NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

CAreFL: Contribution-Aware Federated Learning for Smart Healthcare

Zelei Liu
zelei.liu@outlook.com

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE WeBank 微众银行 医渡云 YIDUCLOUD AIR 香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

About Me

Liu Zelei

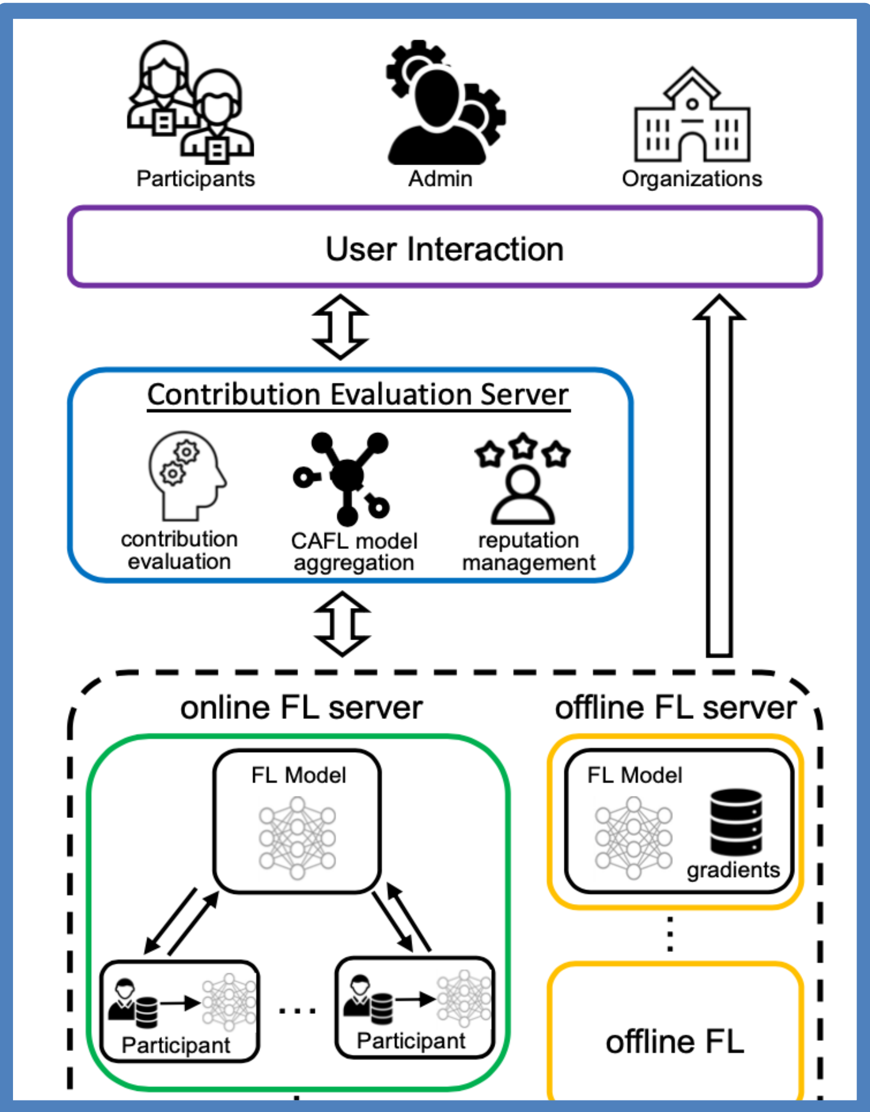
- 2019/10-2022/10, **Research Fellow** @ SCSE, NTU
- 2019: **PhD** in CS from CCST, JLU

Research Interest:

  – Incentive Mechanism Design for FL
  – Fairness in Ethical AI

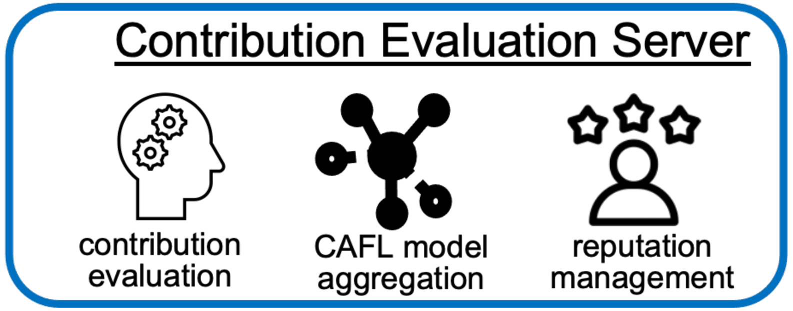
CAreFL Overview

1. FL infrastructure
2. Contribution Evaluation
3. User Interaction



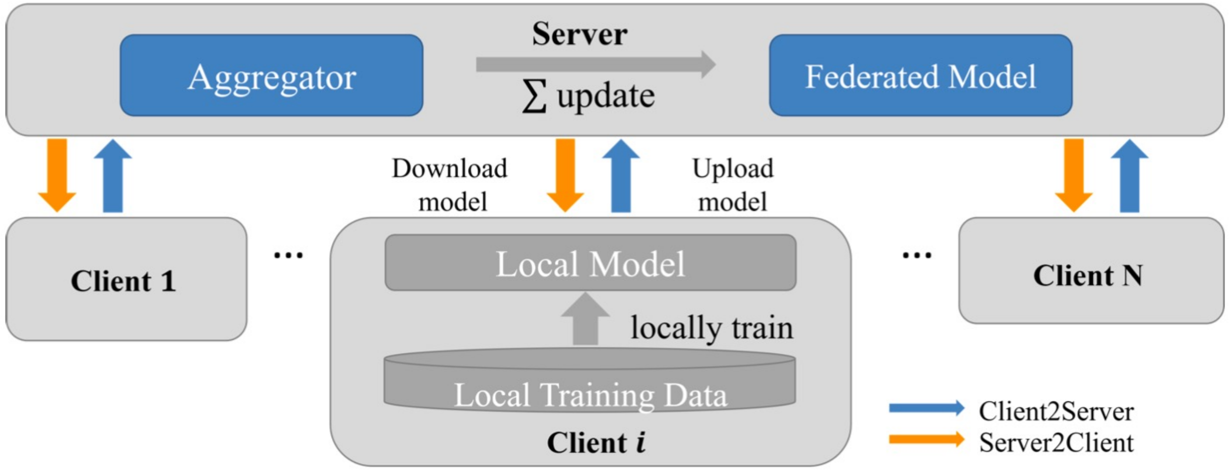
CAreFL Overview

- CAreFL: a HFL framework focusing on Contribution in FL.



