

Classical Planning with LLM-Generated Heuristics: Challenging the State of the Art with Python Code

Augusto B. Corrêa

University of Oxford, UK

André G. Pereira

UFRGS, Brazil

Jendrik Seipp

Linköping University, Sweden

@NeurIPS'25



Source: NASA

```
1 (:action sample_soil
2 :parameters (?x - rover
3             ?s - store
4             ?p - waypoint)
5 :precondition (and (at ?x ?p)
6                  (at_soil_sample ?p)
7                  (equipped_for_soil_analysis ?x)
8                  (store_of ?s ?x)
9                  (empty ?s)
10                )
11 :effect (and (not (empty ?s))
12            (full ?s)
13            (have_soil_analysis ?x ?p)
14            (not (at_soil_sample ?p))
15            )
16 )
```

Solving Planning Tasks

Common approach: **heuristic search**

Domain-independent heuristics:
good & reusable

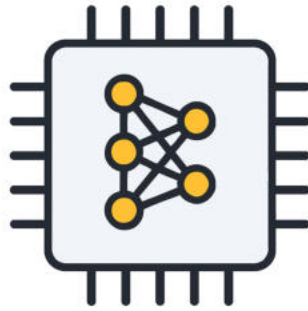
Can we **use LLMs to create heuristics?**



Domain



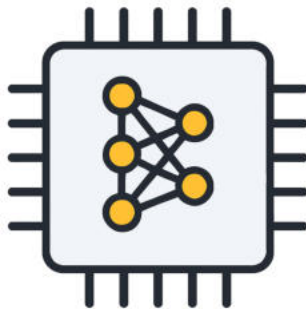
Domain



LLM



Domain



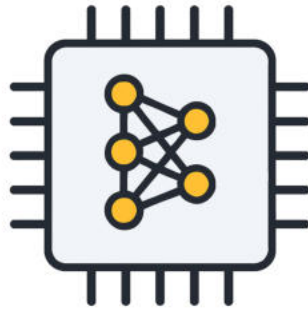
LLM



25 Heuristics



Domain



LLM

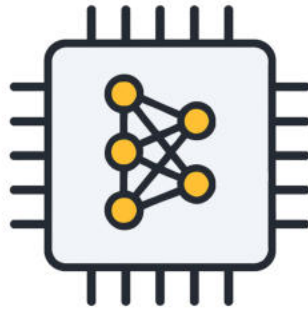


25 Heuristics





Domain



LLM

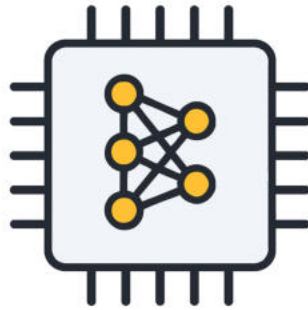


25 Heuristics





Domain



LLM



...

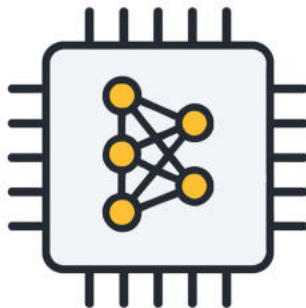


25 Heuristics





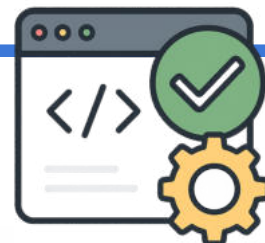
Domain



LLM



25 Heuristics



Test!

Experiments

Eight domains and 720 tasks:

International Planning Competition 2023

Much **harder** than domains usually tested

Plans with **hundreds of steps**

Heuristics from **DeepSeek R1** on **Pyperplan**

Our Method	373
C++ Baseline	295
Python Baseline	243
Gemini 2.0 Pro	202
DeepSeek R1	174
DeepSeek V3	48

Conclusion

LLMs create **good heuristics** for planning

Not only memorization

Can we do the same with **smaller models?**