Federated Learning: Challenges and Solutions

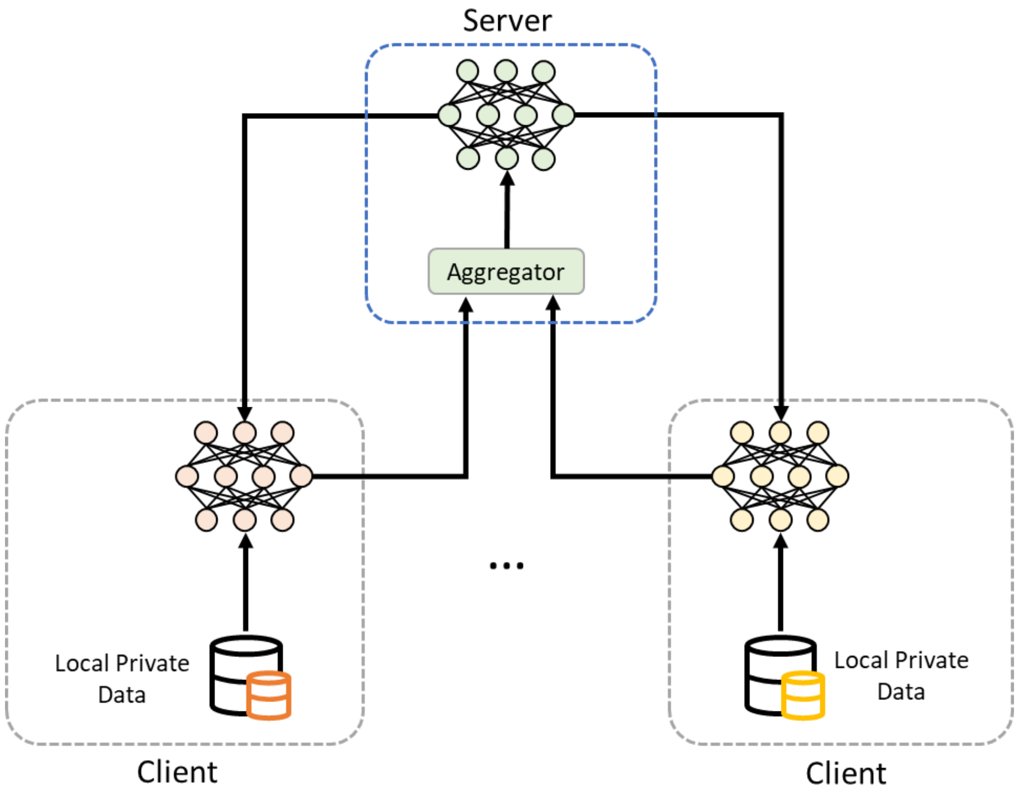
吴方照 微软亚洲研究院 主管研究员

Privacy is Important for AI

- AI relies on data for model training and online serving
  - Highly privacy sensitive in many scenarios
  - Strict laws on user privacy protection

Federated Learning

- Collaboratively learn a shared model while keeping data on device
- Decouple the ability of learning from the need of data centralization



Communication-Efficient Learning of Deep Networks from Decentralized Data, AISTATS 2017

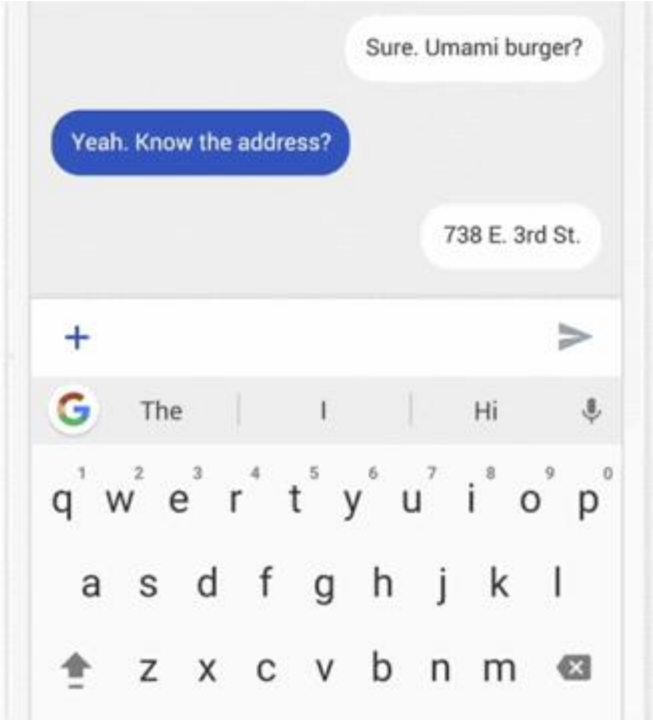
Applications of Federated Learning

- Examples
  - Gboard text prediction
  - Siri personalization

Artificial intelligence / Machine learning

How Apple personalizes Siri without hoovering up your data

The tech giant is using privacy-preserving machine learning to improve its voice assistant while keeping your data on your phone.

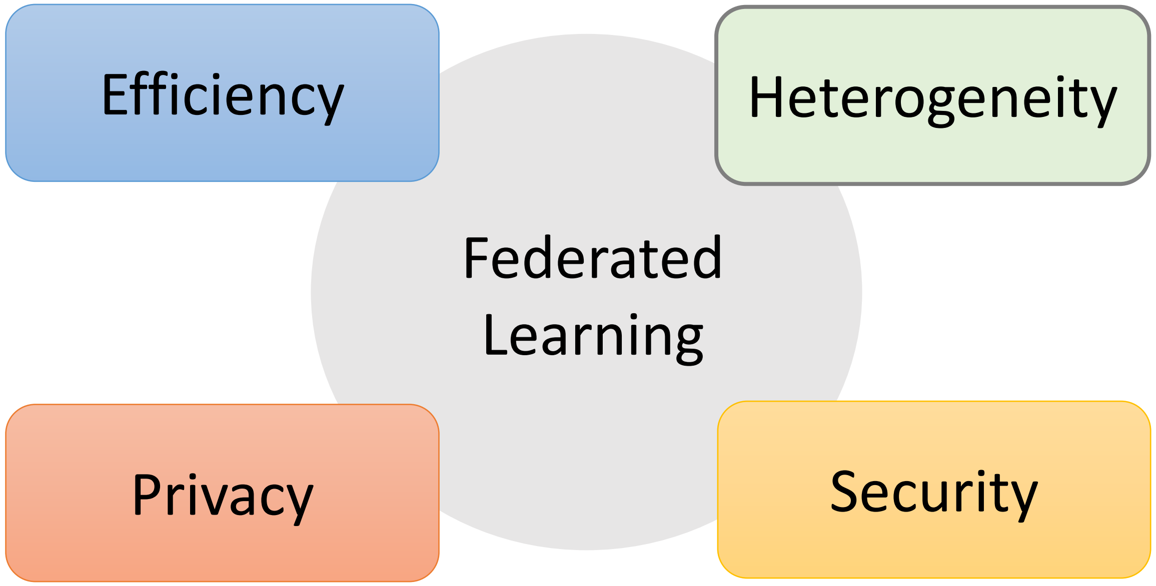


Learn how Gboard gets better

You can help improve voice and typing for everyone when you use the keyboard. A technology called federated learning helps Gboard learn new words and phrases without sending the text you speak or type to Google. What Gboard learns might be sent to Google services, without including what you typed or spoke, where it will be combined with learnings from other users to create better speech and typing models. Gboard only learns when your phone isn't being used, is charging, and is connected to Wi-Fi.

Learn how federated learning works.

