
Recall-AI LongMemEval_S Benchmark Report

Single-system edition - Recall-AI only

System	Recall-AI v8.0.0
Benchmark	LongMemEval_S (ICLR 2025)
Questions evaluated	500 / 500
Report timestamp	2026-04-23 00:50:23
Scope	This PDF summarizes Recall-AI only. External product and baseline comparison sections are intentionally omitted.

This document is derived from the uploaded Recall-AI benchmark report and keeps only Recall-AI measurements, methodology, configuration, analysis, and raw per-question results.

1. Executive Summary

This report presents a single-system evaluation of Recall-AI v8.0.0 on LongMemEval_S. The benchmark measures long-term interactive memory capabilities including information extraction, multi-session reasoning, knowledge updates, temporal reasoning, user preference understanding, and abstention.

91.0% Overall Accuracy	91.9% Task-Averaged Accuracy	455 / 500 Correct Answers	500 / 500 Questions Evaluated
----------------------------------	--	-------------------------------------	---

Metric	Value
Abstention accuracy	76.7%
Mean retrieval latency	105,094 ms (105.1 s)
Mean total latency per question	298,963 ms (299.0 s)
Full benchmark runtime	321.8 min
Average turns per question	249

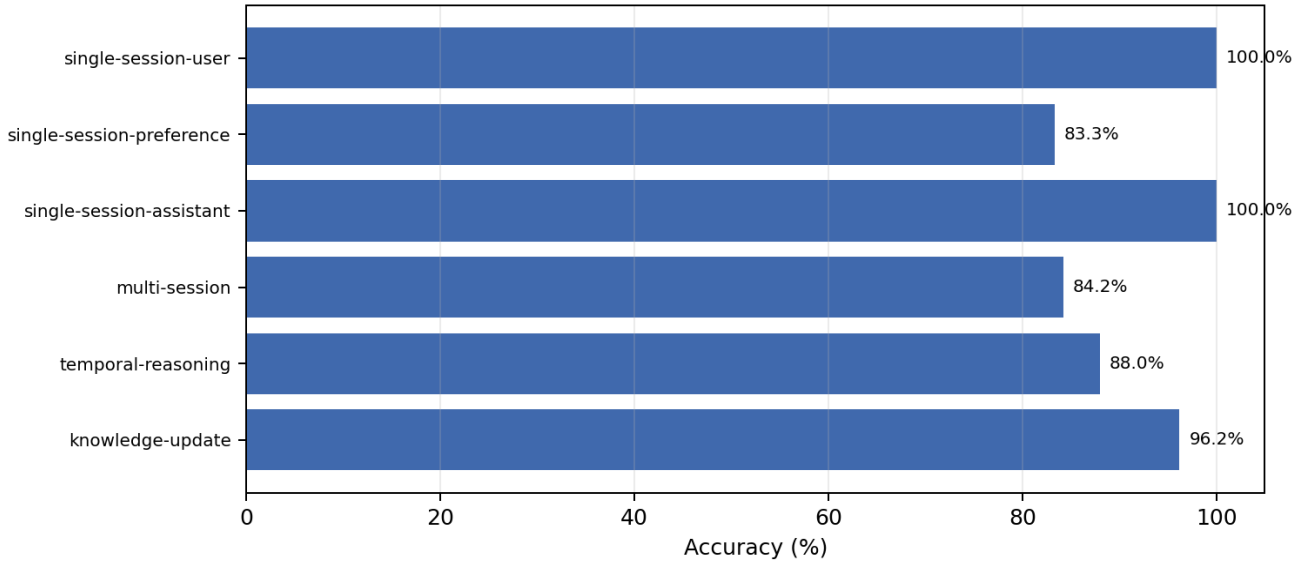
Important caveat: this report is intentionally not a cross-product leaderboard. It is a Recall-AI-only benchmark artifact, suitable for documenting the current system state and tracking future Recall-AI regression or improvement.

2. Per-Task Accuracy

The following table reports Recall-AI accuracy by LongMemEval_S task category. The strongest categories in this run are single-session-user and single-session-assistant, both at 100.0%, followed by knowledge-update at 96.2%. The lower-scoring categories are single-session-preference and multi-session, which remain useful targets for follow-up optimization.

Task category	N	Correct	Accuracy
single-session-user	70	70	100.0%
single-session-preference	30	25	83.3%
single-session-assistant	56	56	100.0%
multi-session	133	112	84.2%
temporal-reasoning	133	117	88.0%
knowledge-update	78	75	96.2%

