Long-term Fairness with Selective Labels

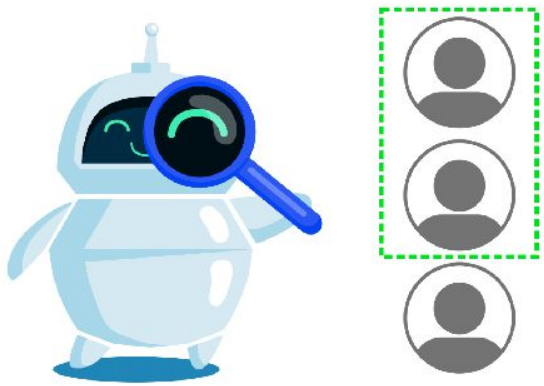
Giovani Valdrighi[1], Isabel Valera[2], Marcos Medeiros Raimundo[1]
[1]Universidade Estadual de Campinas 🇧🇷
[2]Saarland University

July 7th, LatinX in AI @ ICML 2026

UNIVERSITÄT DES SAARLANDES

Fairness in dynamic decision-making

- Accept/reject (A_t) based on features (X_t) and a sensitive attribute (Z)



- a label (Y_t) and decision (A_t) will define rewards and update features (X_{t+1})

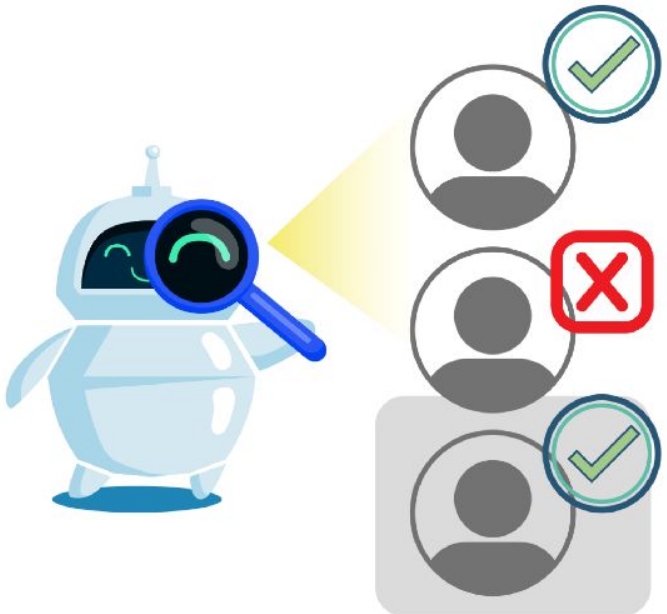
- The objective is to maximize cumulative reward while satisfying a fairness constraint at **every timestamp t**:

$$|\Delta_t| := |\text{avg. utility for } z = 1 - \text{avg. utility for } z = 0|$$

$$|\Delta_t| \leq \omega$$

Fairness metric	Utility
Qualification parity	qualification rate
Accuracy parity	accuracy rate
Equality of opp.	true positive rate

Selective labels



The decision-maker uses a **predictor** to perform label imputation:

$$\tilde{Y}_t = \begin{cases} Y_t \text{ if accepted} \\ \hat{Y}_t \text{ if rejected} \end{cases}$$

The **observed disparity** ($\tilde{\Delta}_t$) is calculated after label imputation.

Labels are only observed for the **accepted population**, however, **they are necessary to evaluate fairness metrics!**

Constraining true disparity

How to use the observed disparity $\tilde{\Delta}_t$ to satisfy a constraint on true disparity $|\Delta_t| \leq \omega$?

Decomposition of observed disparity

- We present a decomposition of the observed disparity:

$$\tilde{\Delta}_t = \Delta_t + \textbf{distortion}$$

- based on the following group-wise quantities:

$$\underset{\textbf{rejection rate}}{r_t^i := P(A_t = 0|Z = i)} \quad \underset{\textbf{observed qualification}}{\tilde{\alpha}_t^i := P(\tilde{Y}_t = 1 \mid Z = i)}$$

$$\underset{\textbf{error rate on rejected population}}{\epsilon_t^i := \mathbb{E}\left[\hat{Y}_t - Y_t | Z = i, A_t = 0\right]}$$

Decomposition of observed disparity

$$\tilde{\Delta}_t = \Delta_t + \mathbf{distortion}$$

Fairness metric	distortion
Qualification parity	$r_t^1 \epsilon_t^1 - r_t^0 \epsilon_t^0$
Accuracy parity	$-(r_t^1 \epsilon_t^1 - r_t^0 \epsilon_t^0)$
Equality of opp.	$-\Delta_t \left(\frac{r_t^1 \epsilon_t^1}{\tilde{\alpha}_t^1} \right) + \mu_t^0 \left(\frac{r_t^0 \epsilon_t^0}{\tilde{\alpha}_t^0} - \frac{r_t^1 \epsilon_t^1}{\tilde{\alpha}_t^1} \right)$

Takeaway: to ensure that $\tilde{\Delta}_t \approx \Delta_t$ ($\mathbf{distortion} \approx 0$), it is necessary to balance rejection and error rates (e.g., by reducing the rejection of the group with higher error).