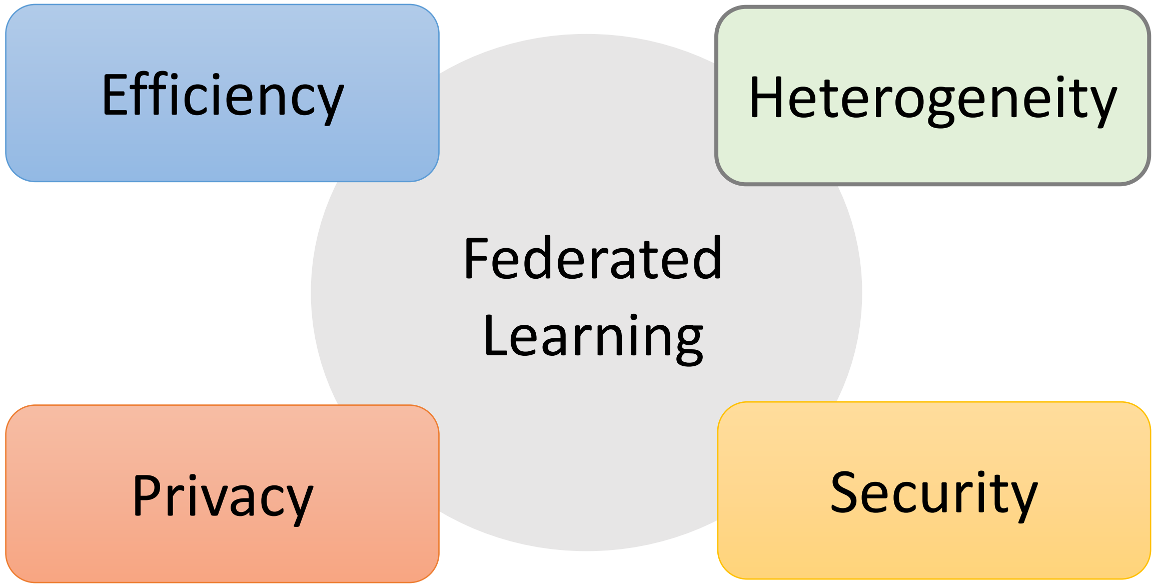
Federated Learning: Key Challenges



Federated Learning: Our Works

- FedKD
- Efficient-FedRec

- FedX
- InclusiveFL
- FedGNN
- FedCTR



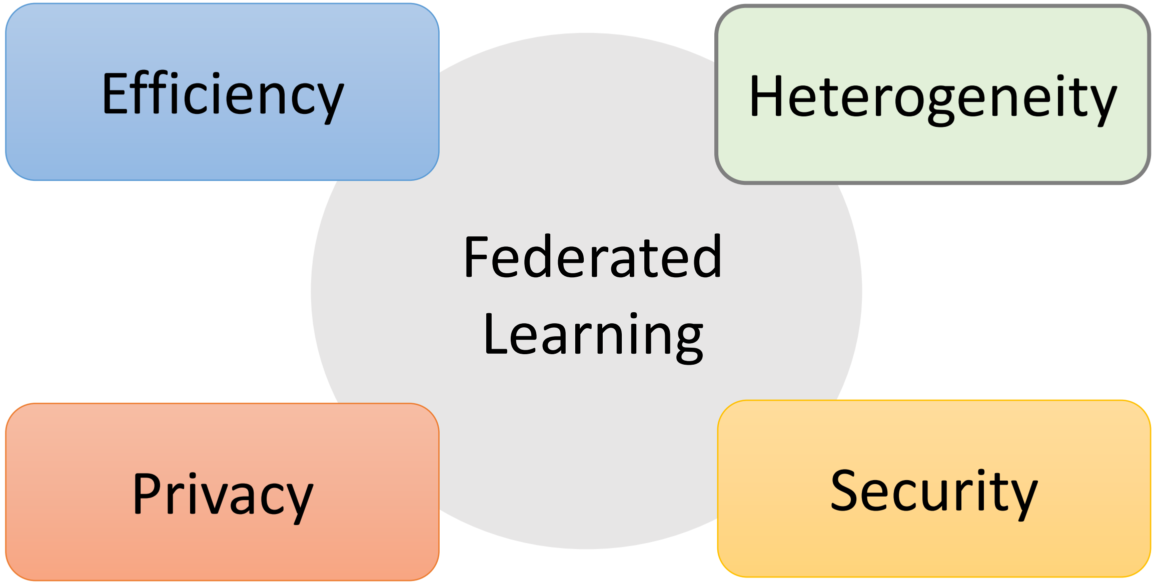
- UA-FedRec
- FedPrompt
- PrivateFL

- FedAttack
- RobustFL

Federated Learning: Our Works

- FedKD
- Efficient-FedRec

- FedX
- InclusiveFL
- FedGNN
- FedCTR

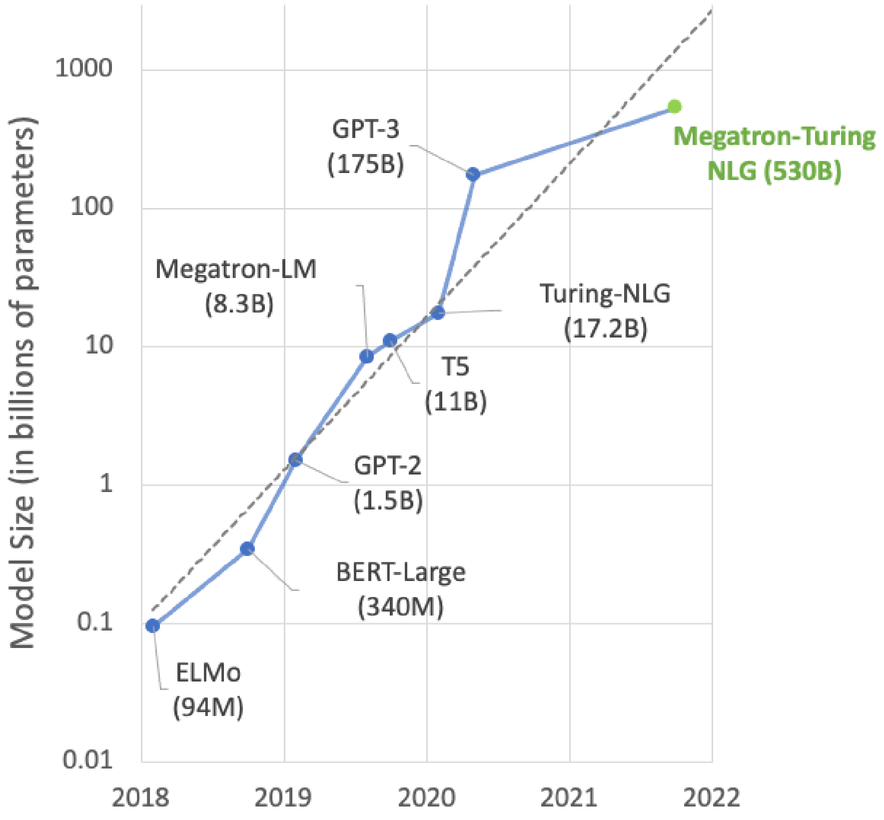


- UA-FedRec
- FedPrompt
- PrivateFL

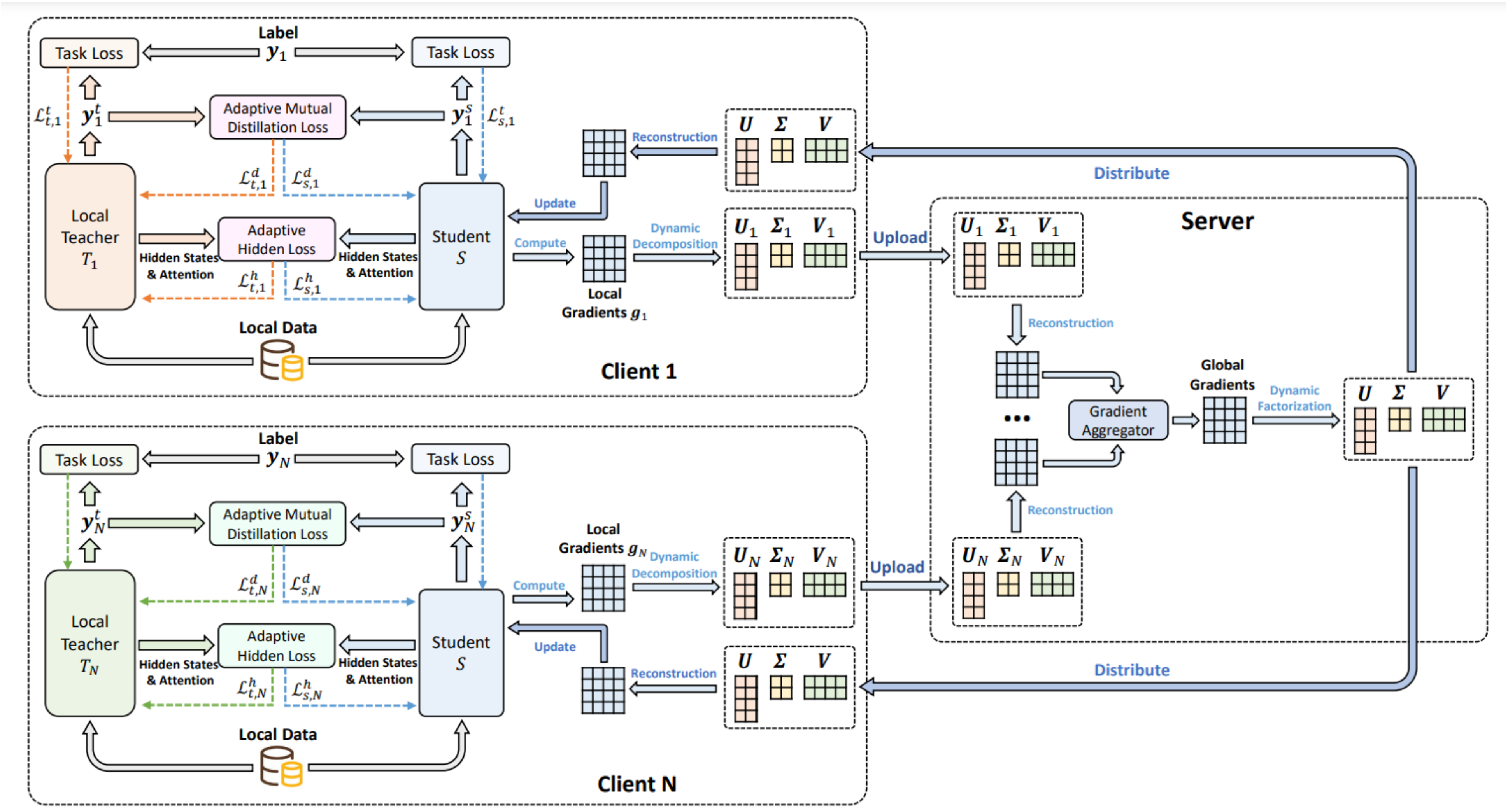
- FedAttack
- RobustFL

FedKD: Motivation

- AI models are bigger and bigger
  - Communication cost between client and server is huge



