FairVFL: A Fair Vertical Federated Learning Framework with Contrastive Adversarial Learning

Tao Qi[1], Fangzhao Wu[2], Chuhan Wu[1], Lingjuan Lyu[3], Tong Xu[4],
Hao Liao[5], Zhongliang Yang[1], Yongfeng Huang[1,6], Xing Xie[2]

[1]Department of Electronic Engineering & BNRist, Tsinghua University, Beijing 100084, China

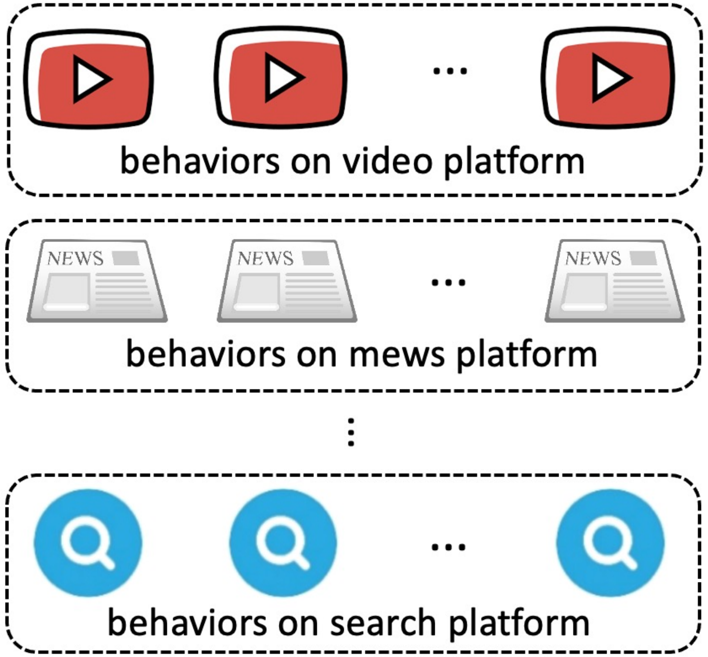
[2]Microsoft Research Asia, Beijing 100080, China

[3]Sony AI, [4]USTC, [5]Shenzhen University, [6]Zhongguancun Laboratory

taoqi.qt@gmail.com

Background

- Data volume explosion has empowered ML models on intelligent tasks
- Feature fields of the same sample may decentralized across platforms
- Centralizing feature fields for model training may arouse privacy concerns

