

Ray3 Evaluation Report – State-of-the-Art Performance for Pro Video Generation

An inside look at how Ray3 turns imagination into motion through realism, control, and creative fidelity

By Sabrina Day and Simran Motwani - Luma AI

October 2025

Contents

1	Introduction	1
2	Methodology - Performance Evaluations	2
3	Ray3 in Practice: Realism, Control, and Creative Fidelity	3
3.1	Physics and Motion Performance	3
3.2	Instruction-Following	4
3.3	Motion Artifacts	5
3.4	Aesthetic Quality	6
3.5	Dynamic Range and Color Reproduction	7
3.6	Temporal Consistency	8
3.7	Image-to-Video Adherence	9
4	Ray3 – The Storytelling Model	10

1 Introduction

The launch of Ray3 marks a new chapter for generative video—one where creators, filmmakers, and visionaries can direct stories with the nuance and realism once reserved for live-action production and use generative video for final pixel production. Built to translate intent into cinematic motion with unprecedented control and fidelity, Ray3 captures the subtleties of human expression, physics, and atmosphere that give life its richness. Independent evaluations place it ahead across nearly every dimension of performance, setting a new benchmark for realism and creative control (see Figure 1).

At the center of every model launch at Luma lies a cycle of iterative performance evaluation. We follow a rigorous evaluation process designed to measure how the

model performs and to surface its most critical opportunities for improvement.

Here, we outline how we evaluate Ray3—for filmmakers, designers, and storytellers who want to understand not only what it can do, but how we measure what makes it great.

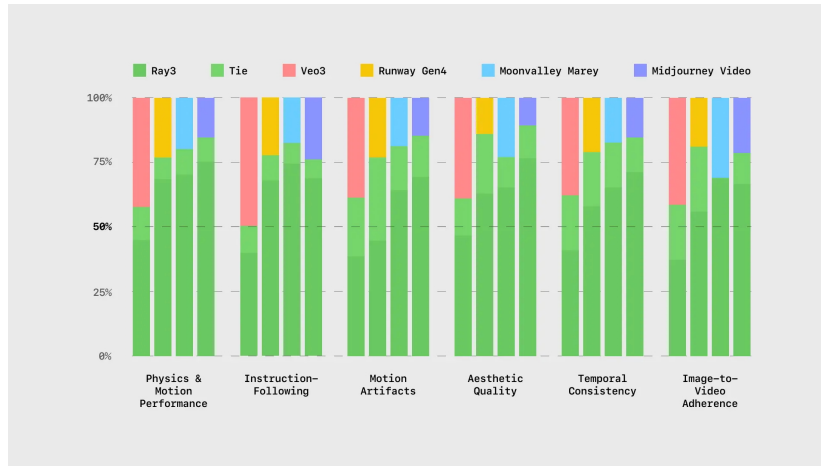


Figure 1: Aggregate performance across evaluation dimensions, as measured through independent review.

2 Methodology - Performance Evaluations

Our evaluation framework measures performance across many axes—from motion accuracy to the subtleties of skin texture and identity preservation. Unlike most existing evaluation methods for video generation—which remain under-specified, relying on frame-level metrics that overlook temporal coherence, physics, and identity consistency—our framework was built from the ground up to capture the full perceptual and physical realism of motion. It is the product of cross-disciplinary collaboration between scientists, filmmakers, artists, and creators, grounded in recurring flaws observed across generative models and validated through independent review to ensure rigor.

To assess a world model’s performance is to measure how faithfully it reproduces the world and physics imagined by the user. It operates a level above prompt understanding—evaluating how well the model situates the subject within a consistent and physically plausible world. Some flaws are overt, as when a body distorts or skin takes on an artificial sheen; other times they are subtle, in the way fabric folds, water behaves, or causality breaks. The weight of each flaw depends on its frequency, severity, and salience within the generated world. In developing Ray3, we measured what is hardest to quantify—but most essential to achieving realism.

We begin by breaking down and understanding the high-dimensional space of world

models—working in close collaboration with creatives, physicists, storytellers, and technicians. Our evaluation process combines structured internal analysis with independent review. Each model checkpoint is tested across thousands of generated videos spanning diverse prompts, styles, and levels of difficulty. Every output is rated by trained evaluators with diverse domain expertise using standardized success metrics.

This dual-layered approach—qualitative precision paired with external validation—ensures that Ray3’s progress is measured objectively and that improvements and regressions can be understood and traced.

3 Ray3 in Practice: Realism, Control, and Creative Fidelity

Third-party quantitative evaluations show Ray3 reaching state-of-the-art performance alongside Veo3, while maintaining a significant and definitive lead over every other industry model— Midjourney Video, Runway Gen-4, and Moonvalley Marey.

