CausalPFN: Amortized Causal Effect Estimation via In-Context Learning

Vahid Balazadeh[*1], Hamid Kamkari[*2], Valentin Thomas[2], Jeremy Ma[2], Benson Li[1], Jesse C. Cresswell[2], Rahul G. Krishnan[1]

[1]University of Toronto, [2]Layer 6 AI, Toronto, Canada






UNIVERSITY OF TORONTO

VECTOR INSTITUTE

layer 6 AI at TD

NEURAL INFORMATION PROCESSING SYSTEMS

Motivation: Heterogeneous Causal Effects

- You're an online retailer deciding whether to offer a 10% discount
- You have historical observational data of the past items and customers

ID	Received Discount (T)	Bought Item (Y)	Customer Features (X_1)	Item Features (X_2)
001	✅ Yes	✅ Yes	Repeat buyer, low income	Mid-range headphones
002	❌ No	✅ Yes	High income, loyal	Premium laptop
003	✅ Yes	❌ No	New visitor, student	Budget smartphone

- Should we offer the new customer a discount?

100	?	?	Student, loyal	Bluetooth Speaker

Motivation: Heterogeneous Causal Effects

- Can we train a predictive model from the features to the outcome?

ID	Received Discount (T)	Customer Features (X_1)	Item Features (X_2)
001	✅ Yes	Repeat buyer, low income	Mid-range headphones
002	❌ No	High income, loyal	Premium laptop
003	✅ Yes	New visitor, student	Budget smartphone

Predictive Model

(Training)

Bought Item (Y)
✅ Yes
✅ Yes
❌ No

UNIVERSITY OF TORONTO

Motivation: Heterogeneous Causal Effects

- Can we train a predictive model from the features to the outcome?

ID	Received Discount (T)	Customer Features (X_1)	Item Features (X_2)
100	❌ No	Student, loyal	Bluetooth Speaker
100	✅ Yes	Student, loyal	Bluetooth Speaker

↓

Predictive Model

↓

(Inference)

Bought Item (Y)
❌ No
❌ No

This Work: Black-Box Heterogeneous Causal Effect Estimation

ID	Received Discount (T)	Bought Item (Y)	Customer Features (X_1)	Item Features (X_2)
001	✅ Yes	✅ Yes	Repeat buyer, low income	Mid-range headphones
002	❌ No	✅ Yes	High income, loyal	Premium laptop
003	✅ Yes	❌ No	New visitor, student	Budget smartphone

ID	Customer Features (X_1)	Item Features (X_2)
100	Student, loyal	Bluetooth Speaker

CausalPFN

ID	Will Buy Under No Discount (Y_0)	Will Buy Under Discount (Y_1)
100	❌ No	✅ Yes

Target Quantities

The goal is to estimate the conditional expected potential outcomes (**CEPOs**) from observational data $(\boldsymbol{X}, T, Y)$

