

PRIM: Meta-Learned Bayesian Root Cause Analysis

Christopher Lohse^{1,2} Anish Dhir³ Amadou Ba²

Bradley Eck² Marco Ruffini¹ Jonas Wahl^{4,5}

October 28, 2026

¹University of Dublin, Trinity College, Dublin, Ireland, ²IBM, ³Gatsby Computational Neuroscience Unit, University College London, ⁴Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Saarbrücken, Germany, ⁵Department of Philosophy, University of Bergen, Norway

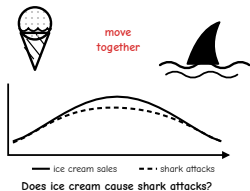
Accepted to the 40th Conference on Neural Information Processing Systems (NeurIPS 2026)

C.L. was partially funded by the EcoTwin project, which has received funding from the European Commission's Horizon Europe programme under grant agreement No 101157369. JW received support from the German Federal Ministry of Research, Technology and Space (BMFTR) as part of the project MAC-MERLin (Grant Agreement No. 01IW24007)

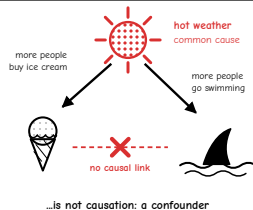
Setup and Motivation

A Brief Introduction to Causality

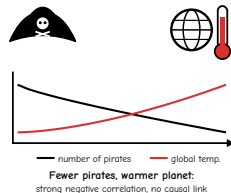
1. Correlation...



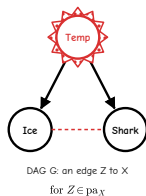
2. Hidden common cause



3. Negative correlation



4. Structural causal model (SCM)



Each variable is a function of its causal parents and noise:

$$X := \phi_X(\text{pa}_X, \epsilon_X)$$

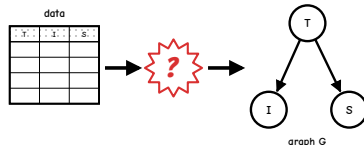
$$\text{Temp} := \epsilon_T$$

$$\text{Ice} := \phi_I(\text{Temp}, \epsilon_I)$$

$$\text{Shark} := \phi_S(\text{Temp}, \epsilon_S)$$

Intervene: $\text{do}(\text{Ice} = 0)$
ban ice cream: shark attacks stay the same,
so ice cream has no causal effect on them

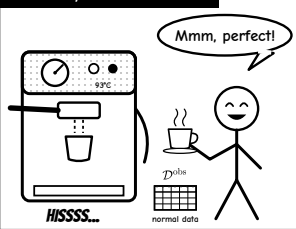
5. Causal discovery



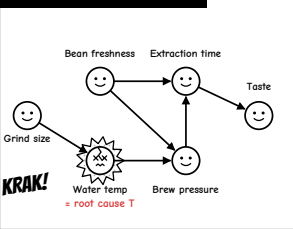
Causal discovery: infer the structure of G
from observational or interventional data

The Root Cause Analysis Problem

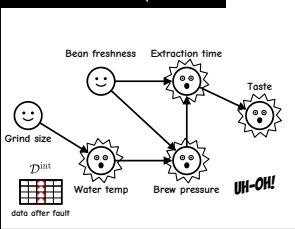
1. Normally: delicious coffee



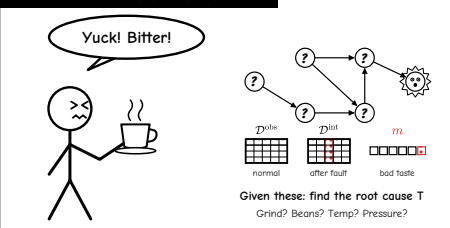
2. Then the heater breaks...



