h/t Lukas Finved	den for help with thi	S.								
Training FLOP (p	hysical FLOP)									
3.00E+25										
k										
1.5	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	5 Longer horizons increase training FLOP without increasing runtime FLOP/s.									
р										
4.71E+11	4.71E+11 Training takes 3hkmp^2 FLOP. [Why the factor of 3? Training takes (FLOP for forward pass)+(FLOP for backward pass)~=3x FLOP for forward pass]									
Runtime FLOP/fc	orward-pass of AI									
7.07E+11										
Duration of training	ng (seconds)									
1.00E+07	4 months									
How many token	s per second, by rea	allocating training	compute to infe	rence?						
4.24E+06										
Token per second, per task										
10										
Number of tasks	in parallel									
4.24E+05										

h/t Lukas Finved	den for help with thi	s.								
Training FLOP (p	hysical FLOP)									
3.00E+27										
k										
1.5	i 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 in	crease training F	OP without incr	easing runtime FL	.OP/s.					
p										
4.71E+12	4.71E+12 Training takes 3hkmp^2 FLOP. [Why the factor of 3? Training takes (FLOP for forward pass)+(FLOP for backward pass)~=3x FLOP for forward pass]									
Runtime FLOP/fc	orward-pass of Al									
7.07E+12										
Duration of training	ng (seconds)									
1.00E+07	4 months									
How many token	s per second, by re	allocating training	compute to infe	rence?						
4.24E+07										
Token per second, per task										
10										
Number of tasks	in parallel									
4.24E+06										

h/t Lukas Finved	den for help with thi	s.								
Training FLOP (p	hysical FLOP)									
3.00E+29										
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 in	crease training F	OP without incr	easing runtime FL	.OP/s.					
p										
4.71E+13	4.71E+13 Training takes 3hkmp^2 FLOP. [Why the factor of 3? Training takes (FLOP for forward pass)+(FLOP for backward pass)~=3x FLOP for forward pass]									
Runtime FLOP/fc	orward-pass of Al									
7.07E+13										
Duration of training	ng (seconds)									
1.00E+07	4 months									
How many token	s per second, by re	allocating training	compute to infe	rence?						
4.24E+08										
Token per second, per task										
10										
Number of tasks	in parallel									
4.24E+07										

h/t Lukas Finved	den for help with th	nis.								
Training FLOP (physical FLOP)										
3 00E+33	nyolour Lor /									
3.00E+33										
К										
1.5	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 i	ncrease training F	LOP without incr	easing runtime FL	.OP/s.					
р										
4.71E+15	4.71E+15 Training takes 3hkmp^2 FLOP. [Why the factor of 3? Training takes (FLOP for forward pass)+(FLOP for backward pass)~=3x FLOP for forward pass]									
Runtime FLOP/fc	orward-pass of AI									
7.07E+15										
Duration of training	ng (seconds)									
1.00E+07	4 months									
How many token	How many tokens per second, by reallocating training compute to inference?									
4.24E+10										
Token per second, per task										
10										
Number of tasks	in parallel									
4.24E+09										