AUTOMATED DISCOVERY OF TRADE-OFF BETWEEN ACCURACY, PRIVACY AND FAIRNESS IN MACHINE LEARNING MODELS

Bogdan Ficiu, Neil D. Lawrence, Andrei Paleyes

Department of Computer Science and Technology, University of Cambridge

BIAS workshop, ECML 2023

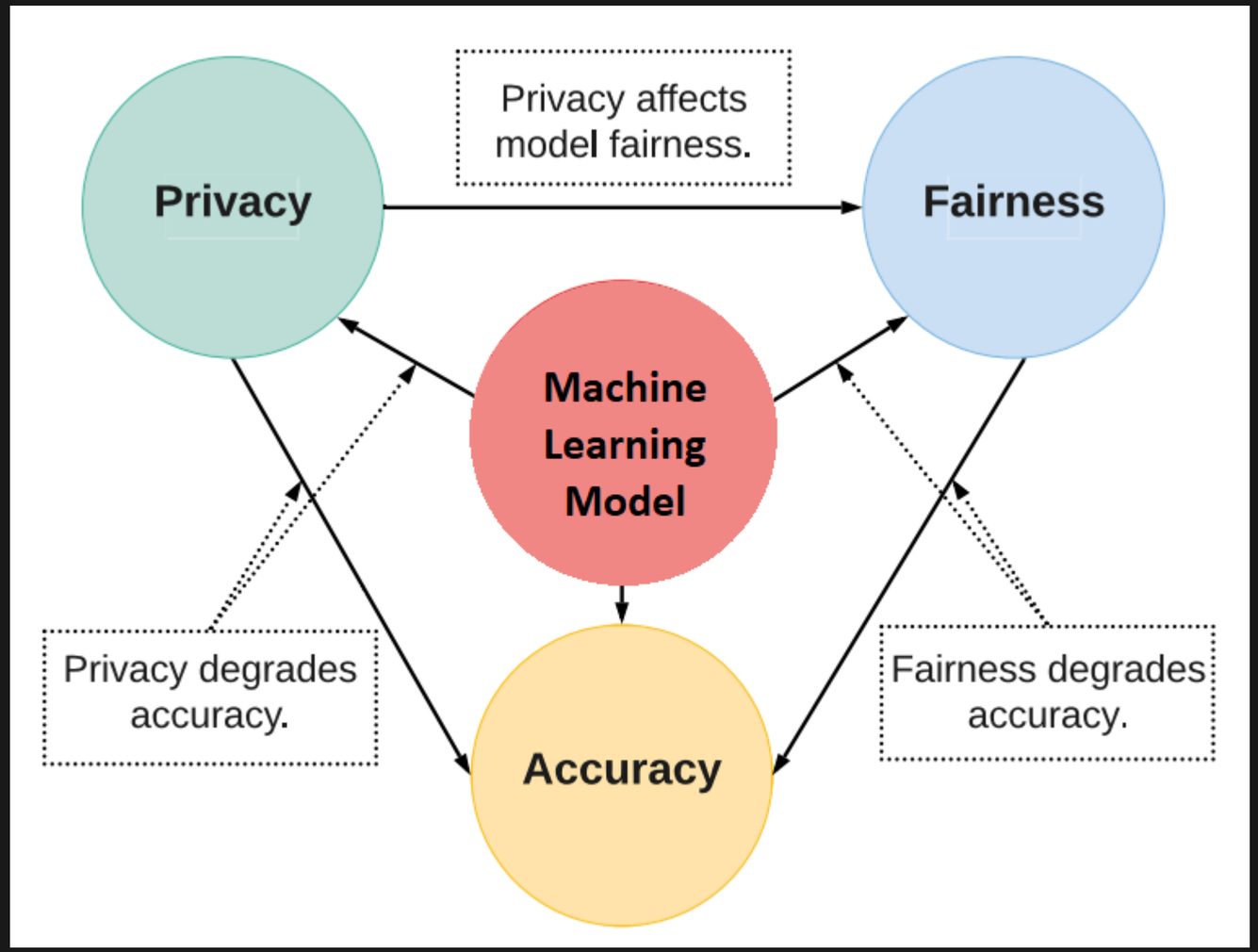
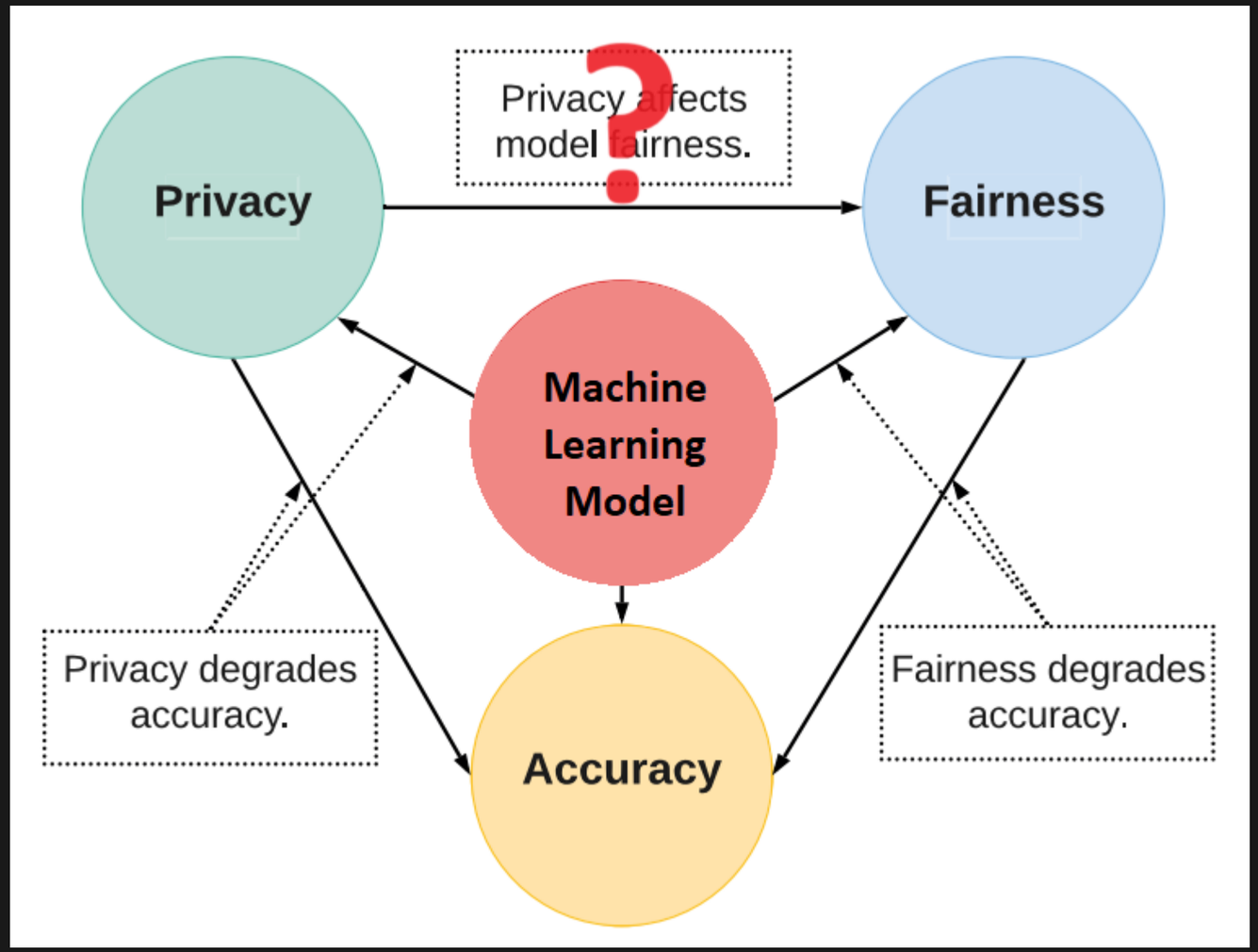


