Cross-Domain Robustness

Evaluation Framework

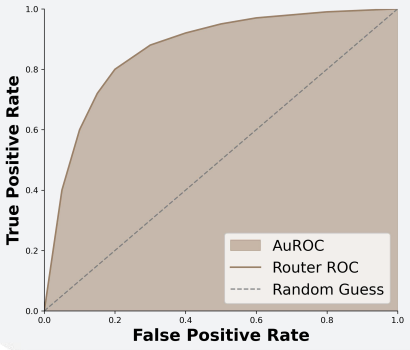
In-Domain

Alpaca | BigMath | MMLU

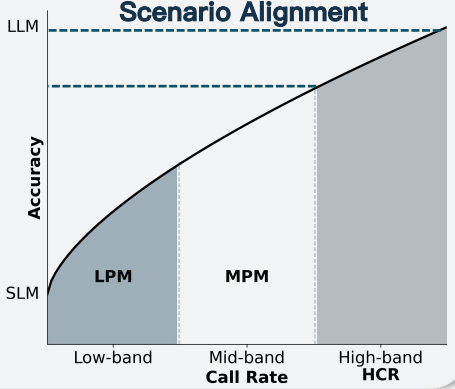
Out-of-Domain

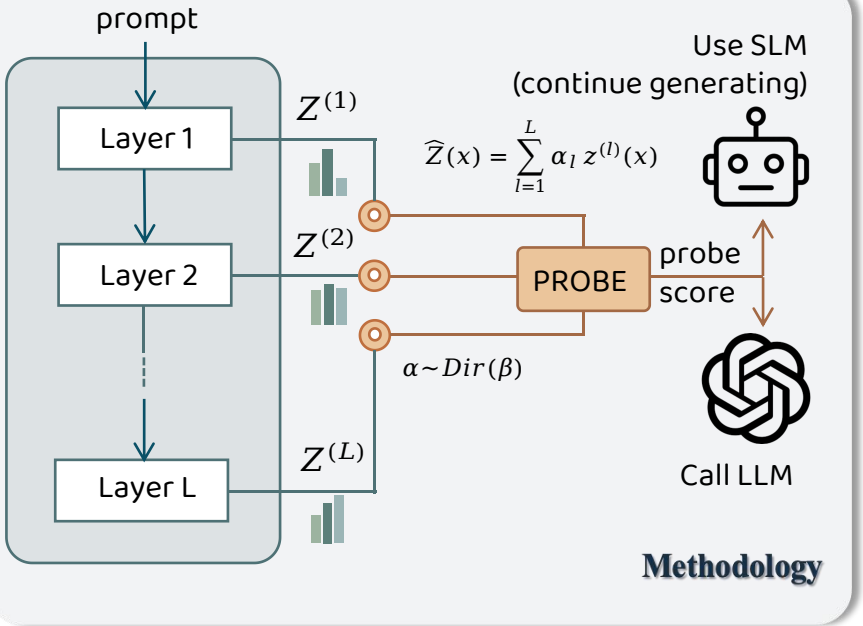
Magpie | Math | MMLU Pro

Router Ability



Scenario Alignment





Methodology