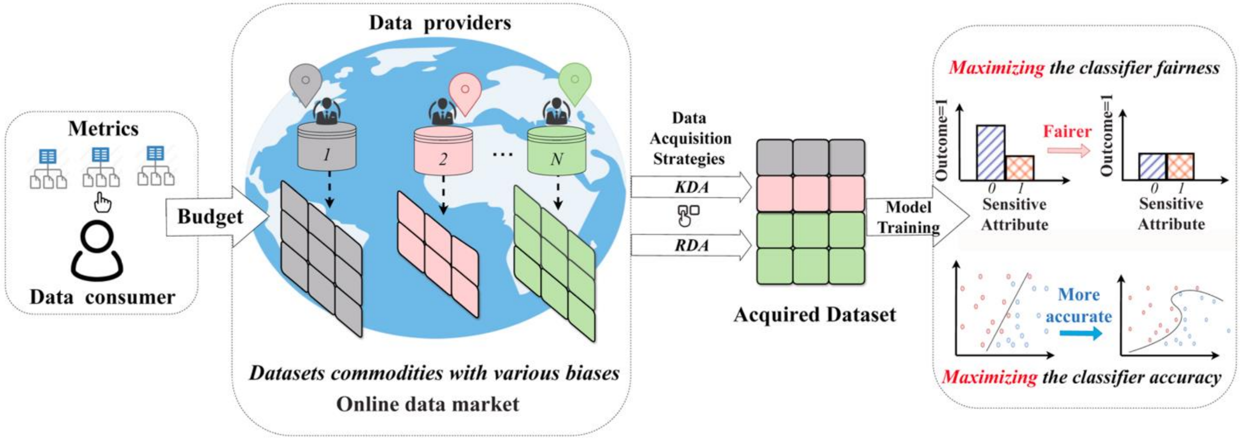
Unbiased Data Acquisition

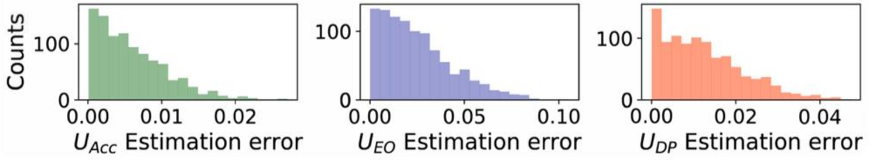
Core Issue	How can we promote fair AI model training in opaque data markets with heterogeneous dataset biases?

To address **opaque and heterogeneous** biases in multi-source data from online markets, a fairness-aware data acquisition framework (FAIRDA) is proposed, utilizing explore-exploit strategies to optimize data selection, maximizing **accuracy and fairness** within **a limited budget** for AI model training

✓KDA: Minimum sample size to explore dataset bias + NLKP instantiation



✓ RDA:Bernoulli bandit problem +dynamic interactions with fairness-aware reward



Method	Budget 1K			Budget 5K			Budget 10K		
	Acc.	EO	DP	Acc.	EO	DP	Acc.	EO	DP
Uniform – Budget	0.604	0.179	0.215	0.741	0.102	0.051	0.770	0.050	0.027
Uniform – Amount	0.602	0.189	0.377	0.751	0.076	0.044	0.772	0.054	0.030
No – fair – KDA	0.727	0.133	0.123	0.771	0.035	0.019	0.782	0.070	0.026
No – fair – RDA	0.695	0.250	0.314	0.775	0.027	0.027	**0.784**	0.055	0.027
KDA	0.728	**0.030**	**0.023**	0.774	**0.020**	**0.011**	0.776	**0.012**	**0.006**
RDA	**0.759**	0.094	0.070	**0.777**	0.026	0.018	0.778	0.022	0.008

Method	Dependence Degree 10%			Dependence Degree 30%			Dependence Degree 50%		
	Acc.	EO	DP	Acc.	EO	DP	Acc.	EO	DP
Uniform – Budget	0.771	0.060	0.020	0.765	0.073	0.116	0.757	0.212	0.075
Uniform – Amount	0.768	0.065	0.033	0.778	0.051	0.019	0.758	0.101	0.041
No – fair – KDA	**0.787**	0.042	0.017	**0.784**	0.040	0.023	0.767	0.084	0.049
No – fair – RDA	0.777	0.076	0.035	0.783	0.075	0.026	0.781	0.043	0.024
KDA	0.775	**0.013**	**0.014**	0.776	0.037	0.017	0.773	0.060	0.024
RDA	0.779	0.017	0.016	0.776	**0.035**	**0.015**	**0.786**	**0.022**	**0.013**

