Oort

Efficient Federated Learning via Guided Participant Selection

Fan Lai, Xiangfeng Zhu,

Harsha V. Madhyastha, Mosharaf Chowdhury

SymbioticLab

University of Michigan

Emerging Trend of Machine Learning

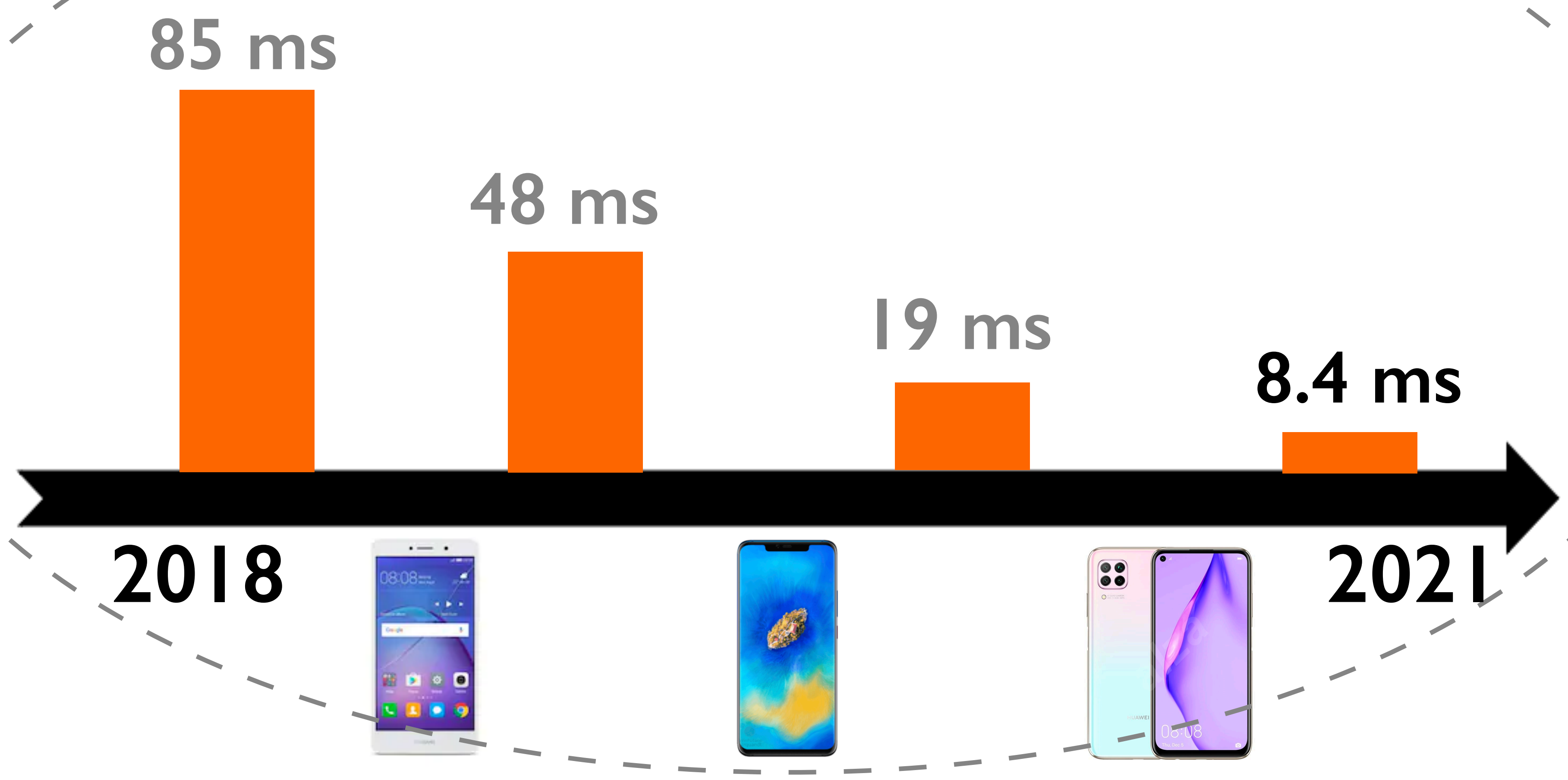
Edge devices generate massive data

Emerging Trend of Machine Learning

Edge devices generate massive data

Increasing resource on edge device

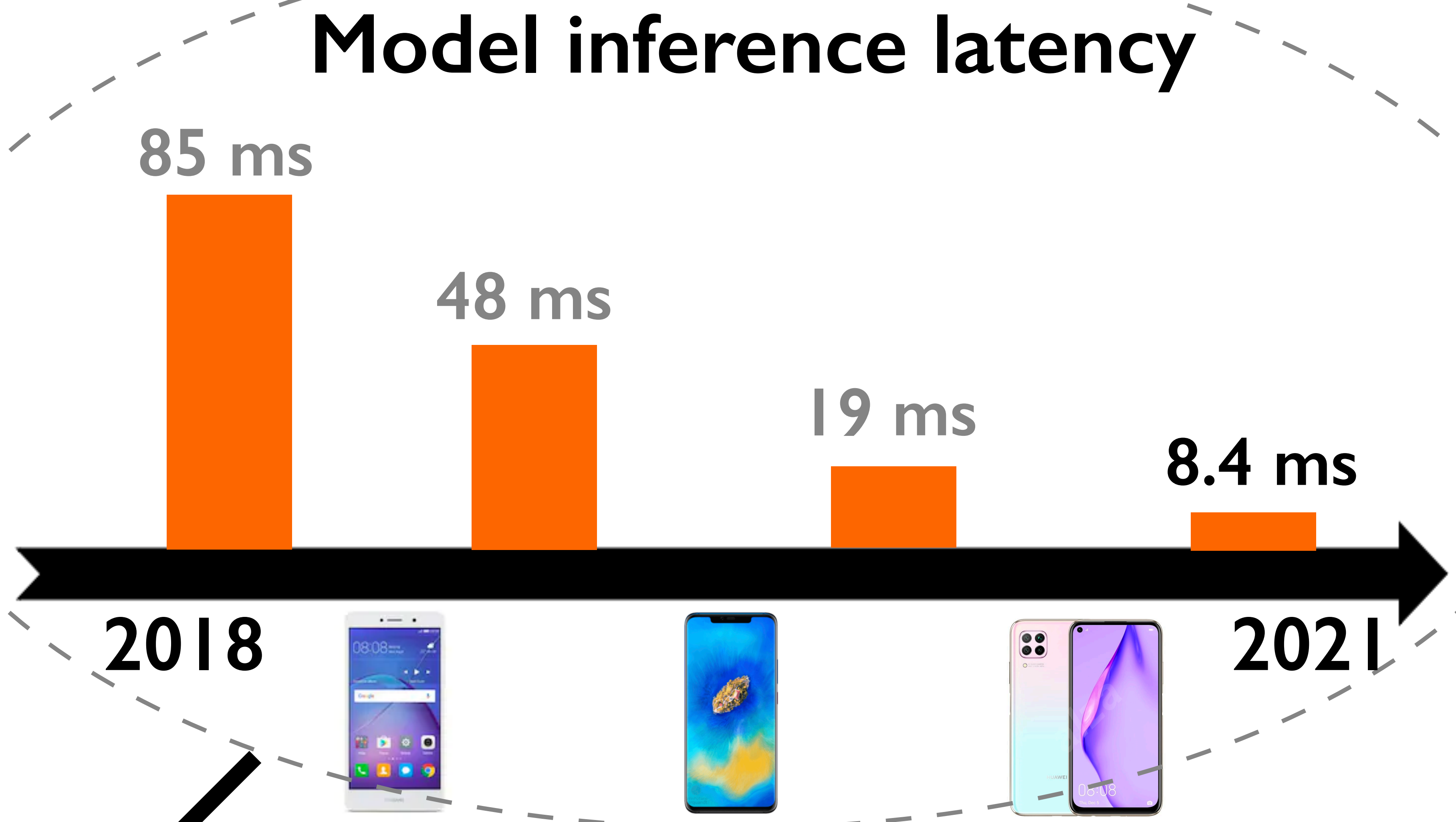
Model inference latency

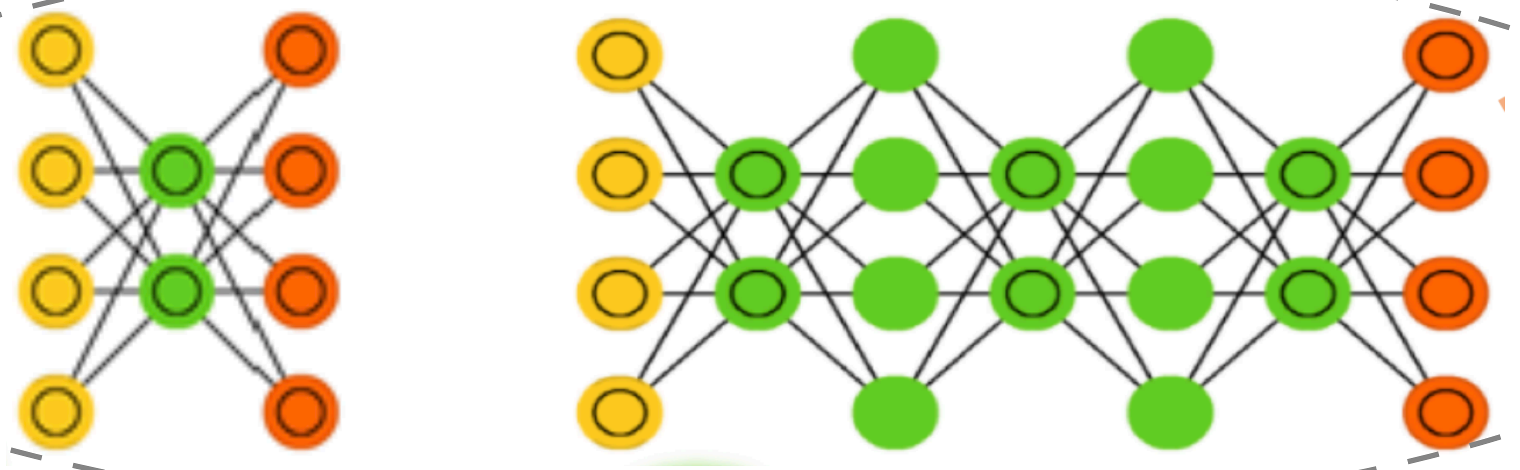


Emerging Trend of Machine Learning

Edge devices generate massive data

Increasing resource on edge device





ML needs fresh and large real-life datasets

Emerging Federated Learning on the Edge

- **On-device machine learning helps**
  - Reduce data migration/privacy risk
  - Learn on fresh real-world data
  - …

***Mistify*: Automating DNN Model Porting for On-Device Inference at the Edge**

TOWARDS FEDERATED LEARNING AT SCALE: SYSTEM DESIGN

MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications

APPLIED FEDERATED LEARNING:
IMPROVING GOOGLE KEYBOARD QUERY SUGGESTIONS

Many others …

Emerging Federated Learning on the Edge

- **On-device machine learning helps**
  - Reduce data migration/privacy risk
  - Learn on fresh real-world data
  - …

- **Federated training and testing**
  - Run model across millions of edge clients

Mistify: **Automating DNN Model Porting for On-Device Inference at the Edge**

TOWARDS FEDERATED LEARNING AT SCALE: SYSTEM DESIGN

MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications

APPLIED FEDERATED LEARNING: IMPROVING GOOGLE KEYBOARD QUERY SUGGESTIONS

Many others …

Syft

NVIDIA Clara

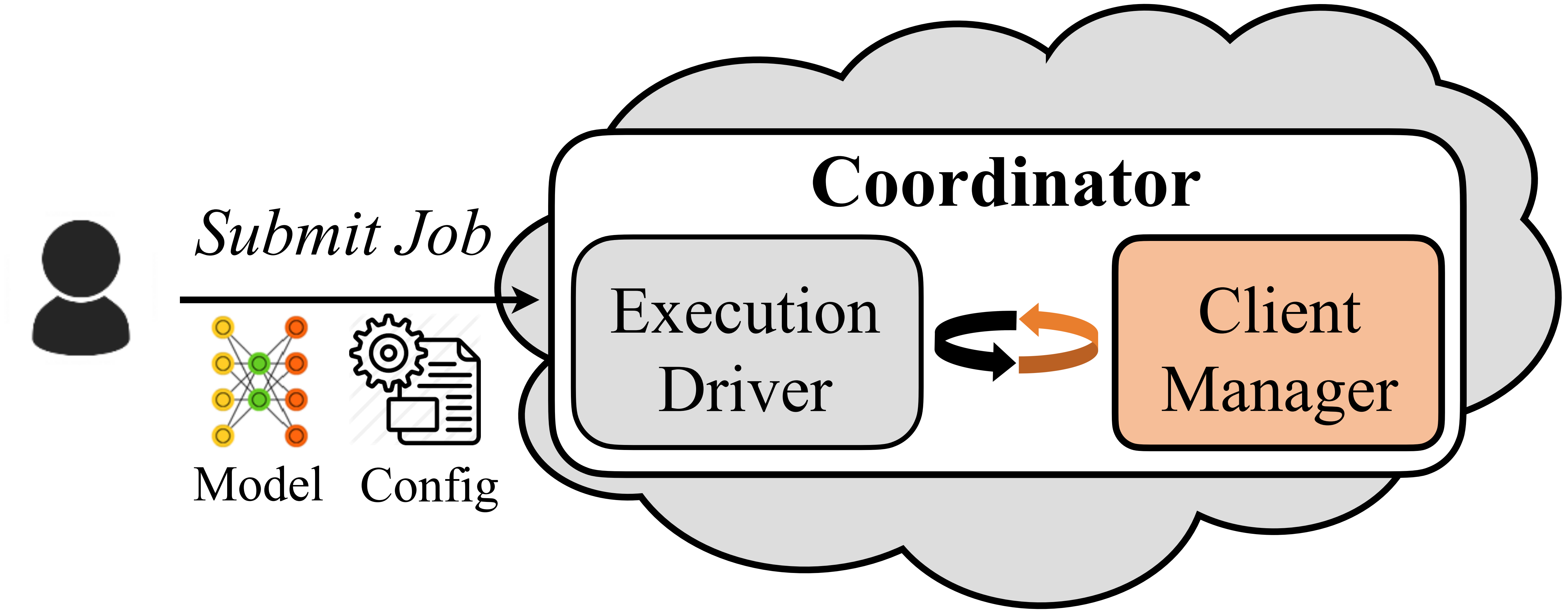
PyTorch Mobile

OpenMined

TensorFlow Federated

Apple CoreML

Execution of Federated Learning (FL)

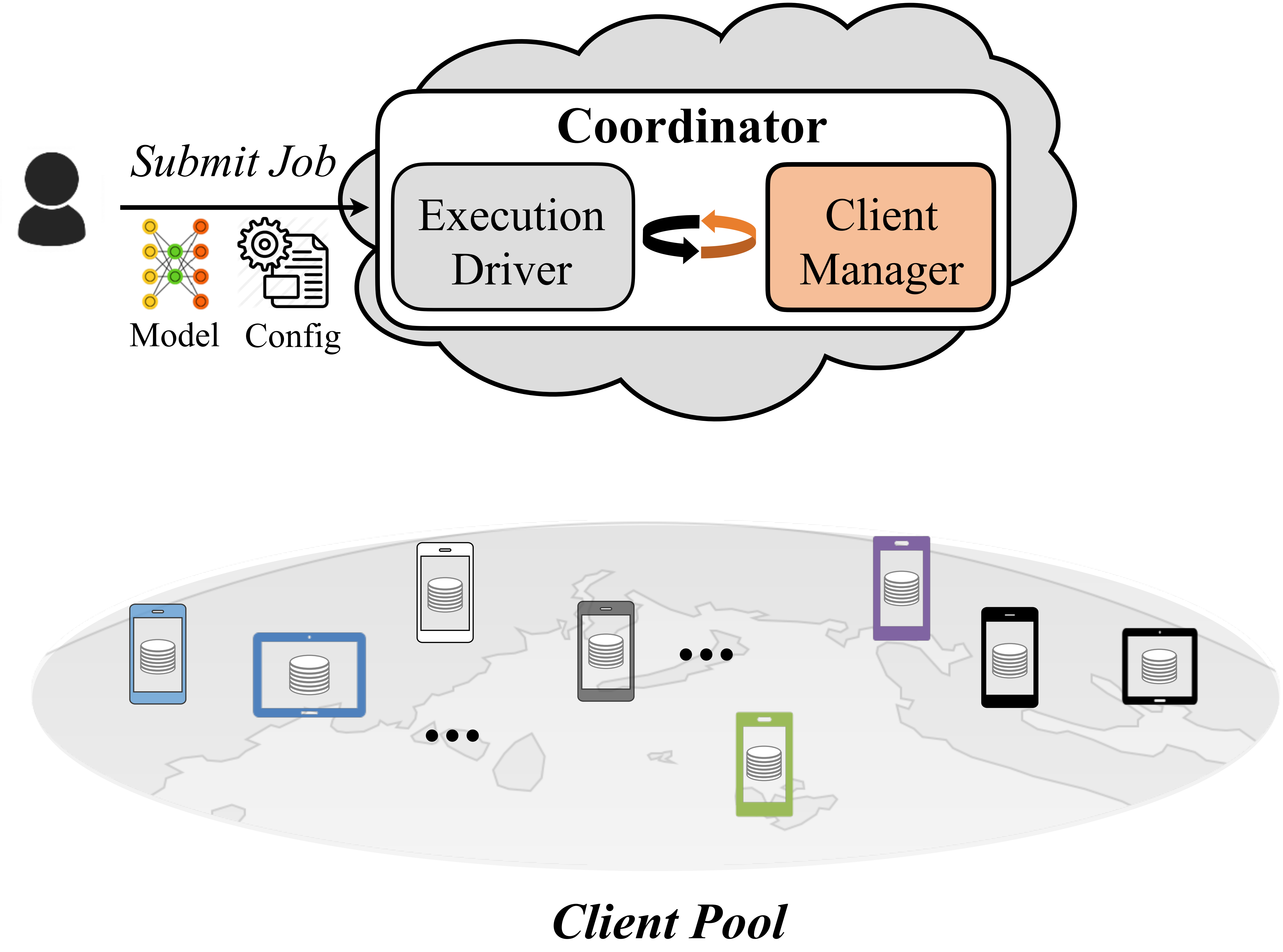


Primary Objective

Better time to accuracy:

- Less time for target acc. under the same setting

Execution of Federated Learning (FL)

