Read-Only Identity Memory and the Boundary of a Generative State

Dennis Coleman (VC), Charles Reynolds (KG), Gary Richardson (IS)

Abstract

An identity reference can remain unchanged while a generated subject drifts. This conceptual analysis distinguishes the integrity of stored evidence from the generator's use of that evidence and from the fidelity of the resulting output. ARGUS motivates the question through its read-only identity-memory design. No memory intervention or ablation is performed here. The proposed evaluation records reference access, generation state, and external outcomes separately, so an architectural property of the conditioning path does not become an unsupported guarantee that the subject is preserved across every generated trajectory.

A stable input does not guarantee stable use

Keeping a reference unchanged protects one part of a generation pipeline. It does not establish how strongly later outputs depend on it or whether other conditions override it. A model can have access to a correct observation and still produce a task-incompatible result. The reverse is also possible: an acceptable output can occur despite a flawed internal explanation of how the reference was used. Input integrity, internal processing, and external fidelity are different scientific objects.

ARGUS describes identity evidence organized as a read-only memory within its generation mechanism (Meng et al., 2026b). That motivates examining the boundary between stored reference information and the evolving output state. The present article does not inspect attention patterns, modify memory tokens, or measure the system's trajectory behavior. It offers a conceptual experimental design. Its contribution is to identify what an ablation or intervention would need to hold fixed before the memory's role could be interpreted from observed outputs.

Three distinct properties of a memory mechanism

The first property is reference integrity: whether the stored evidence remains the declared observation. The second is availability: whether the relevant generation stages can access that evidence under the implemented conditions. The third is effective preservation: whether the output satisfies the intended identity requirement. A successful integrity check does not establish availability or preservation. An internal activation pattern likewise needs an argument connecting it to the external consequence before it can be treated as evidence of useful identity retention.

A proposed record would preserve the selected reference objects, their transformation into the memory representation, the model and sampling settings, and the output trajectory. It would distinguish an immutable reference representation from any state that changes during generation. This does not require publishing every internal tensor in every workflow. It requires identifying the objects relevant to the claimed mechanism and retaining enough information to verify which evidence was available for a scored invocation.

A direct comparison could alter access to a nominated reference component while keeping the remaining information and evaluation task matched. Such a study should separate a missing-evidence condition from a corrupted-evidence condition, because they can require different behavior. It should also report whether the modified system remains within its intended operating assumptions. A severe architectural disruption can show sensitivity without isolating the normal causal contribution of a well-functioning memory.

From internal interventions to an external claim

The outcome should concern the task rather than the appearance of internal agreement. If the claim is preservation through a viewpoint change, the evaluation needs generated outputs under that change and an independently defined identity endpoint. The record should also include viewpoint compliance. A model that avoids the requested condition can maintain a familiar appearance without demonstrating the proposed advantage of its reference memory. The architecture's name does not supply the missing observation.

An ablation needs matched training and information budgets where feasible. Removing a memory path can alter parameter count, optimization, or available reference content. A complete-system comparison may still be useful, but an explanation that attributes every difference to read-only storage would require further controls. The report should state which design component the experiment can actually distinguish. It should preserve unsuccessful and unassessable generations rather than let a narrowed output set determine the conclusion.

The temporal scope also matters. A memory can support short sequences while remaining untested over longer histories or competing conditions. A proposed test should vary that scope deliberately and avoid extrapolating an observed local benefit into unlimited subject persistence. Even a continuously accessible reference is not a proof that every required property will appear correctly in every output. A bounded result linked to a declared trajectory population is more interpretable than a universal guarantee derived from an internal storage property.

Read-only records and observable state after interruption

For the proposed comparison, the key observation is reference-memory transaction. A fixed set of reference tiles remains unchanged while the generator enters a long continuation after an occlusion. This hypothetical case poses a question about whether the system can still realize a recognizable subject after the event. Record same immutable memory with ordinary continuation and same memory restored after a controlled interruption separately. A checksum can confirm that the reference stayed unchanged while the generator ceases to attend to it. Pair integrity checks with outputs from a planned replacement or removal intervention before attributing continuity to memory.

An immutable reference record can verify that its contents stayed constant across an interruption. To test use of that memory, rerun the interruption with one clearly diagnostic view available or deliberately withheld. A change in identity at the expected viewpoint would be interpretable only if the intervention left the other observations comparable. Record the input and output at the event boundary and after recovery. If memory contents are observable but the access path is hidden, describe the integrity check and the behavioral test as different sources of evidence.