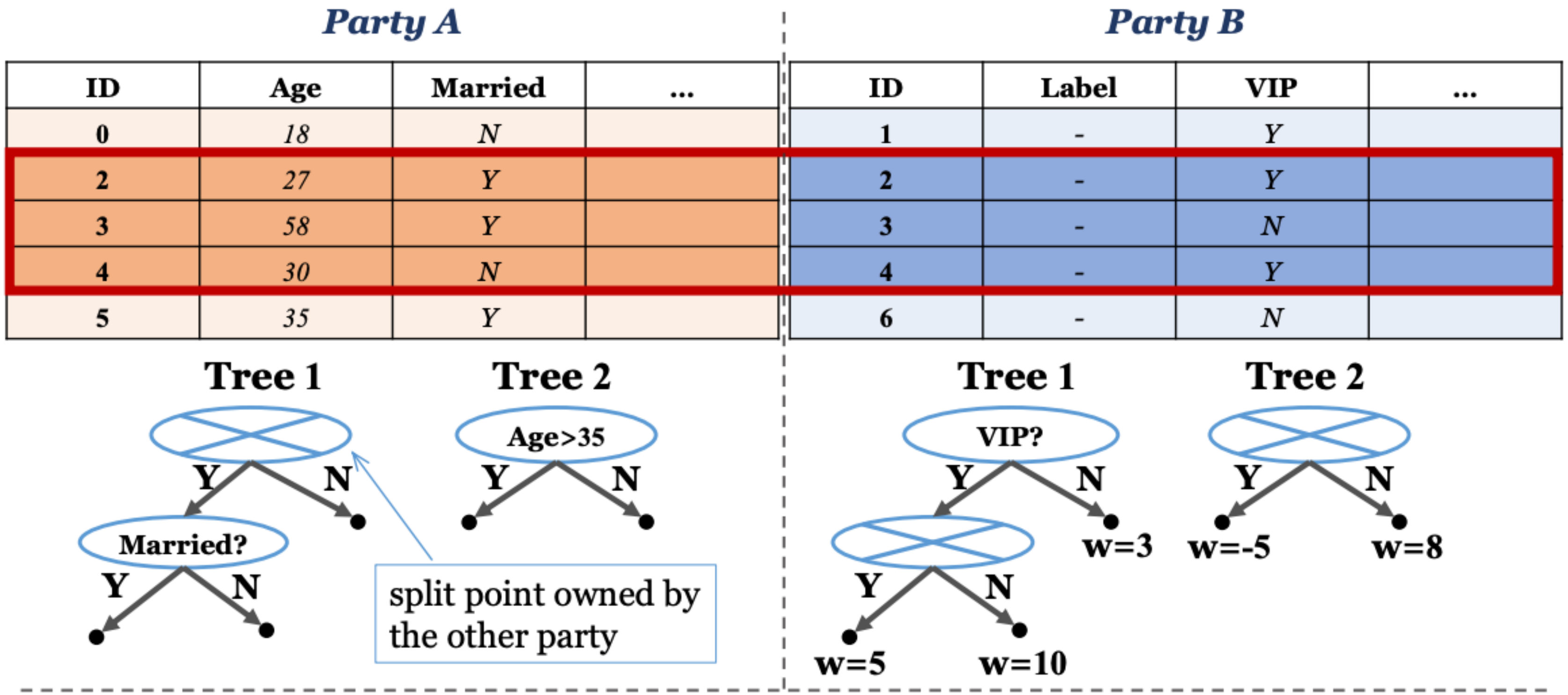


Behaviors of a target user on online platforms.

GDPR: Privacy regulation.

Vertical Federated Learning

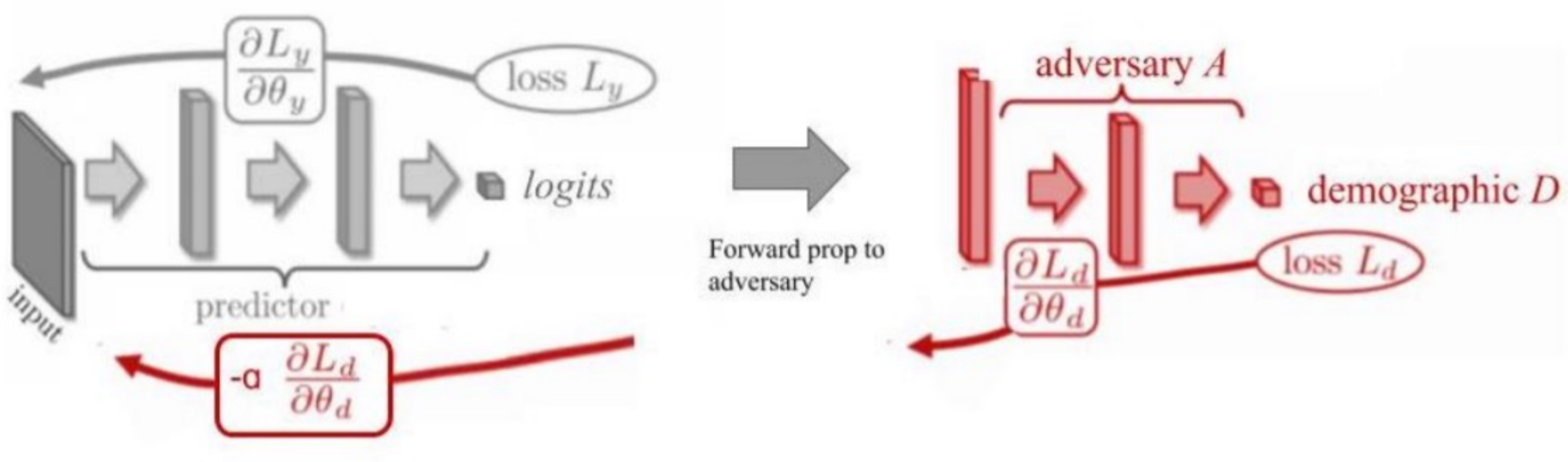
- VFL can utilize decentralized feature fields to learn model



- Challenges
  - Real-world data usually encode bias on sensitive attributes (e.g., gender)
  - VFL models may inherit bias and become unfair for some user groups

