COMPUTER VISION AND CONTINUAL LEARNING

What Did the Wearer Last See? Evidence and Memory in First-Person Video

A synthesis of shared-house recordings, streaming memory architectures and tests of what a wearable assistant can know

Revanth Matha · Matha Labs | Preprint awaiting review 1.0 | 2026-11-20

Abstract

A wearable assistant sees a mug on a counter. Later, while its wearer is away, someone moves it to a shelf. Where did the wearer last see the mug, and what can the assistant say about its current location? We synthesize evidence from Ego4D, EgoLife, recurrent sequence models, Titans, MIRAS, MaLoW and programmable neural memory. We independently analyze source timestamps for 500 released EgoLifeQA questions from one wearer. The median age of the oldest annotated clue was 5.87 hours, and 115 questions (23%) had a clue at least 24 hours old. These ages describe the timing of annotated evidence. In eight constructed worlds, carrying the last visible location forward matched the actual final location six times, including two unsupported guesses. A rule that answered only with visible current evidence matched that physical label four times while answering all eight questions correctly under an evidence-sensitive scoring rule. We specify a prospective test that measures object identity, separate encounters, location revisions, uncertainty and the time to answer using a common first-person stream. The present results are descriptive and diagnostic. Relative performance on the proposed visual task remains open.

Keywords: egocentric video, episodic memory, partial observability, EgoLifeQA, state-space models, fast weights, streaming evaluation

Contributions

1. A definition of three distinct questions about a familiar object: its identity, its last evidenced location for one wearer, and its independently verified location at the query time.

2. A reproducible description of the timing of evidence in one released EgoLifeQA question set, alongside a controlled audit of elementary answering rules.

3. A comparison protocol for searchable video history, recurrent state and writable neural weights under common observation, storage and timing conditions.

1. One mug, two questions

You leave a blue mug on your desk and walk out. Someone puts it on a shelf. Your glasses last recorded it on the desk. A camera in the room recorded the move. Both recordings are accurate. They support different answers. A wearable assistant asked where you last saw the mug should say desk. Asked where it is now, an assistant with only your glasses' footage has grounds to describe the earlier sighting and express uncertainty about the present.

The example gets harder when a second blue mug appears. An assistant might remember the shelf correctly and attribute the move to the wrong mug. It might also see one mug twice and count two objects. These errors matter for any system that claims to learn by turning new observations into weights. A successful write must preserve what was learned, who could observe it and when it was true. An old observation can remain accurate even after the world changes.

This paper asks what published recordings can test, which forms of computer memory make different predictions, and how to score the answers. We calculate how far back annotated evidence lies for questions in one publicly released EgoLife record. Separately, a small controlled example checks whether a scoring rule rewards an assistant for guessing a fact it had no opportunity to see. Together these analyses set up a test that a retrieval-based model and a weight-based model could both take.

2. What we compared and how we chose the evidence

We selected sources to answer three specific questions. What can a first-person camera establish? What information survives between observations in a streaming model? How can a question distinguish a current fact from an earlier observation? For collection and evaluation we used the EgoLife paper, its supplement, the team's collection account, its released annotation file and the Ego4D episodic-memory specification. For computation we selected primary descriptions of recurrent state, fast weights, test-time neural memory and a video model that generates weight updates. The resulting review is purposive. It supports comparison of mechanisms and tasks without estimating the prevalence of any particular method. [1][2][3][4][5]

For each source we recorded the input at inference, the information carried forward, the target used to score the response, and the time or resource unit reported. Where an author timed only part of a pipeline, we preserve that measurement boundary. Accuracy on a multiple-choice video question, localization of an object in earlier frames, and a correct value returned for a supplied symbolic address are different outcomes. Pooling these scores would conceal precisely the differences an experiment needs to explain.