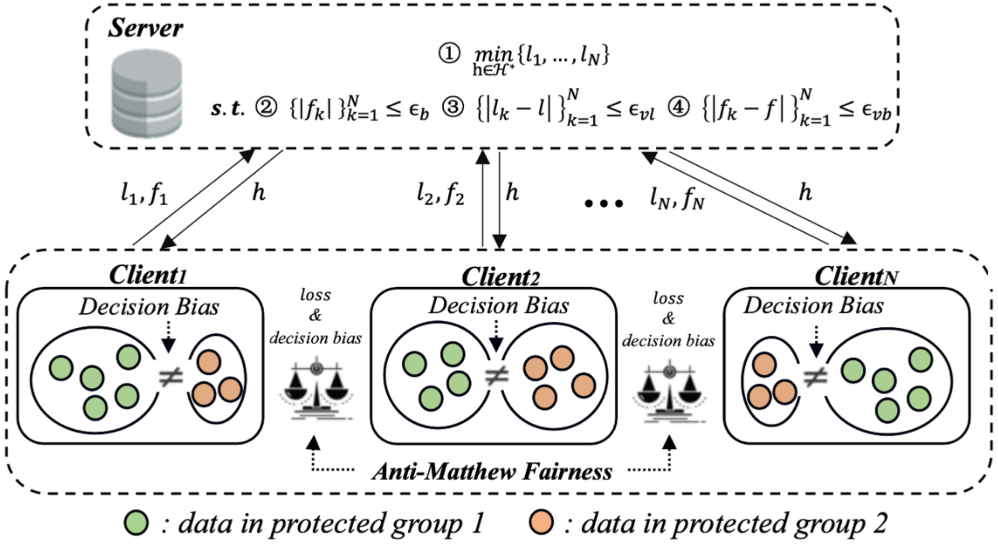
Statistical Inference+Combinatorial Optimization, Multi-Armed Bandit

Surviving in Diverse Biases: Unbiased Dataset Acquisition in Online Data Market for Fair Model Training. AIES 24.

Fairness in Collaborative AI Model Training

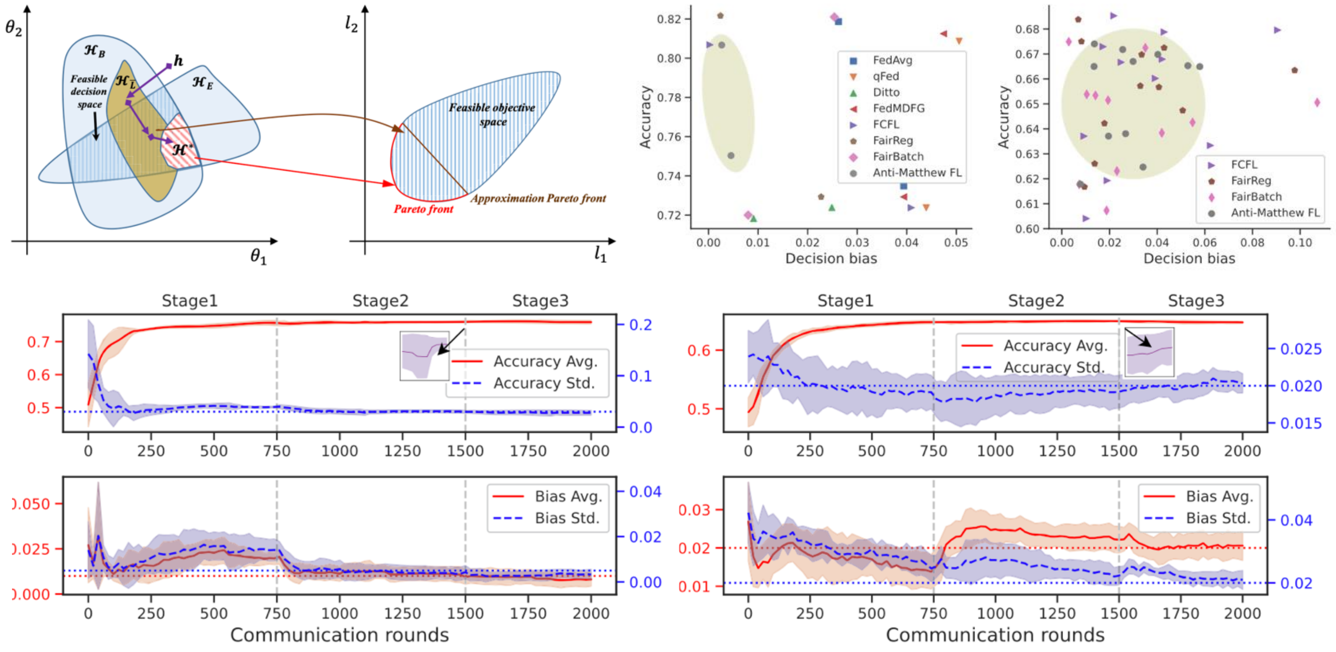
Core Issue	How can we mitigate the Matthew effect in federated learning arising from the heterogeneity of data resources?