Conditions to constrain true disparity

- However, the error rate is not observed under selective labels!

$$\epsilon_t^i := \mathbb{E}\left[\hat{Y}_t - Y_t | Z = i, A_t = 0\right]$$
error rate on rejected population

- We leverage **domain adaptation** theory to use the accepted population and its distance to the rejected one to bound the error:

$$\epsilon_t^i \leq \bar{\epsilon}_t^i := \sum_{j=1}^{N^i} \epsilon(x_j, i) \frac{D_R^i(x)}{D_A^i(x)} + \sqrt{d_2(D_R^i || D_A^i)} / \sqrt{N^i}$$
weighted estimate **Renyi divergence**

Conditions to constrain true disparity

- However, the error rate is not observed under selective labels!

$$\epsilon_t^i := \mathbb{E}\left[\hat{Y}_t - Y_t | Z = i, A_t = 0\right]$$

error rate on rejected population

- We leverage **domain adaptation** theory to use the accepted population and its distance to the rejected one to bound the error:

$$\epsilon_t^i \leq \bar{\epsilon}_t^i \approx \sum_{j=1}^{N^i} \epsilon(x_j, i) \frac{D_R^i(x)}{D_A^i(x)} + \sqrt{d_2(D_R^i || D_A^i)} / \sqrt{N^i}$$

weighted estimate — **Renyi divergence**

$$\textbf{distortion} \leq \overline{\textbf{distortion}}$$

Conditions to constrain true disparity

- **Main result (Theo. 3.6)** We have that $|\Delta_t| \leq \omega$ if:

$$|\tilde{\Delta}_t| \leq \omega/2 \qquad \text{and} \qquad \overline{\mathbf{distortion}} \leq \omega/2$$

which are **fully observable**!

Conditions to constrain true disparity

- **Main result (Theo. 3.6)** We have that $|\Delta_t| \leq \omega$ if:

$$|\tilde{\Delta}_t| \leq \omega/2 \qquad \text{and} \qquad \overline{\textbf{distortion}} \leq \omega/2$$

low error

which are **fully observable**!

Conditions to constrain true disparity

- **Main result (Theo. 3.6)** We have that $|\Delta_t| \leq \omega$ if:

$$|\tilde{\Delta}_t| \leq \omega/2 \qquad \text{and} \qquad \overline{\textbf{distortion}} \leq \omega/2$$

low error — divergence accept/reject pop.

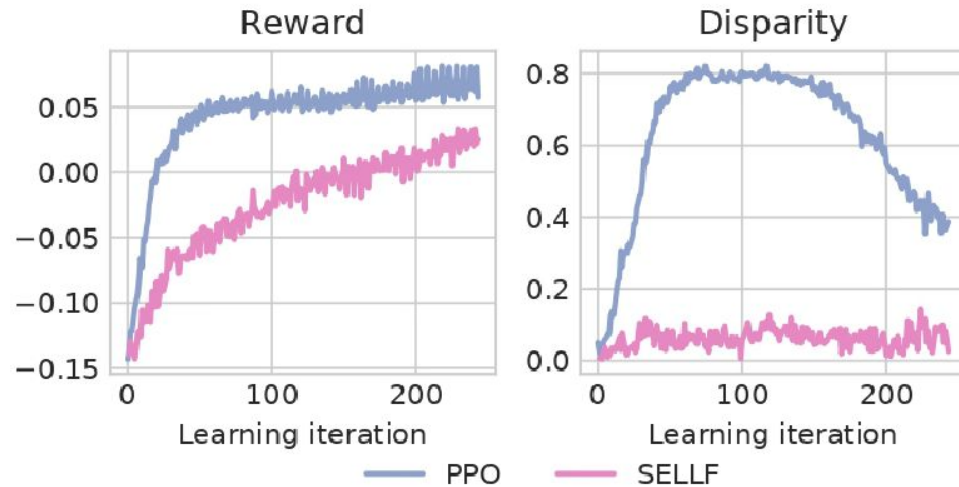
which are **fully observable**!

Proposed algorithm (SELLF)

We adapt PPO to learn a **policy** and a **predictor** simultaneously.

