# Interpreting Neural Combinatorial Optimization via Evolving Programmatic Bottlenecks

Haocheng Duan[1*], Yuxin Guo[1*], Jieyi Bi[2], Anqi Xie[1], Sirui Li[3†], Yining Ma[4‡], Cathy Wu[4]

[1]Carnegie Mellon University [2]Nanyang Technological University
[3]Microsoft Research [4]Massachusetts Institute of Technology

*Equal contribution †Work done outside employment ‡Corresponding author

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

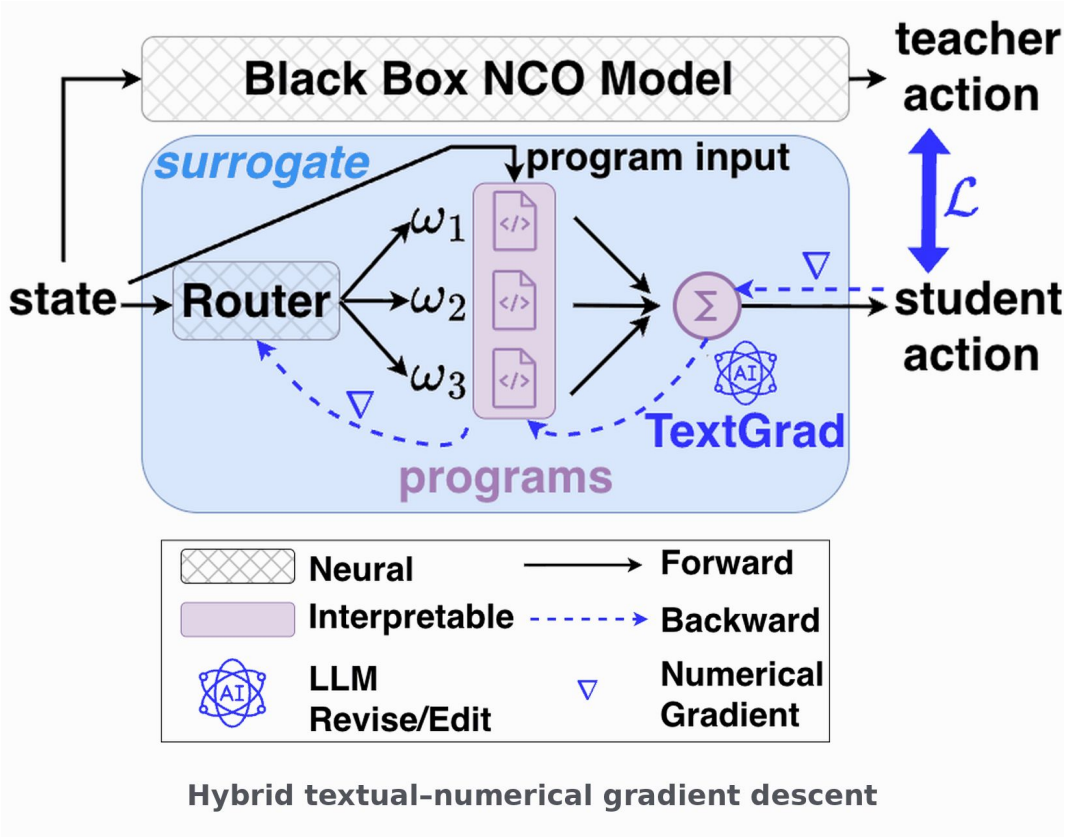
# NCO policies are black boxes

**Black box.** NCO solvers achieve remarkable performance, but their policies are black boxes.

**Surrogate.** Distill the policy into a surrogate, i.e., a concept bottleneck model (CBM) [1], which decides through human-readable concepts we can audit and edit.

