X-RLflow: discussion points

1 Training cost is not reported

- Optimisation time excludes >1000 training episodes per graph
- How many inference passes until the gain over TASO pays back training?

2 RL or measurement?

- X-RLflow optimises measured latency; TASO and TENSAT use a cost model
- Missing ablation: TASO with measured instead of estimated cost

3 Generalisation by design?

- Trained and evaluated on the same graph; only input shapes vary
- In the code, the policy reads 500 fixed candidate slots: shapes transfer, new models don't

4 Coarse, noisy reward

- Latency measured every 5 steps; the steps in between all get +0.1, helpful or not
- Agent can't tell which of 5 rewrites helped; N and +0.1 never ablated

For the room

- When would you pick X-RLflow, and when TENSAT?
- What could each borrow from the other?