & Response

End Point, Detection & Response
+ eXtended Detection & Response... with AI



Network, Detection & Response

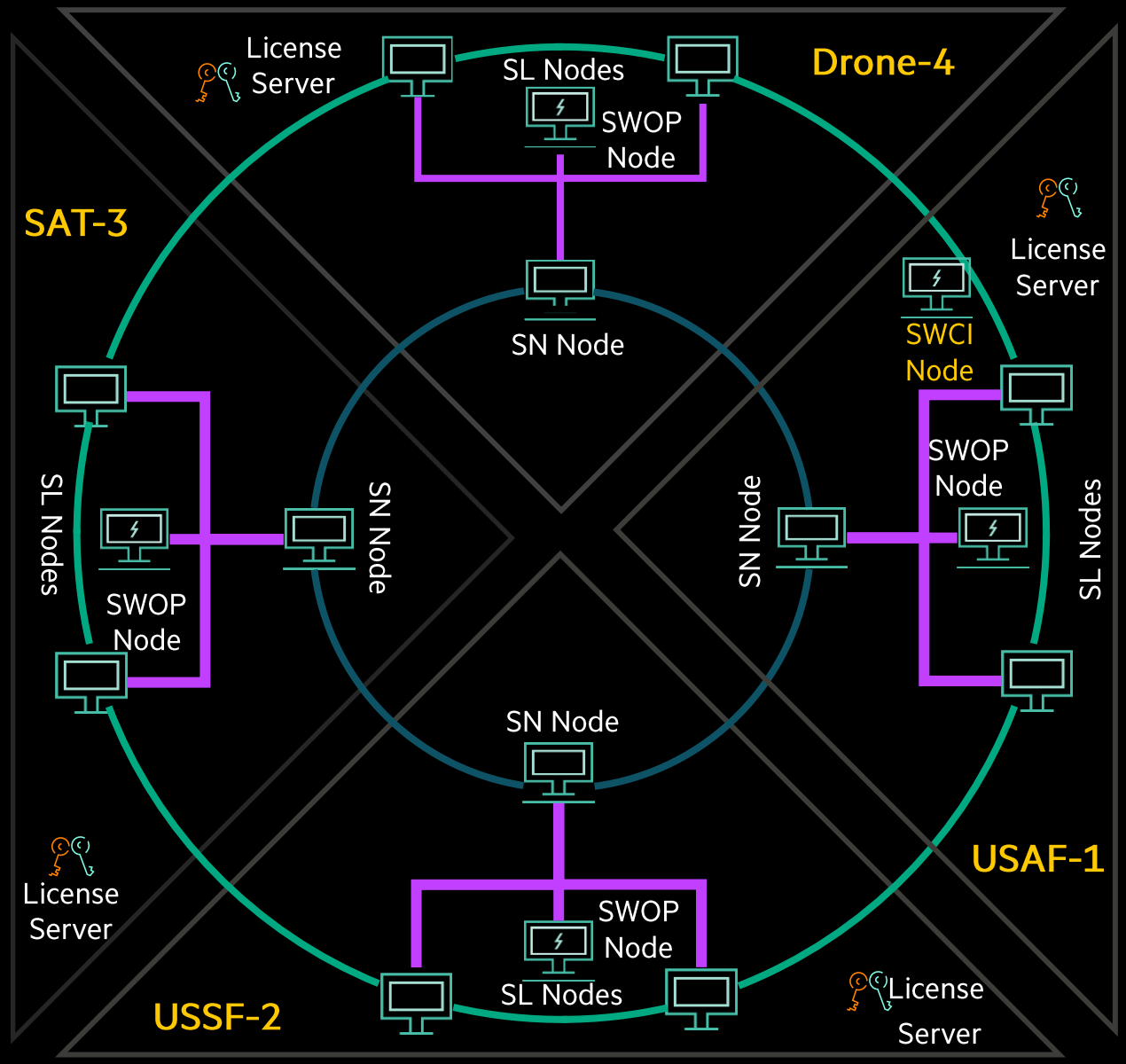
End Point, Detection & Response
+ eXtended Detection & Response... with AI

DEPLOYMENT VIEW

Only Four Components…

1. User interface Swarm Learning - define Machine Learning programs
2. Implements security with Swarm Network Blockchain - SL
3. Command Interface - SWCI
4. Command and Control Agent
   License server – SWOP