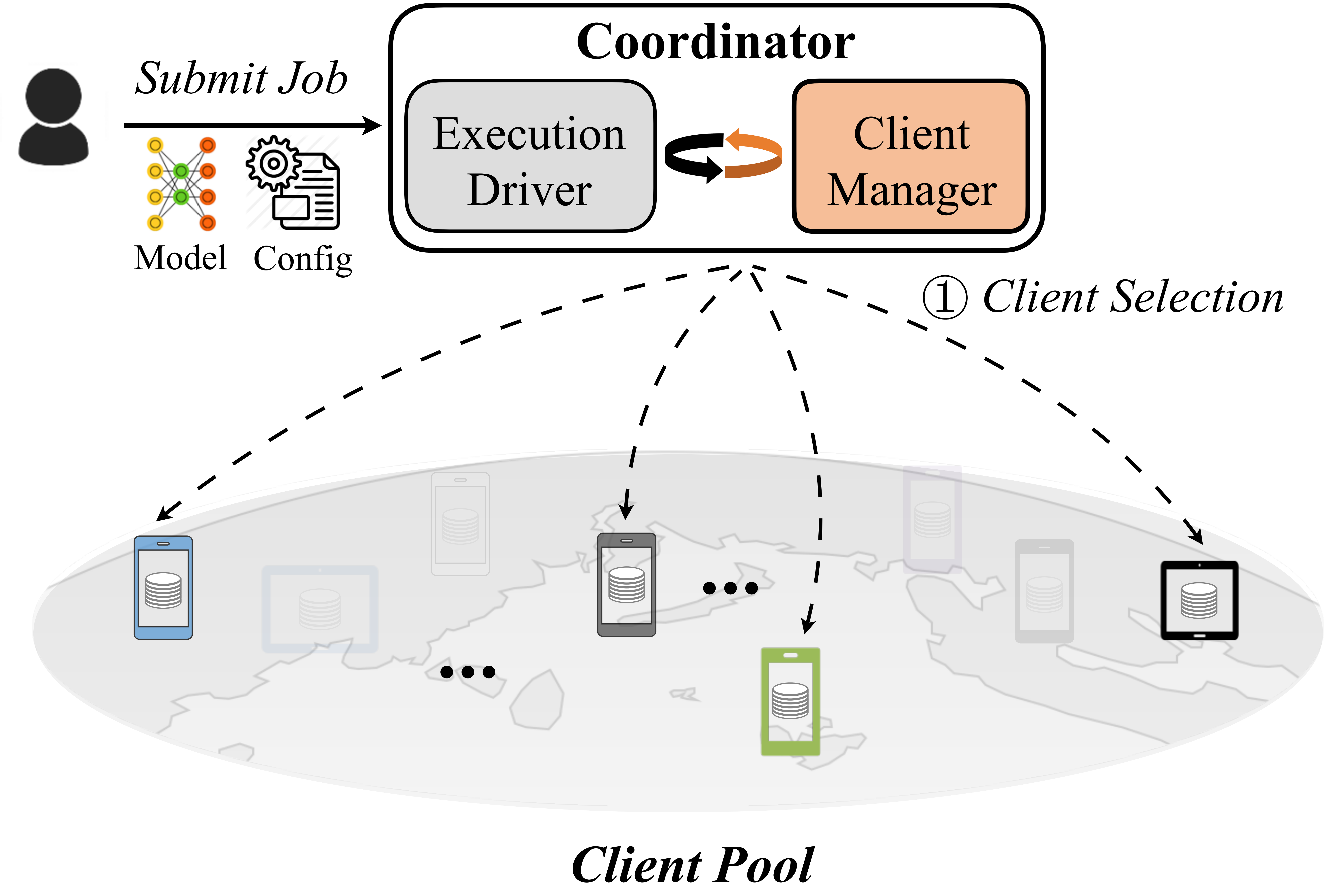


Primary Objective

*Better **time to accuracy**:*

- Less time for target acc. under the same setting

Execution of Federated Learning (FL)



O(100) Rounds:

Round *i*

Client selection

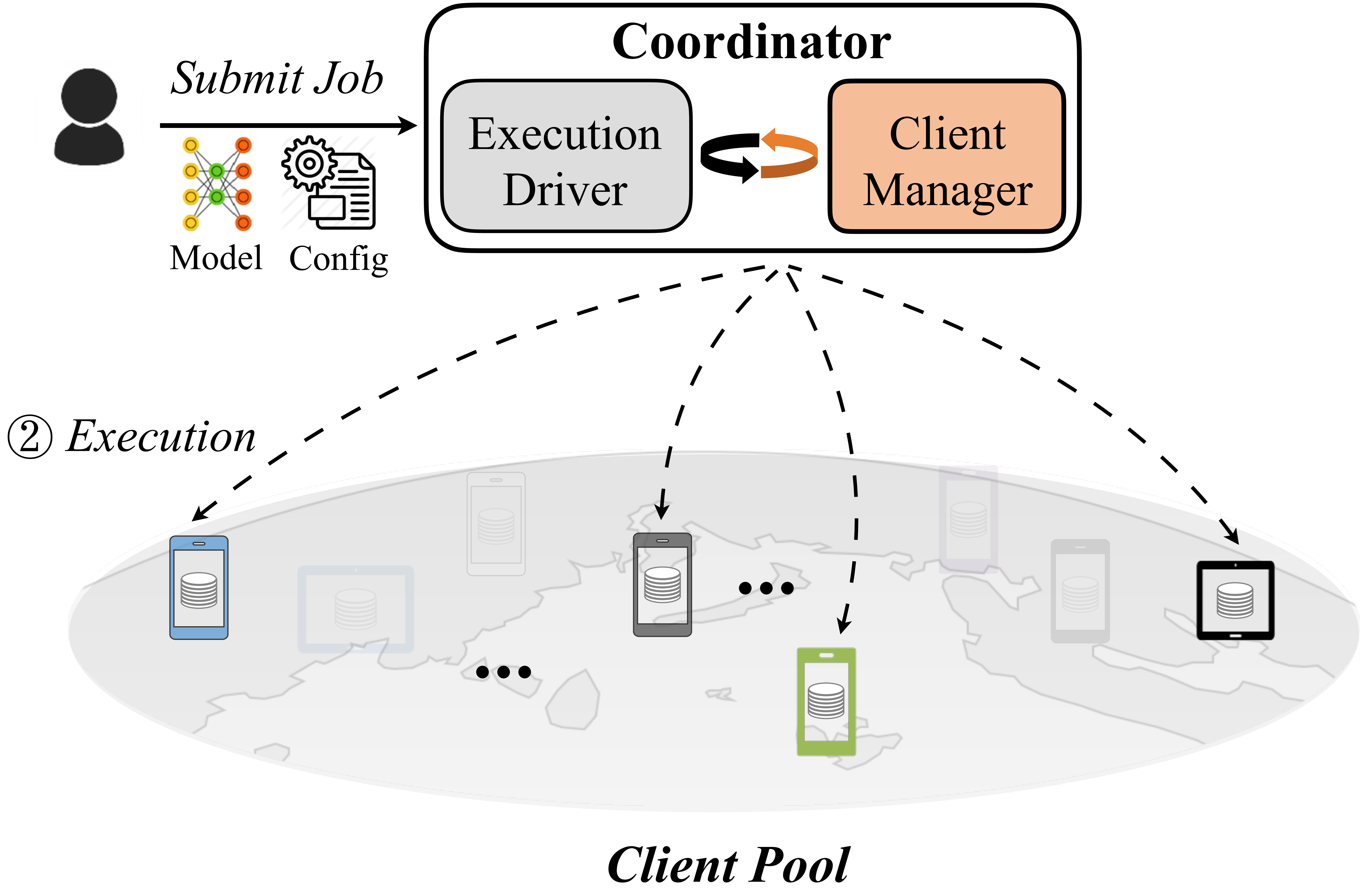
↓

In-situ Execution

↓

Result aggregation

Execution of Federated Learning (FL)



O(100) Rounds:

Round i

Client selection

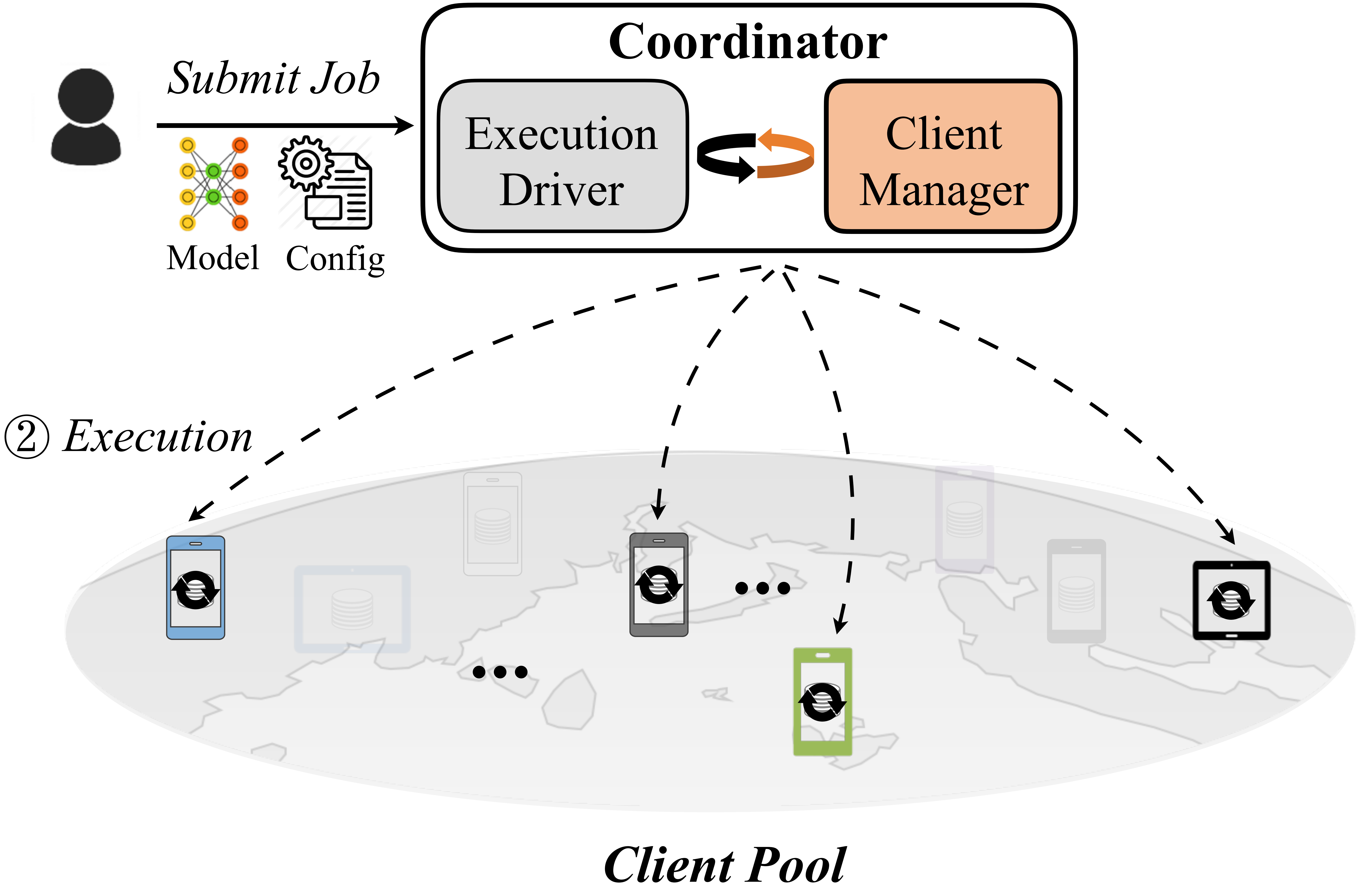
↓

In-situ Execution

↓

Result aggregation

Execution of Federated Learning (FL)

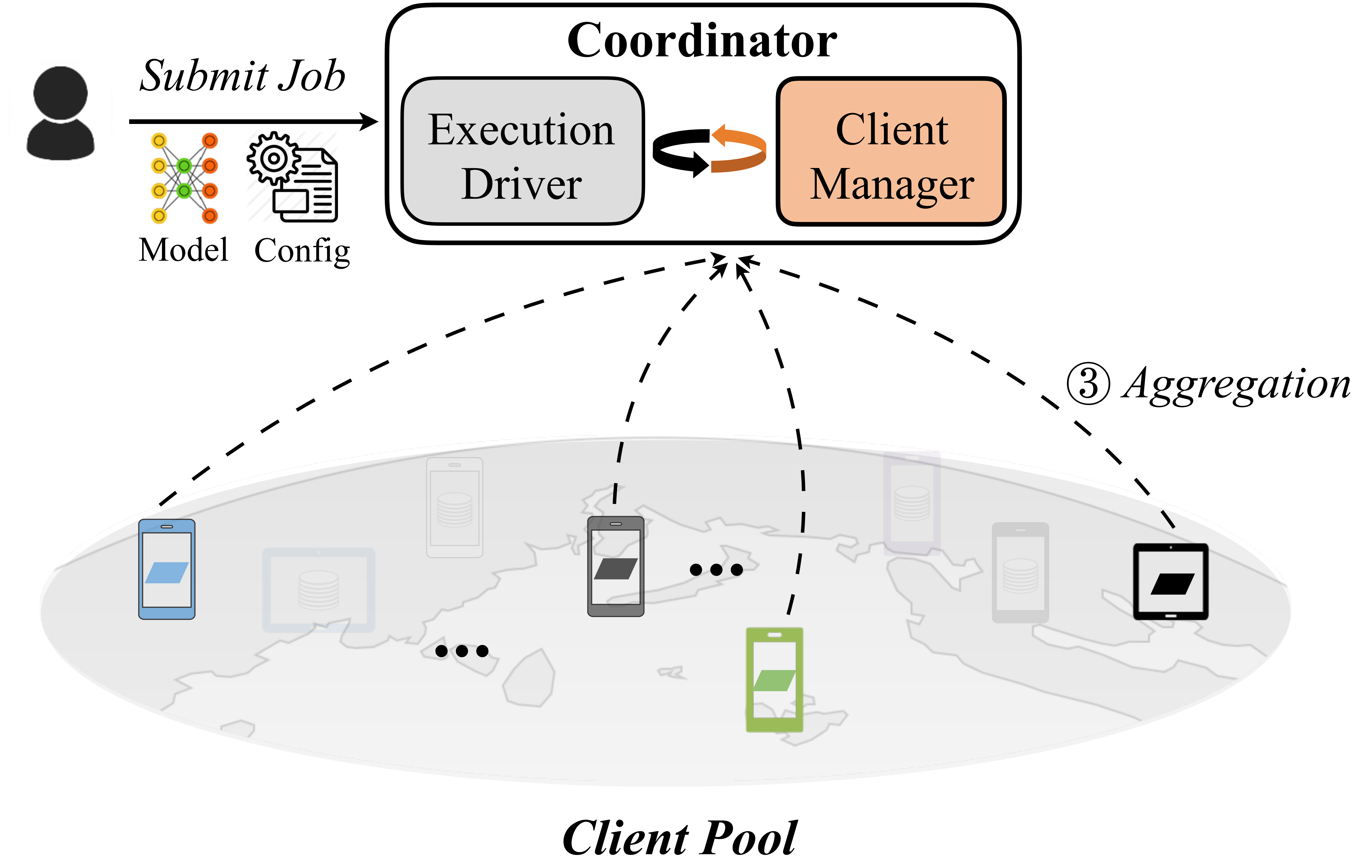


O(100) Rounds:

Round i

- *Client selection*
- ↓
- *In-situ Execution*
- ↓
- *Result aggregation*

Execution of Federated Learning (FL)



O(100) Rounds:

Round *i*

Client selection

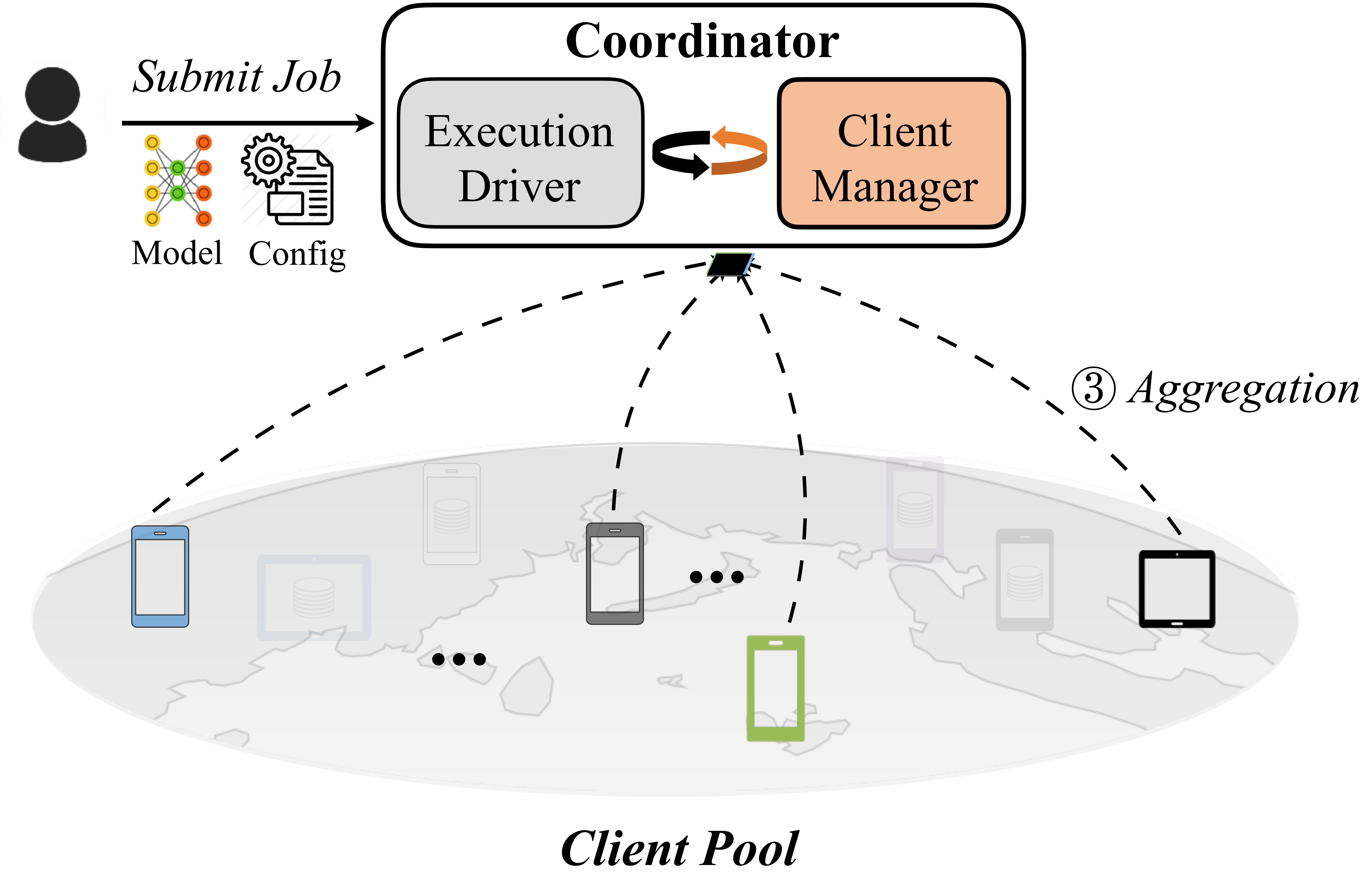
↓

In-situ Execution

↓

Result aggregation

Execution of Federated Learning (FL)

