

TASK	Weakly Supervised		Uses External Resources	Strong Supervision (using supporting facts)						
	<i>N-gram Classifier</i>	<i>LSTM</i>	<i>Structured SVM</i> <i>COREF+SRL features</i>	<i>MemNN</i> <i>Weston et al. (2014)</i>	<i>MemNN</i> <i>ADAPTIVE MEMORY</i>	<i>MemNN</i> <i>AM + N-GRAMS</i>	<i>MemNN</i> <i>AM + NONLINEAR</i>	<i>MemNN</i> <i>AM + NG + NL</i>	<i>No. of ex. req. ≥ 95</i>	<i>MultiTask Training</i>
1 - Single Supporting Fact	36	50	99	100	100	100	100	100	250 ex.	100
2 - Two Supporting Facts	2	20	74	100	100	100	100	100	500 ex.	100
3 - Three Supporting Facts	7	20	17	20	100	99	100	100	500 ex.	98
4 - Two Arg. Relations	50	61	98	71	69	100	73	100	500 ex.	80
5 - Three Arg. Relations	20	70	83	83	83	86	86	98	1000 ex.	99
6 - Yes/No Questions	49	48	99	47	52	53	100	100	500 ex.	100
7 - Counting	52	49	69	68	78	86	83	85	FAIL	86
8 - Lists/Sets	40	45	70	77	90	88	94	91	FAIL	93
9 - Simple Negation	62	64	100	65	71	63	100	100	500 ex.	100
10 - Indefinite Knowledge	45	44	99	59	57	54	97	98	1000 ex.	98
11 - Basic Coreference	29	72	100	100	100	100	100	100	250 ex.	100
12 - Conjunction	9	74	96	100	100	100	100	100	250 ex.	100
13 - Compound Coref.	26	94	99	100	100	100	100	100	250 ex.	100
14 - Time Reasoning	19	27	99	99	100	99	100	99	500 ex.	99
15 - Basic Deduction	20	21	96	74	73	100	77	100	100 ex.	100
16 - Basic Induction	43	23	24	27	100	100	100	100	100 ex.	94
17 - Positional Reasoning	46	51	61	54	46	49	57	65	FAIL	72
18 - Size Reasoning	52	52	62	57	50	74	54	95	1000 ex.	93
19 - Path Finding	0	8	49	0	9	3	15	36	FAIL	19
20 - Agent's Motivations	76	91	95	100	100	100	100	100	250 ex.	100
Mean Performance	34	49	79	75	79	83	87	93		92