Fair Machine Learning

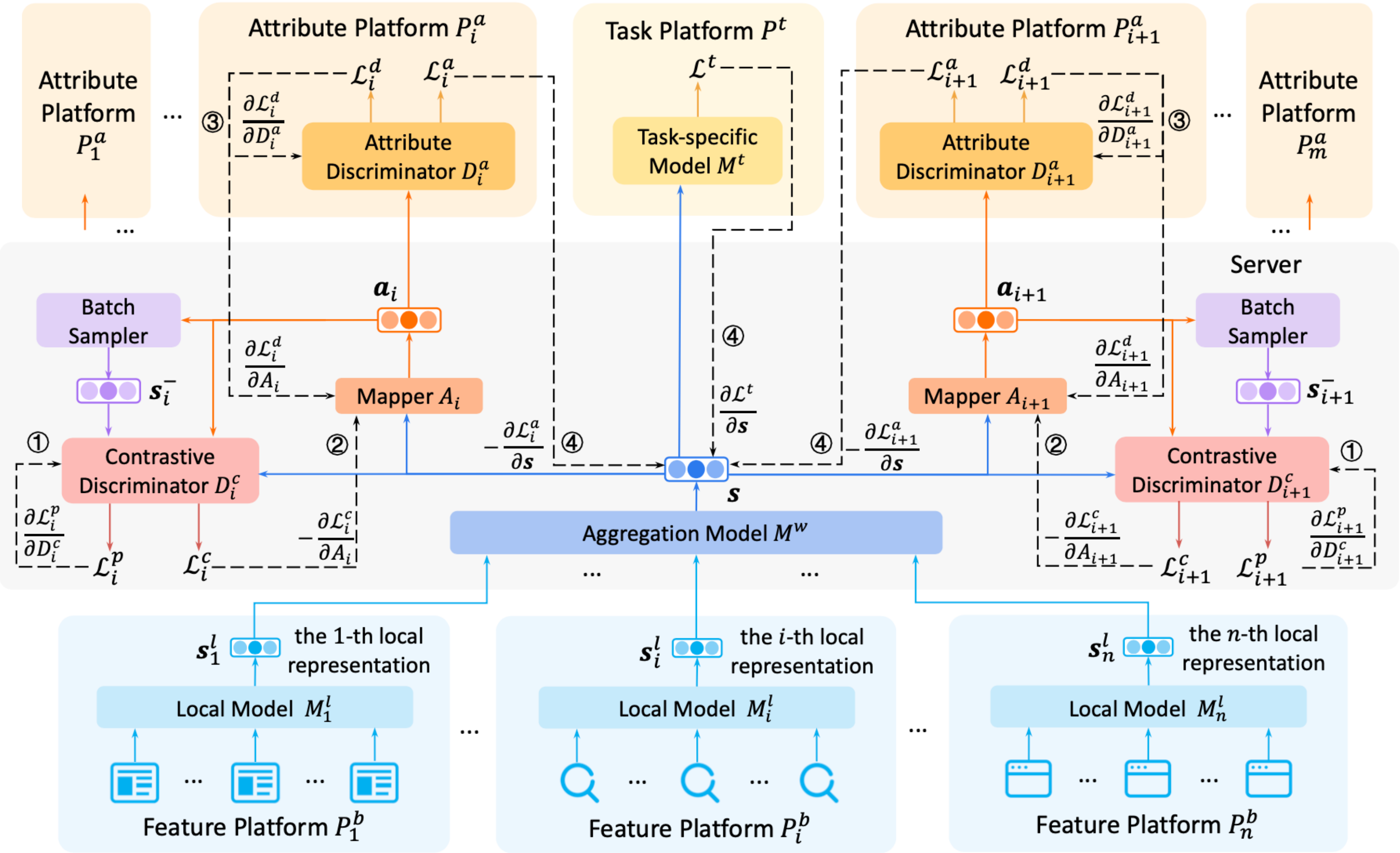
- Aim to eliminate the effect of sensitive user attributes on model decisions



- Challenges
  - Most existing methods rely on centralized storage of feature fields attribute labels