A result about whether the system can still realize a recognizable subject after the event needs a denominator that includes all eligible attempts, including ones without a usable score. For the reference-memory transaction, keep undetectable faces in the log and state which cases remained eligible for recognition.

Store conditioning views, generated frames and visibility annotations separately for reference-memory transaction. A useful second view of same memory restored after a controlled interruption gives each independent group equal weight and retains the row-weighted result for comparison. When examining same memory restored after a controlled interruption, inspect pairs of frames with similar pose before drawing conclusions about a persistent identity error.

For whether the system can still realize a recognizable subject after the event, define one setting for the primary estimate and show the neighboring settings as a separate comparison. The evaluation of same immutable memory with ordinary continuation can distinguish an incorrect subject from a permitted change in expression, pose or scene. Document the sampling boundary for whether the system can still realize a recognizable subject after the event and reserve a new cohort of cases for a distinct replication.

Systems and comparators

A declared software pipeline can preserve the order of transformations and enable held-out validation without refitting preprocessing on the evaluation observations. Separate facial structure and texture checks can identify whether a convincing local detail conceals changes in global appearance over time. One published study separates global facial structure and local detail in identity-conditioned video (Yuan et al., 2024). In practical terms, an existing method describes reusable software interfaces for fitting and evaluating estimators (Pedregosa et al., 2011).

Varying the image-conditioning strength gives a practical control for the trade between copying a reference and following the requested transformation. Keep reference and evaluation views separate so that recognition after adaptation can be distinguished from successful reproduction of a training portrait. In practical terms, a methodological precedent adapts image synthesis to a specific subject using example images (Ruiz et al., 2022). A methodological precedent adds image prompting to a text-conditioned diffusion model (Ye et al., 2023).

A separable motion module offers a useful distinction between the visual quality of a still frame and consistency under movement. For a prospective comparison, use a text-conditioned video baseline to assess generic motion with the same clip duration and language instructions. One approach uses a diffusion transformer for text-conditioned video generation (Yang et al., 2024). One published study adds a motion module to personalized image-generation systems (Guo et al., 2023).

Fixed spatial controls offer a way to vary prompts while holding a geometric scaffold stable, or to vary the scaffold under common language. First-frame conditioning gives a clear initial anchor; later tests can distinguish ordinary appearance change from drift after a break in visibility. In a controlled test, a methodological precedent adds spatial controls to pretrained text-to-image diffusion (Zhang et al., 2023). One study uses first-frame information to guide spatial and temporal consistency (Ren et al., 2024).

Image-conditioned video provides a direct persistence test. Keep the supplied view and later output segment as separate observations. In an evaluation, a one-reference adapter supplies a baseline for evaluating whether added views help temporal persistence when motion and instructions are held constant. Related work investigates image-to-video generation and video training data curation (Blattmann et al., 2023). Related work personalizes video through a face adapter using one reference (He et al., 2024).

Analysis and interpretation

Matching can improve comparability on recorded attributes; an overlap display should identify which observations are lost before the outcome contrast is computed. Related work allows regression covariance estimates to account for unequal error variances (White, 1980). One study reviews matching as a design step for comparing observational units (Stuart, 2010). Unequal residual variance changes uncertainty even under a common mean model, which matters when a few conditions are much noisier than others.

A family of endpoints should be declared before evaluation so that reported discoveries reflect every comparison eligible to appear in the result. A methodological precedent uses causal diagrams to articulate

assumptions about adjustment and selection (Greenland et al., 1999). In an evaluation, related work controls the false discovery rate for a declared family of hypotheses (Benjamini and Hochberg, 1995). A proposed causal diagram can distinguish decisions made before a run from variables created by the system and observed only after it responds.

Single-reference conditioning provides a budget-matched comparison for identifying which errors disappear when extra views are made available to the generator. In practical terms, a multiple-view conditioning baseline is most informative when the views add distinct pose evidence rather than repeated copies of a similar crop. One approach conditions identity-preserving image synthesis on a single reference (Wang et al., 2024). A relevant contribution uses auxiliary views to condition identity-preserving video (Wu et al., 2026).