FedKD: Model



Communication-efficient federated learning via knowledge distillation, Nature Communications

FedKD: Experiments

- News recommendation

Methods	**AUC**	**MRR**	**nDCG@5**	**nDCG@10**	**Comm. Cost per Client**
UniLM (Local)	68.8±0.5	33.5±0.4	36.6±0.5	42.4±0.6	-
UniLM (Cen)	**71.0**±0.1	**35.8**±0.1	**39.0**±0.1	**44.8**±0.1	-
UniLM (Fed)	70.9±0.3	35.7±0.2	38.9±0.3	44.7±0.4	2.05GB
DistilBERT_6	69.3±0.2	34.0±0.2	37.5±0.2	43.0±0.1	1.03GB
DistilBERT_4	69.0±0.2	33.7±0.1	37.0±0.1	42.6±0.2	0.69GB
BERT-PKD_6	69.6±0.2	34.4±0.3	37.7±0.3	43.4±0.2	1.03GB
BERT-PKD_4	69.2±0.2	33.8±0.2	37.1±0.3	42.9±0.3	0.69GB
TinyBERT_6	69.7±0.2	34.5±0.2	37.9±0.1	43.5±0.2	1.03GB
TinyBERT_4	69.4±0.3	33.9±0.3	37.5±0.2	43.1±0.2	0.17GB
UniLM_4	69.6±0.1	34.4±0.2	37.7±0.1	43.4±0.2	0.69GB
UniLM_2	68.9±0.2	33.6±0.2	36.8±0.2	42.5±0.1	0.35GB
FetchSGD	70.5±0.4	35.2±0.3	38.2±0.3	44.0±0.4	0.51GB
FedDropout	70.5±0.2	35.1±0.2	38.3±0.3	44.2±0.3	1.23GB
FedKD_4	**71.0**±0.1	35.6±0.1	38.9±0.1	**44.8**±0.1	0.19GB
FedKD_2	70.5±0.1	35.3±0.2	38.6±0.1	44.3±0.2	**0.11GB**

FedKD: Experiments

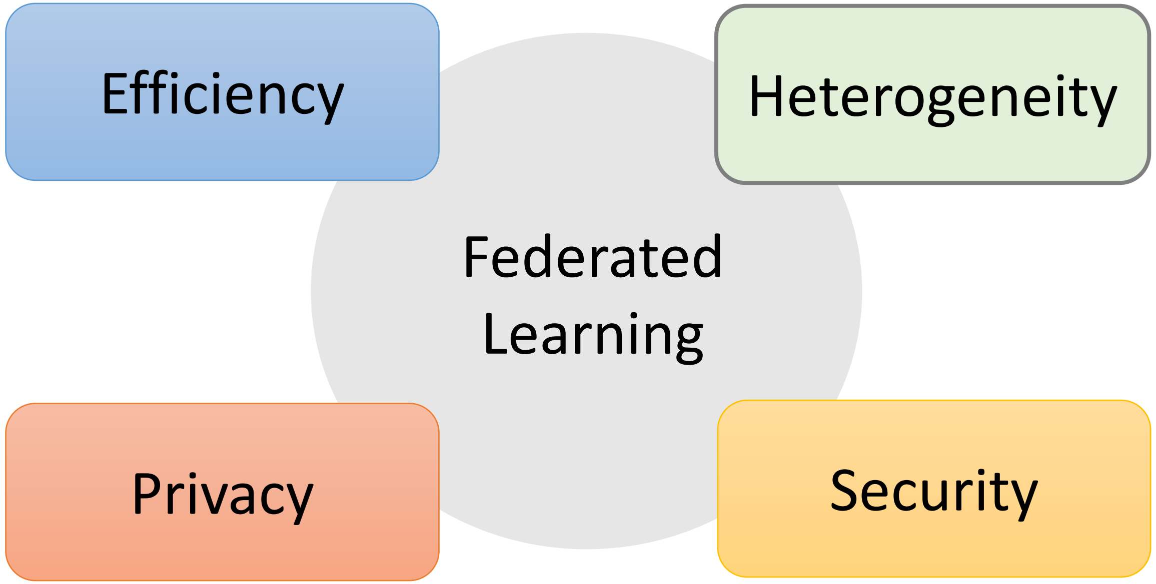
- Medical text classification

Methods	**Precision**	**Recall**	**Fscore**	**Comm. Cost per Client**
UniLM (Local)	53.2±1.3	54.6±1.4	53.9±1.1	-
UniLM (Cen)	**60.3**±0.7	61.6±0.8	**60.8**±0.4	-
UniLM (Fed)	59.1±0.6	62.3±0.6	60.6±0.4	1.37GB
DistilBERT_6	56.8±0.8	59.2±0.8	57.9±0.5	0.69GB
DistilBERT_4	56.5±0.9	58.4±1.1	57.1±0.7	0.46GB
BERT-PKD_6	56.9±0.9	60.4±0.8	58.4±0.6	0.69GB
BERT-PKD_4	56.3±1.1	59.9±0.7	58.0±0.6	0.46GB
TinyBERT_6	57.4±0.8	60.5±0.6	58.6±0.5	0.69GB
TinyBERT_4	57.0±0.7	59.9±1.2	58.3±0.7	0.12GB
UniLM_4	56.1±0.9	60.6±0.9	58.2±0.5	0.46GB
UniLM_2	53.8±0.8	59.1±1.0	56.3±0.6	0.24GB
FetchSGD	57.5±0.9	60.4±1.1	59.0±0.8	0.34GB
FedDropout	57.8±1.0	61.0±0.8	59.4±0.6	0.82GB
FedKD_4	59.4±0.6	**62.8**±0.9	60.7±0.5	0.12GB
FedKD_2	58.2±0.7	62.4±0.9	59.8±0.6	**0.07GB**