Image credit: Zhang et al. 2021 "Balancing Learning Model Privacy, Fairness, and Accuracy With Early Stopping Criteria"



But... (Local) Differential Privacy has NO Disparate Impact on Fairness!

	ε	
Adult	0.1	
	1	
	10	
	100	
Dutch	0.1	
	1	
	10	
	100	

Dataset	Model
Bank (age)	Unconstrained
	Fair ($\delta = 0.1$)
	Fair ($\delta = 0.01$)
	Fair ($\delta = 0.001$)
COMPAS (race)	Unconstrained
	Fair ($\delta = 0.1$)
	Fair ($\delta = 0.01$)
	Fair ($\delta = 0.001$)
COMPAS (gender)	Unconstrained
	Fair ($\delta = 0.1$)
	Fair ($\delta = 0.01$)
	Fair ($\delta = 0.001$)
Law (race)	Unconstrained
	Fair ($\delta = 0.1$)
	Fair ($\delta = 0.01$)
	Fair ($\delta = 0.001$)
Law (gender)	Unconstrained
	Fair ($\delta = 0.1$)
	Fair ($\delta = 0.01$)
	Fair ($\delta = 0.001$)

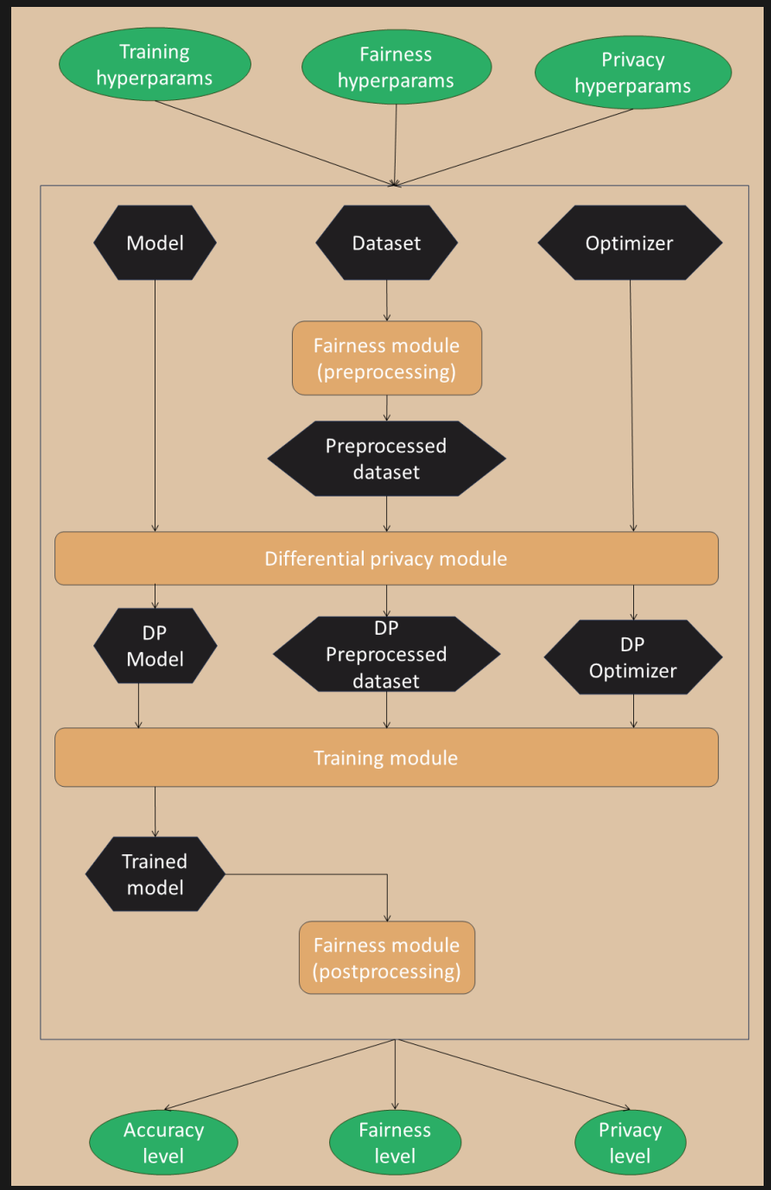
Sources: Xu et al. 2019 "Achieving Differential Privacy and Fairness in Logistic Regression", Chang and Shokri 2020 "On the Privacy Risks of Algorithmic Fairness"

Find a set of Pareto optimal points *x* for:

$$f(x) = \begin{cases} y_{utility} \\ y_{fairness} \\ y_{privacy} \end{cases}$$

$$x = (\theta_{utility}, \theta_{fairness}, \theta_{privacy})$$

f(x): PFairDP


Training hyperparams
Fairness hyperparams
Privacy hyperparams
Model
Dataset
Optimizer
Fairness module (preprocessing)
Preprocessed dataset
Differential privacy module
DP Model
DP Preprocessed dataset
DP Optimizer
Training module
Trained model
Fairness module (postprocessing)
Accuracy level
Fairness level
Privacy level