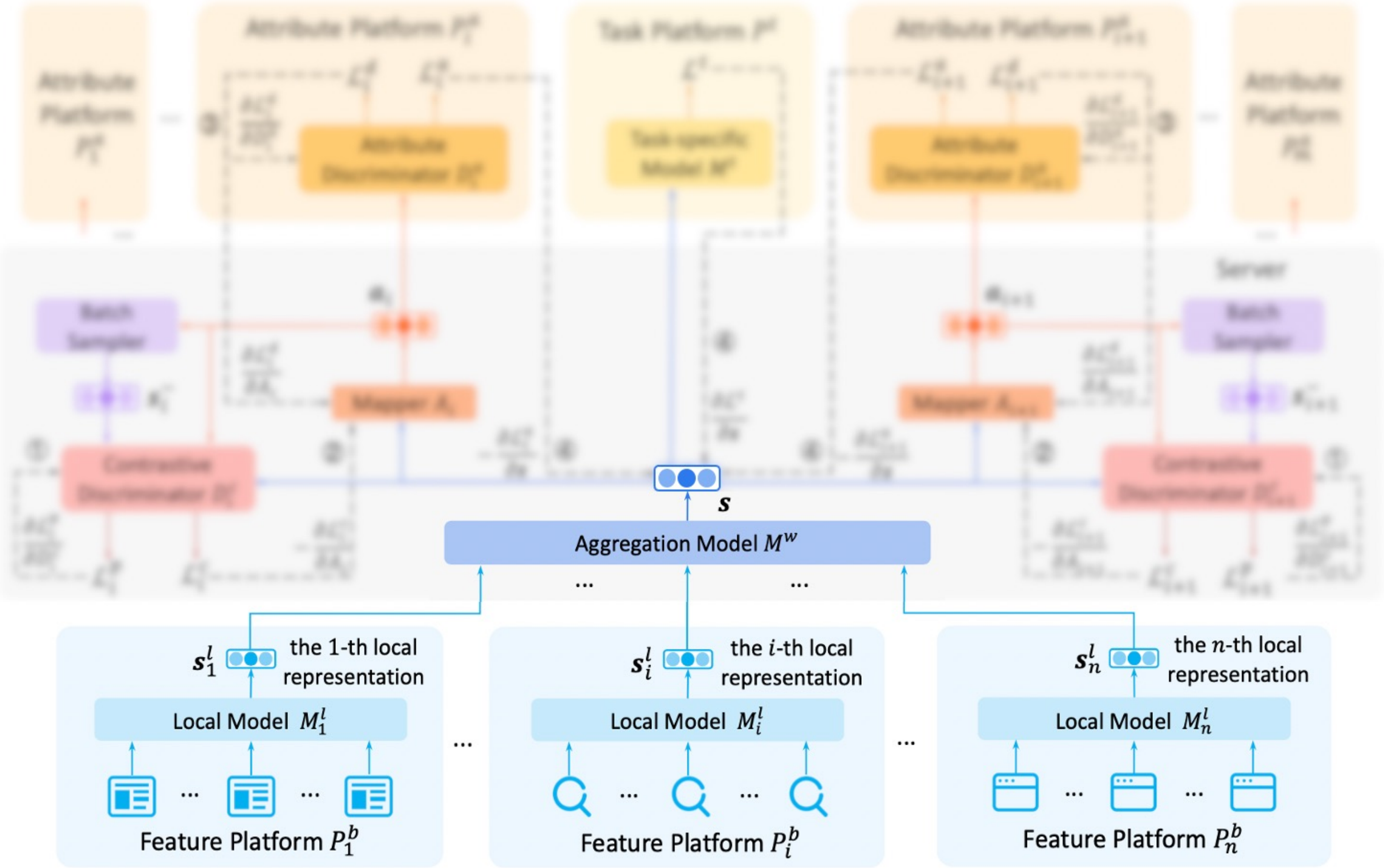
FairVFL: Fair Vertical Federated Learning

- Improve the fairness of VFL models with user privacy well protected



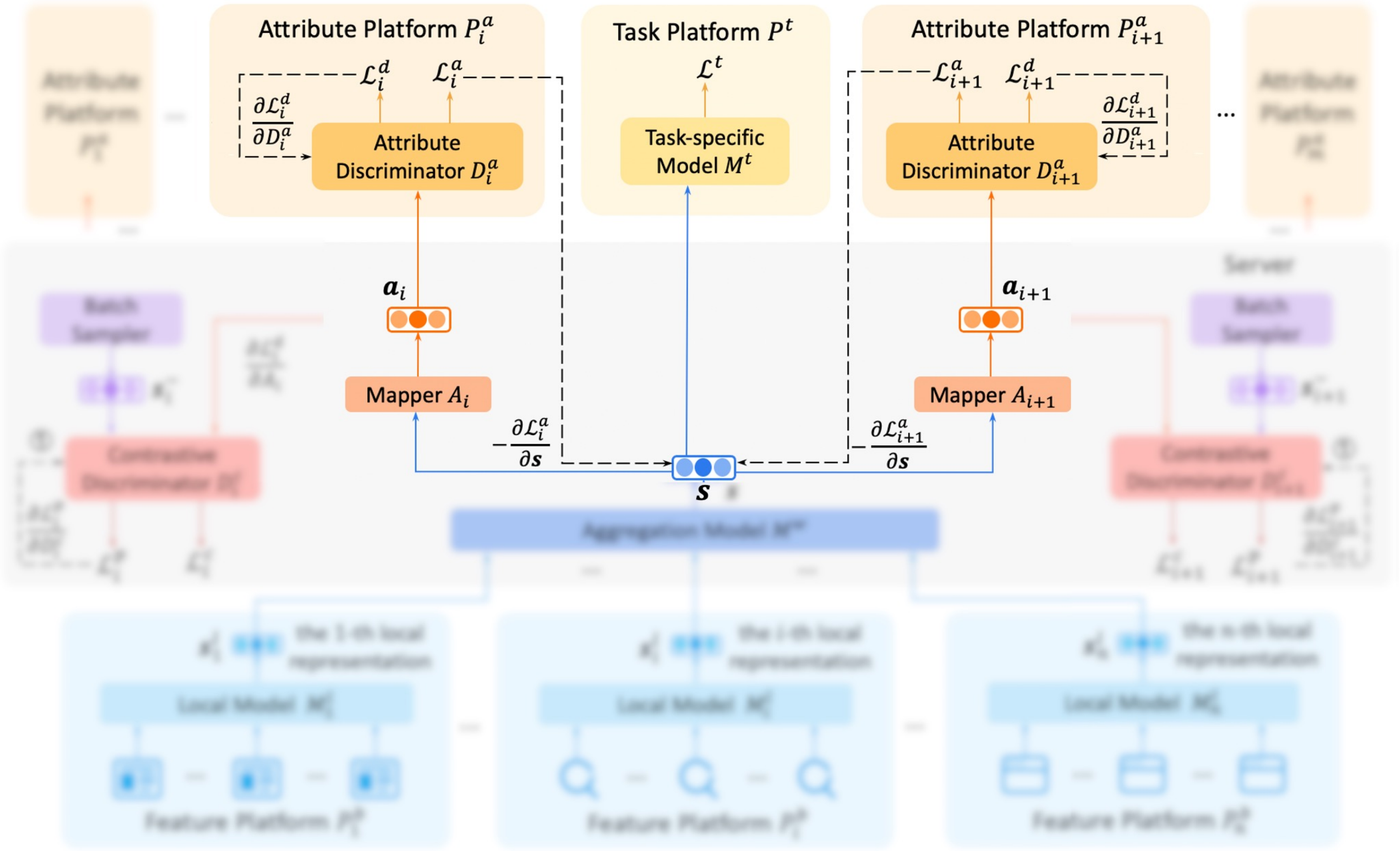
FairVFL: Fair Vertical Federated Learning

