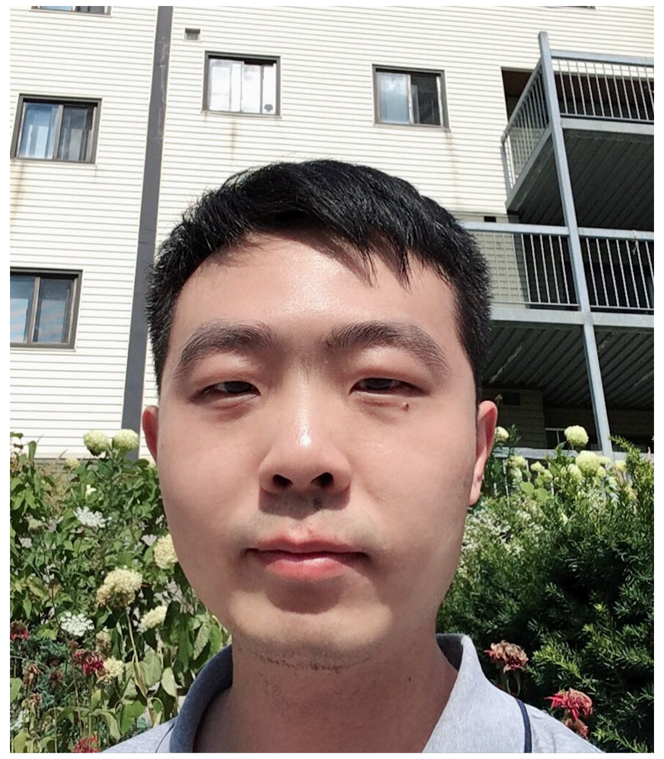
Proportional Fairness in Federated Learning

Guojun Zhang
Noah's Ark Lab
Jan 2023


Guojun Zhang

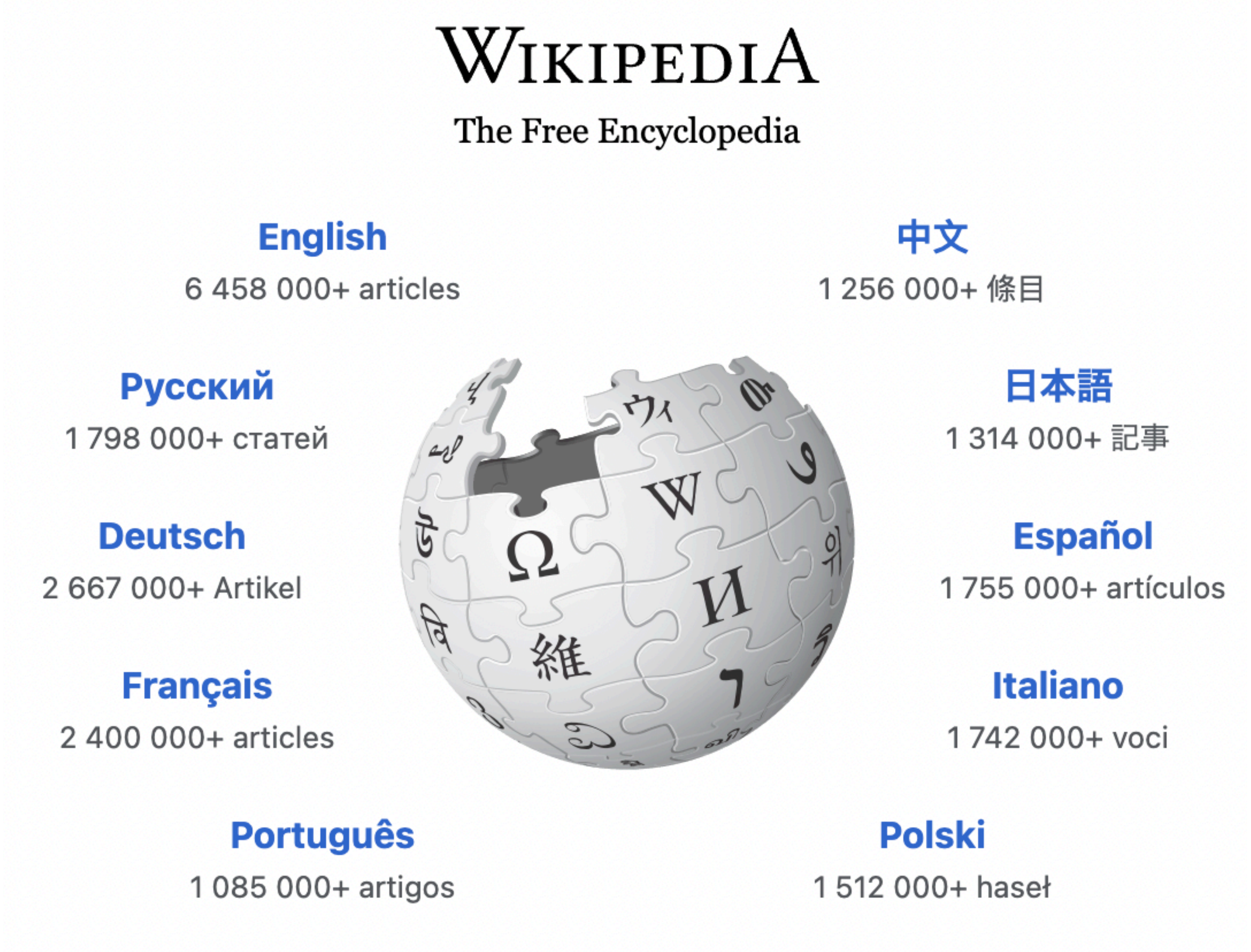

Saber Malekmohammadi


Xi Chen


Yaoliang Yu

Motivations

- Era of a huge amount of data



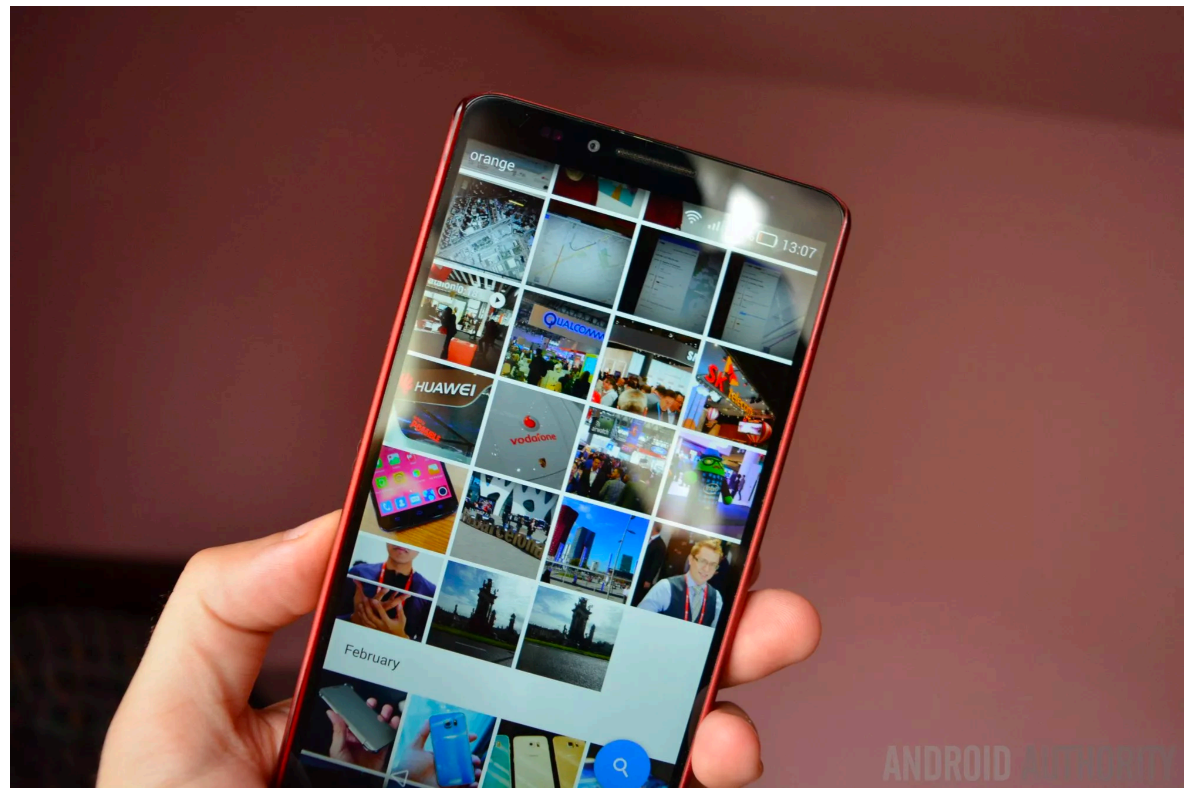
- Era of big models (DALLE-2/GPT-3)