Recall-AI accuracy by LongMemEval_S task category

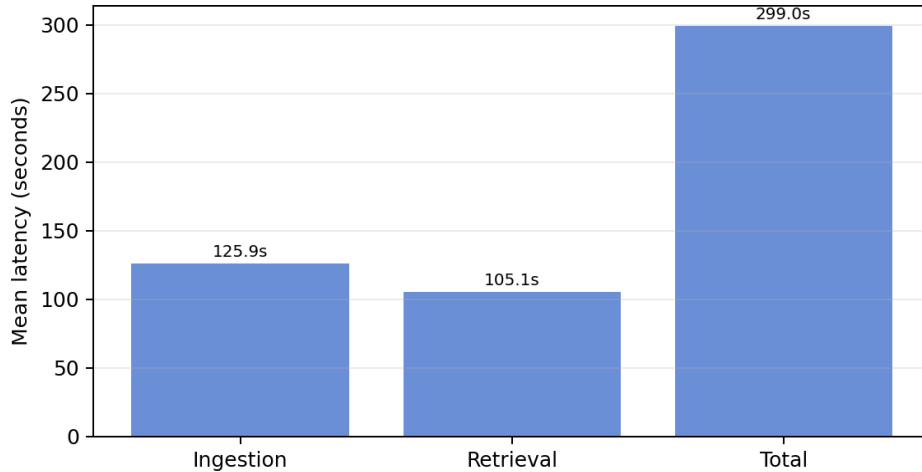


3. Latency Analysis

The current benchmark run shows strong accuracy but heavy processing latency. Mean retrieval latency is 105,094 ms, while mean total latency per question is 298,963 ms. This makes latency optimization a priority before broad production usage.

Stage	Mean	Median / P50	P95
Ingestion	125,915 ms	129,576 ms	153,970 ms
Retrieval	105,094 ms	55,161 ms	424,236 ms
Generation	0 ms	0 ms	0 ms
Total per question	298,963 ms	255,484 ms	598,172 ms

Recall-AI mean processing latency



4. Evaluation Methodology

For each of the 500 questions, the benchmark runner created a fresh, isolated Recall-AI instance and ingested the relevant haystack sessions. Each turn was stored as an individual memory with session metadata and timestamps. This design avoids cross-question contamination.

Step	Description
Memory ingestion	All haystack sessions were ingested into a fresh, isolated Recall-AI instance; each turn was stored with session metadata and timestamps.
Retrieval	Recall-AI used its 11-layer retrieval funnel through <code>build_context()</code> and <code>search()</code> . The route includes temporal, inverted-index, entity, graph, N-gram, vector, reranking, cross-encoder, and LLM filter stages.
Answer generation	Retrieved context was provided to qwen3.5-plus to generate the answer.
Auto-evaluation	Answers were judged with qwen3.5-plus using task-specific evaluation prompts aligned with the LongMemEval protocol.

Isolation guarantee

- Each question used a separate data directory.
- Data directories were cleaned up after each question.
- No cross-contamination between questions was allowed.
- The benchmark did not modify Recall-AI source code or project data.

5. System Configuration

Parameter	Value
Recall-AI version	8.0.0
Embedding backend	BAAI/bge-m3
Embedding dimension	1024
LLM model	qwen3.5-plus
Evaluation model	qwen3.5-plus
Dataset	LongMemEval_S
Questions processed	500
Average turns per question	249
Total runtime	321.8 minutes

6. Recall-AI Result Interpretation

The run demonstrates that Recall-AI can solve most evaluated long-term memory questions under isolated conditions. The 91.0% overall accuracy corresponds to 455 correct answers out of 500.

Area	Observation
Strongest categories	single-session-user and single-session-assistant both reached 100.0%; knowledge-update reached 96.2%.
Harder categories	single-session-preference reached 83.3%; multi-session reached 84.2%; temporal-reasoning reached 88.0%.
Abstention	Abstention accuracy was 76.7% over 30 abstention questions, indicating room for improvement in insufficient-information handling.
Latency	Mean retrieval latency and mean total latency remain heavy, making engineering optimization a clear next step.

7. Limitations and Next Steps

Topic	Note
Benchmark scope	LongMemEval_S is a benchmark signal, not a direct guarantee of production readiness.
Single-system scope	This PDF intentionally omits external product and baseline comparisons. It should not be read as a market ranking.
Evaluation model	The same model family was used for answer generation and automatic judging in this run; future audits may add independent judging configurations.
Latency	Accuracy is currently ahead of usability. Lower retrieval and end-to-end latency should be treated as an engineering priority.
Follow-up tracking	Future reports should keep the same single-system format and track Recall-AI regressions or gains across accuracy, abstention, latency, and category-level performance.

Appendix A. Per-Question Raw Results

This appendix contains 500 question-level rows from the uploaded Recall-AI benchmark report. Source question identifiers are anonymized in this public single-system edition. Correctness is shown as Yes/No, with retrieval and total processing latency in milliseconds.

