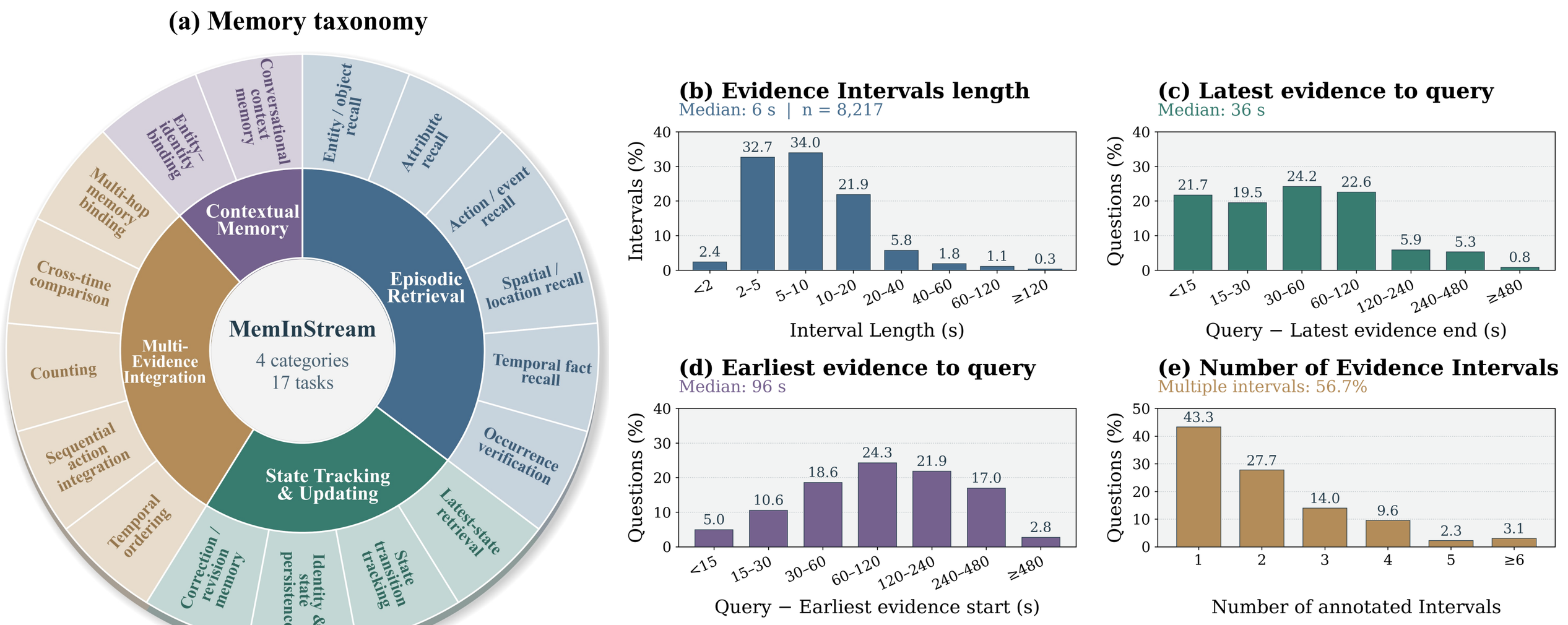

(a) Memory taxonomy
MemInStream
4 categories
17 tasks
Contextual Memory
Episodic Retrieval
State Tracking & Updating
Multi-Evidence Integration
Conversational context memory
Entity–identity binding
Multi-hop memory binding
Entity / object recall
Attribute recall
Action / event recall
Spatial / location recall
Temporal fact recall
Occurrence verification
Latest-state retrieval
State transition tracking
Identity & state persistence
Correction / revision memory
Temporal ordering
Sequential action integration
Counting
Cross-time comparison
(b) Evidence Intervals length
Median: 6 s | n = 8,217
Intervals (%)
2.4
32.7
34.0
21.9
5.8
1.8
1.1
0.3
<2
2–5
5–10
10–20
20–40
40–60
60–120
≥120
Interval Length (s)
(c) Latest evidence to query
Median: 36 s
Questions (%)
21.7
19.5
24.2
22.6
5.9
5.3
0.8
<15
15–30
30–60
60–120
120–240
240–480
≥480
Query − Latest evidence end (s)
(d) Earliest evidence to query
Median: 96 s
Questions (%)
5.0
10.6
18.6
24.3
21.9
17.0
2.8
<15
15–30
30–60
60–120
120–240
240–480
≥480
Query − Earliest evidence start (s)
(e) Number of Evidence Intervals
Multiple intervals: 56.7%
Questions (%)
43.3
27.7
14.0
9.6
2.3
3.1
1
2
3
4
5
≥6
Number of annotated Intervals