enable

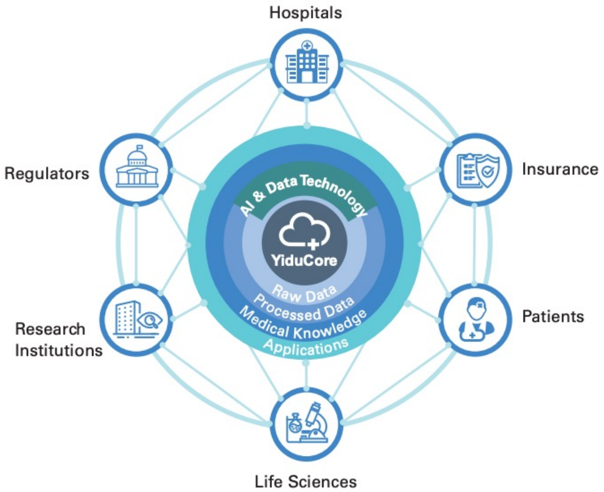
Traditional HFL framework



New Solution for Smart Healthcare

医渡云
YIDUCLOUD

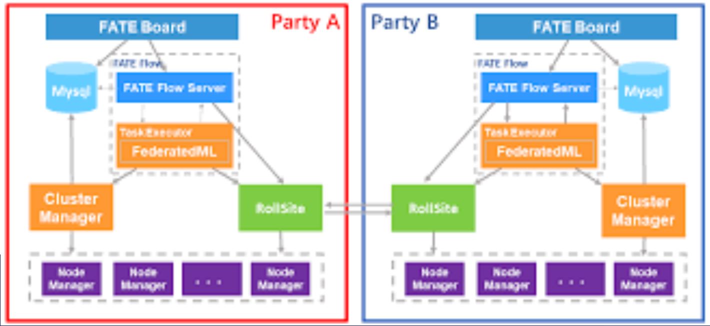
Healthcare Big Data Platform

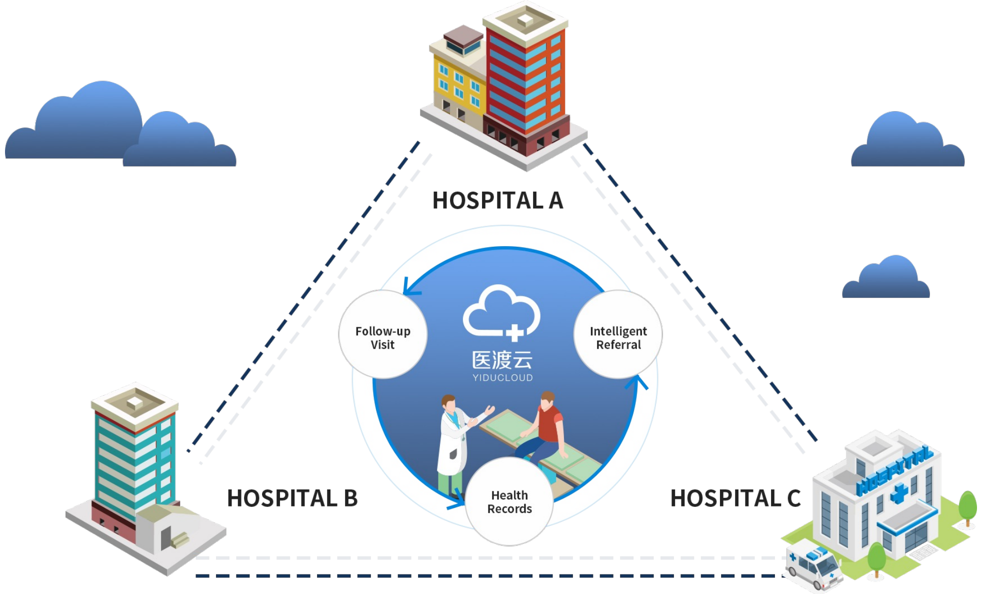


WeBank
微众银行

Federated Learning framework

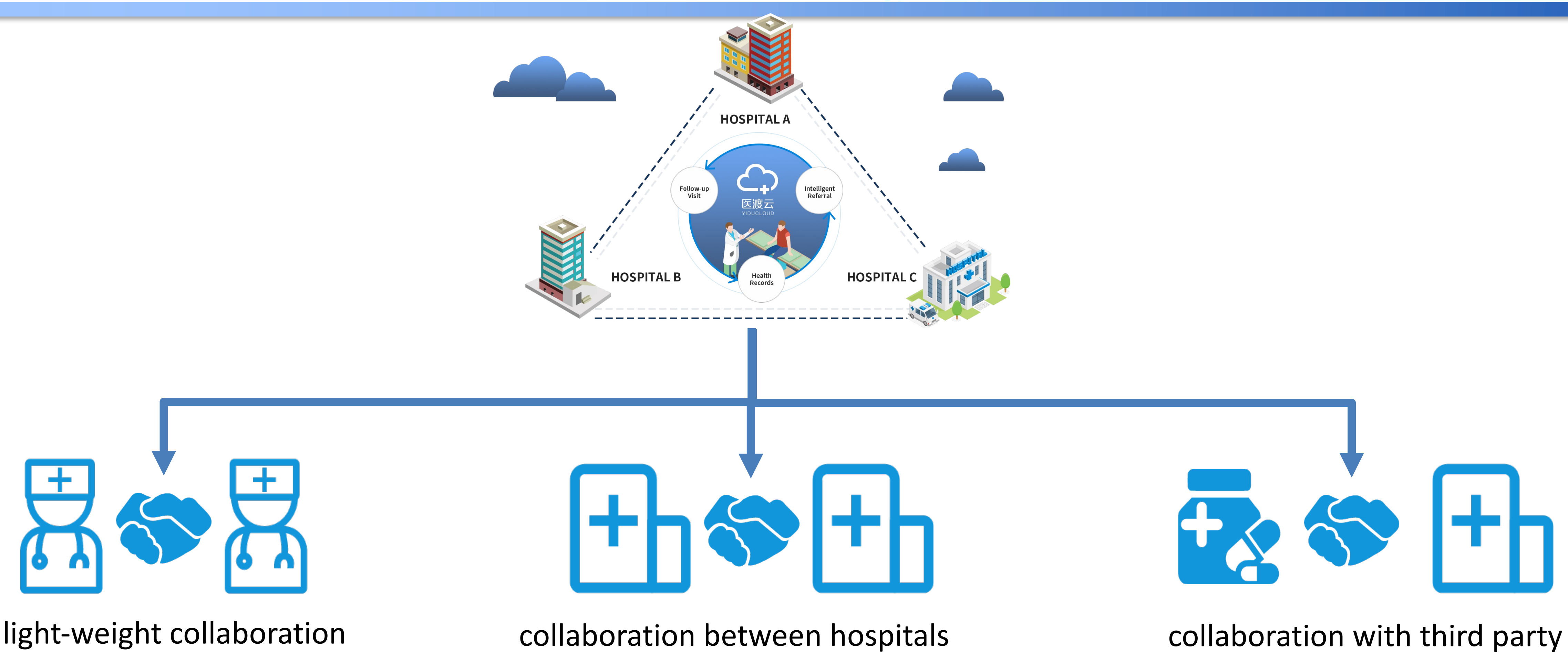
FATE





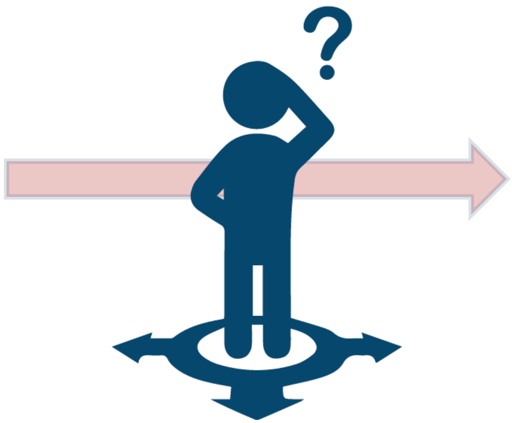
New Collaboration Solution

Diverse Collaborations



Key Concerns - Contribution

For example, a pharmaceutical company may wish to build a model to facilitate drug research by leveraging data from multiple hospitals through FL. In order to compensate the participating hospitals, the pharmaceutical company may need to offer incentive payouts. How to fairly allocate the compensation?



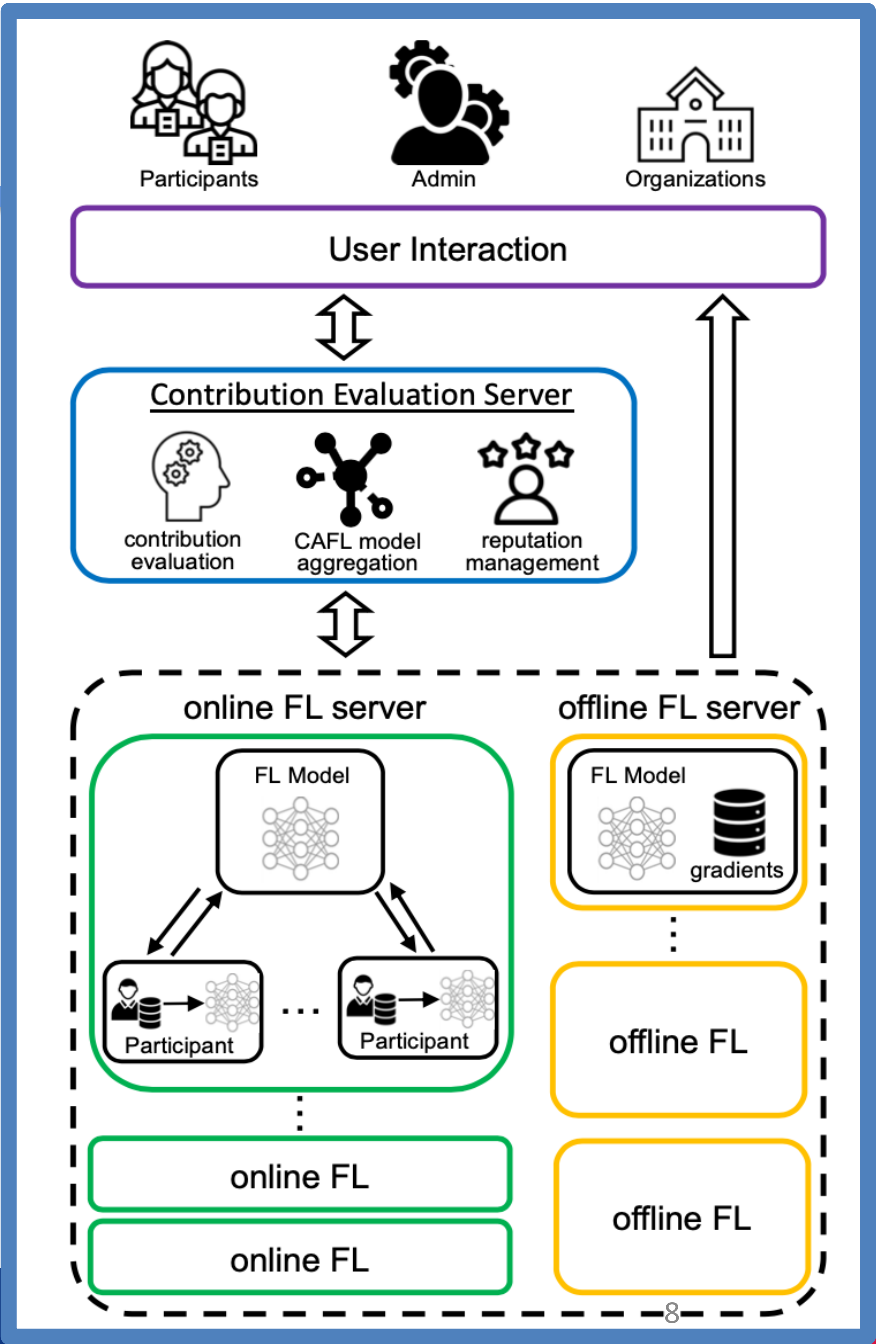
Fair Contribution Evaluation.

AI in CAreFL

- Focus on Contribution Evaluation

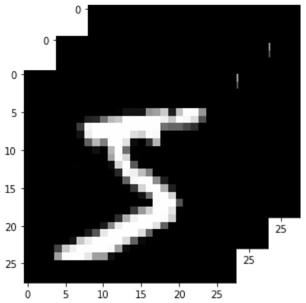
Fast and Accurate Contribution Evaluation

Contribution-Aware FL Model Aggregation



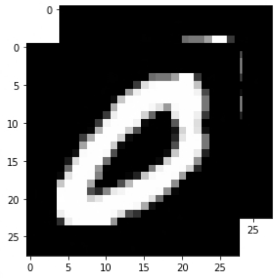
Contribution Evaluation Obstacle

- Quantity, Quality, Label Quality



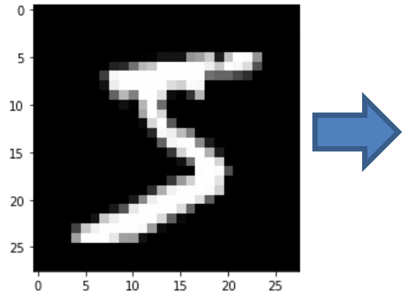
300 images

V.S.



200 images

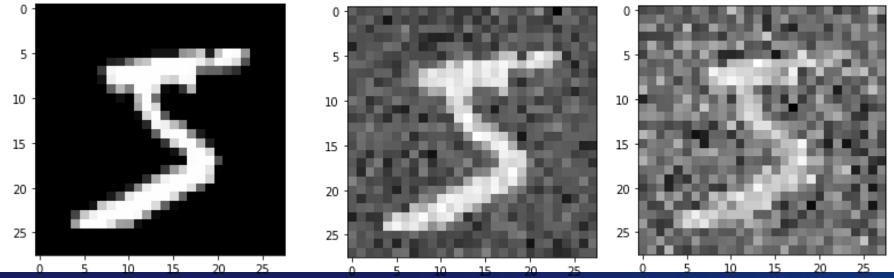
"Quantity"



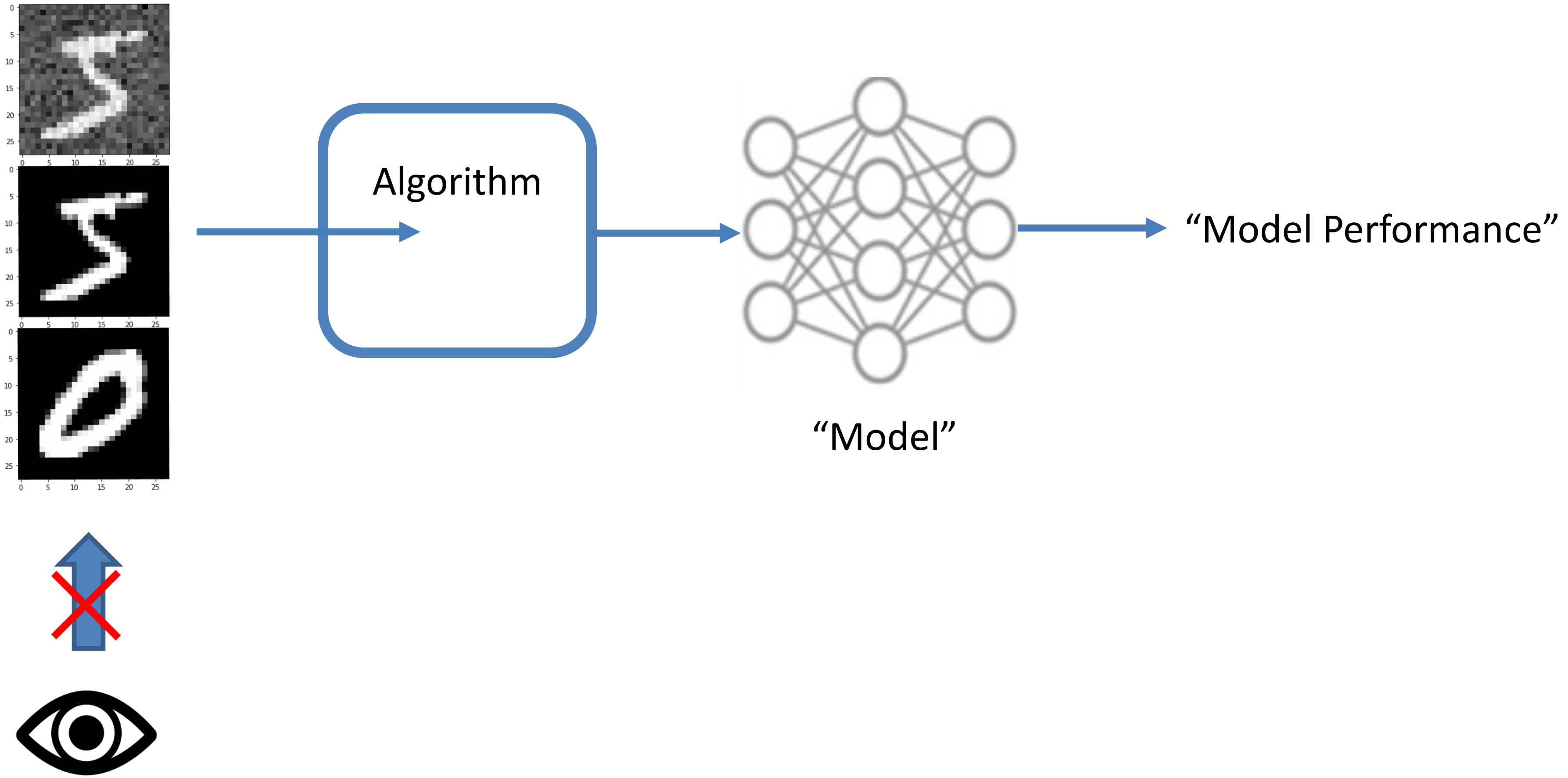
wrong label "1"