- Learn a unified representation to encode decentralized feature fields



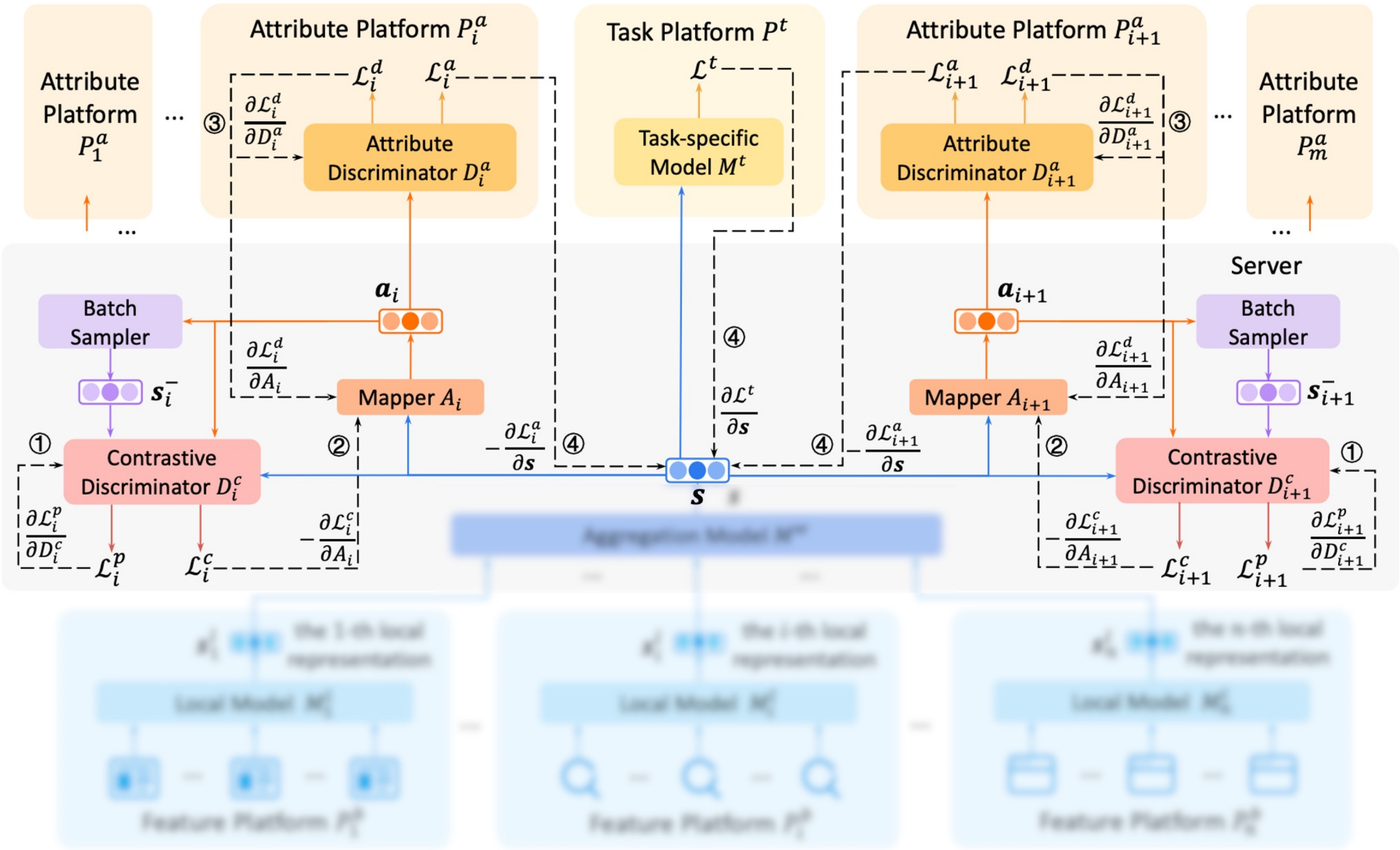
FairVFL: Fair Vertical Federated Learning

