LiLaVe scores accuracy across layers and tokens 1.0 -algebra\_linear\_1d 0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -1.0 -0.9 -GSM-Symbolic 0.8 -0.7 -0.6 -0.5 -0.4 -1.0 layer GSM-Symbolic-p2 0.9 --1 0.8 --2 0.7 --4 0.6 --8 0.5 --16 0.4 -1.0 -0.9 -0.8 -GSM8K 0.7 -0.6 -0.5 -0.4 -1.0 -0.9 -0.8 -MATH 0.7 -0.6 -

0.5 **-**

20

15

10

5

0

25

30

token index

-26