Another study builds a stacked identity representation from input portraits (Li et al., 2023a). A missingness analysis distinguishes failures from observations that were never eligible and makes the assumptions behind complete-case summaries visible to readers. For implementation, a methodological precedent separates missing-data assumptions from observed-data summaries (Little and Rubin, 2002). A stacked portrait input permits a test of view coverage, provided the evaluator retains an independent portrait and records near-duplicate input views.

Measurement and error

One approach constructs a full-reference metric with structure and texture components (Ding et al., 2020). A whole-image embedding reflects scene description and background alongside subject appearance. A local check can distinguish these contributions. Separate structure and texture responses are informative when a change in surface detail is permissible but a change in overall shape is consequential. Another study learns aligned visual and language representations through contrastive training (Radford et al., 2021).

One study learns an embedding space for face verification and recognition (Schroff et al., 2015). Another study organizes video-generation evaluation into distinct measurement dimensions (Huang et al., 2023). Verification should include genuine and impostor pairs with similar acquisition conditions to avoid turning image quality into a surrogate for identity. Its separate video dimensions provide a reason to publish a failure profile instead of combining every observation into a single quality index.

Earlier research jointly localizes faces and facial landmarks (Deng et al., 2019). In a controlled test, a comparison of video systems should record differences in training material and processing before attributing a visible change to the architecture alone. Landmark localization provides a way to identify crop failures before measuring resemblance and to retain a denominator for faces that were never detected. A methodological precedent investigates data and training choices for text-to-video synthesis (Chen et al., 2024).

Linking evidence to the evaluation

One approach examines maintenance costs from machine learning dependencies (Sculley et al., 2015). A methodological precedent specifies principles for discoverable and reusable research artifacts (Wilkinson et al., 2016). One study distinguishes forms of intraclass correlation according to the rating design (Koo and Li, 2016). Record inputs, preprocessing and model versions to locate failures caused by dependencies outside the model. A reproducible result requires identifiers for both the released inputs and the exact intermediate tables from which the paper derives its figures. Report which raters, measurements and agreement definition enter the coefficient, since different designs give superficially similar statistics different meanings. Another study asks dataset creators to record collection, intended use and limitations (Gebru et al., 2021). An existing method tests face recognition with a large distractor collection (Kemelmacher-Shlizerman et al., 2015).

Conclusion

Read-only identity evidence is an architectural property whose external value still requires measurement. The proposed design separates integrity, availability, and output preservation, then connects interventions to task-specific consequences under matched conditions. It reports no new result for ARGUS. Its value is to prevent a correct statement about an input memory from silently becoming a stronger, untested statement about the identity behavior of the generated video.

References

Ren, Weiming, et al. "ConsistI2V: Enhancing Visual Consistency for Image-to-Video Generation." *arXiv preprint arXiv:2402.04324*, 2024. https://arxiv.org/abs/2402.04324.

Blattmann, Andreas, et al. "Stable Video Diffusion: Scaling Latent Video Diffusion Models to Large Datasets." *arXiv preprint arXiv:2311.15127*, 2023. https://arxiv.org/abs/2311.15127.

He, Xuanhua, et al. "ID-Animator: Zero-Shot Identity-Preserving Human Video Generation." *arXiv preprint arXiv:2404.15275*, 2024. https://arxiv.org/abs/2404.15275.

Sculley, D., et al. "Hidden Technical Debt in Machine Learning Systems." *Advances in Neural Information Processing Systems*, vol. 28, 2015. https://papers.nips.cc/paper/2015/hash/86df7dcfd896fcaf2674f757a2463eba-Abstract.html

Yuan, Shenghai, et al. "Identity-Preserving Text-to-Video Generation by Frequency Decomposition." *arXiv preprint arXiv:2411.17440*, 2024. https://arxiv.org/abs/2411.17440.

Pedregosa, Fabian, et al. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research*, vol. 12, 2011, pp. 2825-2830. https://jmlr.org/papers/v12/pedregosa11a.html.

Wu, Mingyang, et al. "ConsID-Gen: View-Consistent and Identity-Preserving Image-to-Video Generation." *arXiv preprint arXiv:2602.10113*, 2026. https://arxiv.org/abs/2602.10113.

Wang, Qixun, et al. "InstantID: Zero-shot Identity-Preserving Generation in Seconds." *arXiv preprint arXiv:2401.07519*, 2024. https://arxiv.org/abs/2401.07519.