Two new analyses accompany the review. The first derives evidence ages from authors' timestamps using a pinned public file. The second uses fully specified synthetic scenes to exercise the scoring logic. Neither requires a downloaded copy of the residents' videos. The results describe those inputs. Statements about real-camera identity tracking, human memory and the relative performance of competing architectures require evidence from their respective tasks.

3. How a shared house became a memory dataset

In the Beijing EgoLife study, six participants wore Meta Project Aria glasses during waking activities over seven days. They prepared an Earth Day party, giving an earlier decision a consequence that might become visible hours later. Fifteen fixed GoPros supplied other views in common areas. The publication reports 266 hours of retained first-person footage, or about 44.3 hours per wearer. Simultaneous views of one dinner add several wearer-hours to that total while describing the same dinner. [1][2]

The authors aligned audio within camera groups and then across groups, with a human checking the automated matches. This matters when a cup changes hands quickly. A clock offset can place a transfer before a person turns toward the table even when the opposite sequence occurred. For a new evaluation, an annotator should inspect the synchronization around each candidate transfer and mark the ordering uncertain if the error could change the label. A fixed camera can confirm a move only while the relevant object and destination are discernible. [2]

The project also collected speech, gaze and movement signals, with procedures for removing sensitive audio and blurring sensitive imagery. The authors describe annotation as the major cost of the project. Their choices make a practical tradeoff visible: more cameras can help label an event, while alignment, review and privacy protection demand work for each added stream. Gaze direction measures orientation. An estimate of what a participant consciously noticed or still remembers would require a separate behavioral measurement with a defined cue and delay. [2][3][14]

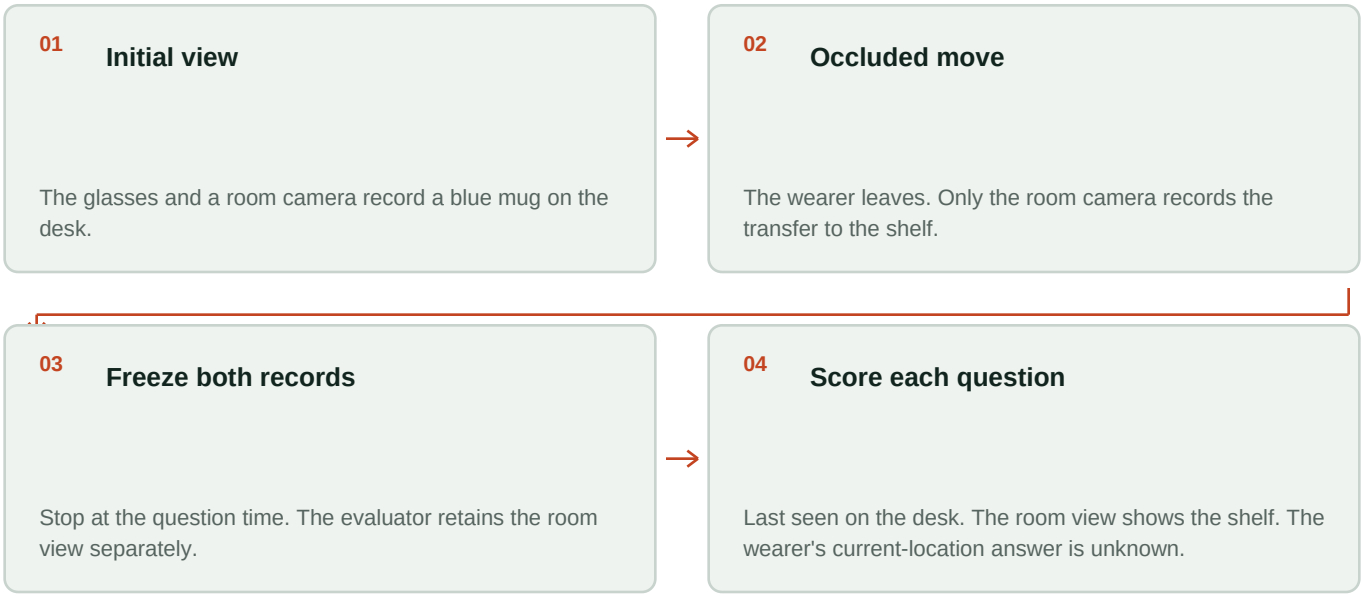
4. Three answers an evaluator must keep distinct