At the core of our evaluation metrics are these aspects critical to creative workflows: physics and motion performance, instruction-following, motion artifacts, aesthetic quality, temporal consistency, and image consistency in reference-based workflows.

A key driver of Ray3’s performance across these axes is its internal reasoning system—its ability to interpret prompts, infer intent, and plan how scenes should unfold before rendering. By modeling spatial and temporal relationships, Ray3 maintains coherence across motion, physics, and emotion. This structured reasoning underpins its realism, making each sequence feel intentional and true to creative direction.

From evaluating Ray3 against these measures, we observe the following patterns:

3.1 Physics and Motion Performance

Across both internal and external evaluations, Ray3 demonstrates exceptional physical and motion realism. Bodies move with coherent mass and inertia; gestures exhibit believable force and follow-through; and environments respond according to consistent physical laws. From fabric settling under gravity to light diffusing through motion, Ray3 sustains realism even under complex dynamic conditions.

Motion in Ray3 is not only accurate but expressive—transforming generated shots into physically grounded, emotionally intelligent movement. Third-party quantitative evaluations place Ray3 ahead of every competitor establishing physical realism as one of Ray3’s defining strengths (see Figure 2).



Figure 2: Results from independent evaluations of motion realism and physical coherence.

3.2 Instruction-Following

Ray3 follows instructions with SOTA precision while capturing the overall intent, and aligning details, actions, and atmosphere with what is asked for. For filmmakers, designers, and storytellers, this means prompts translate into scenes that behave as imagined: performances match direction, emotional tone aligns with the script, and environments move in harmony with the story’s intent. Ray3 adheres to the prompt while leaving space for creative interpretation.

Third-party evaluations place Ray3 ahead of models such as Runway’s Gen-4, Moonvalley, and Midjourney, and matching Veo3, whose outputs generally favor precision over creative expression (see Figure 3).

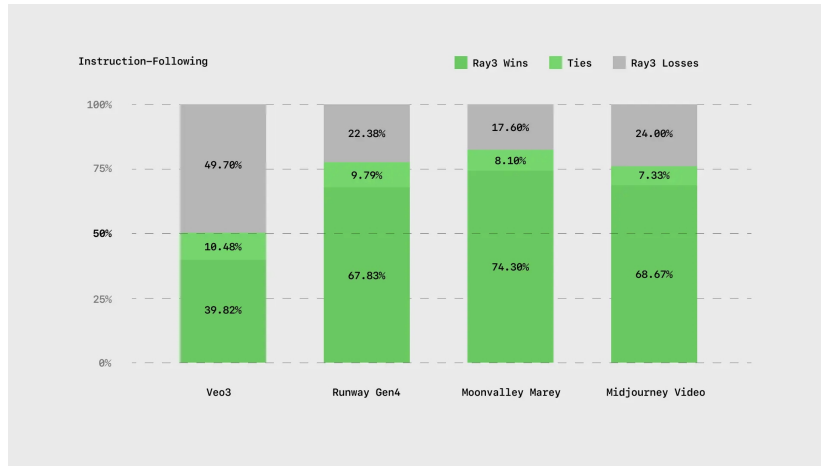


Figure 3: Results from independent evaluations of prompt adherence and intent alignment.

3.3 Motion Artifacts

Motion artifacts are the most egregious factors that break immersion. Distorted limbs, jittering textures, or unnatural blurs remind the viewer that shots are generated. In evaluations, Ray3 exhibits these fractures far less often—and far less severely than any other video model. Movement holds organically, gestures stay natural, and environments remain coherent, allowing attention to stay on the story instead of the seams of generation.

Third-party quantitative results reflect the same trend, placing Ray3 ahead of every other model—a crucial strength for creators aiming to maintain immersion and believability at every frame (see Figure 4).

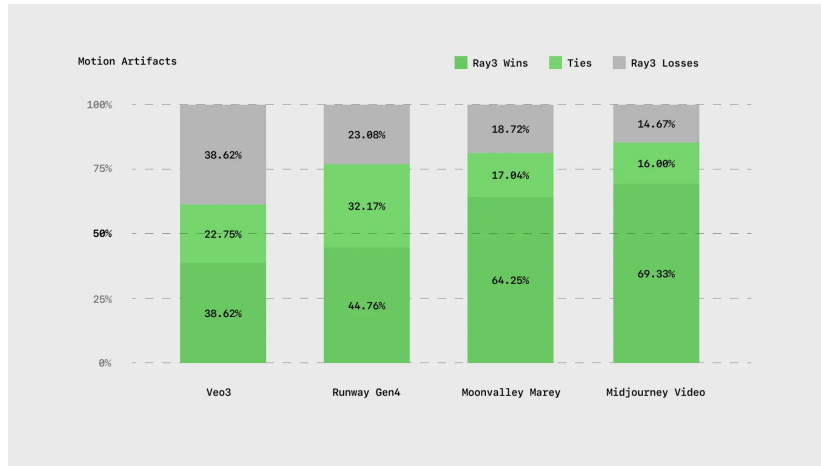


Figure 4: Results from independent evaluations of visual artifact frequency.