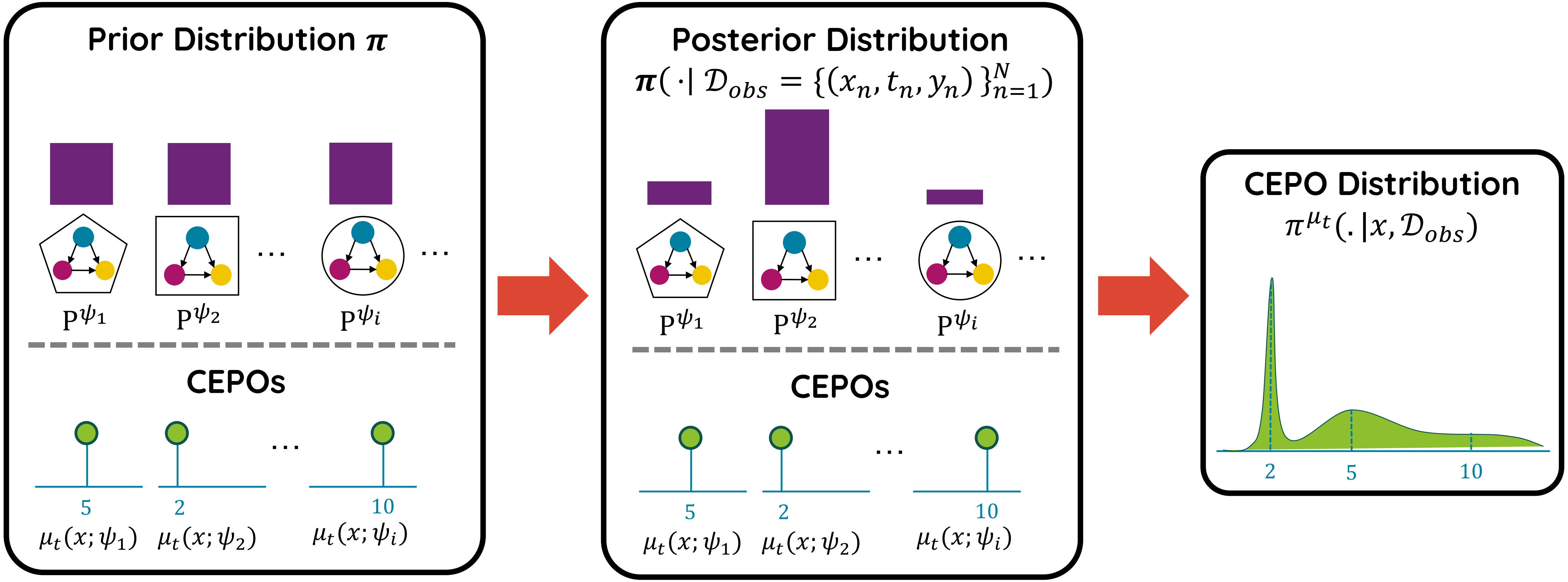
$$\mu_t\left(\mathbf{x}\right) := \mathbb{E}\left[Y_t | \mathbf{X} = \mathbf{x}\right]$$

$$\boldsymbol{\tau}(\boldsymbol{x}) = \boldsymbol{\mu}_{\mathbf{1}}(\boldsymbol{x}) - \boldsymbol{\mu}_{\mathbf{0}}(\boldsymbol{x})$$

Conditional Average Treatment Effect – CATE

UNIVERSITY OF TORONTO

A Formalism for Estimator Design



UNIVERSITY OF TORONTO

This Framework is Elegant but Practically Challenging

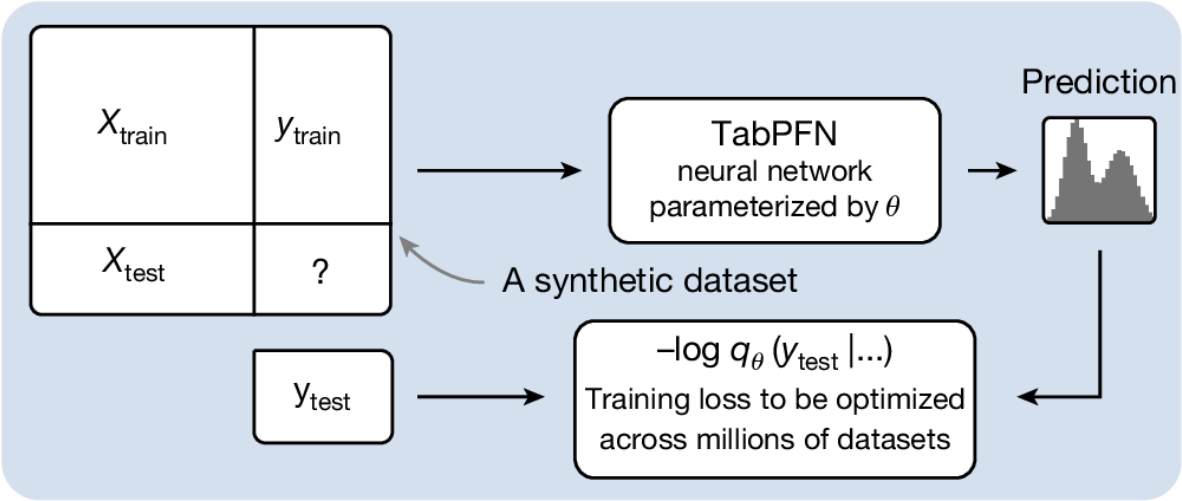
- Expensive Inference
  - Exact sampling from the posterior (e.g. Markov-Chain Monte-Carlo) is **slow**