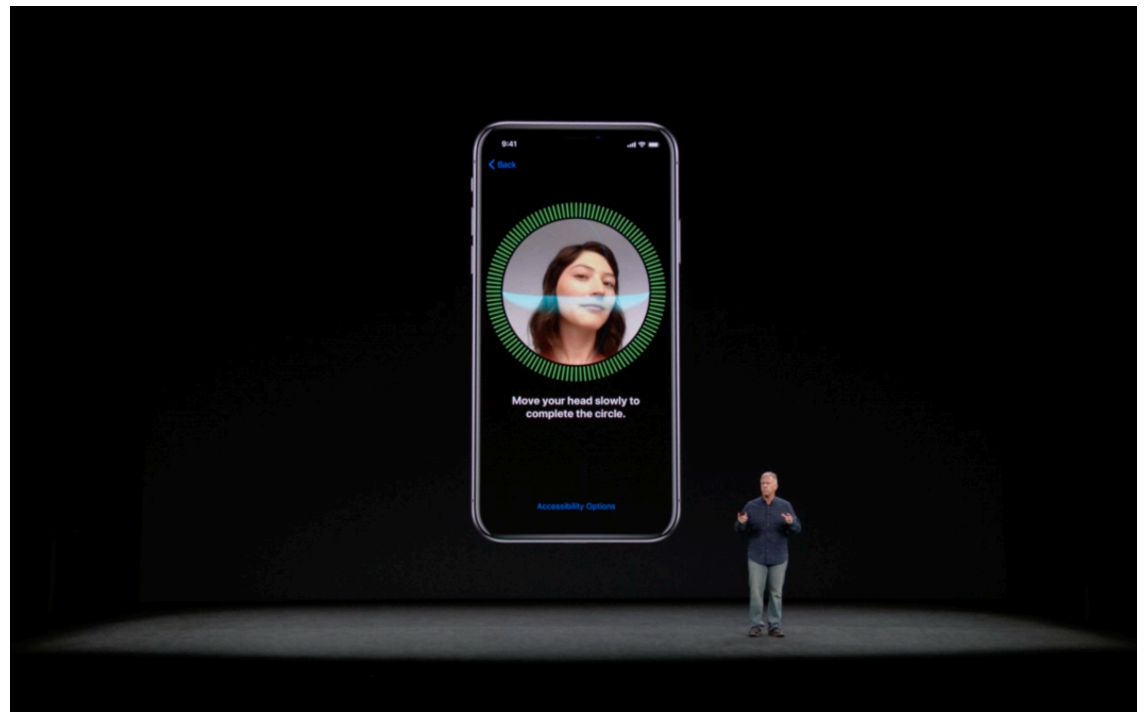
https://image-net.org/; https://www.wikipedia.org/; https://openai.com/dall-e-2/; https://openai.com/api/; https://git.openi.org.cn/PCL-Platform.Intelligence/PanGu-Alpha

Motivations

- Privacy concerns: the huge amount of user data may not be accessible

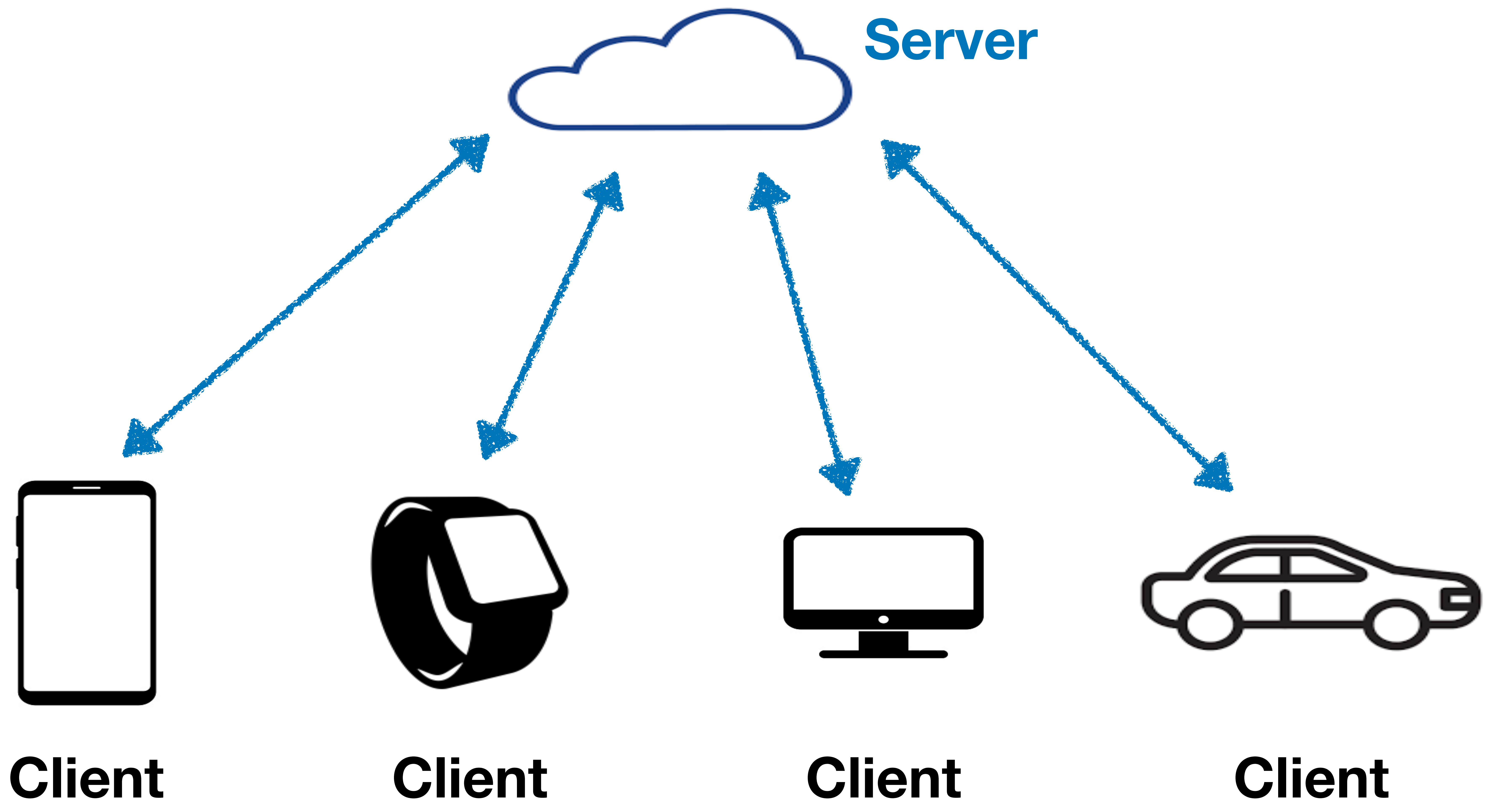
- Sensitive user data should be saved locally





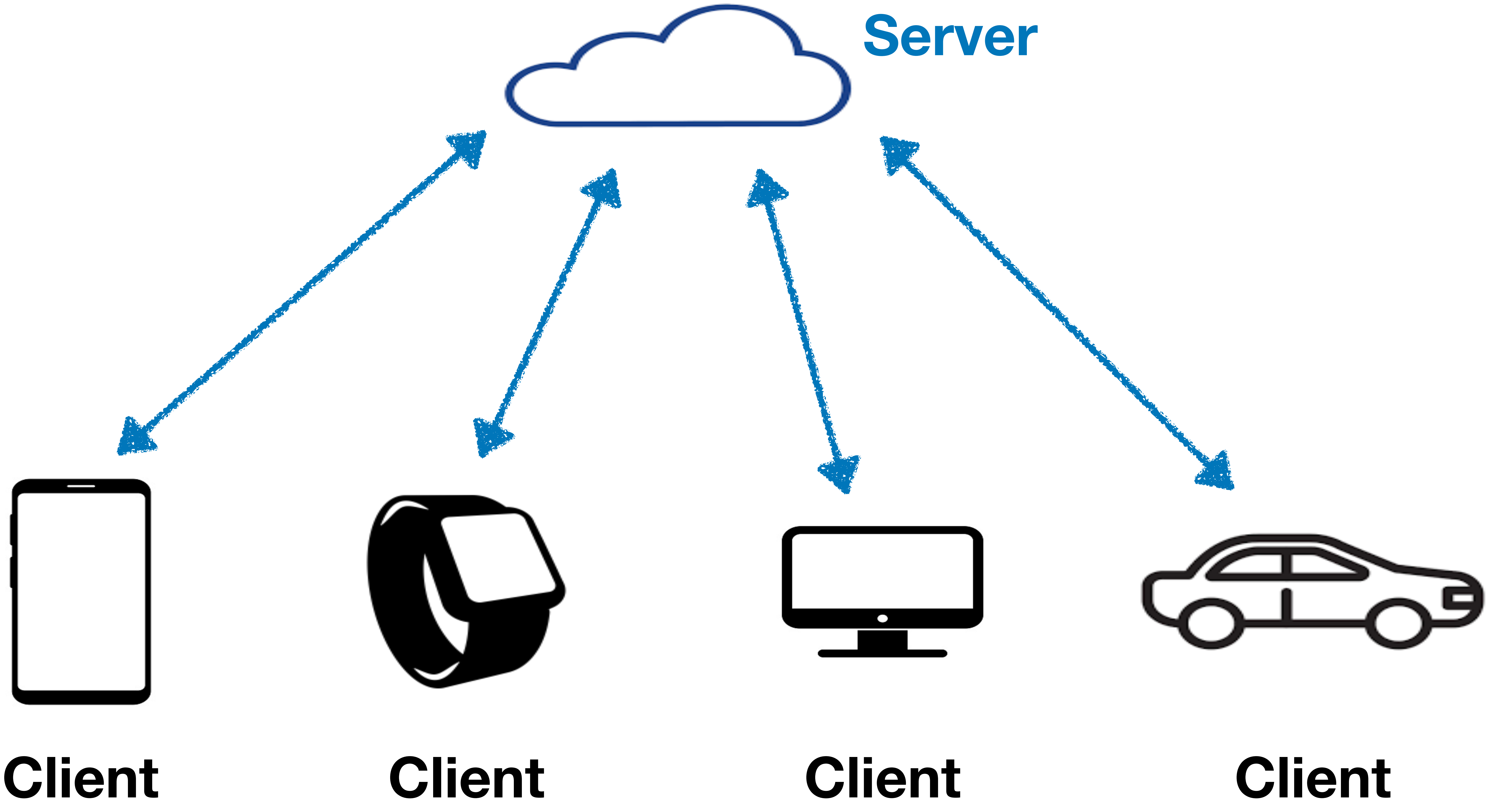
https://www.androidauthority.com/best-gallery-apps-android-687764/; https://www.businessinsider.com/how-to-block-text-messages-on-iphone