Federated Learning: Our Works

- FedKD
- Efficient-FedRec

- FedX
- InclusiveFL
- FedGNN
- FedCTR

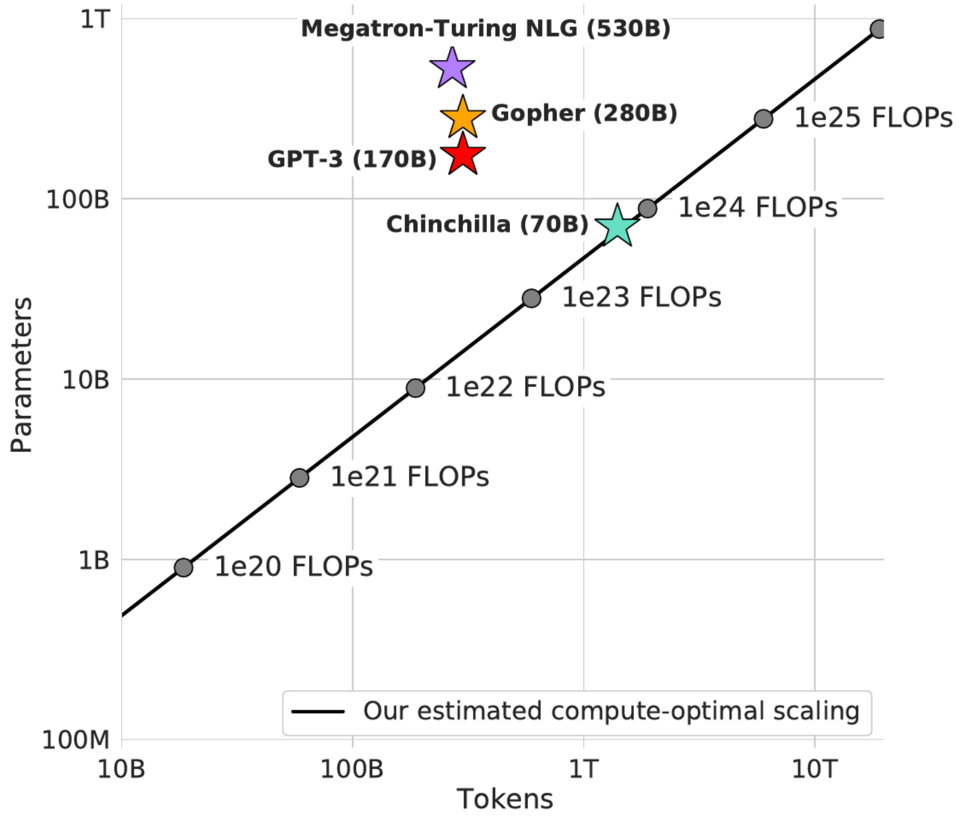


- UA-FedRec
- FedPrompt
- PrivateFL

- FedAttack
- RobustFL

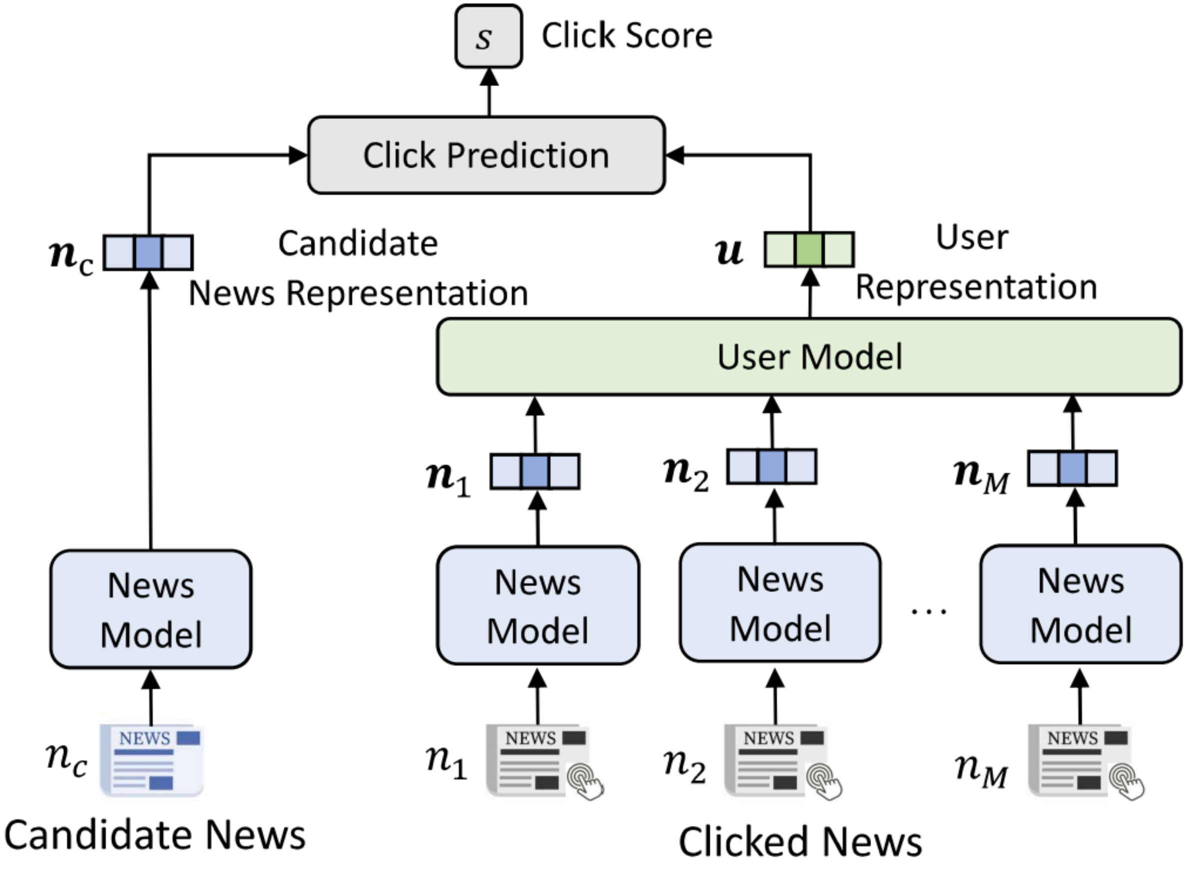
Efficient-FedRec: Motivation

- Big AI models are expensive to learn
  - Clients usually have weak computing capability

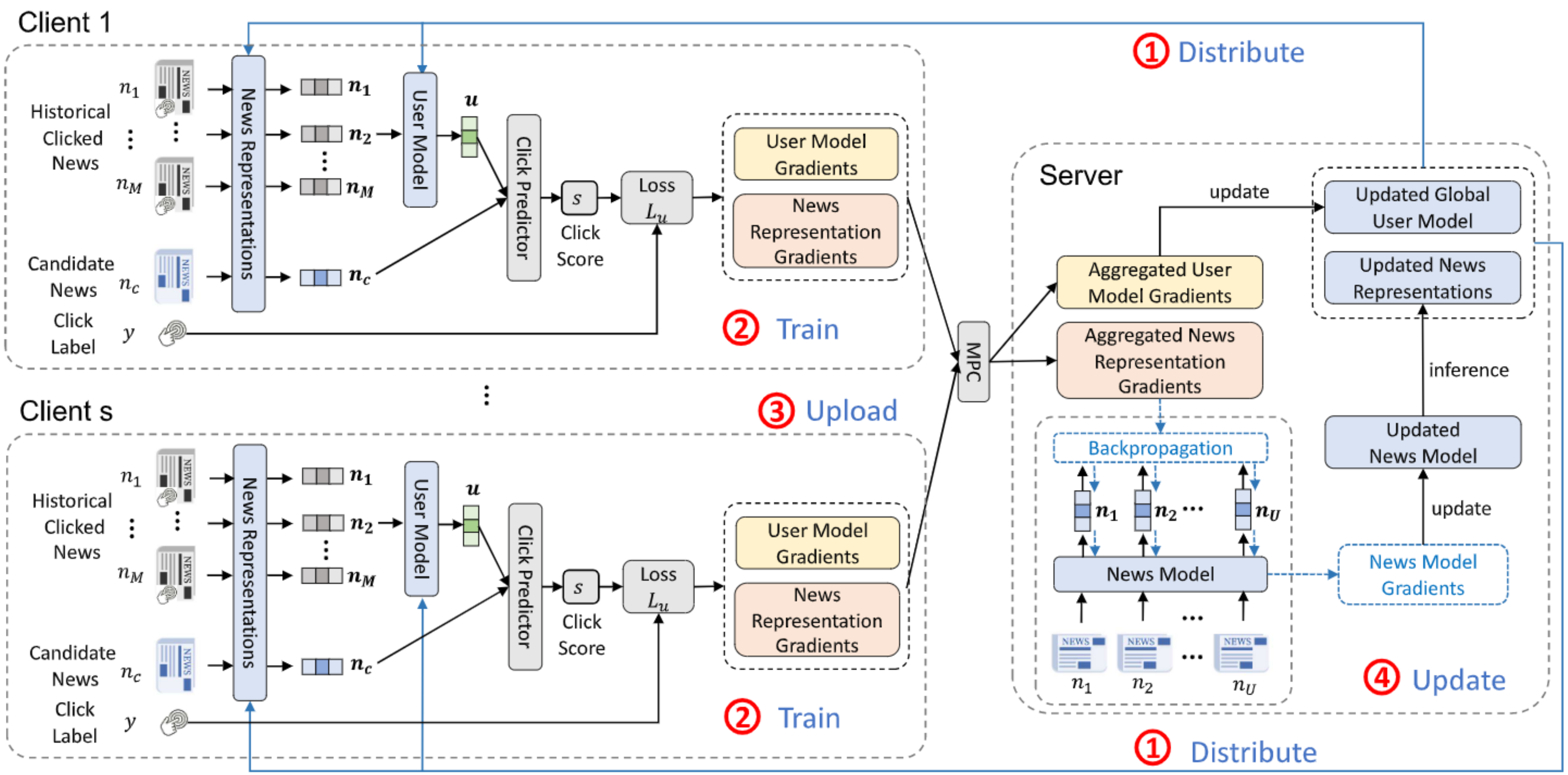


Efficient-FedRec: Motivation

- Sub-models may have different privacy and computing requirements
  - Split learning



