	POPE(MSCOCO)								
	Random			Popular			Adversarial		
InstructBLIP	Accuracy	Precision	F1 Score	Accuracy	Precision	F1 Score	Accuracy	Precision	F1 Score
(a)	59.50	62.90	53.36	62.83	69.15	55.49	62.50	68.47	55.27
(b)	87.83	84.00	88.48	82.10	76.15	83.93	75.50	68.76	79.23
(c)	88.20	91.28	87.74	82.83	81.79	83.11	79.13	76.33	80.19
(d)	88.80	88.19	88.89	84.00	80.58	84.85	78.60	73.44	80.72
(e)	**90.00**	**92.80**	**89.66**	**86.10**	**85.65**	**86.19**	**81.40**	**78.37**	**82.34**

	POPE(A-OKVQA)								
	Random			Popular			Adversarial		
InstructBLIP	Accuracy	Precision	F1 Score	Accuracy	Precision	F1 Score	Accuracy	Precision	F1 Score
(a)	60.37	64.37	53.68	70.23	**89.36**	60.68	67.53	**80.87**	58.59
(b)	85.23	78.68	86.75	79.23	71.68	82.32	69.57	62.69	76.06
(c)	88.57	**86.87**	88.82	82.47	77.80	83.83	73.67	67.61	77.53
(d)	87.40	82.96	88.19	80.90	74.43	83.13	71.43	64.74	76.72
(e)	**88.87**	86.12	**89.27**	**83.33**	78.09	**84.76**	**74.63**	68.10	**78.51**

	POPE(GQA)								
	Random			Popular			Adversarial		
InstructBLIP	Accuracy	Precision	F1 Score	Accuracy	Precision	F1 Score	Accuracy	Precision	F1 Score
(a)	59.20	62.87	52.41	70.70	**92.71**	60.53	66.50	**79.02**	57.29
(b)	83.47	77.92	84.96	72.17	65.56	77.04	67.00	61.13	73.89
(c)	85.93	**85.65**	85.99	77.67	73.58	79.45	72.27	67.38	75.69
(d)	84.70	81.45	85.45	75.47	69.77	78.55	70.33	64.62	75.18
(e)	**86.03**	84.76	**86.28**	**78.33**	73.80	**80.22**	**73.47**	68.22	**76.81**

	Amber			
InstructBLIP	CHAIR ↓	Cover ↑	Hal ↓	Cog ↓
(a)	18.4	**53.4**	57.3	6.6
(b)	19.3	48.0	57.3	7.6
(e)	**10.4**	43.4	**26.1**	**2.6**

	MME				
InstructBLIP	Existence	count	Position	Color	Total
(a)	185.00	88.33	67.72	110.00	451.05
(b)	180.00	62.54	**69.70**	111.19	423.43
(e)	**191.64**	**102.77**	45.61	**140.00**	**480.02**

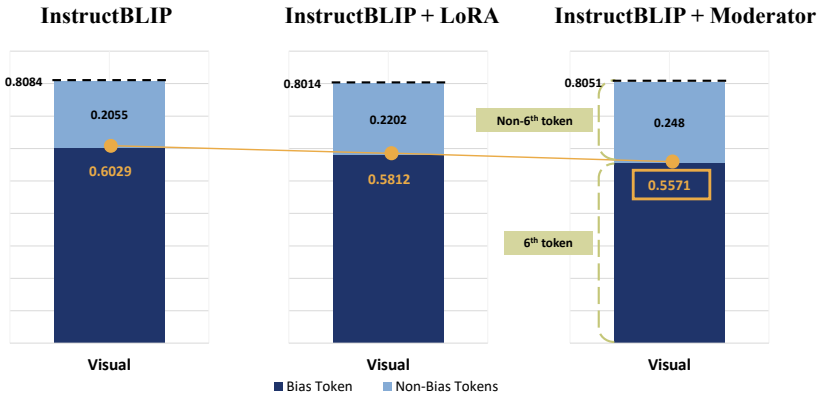

Figure 1: **Comparison of attention distribution** of the baseline, applying LoRA on the baseline and applying Moderator on the baseline.