Algorithm

SELLF is based on three key ideas:

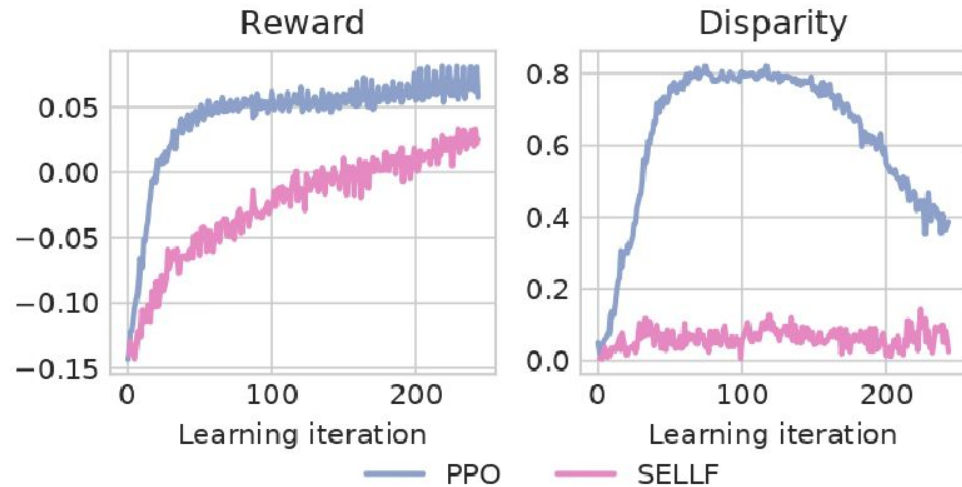


Algorithm

SELLF is based on three key ideas:

Constrain $|\tilde{\Delta}_t|$:

- Advantage penalization with **observed-disparity** (Yu et al., 2022)



Algorithm

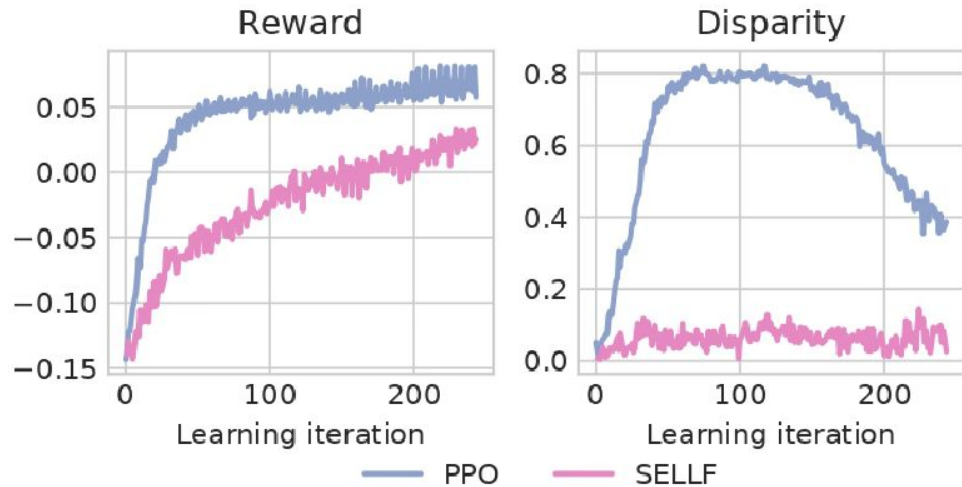
SELLF is based on three key ideas:

Constrain $|\tilde{\Delta}_t|$:

- Advantage penalization with **observed-disparity** (Yu et al., 2022)