Efficient-FedRec: Model



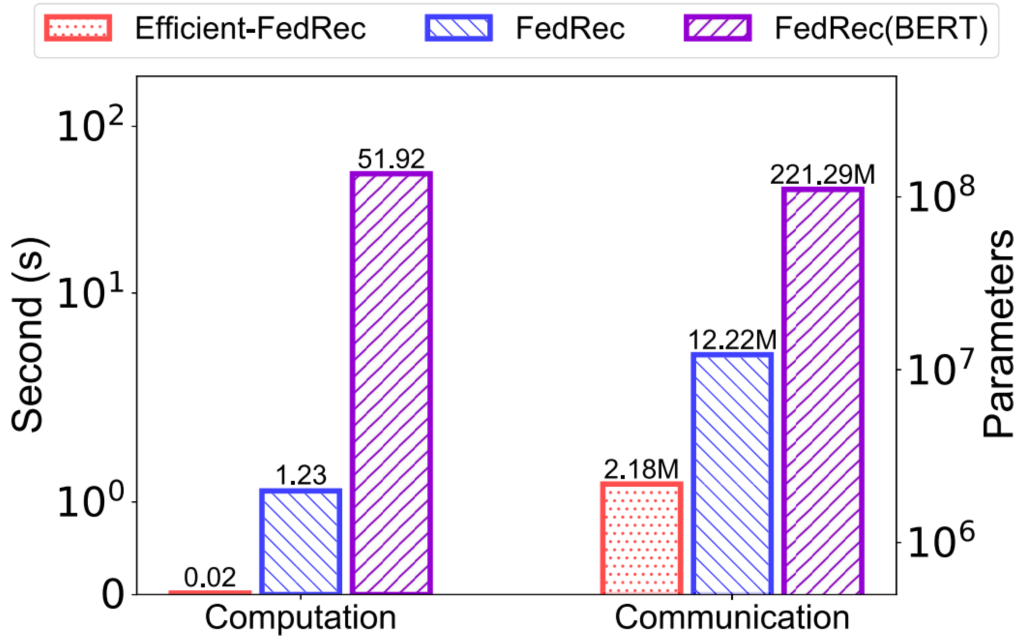
Efficient-FedRec: Experiment

- News Recommendation

Method	MIND				Adressa			
	AUC	MRR	nDCG@5	nDCG@10	AUC	MRR	nDCG@5	nDCG@10
DFM	60.67±0.20	28.08±0.13	29.93±0.13	35.68±0.13	59.90±1.20	32.68±0.75	29.69±0.93	36.43±1.11
DKN	64.72±0.19	30.53±0.13	33.01±0.15	38.70±0.16	73.73±0.48	39.52±1.34	40.98±1.24	47.48±0.86
LSTUR	66.90±0.08	32.45±0.07	35.11±0.07	40.82±0.07	68.37±2.63	38.76±2.14	38.11±2.39	44.33±2.42
NAML	66.10±0.25	31.91±0.23	34.52±0.26	40.21±0.24	73.09±1.53	44.27±1.53	43.51±1.89	50.02±1.71
NRMS	66.67±0.21	32.25±0.09	49.88±0.11	40.74±0.11	75.31±0.94	42.24±0.92	44.66±1.50	48.46±1.19
CenRec	66.92±0.17	32.30±0.11	35.05±0.13	40.78±0.14	72.85±1.53	40.82±1.73	41.62±2.24	47.54±1.47
PLM-NR	67.79±0.29	33.16±0.18	36.08±0.21	41.81±0.21	78.20±1.28	47.26±1.73	48.41±2.10	54.60±1.64
FCF	50.02±0.24	22.37±0.18	22.77±0.17	29.02±0.17	51.39±0.74	18.98±1.57	15.42±1.72	22.94±1.30
FedRec	66.54±0.18	31.96±0.07	34.54±0.09	40.30±0.09	71.73±1.72	41.37±2.21	41.81±2.35	47.18±2.09
FedRec(BERT)	67.45±0.10	32.80±0.10	35.44±0.16	41.35±0.14	78.60±1.82	43.81±0.95	45.76±0.89	52.64±1.68
Efficient-FedRec	67.44±0.20	32.79±0.06	35.62±0.06	41.35±0.07	79.08±1.18	45.09±1.87	47.13±2.35	53.85±1.69

Efficient-FedRec: Experiment

- Efficiency of computation and communication

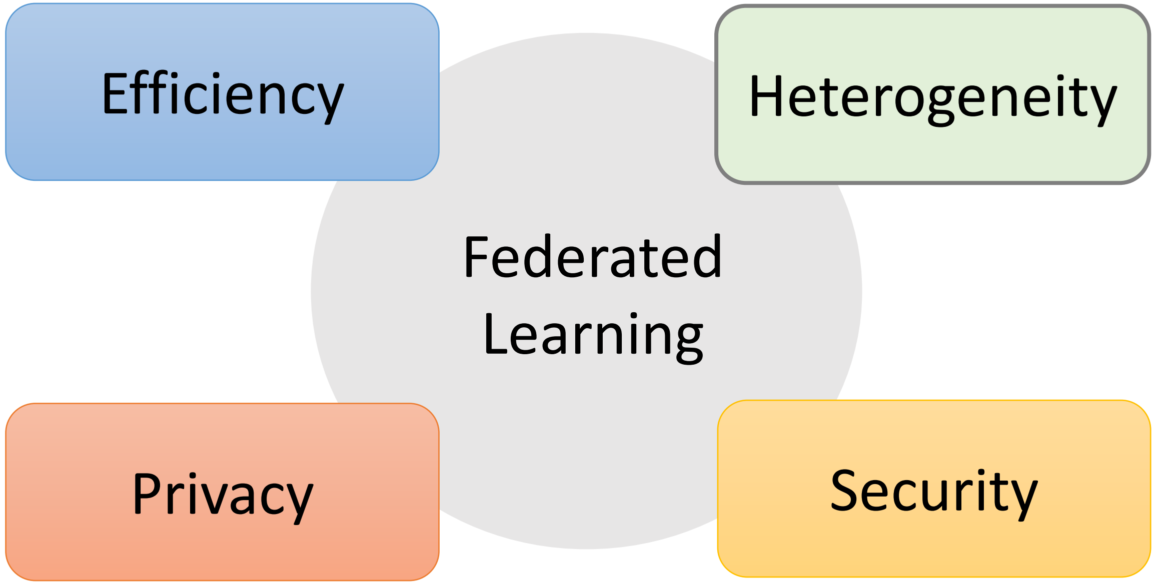


BERT	AUC	Efficient-FedRec Comm. Cost (client)	Efficient-FedRec Comp. Cost (client)	Efficient-FedRec Comp. Cost (server)	FedRec Comm. Cost (client)	FedRec Comp. Cost (client)	FedRec Comp. Cost (server)
Tiny	64.21	2.18M	0.02s	2.05s	10.01M	0.69s	0.01s
Mini	65.55	2.18M	0.02s	3.20s	23.74M	2.44s	0.01s
Small	65.92	2.18M	0.02s	5.88s	59.32M	9.03s	0.01s
Medium	67.05	2.18M	0.02s	6.39s	84.54M	19.55s	0.01s
Base	67.44	2.18M	0.02s	6.74s	221.29M	51.92s	0.02s
Large	67.50	2.18M	0.02s	8.81s	673.28M	117.04s	0.04s

