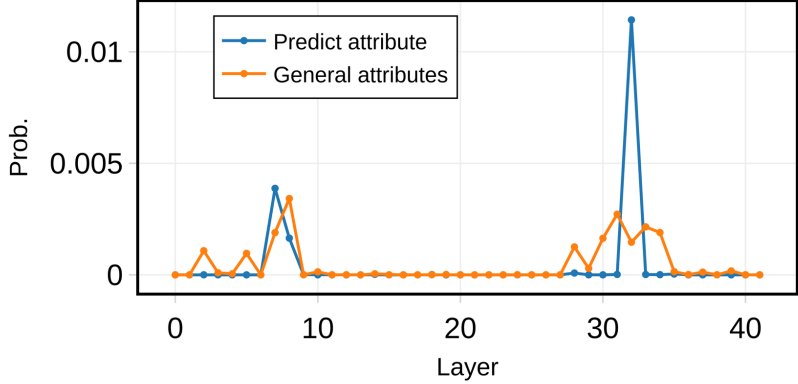
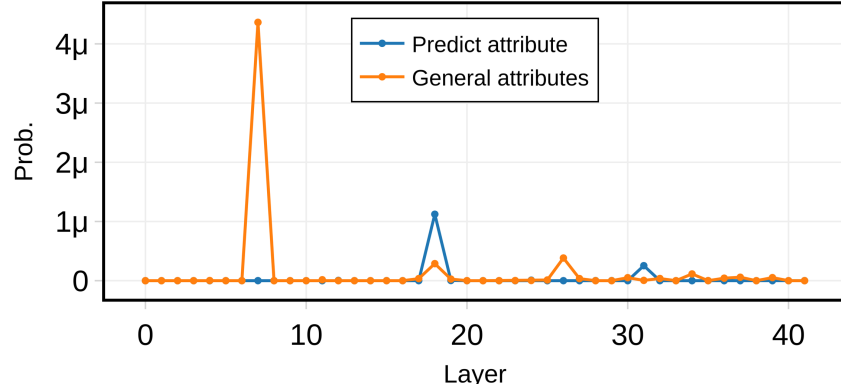
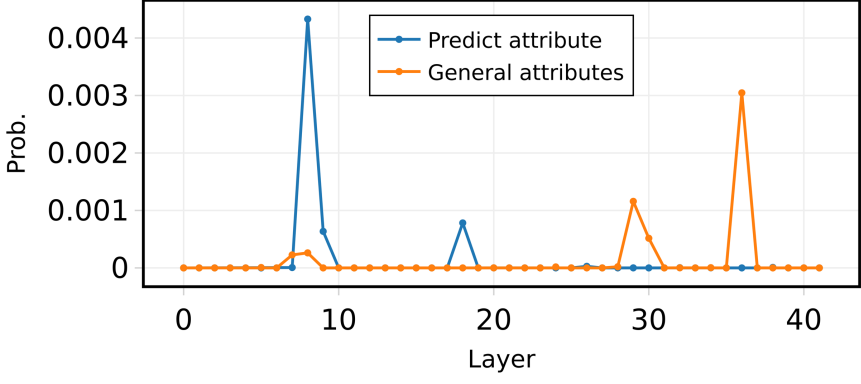
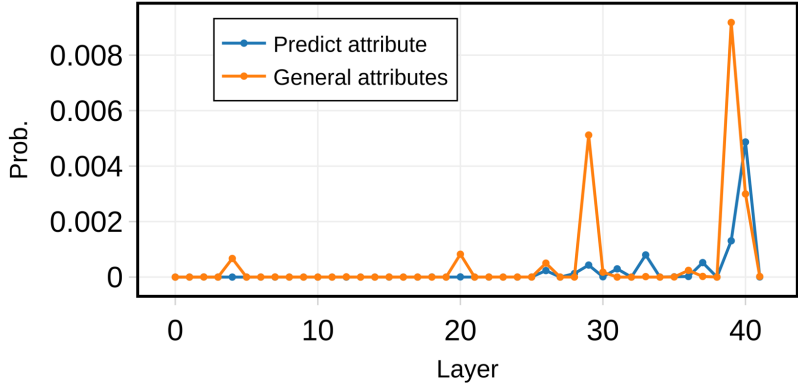
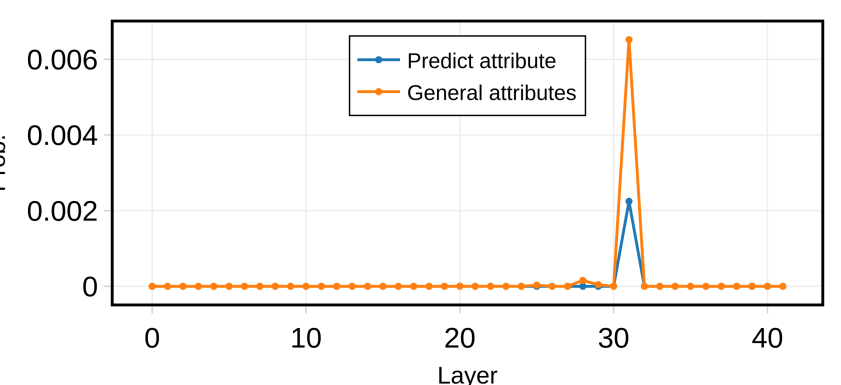
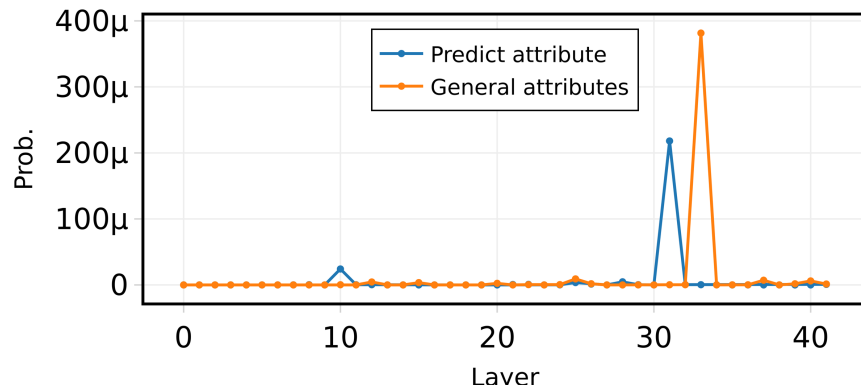
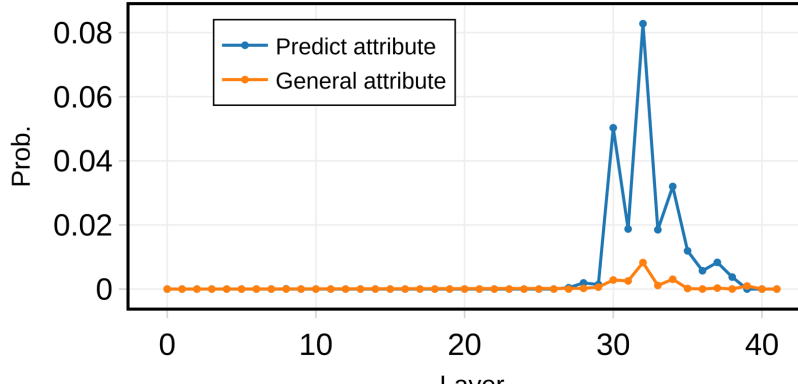
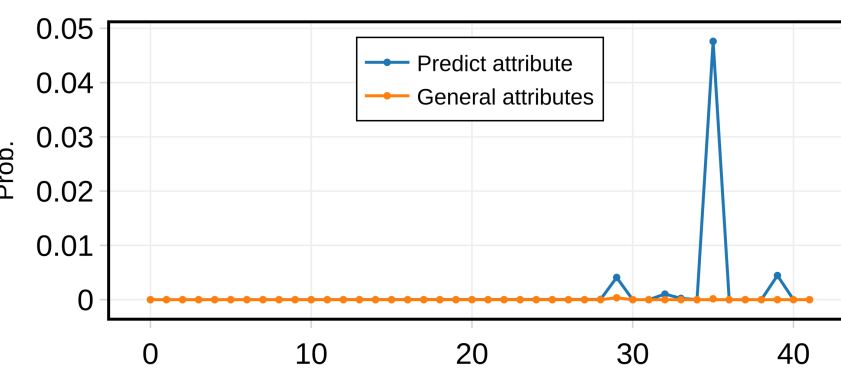
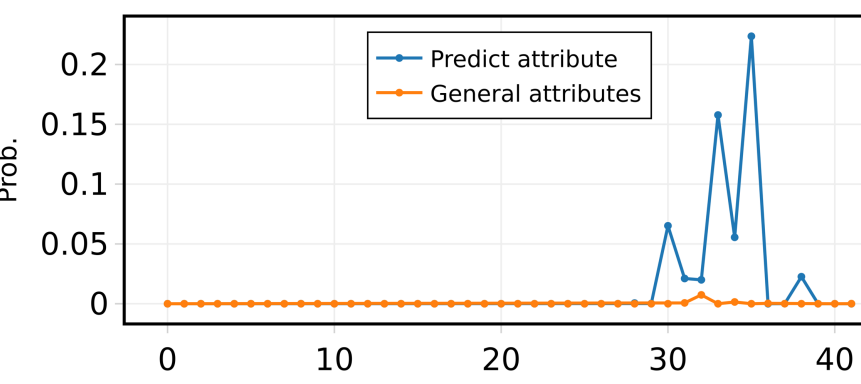
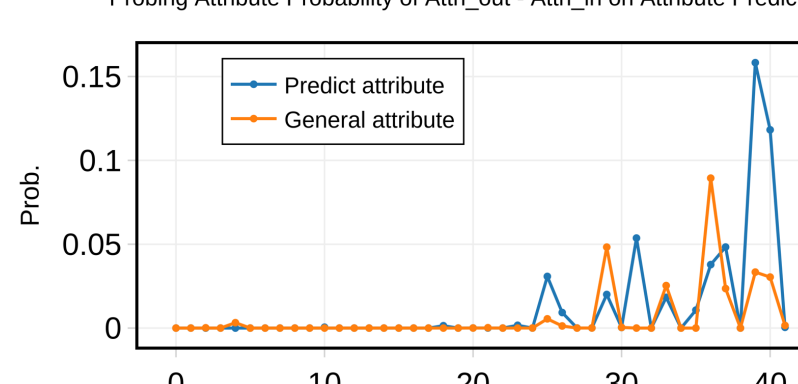
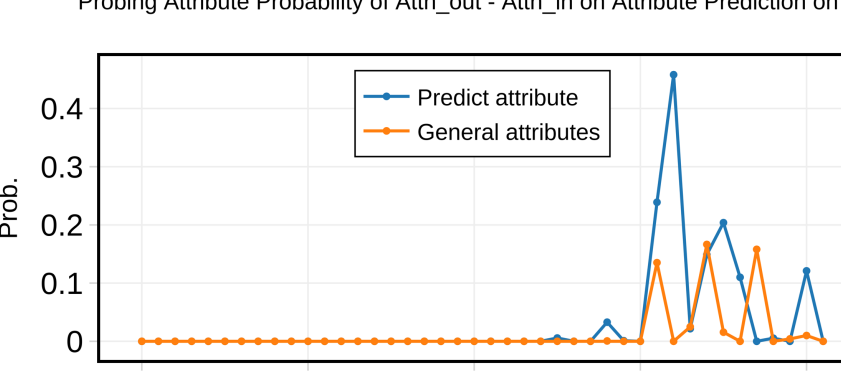
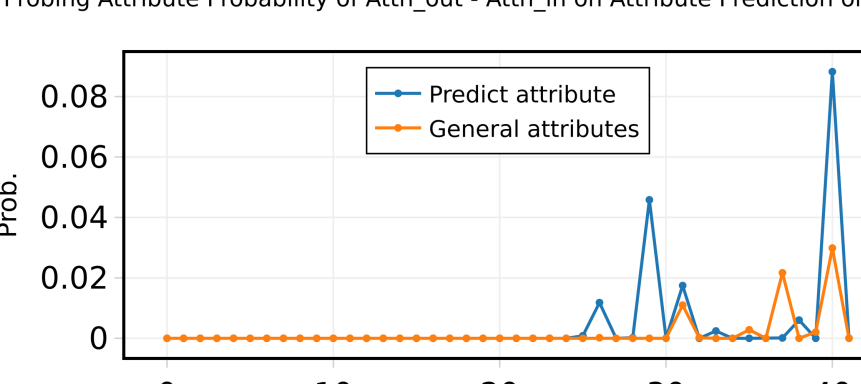
