Qwen2.5-Math-7B | First layer where model's predicted token becomes Top-1 (Shown: 19-28) 3-digit additions | Repetitions: 5 Never Top-1: 0.0 ± 0.0 (SE) 1000.0 1000 800 600 400 200 Avg. 'Never Top-1': 0.0 ± 0.0 (SE) over 5 runs. 22 23 24 19 20 21 25 26 27 28 Layer Index (Range shown: 19 to 28)

over 5 runs)

±SE

Samples (First Top-1 in this layer,

Avg.#