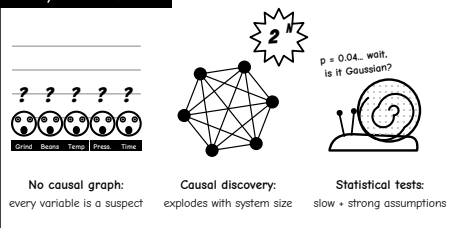
3. ...and the effects spread



4. All we observe: data + bad taste



5. Why is this hard?



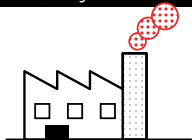
Applications & Existing Challenges

Cloud microservices



latency diagnosis,
fault localization

Manufacturing



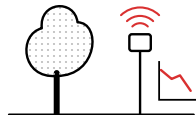
system failures,
anomaly detection

Medicine



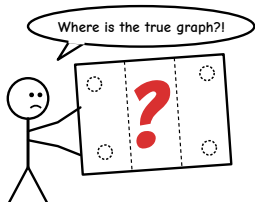
diagnostic reasoning

Sustainability



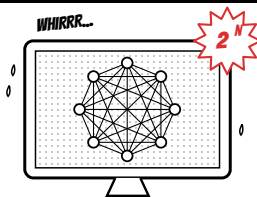
environmental monitoring

Graph-based methods



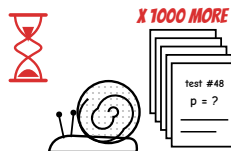
Need accurate prior causal knowledge
...which is often unavailable

Graph-free methods



Must do causal discovery
at test time: expensive

Statistical testing



"Almost done... is it Gaussian?"

Gets slow as the system grows,
and relies on strong assumptions

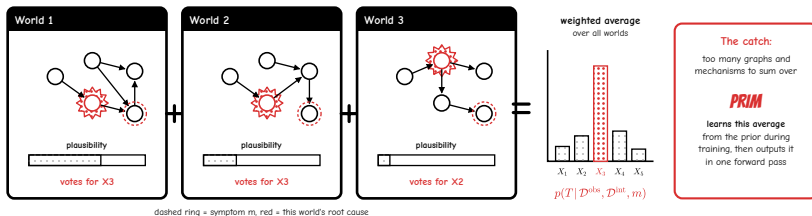
Approach

PRIM Formulation

$$\underbrace{p(T | \mathcal{D}^{\text{obs}}, \mathcal{D}^{\text{int}}, m)}_{\substack{\text{root-cause posterior} \\ \text{what we want}}} = \sum_{\mathcal{G}} \underbrace{p(T | \mathbf{f}, \mathcal{G}, m)}_{\substack{\text{which root cause} \\ \text{explains } m \text{ in this world?}}} \underbrace{p(\mathbf{f} | \mathcal{G}, \mathcal{D})}_{\substack{\text{how plausible are} \\ \text{these mechanisms?}}} \underbrace{p(\mathcal{G} | \mathcal{D})}_{\substack{\text{how plausible} \\ \text{is this graph?}}} d\mathbf{f}$$

average over every possible graph and mechanism
(all the worlds that could have produced the data)

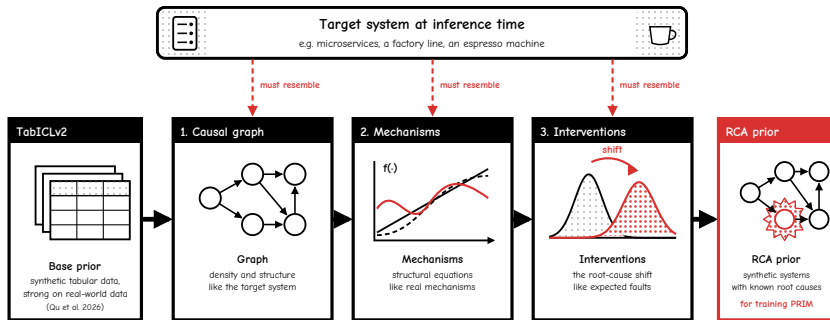
In pictures: many possible worlds



This posterior is identifiable; under standard assumptions on the mass concentration of the prior, posterior updates concentrate asymptotically on the true root cause set as the sample size goes to infinity (Ghosal and Vaart 2017).