"Label Quality"

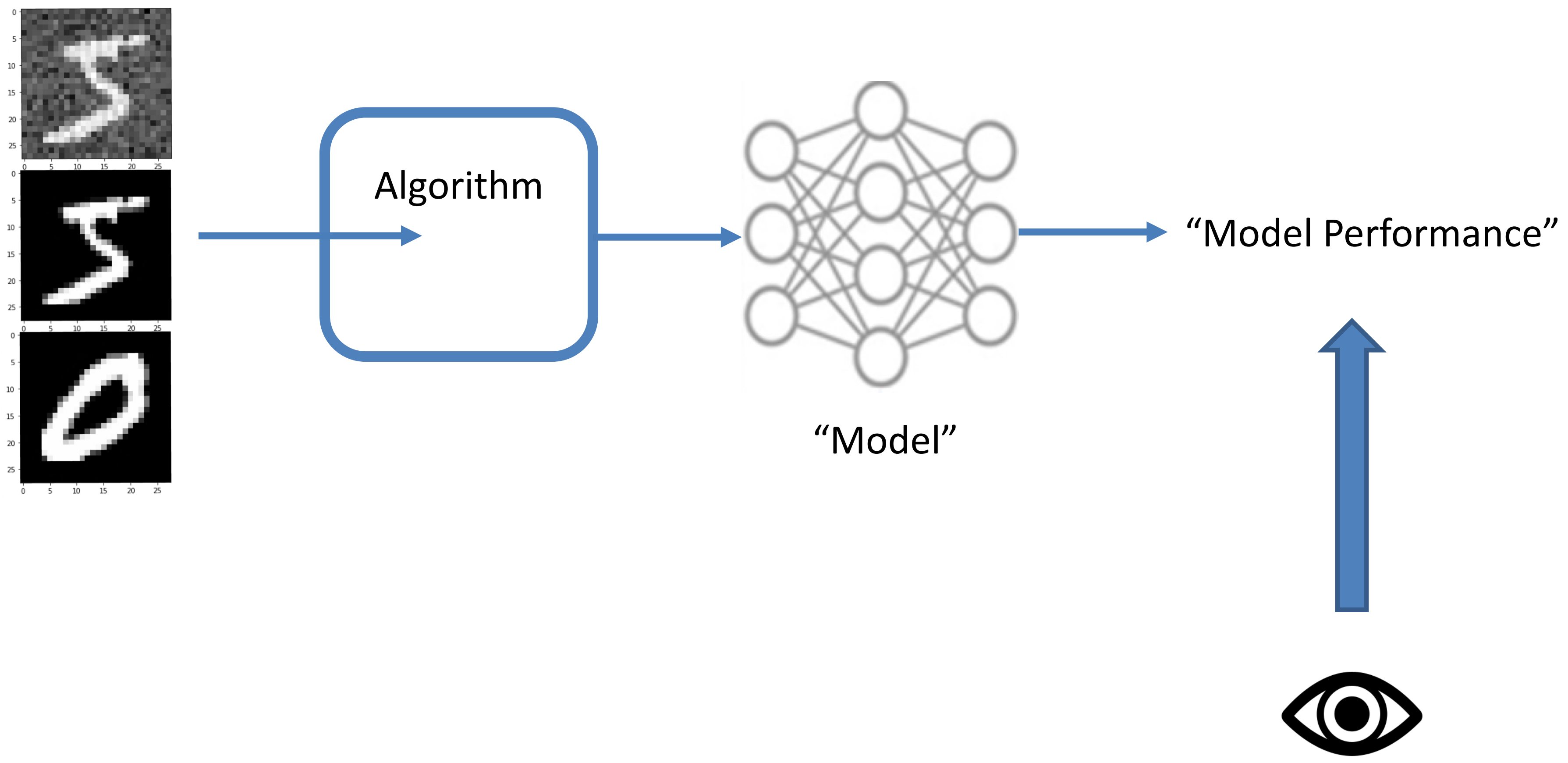
"Quality"



Contribution Evaluation Obstacle

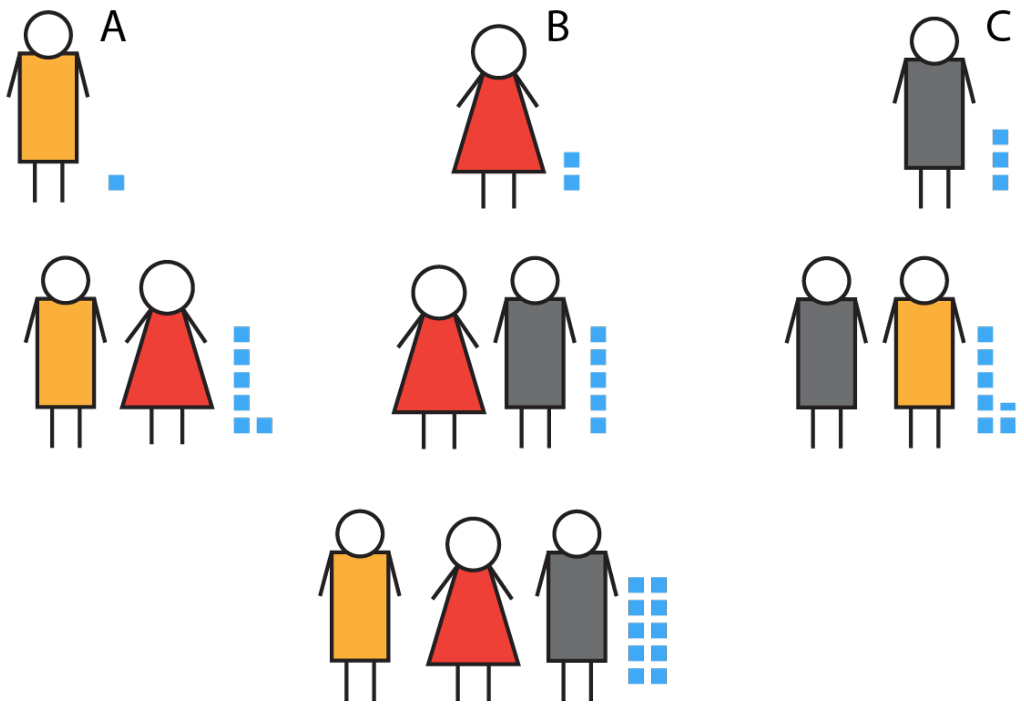


Contribution Evaluation Obstacle



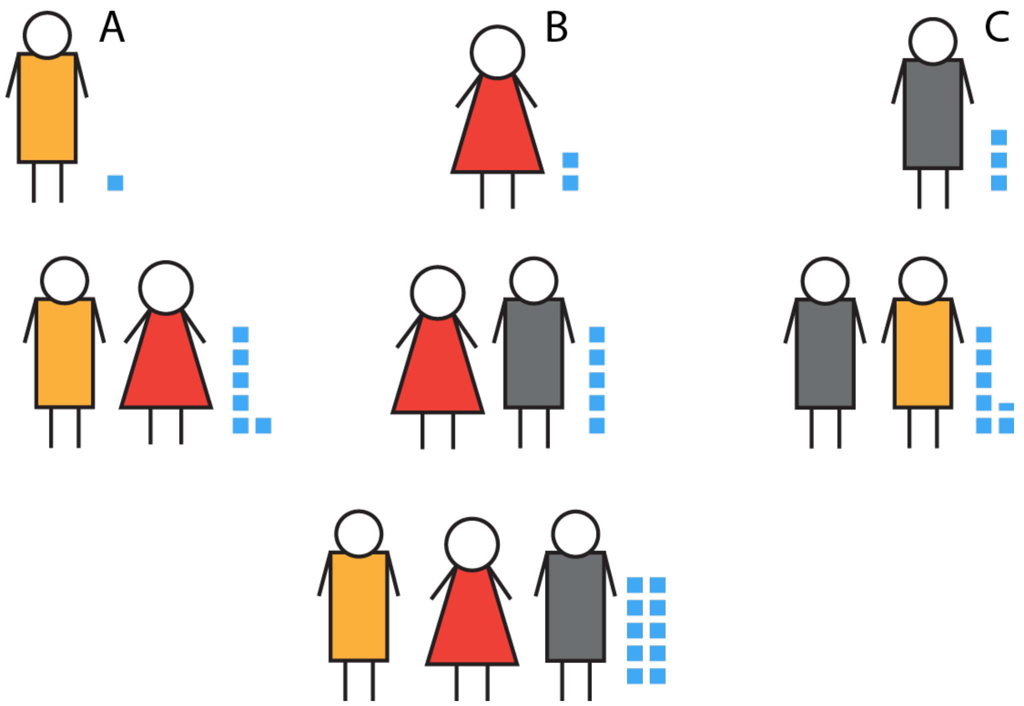
Shapley Value – An Example

- Example: A, B, C works together in a project worth of 100 points. How many points should each of them get?



