cosine T-SNE visualisations (albert-xlarge-v2\_wikitext\_(x,y)\_random-params\_dense-on\_24-layers) layer 0 layer 1 layer 2 layer 3 100 100 100 100 50 50 50 50 0 0 -50-50 -50-50-100-100-1000 50 100 50 100 0 50 100 0 50 100 -50 -100 -500 -50-50 -100 -100-100 layer 4 layer 5 layer 6 layer 7 100 100 100 100 50 50 50 50 0 -50-50 -50 -50 -100 -100-50 0 50 100 -100-500 50 100 -50 0 50 100 -50 50 100 -100-100-100layer 8 layer 10 layer 11 layer 9 100 100 100 100 50 50 50 50 0 0 0 -50-50 -50-100 -100 -10050 100 -10050 100 50 100 50 100 -100-50 -50-100-50-100-50layer 12 layer 13 layer 15 layer 14 100 100 100 100 50 50 50 50 0 -50-50 -50-50-100-100-100-100 50 -100 -50 50 -50 50 -50 0 50 -100 -500 100 0 100 -100 0 100 -100100 layer 16 layer 19 layer 17 layer 18 100 100 100 100 50 50 50 50 0 -50 -50 <del>-</del>50 --50-100 <del>|</del> -100 -100 <del>|</del> -100 -100 --100 -50 0 50 100 -100 -50 0 50 100 -50 0 50 100 -50 0 50 100 -100 layer 20 layer 22 layer 21 layer 23 100 100 100 100 50 50 50 50 0 0 0 0 -50-50 -50 -100 --100 --100 --100-100 -50 0 50 100 -100**-**50 50 100 -100-50 50 100 -100 50 100

-50