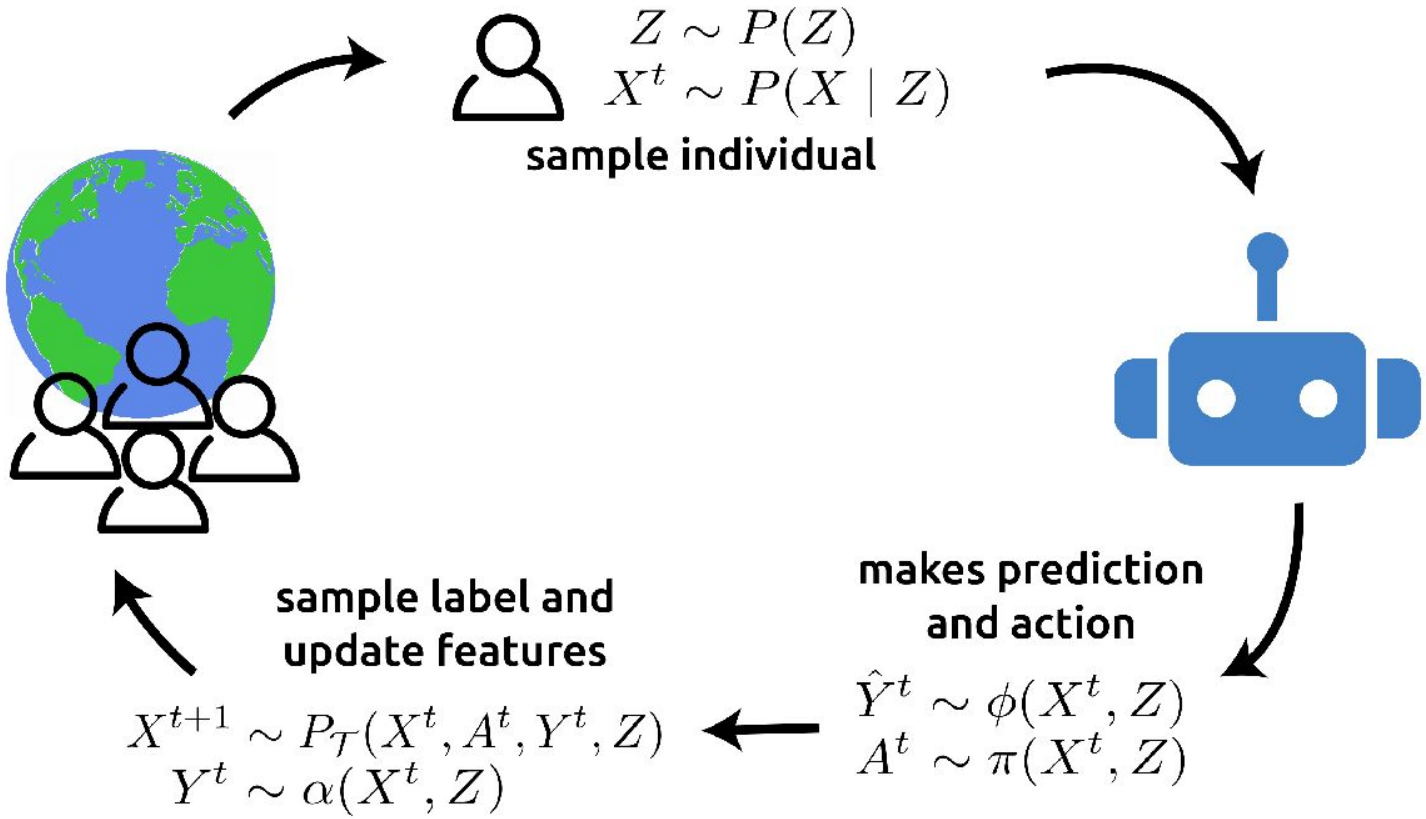
Approximate $|\Delta_t|$ and $|\tilde{\Delta}_t|$:

- Optimization of the predictor using accepted samples and **IPW** on the loss
- Policy regularization with **Renyi divergence** accepted/rejected populations

