

<i>h/t Lukas Finvedden for help with this.</i>							
Training FLOP (physical FLOP)							
3.00E+27							
k							
1.5	If there are p parameters, a forward pass takes k*p FLOP. Bio Anchors uses k=1.5						
Train model for mp tokens							
20	Chinchilla						
Horizon length							
1.5	Longer horizons increase training FLOP without increasing runtime FLOP/s.						
p							
4.71E+12	Training takes 3hkm ^p FLOP. [Why the factor of 3? Training takes (FLOP for forward pass)+(FLOP for backward pass)~3x FLOP for forward pass]						
Runtime FLOP/forward-pass of AI							
7.07E+12							
Duration of training (seconds)							
1.00E+07	4 months						
How many tokens per second, by reallocating training compute to inference?							
4.24E+07							
Token per second, per task							
10							
Number of tasks in parallel							
4.24E+06							