3.4 Aesthetic Quality

Aesthetics is the measure of how composition, light, and color converge to create visual coherence, mood, and a sense of deliberate intent. In evaluating Ray3, we found it sustains aesthetic quality throughout entire sequences, carrying atmosphere and tone with consistency rather than letting clarity degrade over time. Ray3 outshines every other model in pixel-level detail and fidelity, a result of both its scale and architecture.

Ray3 holds clarity, cohesion, and atmosphere from the first frame to the last. Third-party quantitative evaluations confirm this lead in performance, placing Ray3 ahead of every competitor in visual quality (see Figure 5)—proving that realism alone is not enough—it must also be beautiful.

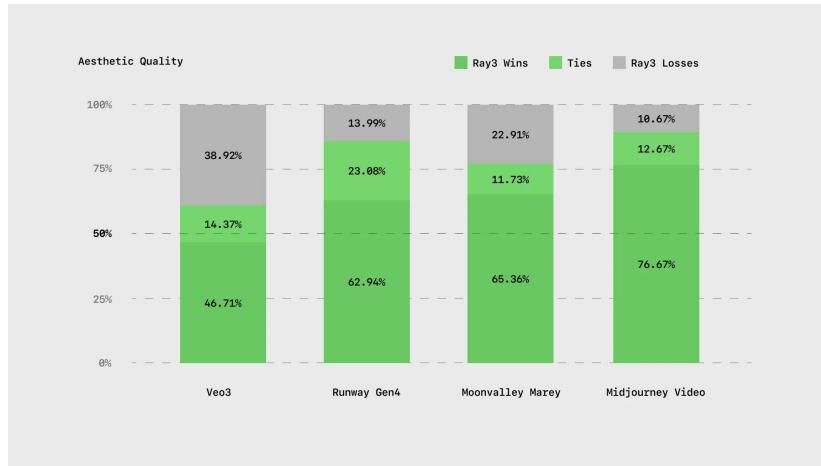


Figure 5: Results from independent evaluations of visual quality, lighting, and composition.

3.5 Dynamic Range and Color Reproduction

Ray3 stands alone as the only generative video model capable of producing true HDR—rendered natively in 10, 12, and 16-bit high dynamic range (see Figure 6). This breakthrough brings unprecedented control and fidelity to creative production pipelines.

HDR is a foundational capability in Ray3. Each frame carries 16bit dynamic range, preserving detail in both shadows and highlights with vivid, accurate color. The result is cinematic contrast and lifelike texture that mirror high-end camera systems used in professional production.

For the first time, creators can generate and export HDR video directly from a generative model—ready for high-end finishing and color workflows. Ray3’s HDR outputs integrate seamlessly into production environments through EXR export, enabling precision control in post-production color grading, compositing, and VFX pipelines.

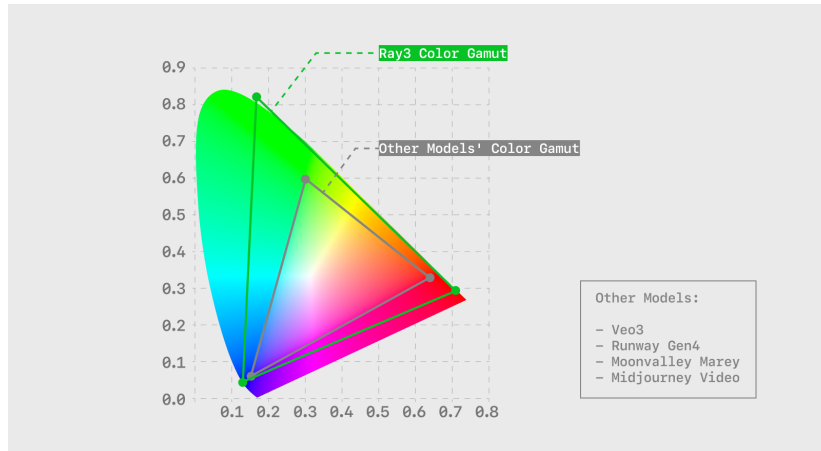


Figure 6: Comparison between Ray3’s wide color gamut versus other models’ standard color gamut.

3.6 Temporal Consistency

A video’s success lies not in a single frame, but in how seamlessly one frame transitions into the next to create a coherent story. Temporal consistency measures the stability and continuity of narrative—how people, objects, and environments evolve through time with coherent form, proportion, and behavior.