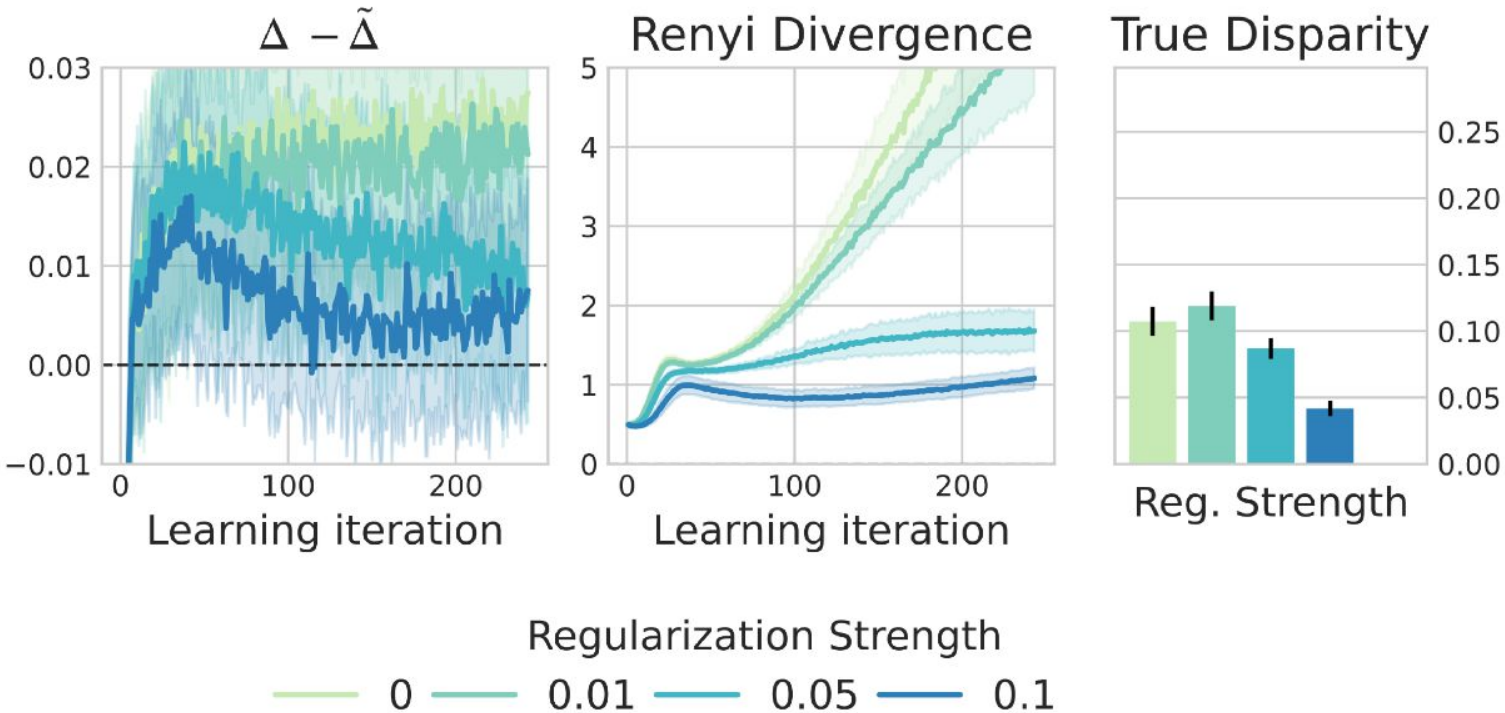


Experiments

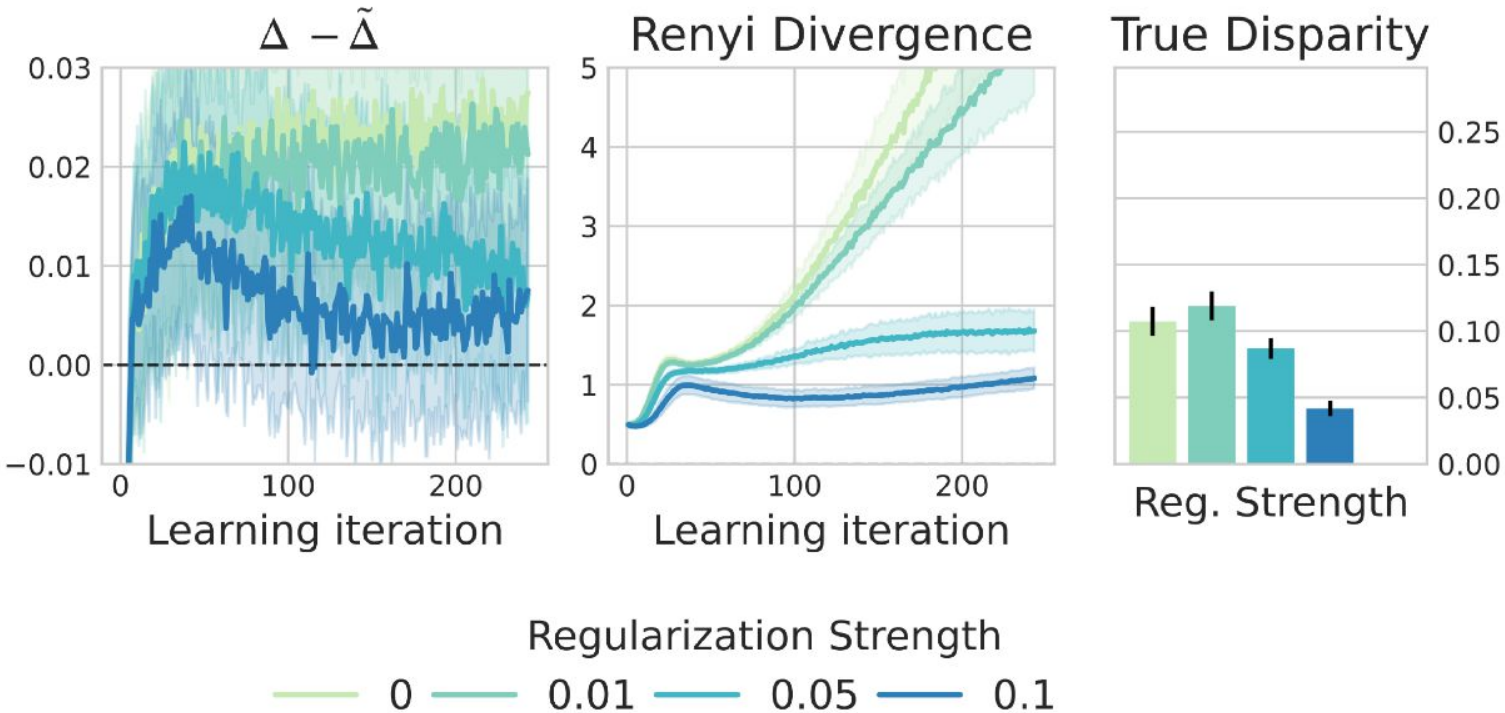
Experiments with 3 semi-synthetic environments and 5 baselines that ignore the partially observed labels. **Check the paper to see the full results!**



Is the Renyi regularization useful?