Discovery: Bayesian optimisation

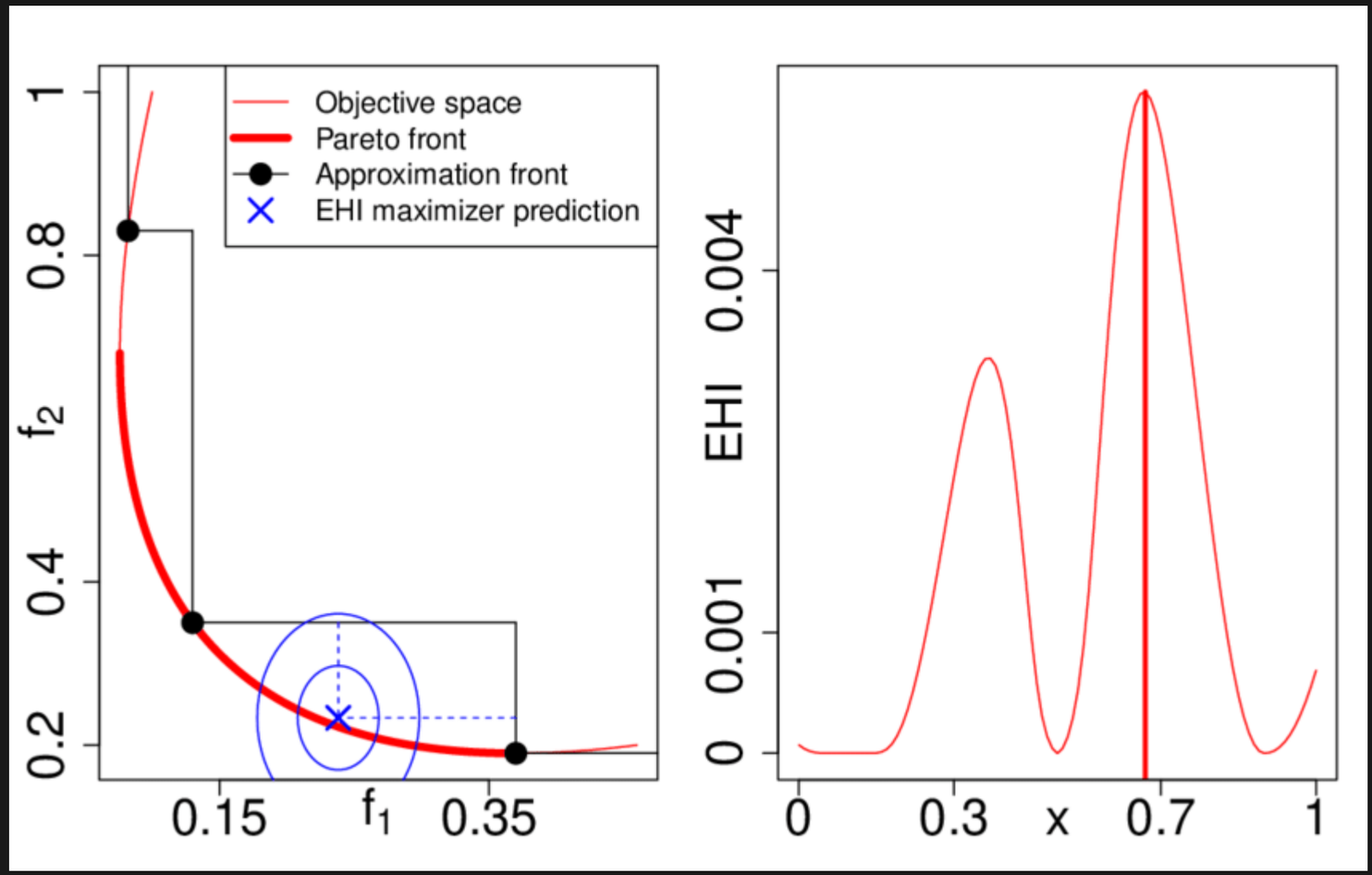


Image source: Gaudrie 2019 "High-Dimensional Bayesian Multi-Objective Optimization"

CONTRIBUTIONS

- Description of full procedure
- Better than grid search a.k.a. constraint setting
- Replication of existing approaches

But wait! What about...

...visualisation of the fronts?

...choice of metrics of utility, fairness and privacy?

...computational efficiency?

...!?!?

But wait! What about...

...visualisation of the fronts?

...choice of metrics of utility, fairness and privacy?

...computational efficiency?

...!?!?

Let's chat at the poster!

✉ ap2169@cl.cam.ac.uk

Thanks!