To mitigate the Matthew effect that **exacerbates performance disparity in data-heterogeneous scenarios**, a fairness-aware Federated Learning framework, anti-MatthewFL, is proposed, achieving a **high-performance** global model while **narrowing performance gaps** among clients



Multi-constrained multi-objectives optimization

Three-stages optimization paths to achieve Pareto stationary

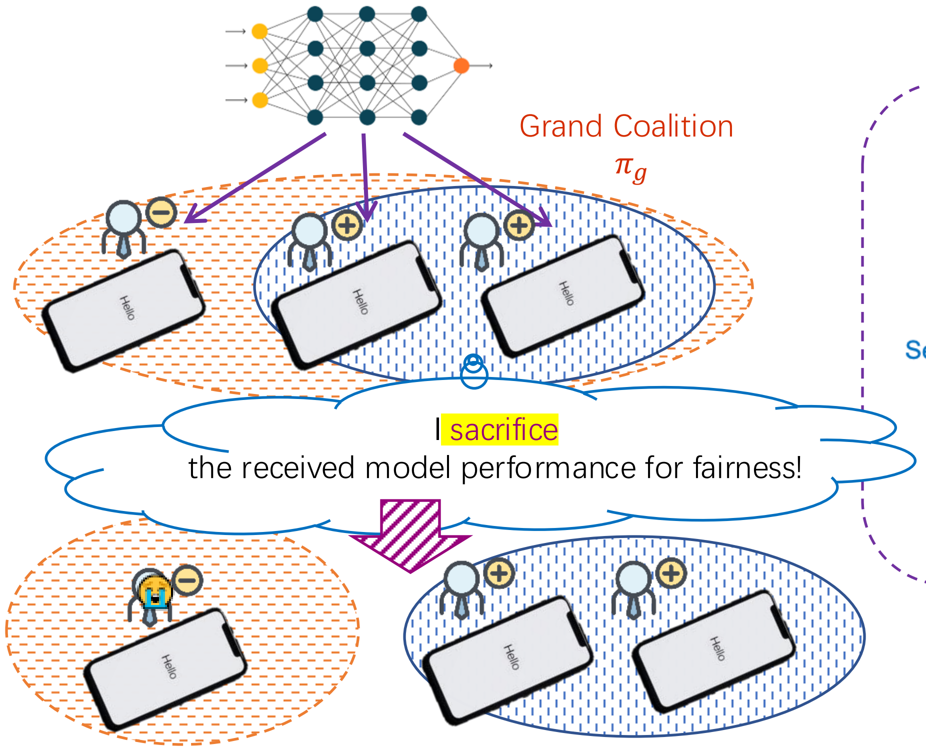


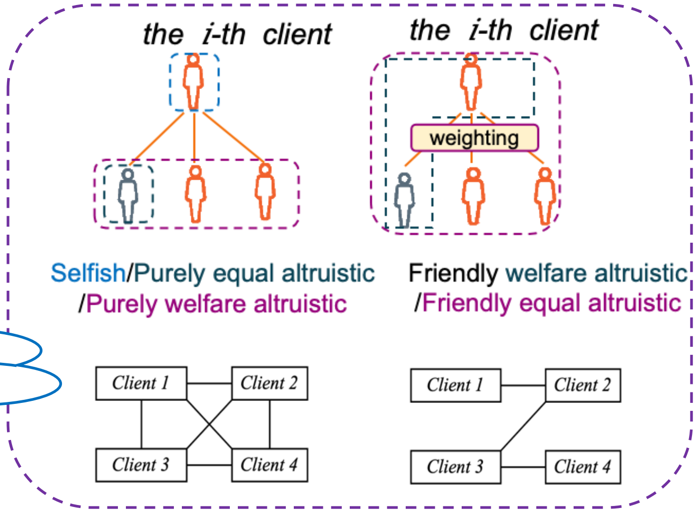
Multi-Objectives Optimization

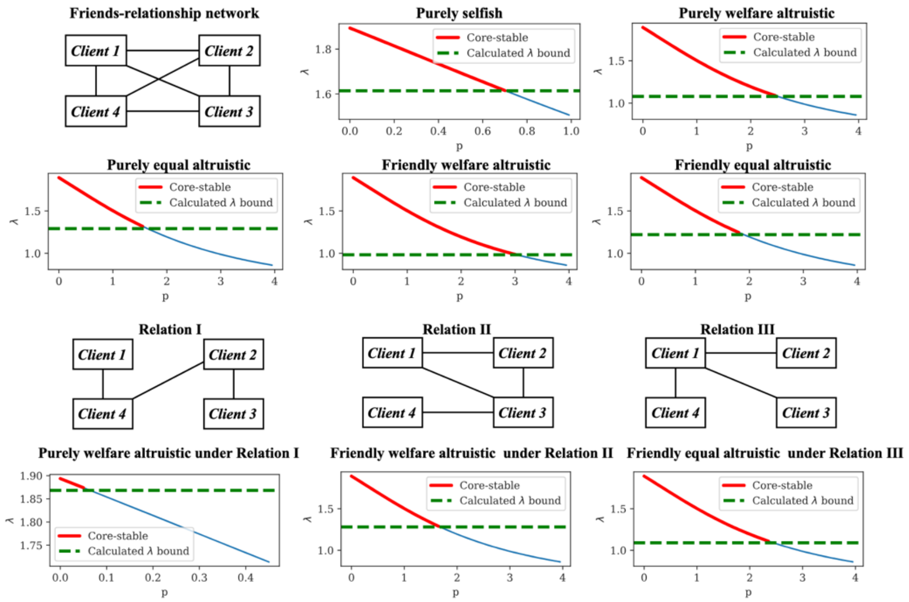
Fairness in Collaborative AI Model Training

Core Issue	Lack of quantitative analysis on the trade-off between egalitarian fairness and collaborative stability in federated learning

To clarify how egalitarian fairness affects the stability of FLs, a theoretical analyzing framework for quantifying **optimal egalitarian fairness bounds that a core-stable FL can obtain** is proposed, unveiling the impact of different **client altruistic behaviors** and their **underlying social relationships**







Setting suitable fairness levels, ensuring both feasibility and positive societal impact

