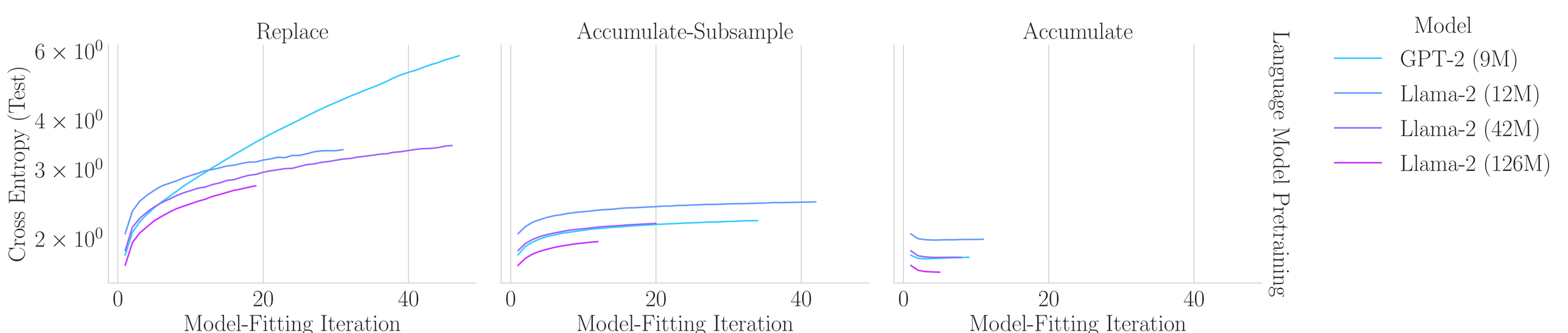
Replace
Accumulate-Subsample
Accumulate
Cross Entropy (Test)
6×10^0
4×10^0
3×10^0
2×10^0
0
20
40
Model-Fitting Iteration
Language Model Pretraining
Model
GPT-2 (9M)
Llama-2 (12M)
Llama-2 (42M)
Llama-2 (126M)