#	Question	Type	Correct	Retrieval ms	Total ms
1	Q001	single-session-user	Yes	27,741	216,378
2	Q002	single-session-user	Yes	30,644	237,114
3	Q003	single-session-user	Yes	29,655	240,612
4	Q004	single-session-user	Yes	32,088	256,337
5	Q005	single-session-user	Yes	33,706	260,479
6	Q006	single-session-user	Yes	31,854	264,000
7	Q007	single-session-user	Yes	29,812	268,898
8	Q008	single-session-user	Yes	31,877	218,342
9	Q009	single-session-user	Yes	232,384	460,274
10	Q010	single-session-user	Yes	33,866	261,719
11	Q011	single-session-user	Yes	47,629	256,769
12	Q012	single-session-user	Yes	30,433	248,595
13	Q013	single-session-user	Yes	40,553	273,311
14	Q014	single-session-user	Yes	33,190	253,260
15	Q015	single-session-user	Yes	28,551	280,207
16	Q016	single-session-user	Yes	39,204	208,077
17	Q017	single-session-user	Yes	27,671	244,746
18	Q018	single-session-user	Yes	28,777	225,250
19	Q019	single-session-user	Yes	28,197	217,952
20	Q020	single-session-user	Yes	31,235	245,759
21	Q021	single-session-user	Yes	47,124	232,836
22	Q022	single-session-user	Yes	31,818	276,500
23	Q023	single-session-user	Yes	31,640	172,307
24	Q024	single-session-user	Yes	107,062	322,645
25	Q025	single-session-user	Yes	29,619	229,254
26	Q026	single-session-user	Yes	28,673	230,289
27	Q027	single-session-user	Yes	55,743	264,314
28	Q028	single-session-user	Yes	31,055	226,317
29	Q029	single-session-user	Yes	29,820	216,033
30	Q030	single-session-user	Yes	31,663	218,205
31	Q031	single-session-user	Yes	44,461	224,536
32	Q032	single-session-user	Yes	44,739	242,228
33	Q033	single-session-user	Yes	29,659	412,489
34	Q034	single-session-user	Yes	156,646	370,406
35	Q035	single-session-user	Yes	41,373	230,034
36	Q036	single-session-user	Yes	36,746	230,424
37	Q037	single-session-user	Yes	102,838	288,750
38	Q038	single-session-user	Yes	31,252	203,137
39	Q039	single-session-user	Yes	47,991	225,746
40	Q040	single-session-user	Yes	33,179	248,523
41	Q041	single-session-user	Yes	58,175	265,373
42	Q042	single-session-user	Yes	33,609	245,529
43	Q043	single-session-user	Yes	27,727	220,997
44	Q044	single-session-user	Yes	29,487	209,484
45	Q045	single-session-user	Yes	32,785	240,973
46	Q046	single-session-user	Yes	41,096	260,668
47	Q047	single-session-user	Yes	31,943	203,163
48	Q048	single-session-user	Yes	27,547	211,416
49	Q049	single-session-user	Yes	40,162	226,471
50	Q050	single-session-user	Yes	32,091	229,255
51	Q051	single-session-user	Yes	43,337	224,652
52	Q052	single-session-user	Yes	41,593	252,301
53	Q053	single-session-user	Yes	48,342	177,678
54	Q054	single-session-user	Yes	141,036	335,654
55	Q055	single-session-user	Yes	27,705	172,600
56	Q056	single-session-user	Yes	246,227	425,602
57	Q057	single-session-user	Yes	31,925	227,041
58	Q058	single-session-user	Yes	70,785	268,163

#	Question	Type	Correct	Retrieval ms	Total ms
59	Q059	single-session-user	Yes	27,689	264,545
60	Q060	single-session-user	Yes	28,812	218,484
61	Q061	single-session-user	Yes	28,386	216,642
62	Q062	single-session-user	Yes	27,867	272,035
63	Q063	single-session-user	Yes	32,497	250,622
64	Q064	single-session-user	Yes	102,067	285,925
65	Q065	single-session-user	Yes	81,850	294,268
66	Q066	single-session-user	Yes	83,988	284,646
67	Q067	single-session-user	Yes	85,595	303,704
68	Q068	single-session-user	Yes	77,564	276,981
69	Q069	single-session-user	Yes	86,042	309,092
70	Q070	single-session-user	Yes	216,554	447,935
71	Q071	multi-session	Yes	128,151	278,620
72	Q072	multi-session	Yes	63,448	285,616
73	Q073	multi-session	No	65,014	645,563
74	Q074	multi-session	Yes	247,904	437,581
75	Q075	multi-session	Yes	85,352	262,197
76	Q076	multi-session	Yes	94,744	288,915
77	Q077	multi-session	Yes	124,277	271,572
78	Q078	multi-session	Yes	377,791	566,366
79	Q079	multi-session	Yes	392,480	598,025
80	Q080	multi-session	Yes	85,180	228,084
81	Q081	multi-session	Yes	459,799	640,790
82	Q082	multi-session	Yes	55,651	224,572
83	Q083	multi-session	Yes	112,426	292,766
84	Q084	multi-session	Yes	140,094	335,249
85	Q085	multi-session	Yes	134,861	310,291
86	Q086	multi-session	No	291,128	451,461
87	Q087	multi-session	Yes	100,496	273,335
88	Q088	multi-session	Yes	115,510	308,251
89	Q089	multi-session	Yes	99,970	319,326
90	Q090	multi-session	No	349,628	534,283
91	Q091	multi-session	No	274,289	455,092
92	Q092	multi-session	Yes	73,938	215,525
93	Q093	multi-session	Yes	375,495	513,717
94	Q094	multi-session	No	478,613	668,386
95	Q095	multi-session	No	753,554	914,570
96	Q096	multi-session	No	495,788	683,567
97	Q097	multi-session	No	288,254	470,456
98	Q098	multi-session	Yes	68,140	198,399
99	Q099	multi-session	Yes	37,751	222,996
100	Q100	multi-session	Yes	722,638	920,091
101	Q101	multi-session	Yes	61,430	200,683
102	Q102	multi-session	Yes	109,715	290,893
103	Q103	multi-session	Yes	34,598	194,952
104	Q104	multi-session	Yes	83,811	231,573
105	Q105	multi-session	Yes	460,175	597,097
106	Q106	multi-session	Yes	291,963	462,123
107	Q107	multi-session	Yes	49,158	182,185
108	Q108	multi-session	Yes	104,796	264,389
109	Q109	multi-session	Yes	84,181	290,074
110	Q110	multi-session	Yes	115,012	296,589
111	Q111	multi-session	No	125,383	306,479
112	Q112	multi-session	Yes	157,985	323,178
113	Q113	multi-session	No	82,557	278,532
114	Q114	multi-session	Yes	449,251	626,223
115	Q115	multi-session	Yes	99,914	247,590
116	Q116	multi-session	Yes	118,791	283,692
117	Q117	multi-session	Yes	84,253	258,526
118	Q118	multi-session	Yes	91,120	254,501
119	Q119	multi-session	Yes	91,780	261,639
120	Q120	multi-session	No	96,013	259,475
121	Q121	multi-session	Yes	451,649	654,779
122	Q122	multi-session	Yes	66,560	231,847
123	Q123	multi-session	Yes	480,058	636,905
124	Q124	multi-session	Yes	303,610	510,528