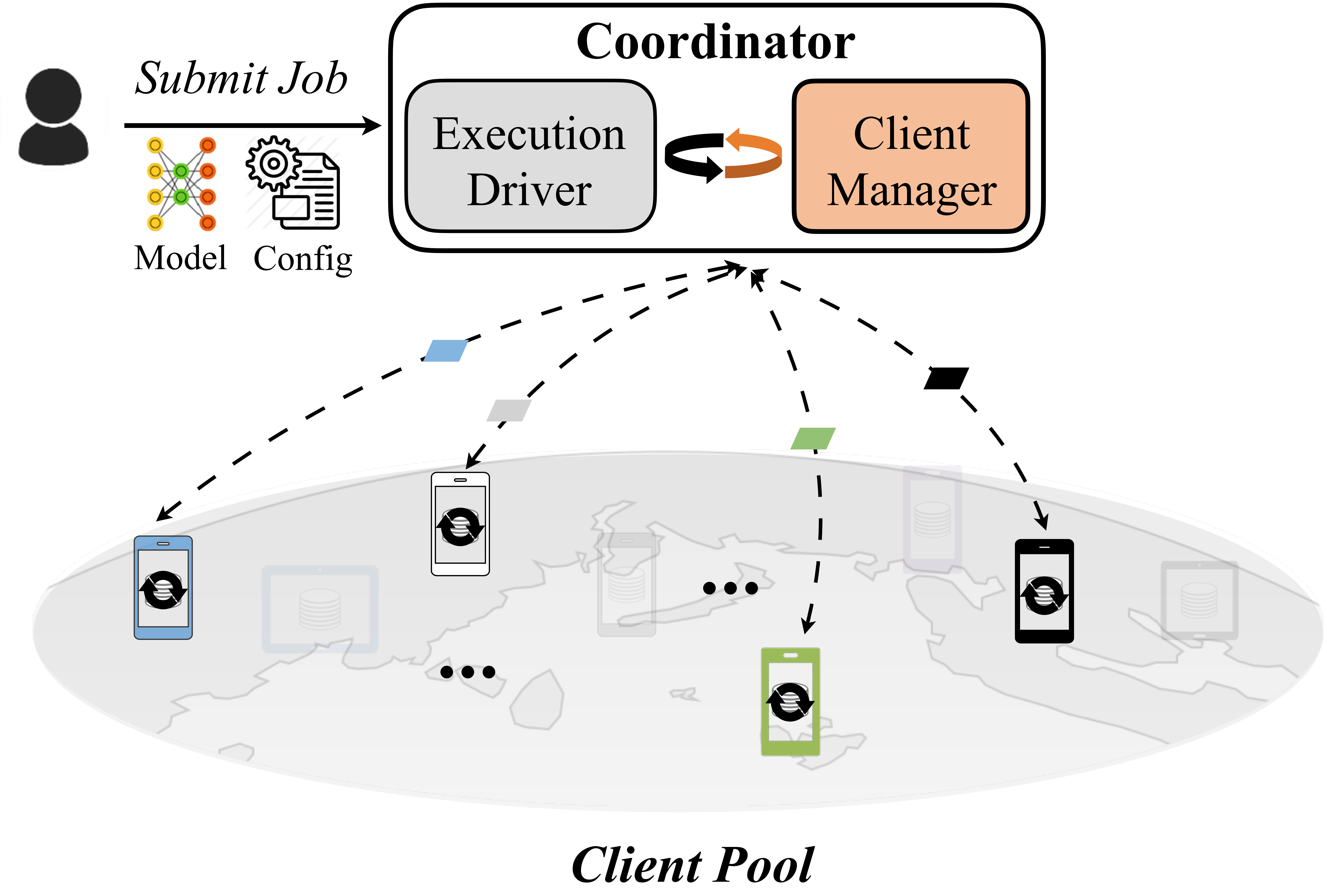


O(100) Rounds:

Round *i*

- *Client selection*
- *In-situ Execution*
- *Result aggregation*

Execution of Federated Learning (FL)



Primary Objective

Better time to accuracy:

- Less time for target acc. under the same setting

O(100) Rounds:

Round i

Client selection

↓

In-situ Execution

↓

Result aggregation

Challenges in Federated Learning

	FL	In-cluster ML
System	Heterogeneous	Homogeneous

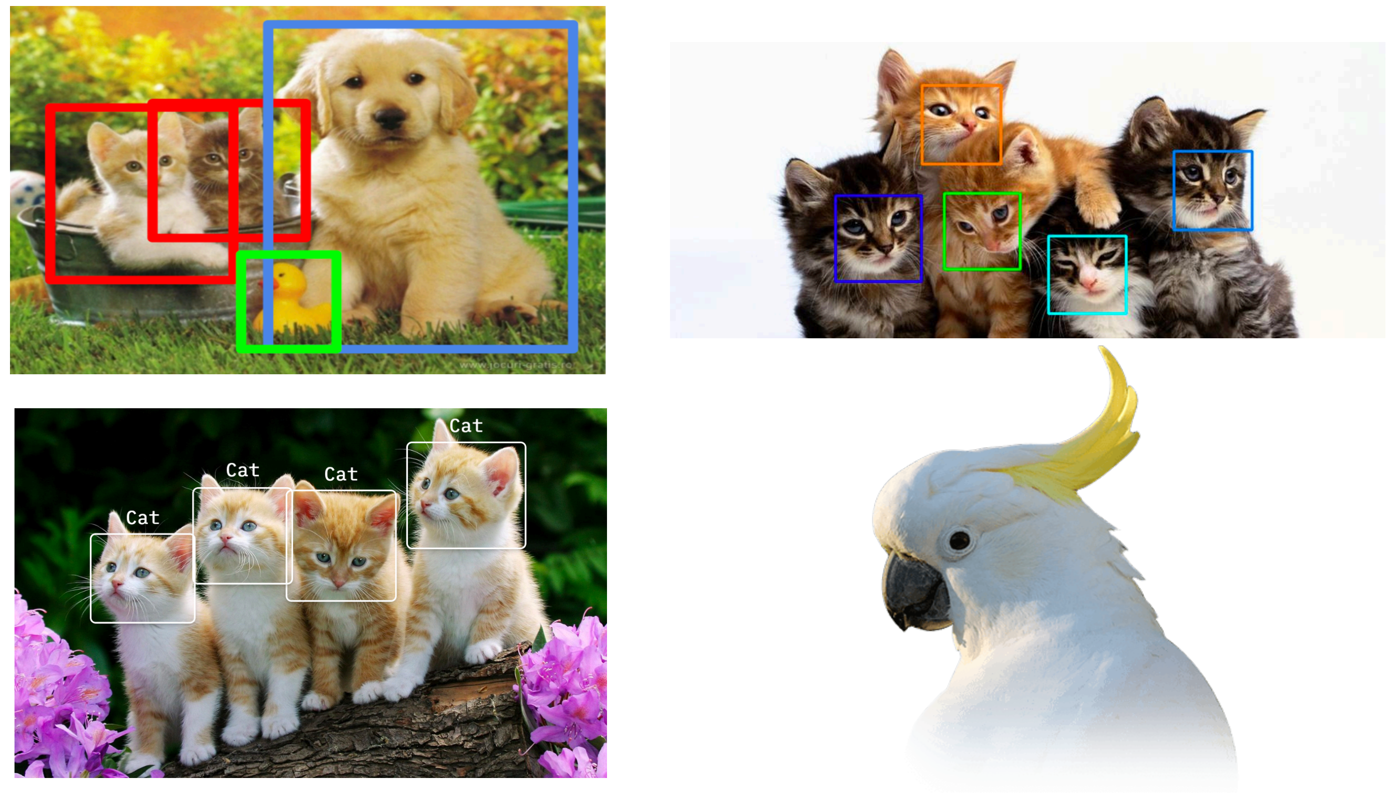
Heterogeneous system speed

Challenges in Federated Learning

	FL	In-cluster ML
System	Heterogeneous	Homogeneous
Data	Heterogeneous	Homogeneous via shuffling
Scale	O(1M)	O(10)
Dynamics	Client can drop out/rejoin	Few
…	…	…

Client A

Client B



Heterogeneous data distribution

Challenges in Federated Learning

	FL	In-cluster ML
System	Heterogeneous	Homogeneous
Data	Heterogeneous	Homogeneous via shuffling
Scale	O(1M)	O(10)
Dynamics	Client can drop out/rejoin	Few
…	…	…

- Existing work optimize for better
  - System efficiency
    - Reduce round duration
  - Statistical efficiency
    - Reduce # of rounds needed
  - …

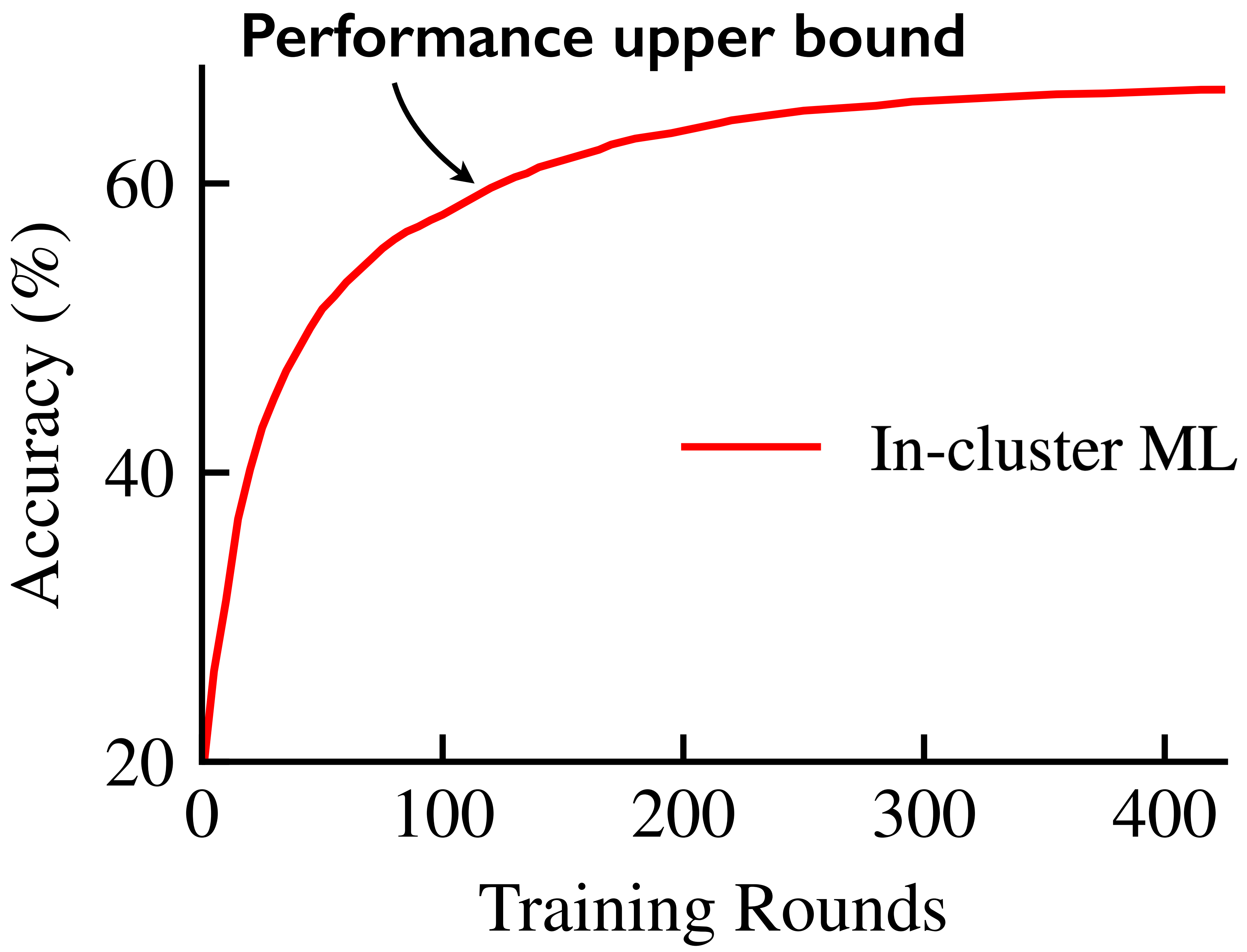
Challenges in Federated Learning

	FL	In-cluster ML
System	Heterogeneous	Homogeneous
Data	Heterogeneous	Homogeneous via shuffling
Scale	O(1M)	O(10)
Dynamics	Client can drop out/rejoin	Few
…	…	…

Existing federated learning relies on random participant selection

Existing Client Selection: Suboptimal Efficiency

Image classification task on OpenImage dataset



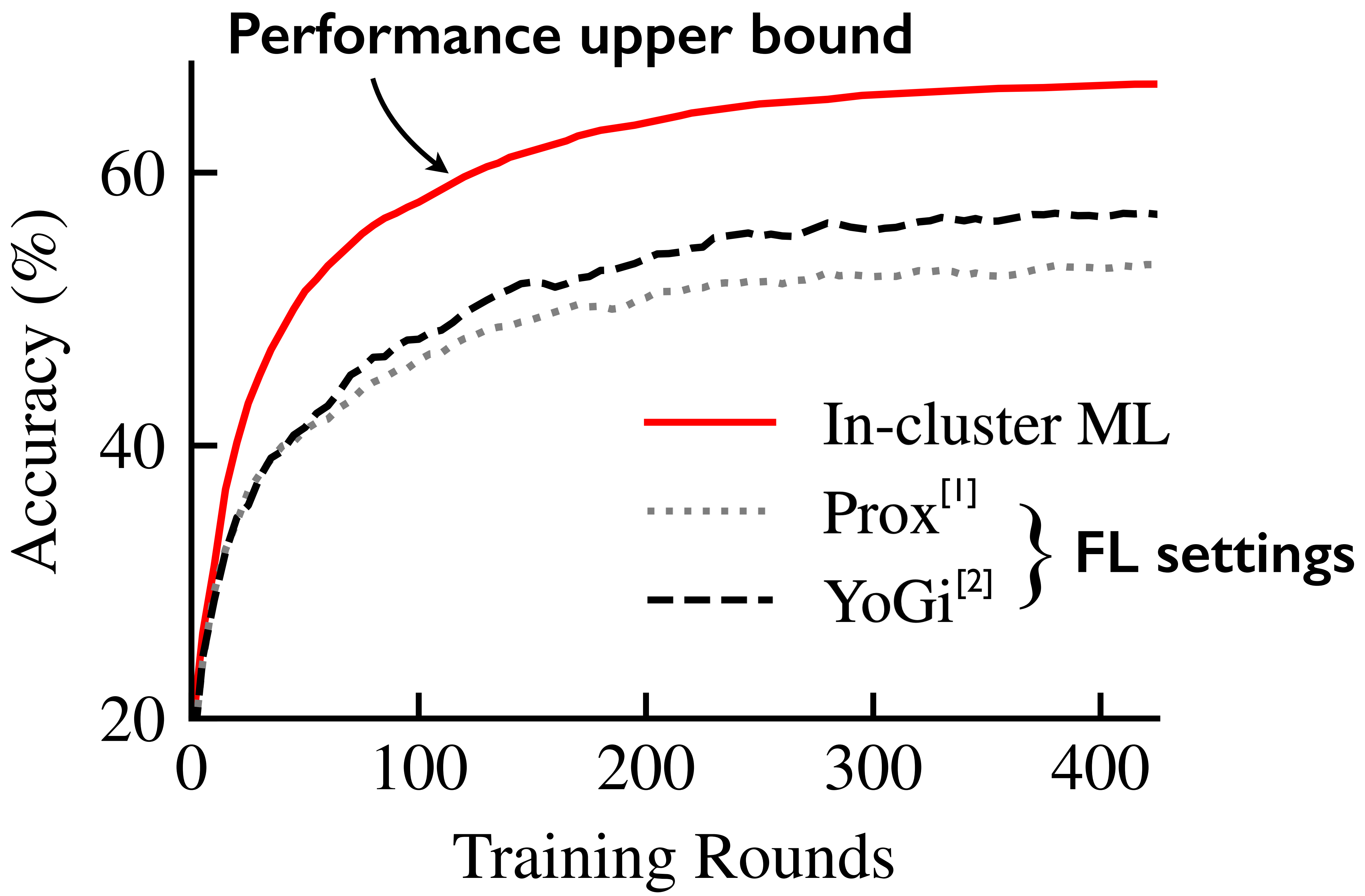
ShuffleNet Model

Problem #1

Overlook heter. client utility

Existing Client Selection: Suboptimal Efficiency

Image classification task on OpenImage dataset



ShuffleNet Model

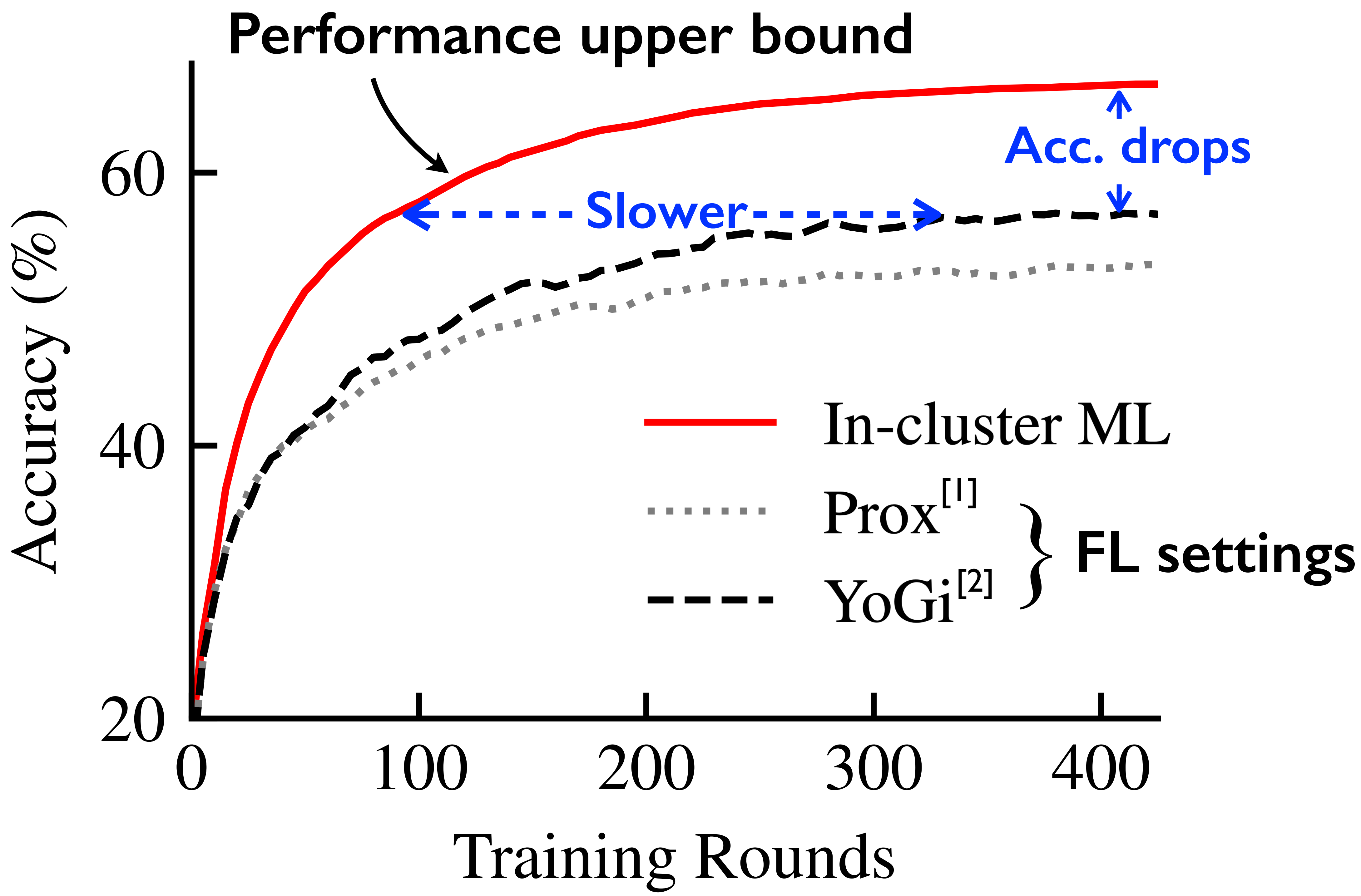
Problem #1

Overlook heter. client utility

[1] "Adaptive Federated Optimization", ICLR'21
[2] "Federated Optimization in Heterogeneous Networks", MLSys'20