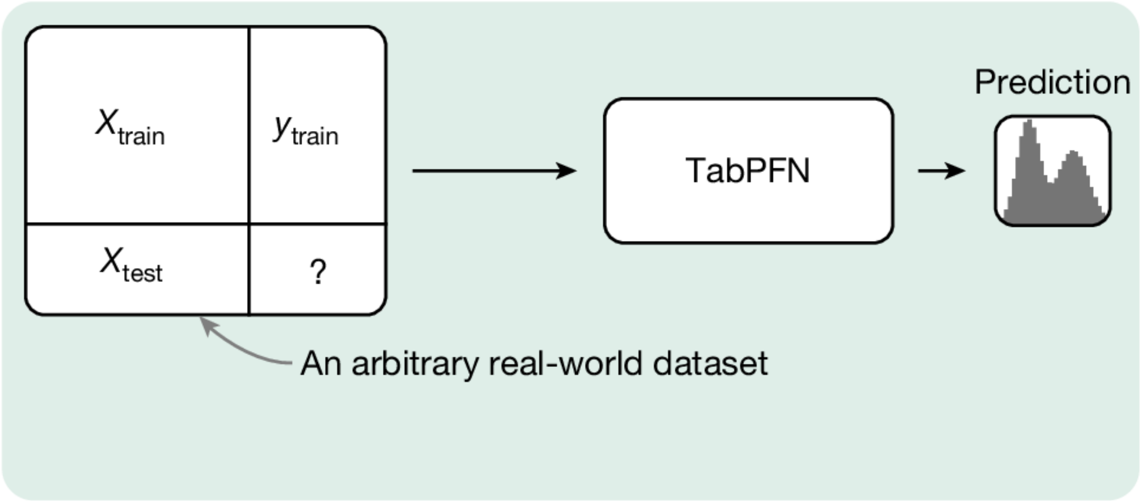
- Approximation Shortcuts
  - Methods such as variational inference **lose concentration guarantees**

UNIVERSITY OF TORONTO

Prior-Data Fitted Networks (PFNs) – E.g., TabPFN [1]

- Directly learn the predictive distributions using **large-scale pre-training of transformers**





UNIVERSITY OF TORONTO

[1] Hollmann, Noah, et al. "Accurate predictions on small data with a tabular foundation model." *Nature* 637.8045 (2025): 319-326.

Challenges of Using PFNs For Causal Effect Estimation

- **Challenge 1 (Distribution Shift):** Typically, it is assumed that train and test data come from the same distribution

$$(X_{train}, T_{train}, Y_{train}) \sim (X_{test}, T_{test}, Y_{test})$$

Not the case in causal inference:

$$(X_{train}, T_{train}, Y_{train}) \sim (X_{test}, T_{test} = 0, Y_0)$$

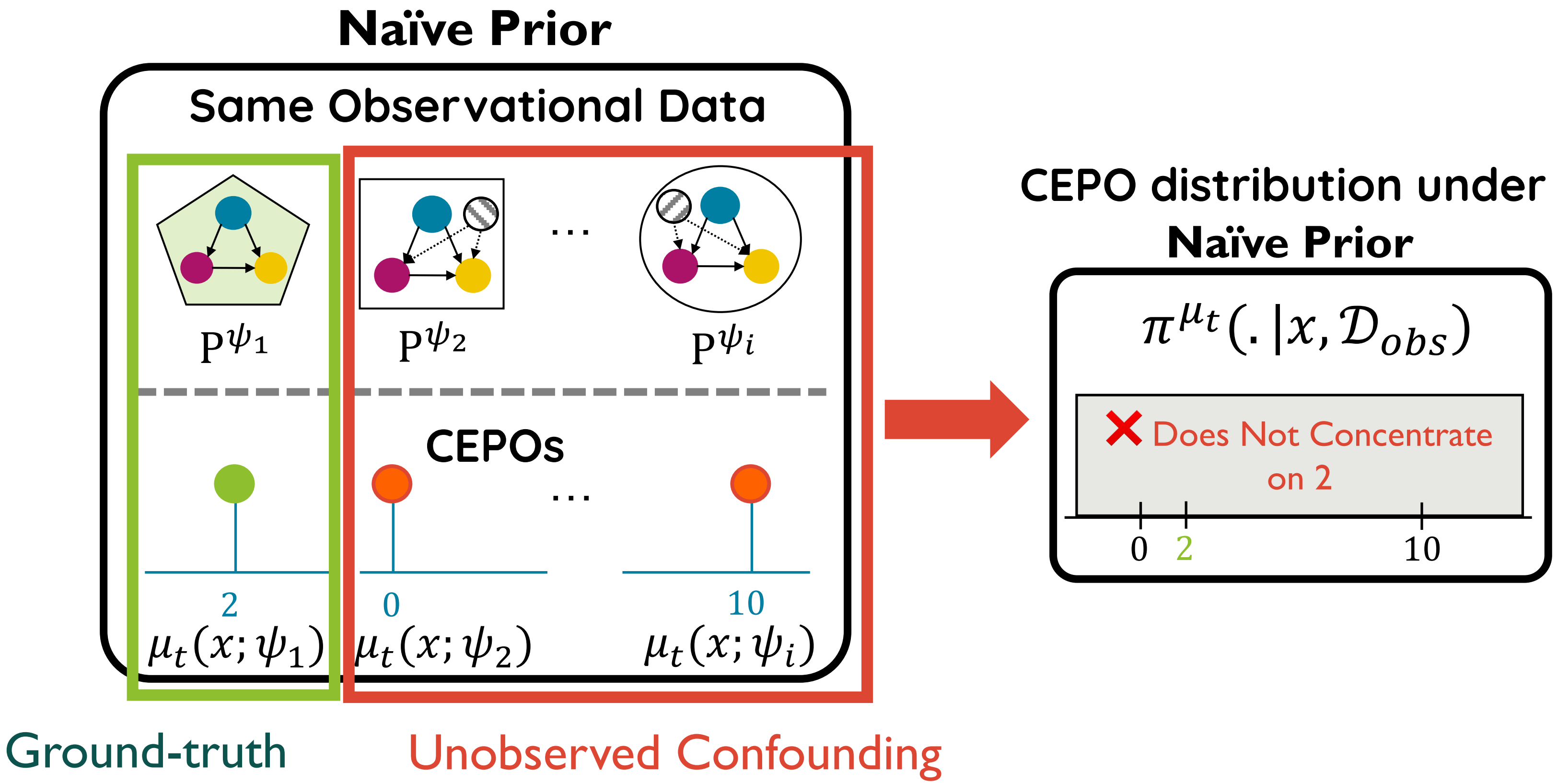
or

$$(X_{train}, T_{train}, Y_{train}) \sim (X_{test}, T_{test} = 1, Y_1)$$

UNIVERSITY OF TORONTO

