SIGMOD · 2026 · BENGALURU

Causal Explanations for Disparate Trends: Where and Why?

A COLLABORATION OF

Ben-Gurion University of the Negev

TECHNION Israel Institute of Technology

THE UNIVERSITY OF UTAH

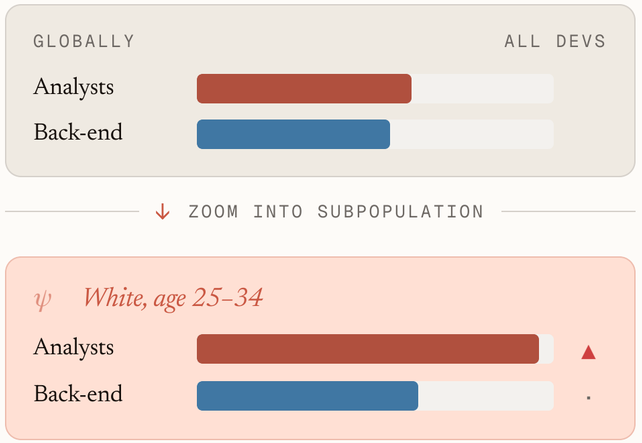
Tal Blau, Brit Youngmann, Anna Fariha, Yuval Moskovitch

Session · Causal & Explanation
May 31 - June 5, 2026

Three datasets. Three *insights* the global view *can't see*.

STACK OVERFLOW **USE CASE 1**

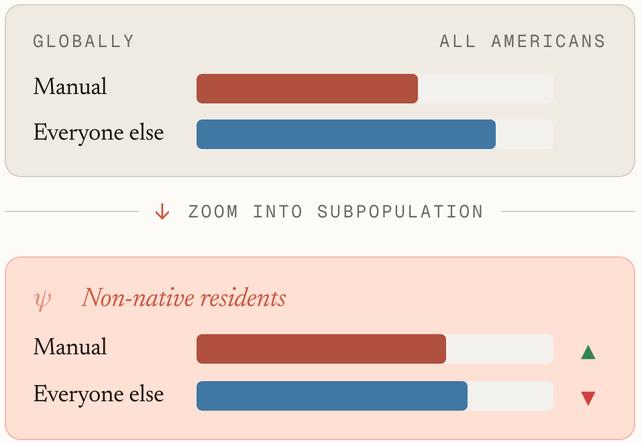
A trend *amplified*.



The same treatment **massively rewards one group**, *barely touches the other.*

ACS **USE CASE 2**

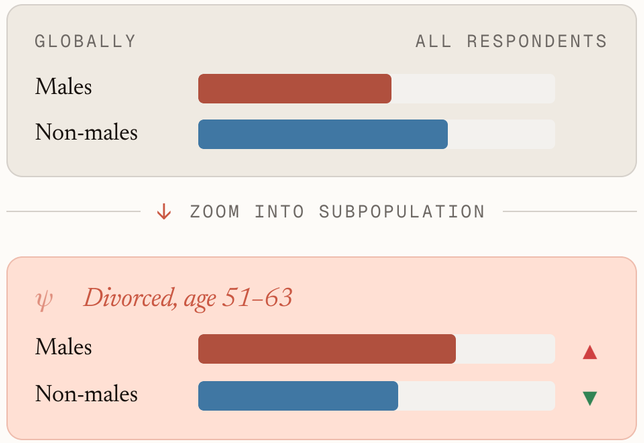
Effects in *opposite* directions.



One treatment, **two opposite causal effects**. *A national policy hurts somebody.*

MEPS **USE CASE 3**

A trend that *reverses*.



Simpson's paradox in the wild — the **local trend reverses** *the global one.*

The basic vocabulary

DEFINITION · PATTERN

A conjunction of equality predicates

$$\psi = \phi_1 \wedge \phi_2 \wedge \cdots \wedge \phi_k , \quad \phi : A_i \text{ op } a_i$$

ψ(D) ⊆ D is the matching subpopulation. e.g. `Ethnicity = Asian ∧ Role = Analyst`

DEFINITION · OUTCOME

The variable whose disparity we explain

$$O \in A \cdot \text{numeric or binary}$$

A designated attribute whose AVG(O) differs across groups — e.g. `TotalCompensation`.

DEFINITION · TWO GROUPS OF INTEREST

The disparity is between g_1 and g_2

$$g_1 = \psi_{g_1} , \; g_2 = \psi_{g_2} \cdot \text{disjoint patterns}$$

The user picks the comparison — e.g. `Role = Analyst` vs. `Role = Backend`.