Existing Client Selection: Suboptimal Efficiency

Image classification task on OpenImage dataset



ShuffleNet Model

Problem #1

Overlook heter. client utility

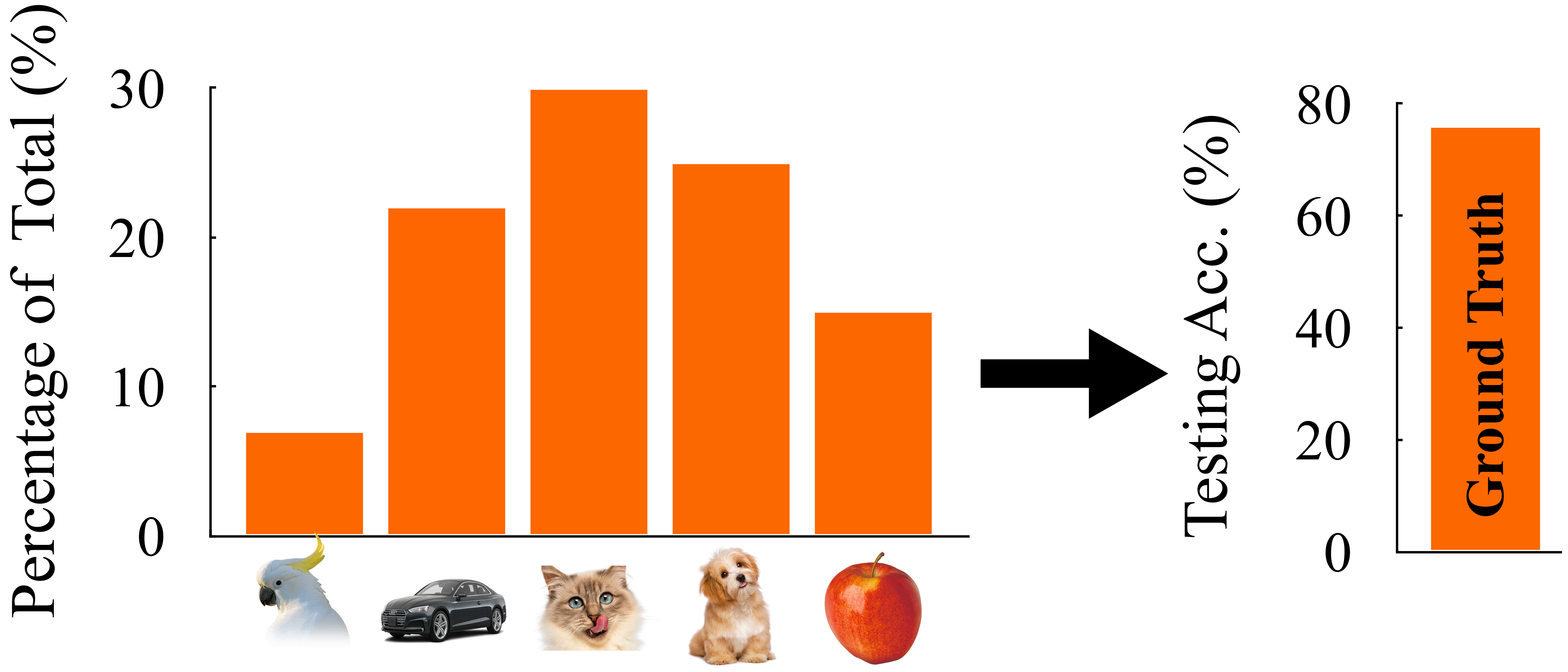
↓

Suboptimal training convergence

[1] "Adaptive Federated Optimization", ICLR'21
[2] "Federated Optimization in Heterogeneous Networks", MLSys'20

Existing Client Selection: Unable for Selection Criteria

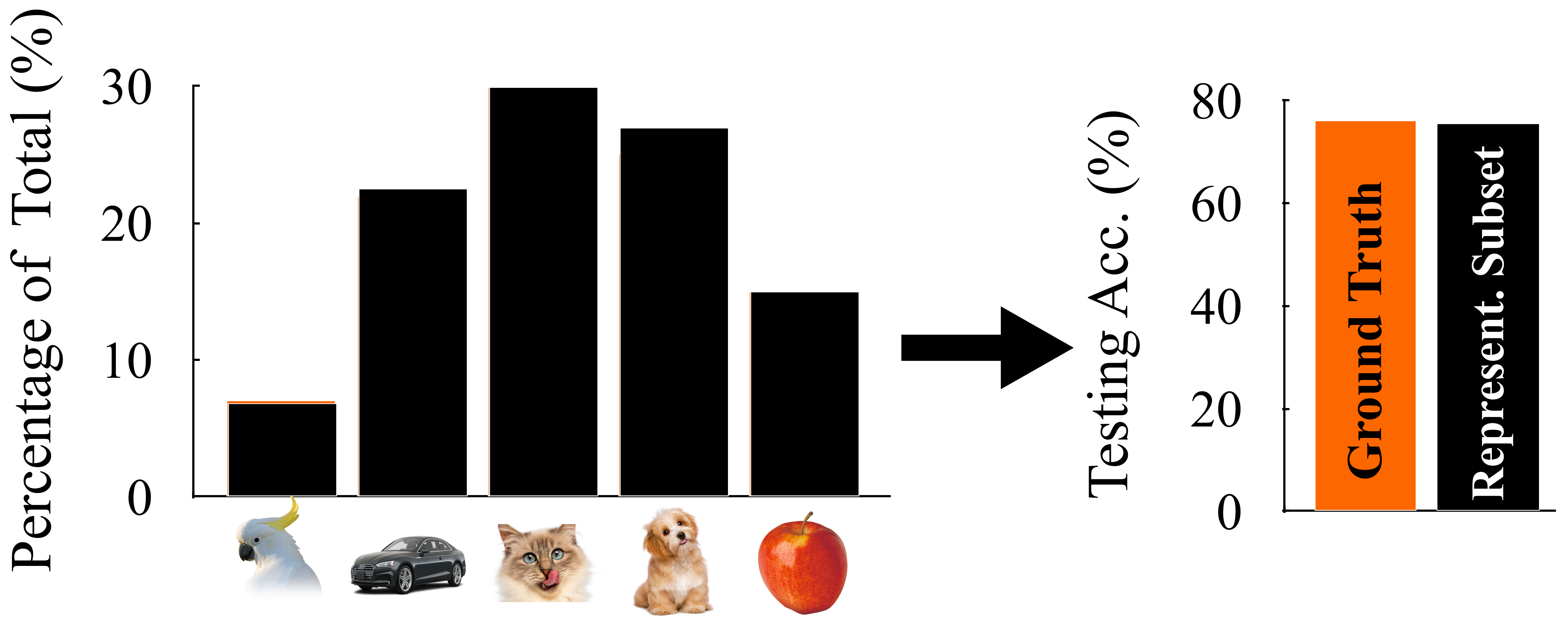
- **Enforcing selection criteria is crucial in FL testing**
  - "Give me 4k representative samples"
  - "Give me x samples of class y"
  - …



(Hypothetical) model testing on all clients ⟶ ground truth

Existing Client Selection: Unable for Selection Criteria

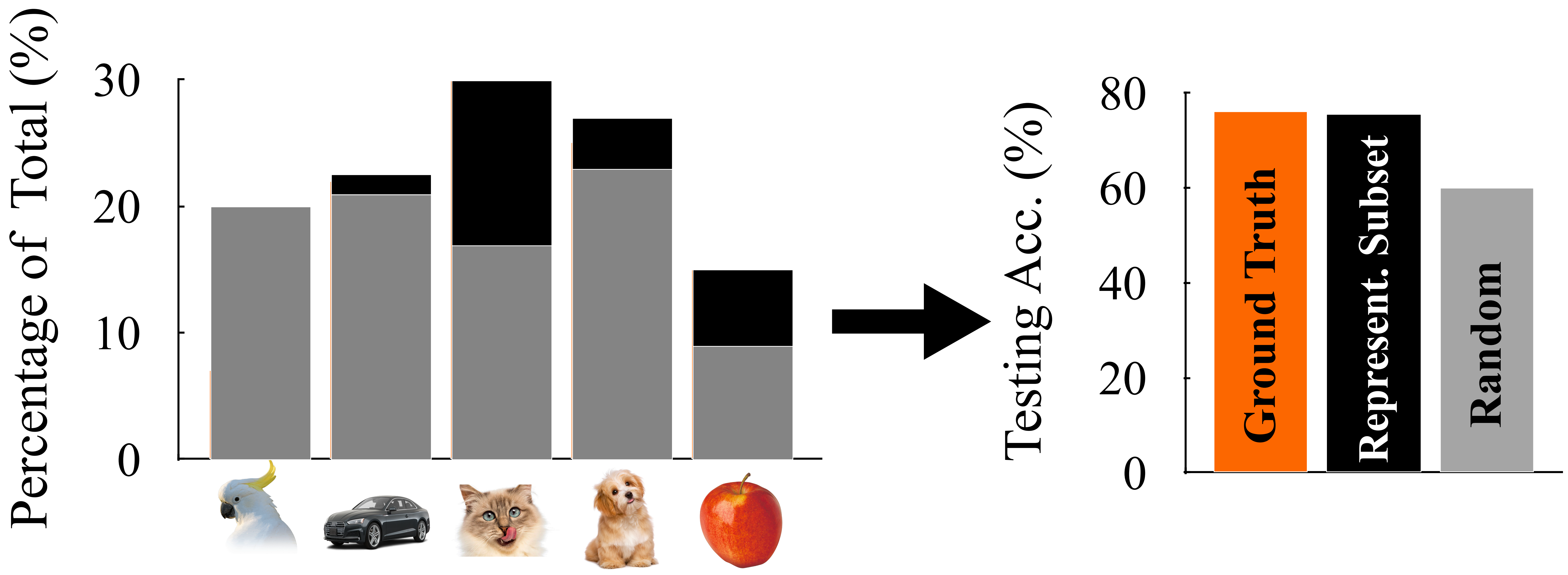
- **Enforcing selection criteria is crucial in FL testing**
  - "Give me 4k representative samples"
  - "Give me x samples of class y"
  - …



Representative Distribution ⟶ useful testing result

Existing Client Selection: Unable for Selection Criteria

- **Enforcing selection criteria is crucial in FL testing**
  - “Give me 4k representative samples”
  - “Give me x samples of class y”
  - …



Random selection ⟶ arbitrary distribution ⟶ useless result

Problem #2

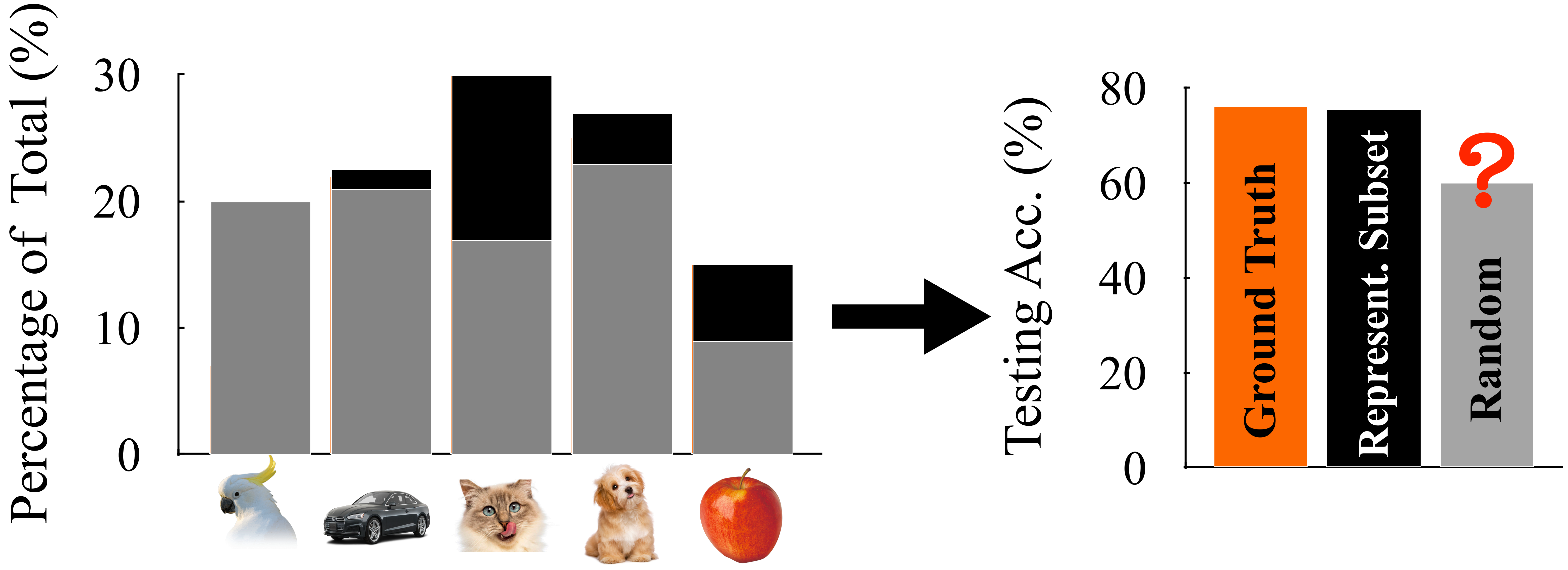
Overlook specified selection criteria

↓

Useless testing results

Existing Client Selection: Unable for Selection Criteria

- **Enforcing selection criteria is crucial in FL testing**
  - "Give me 4k representative samples"
  - "Give me x samples of class y"
  - …