#	Question	Type	Correct	Retrieval ms	Total ms
125	Q125	multi-session	Yes	161,301	318,739
126	Q126	multi-session	No	119,557	255,922
127	Q127	multi-session	No	574,361	1,035,504
128	Q128	multi-session	Yes	390,713	566,873
129	Q129	multi-session	No	237,474	443,334
130	Q130	single-session-preference	Yes	35,596	221,096
131	Q131	multi-session	No	424,120	624,633
132	Q132	multi-session	Yes	452,939	645,403
133	Q133	single-session-preference	Yes	68,190	257,943
134	Q134	single-session-preference	Yes	65,216	306,457
135	Q135	multi-session	Yes	165,236	356,816
136	Q136	single-session-preference	No	142,032	367,167
137	Q137	single-session-preference	No	111,340	267,260
138	Q138	single-session-preference	Yes	72,875	287,212
139	Q139	single-session-preference	Yes	82,386	290,398
140	Q140	single-session-preference	No	52,700	250,059
141	Q141	single-session-preference	Yes	92,427	297,198
142	Q142	single-session-preference	Yes	80,211	277,606
143	Q143	single-session-preference	Yes	81,086	273,912
144	Q144	single-session-preference	Yes	39,078	211,224
145	Q145	single-session-preference	Yes	551,615	730,967
146	Q146	single-session-preference	Yes	46,376	223,761
147	Q147	single-session-preference	Yes	49,759	256,942
148	Q148	single-session-preference	No	68,205	303,970
149	Q149	single-session-preference	Yes	87,555	295,754
150	Q150	single-session-preference	No	56,933	260,165
151	Q151	single-session-preference	Yes	63,528	245,534
152	Q152	single-session-preference	Yes	78,650	269,037
153	Q153	single-session-preference	Yes	69,211	284,331
154	Q154	single-session-preference	Yes	82,981	300,478
155	Q155	single-session-preference	Yes	60,659	293,636
156	Q156	single-session-preference	Yes	43,355	247,216
157	Q157	single-session-preference	Yes	59,886	236,505
158	Q158	single-session-preference	Yes	379,667	562,462
159	Q159	single-session-preference	Yes	87,260	255,704
160	Q160	single-session-preference	Yes	78,651	212,722
161	Q161	single-session-preference	Yes	51,288	242,403
162	Q162	multi-session	Yes	32,965	272,419
163	Q163	multi-session	Yes	46,478	243,885
164	Q164	single-session-preference	Yes	96,024	309,330
165	Q165	multi-session	Yes	44,804	258,343
166	Q166	multi-session	Yes	40,183	181,183
167	Q167	multi-session	Yes	163,285	341,839
168	Q168	multi-session	Yes	66,193	252,079
169	Q169	multi-session	Yes	30,034	219,830
170	Q170	multi-session	Yes	38,762	256,216
171	Q171	multi-session	No	76,822	291,397
172	Q172	multi-session	Yes	38,070	246,585
173	Q173	multi-session	Yes	38,692	272,165
174	Q174	multi-session	Yes	47,486	294,390
175	Q175	multi-session	Yes	43,909	181,367
176	Q176	multi-session	Yes	59,017	229,140
177	Q177	multi-session	Yes	47,967	252,145
178	Q178	multi-session	Yes	36,614	242,070
179	Q179	multi-session	No	507,388	684,783
180	Q180	multi-session	Yes	40,347	201,594
181	Q181	multi-session	Yes	32,842	281,408
182	Q182	multi-session	Yes	95,356	317,537
183	Q183	multi-session	Yes	92,226	248,199
184	Q184	multi-session	Yes	46,946	203,434
185	Q185	multi-session	Yes	73,394	240,522
186	Q186	multi-session	Yes	563,662	787,662
187	Q187	multi-session	Yes	46,369	222,862
188	Q188	multi-session	No	32,761	248,129
189	Q189	multi-session	Yes	50,128	226,866
190	Q190	multi-session	Yes	49,434	265,587