- Reduce bias in unified representation via adversarial learning



FairVFL: Fair Vertical Federated Learning

- Protect privacy in exchanged representations via contrastive adversarial learning



Datasets

- ADULT
  - A public dataset for income prediction task
  - Predicting income of users from various user feature (e.g., education level)
- NEWS
  - Personalized news recommendation task
  - Recommending news based on user's historical news clicks, search, browsing
  - Constructed by user logs on Microsoft News and Bing
- Fairness metric: classification on sensitive attributes via an attack model

ADULT			
# Users	30,000	# Samples	30,000
# Insensitive features fields			12
NEWS			
# Users	151,389	# News	112,052
# Samples	334,612	Avg. # Clicks	77.57
Avg. # Browsing	34.95	Avg. # Search	38.53

Results on Performance and Fairness

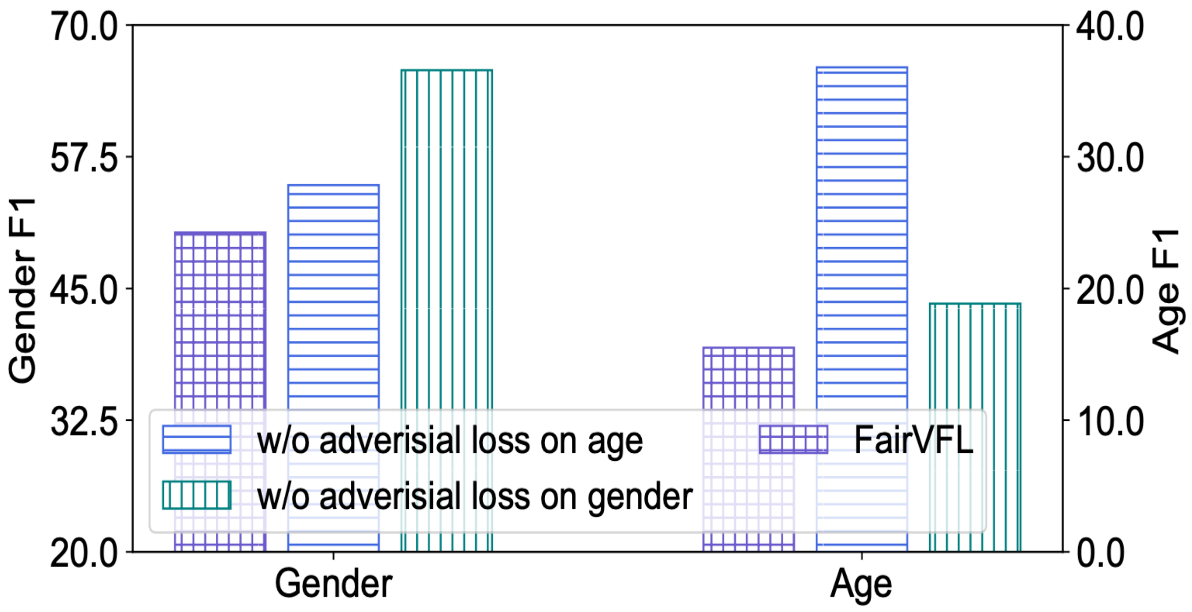
	Income Prediction		Model Fairness	
	Accuracy	F1	Gender F1	Age F1
MLP	82.15±0.86	78.42±0.66	78.27±1.60	47.47±0.90
MLP+FairGo	77.97±1.34	78.49±1.25	50.92±6.25	15.92±3.66
MLP+FairSM	77.57±1.39	78.31±1.06	50.66±6.43	15.55±3.10
MLP+FairRec	77.59±1.42	78.08±1.24	50.94±7.15	15.75±4.62
MLP+VFL	81.47±2.14	77.82±1.57	79.05±0.91	47.48±1.25
MLP+FairVFL	76.74±2.64	77.87±2.14	50.31±4.99	15.50±4.33
TabNet	82.23±1.02	78.50±0.80	79.07±1.79	49.12±1.24
TabNet+FairGo	74.40±1.71	75.33±1.47	50.28±5.32	15.63±3.14
TabNet+FairSM	74.04±1.66	75.07±1.39	50.61±4.44	15.98±3.14
TabNet+FairRec	74.89±1.76	75.65±1.61	50.35±5.79	15.61±3.18
TabNet+VFL	81.77±1.72	78.09±1.35	78.75±0.77	48.36±1.32
TabNet+FairVFL	75.51±0.69	76.06±0.60	50.72±5.72	15.48±2.46
AutoInt	82.31±1.92	78.49±1.50	79.22±0.84	48.99±1.25
AutoInt+FairGo	76.89±1.50	77.31±1.27	50.59±4.41	15.73±2.27
AutoInt+FairSM	76.30±2.56	76.88±2.05	50.53±5.02	15.50±3.10
AutoInt+FairRec	76.60±1.91	77.17±1.54	50.43±7.01	15.83±3.96
AutoInt+VFL	81.65±1.52	78.02±1.17	79.07±1.42	47.98±1.51
AutoInt+FairVFL	76.19±0.99	76.86±0.85	50.53±4.48	15.22±2.93