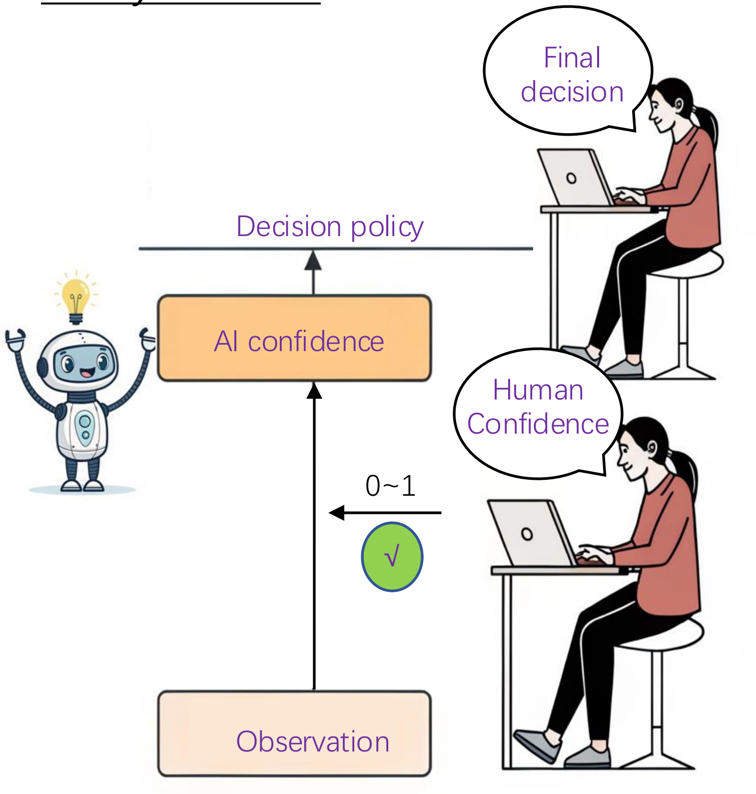
Cooperative Game Theory

Does Egalitarian Fairness Lead to Instability? The Fairness Bounds in Stable Federated Learning Under Altruistic Behaviors. NeurIPS 24.

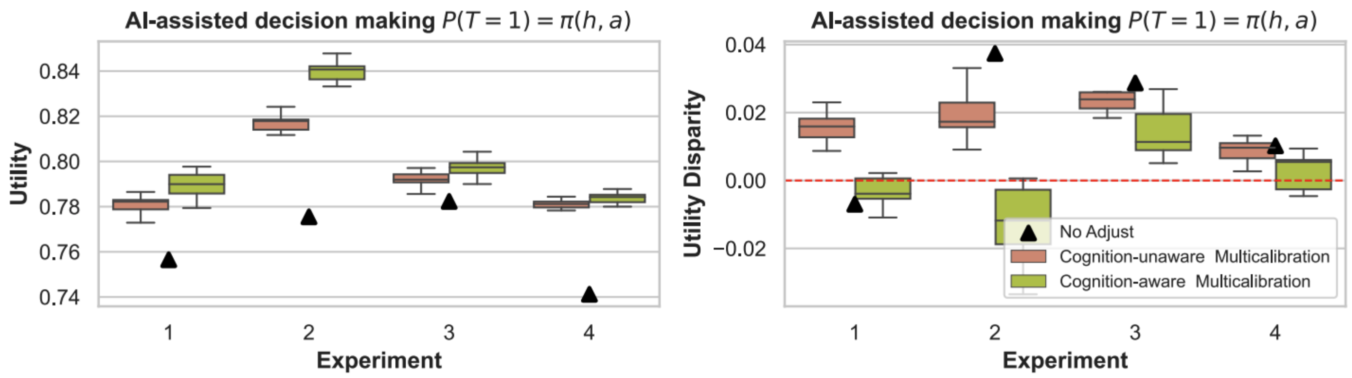
Fairness in Human-AI Collaboration

Core Issue	How does an AI-assisted system provide fair decision utility for experts with heterogeneous cognitive capacities?

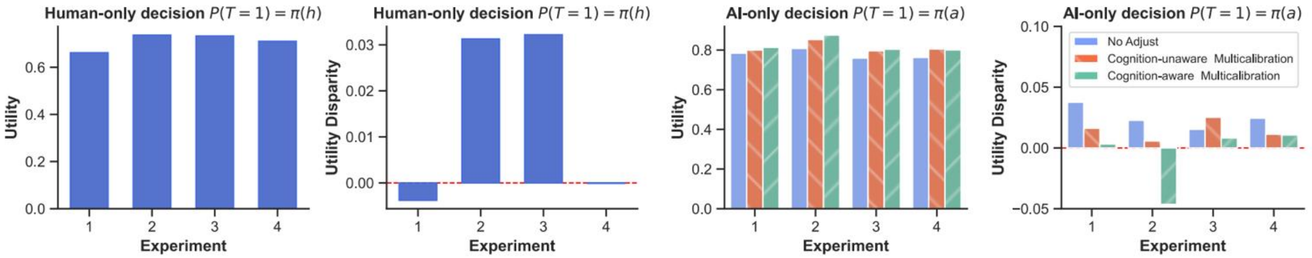
To ensure **fair utility**, an **interpretable** fairness-aware multicalibration-based AI confidence adjustment is proposed for scenarios involving human decision-makers with heterogeneous cognitive capacities, improving utility fairness across human decision-maker groups **without** sacrificing overall utility



Inter-group-alignment → Cognition-Aware AI Confidence Multicalibration



Mitigating human decision utility disparity by reducing AI-decision disparity or by offsetting with the opposite sign



Statistical Modeling+Calibration+Utility Maximization