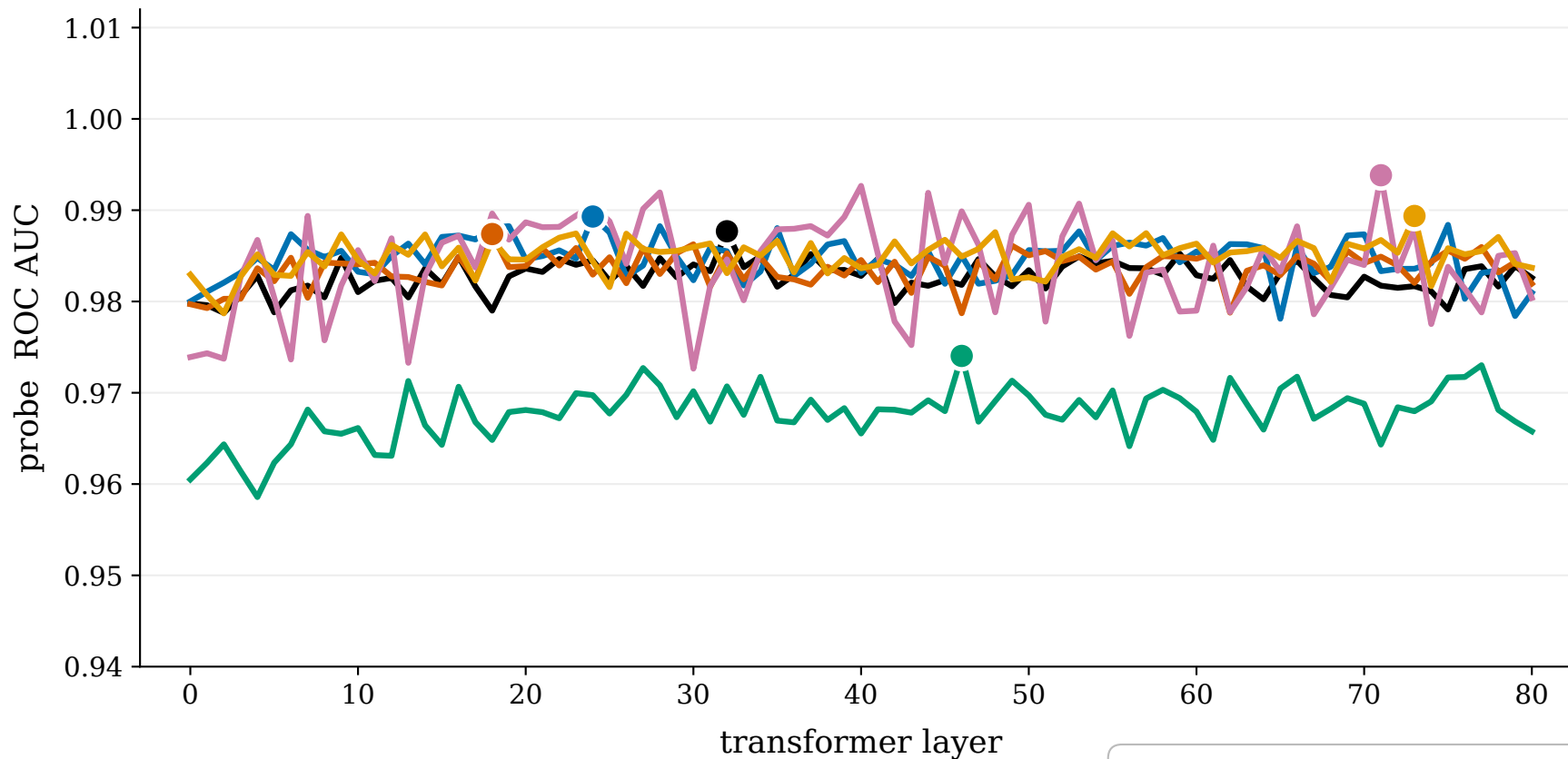


Layer-wise probing accuracy by treatment (mean pooling)



- T_1 stats density
- T_2 question headings
- T_3 schema (JSON-LD)
- T_4 citation authority
- T_5 topical comprehensiveness
- T_6 freshness

peak ROC AUC (mean pooling)			
T_4	L71	0.994	
T_6	L73	0.989	
T_2	L24	0.989	
T_1	L32	0.988	
T_3	L18	0.987	
T_5	L46	0.974	

Linear probe (logistic regression on frozen hidden states) per (treatment, layer); 80/20 stratified split. Llama-3.3-70B + Qwen-2.5-72B; averaged across 4 prompt variants. Shaded band: min-max envelope across variants (invisibly narrow — variant has no measurable effect).