For an object identity and a query time, define the last wearer-visible location as the location in the most recent qualifying first-person sighting at or before that time. Define current world location as the independently corroborated location at the query time. Define wearer-supported current location as the current location justified by the observations available to the assistant under the task's information rules. In the controlled task, occlusion permits an unseen move. A point claim about the present therefore requires fresh evidence. A probability based on the usual fate of objects would answer a different, forecasting question.

Ego4D's visual-query task provides a precise precedent for the first target. Given an image of an object, the benchmark localizes its last appearance in the recording. Its language queries locate a span containing the answer, and its moment queries locate occurrences of an activity. Each anchors evaluation to observable frames. A team with access to synchronized room cameras can add the current-world target while reserving those camera views for annotation. A wearable-only system must still respond according to its own evidence. [5]

Two distinct causes of revision require different history. If the mug really moved from the desk to the shelf, both locations belong in its history. If a later close-up shows that the item on the desk was a different mug, the earlier identity assignment needs correction. Replacing a single location field can solve a current-location question in the first case and leave a false episode in the second. Separate queries about identity and time make those failures visible.

Figure 1. Who could observe the move?



Constructed example. The room view establishes an evaluation label. The assistant receives the wearer's stream. A later shelf sighting would supply new evidence.

5. What the assistant carries from one observation to the next

EgoButler offers an external-history baseline. Its EgoGPT component captions short first-person clips. EgoRAG uses timestamped captions and summaries to retrieve relevant evidence. In the paper's evaluation of 500 questions for one wearer, the retrieval stage selected three 30-second clips. This arrangement makes a failure diagnosable. Inspecting the retrieved clips can reveal whether a caption omitted the mug's handle, retrieval selected the wrong day, or the answerer misread a relevant clip. On the published evaluation subset, Table 6 reports scores of 35.4 with EgoRAG and 25.0 with the segment-limited EgoGPT comparator for queries with a certification length over 24 hours. Those are reported benchmark results, separate from the evidence-age measurements below. [1][12]

MemLife supplies another retrieval baseline. Its writer builds text episodes grounded in entities, and its reader searches them by time without reopening the video at query time. Comparing an addressed neural array with MemLife would hold both methods to previously computed memory while permitting different representations. It would also distinguish the effect of a good episode description from the extra benefit of returning to raw frames. [15]

An S4 layer carries a running state updated by a structured linear transformation and the current input. Mamba introduces input-dependent selection, which changes the influence that new information has on the state. Both give a reader a compact numerical history as the stream advances. Whether a rare mug remains

distinguishable after many intervening scenes depends on the learned representation, the state size and the task used for training. A recurrence describes how information travels. It does not supply an object identity or a definition of a sighting. [6][7]

Titans learns key-value associations in a neural memory during inference. Error on the current association drives updates, with momentum and a forgetting mechanism. MIRAS describes memory architecture, objective, retention and update rule as separable choices. That separation explains why a familiar sighting can be hard to retain as an episode. The occurrence of a mug on a desk might yield little error for an identity predictor, yet it supplies a new time at which the mug was present. A training objective for episode recall can reward storing that time. [8][9]

MaLoW is an especially close visual comparison. Its recurrent writer uses the preceding clip and a compact state to generate low-rank changes to a vision-language model. One path tracks local dynamics. Another adds clip representations when consecutive pooled visual features differ sufficiently. This gate can reduce redundant additions across adjacent views. Identifying the same mug after lunch, or distinguishing it from a near-identical one, requires evidence across nonadjacent clips. A useful test includes both identical successive frames and visually similar objects encountered hours apart. [10]

