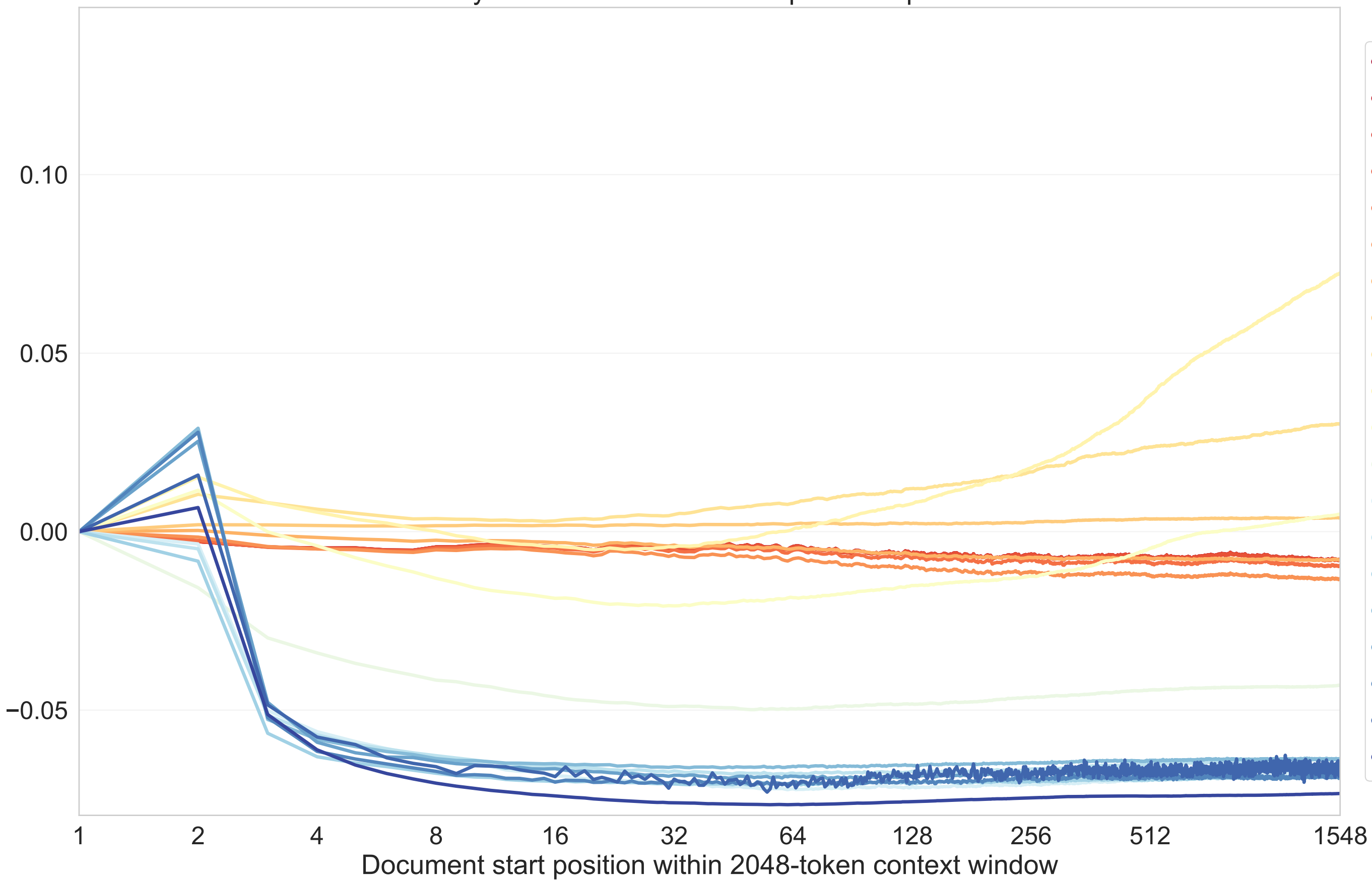


Pythia 31M: 264 docs used per checkpoint

Δ NLL relative to value at beginning of context window



- Step 0 ($NLL_{base} = 11.01$)
- Step 1 ($NLL_{base} = 11.01$)
- Step 2 ($NLL_{base} = 11.01$)
- Step 4 ($NLL_{base} = 11.01$)
- Step 8 ($NLL_{base} = 10.95$)
- Step 16 ($NLL_{base} = 10.77$)
- Step 32 ($NLL_{base} = 10.32$)
- Step 64 ($NLL_{base} = 9.76$)
- Step 128 ($NLL_{base} = 8.57$)
- Step 256 ($NLL_{base} = 6.97$)
- Step 512 ($NLL_{base} = 5.85$)
- Step 1000 ($NLL_{base} = 4.97$)
- Step 2000 ($NLL_{base} = 4.16$)
- Step 4000 ($NLL_{base} = 3.88$)
- Step 8000 ($NLL_{base} = 3.75$)
- Step 16000 ($NLL_{base} = 3.66$)
- Step 32000 ($NLL_{base} = 3.62$)
- Step 64000 ($NLL_{base} = 3.60$)
- Step 128000 ($NLL_{base} = 3.68$)
- Step 143000 ($NLL_{base} = 3.61$)