Random selection ⟶ arbitrary distribution ⟶ useless result

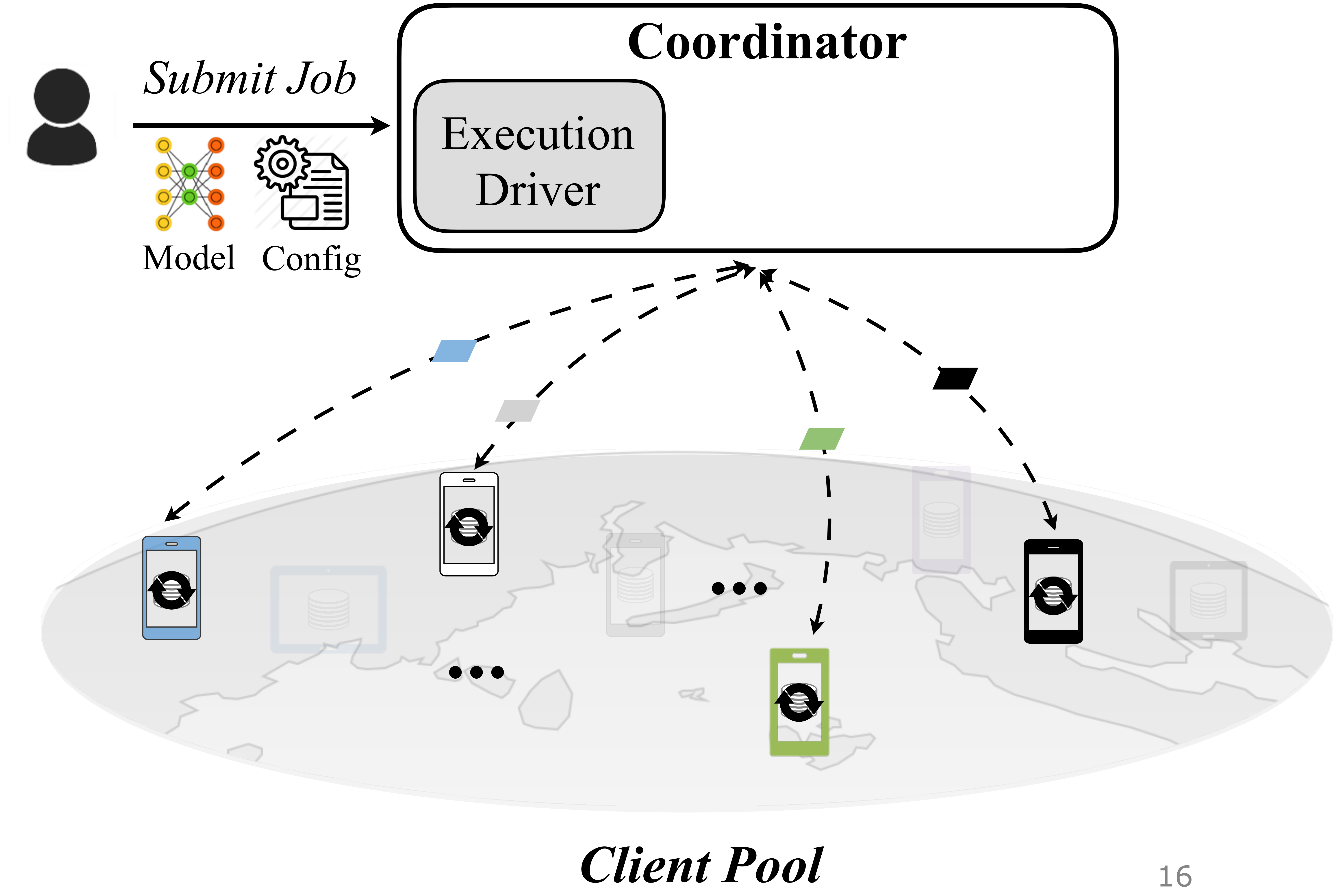
Problem #2

Overlook specified selection criteria

↓

Useless testing results

Oort: Guided Participant Selection for FL

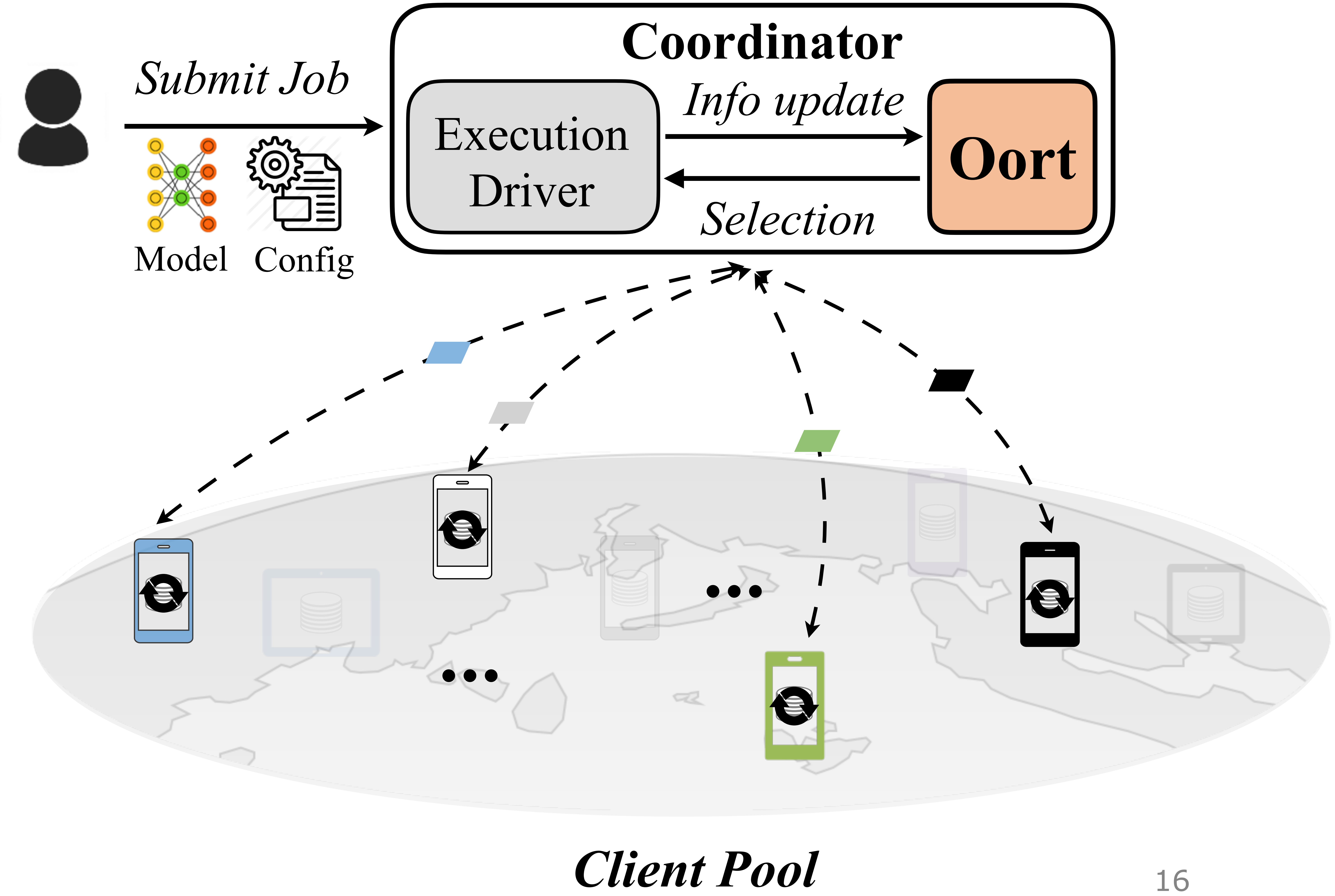


Oort: Guided Participant Selection for FL

Design Overview

- *Enable faster FL training*
  - Adaptively explore and exploit high-utility clients

- *Support interpretable FL testing*
  - Enforce developer-specified data selection criteria at scale

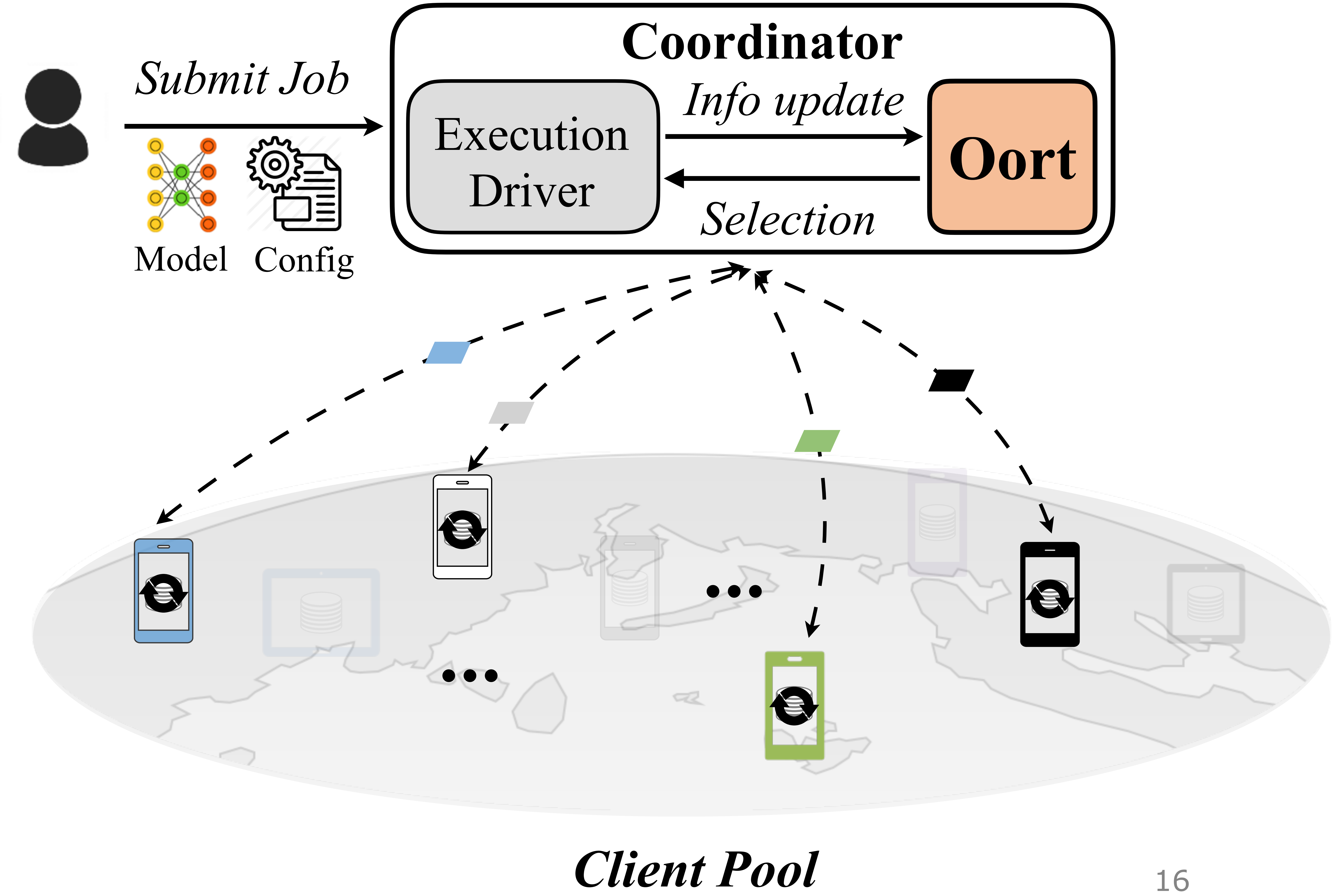


Oort: Guided Participant Selection for FL

Design Overview

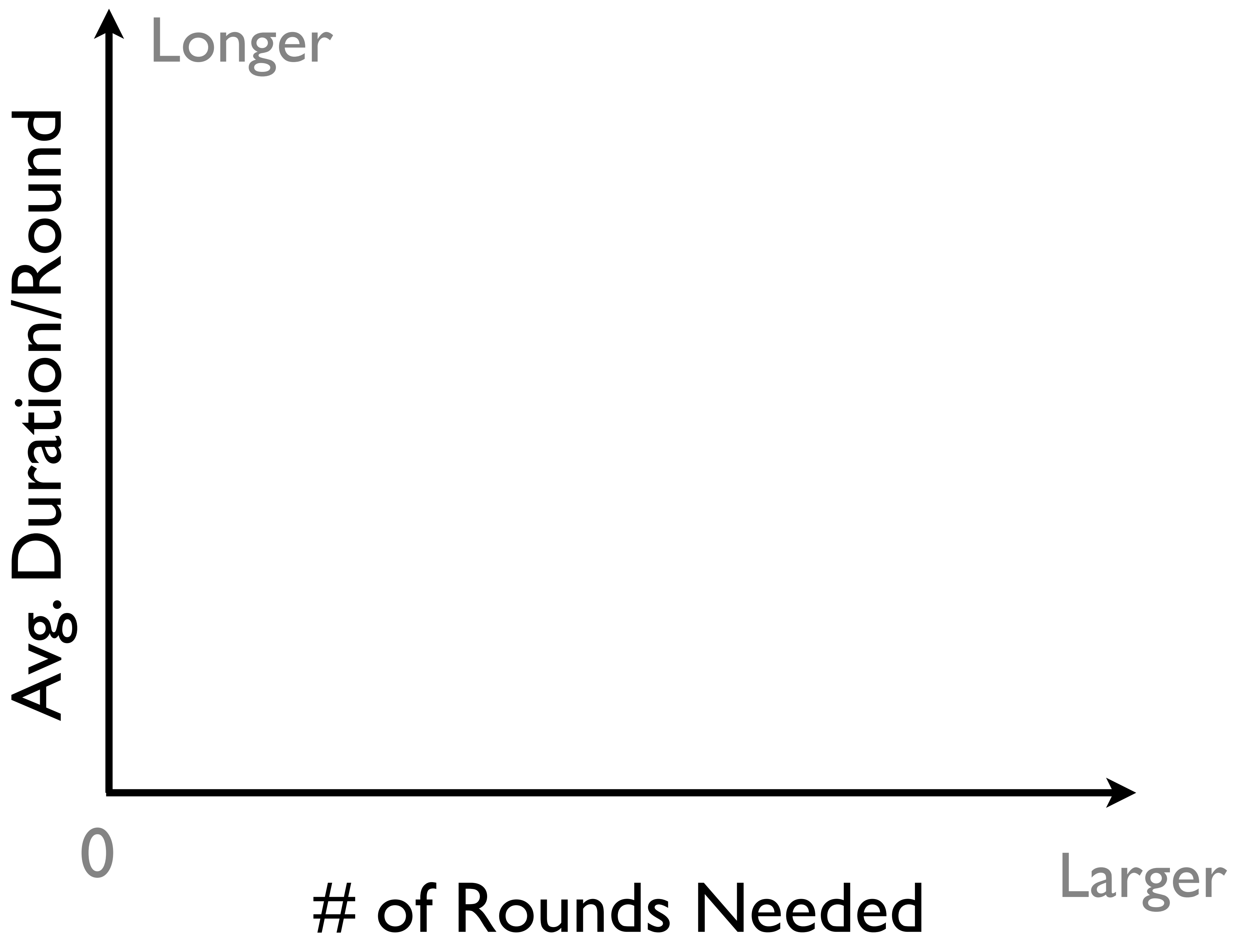
- *Enable faster FL training*
  - Adaptively explore and exploit high-utility clients

- *Support interpretable FL testing*
  - Enforce developer-specified data selection criteria at scale



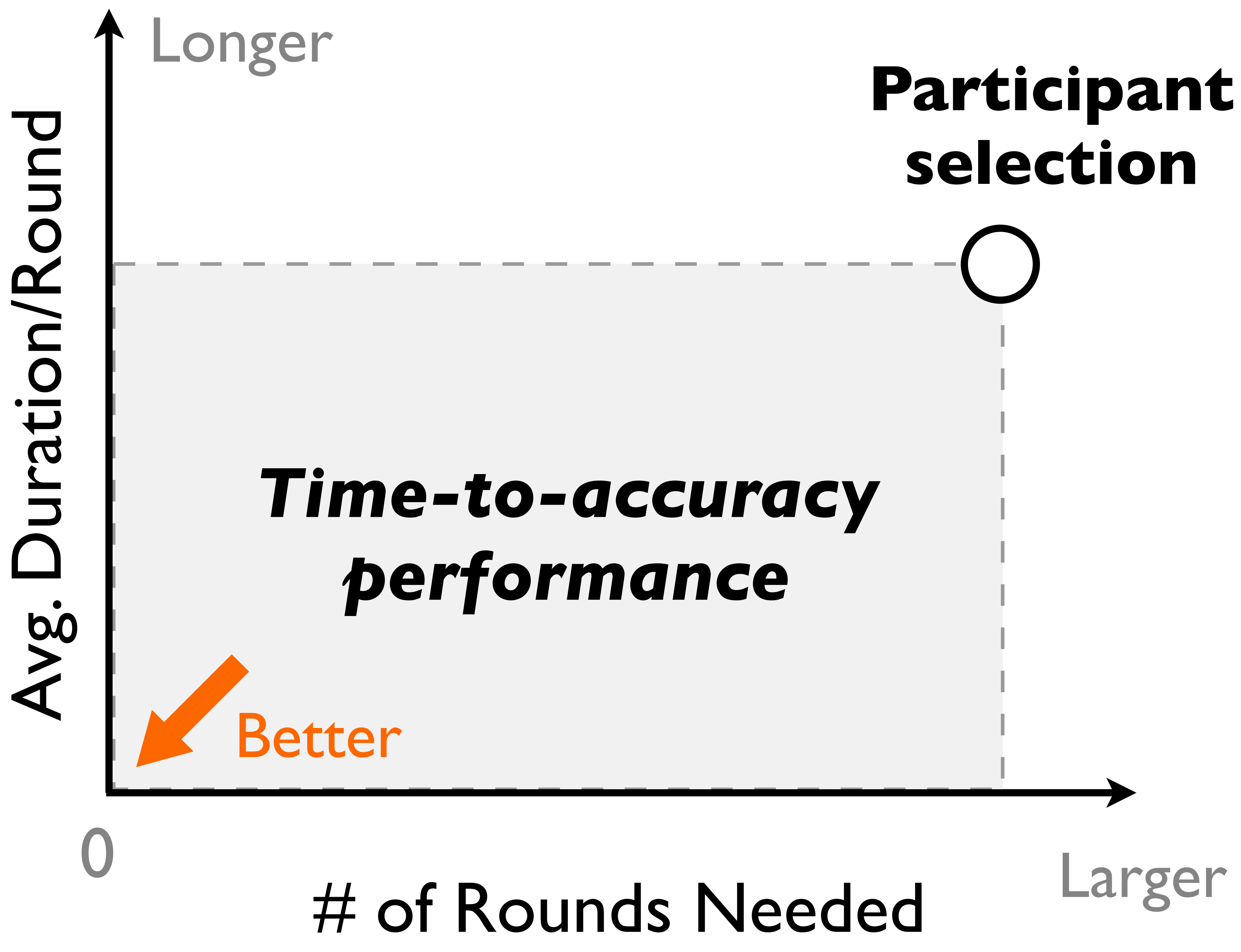
Anatomy of Time to Accuracy in Training

- **System efficiency (round duration)**
  - Determined by client ***system speed***
- **Statistical efficiency (round to accuracy)**
  - Determined by client ***data***



Anatomy of Time to Accuracy in Training

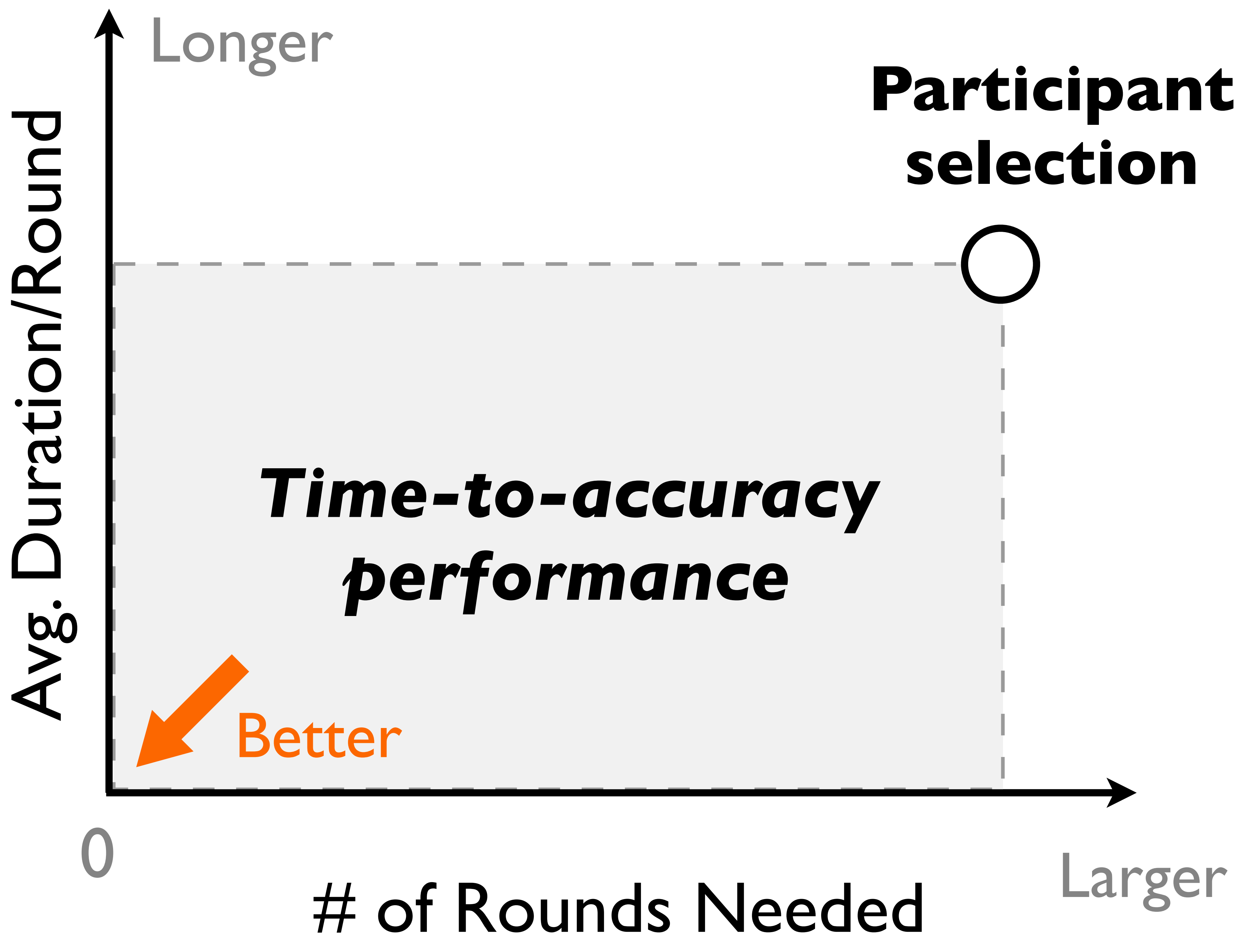
- System efficiency (round duration)
  - Determined by client *system speed*
- Statistical efficiency (round to accuracy)
  - Determined by client *data*