Results on ADULT.

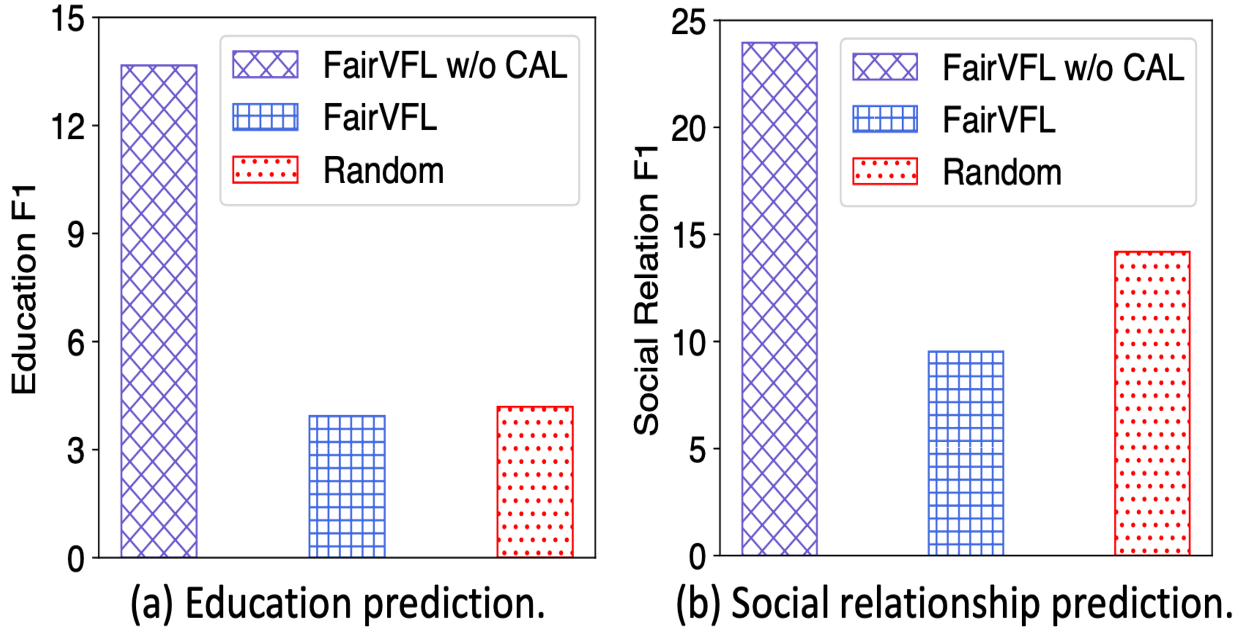
	News Recommendation		Model Fairness	
	AUC	nDCG@10	Gender F1	Age F1
NAML	64.04±0.13	30.80±0.13	70.05±0.21	20.01±3.08
NAML+FairGo	60.73±0.25	28.30±0.16	54.08±5.97	15.93±1.65
NAML+FairSM	60.59±0.19	28.15±0.16	54.13±5.38	15.74±1.52
NAML+FairRec	60.69±0.22	28.21±0.17	54.47±2.79	15.67±1.90
NAML+VFL	63.93±0.45	30.75±0.45	69.72±0.48	20.09±0.86
NAML+FairVFL	60.41±0.18	27.95±0.18	53.38±4.40	15.55±1.41
LSTUR	64.68±0.33	30.97±0.20	70.45±0.37	20.00±0.39
LSTUR+FairGo	61.03±0.24	28.31±0.15	53.57±3.88	15.74±1.22
LSTUR+FairSM	61.11±0.54	28.33±0.34	53.16±4.75	15.24±1.84
LSTUR+FairRec	60.99±0.78	28.31±0.47	53.36±1.37	15.65±0.93
LSTUR+VFL	64.39±0.32	30.85±0.19	70.07±0.37	19.92±1.63
LSTUR+FairVFL	60.98±0.28	28.25±0.36	53.51±3.41	15.23±0.94
NRMS	64.24±0.18	30.78±0.11	70.25±0.24	21.07±0.81
NRMS+FairGo	61.49±0.37	28.83±0.31	53.81±2.94	16.45±1.49
NRMS+FairSM	61.78±0.31	28.95±0.22	54.10±2.42	15.91±1.99
NRMS+FairRec	61.44±0.16	28.76±0.21	53.60±2.84	16.26±2.39
NRMS+VFL	64.38±0.13	30.93±0.11	70.67±0.23	21.41±0.66
NRMS+FairVFL	61.43±0.13	28.81±0.08	53.33±2.35	15.98±1.94

Results on NEWS.