Table 1. Comparators for a shared first-person task

Method	What persists	Failure to investigate
EgoRAG	Timestamped descriptions and retrievable segments	Relevant segment exists but retrieval misses it
S4 or Mamba	Recurrent numerical state	A rare identity is lost after distractors
Titans-style memory	Parameters updated on an association objective	A predictable repeat leaves weak episode evidence
MaLoW-style adapter	Recurrent history used to generate low-rank weights	Similar scenes contain different objects or visits
Addressed weight array	Changes accumulated at selected memory addresses	Incorrect association contaminates one object's history

Failure modes are hypotheses for a matched comparison. The table describes where to look for an error. It contains no cross-paper accuracy ranking.

6. What programmable weights contribute

Paper 06 implements a trained symbolic reader and writer. The writer receives the new payload, the addressed column, an operation and the mean of the current array. The resulting numerical change participates in future reads. Supplying the address gives the writer a destination in advance. Averaging columns gives the writer a global feature, although different assignments of values to columns can yield exactly the same average. Paper 06 separately evaluates recognition of photographed objects across viewpoints. Linking a new photograph to the correct symbolic history is a further learning problem. [11]

Rapid associative writes already have a mathematical precedent in fast-weight programming. An update can carry a value and a key, and a corrective write can adjust the association currently retrieved using that key. In Paper 06, a hand-coded correction tied the learned writer on the bounded SET-only control. This makes the harder claim more specific. A learned writer could earn its complexity by discovering when two observations belong together, protecting historical episodes while revising current beliefs, or allocating limited memory where it improves a later answer. [11][13]

The effect of consolidation also depends on what the model must recover. Algebraically combining numerical updates can preserve every present read while removing the ability to attribute a component to one earlier write. Checking current labels before and after consolidation tests prediction preservation. Asking for each encounter and its time tests preservation of episode-specific information. A request to erase one encounter raises a third question about residual influence on shared weights. We propose to track those outcomes separately and limit the core comparison to recognition, correction and episode recall.

7. Measuring the age of evidence in EgoLifeQA

We analyzed the 500 question records in the A1 release at Hugging Face revision 143fb319be7aa5ae210c936bf4f0f3a86092afb0. The published benchmark describes 500 curated questions for each of six wearers. The file examined here contains one wearer's questions with their annotated source times. Each question contributes one row to our analysis, regardless of how many source times it contains. We checked that all 500 identifiers were unique. This study used annotation metadata and did not rewatch the source clips or rate the correctness of the authors' answers. [1][4]

For a query on a later day, we counted elapsed wall-clock time across midnight. We read the hour, minute and second from each eight-digit clock field and discarded its final two digits. We treated DAY1 through DAY7 as consecutive dates. Some question records store one target time, others list several, and some append another day and clock to the same text field. We decoded all three formats, collapsed identical target times within a question at one-second resolution, and checked that every target preceded its query. All 500 records passed. The calculation includes overnight gaps in recording. It measures distance in time to a cited event, which may differ substantially from the number of video frames available to a model. [4]

For each question we computed two quantities. The oldest-evidence age is the query time minus the earliest cited target. The nearest-evidence age uses the latest cited target. Quartiles use linear interpolation at fractional rank $(n - 1)p$ on sorted values. Our repeat-prompt sensitivity check retained the lowest numeric identifier among records sharing an English question and query clock on the same day. We report distributions

for the observed questions. Their common wearer, household and deliberate selection preclude estimating population uncertainty by resampling these rows as independent lives.

```
oldest age = query time - min(evidence times) nearest age = query time - max(evidence times)
```

Both differences use seconds on the same seven-day clock. A question with one target has equal ages. An evidence timestamp is the reference chosen by an annotator.