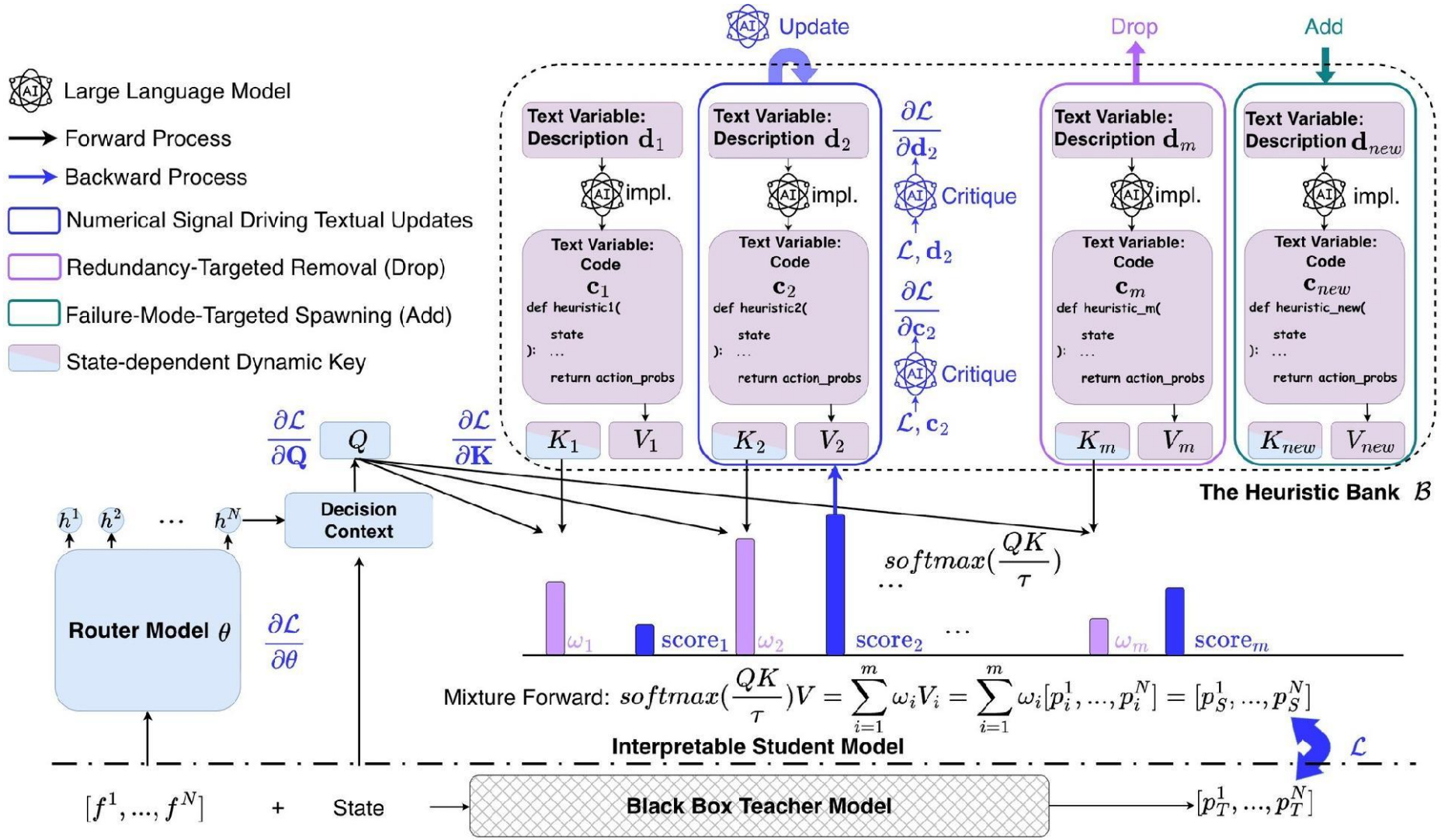
**Challenge:** CBM concepts are usually input attributes; NCO is better explained by decision rules, which are hard to predefine.

**Our solution: EPB.** LLMs evolve human-readable programs as concepts; a state-dependent router combines them.




[1] Koh et al., ICML, PMLR 119, 5338–5348 (2020).

