

Avg Attention Head Flow — variable_value corruption (Top 10 heads by $|\text{marginal } \Delta p|$)