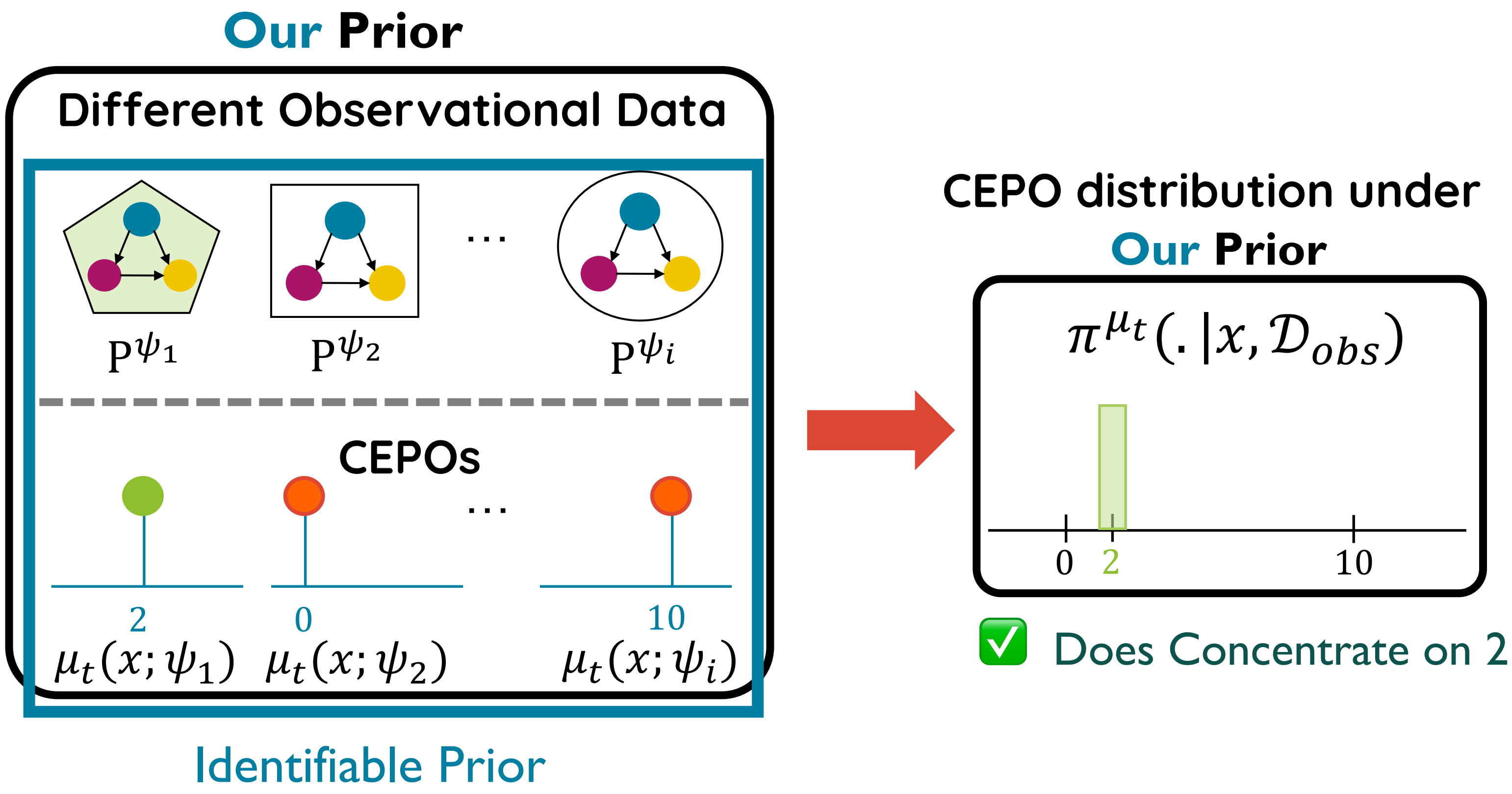
Challenges of Using PFNs For Causal Effect Estimation

- **Challenge 2 (Identifiability):** It's not always possible to identify causal effects from observational data alone.