Ablation Study



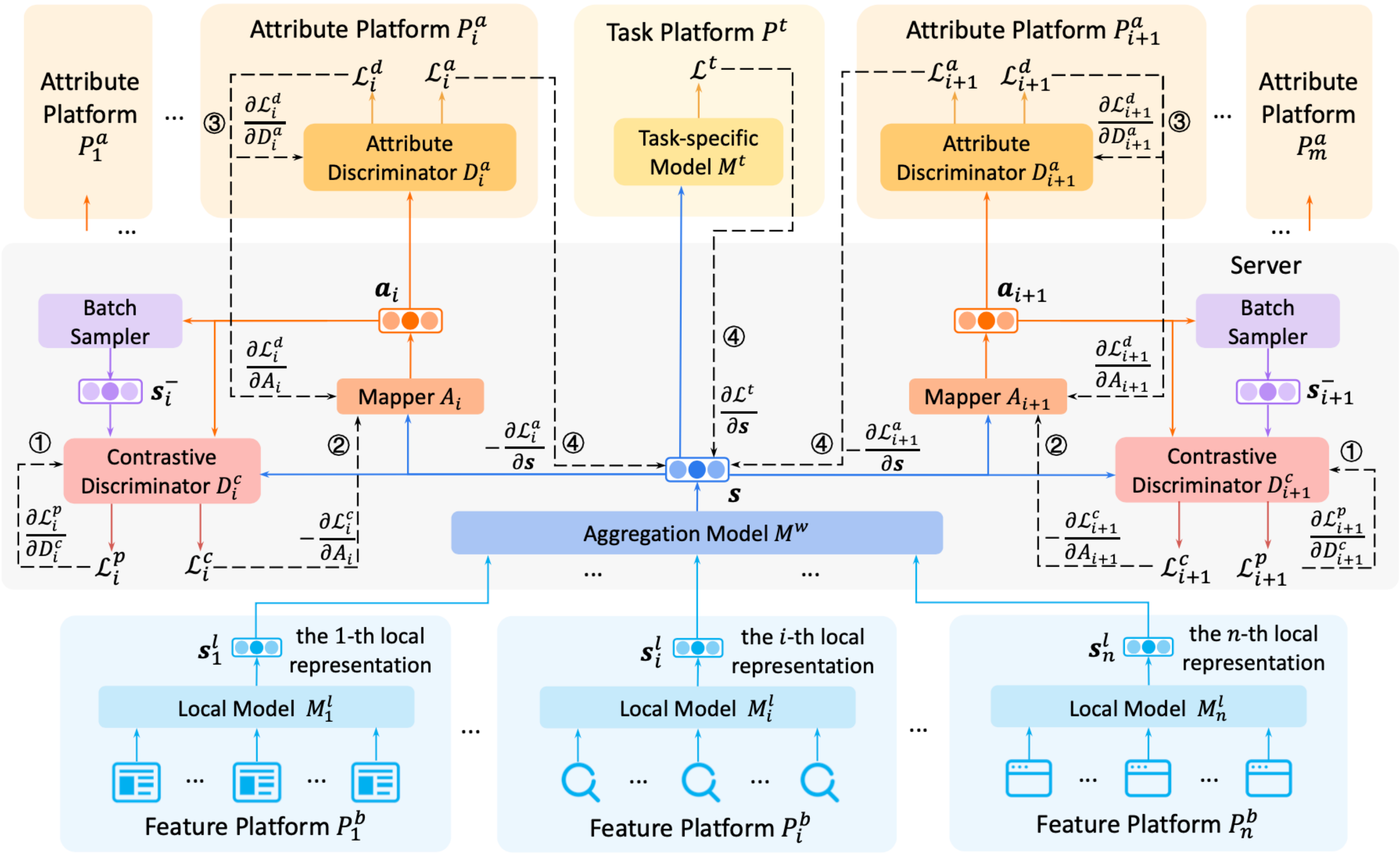
Ablation study on adversarial learning.



Ablation study on contrastive adversarial learning.

Conclusion

- Propose a fair vertical federated learning framework which can improve the fairness of VFL models
- Contrastive adversarial learning for privacy protection in fair VFL.



Tao Qi

taoqi.qt@gmail.com