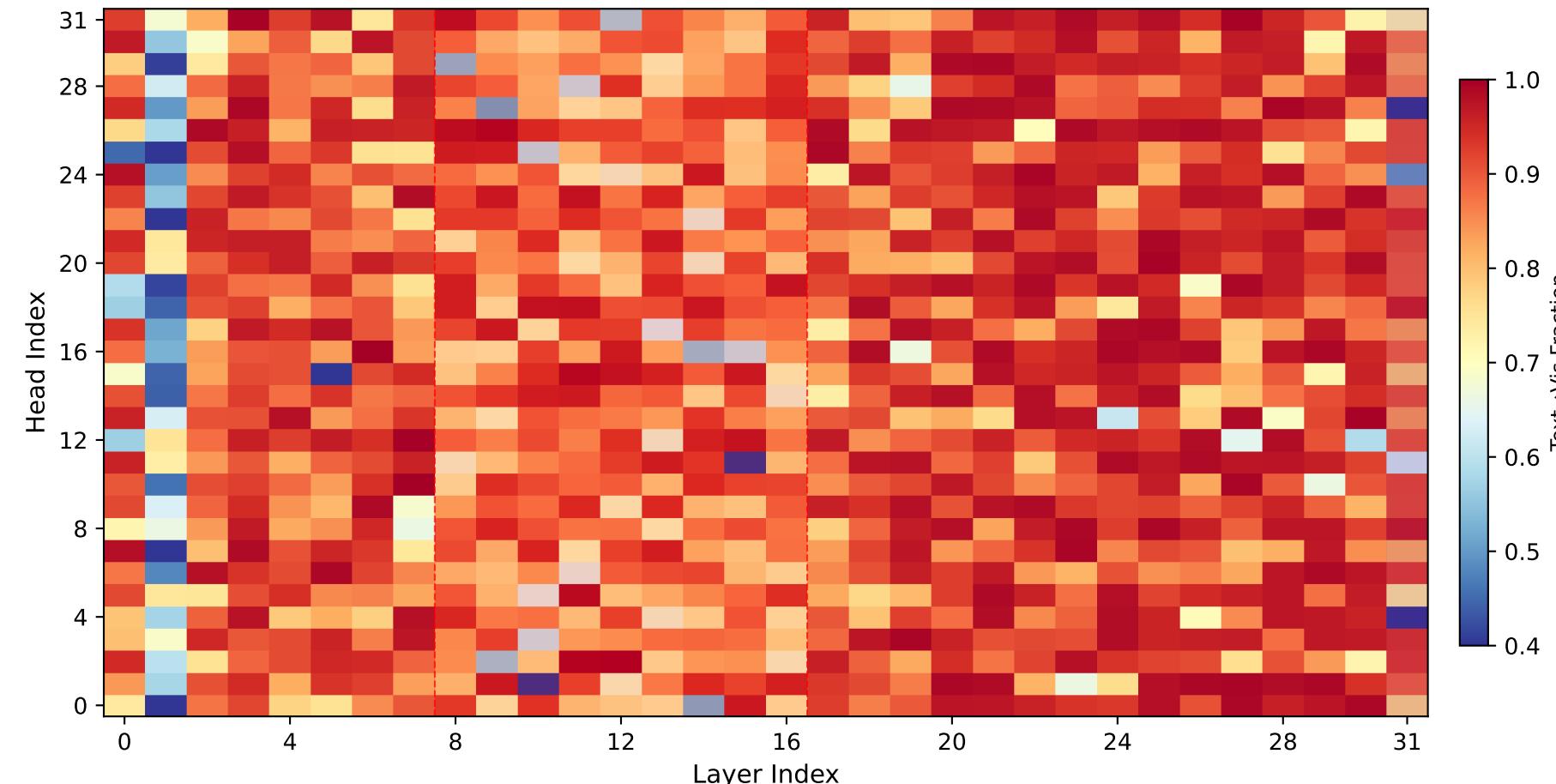**(a) Per-Head Text→Vis Attention Fraction**

(b) Layer Means ± Head Std