

The trend of (E1 - E2) across different pretraining data sizes

2.5M -> 21M

9.8M -> 21M

2.5M -> 38M

9.8M -> 38M

21M -> 38M

