cosine T-SNE visualisations  $(albert-xlarge-v2\_wikitext\_\langle x,y\rangle\_trained-params\_dense-on\_24-layers)$ layer 0 layer 3 layer 1 layer 2 100 100 100 100 50 50 50 50 0 0 0 -50**-**50 **-**50 -50 -100-100-100-1000 50 100 0 50 100 0 50 100 0 50 100 -50 -100-50 -50-50 -100 -100-100layer 7 layer 4 layer 5 layer 6 100 100 100 100 50 50 50 50 0 0 0 -50 **-**50 -50 -50 -100 -100-100-50 0 50 100 -100 -50 0 50 100 -100-50 0 50 100 -50 0 50 100 -100 -100layer 9 layer 10 layer 8 layer 11 100 100 100 100 50 50 50 50 0 0 0 0 -50-50-50 -50-100-100 -100-1000 50 100 50 100 0 50 0 50 100 -100-50 -100-50 0 -100-50 100 -100-50 layer 12 layer 13 layer 14 layer 15 100 100 100 100 50 50 50 50 0 0 0 0 -50 **-**50 -50-50-100 -100 -100-100 -50 -50 -50 0 50 -50 0 -1000 50 -1000 50 100 -100100 -10050 100 100 layer 16 layer 17 layer 18 layer 19 100 100 100 100 50 50 50 50 0 0 0 0 <del>-</del>50 · <del>-50</del> --50 -<del>-</del>50 --100 --100 <del>|</del> -100 -100 -100 Ó -50 0 50 100 -100 -50 0 50 100 -50 50 100 -50 0 50 100 -100 -100 layer 20 layer 22 layer 21 layer 23 100 100 100 100 50 50 50 50 0 0 0 0 -50-50 -50 -50 -100 --100 --100 --100 --100 -50 0 50 100 -100-50 50 100 -100-50 0 50 100 -100 -50 0 50 100