Example starting small Swarm Networks and then growing big by combining them

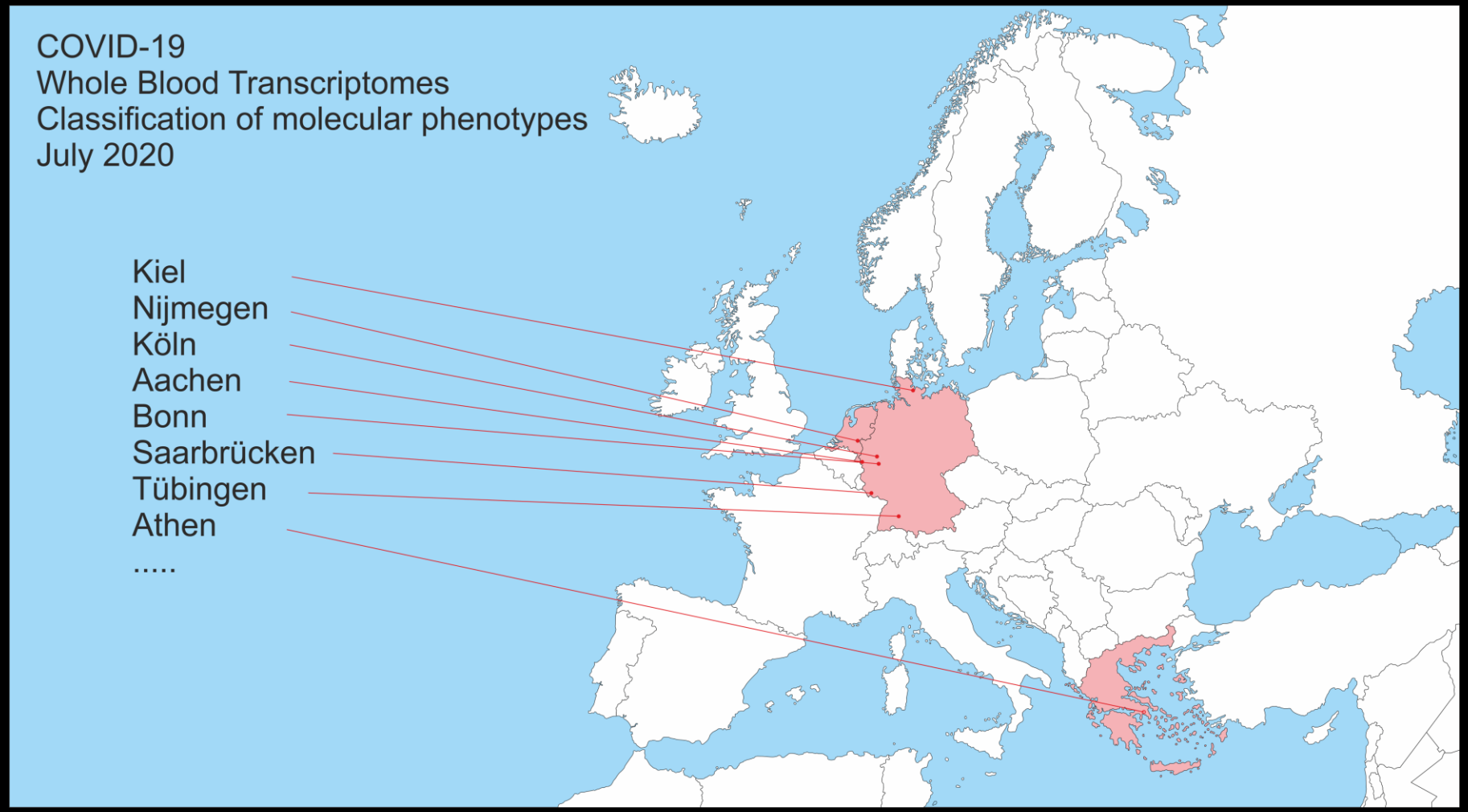


SWARM LEARNING USE CASES

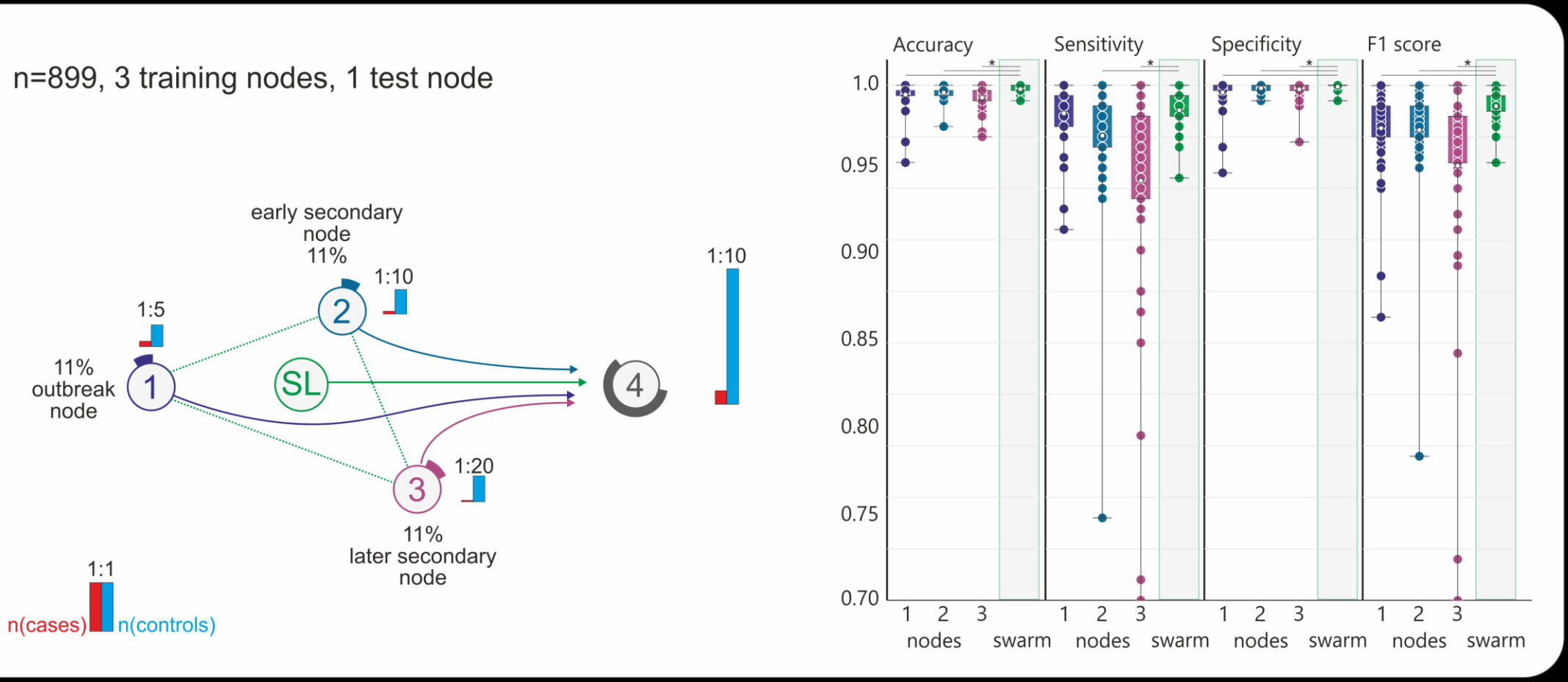
Pandemic Response

SWARM LEARNING – COVID - DISEASE BREAKOUT

Research @ DZNE



SWARM LEARNING AND COVID-19



F1 score is a performance metric for classification and is calculated by the harmonic mean of precision and recall… precision ~ accuracy… recall ~ sensitivity

MODEL METRICS

- **ACCURACY** = TP + TN / TP + FP +FN +TN