Federated Learning



McMahan et al, AISTATS 2017.

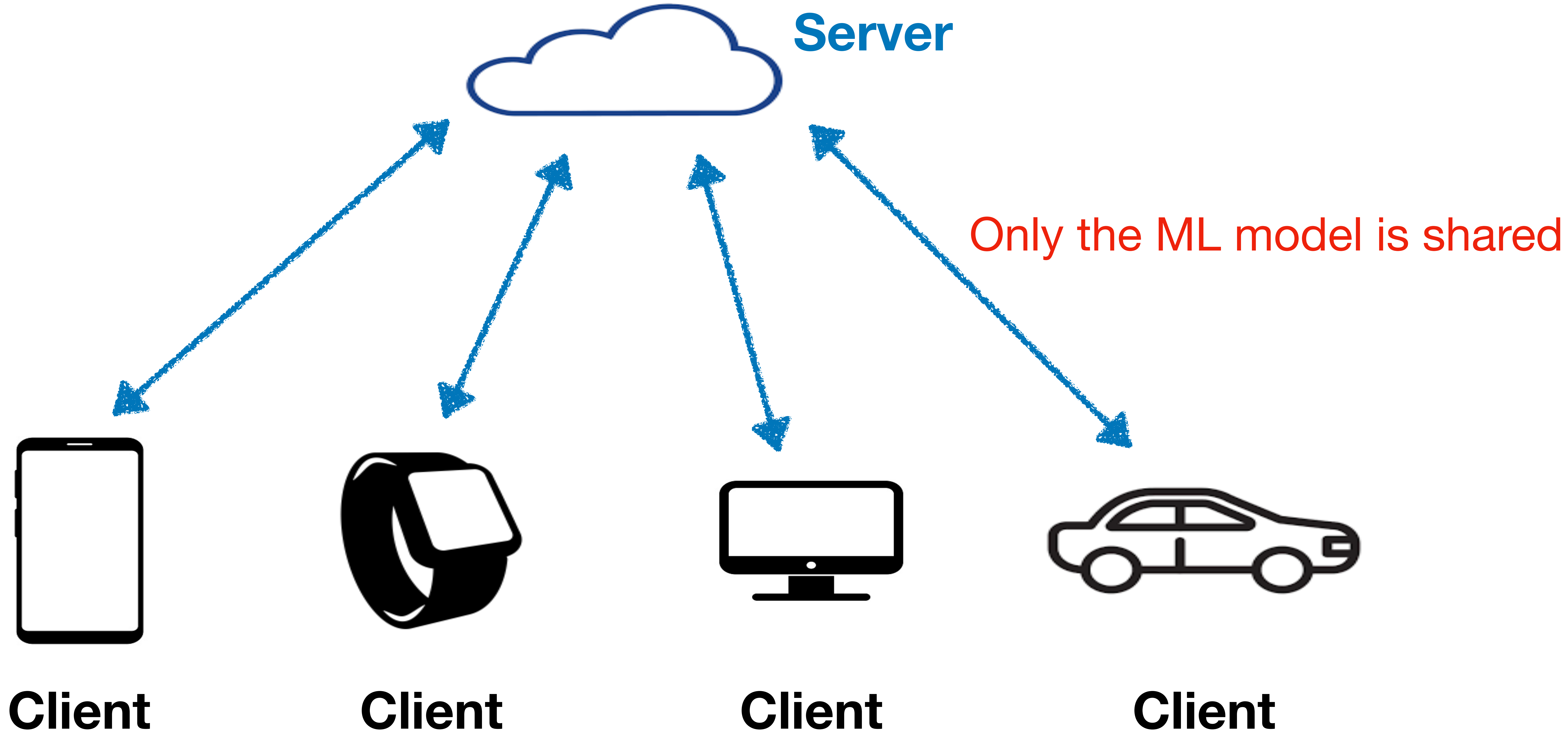
Federated Learning



Client data is kept locally

McMahan et al, AISTATS 2017.

Federated Learning

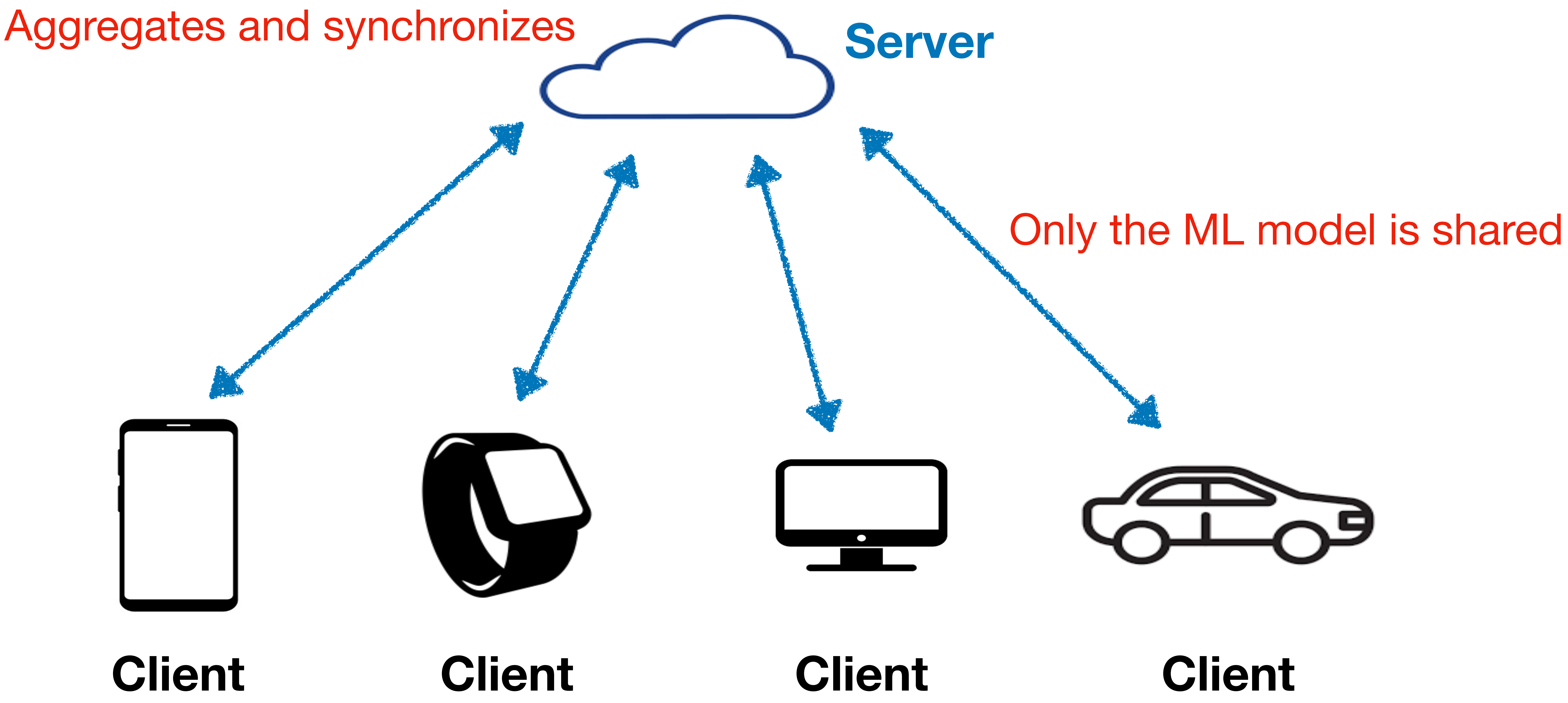


Client data is kept locally

McMahan et al, AISTATS 2017.