Federated Learning: Our Works

- FedKD
- Efficient-FedRec

- FedX
- InclusiveFL
- FedGNN
- FedCTR

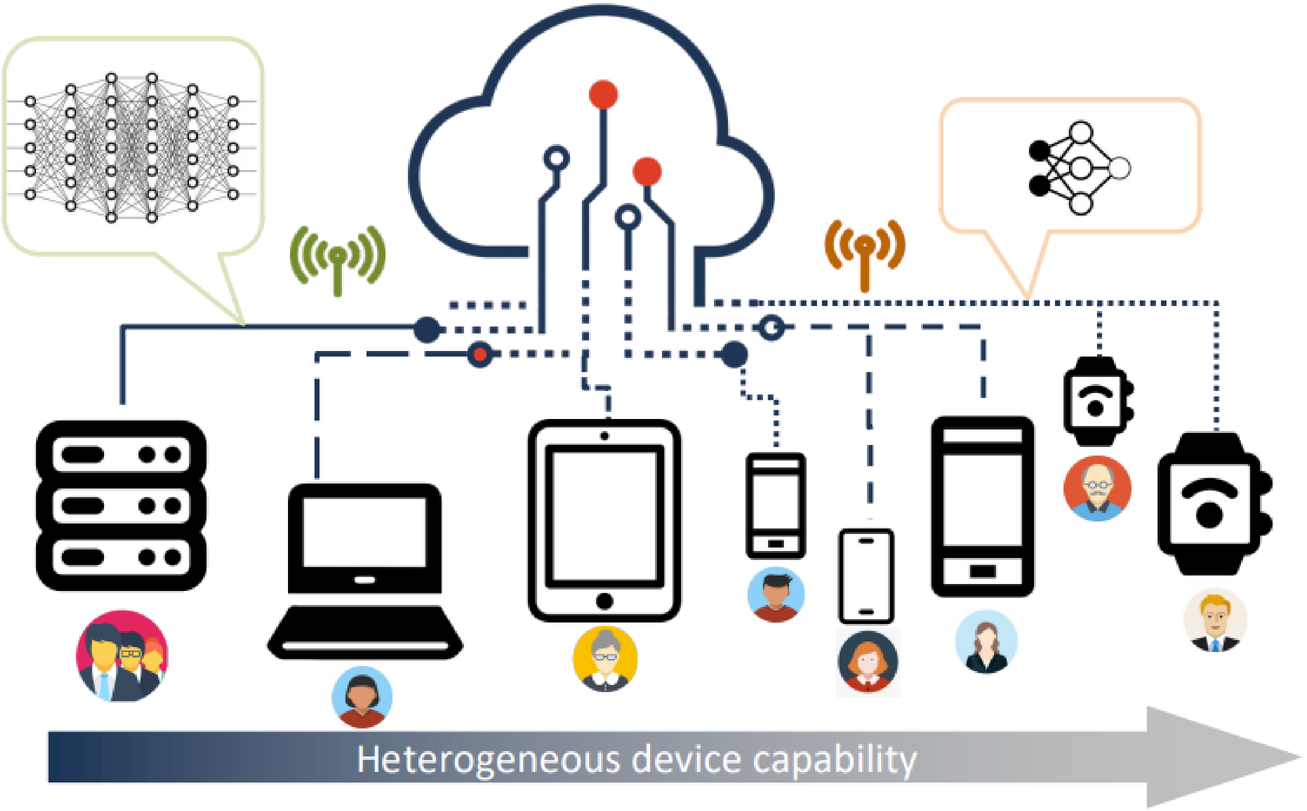


- UA-FedRec
- FedPrompt
- PrivateFL

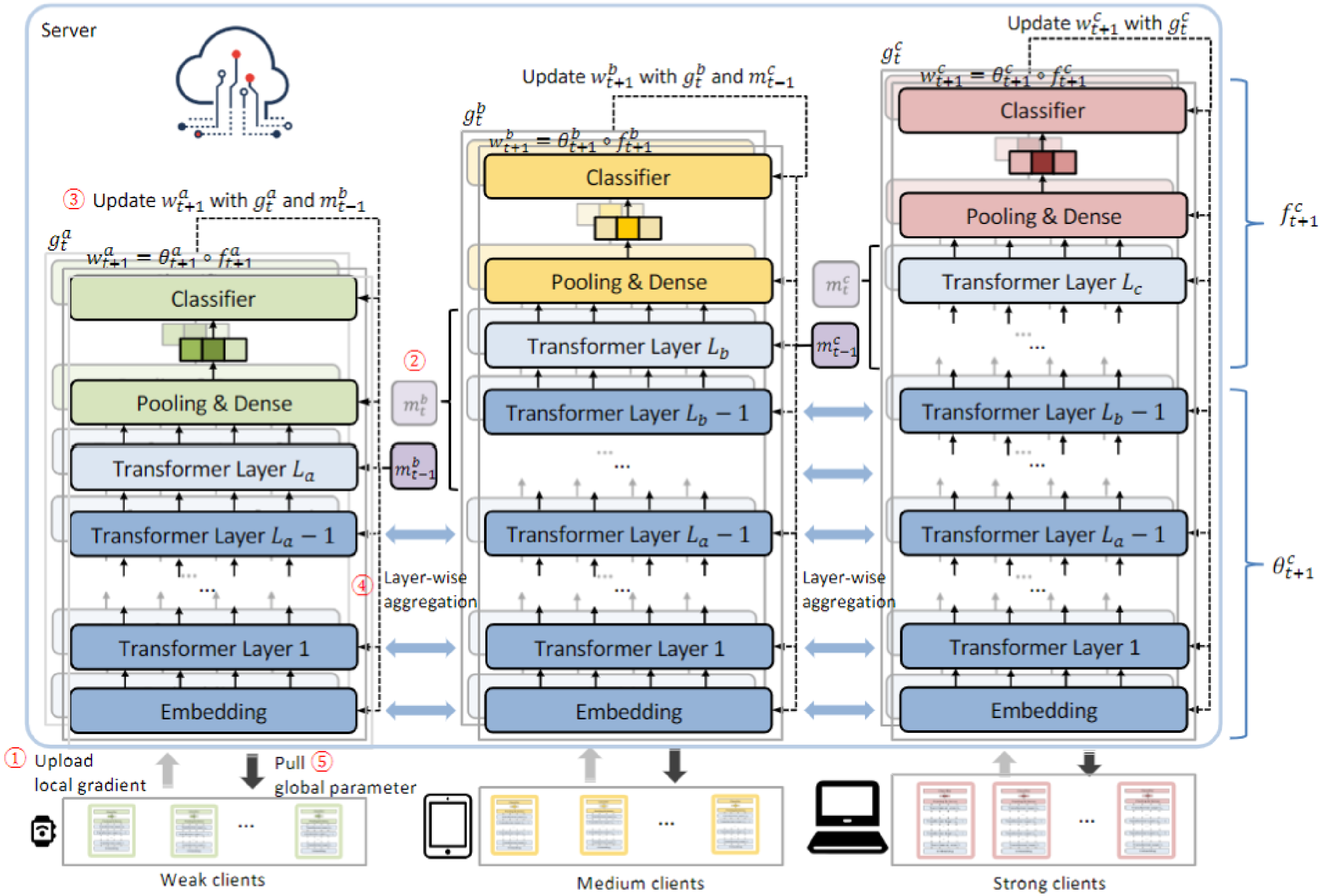
- FedAttack
- RobustFL

InclusiveFL: Motivation

- Heterogeneous client devices have different computing capabilities
  - Use small model for all clients?
  - Exclude weak clients for big model?



InclusiveFL: Model



No One Left Behind: Inclusive Federated Learning over Heterogeneous Devices, KDD 2022

InclusiveFL: Experiments

- Better performance due to contribution from all heterogeneous clients with affordable computing overhead

	Inclusive?	COLA	MNLI	MRPC	QNLI	QQP	RTE	SST2	STSB	Avg.
All-Large	N/A	63.03	86.48	91.50	92.09	91.49	76.12	94.43	90.60	85.72
Exclude-Weak	No	37.77	85.98	89.87	91.24	89.47	62.17	94.06	89.26	79.98
All-Small	Yes	34.91	78.83	82.50	85.93	79.37	58.94	90.14	83.68	74.29
HeteroFL	Yes	8.15	31.83	81.51	62.70	73.79	52.71	84.98	30.54	53.28
InclusiveFL	Yes	54.85	86.36	91.42	91.76	90.55	66.14	94.17	89.94	83.15

Federated Learning: Our Works

Federated Learning

Efficiency
- FedKD
- Efficient-FedRec

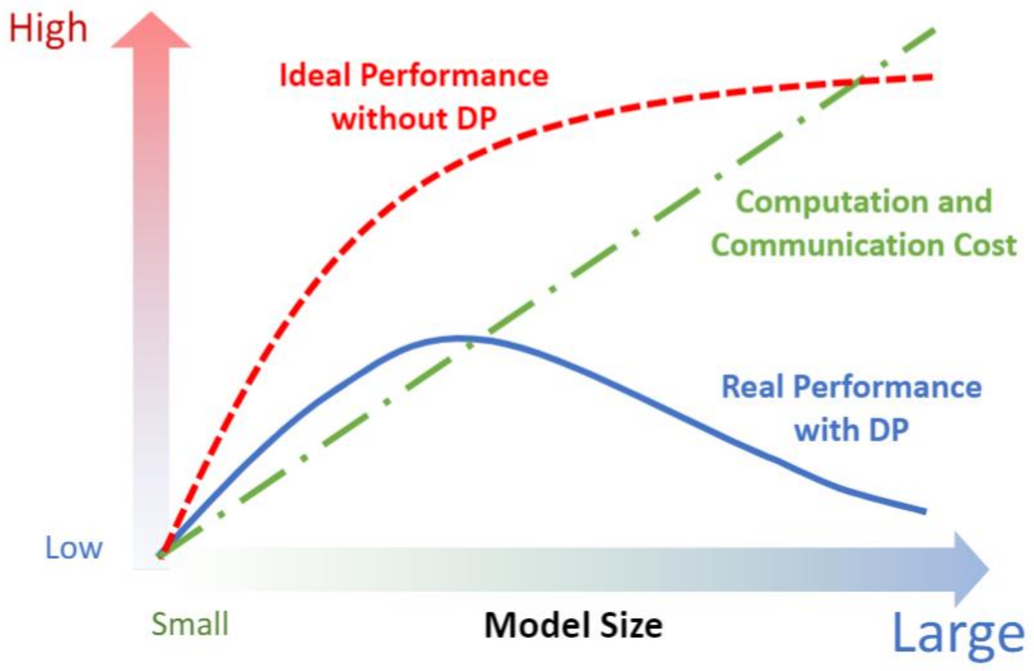
Heterogeneity
- FedX
- InclusiveFL
- FedGNN
- FedCTR