8. How far back the released questions point

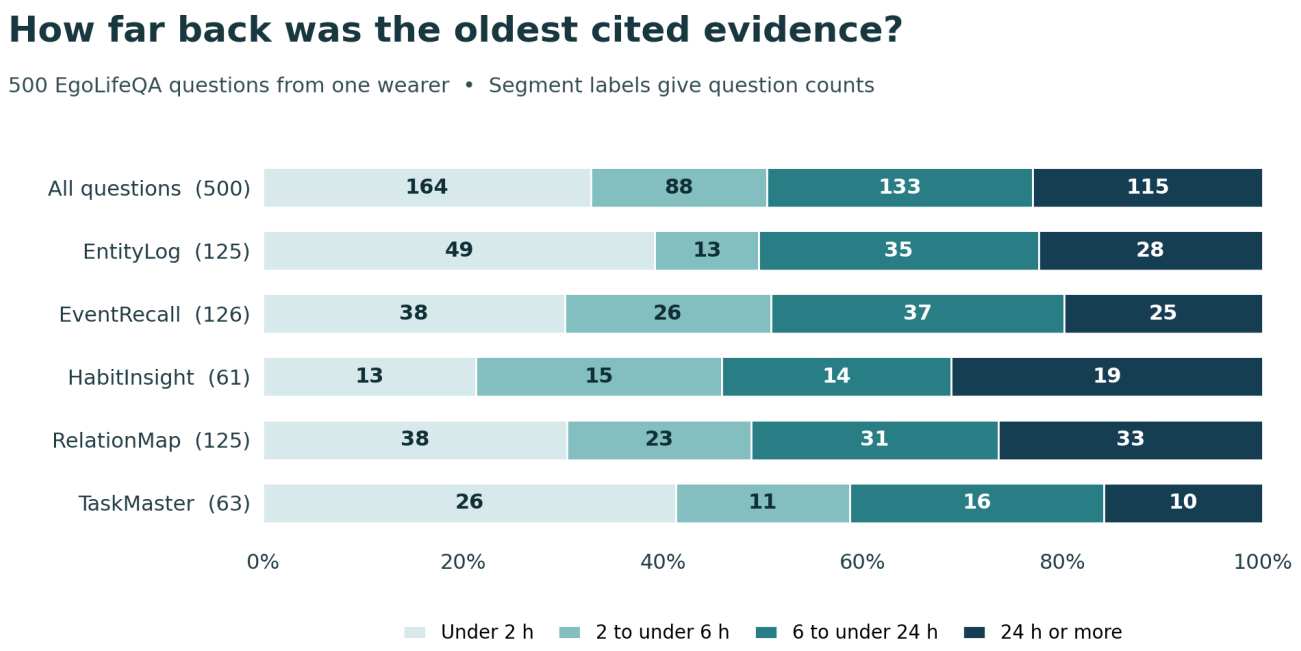
The median oldest cited evidence was 5.87 hours before its question. The middle half of questions spanned 1.06 to 22.21 hours on this measure. In 115 of 500 questions (23.0%), at least one cited target was a full day old. For 110 questions (22.0%), even the nearest cited target was that old. These are distances to particular annotated clues. A model might find another adequate clue closer to the question, infer the answer from an unrelated shortcut, or fail while watching the correct clip. The counts cannot establish a minimum context window for every successful model.

Forty-eight questions named multiple distinct target times. Across all records we found 564 target occurrences and 563 distinct times when deduplicating within each question. The median nearest age was 5.79 hours. The question about an earlier visit to a freezer (record 349) illustrates the clock calculation. The query occurred on day 5 at 15:55:30 and cited day 3 at 17:10:07, an elapsed 46 hours, 45 minutes and 23 seconds. No estimate of what the participant remembered is needed to derive this quantity. [4]

The plot shows substantial variation within each question category. Among 61 questions about habits, 28 referenced multiple distinct times and the median oldest age was 8.08 hours. Among 125 object-focused questions, only three referenced multiple times and the median oldest age was 6.02 hours. The benchmark deliberately uses different question forms. These counts characterize that selection. They cannot tell us how commonly a resident encounters an object or forms a habit.

