trained-from-scratch | compressed

sorting

$F = N$

feature count F

residual stream dimension N

- × accuracy < 1
- ● layer 1
- ● layer 2
- ● layer 3