

cosine T-SNE visualisation - layer 6
(albert-xlarge-v2 wikitext {Qx,Ky} trained-params dense-off 24-layers)

