

1e6

Learning Trajectory of the 582M ByT5 Model

Total Training Examples Seen

1.50
1.25
1.00
0.75
0.50
0.25
0.00

3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28

Addition Problem Length