Federated Learning

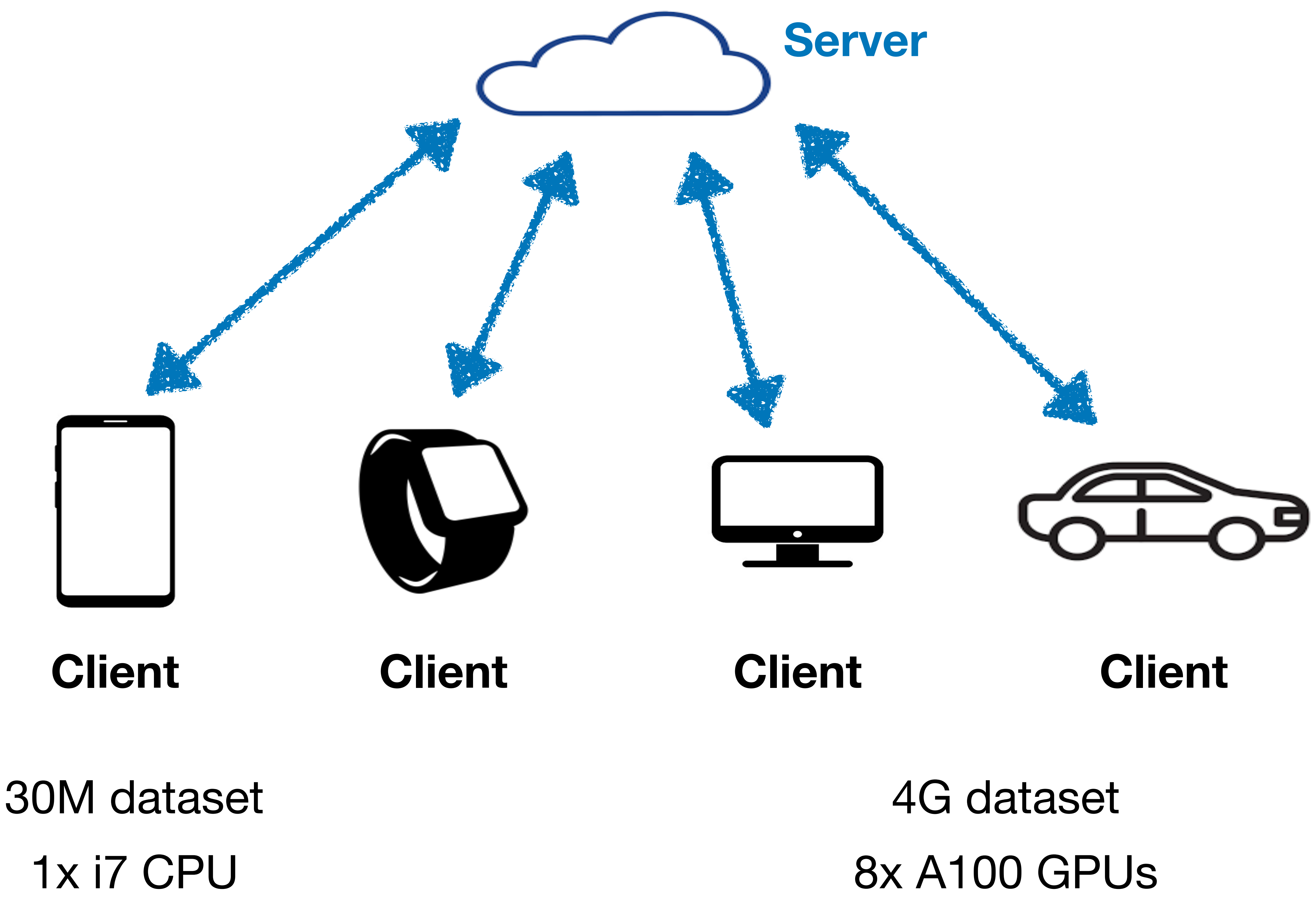
HUAWEI



Client data is kept locally

McMahan et al, AISTATS 2017.

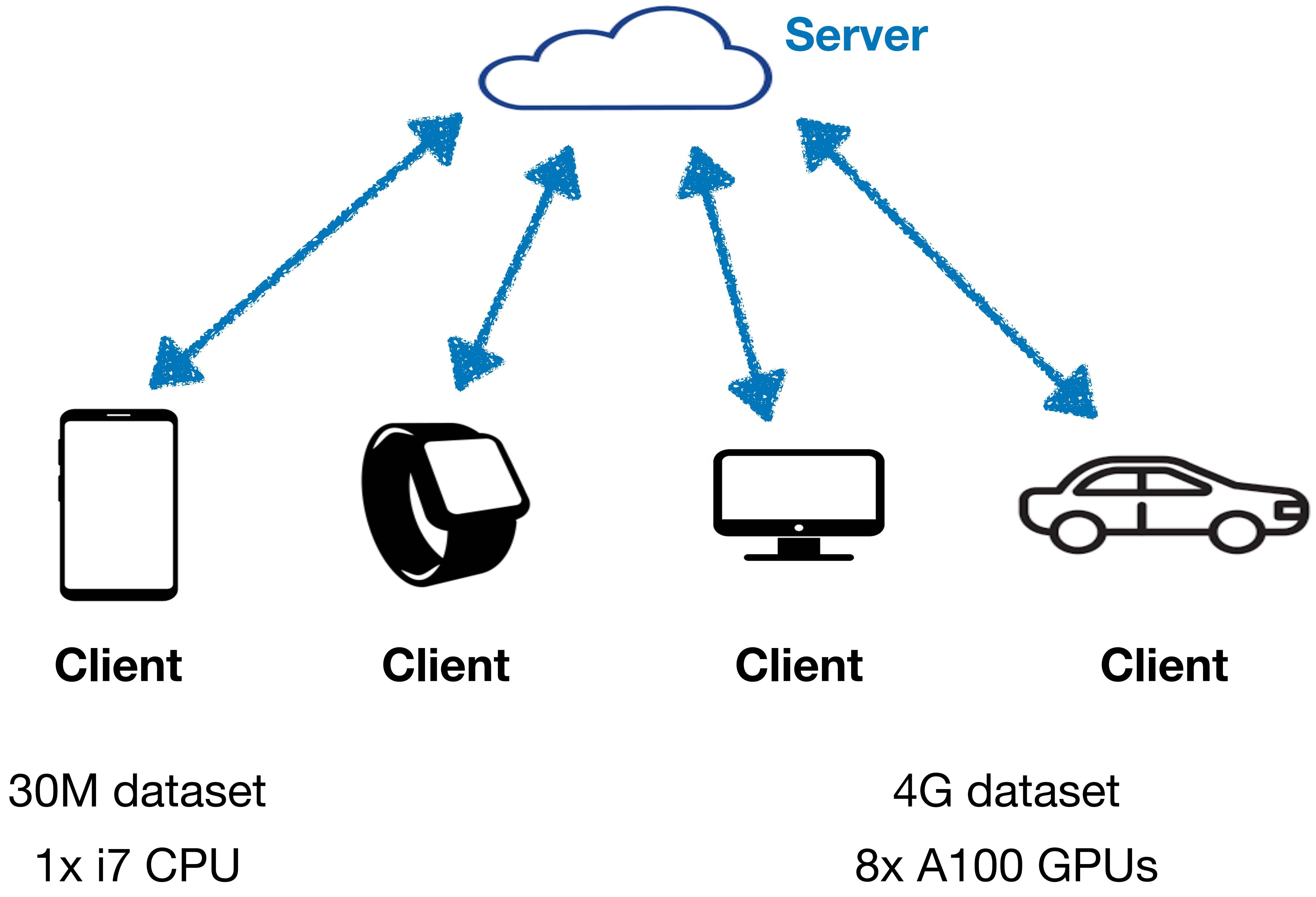
Motivations for Fair FL



Advances and Open Problems in Federated Learning. Kairouz et al, Foundations and Trends® in Machine Learning, 2021.