Shapley Value – An Example

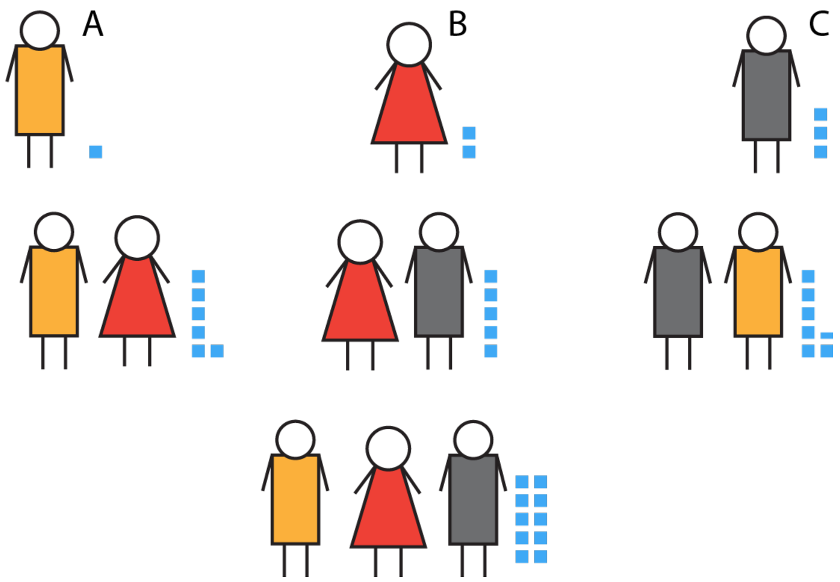
- V(A)=10, V(B)=20, V(C)=30
- V(AB)=60, V(BC)=50, V(AC)=65, V(ABC)=100



Shapley Value – An Example

- V(A)=10, V(B)=20, V(C)=30
- V(AB)=60, V(BC)=50, V(AC)=65, V(ABC)=100

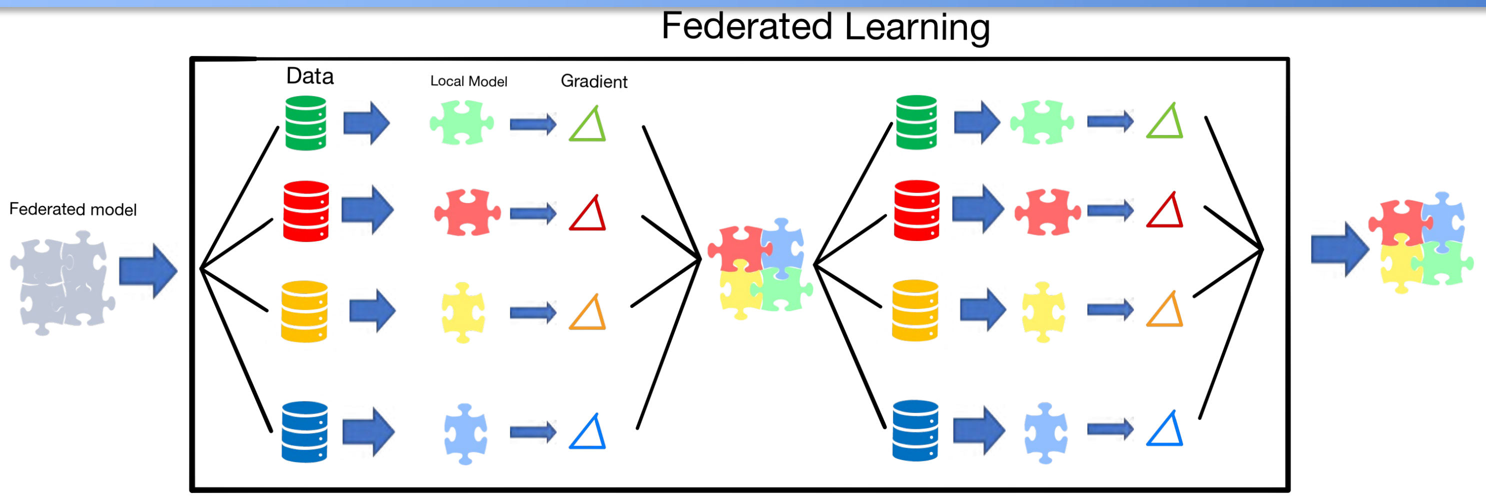
- B-C-A: (A,B,C)=(50,20,30)
- C-A-B: (A,B,C)=(35,35,30)
- A-C-B: (A,B,C)=(10,35,55)
- C-B-A: (A,B,C)=(50,20,30)
- B-A-C: (A,B,C)=(40,20,40)

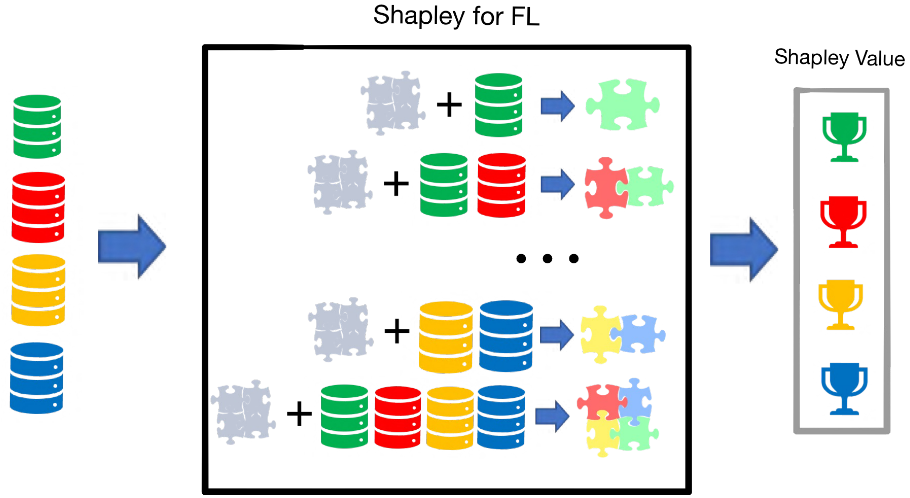


Shapley Value – An Example

- A=(10+50+35+10+50+40)/6=195/6=32.5
- B=(50+20+35+35+20+20)/6=180/6=30
- C=(40+30+30+55+30+40)/6=225/6=37.5

Adopting Shapley Value in FL





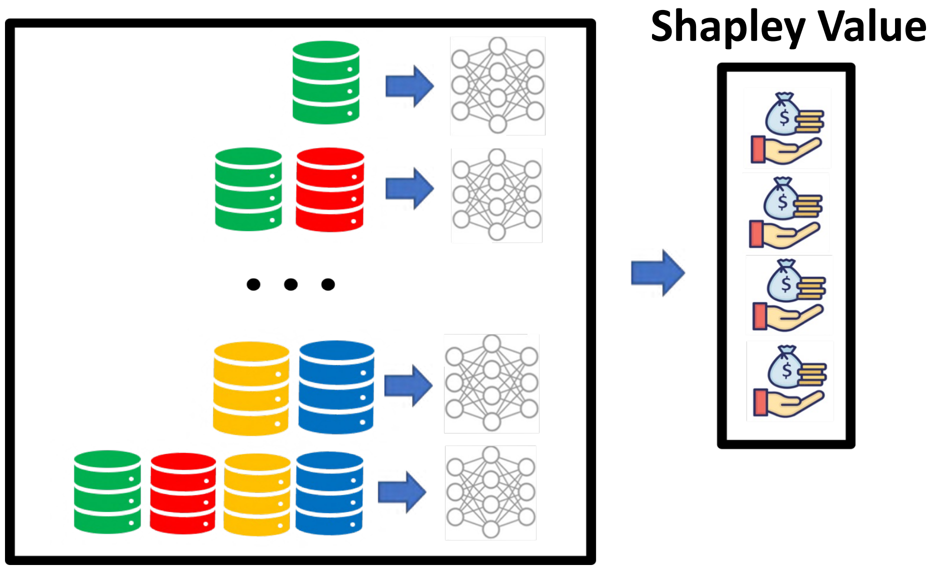
Drawbacks of Shapley Value

Problem:

1. Traditional Shapley requires retraining FL sub-models.

$$V(S) = V(M_S) = V(\mathcal{A}(M^{(0)}, D_S))$$

2. 2^N FL sub-models' utility evaluations $V(S)$ lead to computation overhead.



Solution: GTG-Shapley

- Guided Truncation Gradient Shapley (GTG-Shapley) : Fair, Efficient, and Privacy-preserving.

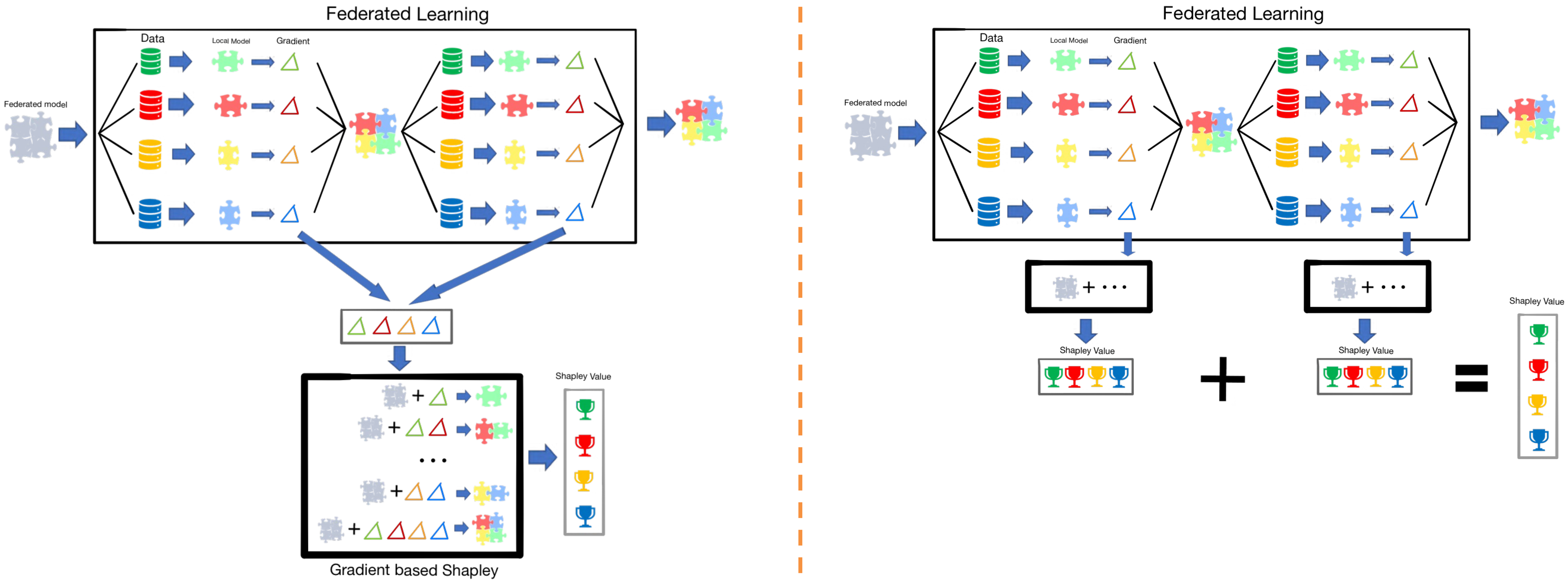
Key Idea:

1. Model Reconstruction, instead of Model Retraining

$$V(S) = V(M_S) = V\left(M + \sum_{i \in S} \frac{|D_i|}{|D_S|} \Delta_i\right)$$

$$\neq V(\mathcal{A}(M^{(0)}, D_S))$$

GTG-Shapley: Model Reconstruction



Solution: GTG-Shapley

- Guided Truncation Gradient Shapley (GTG-Shapley) : Fair, Efficient, and Privacy-preserving.

