Baseline LLM Performance Without Grounding
Multiple-Choice Question Answering on River Dataset

- - - Random chance (20%)

n=12,174

n=4,208

n=7,839

50.1%

42.0%

16.7%

Accuracy (%)

Model

Gemma 3-4B
Instruct

Claude Sonnet
4.5

Gemini 2.5
Flash Lite