Motivations for Fair FL

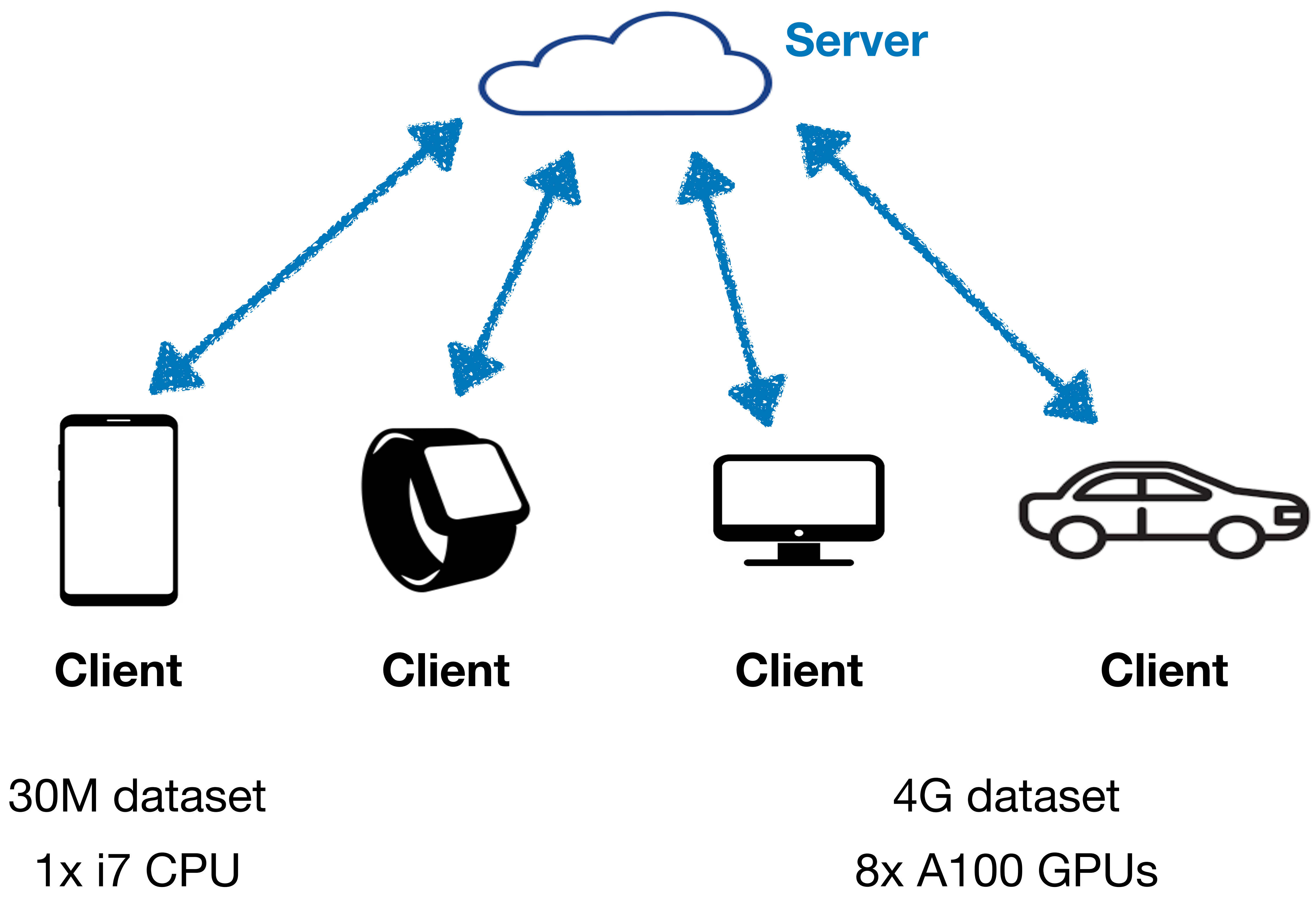
- Clients are so different: data, computing power, storage, bandwidth…



Advances and Open Problems in Federated Learning. Kairouz et al, Foundations and Trends® in Machine Learning, 2021.

Motivations for Fair FL

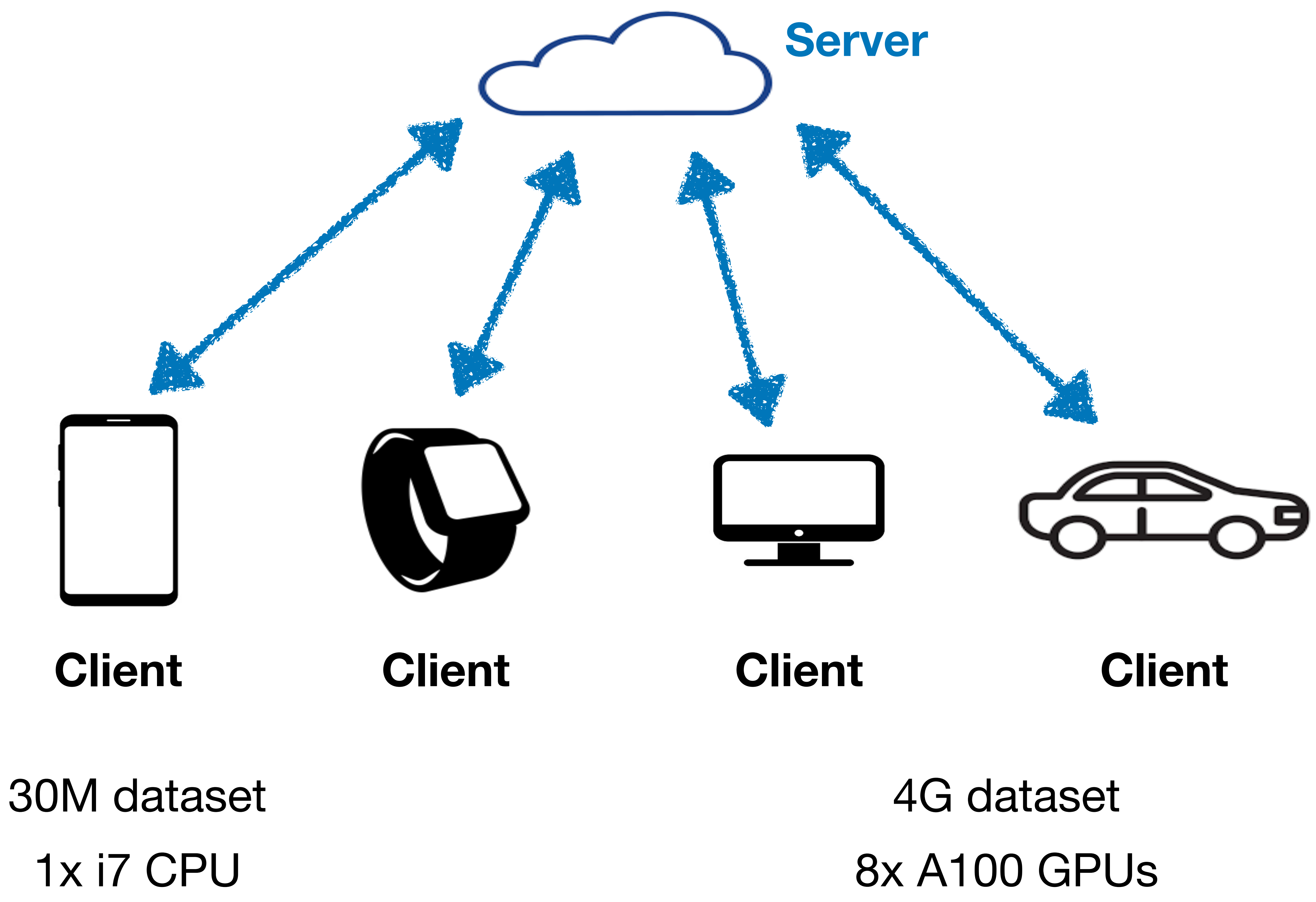
- Clients are so different: data, computing power, storage, bandwidth…
- How to define and implement fairness among various clients?



Advances and Open Problems in Federated Learning. Kairouz et al, Foundations and Trends® in Machine Learning, 2021.

Motivations for Fair FL

- Clients are so different: data, computing power, storage, bandwidth…
- How to define and implement fairness among various clients?
- This has potential social impacts…



Advances and Open Problems in Federated Learning. Kairouz et al, Foundations and Trends® in Machine Learning, 2021.

Fairness from a Social Perspective

HUAWEI

- Given the utility $u_i(\theta)$ of client i, which depends on a public policy θ, how do we define a fair policy?
- Utilitarianism (Bentham 1780): the total utility should be maximized

$$\max_\theta \sum_i p_i u_i(\theta)$$

- Egalitarianism (Rawls 1974): the poorest client should be reasonably happy

$$\max_\theta \min_i u_i(\theta)$$

McMahan et al, AISTATS 2017; Mohri et al ICML 2019.

Fairness from a Social Perspective

HUAWEI

- Given the utility $u_i(\theta)$ of client i, which depends on a public policy θ, how do we define a fair policy?
- Utilitarianism (Bentham 1780): the total utility should be maximized

$$\max_{\theta} \sum_i p_i u_i(\theta)$$

FedAvg

$$\min_{\theta} \sum_i p_i f_i(\theta)$$

- Egalitarianism (Rawls 1974): the poorest client should be reasonably happy

$$\max_{\theta} \min_{i} u_i(\theta)$$

McMahan et al, AISTATS 2017; Mohri et al ICML 2019.

Fairness from a Social Perspective

HUAWEI

- Given the utility $u_i(\theta)$ of client i, which depends on a public policy θ, how do we define a fair policy?

- Utilitarianism (Bentham 1780): the total utility should be maximized

$$\max_{\theta} \sum_i p_i u_i(\theta)$$

FedAvg

$$\min_{\theta} \sum_i p_i f_i(\theta)$$

- Egalitarianism (Rawls 1974): the poorest client should be reasonably happy

$$\max_{\theta} \min_{i} u_i(\theta)$$

AFL

$$\min_{\theta} \max_{i} f_i(\theta)$$

McMahan et al, AISTATS 2017; Mohri et al ICML 2019.

Balance between Utilitarianism and Egalitarianism

- Utilitarianism and Egalitarianism might conflict with each other
- One could improve the worst-case clients, but better-off clients would be degraded
- (Consider sharing the wealth equally to everyone…)
- Can we find some balance between these two?