DEFINITION · ATTRIBUTE PARTITION

Where vs. why — two disjoint sets

$$A \setminus \{O\} = I \sqcup M$$

IMMUTABLE I

defines **where** · ethnicity, age

MUTABLE M

defines **why** · experience, education

The formal language of disparity

DEFINITION 4.2 · DISPARITY EXPLANATION

$$\phi = (\psi_g, \psi_e)$$

ψ_g over I — the subpopulation where g_1 and g_2 diverge. ψ_e over M — the treatment that explains the gap inside $\psi_g(D)$.

DEFINITION 4.3 · DISPARITY SCORE

$$\frac{\left| \mathrm{CATE}_{G_b}(\psi_e, O \mid \psi_g \wedge \psi_{g_1}) - \mathrm{CATE}_{G_b}(\psi_e, O \mid \psi_g \wedge \psi_{g_2}) \right|}{\max\{ |o| : o \in O \}}$$

Difference of two CATE values across g_1 and g_2, normalised by the maximal outcome.

DEFINITION 4.4 · SUPPORT

$$\frac{\left| \psi_{g \wedge g_1}(D) \cup \psi_{g \wedge g_2}(D) \right|}{|D|}$$

Fraction of tuples covered — eliminates explanations on negligible slivers.

Anatomy of a disparity explanation

We observe that **analysts** earn ~$10K more than **back-end developers** on average. Where does this gap come from — and why?

DISPARITY EXPLANATION $\Phi = (\Psi_G, \Psi_E)$

Among `White individuals · age 25-34` having `6-8 years of professional coding` boosts the **total compensation** of **analysts** much more than that of **back-end developers**.

Comparison g_1 vs. g_2	`Analysts` vs. `Back-end developers`
Subpopulation Ψ_g · immutable	`Ethnicity = White ∧ Age = 25-34`
Treatment Ψ_e · mutable	`YrsProfCoding = 6-8`

CATE WITHIN Ψ_G · AVG(TC)

Gap of **$10K** within this subgroup.

QUALITY OF THE EXPLANATION

DISPARITY SCORE $\Delta(\Phi)$

0.016

|44,058 − 10,552| / 2,000,000

SUPPORT

34.6%

16,508 / 47,702 tuples

THE KEY INSIGHT

By **separating** the **where** (immutable) from the **why** (mutable), an explanation becomes **actionable**.

Maximise causal explainability — *under three constraints.*

$$\max \Sigma \Delta(\phi)$$

SUBJECT TO

- k — at most **k** explanations in the set
- σ — each subpopulation covers ≥ **σ** of the data
- τ — pairwise Jaccard similarity ≤ **τ**

⚠ NP-HARD · REDUCTION FROM INDEPENDENT SET

1 Subpopulation miner

One Apriori pass over `I` ⊂ attributes; keep only patterns with support ≥ σ.

2 Explanation miner — *the bottleneck*

For each candidate, lattice-search treatment patterns over `M`. Sampling + DAG caching + parallel scoring make it tractable.

3 Greedy with diversity

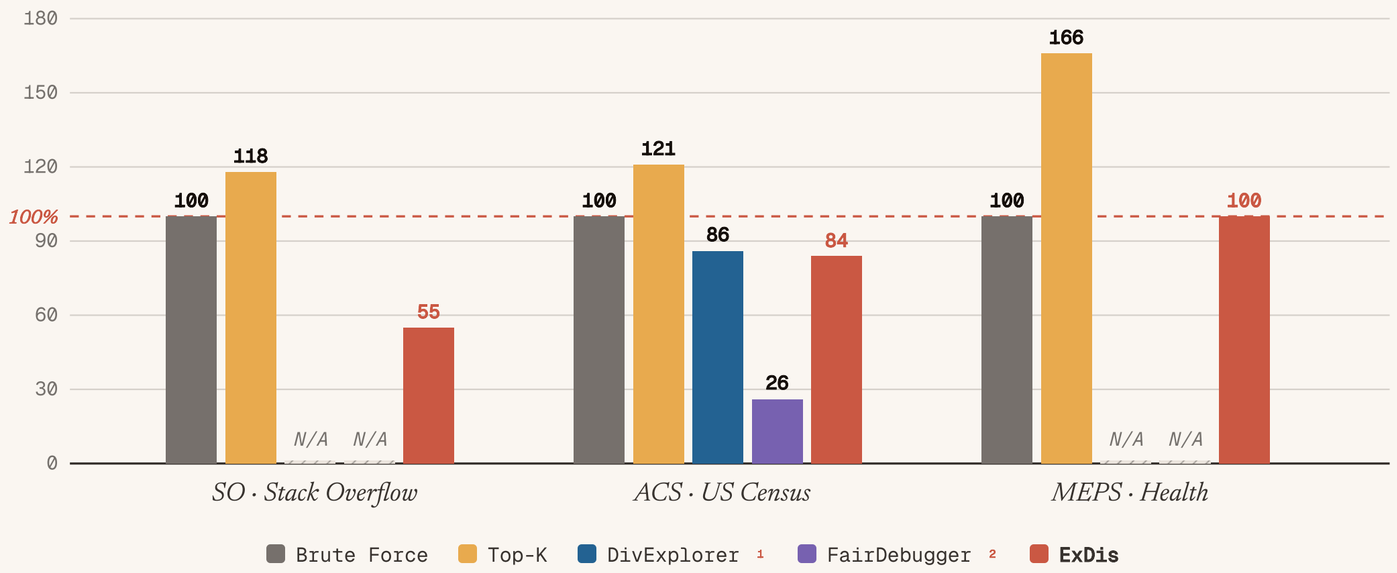
Hierarchical-cluster on subpopulation overlap; iteratively pick the highest-Δ candidate that respects τ.