  - Model's ability to predict correctly
- **SENSITIVITY** = TP / TP + FN
  - Model's ability to predict TP
- Specificity = TN / TN + FP
  - Model's ability to predict TN

TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative

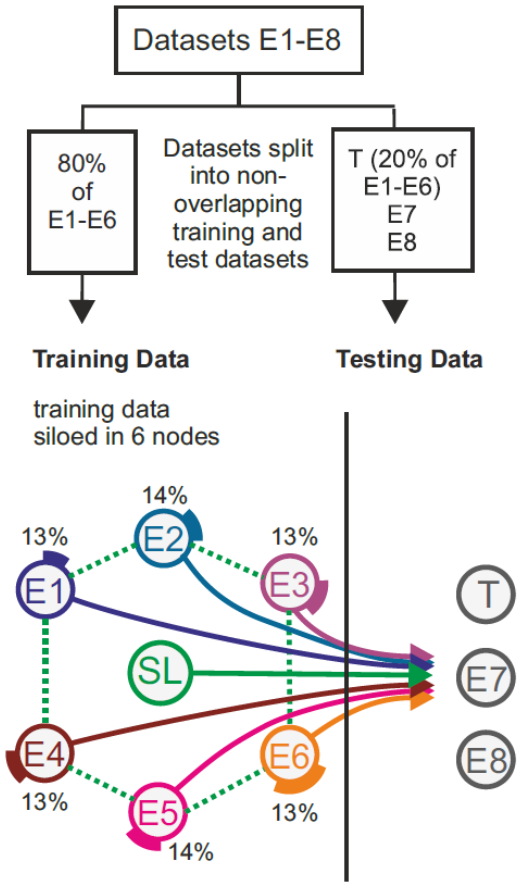
F1 score is a performance metric for classification and is calculated by the harmonic mean of precision and recall… precision ~ accuracy… recall ~ sensitivity

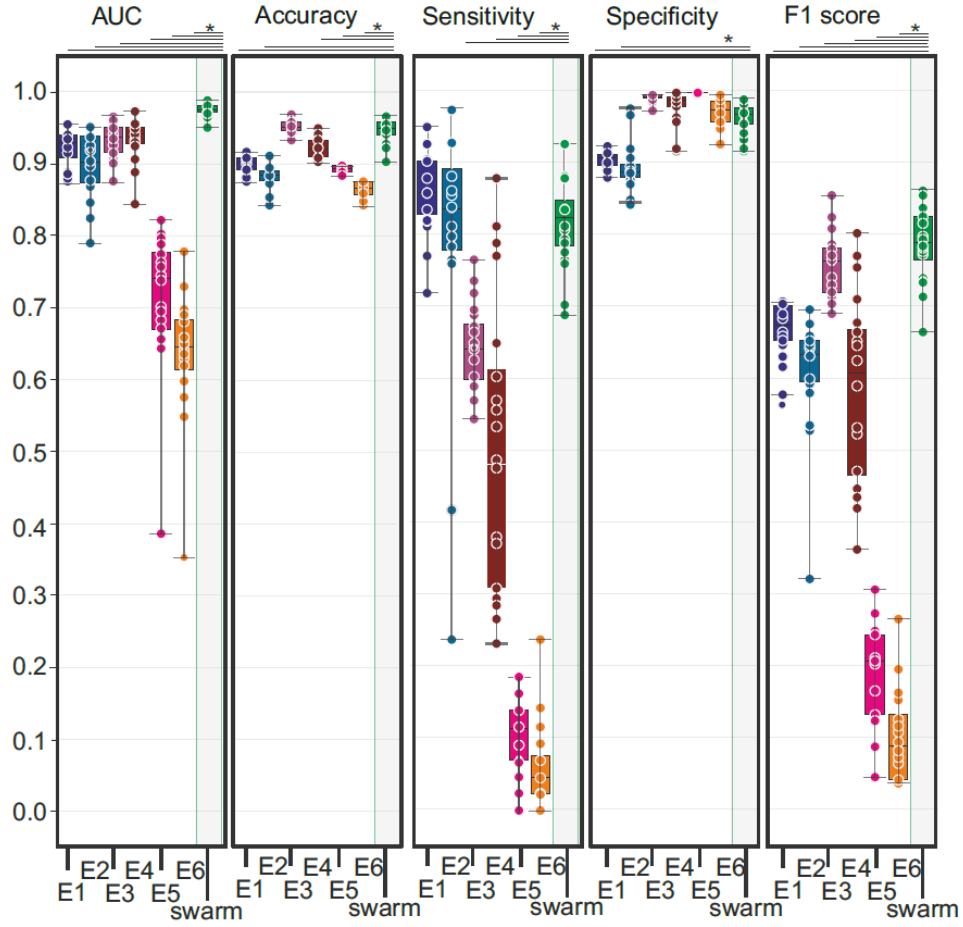
Some "believe" they should all be equal… some don't…

