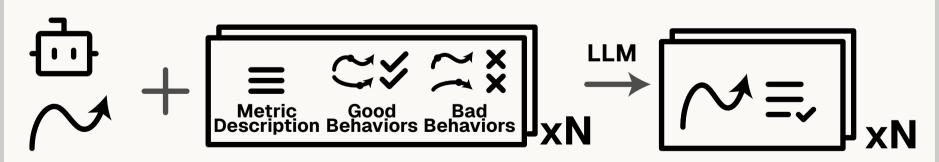
**EVALUATION** STEP 1

## **Evaluate Agents**

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



Evaluate each trajectory with induced **metrics** 

**Traits**