#	Question	Type	Correct	Retrieval ms	Total ms
191	Q191	multi-session	Yes	47,399	197,722
192	Q192	multi-session	Yes	36,918	201,755
193	Q193	multi-session	Yes	44,690	245,259
194	Q194	multi-session	Yes	30,936	223,795
195	Q195	multi-session	Yes	39,247	274,121
196	Q196	multi-session	Yes	34,020	221,717
197	Q197	multi-session	Yes	43,535	270,697
198	Q198	multi-session	Yes	71,922	272,829
199	Q199	multi-session	Yes	69,543	246,037
200	Q200	multi-session	Yes	38,484	189,588
201	Q201	multi-session	Yes	228,590	447,388
202	Q202	multi-session	Yes	58,422	249,296
203	Q203	multi-session	Yes	36,097	237,910
204	Q204	multi-session	Yes	87,070	291,964
205	Q205	multi-session	Yes	53,161	393,180
206	Q206	multi-session	Yes	38,759	200,316
207	Q207	multi-session	Yes	56,395	261,491
208	Q208	multi-session	Yes	113,501	366,199
209	Q209	multi-session	Yes	79,604	230,761
210	Q210	multi-session	Yes	45,394	193,868
211	Q211	multi-session	Yes	42,825	224,106
212	Q212	multi-session	Yes	43,732	220,795
213	Q213	multi-session	Yes	347,123	520,452
214	Q214	multi-session	Yes	63,616	296,330
215	Q215	multi-session	No	429,904	617,041
216	Q216	multi-session	Yes	64,291	248,029
217	Q217	multi-session	Yes	489,256	690,041
218	Q218	multi-session	Yes	95,591	260,688
219	Q219	multi-session	Yes	36,870	230,304
220	Q220	multi-session	Yes	38,047	227,187
221	Q221	multi-session	Yes	81,978	269,701
222	Q222	multi-session	Yes	41,767	264,387
223	Q223	multi-session	Yes	70,941	251,291
224	Q224	multi-session	Yes	157,198	339,759
225	Q225	multi-session	Yes	77,214	311,855
226	Q226	multi-session	Yes	106,475	329,650
227	Q227	multi-session	Yes	263,783	465,442
228	Q228	multi-session	Yes	311,715	529,682
229	Q229	multi-session	Yes	112,328	259,214
230	Q230	multi-session	No	86,168	230,941
231	Q231	multi-session	Yes	203,435	378,142
232	Q232	multi-session	Yes	121,626	282,240
233	Q233	temporal-reasoning	Yes	55,178	246,277
234	Q234	temporal-reasoning	Yes	34,976	237,505
235	Q235	temporal-reasoning	Yes	51,328	264,242
236	Q236	temporal-reasoning	Yes	45,757	260,894
237	Q237	multi-session	No	244,184	427,416
238	Q238	temporal-reasoning	Yes	177,942	346,611
239	Q239	temporal-reasoning	Yes	33,893	205,709
240	Q240	temporal-reasoning	Yes	31,553	222,255
241	Q241	temporal-reasoning	Yes	49,745	249,466
242	Q242	temporal-reasoning	Yes	56,277	234,957
243	Q243	temporal-reasoning	No	315,829	524,565
244	Q244	temporal-reasoning	No	141,247	347,028
245	Q245	temporal-reasoning	Yes	118,139	269,848
246	Q246	temporal-reasoning	Yes	74,851	285,032
247	Q247	temporal-reasoning	Yes	34,469	252,920
248	Q248	temporal-reasoning	Yes	58,411	291,836
249	Q249	temporal-reasoning	Yes	74,639	293,865
250	Q250	temporal-reasoning	Yes	59,419	280,181
251	Q251	temporal-reasoning	Yes	78,492	250,313
252	Q252	temporal-reasoning	Yes	49,530	227,032
253	Q253	temporal-reasoning	Yes	36,688	215,756
254	Q254	temporal-reasoning	Yes	37,417	215,973
255	Q255	temporal-reasoning	Yes	43,219	251,041
256	Q256	temporal-reasoning	Yes	44,296	220,219

