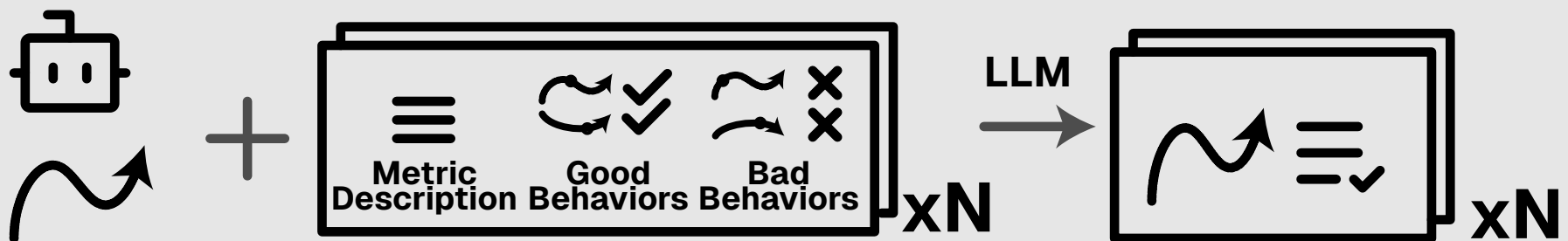


EVALUATION STEP 1

LLM-as-a-Judge

The extracted metrics can be used for prompting LLMs to produce ratings (+/-/NA) for each metric. Positive metrics → positive **traits**, negative metrics → negative **traits**.



Evaluate each trajectory
with induced **metrics**

Traits