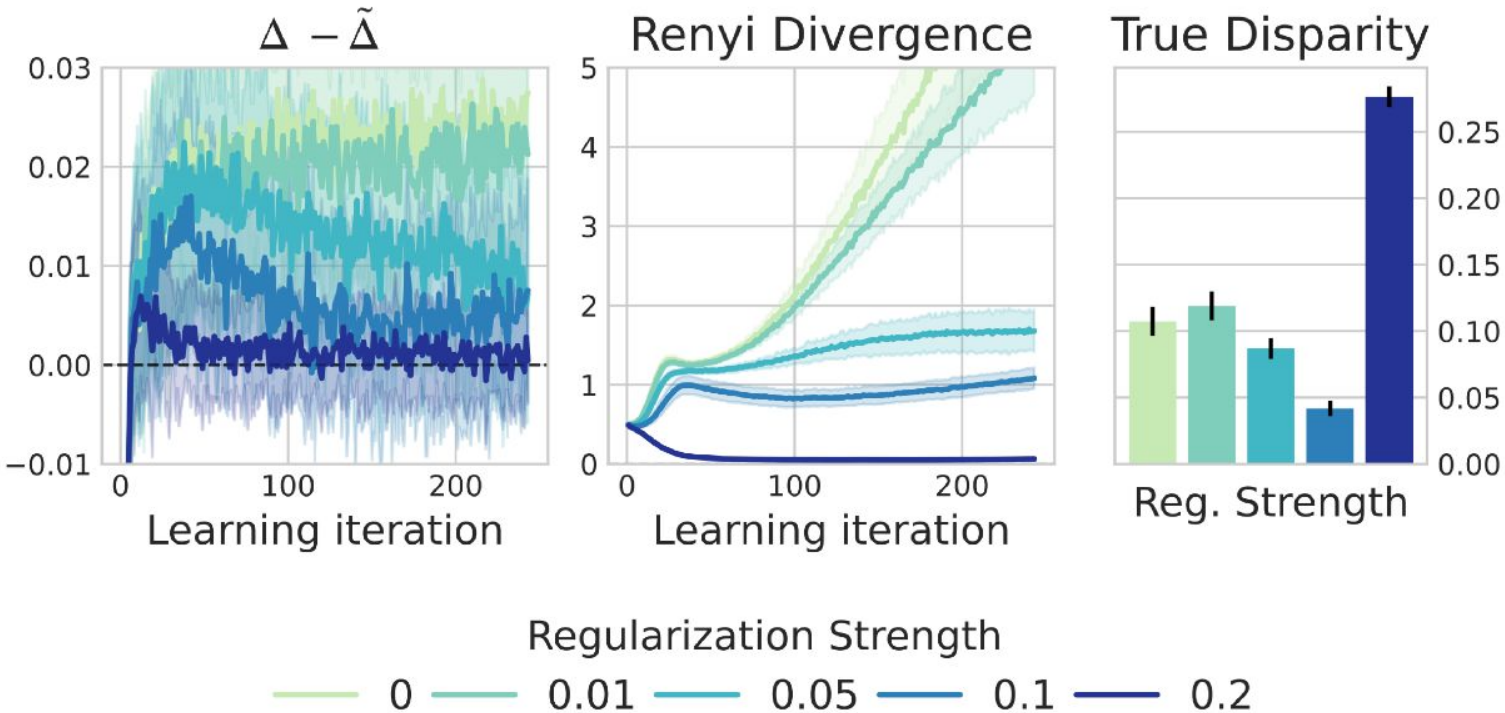


Is the Renyi regularization useful?



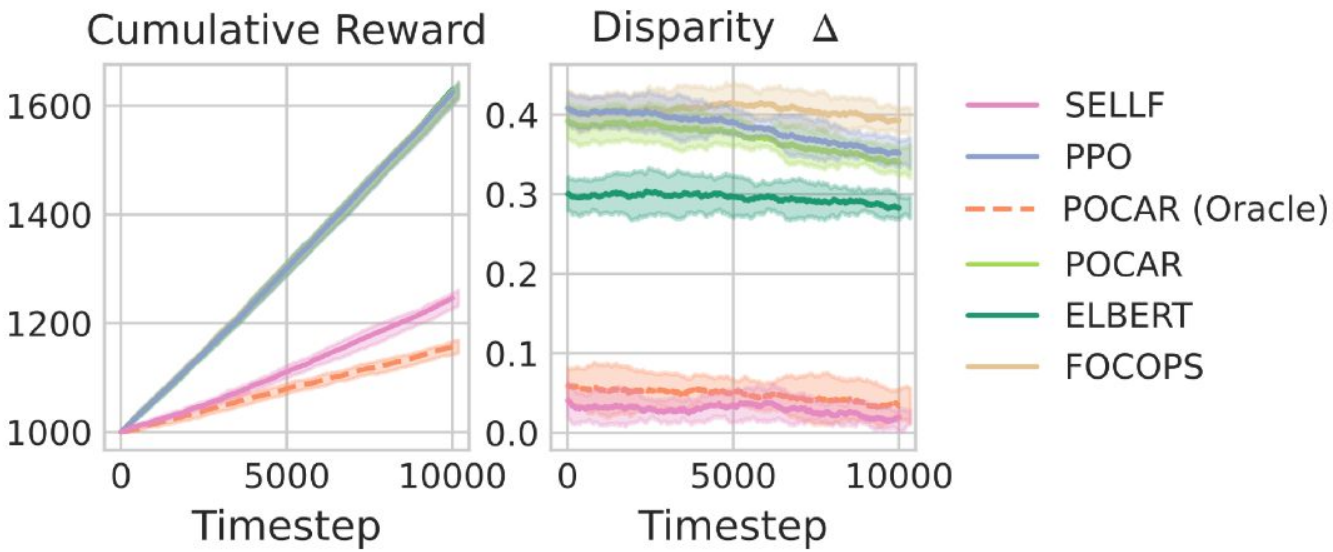
- Higher strength $\rightarrow$ lower gap