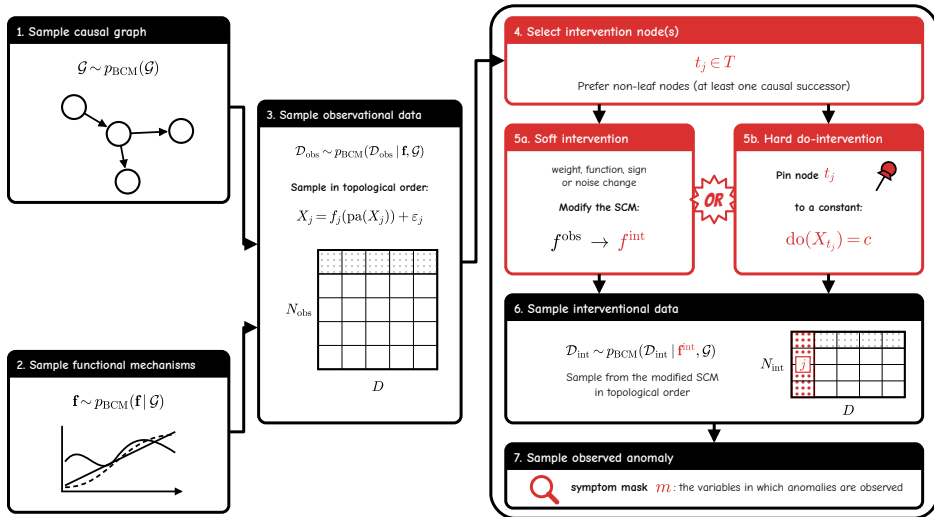
Method

Adapting a Prior for the RCA setting

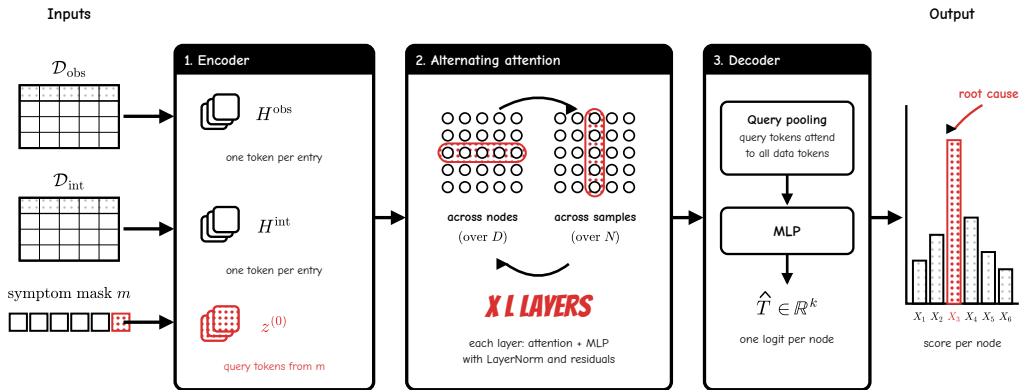


Adapting the TabICLv2 prior Qu et al. 2026 to root cause analysis. The causal graph, mechanisms, and root-cause interventions are each chosen to resemble the target system, yielding an RCA prior of synthetic systems with known root causes.

Data Generation and Prior Design

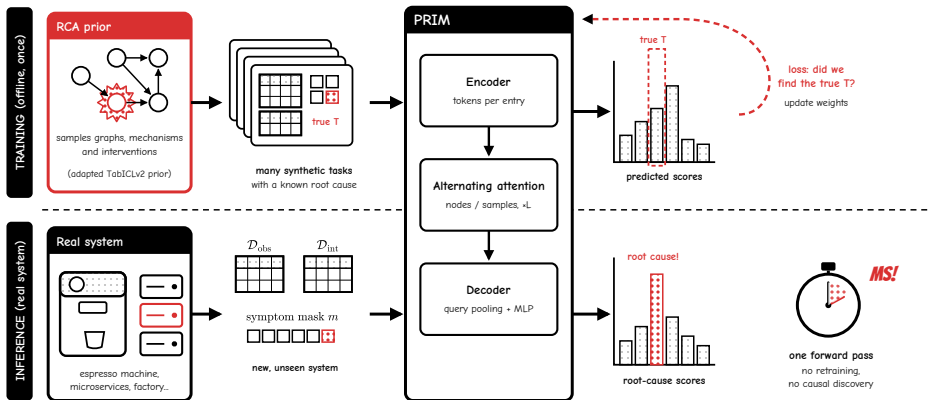


PRIM Architecture



Drawing inspiration from Dhir et al. 2025; Ke et al. 2023

PRIM PFN Training

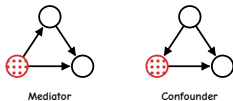


Following standard PFN training procedure Hollmann, Müller, Eggenberger, et al. 2023; Hollmann, Müller, Purucker, et al. 2025

