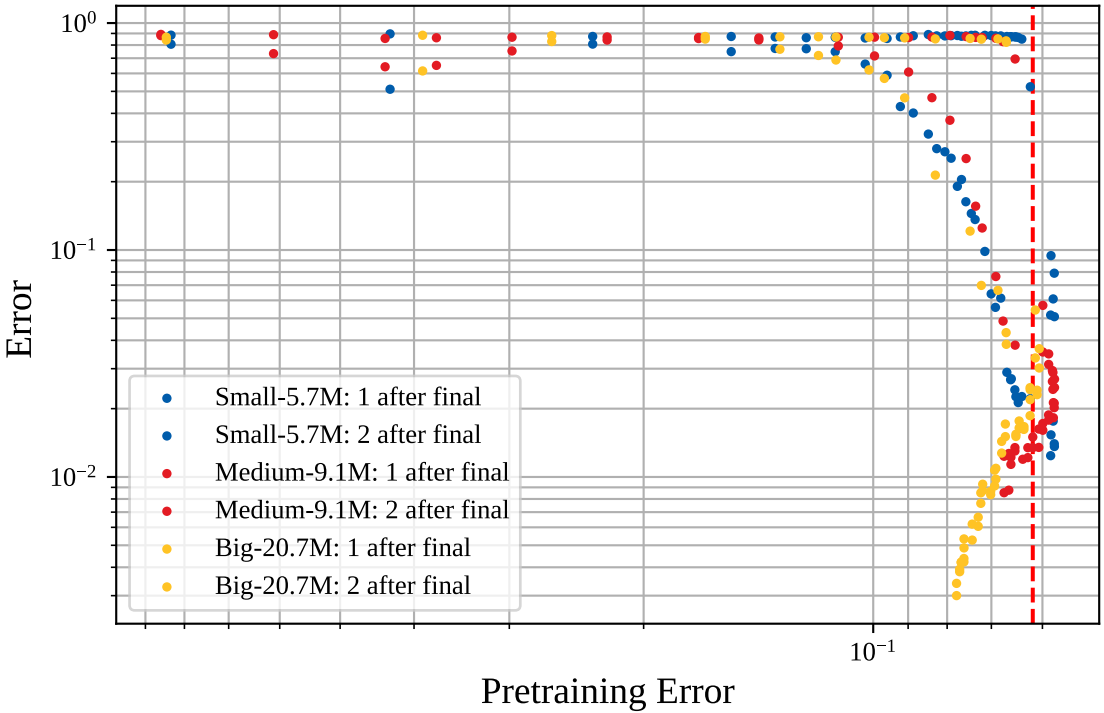
Error
Pretraining Error
10^0
10^{-1}
10^{-2}
10^{-1}
Small-5.7M: 1 after final
Small-5.7M: 2 after final
Medium-9.1M: 1 after final
Medium-9.1M: 2 after final
Big-20.7M: 1 after final
Big-20.7M: 2 after final