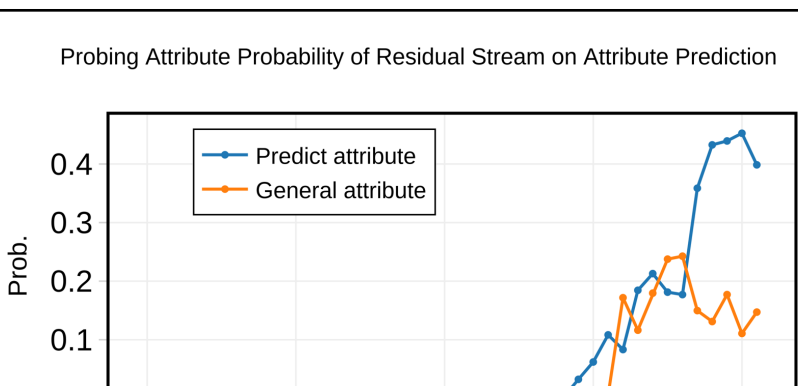
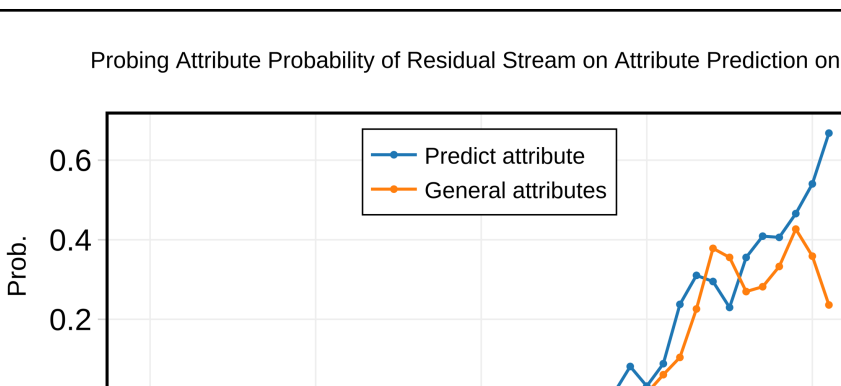


Stage	Probing Module	Dataset 1: Strategyqa	Dataset 2: SocialIQA	Dataset 3: Winogrande
Knowledge Augmentation	MLP	Probing Attribute Probability of MLP_out - MLP_in on Concept 	Probing Attribute Probability of MLP_out - MLP_in on Concept on siqa 	Probing Attribute Probability of MLP_out - MLP_in on Concept on winogrande 
	Attention	Probing Attribute Probability of Attn_out - Attn_in on Concept 	Probing Attribute Probability of Attn_out - Attn_in on Concept on siqa 	Probing Attribute Probability of Attn_out - Attn_in on Concept on winogrande 
Knowledge Retrieval	MLP	Probing Attribute Probability of MLP_out - MLP_in on Attribute Prediction 	Probing Attribute Probability of MLP_out - MLP_in on Attribute Prediction on siqa 	Probing Attribute Probability of MLP_out - MLP_in on Attribute Prediction on winogrande 
	Attention	Probing Attribute Probability of Attn_out - Attn_in on Attribute Prediction 	Probing Attribute Probability of Attn_out - Attn_in on Attribute Prediction on siqa 	Probing Attribute Probability of Attn_out - Attn_in on Attribute Prediction on winogrande 
Re-ranking	Residual Stream	Probing Attribute Probability of Residual Stream on Attribute Prediction 	Probing Attribute Probability of Residual Stream on Attribute Prediction on siqa 	Probing Attribute Probability of Residual Stream on Attribute Prediction on winogrande 