Existing Approaches to the Balance

HUAWEI

- Tilted Empirical Risk Minimization:

$$\min_{\theta} \log\left(\sum_i p_i \exp(\alpha f_i(\theta))\right)$$

$\alpha \geq 0$

- $\alpha = 0$, FedAvg; $\alpha \to \infty$, AFL
- q-FFL:

$$\min_{\theta} \frac{1}{q+1} \sum_i p_i f_i^{q+1}(\theta)$$

$q \geq 0$

- $q = 0$, FedAvg; $q \to \infty$, AFL

Li et al 2020a, b

α-Fairness

- q-FFL:

$$\min_{\theta} \frac{1}{q+1} \sum_i p_i f_i^{q+1}(\theta)$$

$q \geq 0$

- Comes from α-Fairness (Mo & Walrand 2000)

$$\max_{\theta} \sum_i p_i \phi_\alpha(u_i)$$

$$\phi_\alpha(u_i) = \begin{cases} u_i^{1-\alpha}/(1-\alpha) \text{ if } \alpha > 0, \alpha \neq 1; \\ \log u_i \text{ if } \alpha = 1. \end{cases}$$

- q-FFL takes 1) $\alpha = -q$; 2) $u_i \to f_i$

α-Fairness

- q-FFL:

$$\min_{\theta} \frac{1}{q+1} \sum_i p_i f_i^{q+1}(\theta)$$

$q \geq 0$

- Comes from α-Fairness (Mo & Walrand 2000)

$$\max_{\theta} \sum_i p_i \phi_\alpha(u_i) \qquad \phi_\alpha(u_i) = \begin{cases} u_i^{1-\alpha}/(1-\alpha) \textbf{ if } \alpha > 0, \alpha \neq 1; \\ \log u_i \textbf{ if } \alpha = 1. \end{cases}$$

Proportional Fairness

- q-FFL takes 1) $\alpha = -q$; 2) $u_i \to f_i$

Proportional Fairness: relative changes

HUAWEI

- The perspective of relative changes is quotidian
- In finance, people care more about the interest rates
- In telecommunication, people care more the data transmission speed compared to the bandwidth

$$\max_u \sum_i p_i \log u_i$$

$\longleftrightarrow$

$$\textbf{find } u_i^*, \forall u \quad \sum_i p_i \frac{u_i - u_i^*}{u_i^*} \leq 0$$

Nash bargaining solution (NBS)

Proportional fairness

- NBS achieves some balance between utilitarian and egalitarian

$$\sum_i p_i \log u_i \leq \log \sum_i p_i u_i$$

q-FFL misses Proportional Fairness

HUAWEI

- q-FFL:

$$\min_{\theta} \frac{1}{q+1} \sum_i p_i f_i^{q+1}(\theta) \qquad q \geq 0$$

- q-FFL takes 1) $\alpha = -q$; 2) $u_i \to f_i$
- Proportional fairness corresponds to $\alpha = 1, q = -1$
- Therefore, the objective becomes

$$\min_{\theta} \sum_i p_i \log f_i$$

- This encourages the clients to be more disparate!
- $(f_1, f_2) = (1/3, 2/3)$ vs. $(f_1, f_2) = (1/2, 1/2)$

PropFair Algorithm

HUAWEI

- Treat the utility as the loss:

$$\max_{\theta} \sum_i p_i \log f_i(\theta)$$

doesn't work

- Treat the utility as the negated loss, with some constant:

$$\max_{\theta} \sum_i p_i \log(M - f_i(\theta))$$

- Huberization for robustness:

$$\log_{[\epsilon]}(M - f_i(\theta)) = \begin{cases} \log(M - t) \textbf{ if } t \leq M - \epsilon; \\ \log \epsilon - \frac{1}{\epsilon}(t - M + \epsilon) \textbf{ if } t > M - \epsilon\,. \end{cases}$$

Linear extrapolation

Huber 1964

PropFair Algorithm

HUAWEI

- Treat the utility as the negated loss, with some constant:

$$\max_{\theta} \sum_i p_i \log_{[\epsilon]}(M - f_i(\theta))$$

- Huberization for robustness:

$$\log_{[\epsilon]}(M - f_i(\theta)) = \begin{cases} \log(M - t) \textbf{ if } t \le M - \epsilon; \\ \log \epsilon - \frac{1}{\epsilon}(t - M + \epsilon) \textbf{ if } t > M - \epsilon\,. \end{cases}$$

- Drop-in replacement of FedAvg; can be easily adapted to existing FL frameworks (e.g. Flower, TensorFlow Federated)

Fair FL Algorithms

HUAWEI

utilitarian	$\max_\theta \sum_i p_i u_i(\theta)$	$\min_\theta \sum_i p_i f_i(\theta)$	**FedAvg**
egalitarian	$\max_\theta \min_i u_i(\theta)$	$\min_\theta \max_i f_i(\theta)$	**Agnostic FL (AFL)**
α-fairness	$\max_\theta \frac{1}{1-\alpha} \sum_i p_i u_i(\theta)^{1-\alpha}$	$\min_\theta \frac{1}{q+1} \sum_i p_i f_i^{q+1}(\theta)$	**q-FFL**
proportional	$\max_\theta \sum_i p_i \log u_i(\theta)$	$\min_\theta - \sum_i p_i \log(M - f_i(\theta))$	**PropFair**

Maskin 1978; Rawls 1974; Lan et al 2010; Kelly et al 1998
McMahan et al AISTATS 2017; Mohri et al ICML 2019; Li et al ICLR 2020

Generalized Means

HUAWEI

$$\mathrm{A}_\varphi(f) = \varphi^{-1}\left(\sum_i p_i \varphi(f_i)\right)$$

Kolmogorov's generalized mean

$$\min_\theta \sum_i \mathrm{A}_\varphi(f(\theta))$$

$$\min_\theta \sum_i p_i f_i(\theta)$$ **FedAvg**

$$\min_\theta \max_i f_i(\theta)$$ **AFL**

$$\min_\theta \frac{1}{q+1} \sum_i p_i f_i^{q+1}(\theta)$$ **q-FFL**

$$\min_\theta - \sum_i p_i \log(M - f_i(\theta))$$ **PropFair**

Kolmogorov, On the notion of mean, *Anti della Academia Nazionale dei Lincei,* 1930.

Generalized Means

HUAWEI

$$\mathrm{A}_\varphi(f) = \varphi^{-1}\left(\sum_i p_i \varphi(f_i)\right)$$

Kolmogorov's generalized mean

$$\min_\theta \sum_i \mathrm{A}_\varphi(f(\theta))$$

$$\varphi(t) = t$$

$$\min_\theta \sum_i p_i f_i(\theta)$$

FedAvg

$$\min_\theta \max_i f_i(\theta)$$

AFL

$$\min_\theta \frac{1}{q+1} \sum_i p_i f_i^{q+1}(\theta)$$

q-FFL

$$\min_\theta -\sum_i p_i \log(M - f_i(\theta))$$

PropFair

Kolmogorov, On the notion of mean, *Anti della Academia Nazionale dei Lincei,* 1930.

Generalized Means

HUAWEI

$$\mathrm{A}_\varphi(f) = \varphi^{-1}\left(\sum_i p_i \varphi(f_i)\right)$$

Kolmogorov's generalized mean

$$\min_\theta \sum_i \mathrm{A}_\varphi(f(\theta))$$

$\varphi(t) = t$

$$\min_\theta \sum_i p_i f_i(\theta)$$ **FedAvg**

$$\min_\theta \max_i f_i(\theta)$$ **AFL**

$\varphi(t) = t^{q+1}$

$$\min_\theta \frac{1}{q+1} \sum_i p_i f_i^{q+1}(\theta)$$ **q-FFL**

$$\min_\theta - \sum_i p_i \log(M - f_i(\theta))$$ **PropFair**

Kolmogorov, On the notion of mean, *Anti della Academia Nazionale dei Lincei,* 1930.

Generalized Means

$$A_\varphi(f) = \varphi^{-1}\left(\sum_i p_i \varphi(f_i)\right)$$

Kolmogorov's generalized mean

$$\min_\theta \sum_i A_\varphi(f(\theta))$$

$\varphi(t) = t$ | $\min_\theta \sum_i p_i f_i(\theta)$ | **FedAvg**

$\varphi(t) = t^\infty$ | $\min_\theta \max_i f_i(\theta)$ | **AFL**

$\varphi(t) = t^{q+1}$ | $\min_\theta \frac{1}{q+1} \sum_i p_i f_i^{q+1}(\theta)$ | **q-FFL**

$\min_\theta - \sum_i p_i \log(M - f_i(\theta))$ | **PropFair**

Kolmogorov, On the notion of mean, *Anti della Academia Nazionale dei Lincei,* 1930.

Generalized Means

$$\mathrm{A}_{\varphi}(f) = \varphi^{-1}\left(\sum_i p_i \varphi(f_i)\right)$$

Kolmogorov's generalized mean

$$\min_{\theta} \sum_i \mathrm{A}_{\varphi}(f(\theta))$$

$\varphi(t) = t$	$\min_{\theta} \sum_i p_i f_i(\theta)$	**FedAvg**
$\varphi(t) = t^{\infty}$	$\min_{\theta} \max_{i} f_i(\theta)$	**AFL**
$\varphi(t) = t^{q+1}$	$\min_{\theta} \frac{1}{q+1} \sum_i p_i f_i^{q+1}(\theta)$	**q-FFL**
$\varphi(t) = -\log(M - t)$	$\min_{\theta} - \sum_i p_i \log(M - f_i(\theta))$	**PropFair**

Kolmogorov, On the notion of mean, *Anti della Academia Nazionale dei Lincei,* 1930.