Collapsing repeated question-and-clock pairs left 496 records. The median oldest age was 5.82 hours and 113 of 496 questions had oldest evidence at least one day old. The full analysis retained two cases with oldest ages just under five minutes despite the paper's stated five-minute selection criterion. The shortest was 292 seconds. Keeping those rows makes the descriptive result reproducible from the released file and avoids silently applying an eligibility rule after seeing the data. [1][4]

Figure 2. Age of the oldest cited evidence by question type



Each row uses its own question count as the denominator. The bands are under 2 hours, 2 to under 6 hours, 6 to under 24 hours, and at least 24 hours. The figure describes the complete analyzed A1 file. Bars show proportions of curated questions. [4]

Table 2. Question-level evidence age in the released A1 annotations

Question type	Questions	Median oldest (h)	Oldest at least 24 h
All	500	5.87	115
EntityLog	125	6.02	28
EventRecall	126	5.66	25
HabitInsight	61	8.08	19
RelationMap	125	6.77	33
TaskMaster	63	3.24	10

The category rows sum to 500. Each time is measured at one-second resolution before conversion to hours and rounding. [4]

9. What an unseen move does to a score

We enumerated eight synthetic worlds. One identified mug begins on a counter or shelf. It then moves to the other location or stays put. At the final time the glasses either see it clearly or have no view of it. The room observer sees both states. Each of the two possible initial locations gives a pair of hidden worlds with identical wearer observations and different final room labels. We deliberately allowed arbitrary unseen movement. Without another information channel, a wearer-only rule has the same inputs in both members of a pair. It can return a supported last-seen answer in both, while the hidden final location remains unresolved.

Three hand-coded controls receive the same observations. Each correctly returns the last visible location in all eight worlds by construction. For the current-location question, one rule carries the last sighting forward, one always abstains, and one answers only when the final location is visible. The carry-forward rule matches the physical location in six cases, including two correct guesses under occlusion. The visibility-aware rule matches the physical location in four cases, and earns all eight available points when the target is the answer supported by the wearer's evidence. A score rewarding only the physical label reverses the ordering of these two rules.

The always-abstaining control exposes another shortcut. It makes zero unsupported assertions and still fails every visible current-location question. A useful evaluation therefore counts both unsupported assertions under occlusion and correct answers on supported cases. These counts exhaust the toy construction. They measure properties of the scoring rule and elementary policies, with visibility and object identity given. A learned visual system would additionally have to detect occlusion and identity uncertainty from pixels.

Table 3. Eight-world audit of the answering rules

Current-answer rule	Supported current correct / 8	Unsupported assertions / 4 hidden	Physical location matches / 8
Carry last location forward	4	4	6
Always answer unknown	4	0	0
Answer only with a visible final state	8	0	4

All rules return the correct last-seen location in 8 of 8 cases. The last column counts unknown as a nonmatch even when uncertainty is the only supported answer. These are controlled logical cases.

10. A fair streaming experiment with real objects

We propose two evaluation tracks. First, every method receives the same causally ordered sequence of annotated observations with the object identity supplied. This isolates the memory mechanism. Second, every method receives the same visual and audio inputs at the same sampling times from one wearer. An identical frozen visual encoder may be used across memory variants to reduce confounding. Hold out physical object instances and contiguous episodes. A second view of a held-out encounter belongs on the same side of the split as the first view. Otherwise a learner could recognize the evaluation event from training.

Use synchronized fixed-camera evidence solely to construct and audit reference labels. Annotators should mark the wearer, candidate object, encounter interval, location and latest time at which a reference camera can corroborate that location. Two annotators independently assess hard cases such as two similar mugs crossing paths. Record disagreements, occlusions, dropped frames and redacted regions. Mark unresolved physical labels as unavailable for the world-location score and retain them where a supported uncertainty response can still be labeled. Synchronization alone cannot reveal a transfer that every camera missed.

