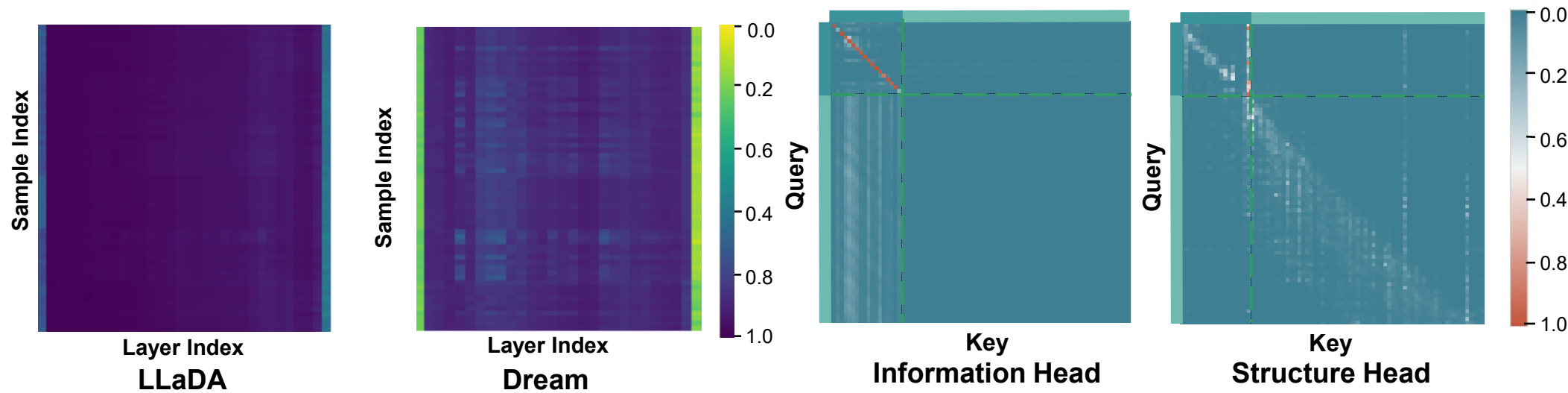


(a) Attention map across steps



(b) Cosine similarity across layers

(c) Attention map across heads