Confusion Matrix

OUTBREAK SCENARIO COVID-19

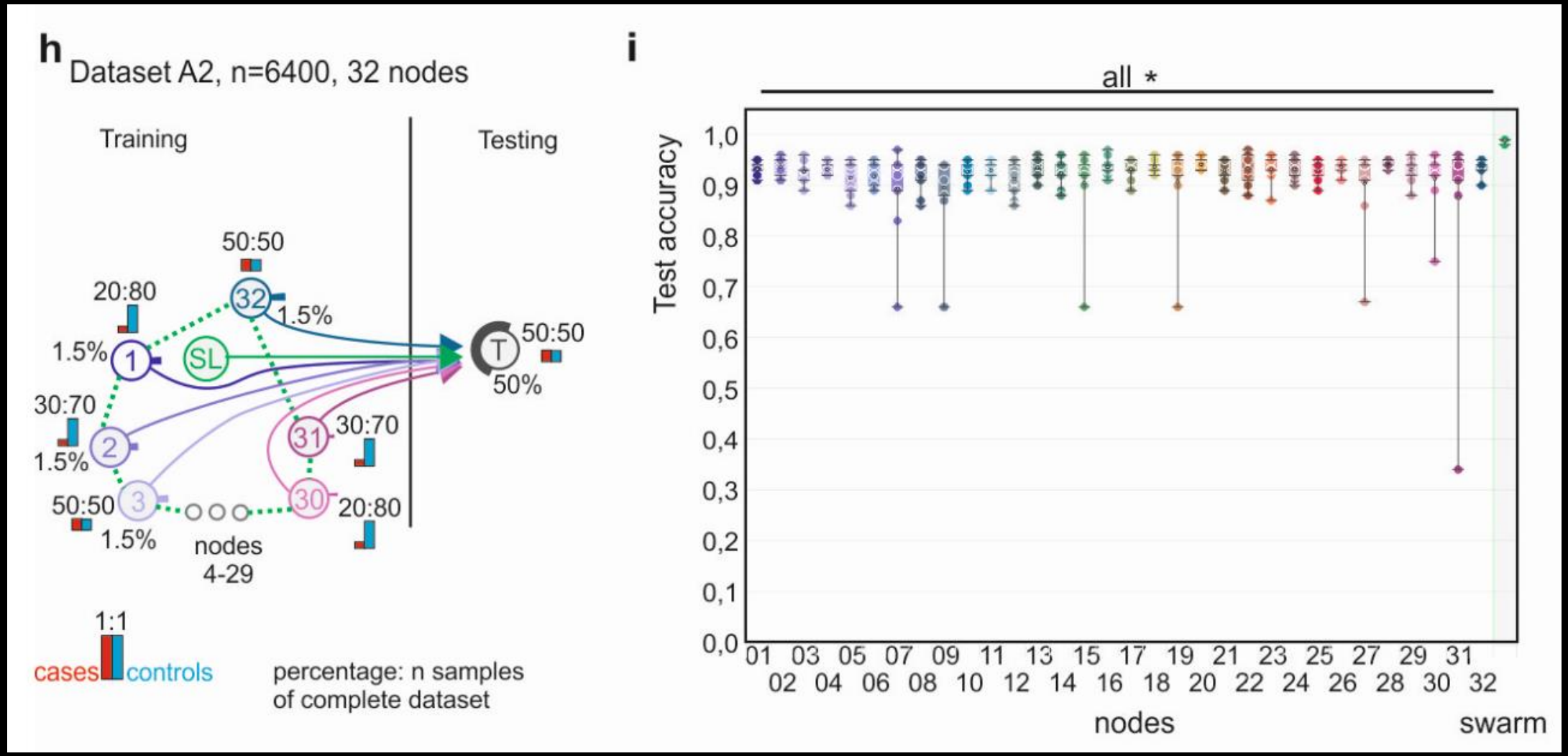
Whole blood-derived transcriptome



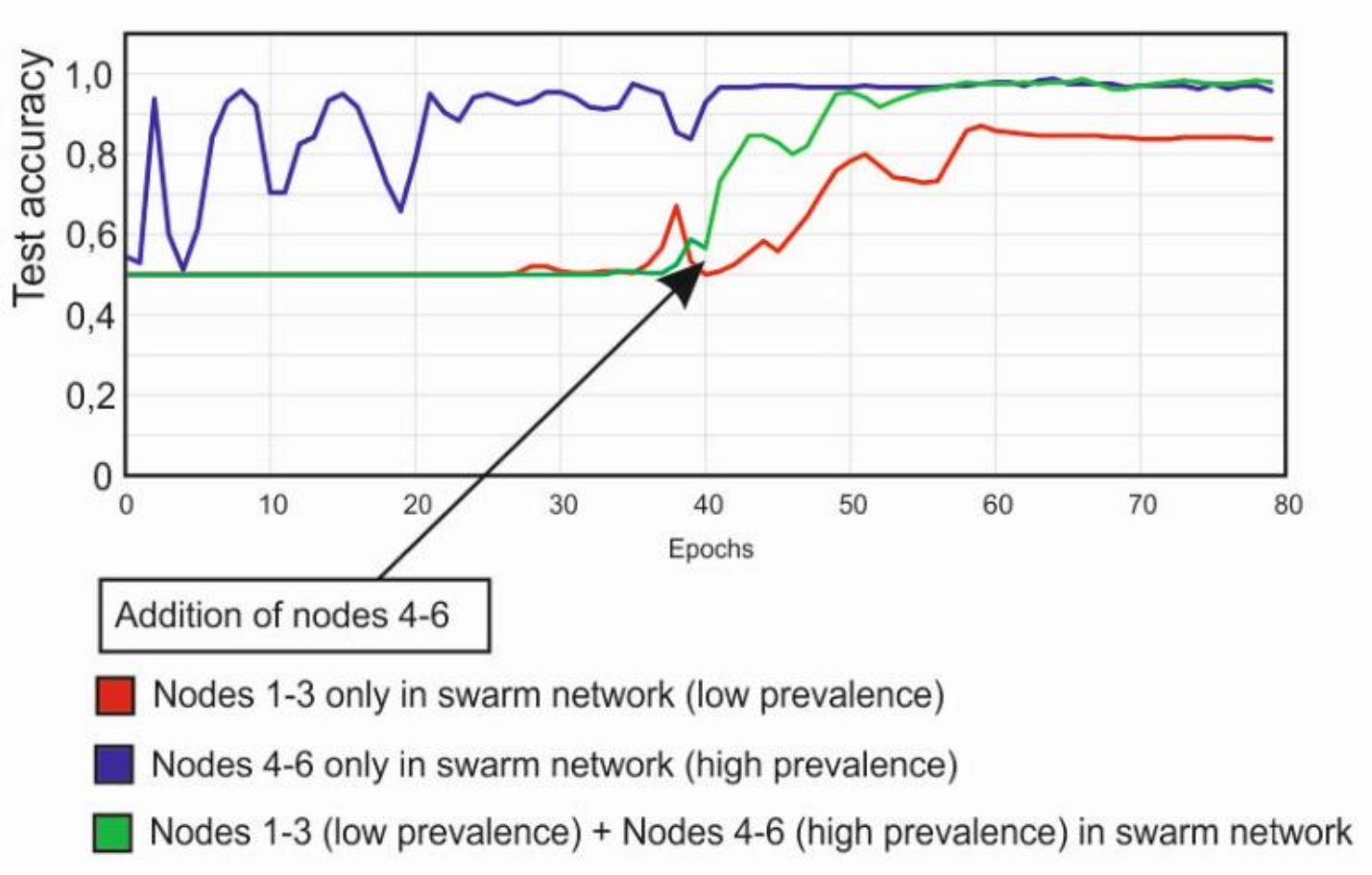


Transcriptome: a collection of all the gene readouts present in a cell… the entire collection of RNA sequences in a cell…. AUC = Model's ability to distinguish between classes (positive and negative)