Results & Future Work

Results

Convergence: 3-node graphs



0.9 – 1.0

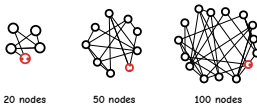
Recall@1 with enough data
(chance level: 0.33)

More interventional samples:

$n_{\text{int}} = 1$ **0.7** $\longrightarrow$ $n_{\text{int}} = 50$ **~1.0**

PRIM converges to the true
root cause on both topologies

Scalability: 20-100 nodes (in-distribution)



PRIM stays stable as systems grow:

NN functions **0.75 – 0.84**

GP functions **0.63 – 0.76**

best baselines: 0.57 or less, and dropping



7 ms

inference at every size

Real-world (out-of-distribution)



PetShop
microservices



CausRCA
manufacturing

PetShop (top-3 recall)



BEST

CausRCA full (MAP@3)



BEST

CausRCA sub (MAP@3)

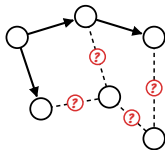


No causal graph given.

Best on 2 of 3 settings, competitive on the third

Future Work

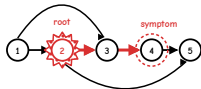
1. Partial graph knowledge



— known - - - unknown

Use what is known
plug in partial graph information

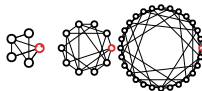
2. Interpretability



red path = the explanation

Commit to a causal order
so the answer comes with a why

3. Scale

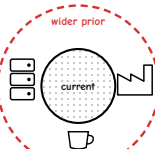


today

next: far bigger

Scale to bigger systems
beyond 100 nodes

4. Wider priors



Apply wider priors
more mechanisms, graphs and fault types

Thank you!

-  Budhathoki, Kailash, Dominik Janzing, et al. (2021). **“Why did the distribution change?”** In: *International Conference on Artificial Intelligence and Statistics*. PMLR, pp. 1666–1674.
-  Budhathoki, Kailash, Lenon Minorics, et al. (17–23 Jul 2022). **“Causal structure-based root cause analysis of outliers”**. In: *Proceedings of the 39th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri et al. Vol. 162. Proceedings of Machine Learning Research. PMLR, pp. 2357–2369. URL: <https://proceedings.mlr.press/v162/budhathoki22a.html>.
-  Dhir, Anish et al. (2025). **“Continuous Bayesian Model Selection for Multivariate Causal Discovery”**. In: *Forty-second International Conference on Machine Learning*. URL: <https://openreview.net/forum?id=zydNWJzoVd>.

-  Ghosal, Subhashis and Aad van der Vaart (2017). **“Consistency: General Theory”**. In: *Fundamentals of Nonparametric Bayesian Inference*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. Chap. 6, pp. 123–164.
-  Hardt, Michaela et al. (Jan. 2024). **“The PetShop Dataset — Finding Causes of Performance Issues across Microservices”**. In: *Proceedings of the Third Conference on Causal Learning and Reasoning*. Ed. by Francesco Locatello and Vanessa Didelez. Vol. 236. Proceedings of Machine Learning Research. PMLR, pp. 957–978. URL: <https://proceedings.mlr.press/v236/hardt24a.html>.

-  Hollmann, Noah, Samuel Müller, Katharina Eggensperger, et al. (2023). **“TabPFN: A Transformer That Solves Small Tabular Classification Problems in a Second”**. In: *The Eleventh International Conference on Learning Representations*.
-  Hollmann, Noah, Samuel Müller, Lennart Purucker, et al. (Jan. 2025). **“Accurate predictions on small data with a tabular foundation model”**. en. In: *Nature* 637.8045, pp. 319–326. ISSN: 1476-4687. DOI: 10.1038/s41586-024-08328-6. URL: <https://www.nature.com/articles/s41586-024-08328-6> (visited on 01/16/2026).
-  Ke, Nan Rosemary et al. (2023). **“Learning to Induce Causal Structure”**. In: *International Conference on Learning Representations*. URL: https://openreview.net/forum?id=hp_RwhKDJ5.