Privacy
- UA-FedRec
- FedPrompt
- PrivateFL

Security
- FedAttack
- RobustFL

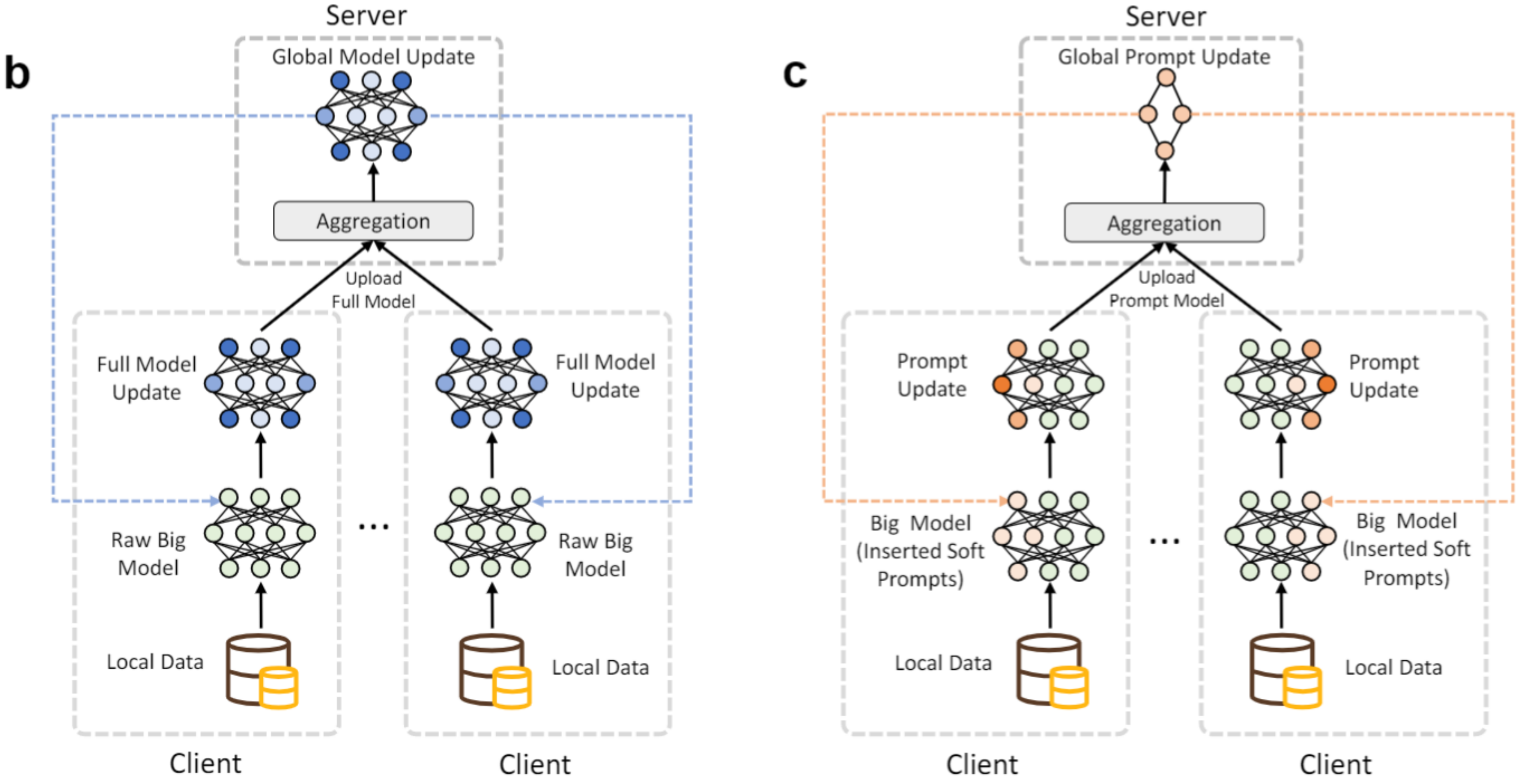
FedPrompt: Motivation

- Federated learning cannot provide strict privacy protection guarantee
- Solution: DP/LDP
  - Challenge: lower accuracy



FedPrompt: Model

- LDP+Prompt-tuning



FedPrompt: Experiments

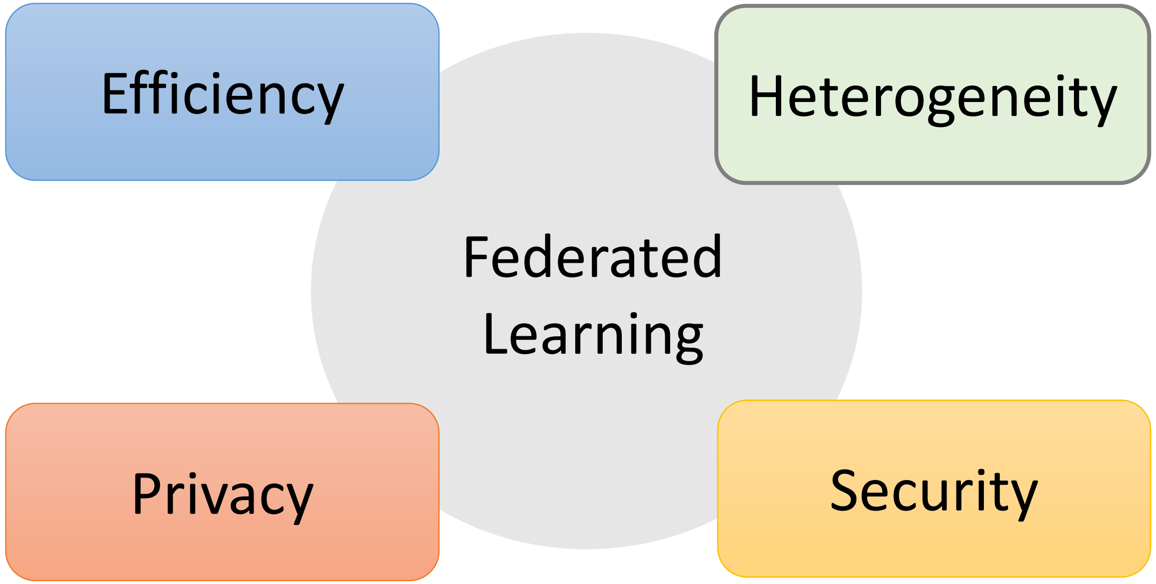
- NLP tasks

Basic Model	Finetuning Method	MNLI		QNLI		QQP		SST-2		Learnable Parameters
		w/o LDP	w/ LDP	w/o LDP	w/ LDP	w/o LDP	w/ LDP	w/o LDP	w/ LDP	
BERT-Base	Full	**84.3**±0.2	42.1±0.5	91.6±0.1	81.4±0.4	91.2±0.1	83.9±0.3	**93.1**±0.3	86.9±0.5	100%
	Prefix	80.5±0.1	65.0±0.7	86.5±0.2	81.2±0.4	87.4±0.1	79.2±0.5	91.5±0.2	84.8±0.6	0.01%
	P-Tuning	82.9±0.2	68.2±0.6	91.4±0.1	82.6±0.5	89.6±0.2	81.4±0.4	92.1±0.3	86.2±0.6	0.1%
	LoRA	84.0±0.3	**70.4**±0.6	91.8±0.1	83.1±0.6	90.5±0.1	84.4±0.3	93.0±0.2	87.5±0.4	1.0%
	P-Tuning v2	83.6±0.2	70.3±0.5	**92.0**±0.2	**83.3**±0.5	**91.3**±0.2	**84.9**±0.5	92.9±0.2	**87.7**±0.5	1.1%
BERT-Large	Full	86.2±0.3	40.7±0.8	**92.1**±0.2	80.2±0.4	91.4±0.2	83.0±0.3	93.3±0.3	85.8±0.6	100%
	Prefix	81.4±0.2	70.2±0.6	88.0±0.1	83.3±0.5	88.2±0.2	81.5±0.5	91.8±0.4	85.2±0.7	0.01%
	P-Tuning	84.0±0.3	71.3±0.7	91.7±0.2	83.6±0.4	89.9±0.1	82.0±0.4	92.4±0.3	86.4±0.7	0.1%
	LoRA	**86.3**±0.4	**72.5**±0.7	92.0±0.2	84.2±0.5	91.2±0.2	84.7±0.3	93.2±0.4	87.8±0.8	1.0%
	P-Tuning v2	85.9±0.3	72.0±0.7	**92.1**±0.2	**84.4**±0.5	**91.6**±0.2	**85.0**±0.5	**93.4**±0.2	**88.0**±0.6	1.0%
RoBERTa-Base	Full	**87.4**±0.2	44.1±0.8	**92.6**±0.1	82.5±0.5	**91.7**±0.1	84.3±0.3	94.7±0.2	87.4±0.5	100%
	Prefix	82.5±0.3	69.2±0.6	88.2±0.2	82.1±0.6	88.8±0.2	81.8±0.2	92.2±0.2	86.0±0.6	0.01%
	P-Tuning	84.9±0.3	73.6±0.5	92.0±0.2	83.3±0.5	90.2±0.1	82.9±0.2	93.5±0.3	87.0±0.6	0.1%
	LoRA	**87.4**±0.2	**75.5**±0.6	92.5±0.1	84.5±0.4	91.0±0.1	84.8±0.3	**94.6**±0.2	**88.1**±0.5	0.9%
	P-Tuning v2	87.0±0.3	75.2±0.6	92.4±0.1	**84.7**±0.6	91.6±0.1	**85.1**±0.2	94.5±0.2	87.9±0.7	1.0%
RoBERTa-Large	Full	90.0±0.3	42.3±0.5	94.4±0.2	81.0±0.5	**92.0**±0.1	83.5±0.2	96.1±0.3	86.1±0.6	100%
	Prefix	84.4±0.2	71.1±0.7	91.6±0.2	83.4±0.5	89.2±0.1	82.3±0.3	92.9±0.4	87.4±0.7	0.01%
	P-Tuning	87.8±0.3	74.3±0.6	93.9±0.1	84.9±0.3	91.0±0.1	83.6±0.2	94.0±0.3	87.5±0.5	0.1%
	LoRA	**90.4**±0.2	**77.2**±0.7	**94.5**±0.2	86.1±0.6	91.8±0.2	85.3±0.3	96.0±0.2	**88.4**±0.5	0.9%
	P-Tuning v2	90.2±0.2	77.0±0.8	94.2±0.2	**86.2**±0.4	91.9±0.1	**85.5**±0.3	**96.1**±0.2	88.3±0.4	1.0%

