Config	S_{instr}	Contr	S_{res}	Magn	Mult	Syn	Anti	S_{loc}	V_0	Gener	A_0	Adv_0	Sing	Redun	S_0
ChatGPT + k:0-s:0 +EM	27	64	32	26	15	39	86	46	87	45	91	94	50	4	92
ChatGPT + k:0-s:1 + EM	23	72	28	14	1	32	84	44	87	38	88	95	61	4	91
ChatGPT + k:1-s:0 + EM	56	62	42	7	4	43	83	59	83	48	91	92	44	13	90
ChatGPT + k:1-s:1+ EM	37	65	39	1	1	29	84	53	87	39	87	94	54	13	95
Llama3.0 + k:0-s:0 + EM	0	44	28	0	0	0	65	19	75	0	74	71	18	0	69
Llama3.0 + k:0-s:1 + EM	0	48	22	0	0	0	68	25	71	0	71	79	21	0	57
Llama3.0 + k:1-s:0 + EM	0	38	23	0	0	0	61	13	64	0	65	67	0	0	75
Llama3.0 + k:1-s:1 + EM	0	33	16	0	0	0	52	6	57	0	45	35	0	0	40
Llama3.1 + k:0-s:0 + EM	12	56	33	3	0	15	67	32	79	27	80	82	20	0	71
Llama3.1 + k:0-s:1 + EM	0	49	35	0	0	1	69	32	70	1	71	77	25	0	54
Llama3.1 + k:1-s:0 + EM	0	51	39	0	0	0	72	19	69	0	68	77	0	0	77
Llama3.1 + k:1-s:1 + EM	0	51	31	0	0	0	67	11	69	0	69	81	2	0	71
Qwen + k:0-s:0 + EM	28	54	43	13	12	29	79	33	81	42	85	87	42	0	89
Qwen + k:0-s:1 + EM	2	53	27	1	0	2	77	28	79	2	87	83	44	0	83
Qwen + k:1-s:0 + EM	35	51	46	7	6	24	80	36	85	47	86	86	35	5	89
Qwen + k:1-s:1 + EM	3	49	24	1	0	2	76	30	83	2	83	85	38	0	84
ChatGPT + k:0-s:0 + CM	58	68	52	35	22	38	86	57	87	46	91	97	53	13	94
ChatGPT + k:0-s:1 + CM	65	77	70	45	38	53	84	64	87	63	88	98	65	29	92
ChatGPT + k:1-s:0 + CM	57	69	52	31	39	42	83	66	83	49	91	95	52	19	91
ChatGPT + k:1-s:1 + CM	56	71	62	42	46	53	86	64	88	57	88	97	62	37	96
Llama3.0 + k:0-s:0 + CM	40	62	77	33	35	32	73	53	74	46	84	75	45	5	75
Llama3.0 + k:0-s:1 + CM	50	64	68	39	36	41	78	55	75	58	82	85	57	32	70
Llama3.0 + k:1-s:0 + CM	47	60	70	20	35	21	73	56	77	46	79	90	55	3	88
Llama3.0 + k:1-s:1 + CM	44	65	80	31	39	36	80	62	83	61	86	94	58	16	91
Llama3.1 + k:0-s:0 + CM	45	63	77	31	32	26	70	61	79	46	84	86	48	6	77
Llama3.1 + k:0-s:1 + CM	54	69	79	36	28	46	78	63	76	59	86	85	62	33	71
Llama3.1 + k:1-s:0 + CM	45	60	78	24	32	16	76	60	80	55	80	86	52	16	90
Llama3.1 + k:1-s:1 + CM	47	62	82	34	40	43	80	60	83	59	88	88	63	26	93
Qwen + k:0-s:0 + CM	45	56	55	27	31	32	79	42	82	48	87	92	60	7	93
Qwen + k:0-s:1 + CM	67	66	75	43	57	56	85	57	82	68	90	89	68	35	92
Qwen + k:1-s:0 + CM	42	52	60	35	36	27	80	59	85	53	88	92	60	12	92
Qwen + k:1-s:1 + CM	66	67	79	60	58	59	84	61	85	72	86	92	71	43	94

Table 5: Exact Match (EM) and Contain Match (CM) scores (%) across all models and configurations.