Arxiv, GPT-2, context 128, mem 2 Trained on 1.15 ..■. 1 seg 2 sea - 3 seg 4 seg 5 seq 1.10 O 1.05 1.00 0.95 124 8 16 32 Evaluated on, segments