Is the Renyi regularization useful?



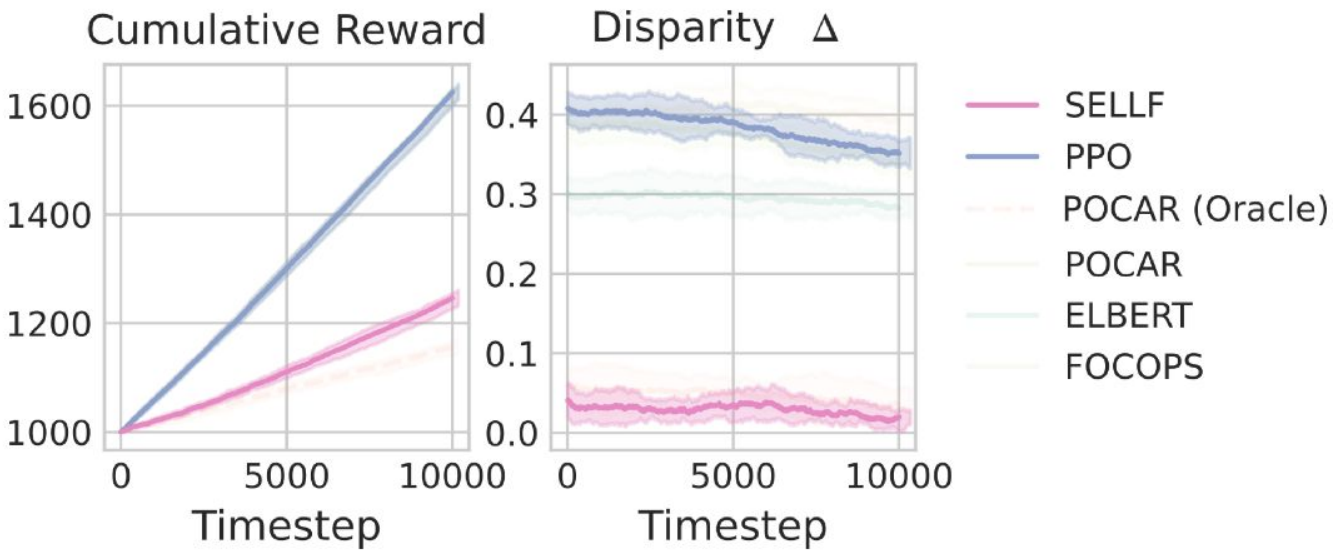
- Higher strength $\rightarrow$ lower gap
- Very high strength $\rightarrow$ breaks optimization

Is SELLF better than baselines?



Lending experiment with equality of opportunity

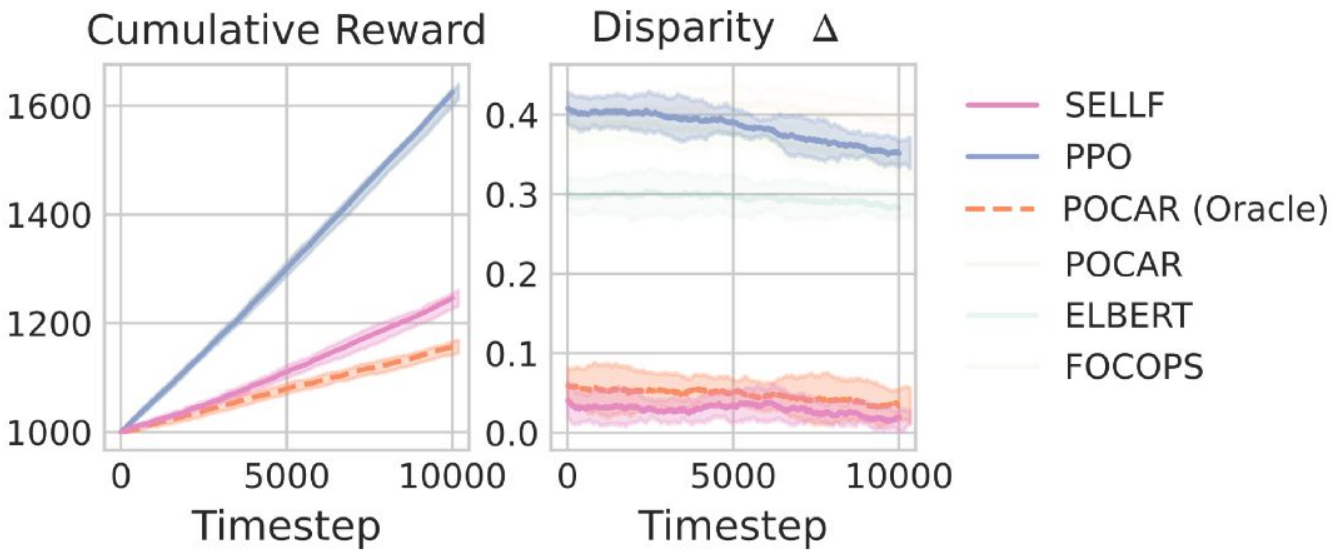
Is SELLF better than baselines?