Schroff, Florian, et al.. "FaceNet: A Unified Embedding for Face Recognition and Clustering" *arXiv preprint arXiv:1503.03832*, 2015. https://arxiv.org/abs/1503.03832.

Little, Roderick J. A., and Donald B. Rubin. *Statistical Analysis with Missing Data*. Wiley, 2002. https://doi.org/10.1002/9781119013563.

Wilkinson, Mark D., et al. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data*, vol. 3, 2016, article 160018. https://doi.org/10.1038/sdata.2016.18.

Kemelmacher-Shlizerman, Ira, et al.. "The MegaFace Benchmark: 1 Million Faces for Recognition at Scale" *arXiv preprint arXiv:1512.00596*, 2015. https://arxiv.org/abs/1512.00596.

Ye, Hu, et al.. "IP-Adapter: Text Compatible Image Prompt Adapter for Text-to-Image Diffusion Models" *arXiv preprint arXiv:2308.06721*, 2023. https://arxiv.org/abs/2308.06721.

Li, Zhen, et al. "PhotoMaker: Customizing Realistic Human Photos via Stacked ID Embedding." *arXiv preprint arXiv:2312.04461*, 2023. https://arxiv.org/abs/2312.04461.

Ruiz, Nataniel, et al.. "DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation" *arXiv preprint arXiv:2208.12242*, 2022. https://arxiv.org/abs/2208.12242.

White, Halbert. “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.” *Econometrica*, vol. 48, no. 4, 1980, pp. 817-838. https://doi.org/10.2307/1912934.

Chen, Haoxin, et al. "VideoCrafter2: Overcoming Data Limitations for High-Quality Video Diffusion Models." *arXiv preprint arXiv:2401.09047*, 2024. https://arxiv.org/abs/2401.09047.

Yang, Zhuoyi, et al. "CogVideoX: Text-to-Video Diffusion Models with An Expert Transformer." *arXiv preprint arXiv:2408.06072*, 2024. https://arxiv.org/abs/2408.06072.

Benjamini, Yoav and Yosef Hochberg. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 57, no. 1, 1995, pp. 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Gebru, Timnit, et al.. "Datasheets for datasets" *Communications of the ACM*, 2021. https://doi.org/10.1145/3458723.

Deng, Jiankang, et al. "RetinaFace: Single-stage Dense Face Localisation in the Wild." *arXiv preprint arXiv:1905.00641*, 2019. https://arxiv.org/abs/1905.00641.

Guo, Yuwei, et al. "AnimateDiff: Animate Your Personalized Text-to-Image Diffusion Models without Specific Tuning." *arXiv preprint arXiv:2307.04725*, 2023. https://arxiv.org/abs/2307.04725.

Koo, Terry K., and Mae Y. Li. "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research." *Journal of Chiropractic Medicine*, vol. 15, no. 2, 2016, pp. 155-163. https://doi.org/10.1016/j.jcm.2016.02.012.

Greenland, Sander, Judea Pearl, and James M. Robins. “Causal Diagrams for Epidemiologic Research.” *Epidemiology*, vol. 10, no. 1, 1999, pp. 37-48. https://doi.org/10.1097/00001648-199901000-00008.

Stuart, Elizabeth A. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science*, vol. 25, no. 1, 2010, pp. 1-21. https://doi.org/10.1214/09-sts313.

Meng, Zijie, et al. "Argus: Stacked multi-view identity mosaic injection for subject-preserving video generation." *arXiv preprint arXiv:2606.11670* (2026).

Ding, Keyan, et al. "Image Quality Assessment: Unifying Structure and Texture Similarity." *arXiv preprint arXiv:2004.07728*, 2020. https://arxiv.org/abs/2004.07728.

Zhang, Lvmin, et al. "Adding Conditional Control to Text-to-Image Diffusion Models." *arXiv preprint arXiv:2302.05543*, 2023. https://arxiv.org/abs/2302.05543.

Radford, Alec, et al.. "Learning Transferable Visual Models From Natural Language Supervision" *arXiv preprint arXiv:2103.00020*, 2021. https://arxiv.org/abs/2103.00020.

Huang, Ziqi, et al.. "VBench: Comprehensive Benchmark Suite for Video Generative Models" *arXiv preprint arXiv:2311.17982*, 2023. https://arxiv.org/abs/2311.17982.