-  Mehling, Carl Willy et al. (Jan. 2026). “Enabling Joint Benchmarking of Automated Root Cause Analysis and Causal Discovery in Manufacturing Using the *causRCA* Dataset”. In: *Procedia CIRP*. 13th CIRP Global Web Conference 139, pp. 114–120. ISSN: 2212-8271. DOI: 10.1016/j.procir.2025.09.010. URL: <https://www.sciencedirect.com/science/article/pii/S221282712500976X> (visited on 03/13/2026).
-  Pearl, Judea (1995). “Causal diagrams for empirical research”. In: *Biometrika* 82.4, pp. 669–688.
-  — (2009). *Causality*. Cambridge university press.

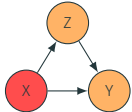
-  Pham, Luan et al. (2025). **“Rcaeval: A benchmark for root cause analysis of microservice systems with telemetry data”**. In: *Companion Proceedings of the ACM on Web Conference 2025*, pp. 777–780.
-  Qu, Jingang et al. (2026). **“TabICLv2: A Better, Faster, Scalable, and Open Tabular Foundation Model”**. In: *Forty-third International Conference on Machine Learning*. URL: <https://openreview.net/forum?id=SxsyLjIfWB>.

Appendix

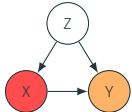
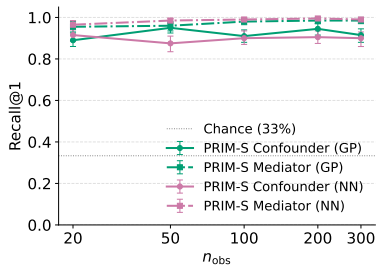
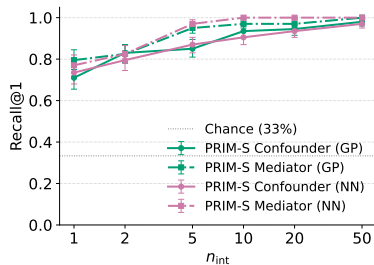
Synthetic Validation: 3-Node Structures

Test Topologies:

Mediator



Confounder

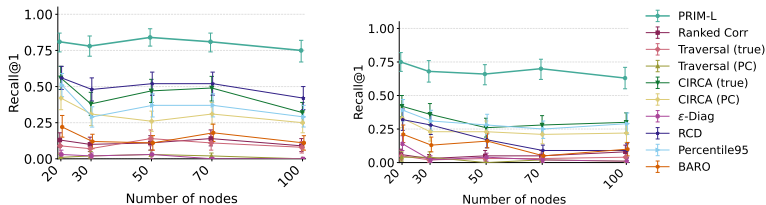


Left: Varying n_{obs} , keeping $n_{int} = 10$ Right: Varying n_{int} , keeping $n_{obs} = 100$

Legend: ■ Root cause (X) ■ Descendants ■ Unaffected.

Result: PRIM competitive with graph-not-given baselines; accurately identifies root cause with sufficient data across both topologies.

Synthetic Validation: Scalability



Recall@1 vs. number of nodes (4–100 nodes). Left: Neural Network Functions, Right: Gaussian Process Functions

Key findings:

1. **PRIM outperforms all baselines** (Traversal, CIRCA, RCD, ϵ -diagnosis) as systems grow
2. Constant 17ms inference (independent of size)
3. RCD loses ground with scale (causal discovery is expensive) (Pham et al. 2025)

Real World Benchmarks

PetShop (microservices)

- 44 services, 48 fault injections
- 4 traffic scenarios, latency & availability metrics

CausRCA (manufacturing)

- 100 runs, 19 fault experiments
- 3 subsystems: coolant, hydraulics, probe

		ϵ -diag	RCD	TimeRank	Baro	Corr	CausTR+PC	PRIM [†]
PetShop		0.10 [*]	0.39 [*]	0.15	0.08	0.78 [*]	–	0.79
CausRCA	Full	0.02	0.04	0.24	0.09	0.05	–	0.27
	Sub	0.11	0.24	0.60	0.75	0.33	0.57 [*]	0.61

Table 1: Average RCA accuracy without a given graph. PetShop: top-3 recall; CausRCA: MAP@3. * from original papers Hardt et al. 2024; Mehling et al. 2026; – not reported; **bold** best per row; [†] ours.