RESULTS · I EXPLANATION QUALITY

Power of *local* causal explanations.

Disparity score (Δ) by algorithm — % of brute-force ground truth

The dashed **100%** line is the brute-force ground truth. Closer to it = better. Top-K can exceed 100% because redundant explanations stack. DivExplorer & FairDebugger return N/A on SO & MEPS (no causal "why").



EXDIS · THIS WORK

Closest to ground truth, every time.

SO	ACS	MEPS
55%	84%	100%
2.9× faster	2.7× faster	1.1× faster
VS. BRUTE FORCE	VS. BRUTE FORCE	VS. BRUTE FORCE

1 Pastor, De Alfaro, Baralis. *Looking for trouble: Analyzing classifier behavior via pattern divergence.* SIGMOD 2021. dl.acm.org ↗

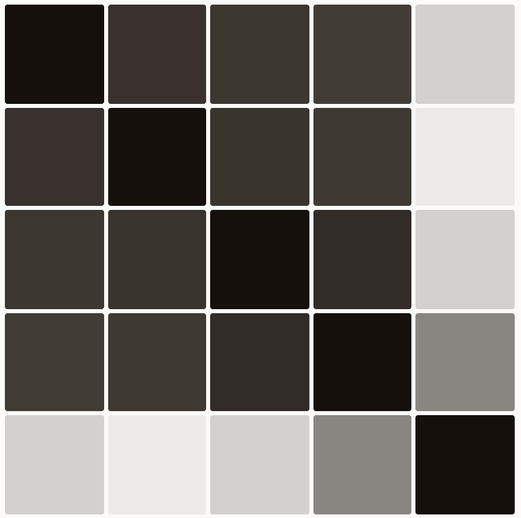
2 Surve, Pradhan. *Example-based Explanations for Random Forests using Machine Unlearning.* arXiv:2402.05007, 2024. arxiv.org ↗

RESULTS · II DIVERSITY MATTERS

Why the diversity constraint *pays off*.

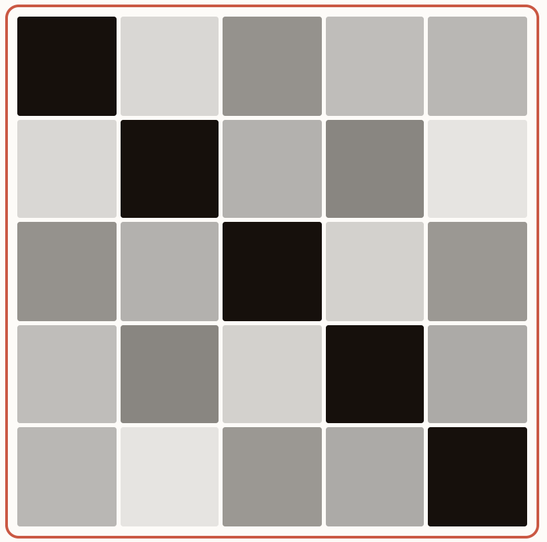
Two pairwise Jaccard-similarity matrices over the top-5 explanations each method returns. Lighter = less overlap = more new information per explanation.

Top-K · *no diversity enforcement*



Dark off-diagonals — **high redundancy**. The 5 explanations restate the same insight; readers learn little new.

ExDis · *Jaccard similarity τ enforced*



A clean checkerboard — **five truly distinct insights**. The greedy clustering guarantees broad, non-overlapping coverage.

IDENTICAL DISJOINT

A NEW PARADIGM FOR FOCUSED ACTION

Beyond the global average.

01

New resolution.

Global averages hide the truth. ExDis surfaces the hidden corners of the data — **where** disparity actually lives.

02

Causality, not correlation.

Stop chasing spurious correlations. Identifying the true **why** enables ***effective***, targeted intervention.

03

Automatic diversity.

Receive a curated set of distinct, non-overlapping explanations — saving hours of manual filtering.

ExDis is not just an algorithm – it is a paradigm shift. From "proving a bias exists" to diagnosing and acting on its causal roots.

THANK YOU

Questions?