#	Question	Type	Correct	Retrieval ms	Total ms
257	Q257	temporal-reasoning	Yes	102,311	285,982
258	Q258	temporal-reasoning	Yes	37,125	198,116
259	Q259	temporal-reasoning	Yes	210,588	399,925
260	Q260	temporal-reasoning	No	66,277	243,816
261	Q261	temporal-reasoning	Yes	55,145	226,478
262	Q262	temporal-reasoning	No	780,471	993,438
263	Q263	temporal-reasoning	No	57,021	198,690
264	Q264	temporal-reasoning	No	405,520	590,990
265	Q265	temporal-reasoning	No	594,010	809,874
266	Q266	temporal-reasoning	Yes	89,223	285,031
267	Q267	temporal-reasoning	Yes	52,049	199,071
268	Q268	temporal-reasoning	Yes	42,988	217,544
269	Q269	temporal-reasoning	No	180,885	332,651
270	Q270	temporal-reasoning	Yes	34,619	197,838
271	Q271	temporal-reasoning	Yes	38,522	266,626
272	Q272	temporal-reasoning	Yes	594,826	766,059
273	Q273	temporal-reasoning	Yes	54,275	299,778
274	Q274	temporal-reasoning	Yes	60,047	257,026
275	Q275	temporal-reasoning	Yes	656,432	832,764
276	Q276	temporal-reasoning	Yes	43,111	190,687
277	Q277	temporal-reasoning	Yes	38,423	207,963
278	Q278	temporal-reasoning	Yes	44,475	227,302
279	Q279	temporal-reasoning	Yes	33,053	234,481
280	Q280	temporal-reasoning	Yes	35,349	260,649
281	Q281	temporal-reasoning	No	59,506	292,184
282	Q282	temporal-reasoning	Yes	50,799	247,774
283	Q283	temporal-reasoning	Yes	50,717	243,856
284	Q284	temporal-reasoning	Yes	30,849	408,101
285	Q285	temporal-reasoning	Yes	48,242	253,119
286	Q286	temporal-reasoning	Yes	34,877	230,614
287	Q287	temporal-reasoning	Yes	37,499	213,998
288	Q288	temporal-reasoning	Yes	111,241	304,904
289	Q289	temporal-reasoning	Yes	109,666	278,770
290	Q290	temporal-reasoning	No	42,760	241,449
291	Q291	temporal-reasoning	Yes	174,756	367,166
292	Q292	temporal-reasoning	Yes	110,255	323,083
293	Q293	temporal-reasoning	Yes	127,451	296,602
294	Q294	temporal-reasoning	No	319,465	490,732
295	Q295	temporal-reasoning	No	159,783	371,655
296	Q296	temporal-reasoning	Yes	159,166	319,819
297	Q297	temporal-reasoning	Yes	103,437	310,366
298	Q298	temporal-reasoning	Yes	72,907	287,779
299	Q299	temporal-reasoning	Yes	114,311	313,133
300	Q300	temporal-reasoning	Yes	305,291	471,014
301	Q301	temporal-reasoning	Yes	100,810	246,762
302	Q302	temporal-reasoning	Yes	361,000	535,455
303	Q303	temporal-reasoning	Yes	188,635	322,298
304	Q304	temporal-reasoning	Yes	306,937	492,037
305	Q305	temporal-reasoning	Yes	68,385	202,010
306	Q306	temporal-reasoning	Yes	432,317	600,951
307	Q307	temporal-reasoning	Yes	48,722	245,706
308	Q308	temporal-reasoning	Yes	47,489	246,555
309	Q309	temporal-reasoning	Yes	34,690	262,175
310	Q310	temporal-reasoning	No	499,577	638,015
311	Q311	temporal-reasoning	Yes	48,682	213,195
312	Q312	temporal-reasoning	Yes	41,815	190,488
313	Q313	temporal-reasoning	Yes	426,442	587,821
314	Q314	temporal-reasoning	Yes	51,335	217,235
315	Q315	temporal-reasoning	Yes	277,989	471,103
316	Q316	temporal-reasoning	Yes	41,731	246,054
317	Q317	temporal-reasoning	Yes	45,022	244,290
318	Q318	temporal-reasoning	Yes	54,235	273,422
319	Q319	temporal-reasoning	Yes	106,582	302,325
320	Q320	temporal-reasoning	Yes	52,719	242,991
321	Q321	temporal-reasoning	Yes	53,006	220,768
322	Q322	temporal-reasoning	Yes	120,509	285,446