Lending experiment with equality of opportunity

- PPO is highly unfair

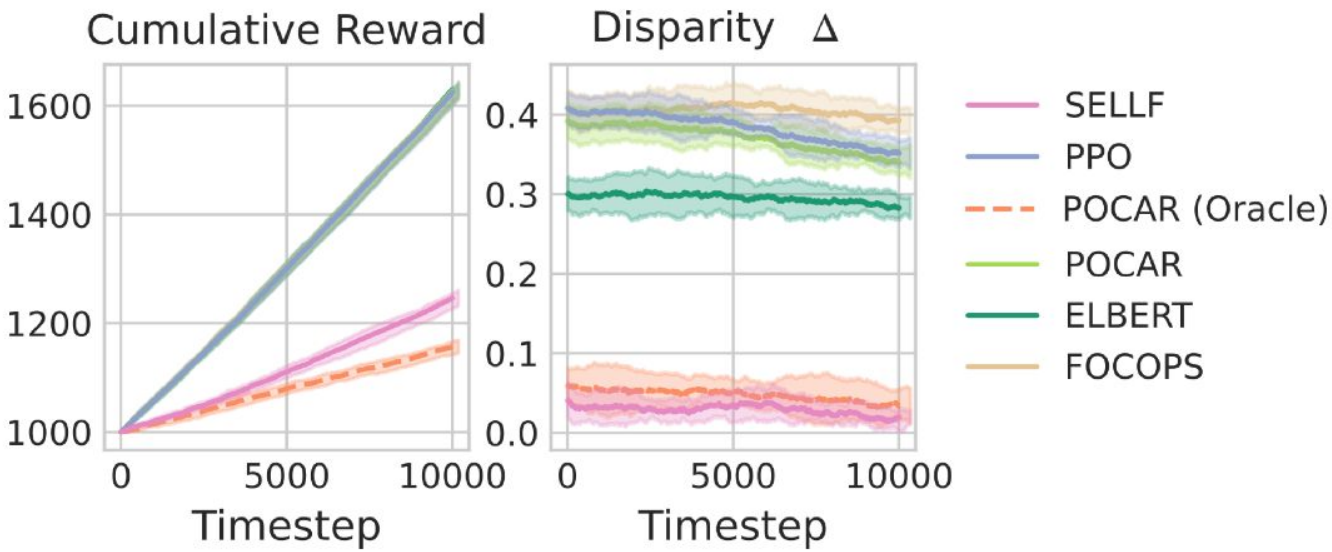
Is SELLF better than baselines?



Lending experiment with equality of opportunity

- PPO is highly unfair
- SELLF achieves a performance comparable with Oracle access to labels

Is SELLF better than baselines?



Lending experiment with equality of opportunity

- PPO is highly unfair
- SELLF achieves a performance comparable with Oracle access to labels
- Related methods fail to achieve fairness due to missing labels

Summary

In conclusion, our work shows that:

Summary

In conclusion, our work shows that:

1. **selective labels** are an important aspect to ensure fair decision-making

Summary

In conclusion, our work shows that:

1. **selective labels** are an important aspect to ensure fair decision-making
2. to ensure bounded disparity, it is necessary to consider the **group-wise prediction error and uncertainty of estimating it**

Summary

In conclusion, our work shows that:

1. **selective labels** are an important aspect to ensure fair decision-making
2. to ensure bounded disparity, it is necessary to consider the **group-wise prediction error and uncertainty of estimating it**
3. with the proposed algorithm, we can **train a policy to satisfy long-term fairness with selective labels.**

Thanks! Questions?

Contact: giovani.valdrighi@ic.unicamp.br

Check the paper!

H.IAAC recod.ai

UNICAMP

UNIVERSITÄT DES SAARLANDES