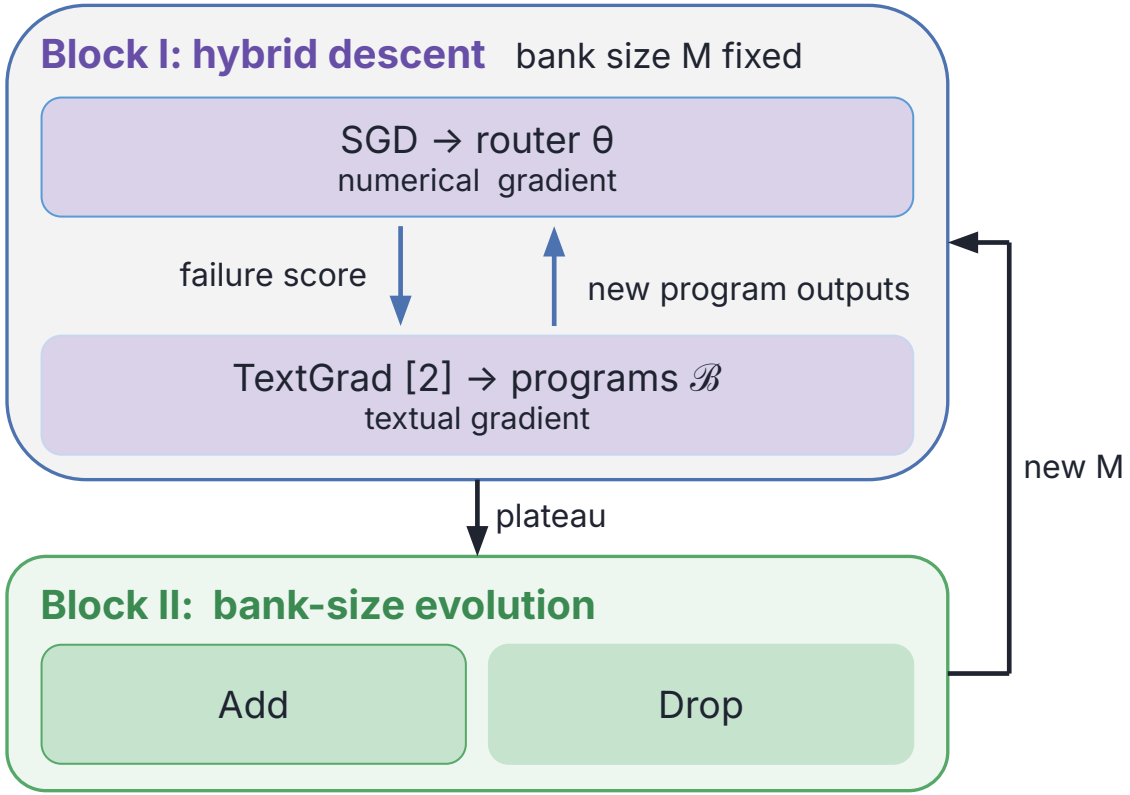
# EPB: readable programs plus a router



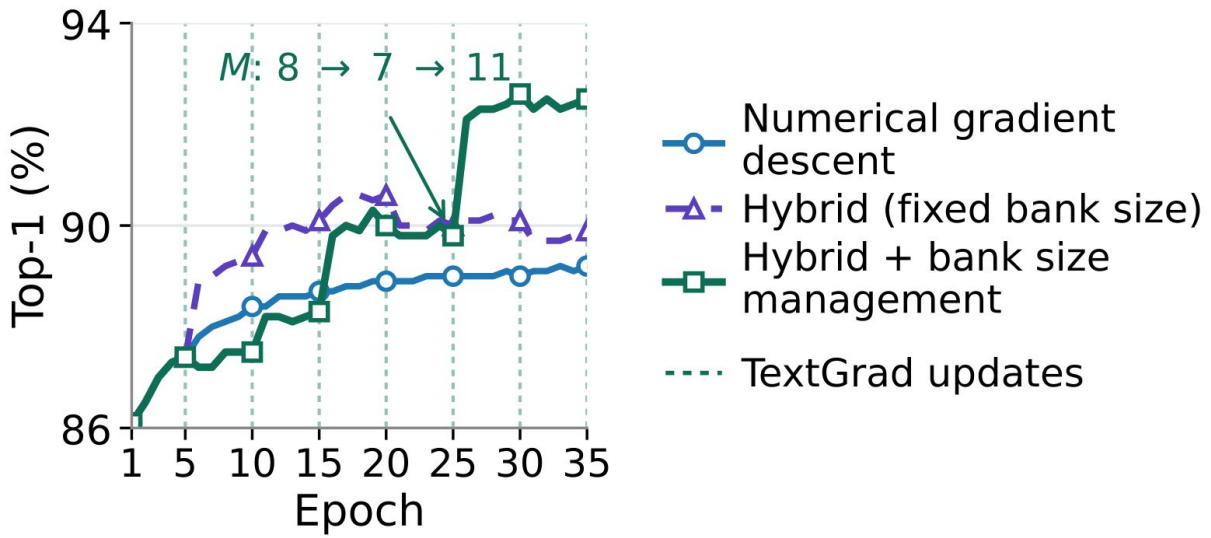
# Two-block optimization to train EPB

**Objective:** $\min_{\theta, \mathcal{B}, M} \mathbb{E}_s \left[ \mathrm{KL}\left( \pi_T(\cdot \mid s) \,\|\, \pi_S(\cdot \mid s) \right) \right]$

**Challenge:**
symbolic programs $\mathcal{B}$,
continuous router θ,
discrete M




Loop until Convergence




Ablation: SGD alone < Block I < Block I + Block II

[2] Yuksekgonul et al., Nature 639, 609–616 (2025).

# EPB matches its teacher with high fidelity

**Table 1. Fidelity comparison of the surrogate to the teacher model.** Teachers: POMO [3] and LEHD [4].

| | TSP | | | | CVRP | | | |
|---|---|---|---|---|---|---|---|---|
| | Top-1 agreement↑ | | Search gap ↓ | | Top-1 agreement↑ | | Search gap ↓ | |
| **Teacher** | Baseline | **EPB** | Base | **EPB** | Baseline | **EPB** | Baseline | **EPB** |
| POMO-50 | 87.2% | **94.8%** | 2.45% | **0.03%** | 71.6% | **85.8%** | 7.52% | **1.13%** |
| POMO-100 | 86.5% | **94.1%** | 3.90% | **0.07%** | 71.0% | **86.5%** | 9.12% | **1.25%** |
| LEHD-100 | 86.9% | **94.7%** | 1.64% | **0.34%** | 68.7% | **85.1%** | 11.53% | **2.16%** |

Baseline = the best of 5 interpretable baselines
two LLM coding agents (GPT-6, Fable 5.1), Tree, NAM [5] and MoËT [6].

[3] Kwon et al., NeurIPS 33, 21188–21198 (2020). [4] Luo et al., NeurIPS 36, 8845–8864 (2023).
[5] Agarwal et al., NeurIPS 34, 4699–4711 (2021). [6] Vasić et al., Neural Networks 151, 34–47 (2022).

# Even simple programs stay close to the teacher

**A simple program** · POMO-TSP-100 · 2 primitives, 3 statements

```python
def heuristic(locs, current_node, mask, visited, step_norm, first_node):
    # Uses 2 concept(s): dist_cur, heading_cos
    d = feat.dist_cur(locs, current_node)
    h = feat.heading_cos(locs, current_node, visited)
    # Combines proximity with a heading-alignment bonus.
    logits = -6.0 * d + 1.5 * h
    return logits
```