Construct matched sequences that independently vary the number of physical objects and the number of encounters. Include one object on two visits, two similar objects in one visit, and the same object moving under visible or hidden conditions. Define a new encounter by an annotated departure and re-entry or a documented episode boundary. Consecutive frames of a stationary mug belong to one encounter. A separately labeled mistaken-identity correction requires the system to revise the assignment of an earlier sighting without counting an extra physical move.

Run each method on identical observation prefixes and ask current, last-seen and episode questions at the same instants. Score answers before exposing labels or subsequent frames. Fix the online-write budget and report what each model can retain, including caption indexes, any video cache, recurrent state, update factors and optimizer state. Flag a reference retrieval system that can reopen the entire archive as an archive-access condition. Evaluate the archive-backed variant separately and report its storage and read bandwidth. A text-only retrieval system also provides a bounded-memory comparison once its retained bytes are capped. [15]

11. Accuracy, abstention and the clock

For identity, count a false merge when two physical objects receive one identity and a false split when one physical object receives several identities. Report both numerators and the eligible pair counts. For episode retrieval, score the predicted encounter against a labeled time interval with a declared overlap tolerance. For current and last-seen queries, report accuracy separately for visible, occluded and disputed cases. Publish answer coverage, the fraction of unsupported claims among hidden trials, and accuracy among answered questions so a system cannot earn a favorable summary by systematically refusing difficult cases.

Dependence is built into this experiment. Several cameras record a single interaction, and several questions can refer to one underlying event. Reserve all views of a shared event together. Summarize performance by event and wearer as well as by question, and compare competing models on matched queries. A generalization claim about homes would require data from held-out homes. With only one house, an interval based on thousands of adjacent frames can give a misleading impression of precision.

Define the proposed 100-millisecond deadline from availability of the final required camera sample to emission of a complete structured answer. Record ingestion, encoding, memory lookup, decoding, queuing and background consolidation separately. Report the fraction of misses including dropped requests, as well as p50, p95 and p99 latency and the oldest pending update. Verify that the answer uses a snapshot committed no later than the query. Fast reads from a bank lagging several seconds behind the stream can give a consistently obsolete location.

Published numbers use different boundaries. MaLoW reports a 152.89 ms profile through the first output token for a Qwen3.5-9B setup on 100 sampled OVO-Bench questions. Video decoding, resizing and other input preprocessing were excluded. Paper 06 times small symbolic reads during background consolidation on an Apple M1 Max. Both measurements describe their respective workloads. Meeting the proposed wearable deadline requires timing the entire agreed input-to-answer path on named hardware under a sustained arrival rate. [10][11]

12. Reproduction and limits of inference

The private review package includes the annotation analysis, the scoring-control program and their saved outputs. Running the analysis fetches the pinned A1 annotations, checks their SHA-256 digest and generates the question-level lag file, category summaries and plotted data. A reviewer can recalculate the oldest and nearest ages for a few records, compare the category totals with Table 2, and inspect the duplicate-prompt sensitivity result. Running the scoring control separately regenerates the eight possible worlds and all entries in Table 3. The data and experiment serve different purposes and have separate denominators. [4]

Several uncertainties survive these checks. A target time marks a reference chosen during annotation and can omit alternative valid clues. Clock time across a night can be long even when the model received only a few intervening clips. Curated multiple-choice questions restrict the answer space, and another household might generate substantially different interactions. A camera can show that an object was potentially visible while its wearer happened to attend to something else. The digital evidence can support camera-level claims. A test of the wearer's memory would require an actual later response from that person under a controlled cue. [3][14]

13. Conclusions

The released questions establish a measurable reason to study long-lived memory in first-person video. For one wearer's 500 curated questions, the median age of the earliest cited clue was almost six hours, and 115 questions reached back at least a day. That is a benchmark requirement expressed in wall-clock time. It leaves open how efficiently a model could satisfy it and which intervening observations it could safely discard.