MULTIPLE NODES

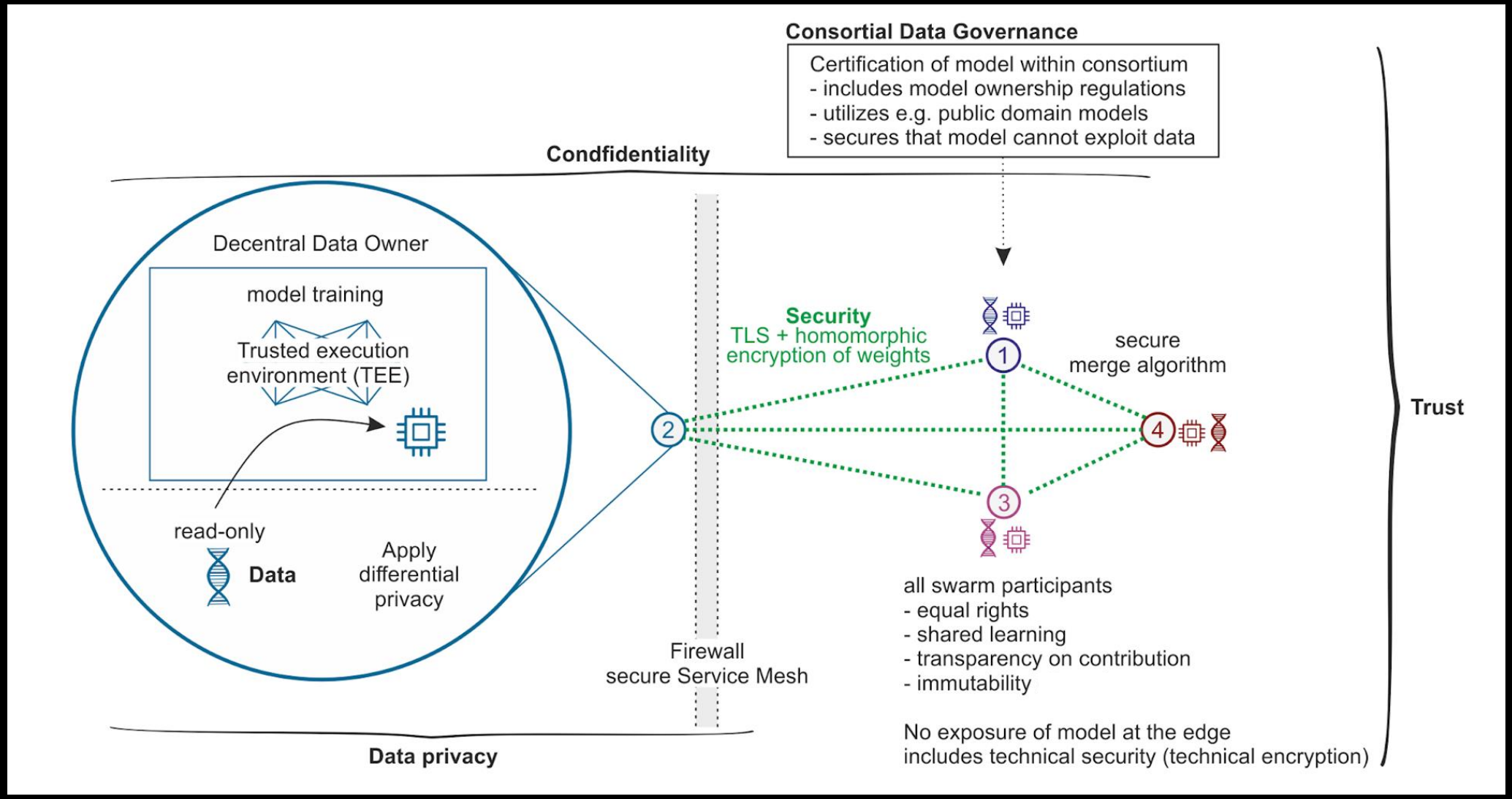


DYNAMIC ADDITION OF NODES



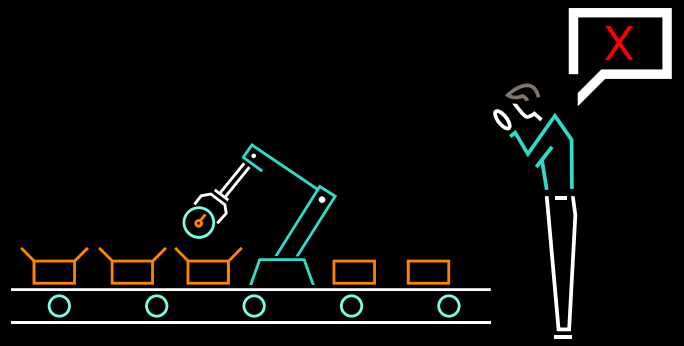
Node	Cases:Controls
Node 1	1:99
Node 2	5:95
Node 3	10:90
Node 4	30:70
Node 5	50:50
Node 6	70:30

CONSORTIAL DATA GOVERNANCE



SWARM LEARNING IN MANUFACTURING

Challenge

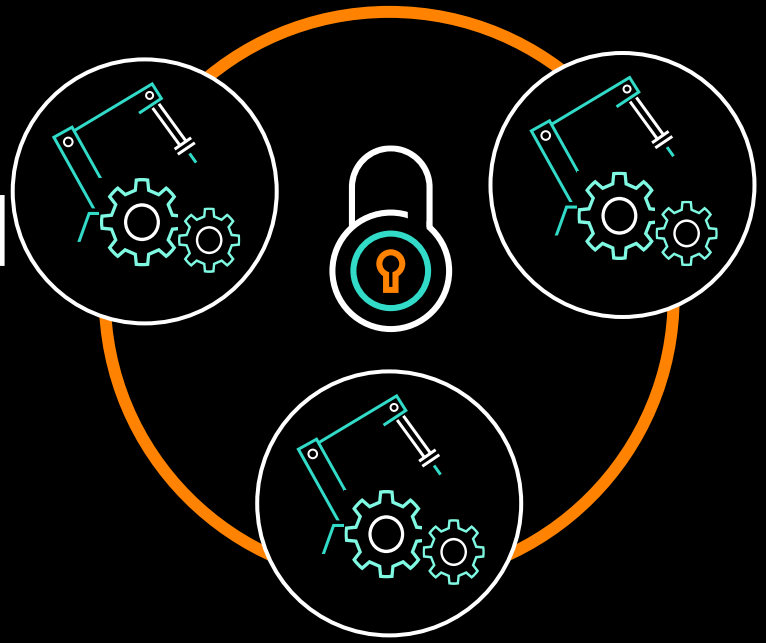


- Customer has a high scrap-rate
- They want to reduce it

Approach

- Customer offers an AI-based software that reduces scrap based on the machine's sensors
- The model learns with the customer's data
- After it can be offered to other customers aaS where it further learns based on their data