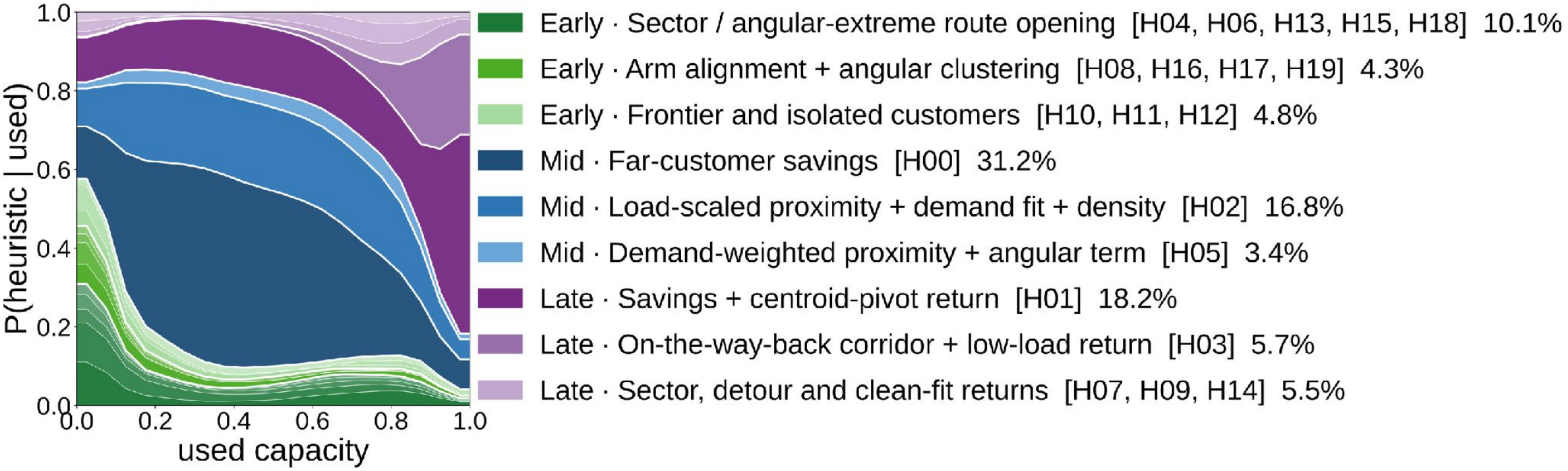
## Takeaway: a few simple programs already retain near-teacher solution quality

With at most 2 primitives per program, the 8-program POMO bank stays within 0.58% of POMO's search cost.

# Each decision step is explained by a readable program



# Program usage shifts as routes open, grow and close




## Aggregated decision patterns

POMO-CVRP-50, 200 instances:
routes open around spatial structure, grow by savings and demand fit, and close toward the depot.

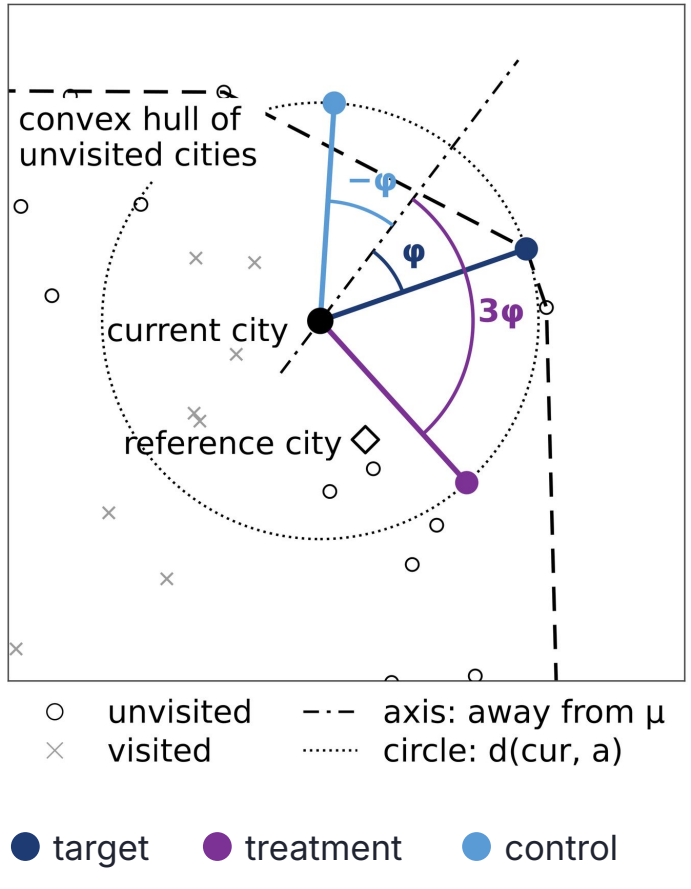
# The explanations are faithful to the student and the teacher

**Table 3. Dominance knockout.** Flip rate of the next action when one program is removed.

| Setting | Remove top-weighted program ($F_m$) | Remove a random program ($F_{rand}$) | Programs with $F_m > F_{rand}$ |
|---|---|---|---|
| POMO TSP | **94.3% [93.8, 94.8]** | 0.6% [0.4, 0.8] | 14 / 14 |
| LEHD TSP | **98.1% [97.6, 98.5]** | 0.2% [0.1, 0.4] | 11 / 11 |
| POMO CVRP | **65.2% [64.6, 65.7]** | 0.5% [0.5, 0.6] | 19 / 20 |