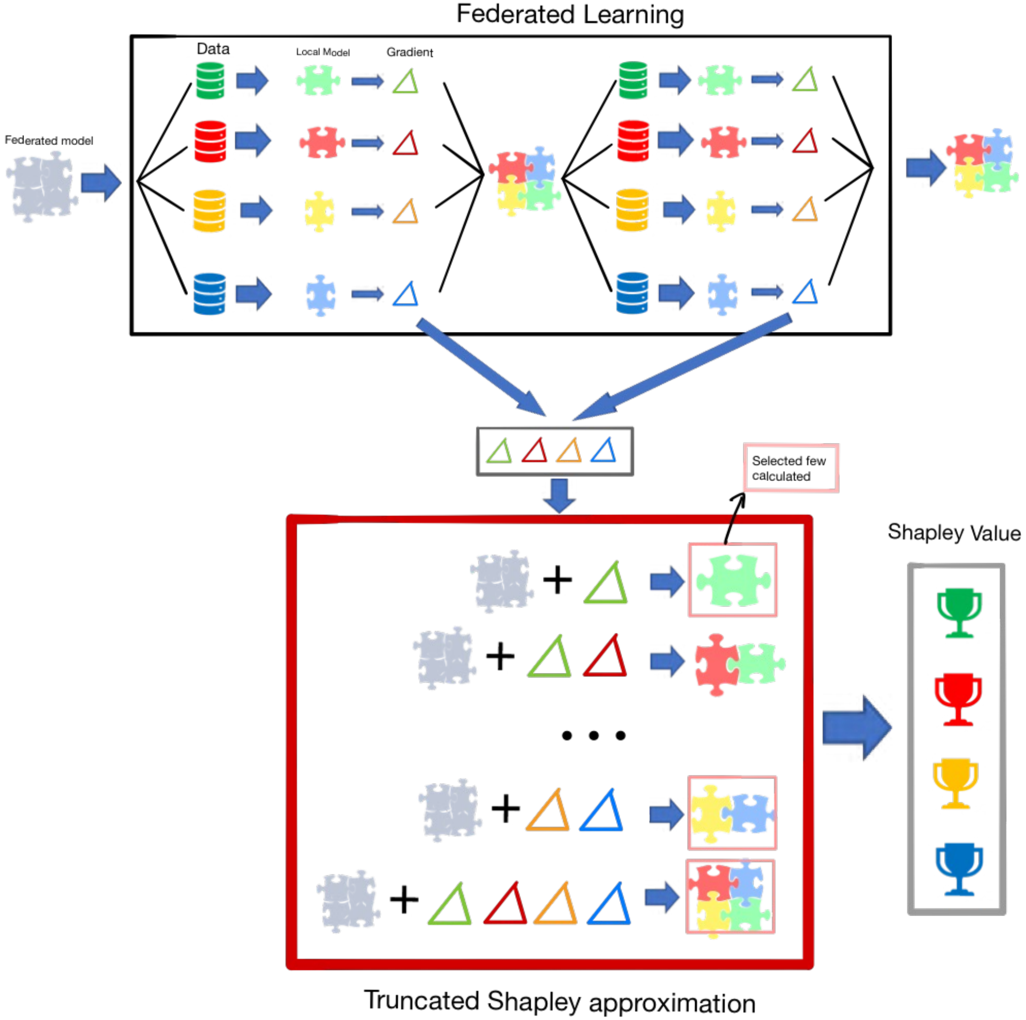
Key Idea:

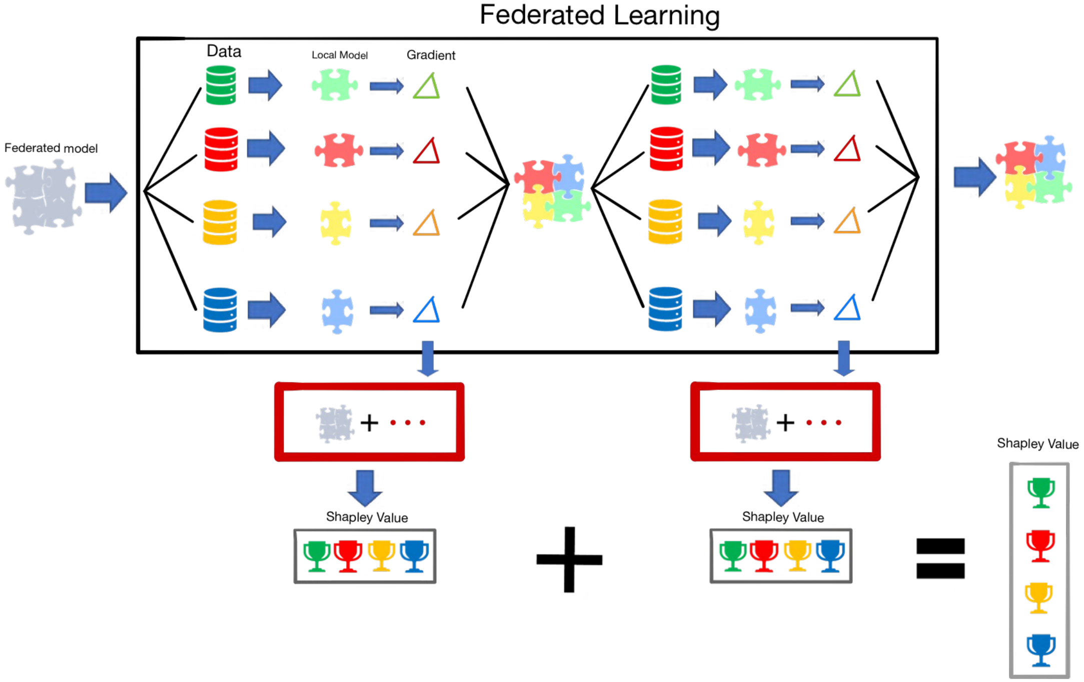
1. Model Reconstruction, instead of Model Retraining

$$V(S) = V(M_S) = V\left(M + \sum_{i \in S} \frac{|D_i|}{|D_S|} \Delta_i\right)$$
$$\neq V(\mathcal{A}(M^{(0)}, D_S))$$

2. Truncating unnecessary sub-model, instead of 2^N sub-models.

GTG-Shapley: Monte-Carlo Truncation





GTG-Shapley Performance

Empirical studies on 7 existing SV-based FL participant contribution evaluation approaches under i.i.d. and non-i.i.d settings.

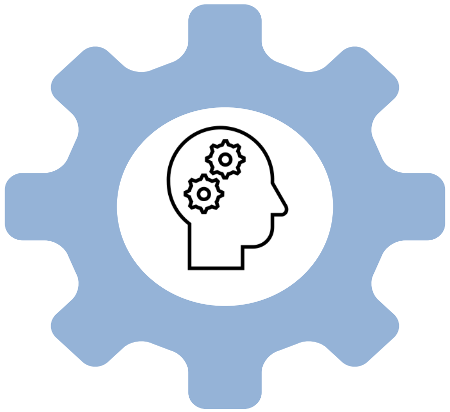
GTG-Shapley consistently achieves the highest efficiency and accuracy under both i.i.d. and non- i.i.d. settings.

	i.i.d		non-i.i.d	
	Duration	ED	Duration	ED
Canonical SV	4.615	-	4.615	-
MR	3.833	-2.35	3.733	-2.148
TMC	4.168	-1.687	4.213	-1.369
TMR	3.531	-2.353	3.678	-2.27
GroupTesting	4.583	-0.894	4.557	-0.667
Fed-SV	3.784	-0.757	3.711	-0.789
GTG-Shapley	**2.662**	**-2.427**	**2.733**	**-2.323**