Anatomy of Time to Accuracy in Training

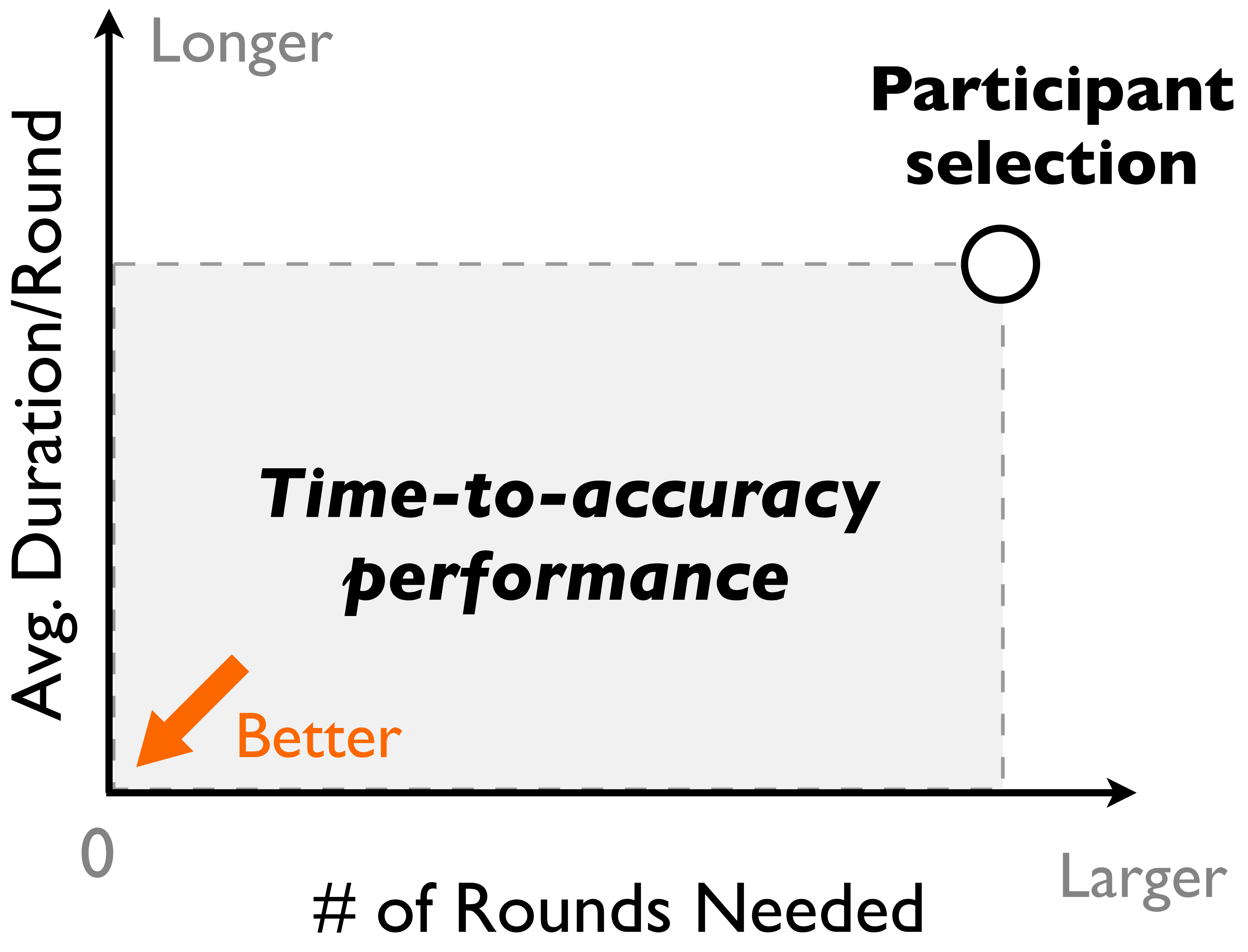
- System efficiency (round duration)
  - Determined by client *system speed*
- Statistical efficiency (round to accuracy)
  - Determined by client *data*

Client utility
- **System utility** (round duration)
- **Statistical utility**



Anatomy of Time to Accuracy in Training

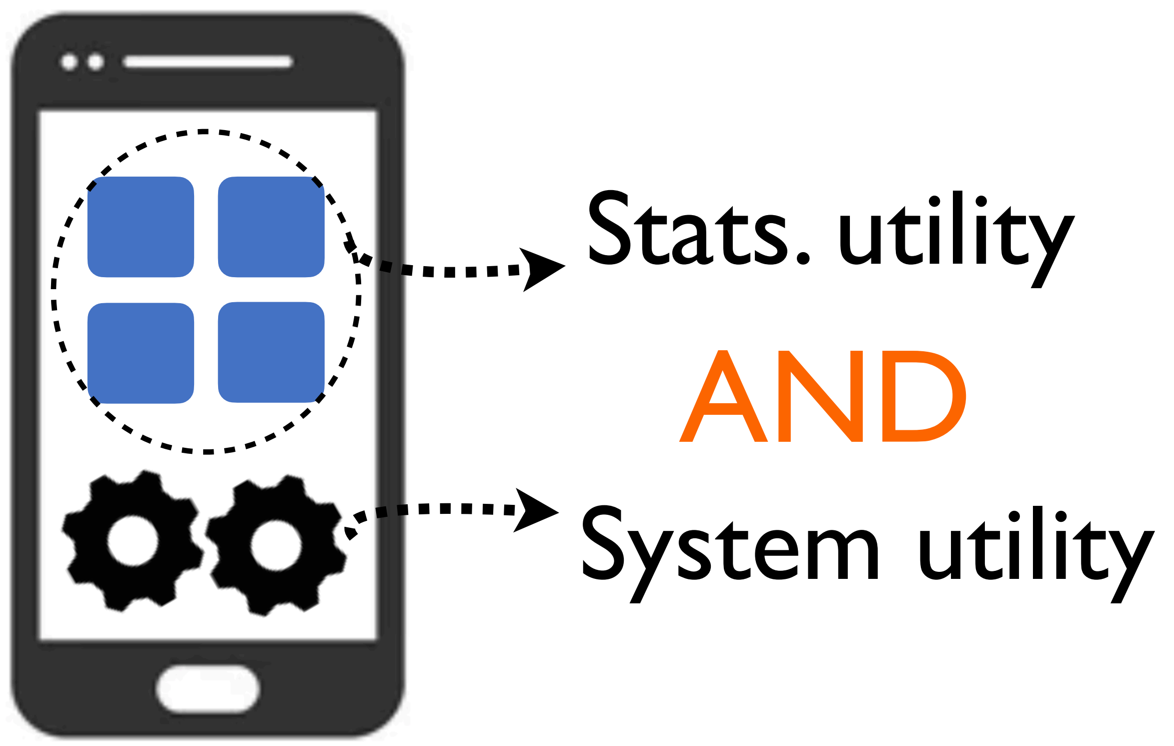
- System efficiency (round duration)
  - Determined by client *system speed*
- Statistical efficiency (round to accuracy)
  - Determined by client *data*



Client utility
- **System utility** (round duration)
- **Statistical utility**: how data helps round to accuracy?

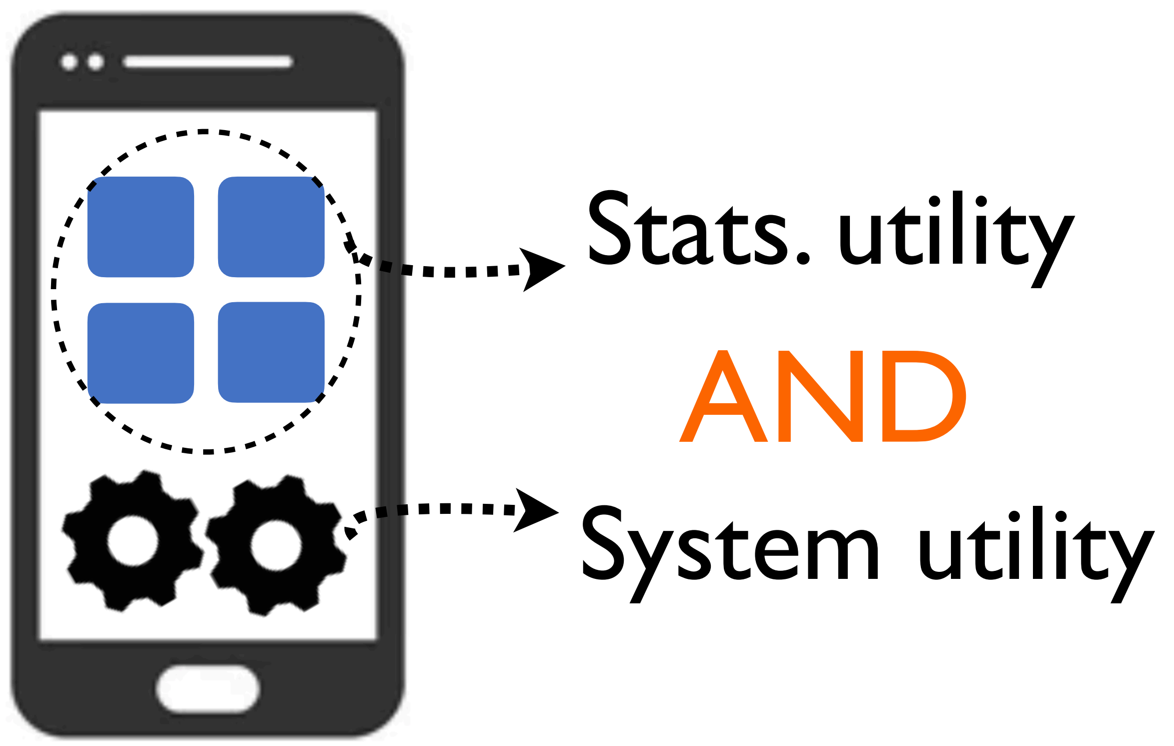
Challenge 1: Identify Heterogeneous Client Utility

- **Statistical utility**
  - Capture how the client data can help to improve the model



Challenge 1: Identify Heterogeneous Client Utility

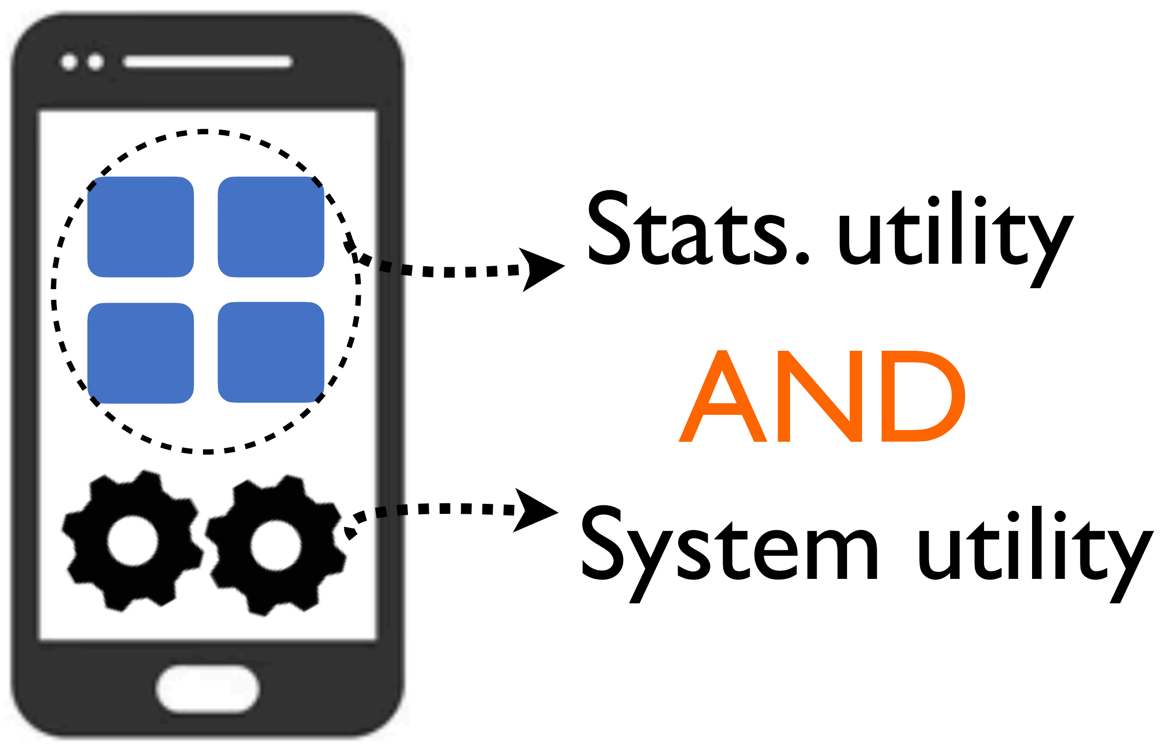
- **Statistical utility**
  - Capture how the client data can help to improve the model
  - **Metric**: ***aggregate training loss*** of client data
    - Higher loss ⟶ higher stats utility (proof in paper)



Challenge 1: Identify Heterogeneous Client Utility

- **Statistical utility**
  - Capture how the client data can help to improve the model
  - **Metric**: ***aggregate training loss*** of client data
    - Higher loss ⟶ higher stats utility (proof in paper)

- **Utility of a client =** $\dfrac{stats\_util\ (i)}{round\_duration\ (i)}$
  - i.e., *speed* of accumulating stats utility in ***round i***



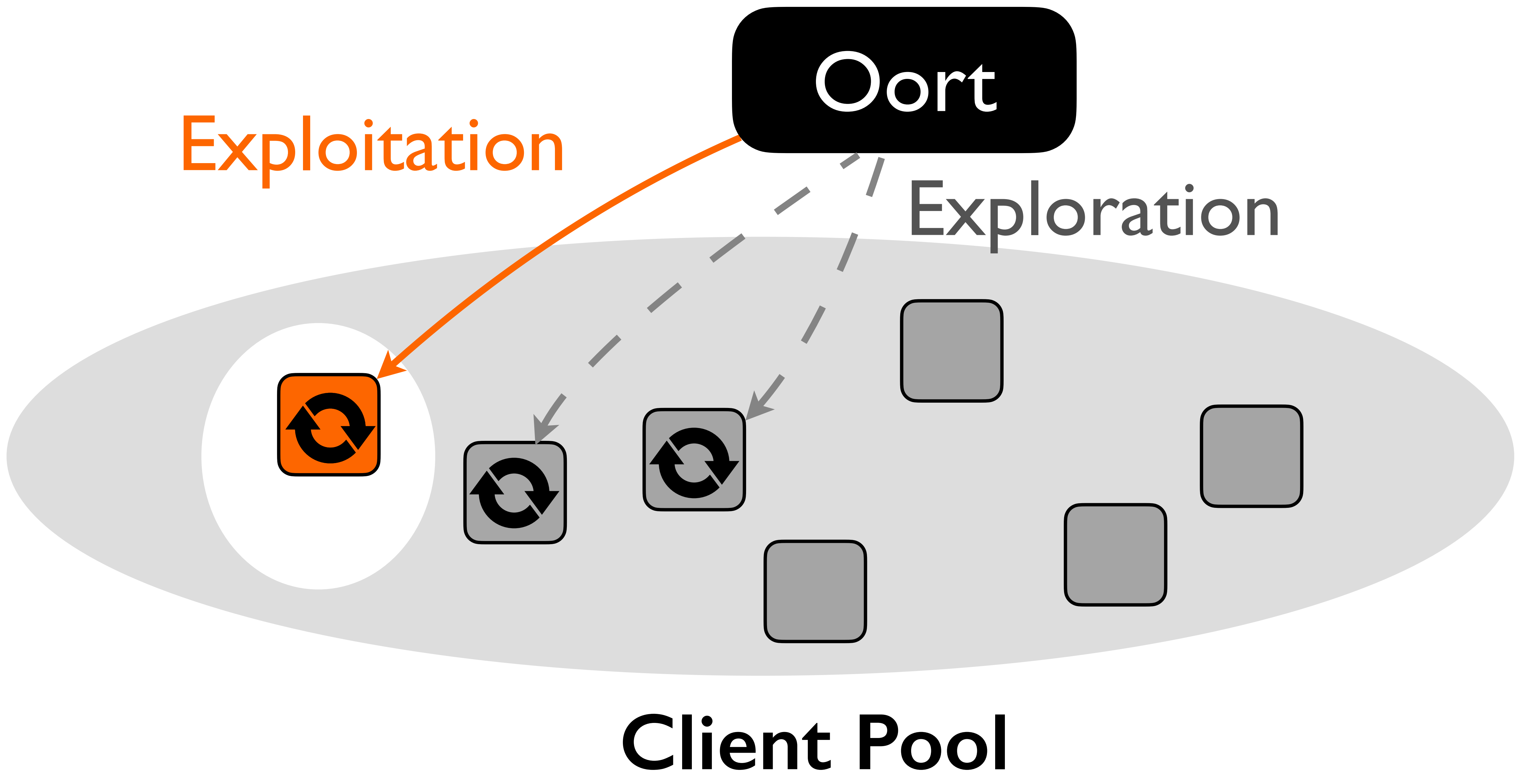
Challenge 2: Select High-Utility Clients at Scale

- **How to identify high-utility clients from millions of clients?**
  - *Spatiotemporal* variation: heterogeneous utility across clients over rounds

Challenge 2: Select High-Utility Clients at Scale

- **How to identify high-utility clients from millions of clients?**
  - *Spatiotemporal* variation: heterogeneous utility across clients over rounds

