Tuned Lens Accuracy Across Layers of EleutherAl/pythia-410m-deduped Probability of token '44' for question 1+43 90% Probability of token '20' for question 3+17 Probability of token '6' for question 3+3 80% Probability of token '26' for question 5+21 Probability of token '30' for question 26+4 70% Probability of token '57' for question 9+48 Probability (%) 60% 50% 40% Probability of token '47' for question 44+3 Probability of token '57' for question 12+45 Probability of token '43' for question 15+28 Probability of token '77' for question 29+48 Average Probability Across 2499 prompts 30% 20% 10% 0% **Block Number**