Challenges of Using PFNs For Causal Effect Estimation

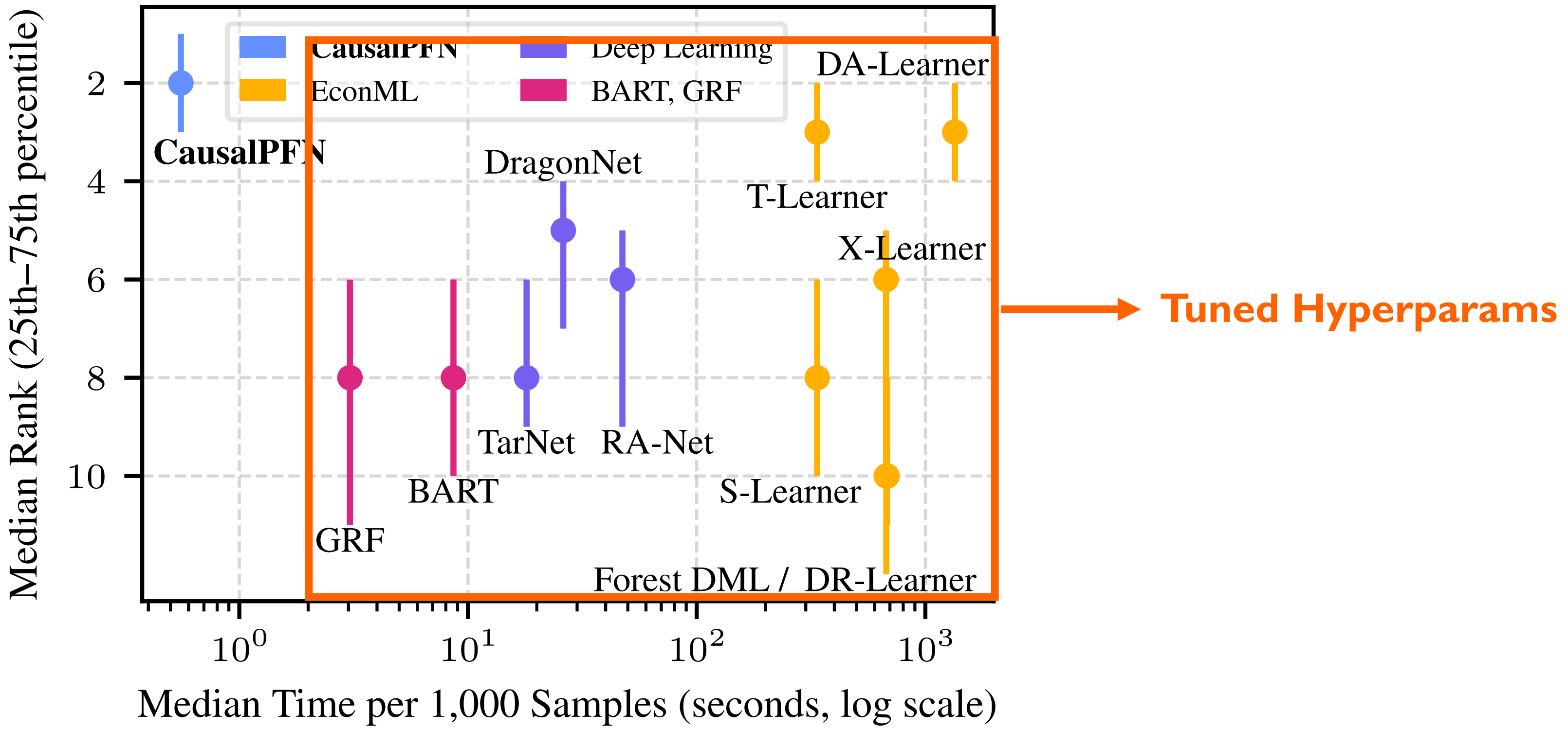
- **Challenge 2 (Identifiability):** It's not always possible to identify causal effects from observational data alone.



✅ Does Concentrate on 2

Empirical Results

- Time vs. Performance comparison across 310 causal inference tasks from IHDP, ACIC, and Lalonde.
- CausalPFN achieves the best average rank (by the error in CATE estimation).



Summary

- Bayesian causal inference with the large-scale training paradigm of PFNs
- Theoretical characterization of rationale for prior identifiability
- Strong empirical results; a practical tool for automated causal inference

UNIVERSITY OF TORONTO

https://github.com/vdblm/CausalPFN

Thank you!

Try the Model! `pip install causalpfn`

UNIVERSITY OF TORONTO