**Table 4. Counterfactual faithfulness tests.** Effect g of weakening one concept (g > 0 as predicted); share of samples with g > 0 in parentheses. †Manipulation check.

| Channel | POMO-TSP hull | LEHD-TSP isolation | POMO-CVRP corridor |
|---|---|---|---|
| Carrier score† | +0.38 [0.34, 0.43] (100%) | +1.20 [1.18, 1.23] (100%) | +0.92 [0.90, 0.94] (100%) |
| Routing log-weight | +1.54 [1.08, 1.99] (73%) | +0.29 [0.20, 0.37] (72%) | +0.39 [0.32, 0.46] (61%) |
| Program contribution | +0.38 [0.33, 0.43] (92%) | +0.08 [0.06, 0.09] (74%) | +0.25 [0.24, 0.26] (91%) |
| Student log-odds | +1.09 [0.80, 1.38] (83%) | +0.54 [0.45, 0.63] (71%) | +3.14 [3.00, 3.28] (87%) |
| Teacher log-odds | +9.57 [7.69, 11.37] (84%) | +1.47 [1.24, 1.74] (76%) | +9.79 [9.29, 10.29] (83%) |




Hull test: treatment moves the target toward μ (less peripheral); control mirrors the move.

# EPB can outscale its teacher and be edited to fix its failures

**Table 5. Out-of-distribution generalization results on unseen problem sizes.**
Cost gap of EPB vs. POMO.

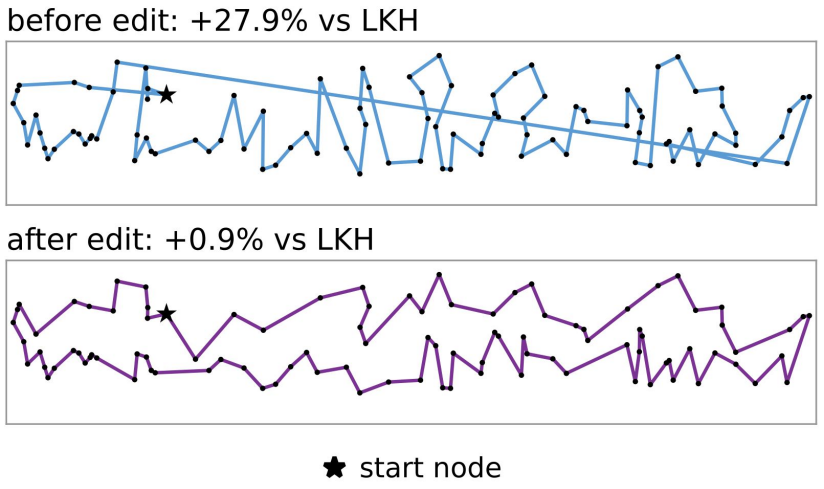
| | Greedy | Search |
|---|---|---|
| TSP-500 | **−14.58%** | **−13.27%** |
| CVRP-500 | **−12.81%** | **−14.72%** |

**−13% to −15%**

cost vs. POMO on 500-node TSP and CVRP
zero-shot, greedy and search decoding

**Table 6. Inference time edits on EPB surrogates.**
Gap on diagonal bands (vs. LKH / HGS).

| | TSP-100 | | CVRP-50 | |
|---|---|---|---|---|
| | Greedy | Search | Greedy | Search |
| Teacher (POMO) | 38.18% | 20.90% | 13.34% | 7.66% |
| EPB | 24.16% | 8.70% | 10.88% | 4.17% |
| Edited EPB | **7.01%** | **1.78%** | **5.94%** | **2.43%** |




EPB on one TSP-100 diagonal band, before and after editing the program bank

# Takeaways

- **EPB:** LLM-evolved programs + a state-dependent router
- **High fidelity** in all 6 settings, even with simple code
- **Explanations are readable and faithful**
- **Beyond the teacher:** better scaling, editable fixes

**Paper**
arxiv.org/abs/2606.19741

**Code**
github.com/HaochengDuan/EPB