

# OLM Pipeline: Color Grounding Experiment Report

Generated: 2025-10-14 19:00:10

## Project Context

**OLM Pipeline - Frozen VAE Latent Dynamics** 1.0.0

First functioning model on the OLM (Object-Level Manipulation) pipeline that turns raw video into predictive latent state using a frozen VAE and three small LSTMs. This represents a major stepping stone towards completing the OLM project.

## Architecture Summary

- Frozen VAE for stable frame $\leftrightarrow$ latent mapping
- PatternLSTM  $\rightarrow$  CompressionLSTM  $\rightarrow$  CentralLSTM stack
- CentralLSTM predicts  $\Delta z$  for next-step latent
- Color-token path injects text-encoded color into CentralLSTM to test grounding

## Experimental Procedure

- Training: paired solid-color frames with corresponding color tokens; backprop on next-latent prediction.
- Evaluation (Color Eval Mode): feed token only; no image input; decode predicted latent.
- Metrics: latent MSE, cosine similarity, L2 distance; optional PSNR/SSIM for decoded frames.

No eval\_results.jsonl found under ai\_logs/. Add a pass in Color Eval Mode to populate metrics.

## Implementation Notes

- Color capture and eval toggles are exposed in the GUI. The pipeline pre-encodes the hex string and caches the solid-color latent for comparison.
- The Central LSTM accepts a text-encoded override to test text $\rightarrow$ latent grounding without image input.
- Eval runs append JSONL summaries to ai\_logs/eval\_results.jsonl for offline analysis.

## Next Steps

- Introduce a second color token and alternate batches 50/50; verify minimal cross-color interference.
- Track cosine per-color and latent overlap; target  $\text{Cos} \geq 0.98$  for both with  $<15\%$  overlap.
- Then extend to shapes and simple motion to begin temporal compositionality.

End of report.