Evaluations of Ray3 show that this stability is one of the defining strengths of the model. Faces remain consistent, bodies retain proportion, and gestures progress with smooth, believable motion, creating the impression of genuine time unfolding rather than frames being stitched together. This consistency supports longer, more expressive sequences, allowing emotion, motion, and narrative to develop with cohesion and intent.

Independent evaluations confirm this strength, with Ray3 achieving the highest temporal consistency scores across all measured competitors (see Figure 7)—setting a new benchmark for cinematic flow and visual continuity.

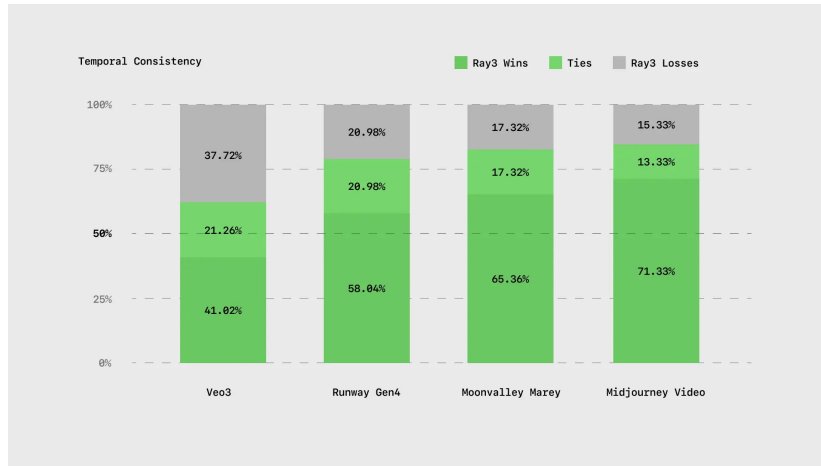


Figure 7: Results from independent evaluations of stability and continuity across frames.

3.7 Image-to-Video Adherence

Reference-based workflows only

Image reference is a dominant workflow in professional generative video pipelines. Image-to-Video Adherence measures how faithfully a generated video maintains alignment with its reference image or start frame—preserving identity, layout, color palette, style, and structure.

In evaluations, Ray3 demonstrates exceptional reference fidelity (see Figure 8). It sustains character identity, spatial layout, and style with high precision, capturing the visual and logical essence of the source image. Faces remain much more recognizable, environments stay 3D consistent, original color grading retains throughout the shot much more consistently, and the model is able to predict correct motion from just an image.

Rather than drifting or reinterpreting source material, Ray3 preserves the creative baseline—ensuring that the final sequence reflects the original composition and design intent. Independent evaluations place Ray3 at near parity with SOTA in this category, while maintaining a decisive lead over all other models.

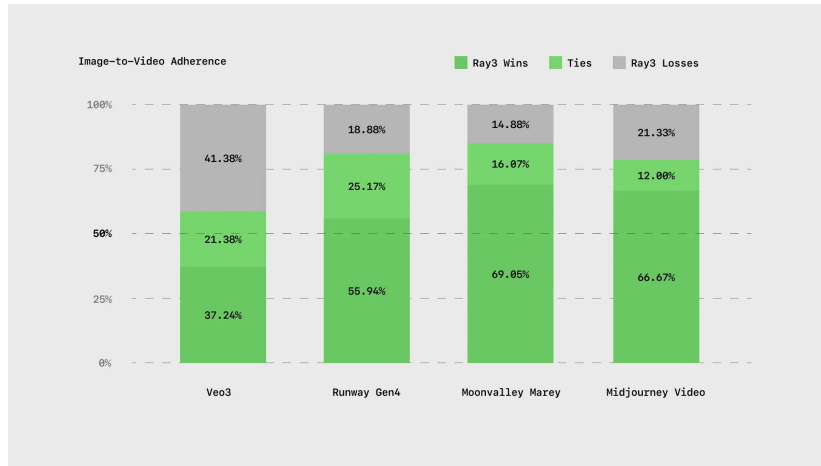


Figure 8: Results from independent evaluations of reference image fidelity and style preservation.

4 Ray3 – The Storytelling Model

Evaluating Ray3 reveals more than technical progress—it reflects a model advancing toward perceptual realism and new narrative capabilities. Ray3 shows measurable gains in its capacity to represent the physics and emotions of the real world. These evaluations highlight not only where the model performs well, but how it aligns its outputs with creative intent—responding to direction with precision while preserving flexibility for interpretation.

The outcome extends beyond fidelity. Ray3 represents a shift toward models that understand and execute intent—systems capable of reasoning about narrative structure, spatial relationships, and emotional tone. This evolution marks an early step toward generative models that can be assessed not merely by their technical accuracy, but by their ability to produce coherent, believable, and expressive visual storytelling experiences.

Try it in Dream Machine - <https://dream-machine.lumalabs.ai/>

Reach out to the Luma team to deploy Ray3 in your organization - <https://lumalabs.ai/contact-sales>