#	Question	Type	Correct	Retrieval ms	Total ms
323	Q323	temporal-reasoning	Yes	67,663	243,775
324	Q324	temporal-reasoning	Yes	152,634	300,565
325	Q325	temporal-reasoning	Yes	41,272	239,960
326	Q326	temporal-reasoning	Yes	76,140	268,565
327	Q327	temporal-reasoning	No	143,898	336,838
328	Q328	temporal-reasoning	Yes	54,191	273,198
329	Q329	temporal-reasoning	Yes	94,337	303,966
330	Q330	temporal-reasoning	Yes	48,259	233,627
331	Q331	temporal-reasoning	Yes	69,872	282,033
332	Q332	temporal-reasoning	Yes	74,044	298,909
333	Q333	temporal-reasoning	Yes	43,005	239,971
334	Q334	temporal-reasoning	Yes	99,099	288,993
335	Q335	temporal-reasoning	Yes	45,890	231,228
336	Q336	temporal-reasoning	Yes	55,574	253,605
337	Q337	temporal-reasoning	Yes	51,255	206,399
338	Q338	temporal-reasoning	Yes	40,890	209,810
339	Q339	temporal-reasoning	Yes	42,214	243,840
340	Q340	temporal-reasoning	Yes	40,797	232,656
341	Q341	temporal-reasoning	Yes	105,554	288,486
342	Q342	temporal-reasoning	Yes	40,591	244,978
343	Q343	temporal-reasoning	Yes	55,518	192,857
344	Q344	temporal-reasoning	Yes	48,278	193,713
345	Q345	temporal-reasoning	Yes	55,000	250,645
346	Q346	temporal-reasoning	Yes	44,330	230,176
347	Q347	temporal-reasoning	Yes	46,249	213,579
348	Q348	temporal-reasoning	Yes	68,518	271,165
349	Q349	temporal-reasoning	Yes	357,769	570,526
350	Q350	temporal-reasoning	Yes	97,293	340,491
351	Q351	temporal-reasoning	Yes	50,532	218,917
352	Q352	temporal-reasoning	Yes	46,703	191,274
353	Q353	temporal-reasoning	Yes	61,320	201,816
354	Q354	temporal-reasoning	Yes	56,708	215,403
355	Q355	temporal-reasoning	Yes	195,027	350,663
356	Q356	temporal-reasoning	Yes	57,128	243,544
357	Q357	temporal-reasoning	Yes	873,186	1,079,990
358	Q358	temporal-reasoning	Yes	59,121	271,490
359	Q359	temporal-reasoning	Yes	45,976	215,919
360	Q360	temporal-reasoning	No	122,296	373,675
361	Q361	temporal-reasoning	Yes	129,879	294,980
362	Q362	temporal-reasoning	Yes	95,415	295,573
363	Q363	temporal-reasoning	Yes	133,854	316,143
364	Q364	knowledge-update	Yes	74,219	262,870
365	Q365	temporal-reasoning	Yes	138,630	329,078
366	Q366	temporal-reasoning	No	109,150	315,877
367	Q367	knowledge-update	Yes	42,435	216,696
368	Q368	knowledge-update	Yes	56,509	223,939
369	Q369	knowledge-update	Yes	88,097	276,888
370	Q370	knowledge-update	Yes	50,374	240,929
371	Q371	knowledge-update	Yes	33,195	226,148
372	Q372	knowledge-update	Yes	41,193	208,554
373	Q373	knowledge-update	Yes	49,087	241,987
374	Q374	knowledge-update	Yes	64,520	265,721
375	Q375	temporal-reasoning	Yes	430,577	637,259
376	Q376	knowledge-update	Yes	35,881	242,766
377	Q377	knowledge-update	Yes	77,955	278,717
378	Q378	knowledge-update	Yes	47,980	215,845
379	Q379	knowledge-update	Yes	134,131	320,373
380	Q380	knowledge-update	Yes	41,876	216,293
381	Q381	knowledge-update	Yes	39,114	229,819
382	Q382	knowledge-update	Yes	120,948	325,086
383	Q383	knowledge-update	Yes	34,409	352,067
384	Q384	knowledge-update	Yes	92,205	308,769
385	Q385	knowledge-update	Yes	57,264	278,226
386	Q386	knowledge-update	Yes	29,847	202,392
387	Q387	knowledge-update	Yes	127,479	339,502
388	Q388	knowledge-update	Yes	45,918	278,572

#	Question	Type	Correct	Retrieval ms	Total ms
389	Q389	knowledge-update	Yes	41,307	234,328
390	Q390	knowledge-update	Yes	33,850	198,140
391	Q391	knowledge-update	Yes	310,084	513,662
392	Q392	knowledge-update	Yes	67,285	267,595
393	Q393	knowledge-update	Yes	45,572	233,260
394	Q394	knowledge-update	Yes	77,893	245,086
395	Q395	knowledge-update	Yes	109,273	296,709
396	Q396	knowledge-update	Yes	51,433	202,432
397	Q397	knowledge-update	Yes	30,474	211,636
398	Q398	knowledge-update	Yes	42,004	237,233
399	Q399	knowledge-update	Yes	345,347	537,038
400	Q400	knowledge-update	Yes	31,893	237,812
401	Q401	knowledge-update	Yes	31,578	223,879
402	Q402	knowledge-update	Yes	42,889	246,562
403	Q403	knowledge-update	Yes	31,759	201,338
404	Q404	knowledge-update	Yes	47,162	240,189
405	Q405	knowledge-update	Yes	76,113	414,218
406	Q406	knowledge-update	Yes	30,178	245,275
407	Q407	knowledge-update	Yes	42,195	263,492
408	Q408	knowledge-update	Yes	36,777	252,954
409	Q409	knowledge-update	Yes	50,370	217,101
410	Q410	knowledge-update	Yes	108,517	343,122
411	Q411	knowledge-update	Yes	179,497	393,225
412	Q412	knowledge-update	Yes	70,769	220,855
413	Q413	knowledge-update	Yes	197,450	380,249
414	Q414	knowledge-update	Yes	68,185	236,372
415	Q415	knowledge-update	Yes	36,727	244,445
416	Q416	knowledge-update	No	91,065	303,042
417	Q417	knowledge-update	Yes	47,239	234,172
418	Q418	knowledge-update	Yes	48,404	227,840
419	Q419	knowledge-update	Yes	30,666	234,897
420	Q420	knowledge-update	Yes	228,239	461,611
421	Q421	knowledge-update	Yes	71,110	254,523
422	Q422	knowledge-update	Yes	62,879	234,909
423	Q423	knowledge-update	Yes	67,009	271,609
424	Q424	knowledge-update	Yes	81,758	273,343
425	Q425	knowledge-update	Yes	43,775	277,530
426	Q426	knowledge-update	Yes	301,587	487,212
427	Q427	knowledge-update	Yes	49,112	210,890
428	Q428	knowledge-update	Yes	139,778	388,789
429	Q429	knowledge-update	Yes	93,376	259,311
430	Q430	knowledge-update	Yes	143,782	319,131
431	Q431	knowledge-update	Yes	43,839	197,995
432	Q432	knowledge-update	Yes	48,072	214,815
433	Q433	knowledge-update	Yes	33,843	220,814
434	Q434	knowledge-update	Yes	39,334	213,927
435	Q435	knowledge-update	Yes	67,986	258,154
436	Q436	knowledge-update	Yes	72,338	286,864
437	Q437	knowledge-update	Yes	95,012	297,341
438	Q438	knowledge-update	Yes	91,715	314,495
439	Q439	knowledge-update	No	156,650	364,344
440	Q440	knowledge-update	Yes	83,549	291,157
441	Q441	knowledge-update	Yes	280,095	438,792
442	Q442	knowledge-update	No	104,857	274,391
443	Q443	knowledge-update	Yes	72,456	313,267
444	Q444	knowledge-update	Yes	44,317	207,698
445	Q445	single-session-assistant	Yes	34,346	227,301
446	Q446	single-session-assistant	Yes	33,043	212,300
447	Q447	single-session-assistant	Yes	31,340	233,833
448	Q448	single-session-assistant	Yes	32,515	230,237
449	Q449	single-session-assistant	Yes	34,327	227,566
450	Q450	single-session-assistant	Yes	33,730	242,393
451	Q451	single-session-assistant	Yes	33,913	256,073
452	Q452	single-session-assistant	Yes	31,068	196,170
453	Q453	single-session-assistant	Yes	33,098	224,109
454	Q454	single-session-assistant	Yes	30,579	224,504