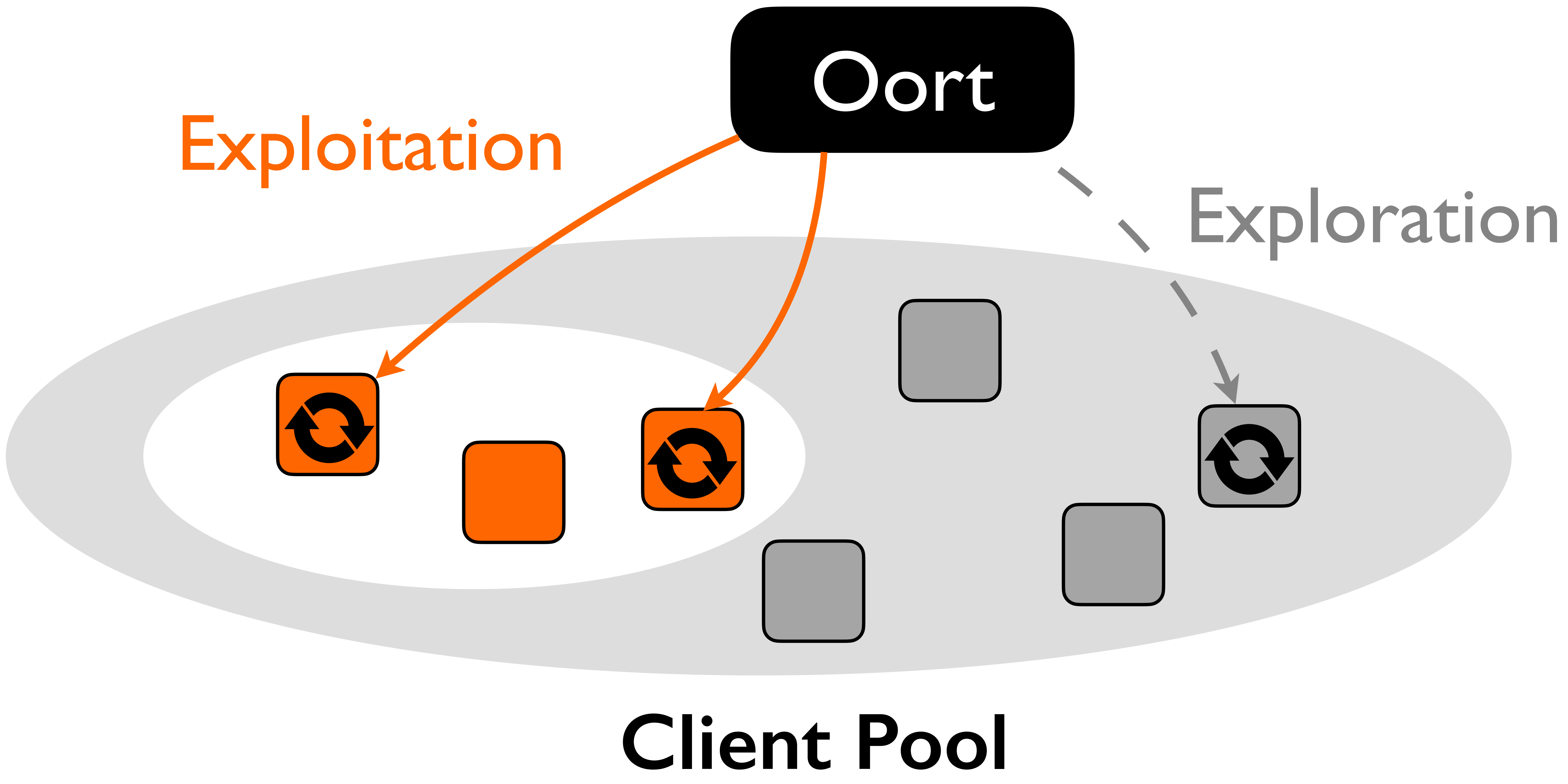
- **Exploration + Exploitation**
  - Explore not-tried clients



Challenge 2: Select High-Utility Clients at Scale

- **How to identify high-utility clients from millions of clients?**
  - *Spatiotemporal* variation: heterogeneous utility across clients over rounds

- **Exploration + Exploitation**
  - Explore not-tried clients
  - Exploit known *high-utility* clients



Challenge 3: Select High-Utility Clients Adaptively

- **How to account for stale utility since last participation?**
  - Utility changes due to dynamics

Challenge 3: Select High-Utility Clients Adaptively

- **How to account for stale utility since last participation?**
  - Utility changes due to dynamics

1. Aging: add uncertainty to utility
   - current_utility = last_observed_utility + *observation_age*

Challenge 3: Select High-Utility Clients Adaptively

- **How to account for stale utility since last participation?**
  - Utility changes due to dynamics

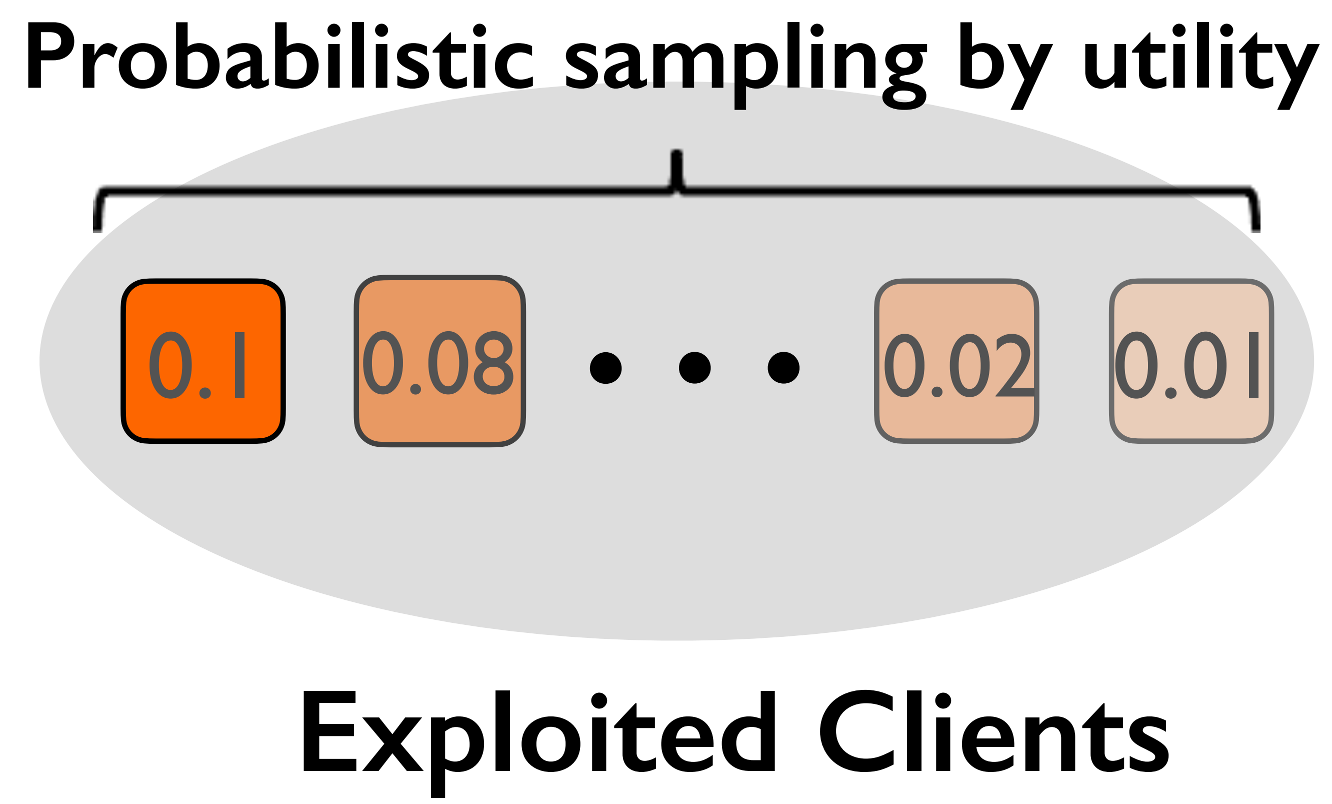
1. Aging: add uncertainty to utility ⟶ ***Re-discover missed good clients***
   - current_utility = last_observed_utility + *observation_age*

Challenge 3: Select High-Utility Clients Adaptively

- **How to account for stale utility since last participation?**
  - Utility changes due to dynamics

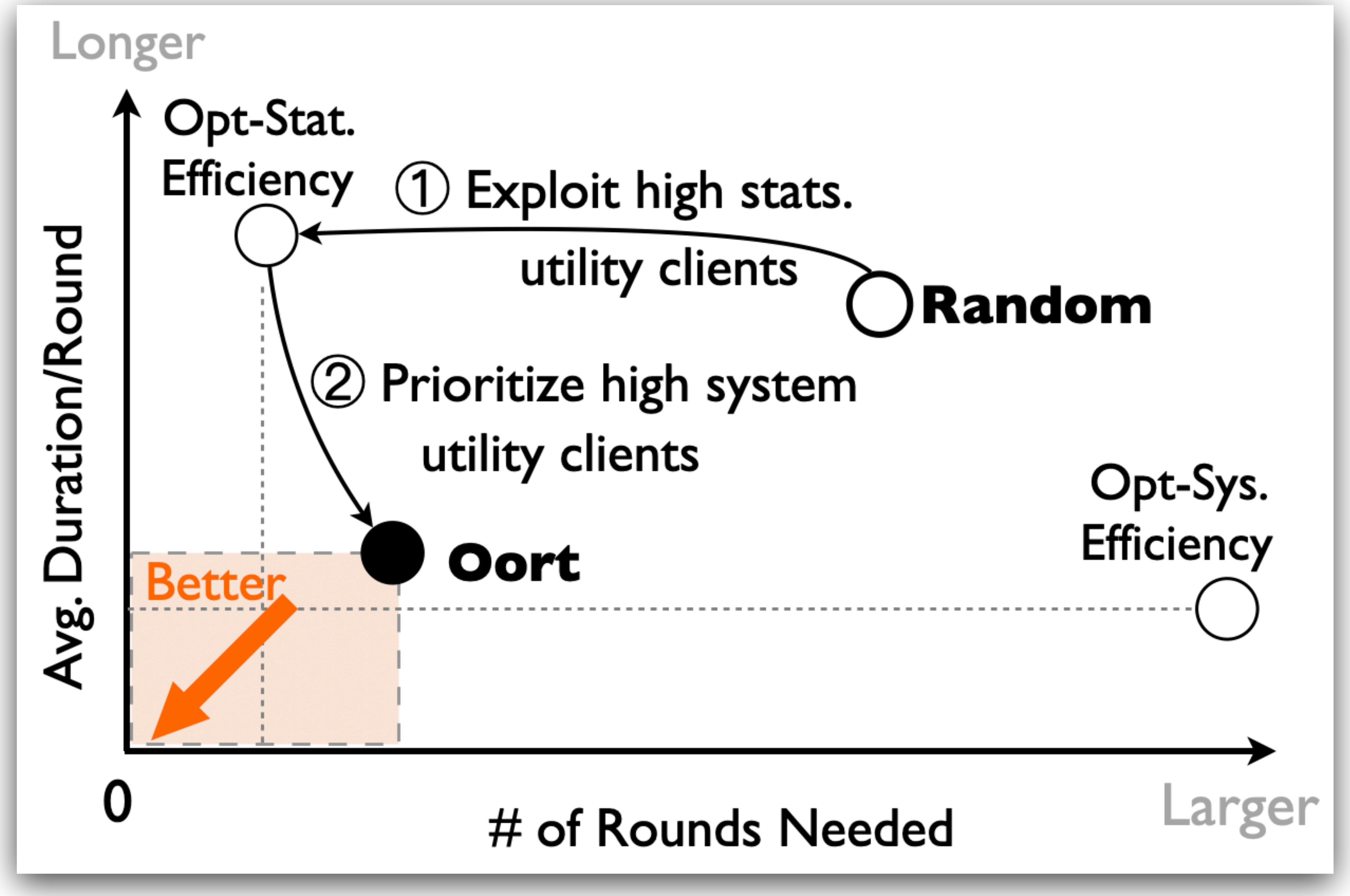
1. Aging: add uncertainty to utility ⟶ *Re-discover missed good clients*
   - current_utility = last_observed_utility + *observation_age*

2. **Probabilistic selection** by utility values
   - Prioritize high-utility clients
   - Robust to outliers and uncertainties



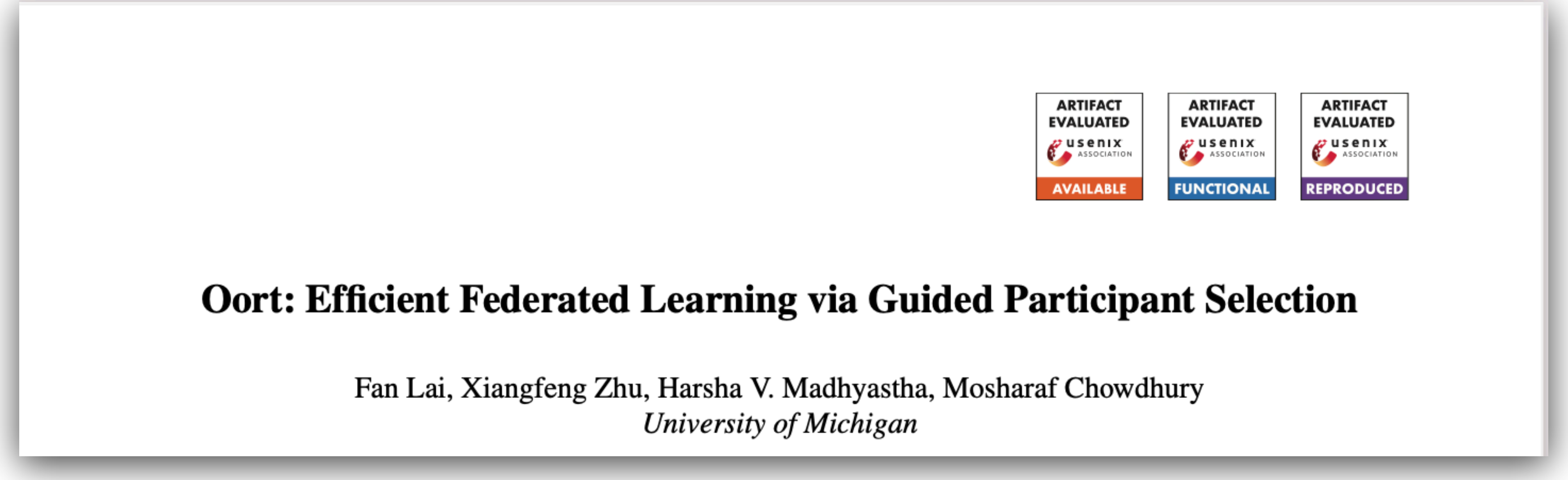
More in Our Paper

- How to respect privacy
- How to be robust to corrupted clients
- How to enforce diverse selection criteria
  - Fairness, data distribution for FL testing



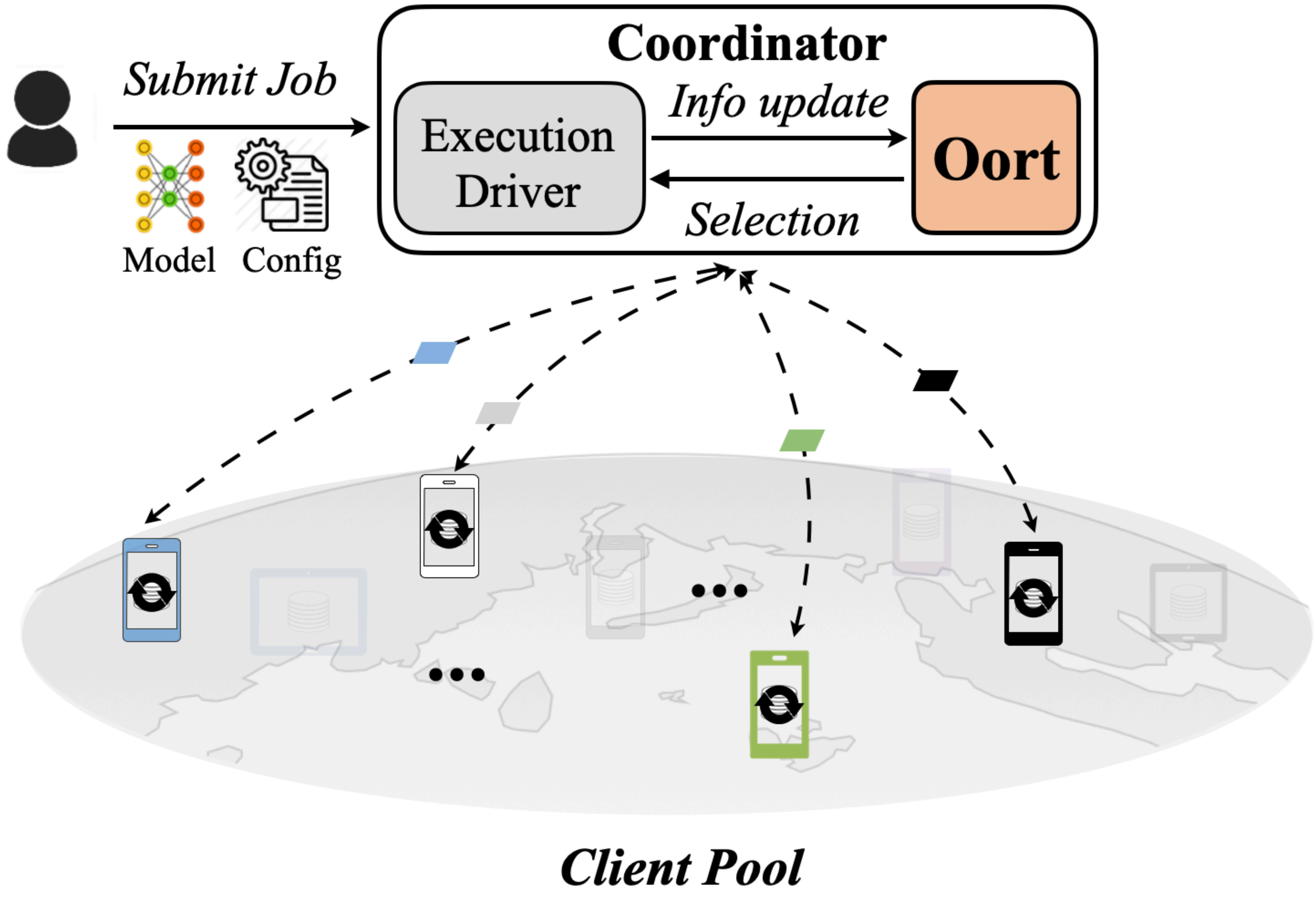
Evaluation

Oort as a lib to support TensorFlow Federated / PySyft



Oort: Efficient Federated Learning via Guided Participant Selection

Fan Lai, Xiangfeng Zhu, Harsha V. Madhyastha, Mosharaf Chowdhury
University of Michigan