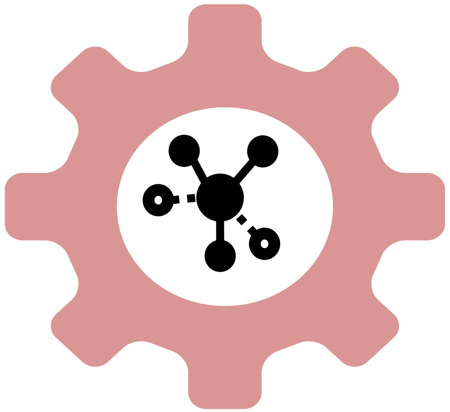
present in $\log_{10}$ scale

AI in CAreFL

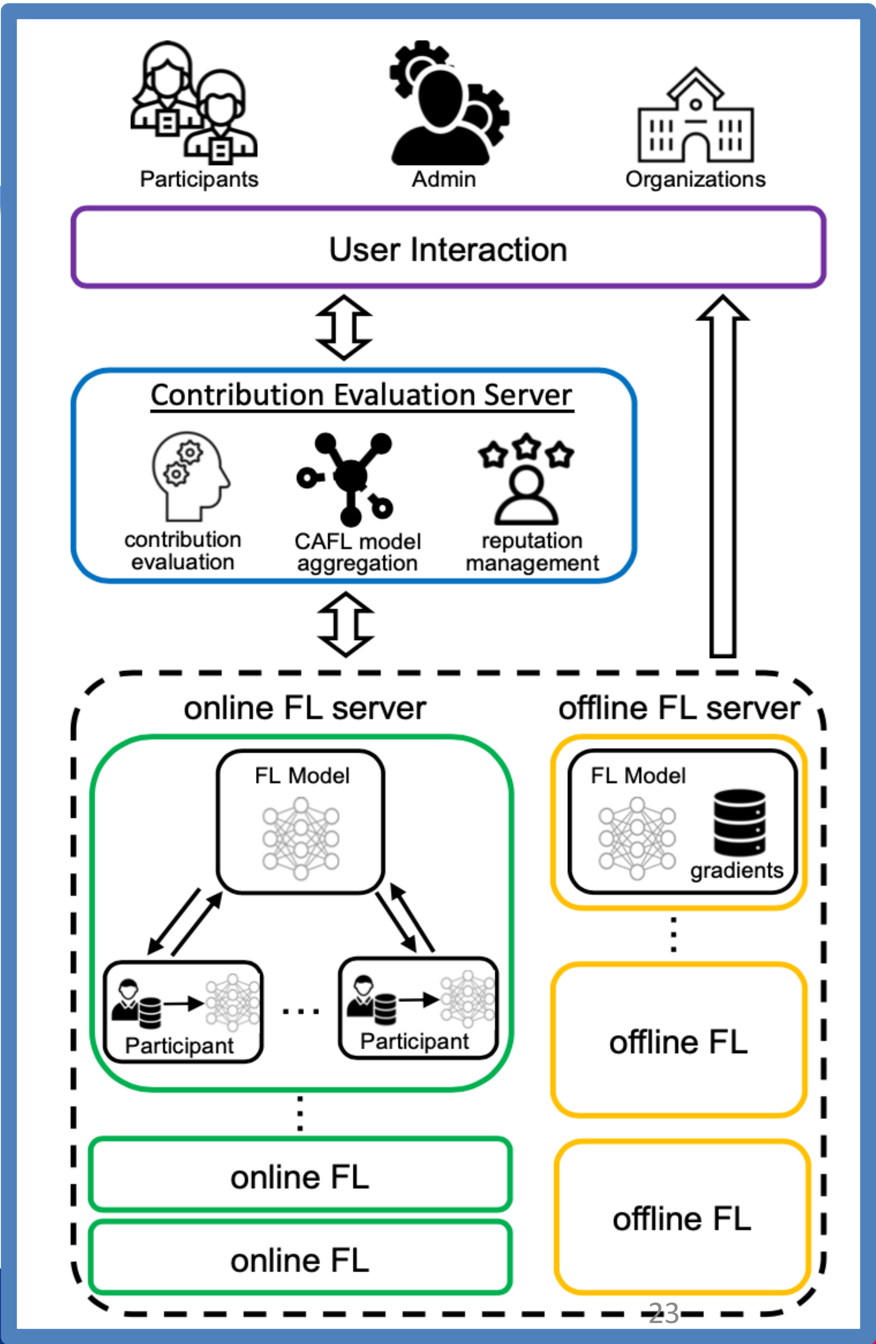
- Focus on Contribution Evaluation



Fast and Accurate Contribution Evaluation

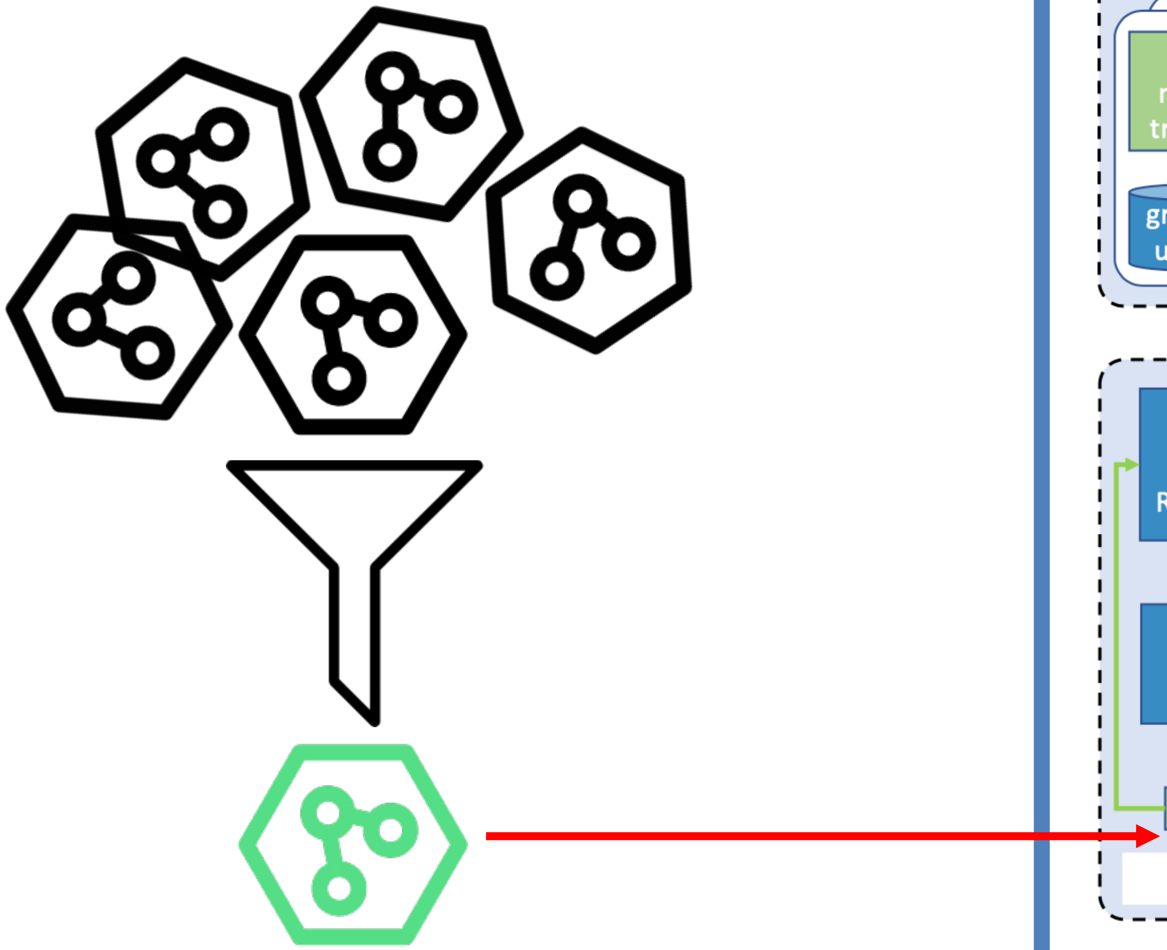


Contribution-Aware FL Model Aggregation

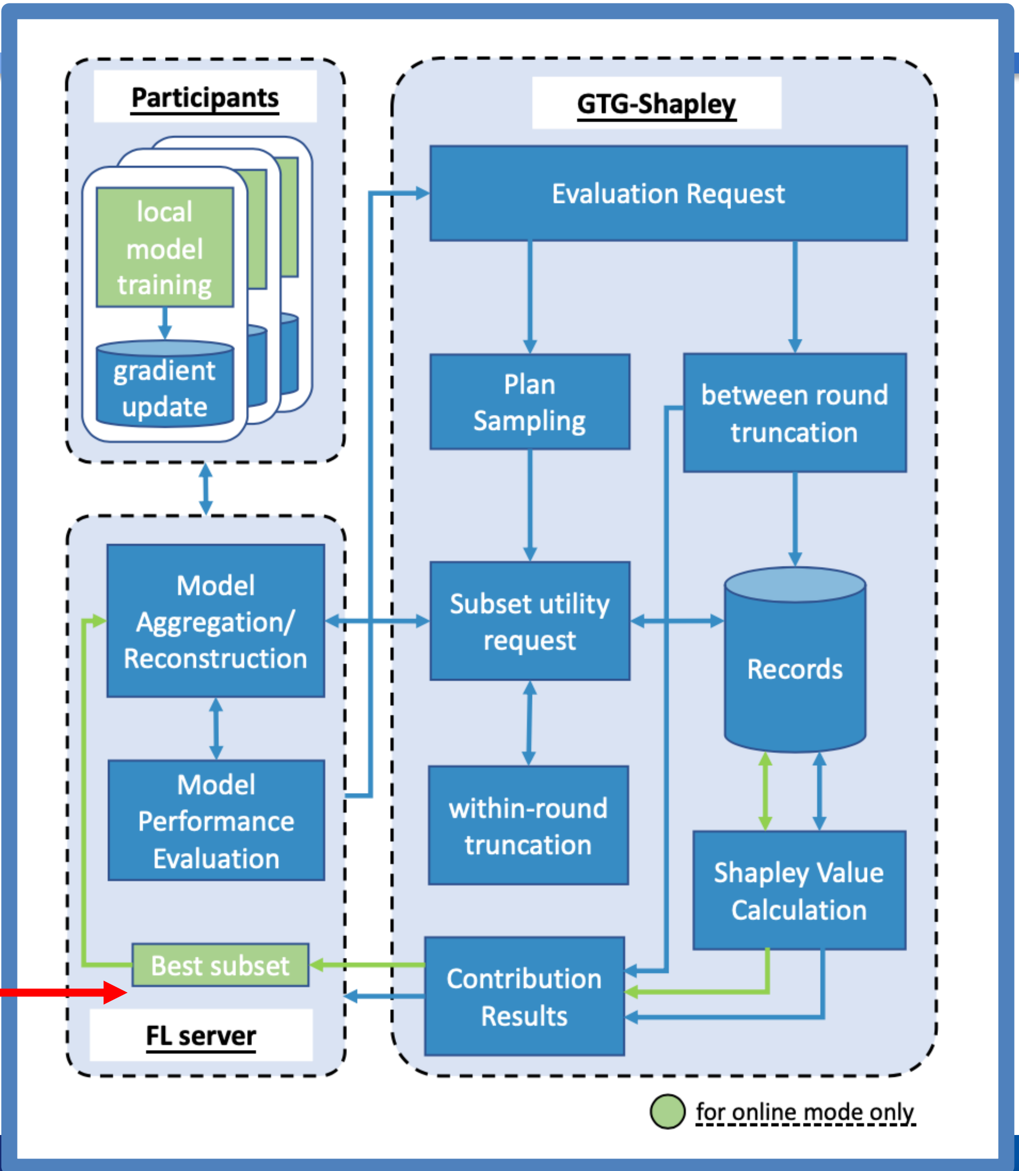


CAreFL model aggregation

FL sub-models

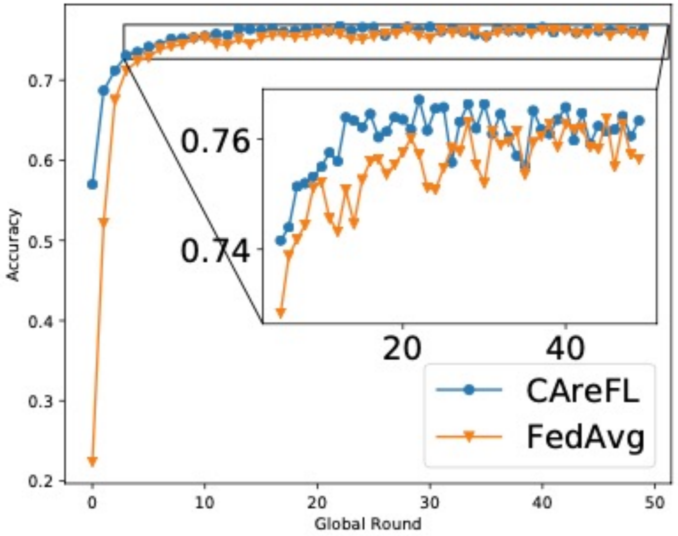


"Best Subset" FL Model

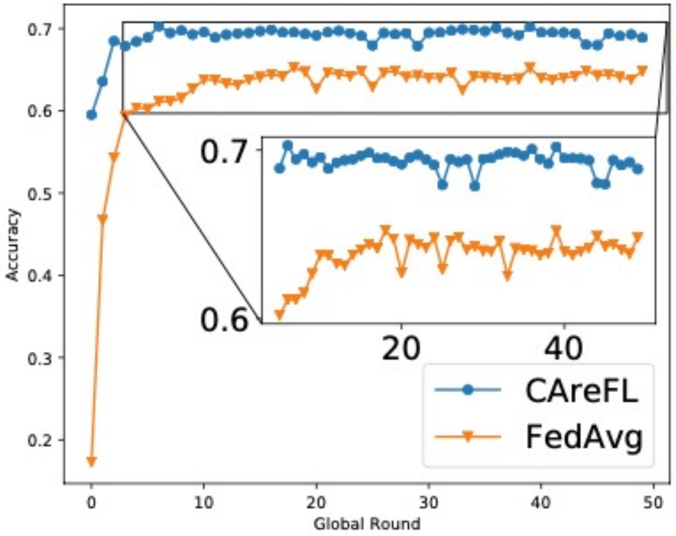


Results on public benchmark

- Empirical Studies on CAreFL model aggregation with FedAvg under i.i.d and non-i.i.d settings (CIFAR-10 dataset).