Dual Perspective

- Many fair FL algorithms can be reduced to minimizing the generalized mean:

$$\min_{\theta} \sum_i \mathrm{A}_{\varphi}(f(\theta))$$

- With convex conjugate, one can derive the dual form:

$$\min_{\theta} \max_{\lambda \geq \mathbf{0}} \sum_i \lambda_i f_i(\theta) - \mathrm{A}^*_{\varphi}(\lambda)$$

convex conjugate

- Fair FL algorithms can be considered as minimizing a linear combination of client losses

Dual Perspective

- Each convex conjugate leads to a different requirement for λ

$$\min_{\theta} \max_{\lambda \geq \mathbf{0}} \sum_i \lambda_i f_i(\theta) - \mathrm{A}_{\varphi}^{*}(\lambda)$$

FedAvg

$$\lambda_i = p_i$$

AFL

$$\lambda_i \geq 0, \sum_i \lambda_i = 1$$

q-FFL

$$\lambda_i \propto p_i f_i^q$$

PropFair

$$\lambda_i \propto \frac{p_i}{M - f_i}$$

Higher loss with higher weight

Optimization of PropFair

- PropFair reduces to FedAvg when the baseline M is large

$$\min_{\theta} - \sum_i p_i \log(M - f_i(\theta)) \approx \sum_i p_i \frac{f_i(\theta)}{M} - n \log M$$

$$\nabla \left(-\log(M - f_i(\theta)) \right) = \frac{\nabla f_i(\theta)}{M - f_i}$$

Optimization of PropFair

- PropFair reduces to FedAvg when the baseline M is large

$$\min_{\theta} - \sum_i p_i \log(M - f_i(\theta)) \approx \sum_i p_i \frac{f_i(\theta)}{M} - n \log M$$

- Adaptive learning rate for heterogeneous clients:

$$\nabla \left(-\log(M - f_i(\theta))\right) = \frac{\nabla f_i(\theta)}{M - f_i}$$

Optimization of PropFair

- PropFair reduces to FedAvg when the baseline M is large

$$\min_{\theta} - \sum_i p_i \log(M - f_i(\theta)) \approx \sum_i p_i \frac{f_i(\theta)}{M} - n \log M$$

- Adaptive learning rate for heterogeneous clients:

$$\nabla \left(-\log(M - f_i(\theta))\right) = \frac{\nabla f_i(\theta)}{M - f_i}$$

- With Lipschitz smoothness and bounded variance assumptions, we can prove the convergence of PropFair to Nash bargaining solutions

Optimization of FedAvg

HUAWEI

Theorem 4.2 (**FedAvg**). *Given Assumption 4.1, assume that the local learning rate satisfies* $\eta K_i \leq \frac{1}{6L}$ *for any* $i \in [n]$ *and*

$$\eta \leq \frac{1}{L}\sqrt{\frac{1}{24(e-2)(\sum_i p_i^2)(\sum_i K_i^4)}}. \tag{4.2}$$

Running Algorithm 2 for T *global epochs we have:*

$$\min_{0 \leq t \leq T-1} \mathbf{E}\|\nabla F(\boldsymbol{\theta}_t)\|^2 \leq \frac{12}{(11\mu - 9)\eta}\left(\frac{F_0 - F^*}{T} + \Psi_\sigma\right),$$

with $\mu = \sum_i p_i K_i$ *for full participation and* $\mu = \min_i K_i$ *for partial participation,* $F_0 = F(\boldsymbol{\theta}_0)$*,* $F^* = \min_{\boldsymbol{\theta}} F(\boldsymbol{\theta})$ *the optimal value, and*

$$\Psi_\sigma = \eta\|\boldsymbol{p}\|^2\left[\sum_{i=1}^{n} K_i^2\left(\frac{L\eta\sigma_i^2}{2m} + \sigma^2\right) + (e-2)\eta^2 L^2 \sum_{i=1}^{n} K_i^3\left(\frac{\sigma_i^2}{m} + 6K_i\sigma^2\right)\right], \boldsymbol{p} = (p_1, \ldots, p_n).$$

Retrieves the convergence rate of SGD when the global variance $\sigma = 0$

Optimization of PropFair

Theorem 4.4 (PropFair). *Denote* $\widetilde{L} = \frac{4}{M^2}(\frac{3}{2}ML + L_0^2)$ *and* $p_i = \frac{n_i}{N}$. *Given Assumptions 4.1 and 4.3, assume that the local learning rate satisfies:*

$$\eta \leq \min\left\{\min_{i\in[n]} \frac{1}{6\widetilde{L}K_i}, \frac{1}{8\widetilde{L}}\sqrt{\frac{1}{(e-2)(\sum_i p_i^2)(\sum_i K_i^4)}}\right\}. \quad (4.3)$$

By running Algorithm 1 for T *global epochs we have:*

$$\min_{0\leq t\leq T-1} \mathbb{E}\|\nabla\pi(\boldsymbol{\theta}_t)\|^2 \leq \frac{12}{(11\mu-9)\eta}\left(\frac{\pi_0-\pi^*}{T} + \widetilde{\Psi}_\sigma\right),$$

with $\mu = \sum_i p_i K_i$ *for full participation and* $\mu = \min_i K_i$ *for partial participation,* $\pi_0 = \pi(\boldsymbol{\theta}_0)$, $\pi^* = \min_{\boldsymbol{\theta}} \pi(\boldsymbol{\theta})$ *the optimal value, and*

$$\widetilde{\Psi}_\sigma = \eta\|\boldsymbol{p}\|^2\left[\sum_{i=1}^n K_i^2\left(\frac{\widetilde{\sigma}_i^2}{m} + 2\widetilde{\sigma}^2\right) + 16(e-2)\eta^2\widetilde{L}^2\sum_{i=1}^n K_i^4\left(\frac{\widetilde{\sigma}_i^2}{m} + \widetilde{\sigma}^2\right)\right]$$

where $\widetilde{\sigma}_i^2 = \frac{8}{M^4}(9M^2\sigma_i^2 + 4L_0^2\sigma_{0,i}^2)$ *and* $\widetilde{\sigma} = \frac{4}{M}\left(\frac{3}{2}\sigma + \frac{L_0}{M}\sigma_0\right)$.

Experiments

- Experimental Setup:

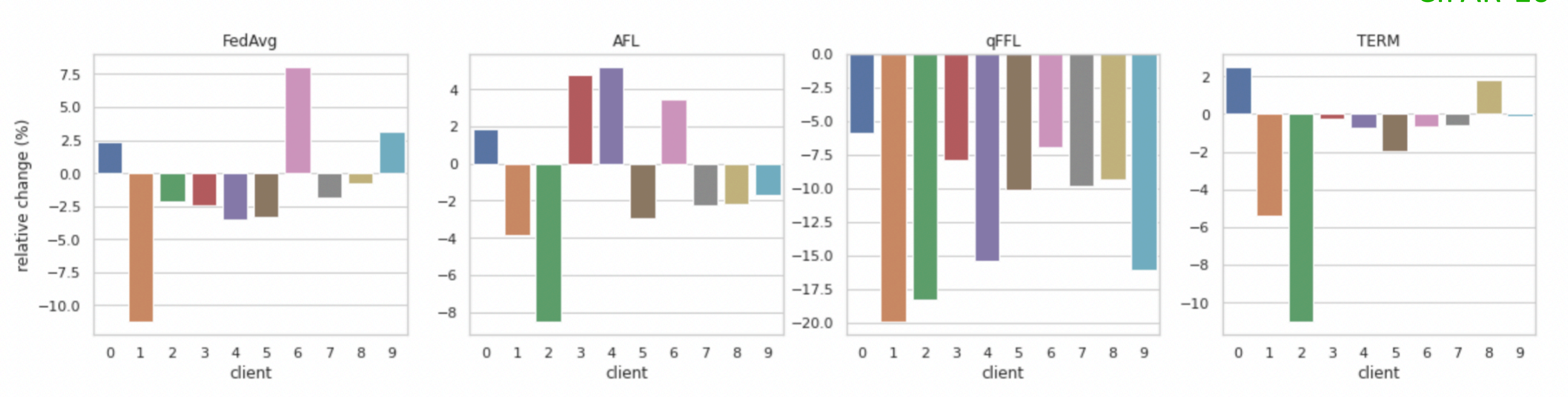
Dataset	Task	Size	Partition	Client number	Model
CIFAR-10	Vision	50,000	Dirichlet allocation	10	ResNet18
CIFAR-100	Vision	50,000	Dirichlet allocation	10	ResNet18
TinyImageNet	Vision	100,000	Dirichlet allocation	20	ResNet18
Shakespeare	Language	356,027	Character	20	LSTM

Experiments: Proportional Fairness

$$\textbf{find } u_i^*, \ \forall u \ \sum_i p_i \frac{u_i - u_i^*}{u_i^*} \leq 0$$