The scoring audit establishes a second requirement. An assistant can supply the correct earlier location after an occluded move, yet conditioning its present answer on that same location can turn a verified sighting into an unsupported assertion. Room cameras make that mistake testable. A stronger experiment would compare retrieval, recurrent state and inference weights on identical streams, with one score for identity, another for retained encounters, and separate scores for past and current locations. This would show which memories survive repeated observations, what evidence licenses a correction, and how quickly the system can answer while new events continue to arrive.

References

[1] Yang J and colleagues. EgoLife: Towards Egocentric Life Assistant. CVPR 2025. Sections 3-5 and Tables 1, 2 and 6 describe collection, question construction, retrieval and evaluation. Accessed 2026-11-20.

[2] EgoLife project team. Collection, synchronization and behind-the-scenes account. Authors' account of the Beijing house, recruitment, common-area cameras, synchronization and annotation costs. Accessed 2026-11-20.

[3] Yang J and colleagues. EgoLife supplemental material. Appendices A, B, D and E give authorship, consent procedures, the dataset card and the schedule of activities. Accessed 2026-11-20.

[4] LMMs-Lab. EgoLifeQA annotations for wearer A1 at a fixed dataset revision. Public question and evidence records used for the descriptive analysis. Revision 143fb319be7aa5ae210c936bf4f0f3a86092afb0. Accessed 2026-11-20.

[5] Ego4D consortium. Episodic Memory benchmark specification. Defines the last-seen visual query response track and the separate natural-language and moment-localization tasks. Accessed 2026-11-20.

[6] Gu A, Goel K, Ré C. Efficiently Modeling Long Sequences with Structured State Spaces. ICLR 2022. Primary description of S4 and its structured linear state-space layer. Accessed 2026-11-20.

[7] Gu A, Dao T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. Primary paper defining input-dependent selection within an efficient recurrent sequence model. Accessed 2026-11-20.

[8] Behrouz A, Zhong P, Mirrokni V. Titans: Learning to Memorize at Test Time. Introduces neural associative memory updated at test time and architectures combining that memory with attention. Accessed 2026-11-20.

[9] Behrouz A, Razaviyayn M, Zhong P, Mirrokni V. It's All Connected: A Journey Through Test-Time Memorization, Attentional Bias, Retention, and Online Optimization. The MIRAS framework specifies memory architecture, learning objective, retention and update algorithm. Accessed 2026-11-20.

[10] Liu Z, Lin J, Zhang L, Zhong Y, Li Y. Memory as Weights: Internalizing Long-Term History for Streaming Videos. October 2026 preprint. Sections 3-4 and Appendices C-D describe MaLoW, delayed video queries, update intervals and timing. Accessed 2026-11-20.

[11] Matha R. Programmable Neural Memory: Recognition and Learning in a Continuously Updated Weight Array. Matha Labs preprint 06. Reports addressed symbolic writes, a separate visual recognition arm and snapshot-read timing. Accessed 2026-11-20.

[12] EvolvingLMMs-Lab. EgoLife reference implementation. Released implementations of EgoGPT and EgoRAG and entry points to the published media and model resources. Accessed 2026-11-20.

[13] Schlag I, Irie K, Schmidhuber J. Linear Transformers Are Secretly Fast Weight Programmers. ICML 2021. Derives additive and corrective fast-weight rules for learned key-value mappings. Accessed 2026-11-20.

[14] Matha R. How Much of a Day Do We Remember? Sensory Input, Attention, and the Development of Memory. Matha Labs preprint 07. Discusses differences between sensory availability, attention, cued retrieval and measurement. Accessed 2026-11-20.

[15] Xiong G and colleagues. MemLife: Curating and Reasoning over Long-Term Egocentric Video Memories. September 2026 preprint. Describes entity-grounded text episodes and a reader indexed by time, including use without query-time video. Accessed 2026-11-20.

Reproducibility record

Private review materials: https://github.com/RevanthM/egocentric-memory | Research snapshot: f8f8af705ee62b45478a081876878ebacebc7fe2