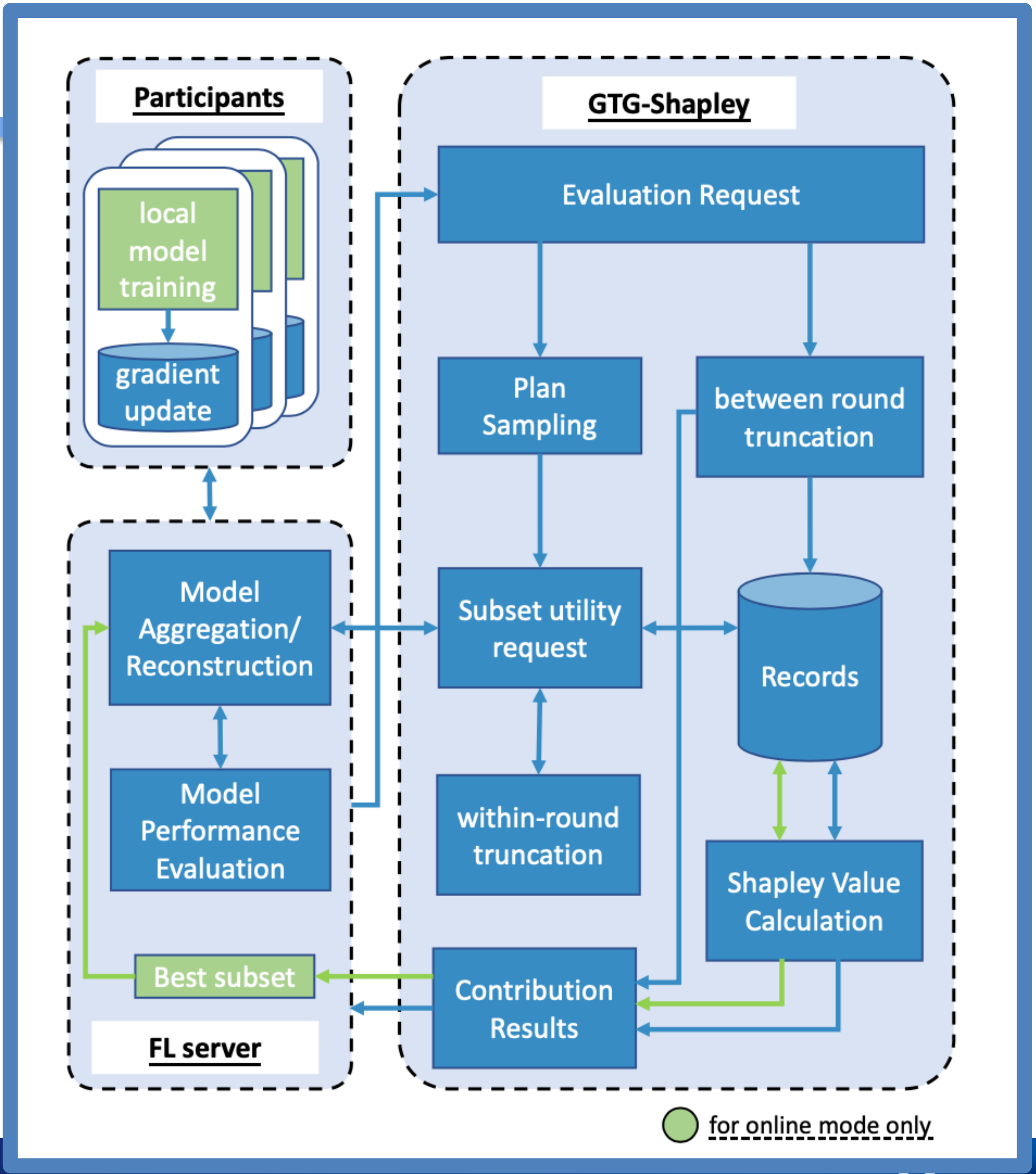
(a) i.i.d. case



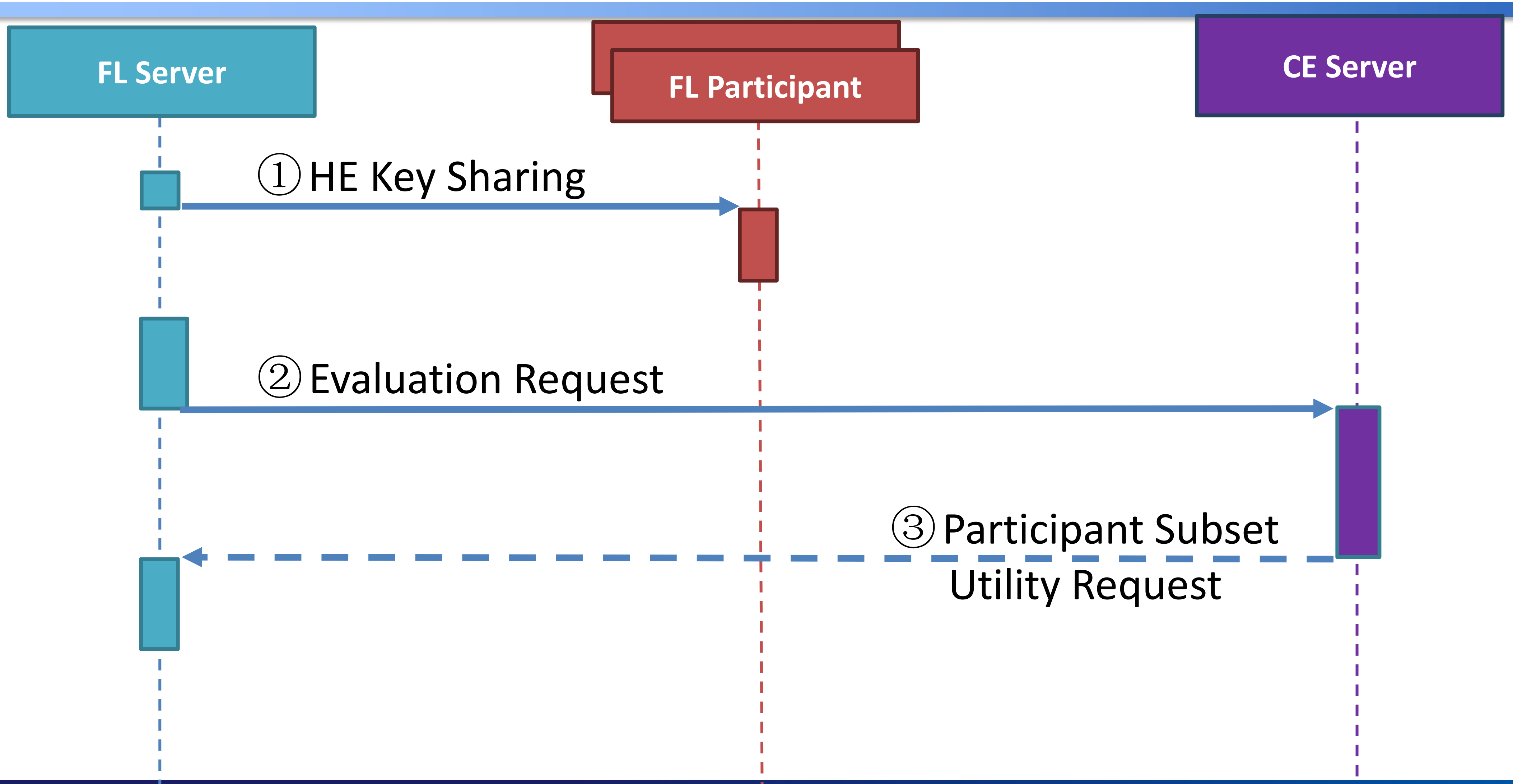
(b) non-i.i.d. case

CAreFL in details

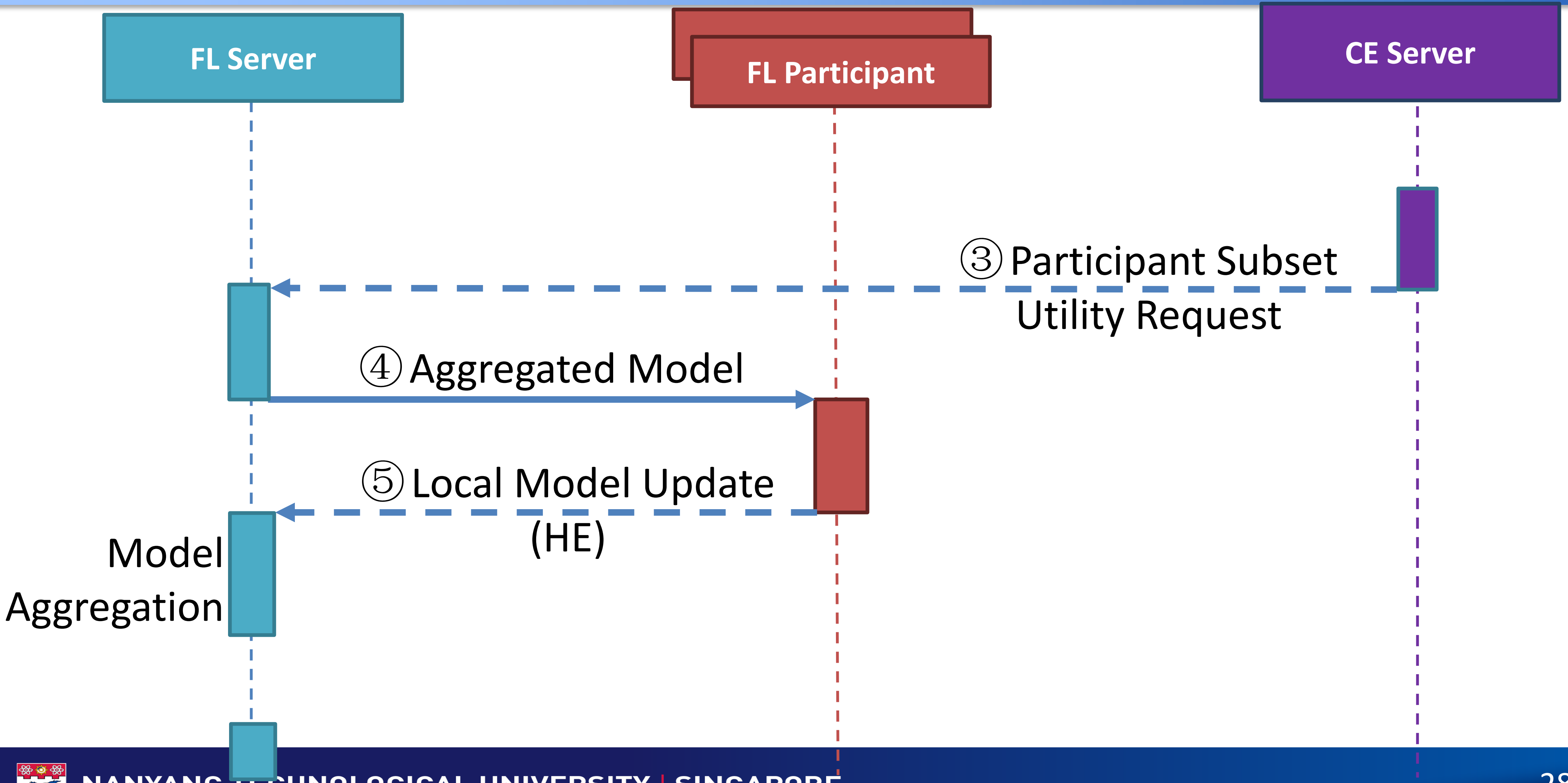
- GTG-Shapley only requires a list of unique participants' IDs and computes the participants' contributions in an efficient manner and returns the results to the FL server.
- In addition, it also identifies the "best subset" and passes this information to the FL server to improve model aggregation.
- The aggregation function is only relevant for online FL training during which the global FL model is still in the process of being established.