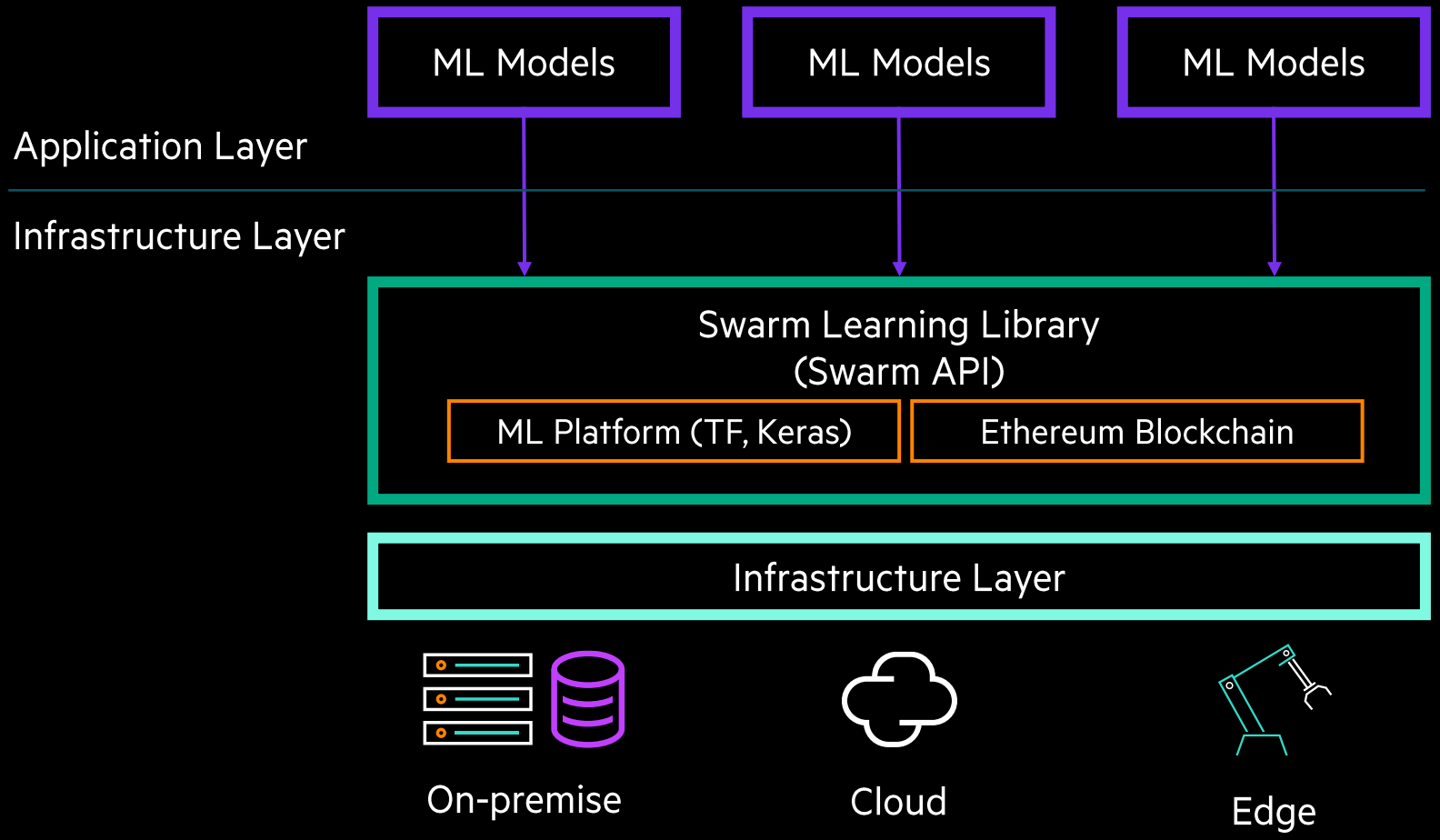
Opportunity



- Train a machine learning model in a decentralized environment
- Ensure data protection & sovereignty
- Enable a new business model

SWARM LEARNING STACK

- Swarm Learning Library (SLL) provided as containers
- Simple callback API for integration with ML models
- Tunable hyper parameters
- Management commands to control the network



Hewlett Packard
Enterprise

SWARM LEARNING RESOURCES

Jeff Winterich – DoD Account Chief Technologist

RESEARCH STUDY: DISEASE DIAGNOSIS USING HPE SWARM LEARNING

- Medical data is strongly protected by privacy laws so HPE developed Swarm Learning—a decentralized machine-learning approach that unites edge computing, blockchain-based peer-to-peer networking and coordination -- while maintaining confidentiality without the need for a central coordinator, thereby *going beyond federated learning.*

- This Swarm Learning study was executed to answer the need for fast and reliable detection of patients with severe and heterogeneous illnesses.

- For instance, patients with leukemia, TB or Covid-19 can be identified using machine learning based on their blood transcriptomes.

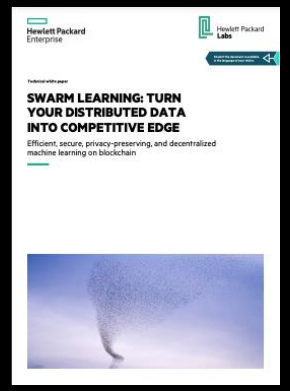
- Swarm Learning addresses the increasing divide between what is technically possible and what is allowed, due to privacy legislation.