- PropFair can achieve proportional fairness approximately:

CIFAR-10



-2.21% -1.32% -12.95% -1.79%

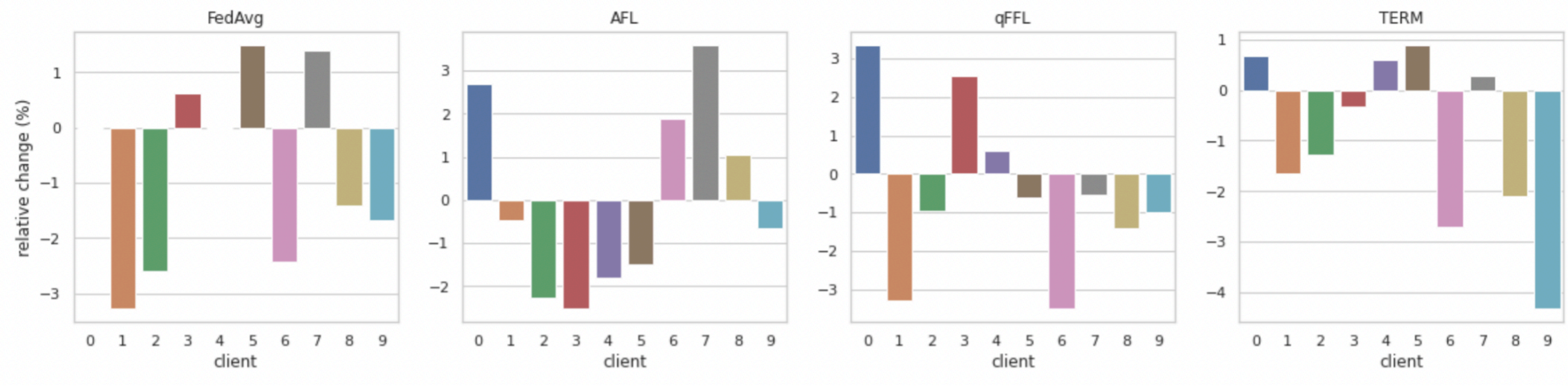
- Vanilla comparison of other baselines vs. PropFair

Experiments: Proportional Fairness

$$\textbf{find } u_i^*, \; \forall u \; \sum_i p_i \frac{u_i - u_i^*}{u_i^*} \leq 0$$

- PropFair can achieve proportional fairness approximately:

CIFAR-100



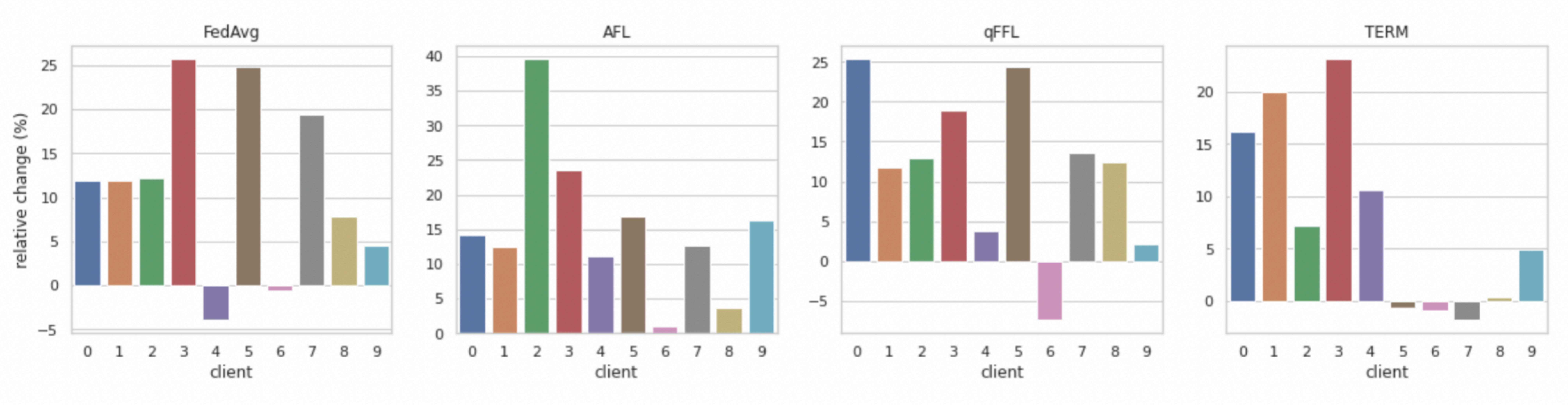
-0.86% +0.05% -0.67% -1.01%

- Baseline algorithms pertained from PropFair

Experiments: Proportional Fairness

- Other baseline algorithms cannot achieve the same level of proportional fairness

CIFAR-100

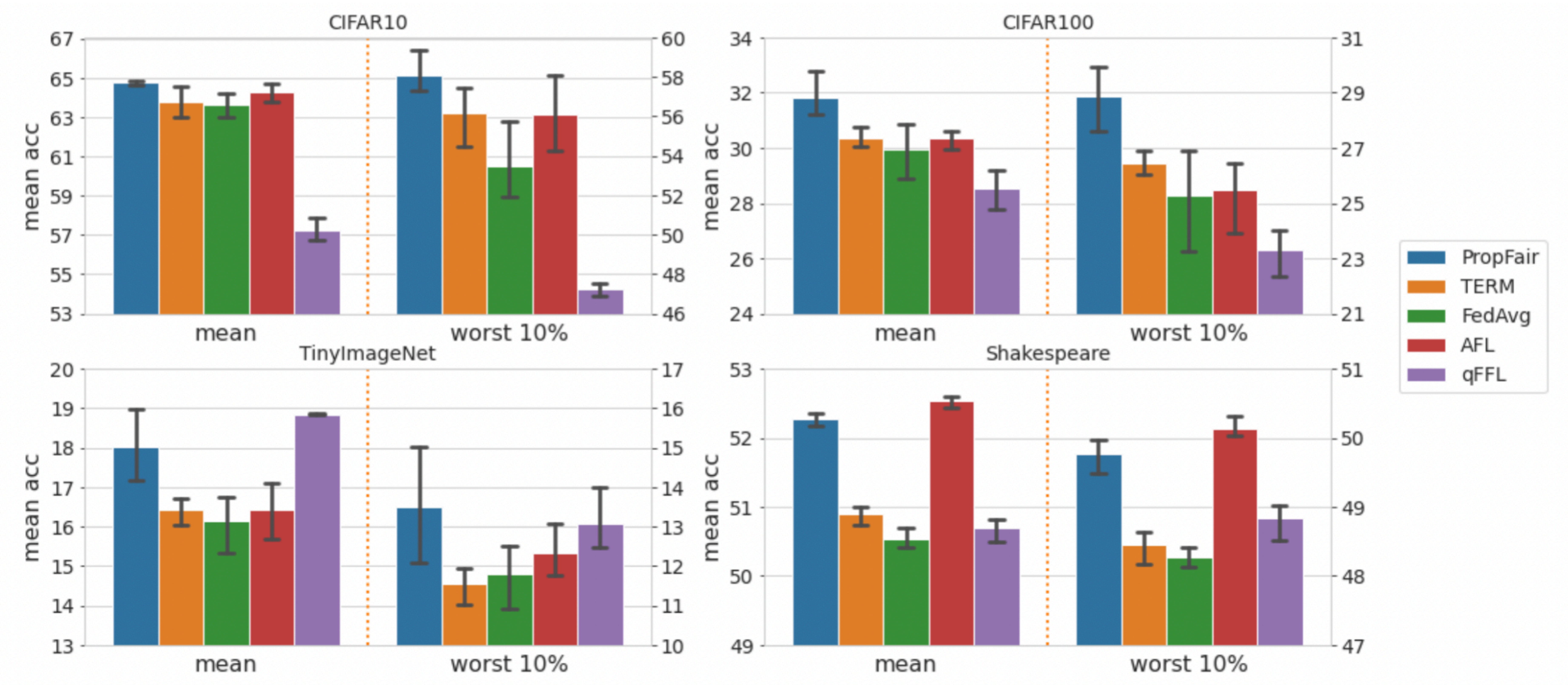


+10.88% +13.93% +11.58% +8.67%

- PropFair can achieve improved relative performance over pertained baselines

Experiments: Other Metrics

- Even on other metrics, PropFair show competitive performance:



- Left: Average performance; Right: Worst 10% performance

Conclusions

Conclusions

- Goal: Find proportionally fair solutions in FL

Conclusions

- Goal: Find proportionally fair solutions in FL
- Propose the PropFair algorithm based on Nash bargaining solution, the right choice of utility function and huberization

Conclusions

- Goal: Find proportionally fair solutions in FL
- Propose the PropFair algorithm based on Nash bargaining solution, the right choice of utility function and huberization
- Simple modification of FedAvg, but effective

Conclusions

- Goal: Find proportionally fair solutions in FL
- Propose the PropFair algorithm based on Nash bargaining solution, the right choice of utility function and huberization
- Simple modification of FedAvg, but effective
- Future directions: explore the generalized means and study the tradeoff among FL clients

Thank you!