

Figure 1 displays two line plots showing the performance of different models on the LLaMA dataset. The top plot is for the '1' model and the bottom plot is for the '3' model. Both plots show accuracy (y-axis) versus the number of samples (x-axis: 8M, 10M, 12M, 14M). The red line with circles represents the 'Ours' model, the green line with triangles represents the 'Ours + GPT-4' model, the blue line with squares represents the 'Ours + GPT-4 + GPT-4o' model, and the purple line with pluses represents the 'Ours + GPT-4 + GPT-4o + GPT-4o' model. In both plots, the 'Ours' model shows the highest accuracy, followed by 'Ours + GPT-4', 'Ours + GPT-4 + GPT-4o', and 'Ours + GPT-4 + GPT-4o + GPT-4o'.