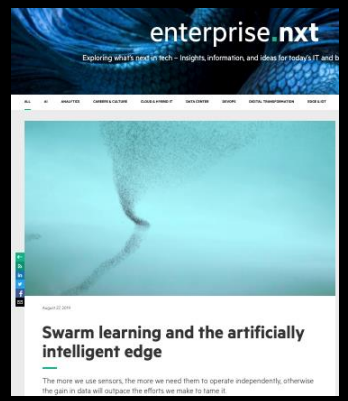
Collaboration

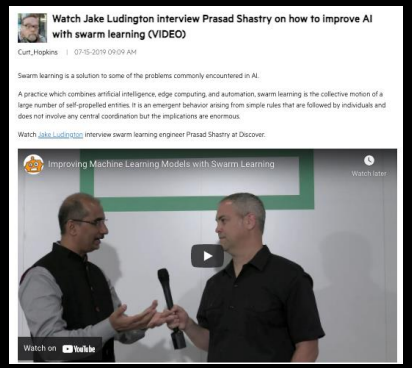
- DZNE - Systems Medicine, Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE), Bonn, Germany
- U of Bonn - Genomics and Immunoregulation, Life & Medical Sciences (LIMES) Institute, University of Bonn, Bonn, Germany
- HPE - Hewlett Packard Enterprise, Houston, TX, USA

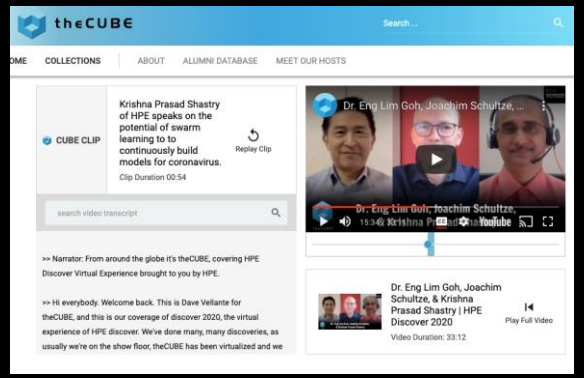
Article citation: Warnat-Herresthal, S., Schultze, H., Shastry, K.L. *et al.* Swarm Learning for decentralized and confidential clinical machine learning. *Nature* **594,** 265–270 (2021). https://doi.org/10.1038/s41586-021-03583-3

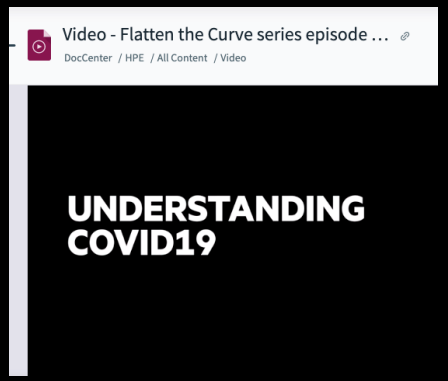
SWARM LEARNING LINK LIBRARY (THUMBNAILS ARE HYPERLINKED)

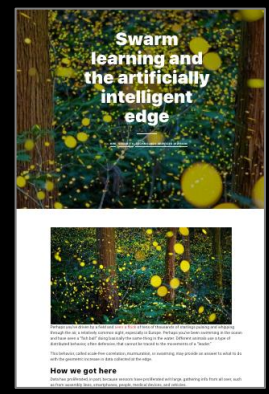


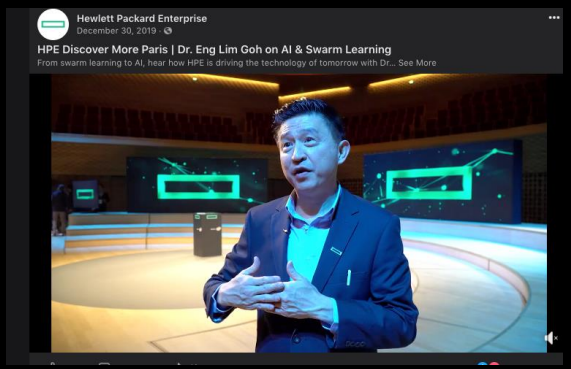


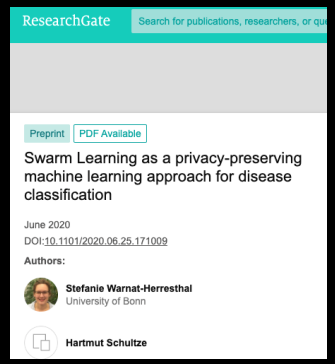


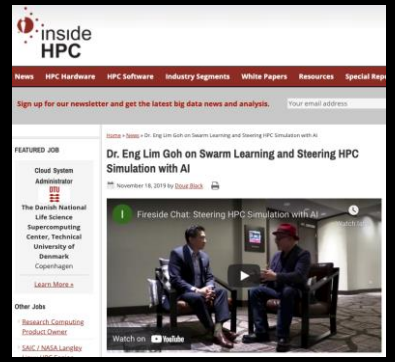


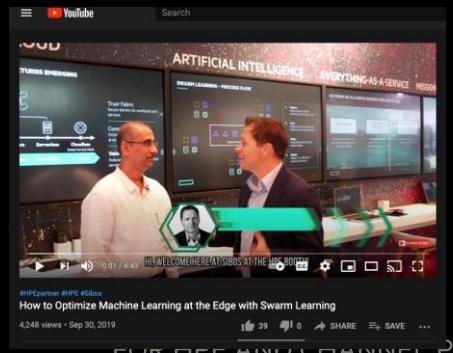












SWARM LEARNING—ADDITIONAL RESOURCES

EXPLORE THESE VIDEOS

- The Big Shift: Healthcare Research Redefined
- The Big Shift: What is Swarm Learning?
- The Big Shift: Meet the Car of the Future

Nature Article: Swarm Learning for decentralized and confidential clinical machine learning

https://www.nature.com/articles/s41586-021-03583-3

ACCESS TO SWARM LEARNING EVALUATION LICENSE

Swarm Learning Evaluation License: https://myenterpriselicense.hpe.com/


THANK YOU
Jeff Winterich, HPE Chief Technologist, AI Ambassador