Data collected by Kartik Hegde (kvhegde2@illinois.edu). Training FLOPS for transformers is based on Narayanan et al. "Efficient large-scale language model training on gpu clusters using megatron-lm". In SC'22, for others it is calculated as FLOPS/forward pass* #dataset * #epochs * 3.