#	Question	Type	Correct	Retrieval ms	Total ms
455	Q455	single-session-assistant	Yes	33,478	229,356
456	Q456	single-session-assistant	Yes	32,221	212,697
457	Q457	single-session-assistant	Yes	34,336	255,265
458	Q458	single-session-assistant	Yes	35,436	217,257
459	Q459	single-session-assistant	Yes	31,844	214,318
460	Q460	single-session-assistant	Yes	32,453	232,876
461	Q461	single-session-assistant	Yes	32,362	215,867
462	Q462	single-session-assistant	Yes	33,304	237,058
463	Q463	single-session-assistant	Yes	49,822	250,550
464	Q464	single-session-assistant	Yes	39,299	233,793
465	Q465	single-session-assistant	Yes	30,105	224,509
466	Q466	single-session-assistant	Yes	31,686	246,685
467	Q467	single-session-assistant	Yes	37,425	283,164
468	Q468	single-session-assistant	Yes	36,576	227,761
469	Q469	single-session-assistant	Yes	32,953	279,150
470	Q470	single-session-assistant	Yes	35,783	239,897
471	Q471	single-session-assistant	Yes	32,721	209,598
472	Q472	single-session-assistant	Yes	29,821	240,357
473	Q473	single-session-assistant	Yes	29,443	227,768
474	Q474	single-session-assistant	Yes	30,161	248,506
475	Q475	single-session-assistant	Yes	34,214	227,468
476	Q476	single-session-assistant	Yes	33,623	247,378
477	Q477	single-session-assistant	Yes	29,022	216,997
478	Q478	single-session-assistant	Yes	44,094	248,953
479	Q479	single-session-assistant	Yes	46,190	247,779
480	Q480	single-session-assistant	Yes	32,568	227,225
481	Q481	single-session-assistant	Yes	56,243	244,019
482	Q482	single-session-assistant	Yes	30,389	232,333
483	Q483	single-session-assistant	Yes	33,492	205,877
484	Q484	single-session-assistant	Yes	31,425	224,917
485	Q485	single-session-assistant	Yes	31,885	227,269
486	Q486	single-session-assistant	Yes	31,513	260,919
487	Q487	single-session-assistant	Yes	30,631	243,098
488	Q488	single-session-assistant	Yes	34,203	229,652
489	Q489	single-session-assistant	Yes	34,360	213,587
490	Q490	single-session-assistant	Yes	38,392	234,364
491	Q491	single-session-assistant	Yes	36,257	239,405
492	Q492	single-session-assistant	Yes	33,443	193,786
493	Q493	single-session-assistant	Yes	35,482	233,636
494	Q494	single-session-assistant	Yes	32,816	225,240
495	Q495	single-session-assistant	Yes	81,027	259,552
496	Q496	single-session-assistant	Yes	32,527	217,034
497	Q497	single-session-assistant	Yes	31,519	230,377
498	Q498	single-session-assistant	Yes	31,107	233,650
499	Q499	single-session-assistant	Yes	31,458	247,895
500	Q500	single-session-assistant	Yes	218,325	414,431

Appendix B. Benchmark Reference

Wu, D., Wang, H., Yu, W., Zhang, Y., Chang, K.W., and Yu, D. (2025). LongMemEval: Benchmarking Chat Assistants on Long-Term Interactive Memory. ICLR 2025. arXiv:2410.10813.

Report generated by LongMemEval Benchmark Runner for Recall-AI. This PDF version was prepared as a Recall-AI-only edition with external product and baseline comparisons removed.