

Progression on project in last week:

1. I have deployed the fast downwards on the computer, with its accompanying tools (VAL, report generating...);
2. I worked to be more familiar with PDDL syntax;
3. I have given the prompt (NL instruction, PDDL domain predicates) to Gemma3 and Devstral, and asked them to generate action:

Prompt: Given a description of an action in some domain, convert it to Planning Domain Definition Language (PDDL) action. You may only use the allowed predicates for each action. Allowed Predicates: - (handempty) : the hand is empty. - (holding ?x - block): block x is held. - (clear ?x - block): block x is clear. - (on ?x - block ?y - block) : block x is on block y. - (ontable ?x - block): block x is on the table. Input: The action, “put-down” will have the hand put down a block. PDDL Action:

Result of devstral:

```
(:action put-down
  :parameters (?x - block)
  :precondition (holding ?x)
  :effect (and
    (not (holding ?x))
    (handempty)
    (ontable ?x)
  )
)
```

Explanation: - Parameters: The action takes one parameter, ?x, which represents a block. - Precondition: The hand must be holding the block ((holding ?x)). - Effect: After performing the action, the block is no longer being held by the hand ((not (holding ?x))), the hand becomes empty ((handempty)), and the block is now on the table ((ontable ?x)).

Result of Gemma3

```
(action put-down)
:parameters (block ?x) //syntax problem?
:precondition (and (handempty) (holding ?x - block))
:effect (and (handempty) (not (holding ?x - block)) (ontable ?x -
block))
```

4. Reading relevant past paper (<https://arxiv.org/abs/2502.20175>);

Question:

1. What criteria or benchmark should I use to evaluate the result of the PDDL code LLMs generated?