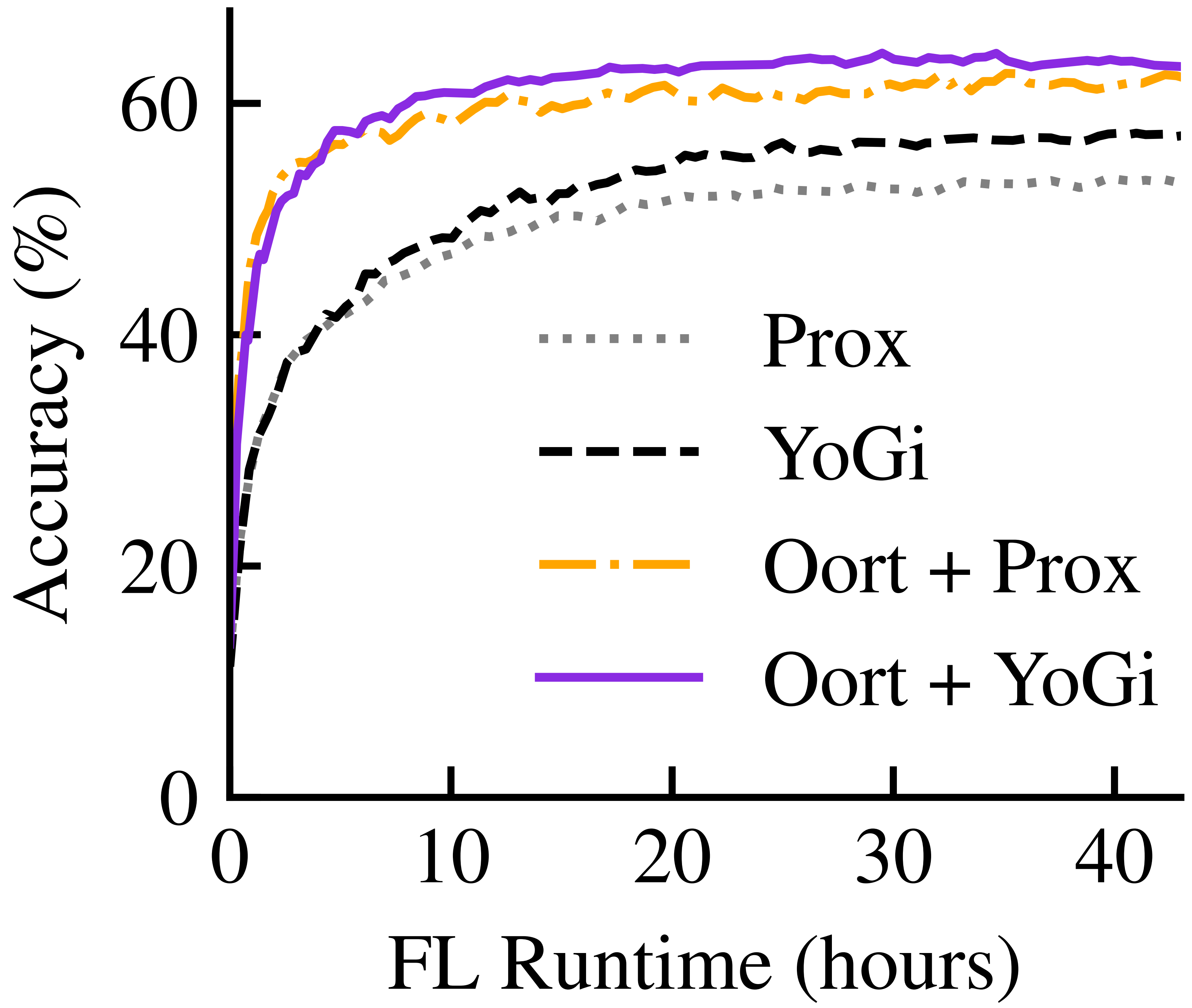
Experiment setting

- Testbed w/ 68 GPUs
- Realistic FL Benchmark[1]
  - Heter. speed/data
  - Dynamics of devices
- 1300 participants/round

[1] FedScale: Benchmarking Model and System Performance of Federated Learning

Time-to-Accuracy (TTA) Performance

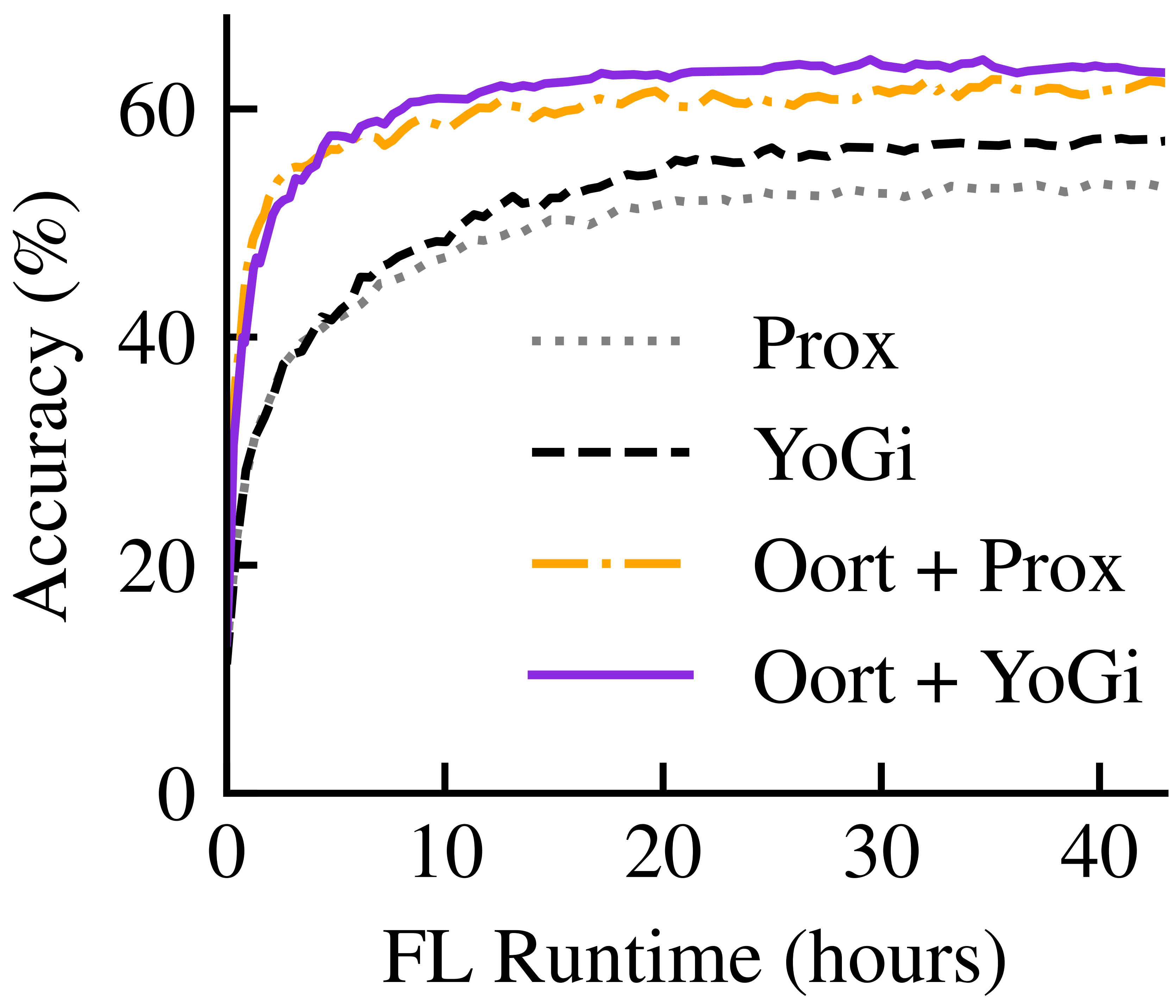
Image classification (OpenImage dataset)



ShuffleNet Model

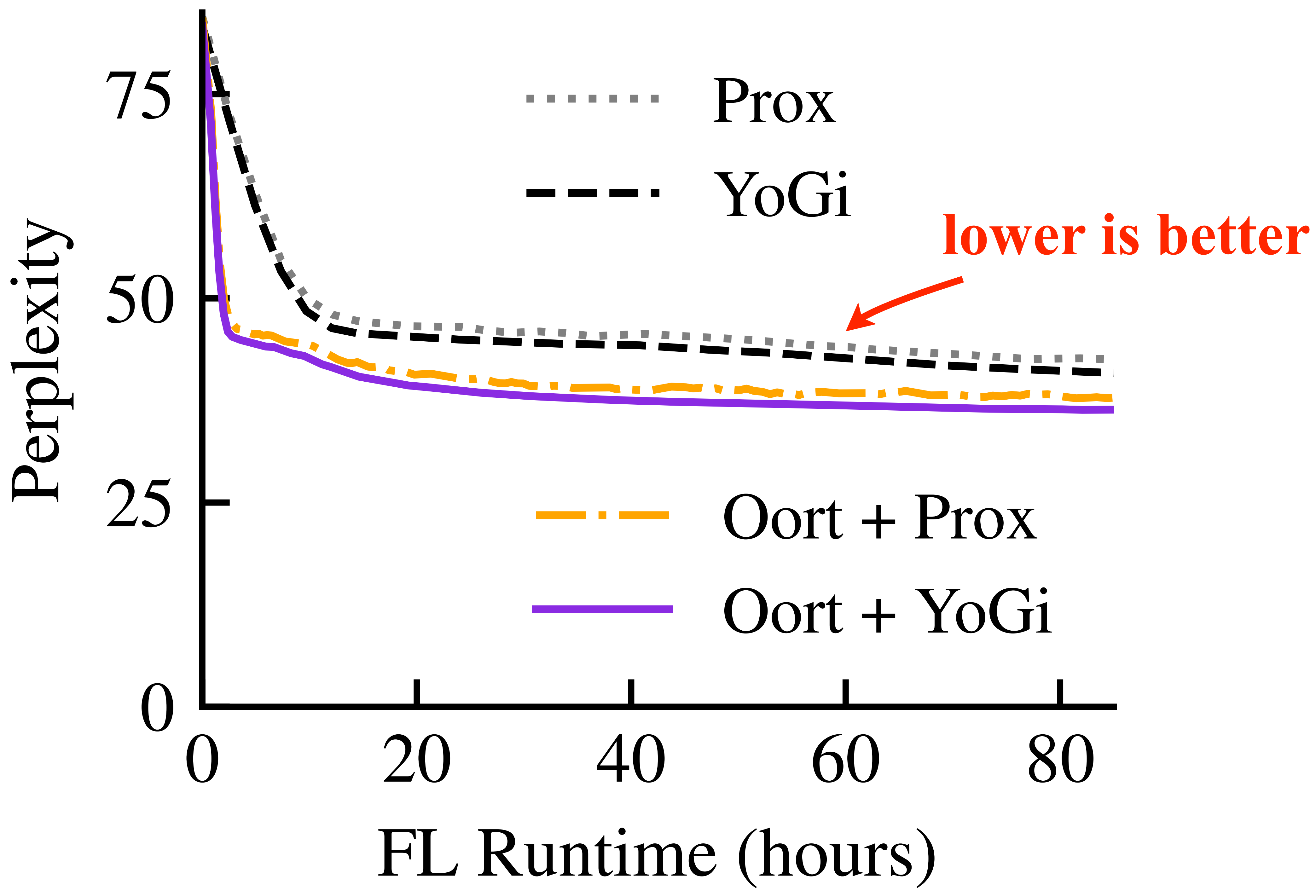
Time-to-Accuracy (TTA) Performance

Image classification (OpenImage dataset)



ShuffleNet Model

Next-word prediction (Reddit Corpus)

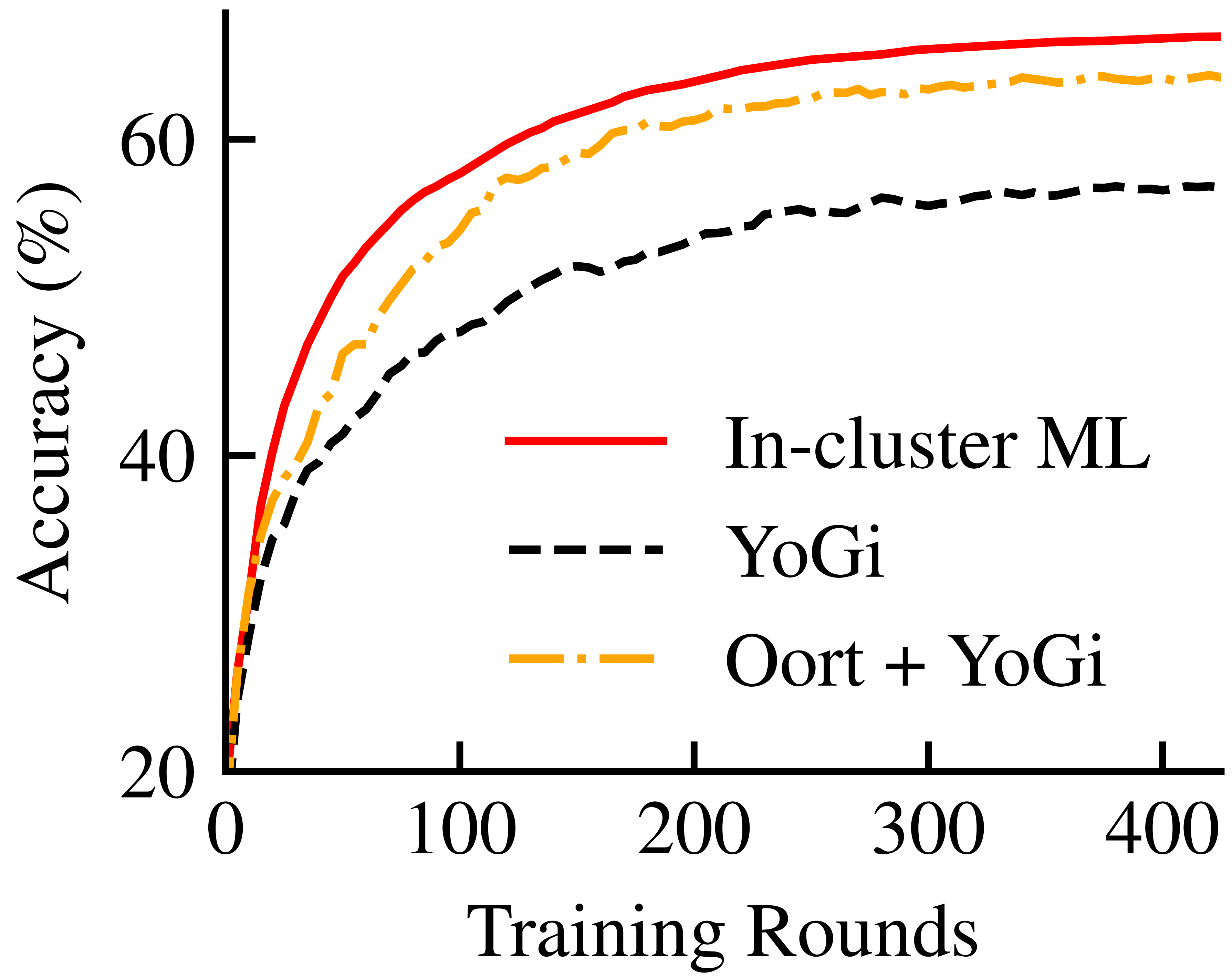


Albert Model

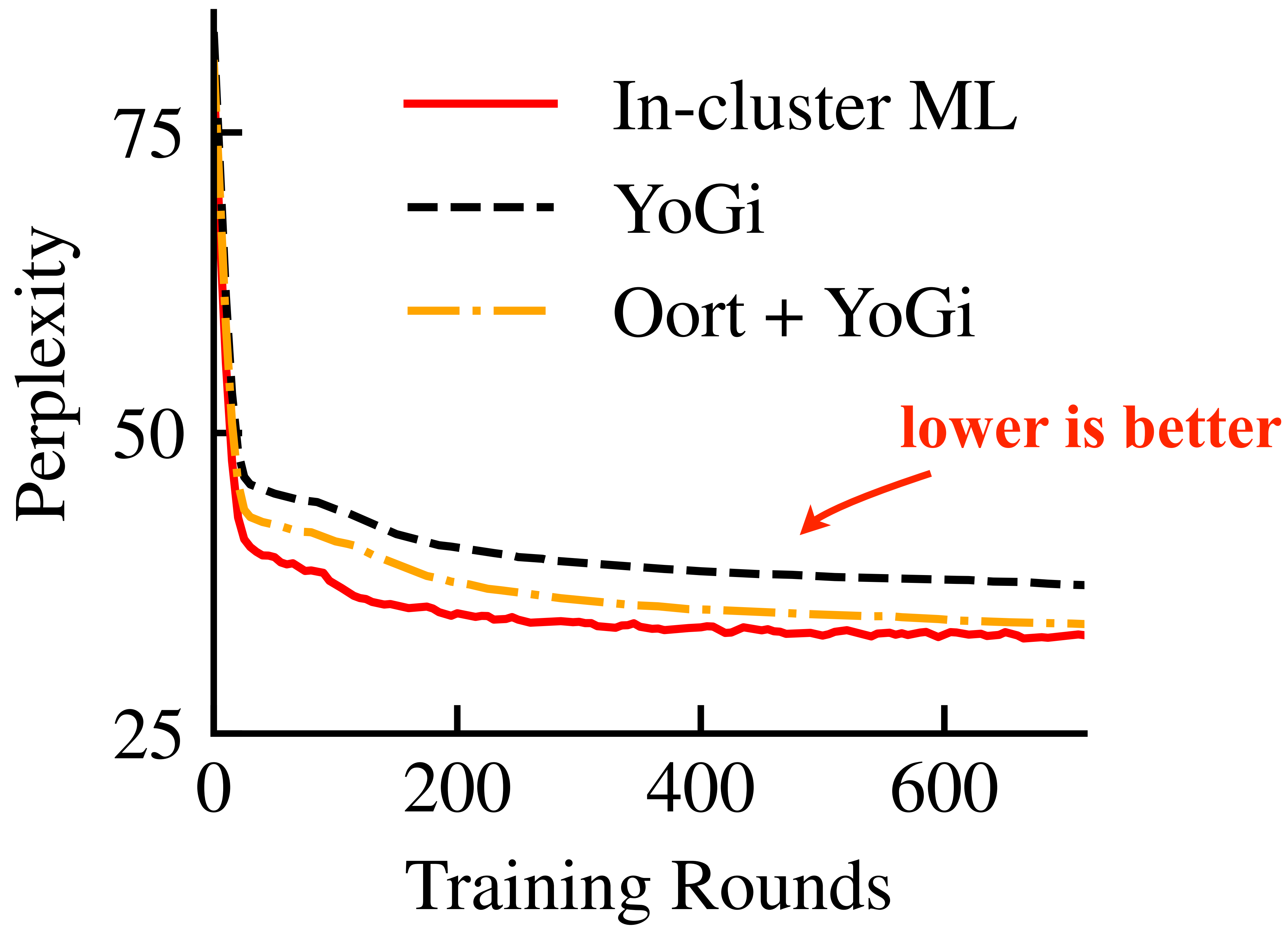
Oort improves TTA by 14X and final accuracy by 9%

Zoom into Statistical Performance

Image classification (ShuffleNet Model)



Next-word prediction (Albert Model)



Oort achieves close to upper-bound statistical performance

Oort

https://github.com/SymbioticLab/Oort

Participant selection framework for
- *Faster* convergence in *FL training*
- *Interpretable* data *selection* in *FL testing*

Client selection for $\begin{cases} \text{\textit{utility-aware FL training} w/ adaptive \textit{exploration-exploitation}} \\ \text{\textit{criteria-aware FL testing to enforce specified data selection}} \end{cases}$

Thank you!

UNIVERSITY OF MICHIGAN