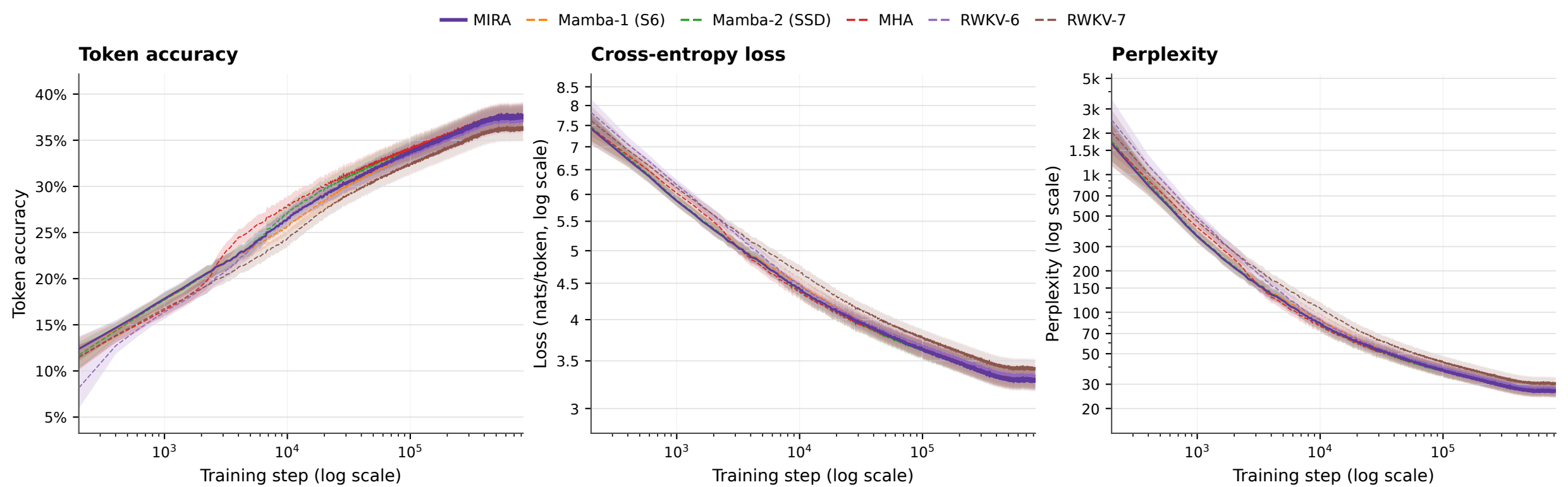

MIRA
Mamba-1 (S6)
Mamba-2 (SSD)
MHA
RWKV-6
RWKV-7
Token accuracy
Token accuracy
40%
35%
30%
25%
20%
15%
10%
5%
10^3
10^4
10^5
Training step (log scale)
Cross-entropy loss
Loss (nats/token, log scale)
8.5
8
7.5
7
6.5
6
5.5
5
4.5
4
3.5
3
10^3
10^4
10^5
Training step (log scale)
Perplexity
Perplexity (log scale)
5k
3k
2k
1.5k
1k
700
500
300
200
150
100
70
50
30
20
10^3
10^4
10^5
Training step (log scale)