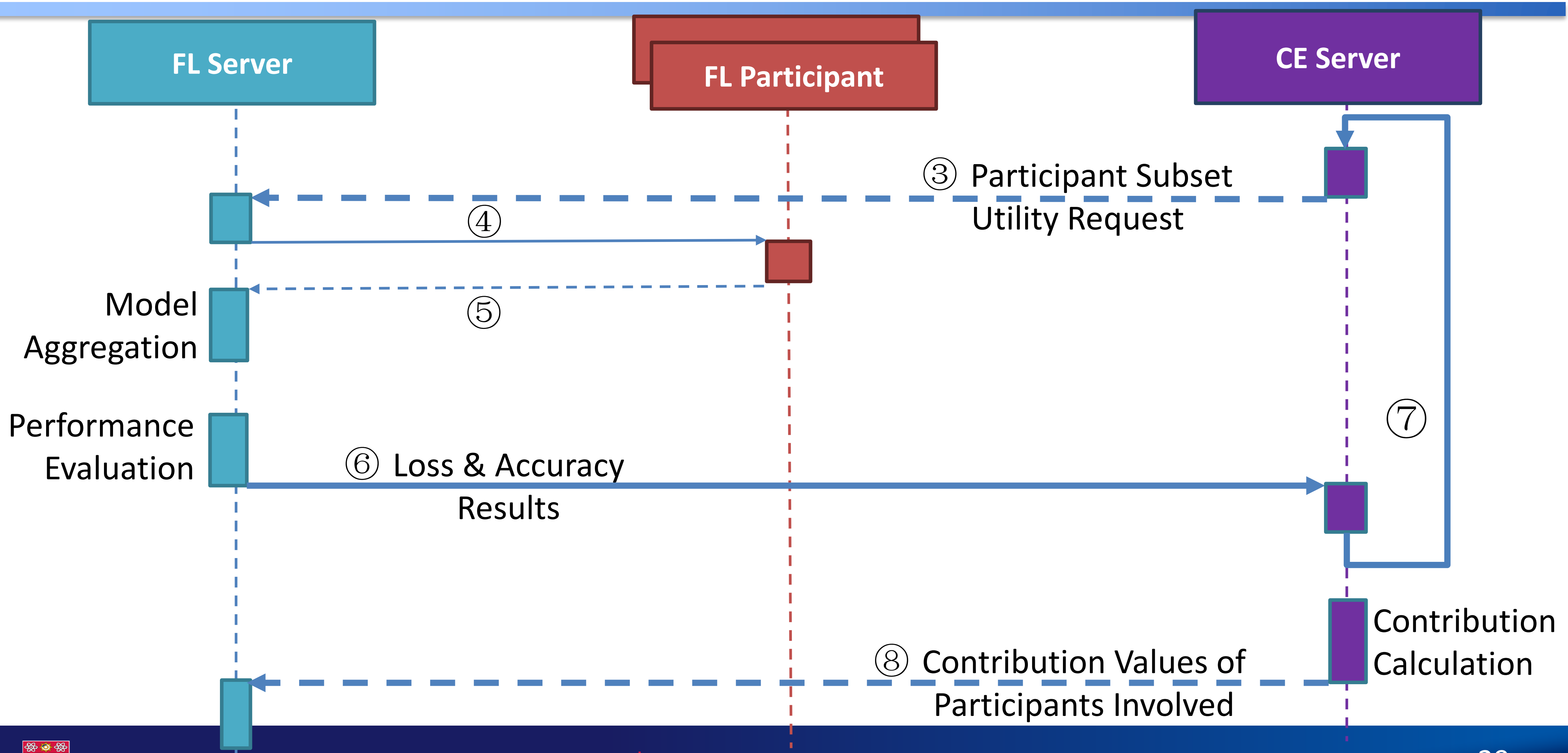
Contribution Evaluation workflow



Contribution Evaluation workflow



Contribution Evaluation workflow



Deployment and Payoff

- The **CAreFL** framework has been **deployed** in Yidu Cloud Technology Inc. since March 2021 in two lines of their business: 1) clinical research services, and 2) real-world trial research services.

Leukemia

- Clinical research.
- A total of 62,000 patients.

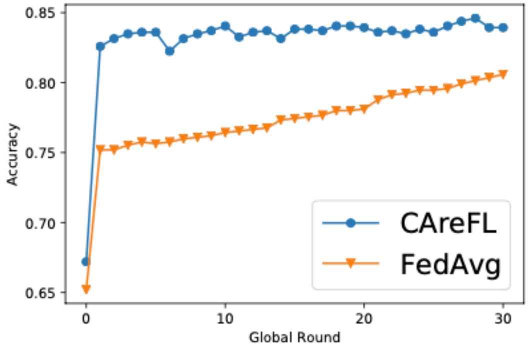
Biopsy

- Real-world trial.
- A total of 5,978 patients screened, and 2,426 patients selected.

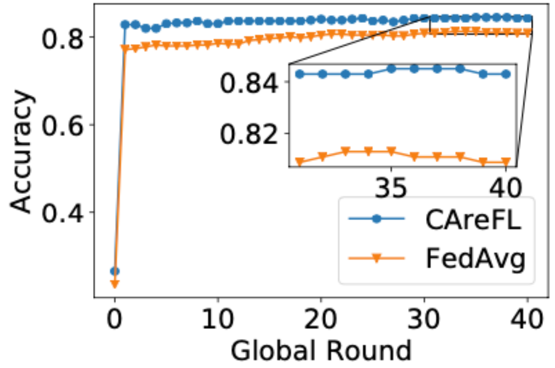
Pneumonia

- Real-world trial.
- A total of 103,455 sample data.

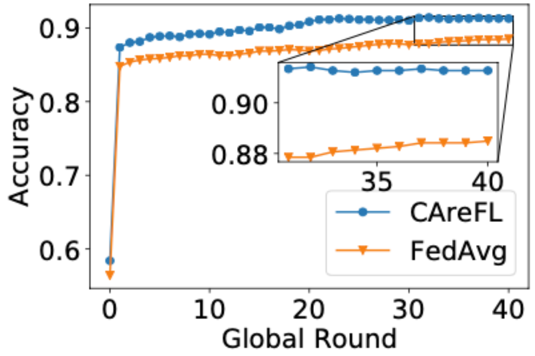
Deployment and Payoff



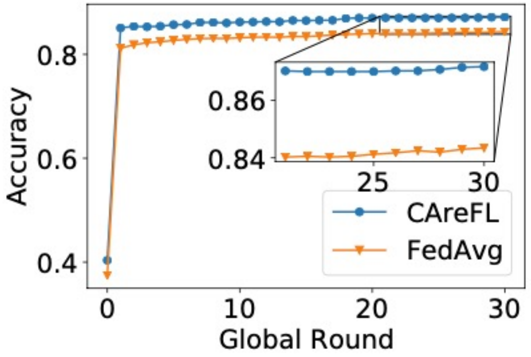
(a) Leukemia LR



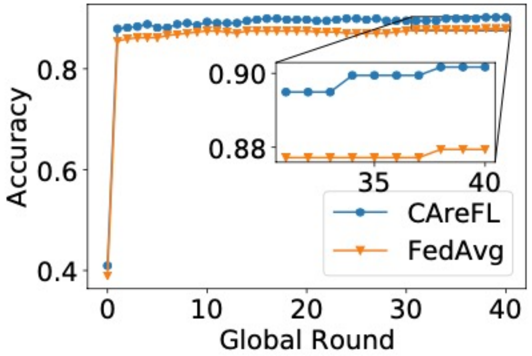
(c) Biopsy LR



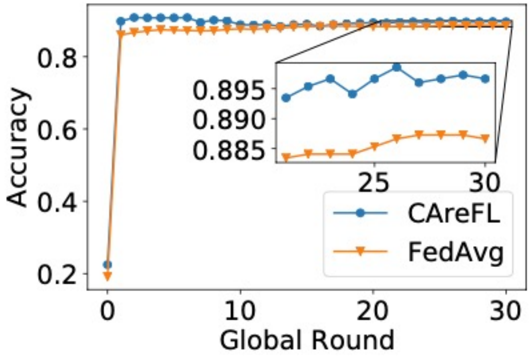
(e) Pneumonia LR



(b) Leukemia SBT

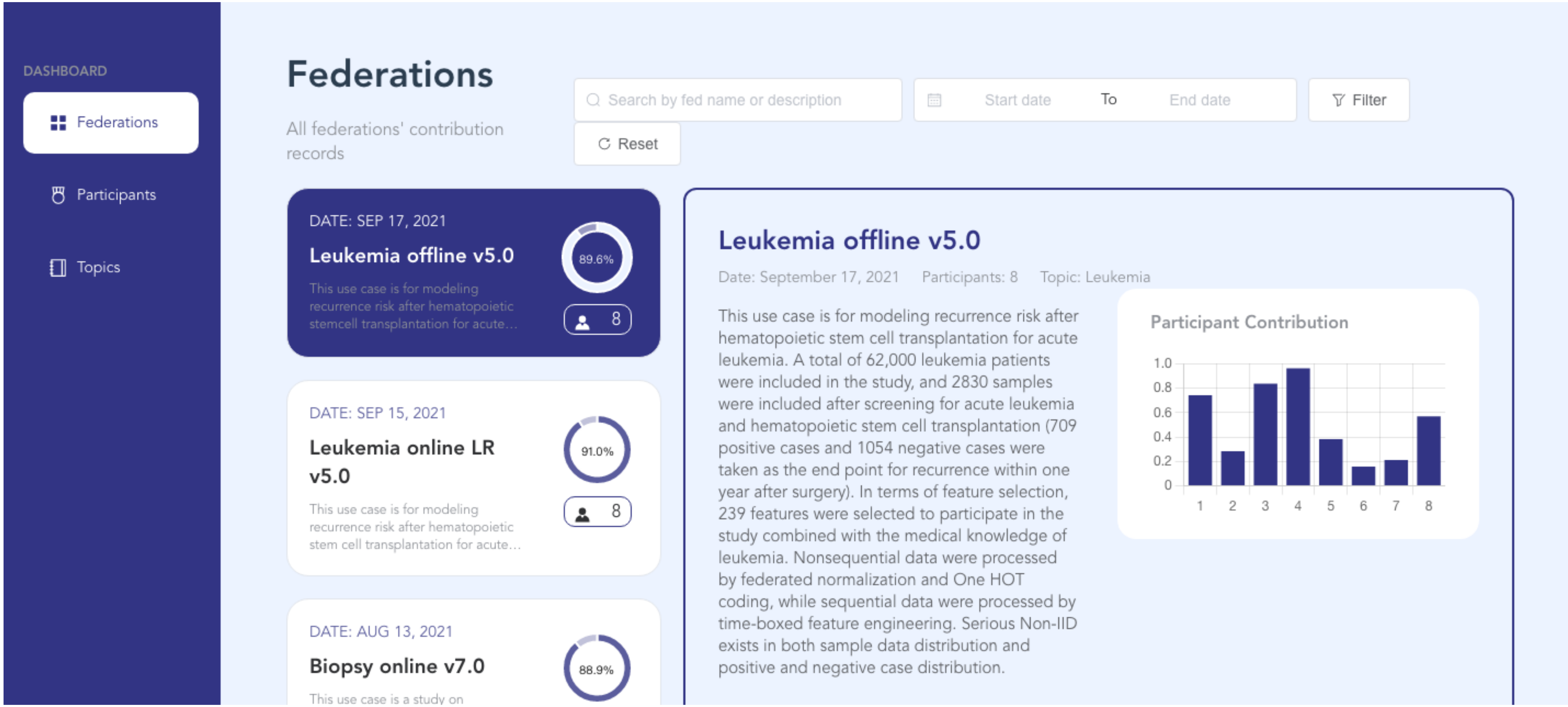


(d) Biopsy SBT



(f) Pneumonia SBT

User Interface



demo video @ link

Award

Innovative Applications of Artificial Intelligence

CERTIFICATE

Innovative Application Award

For the Paper Entitled

"Contribution-Aware Federated Learning for Smart Healthcare"

By

Zelei Liu, Yuanyuan Chen, Yansong Zhao, Han Yu, Yang Liu, Renyi Bao, Jinpeng Jiang, Zaiqing Nie, Qian Xu, and Qiang Yang

Meinolf Sellmann – Program Co-Chair

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

WeBank 微众银行

医渡云 YIDUCLOUD

AIR

香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY