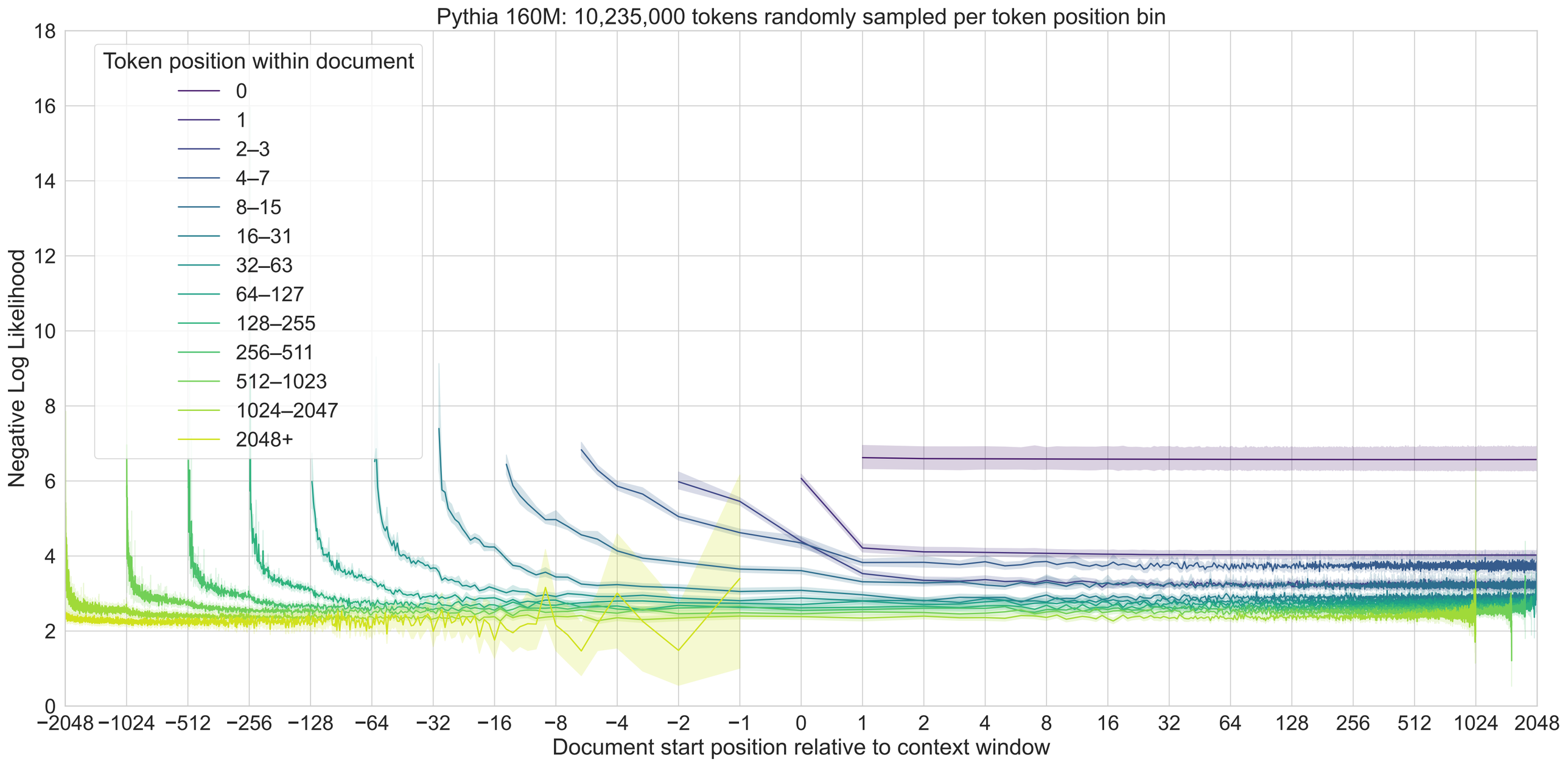
Pythia 160M: 10,235,000 tokens randomly sampled per token position bin
Token position within document
0
1
2–3
4–7
8–15
16–31
32–63
64–127
128–255
256–511
512–1023
1024–2047
2048+
Negative Log Likelihood
0
2
4
6
8
10
12
14
16
18
−2048
−1024
−512
−256
−128
−64
−32
−16
−8
−4
−2
−1
0
1
2
4
8
16
32
64
128
256
512
1024
2048
Document start position relative to context window