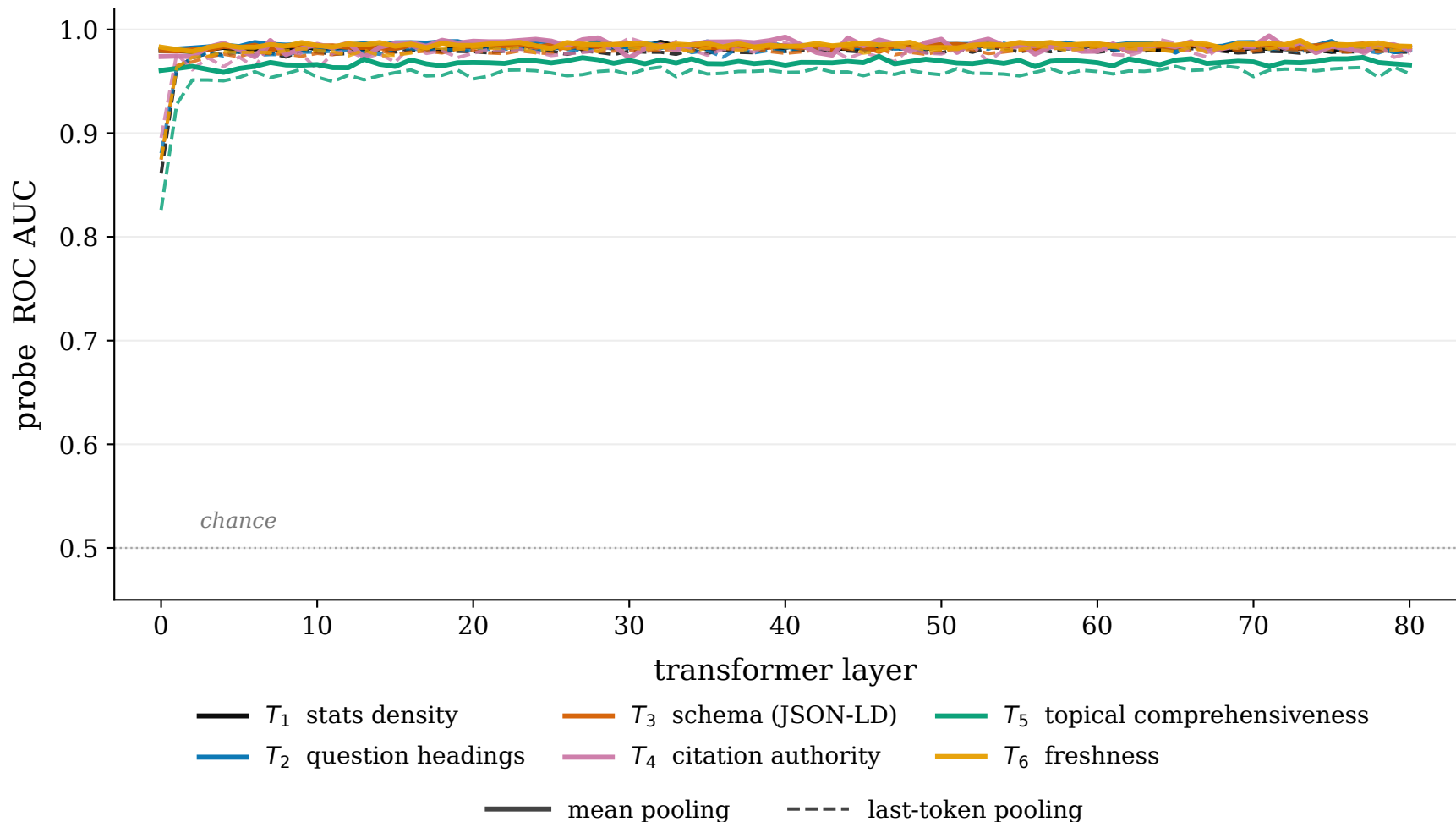


Pooling comparison (mean vs. last-token, per treatment)



Mean pooling (solid) averages hidden states over all tokens of the page; last-token pooling (dashed) takes the final token only. Mean pooling near-saturates from layer 0; last-token requires several layers to absorb context before reaching parity.