Federated Learning: Our Works

- FedKD
- Efficient-FedRec

- FedX
- InclusiveFL
- FedGNN
- FedCTR

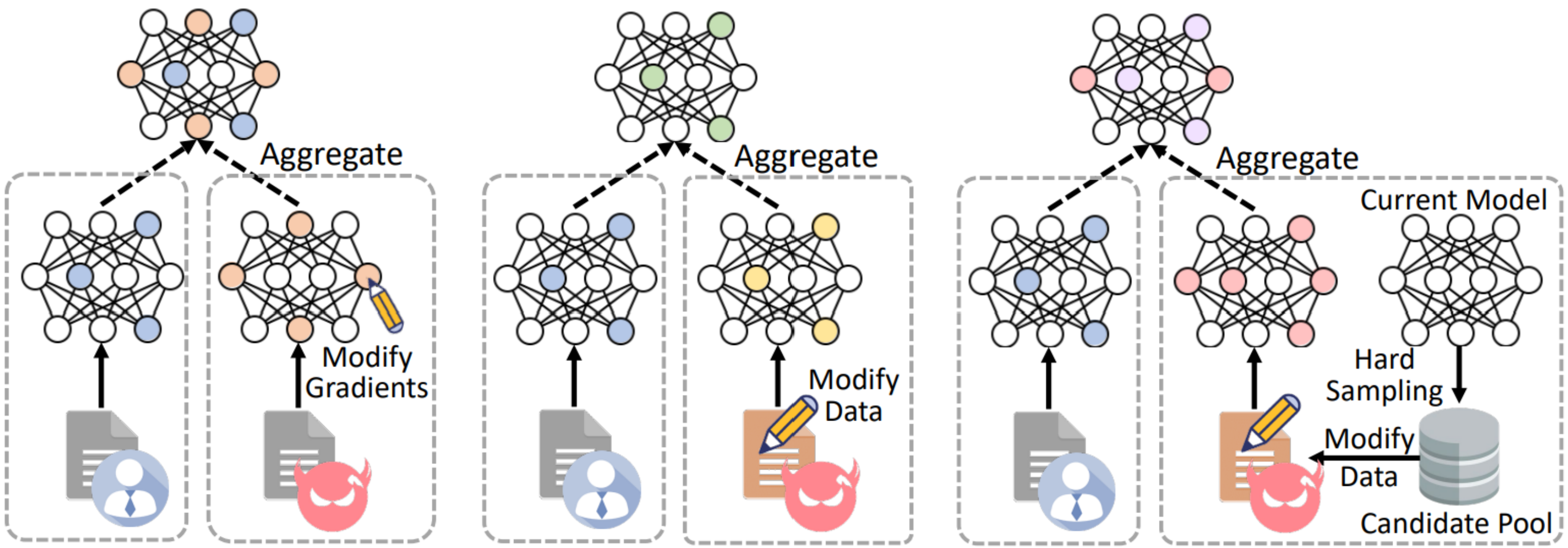


- UA-FedRec
- FedPrompt
- PrivateFL

- FedAttack
- RobustFL

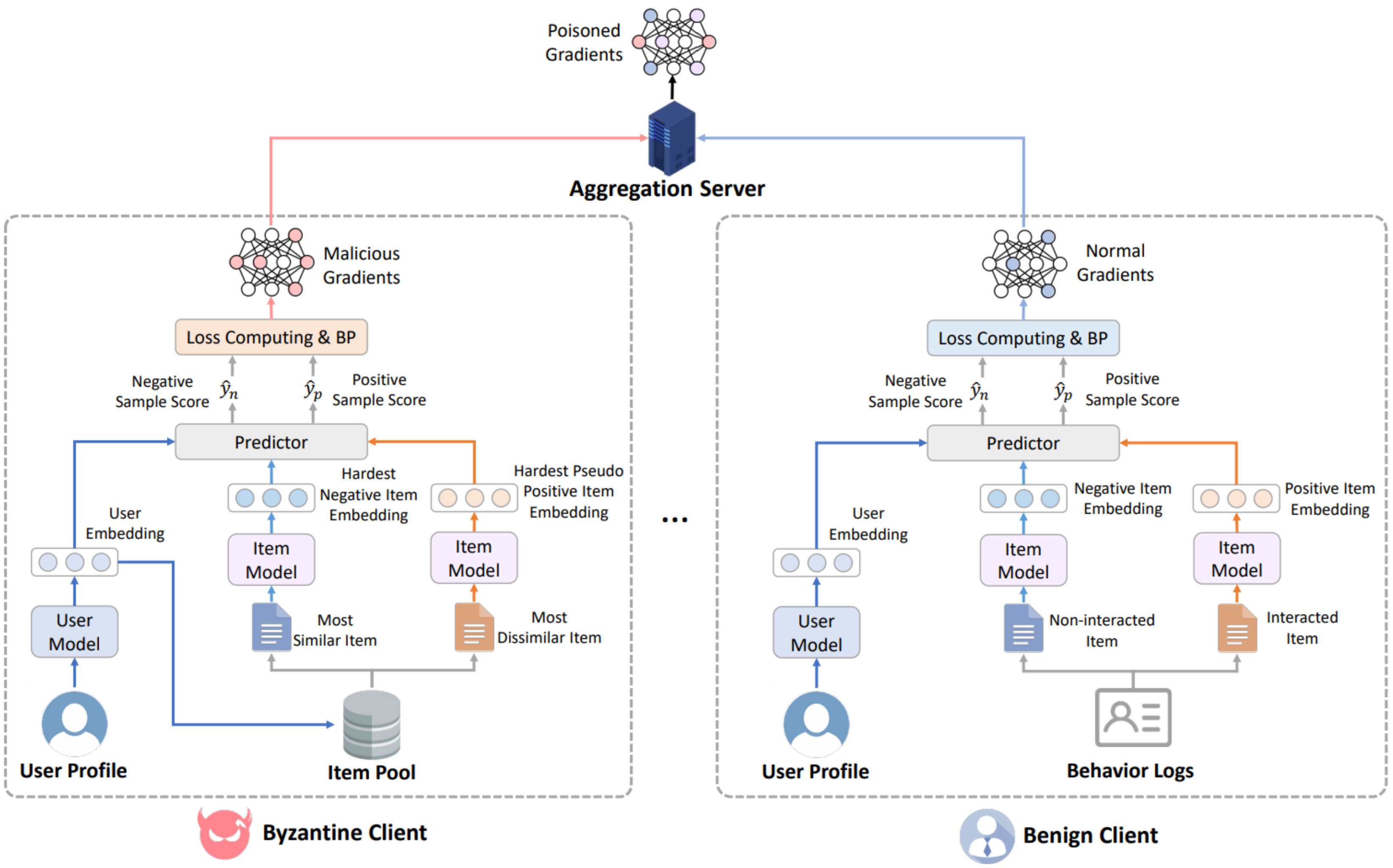
FedAttack: Motivation

- Federated learning is vulnerable
  - Data poison attack
  - Model poison attack



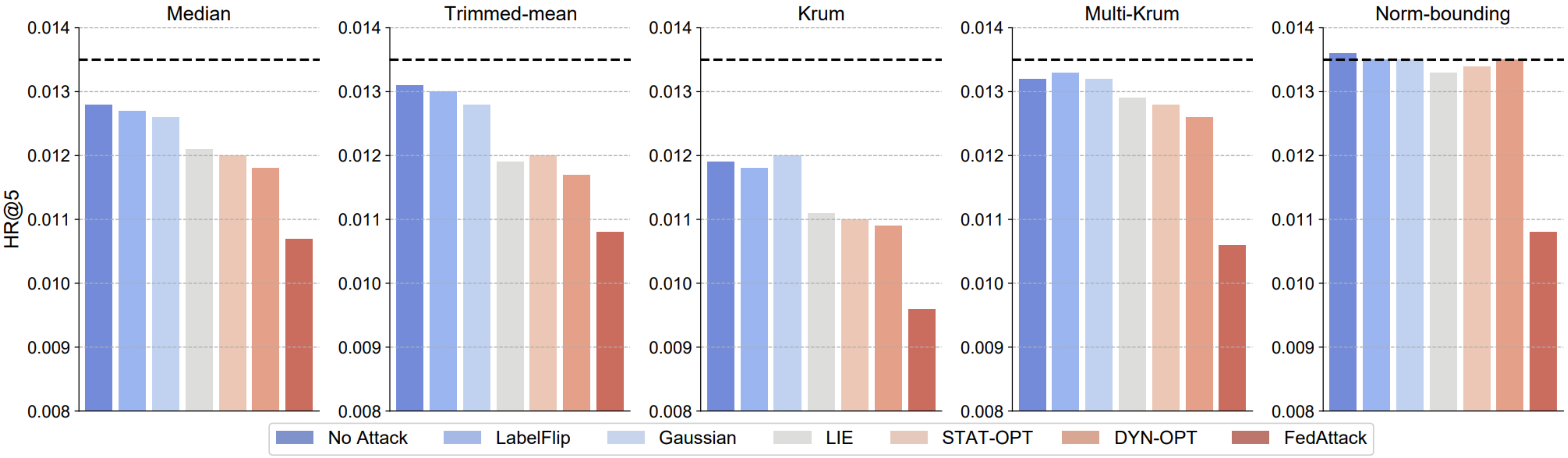
(a) Gradient poisoning **(b) Data poisoning** **(c) FedAttack: hard sampling poisoning**

FedAttack: Model



FedAttack: Effective and Covert Poisoning Attack on Federated Recommendation via Hard Sampling, KDD 2022

FedAttack: Experiments



非常感谢您的观看

Microsoft Research 微软亚洲研究院 | DataFun.