**Getting Started**

Odyssey-2 Pro Overview   ⧉ Copy page   ⌄

A frontier world model that brings video to life.

At its core, Odyssey-2 Pro is an action-conditioned world model. Given the current state, an incoming action, and a history of states and actions, the model predicts the next state in the form of a video frame. You interact with Odyssey-2 Pro much like a language model: type a few words and it begins streaming minutes of imagined video to any screen or device. Type a few more words, and the video adapts to your new request in that moment.
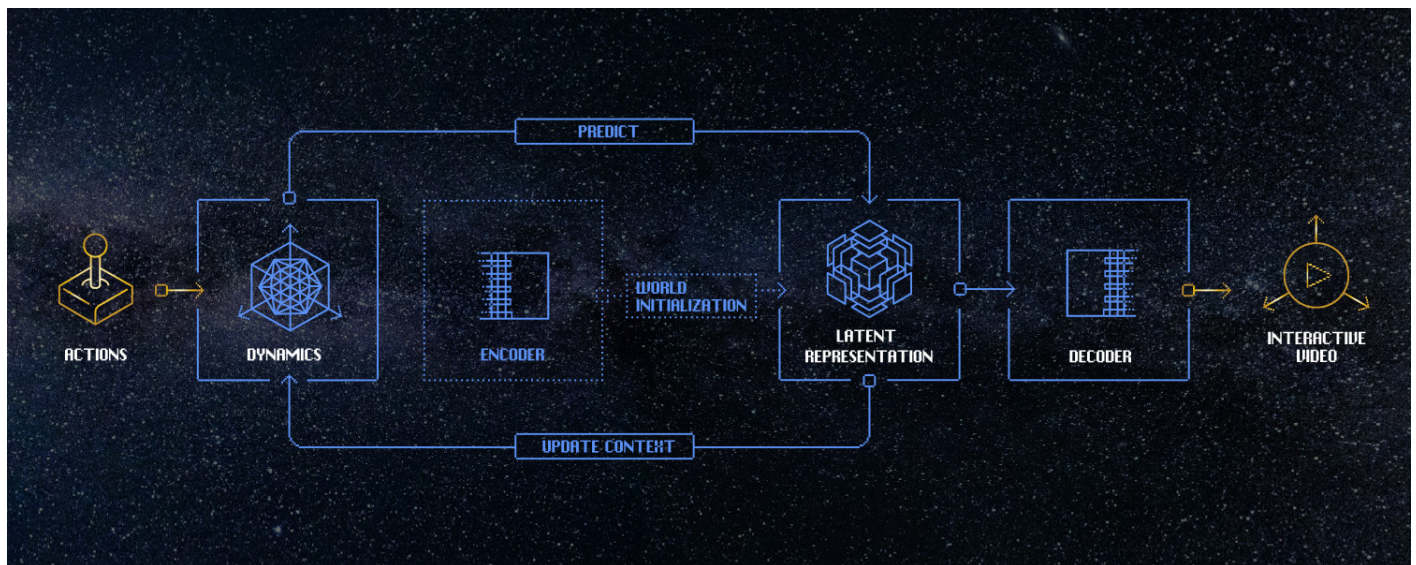
A world model, not a video model

Traditional bidirectional video models today take 1-2 minutes to generate only 5 seconds of footage. Odyssey-2 Pro, on the other hand, begins generating and streaming video instantly —producing a new frame of video every 50 milliseconds. As Odyssey-2 Pro's video streams,

you can shape it in real time with additional inputs. The result is a continuous stream of video that listens, adapts, and reacts.

World Models	Video Models
Predicts one frame at a time, reacting to what happens.	Generates a full video in one go.
Every future is possible.	The model knows the end from the start.
Fully interactive—responds instantly to user input at any time.	No interactivity—the clip plays out the same every time.

A face painted

0:00 / 0:31

A painting come to life

0:00 / 0:24

1   **Prompt**

A close-up portrait of a woman, illuminated by soft, directional lighting. The background is softly blurred. The camera remains steady.

2   **Midstream**

The woman paints her face with green paint.

1   **Prompt**

A close shot of a painter using oil paint to paint a fireplace hearth on canvas.

2   **Midstream**

The flames painted on the canvas come to life. The canvas is on fire.

World models are burgeoning world simulators

World models predict the next frame using only the past, and then roll forward. Long rollouts punish bad guesses, so the model must internalize the dynamics of how motion, lighting, and contact evolve over time. In doing so, they become implicit world simulators—systems that learn to model and generate the world in real time.

Waves on the ocean are a great example. From prior video frames of a wave, you can infer surface slope, curvature, and a velocity field. With that, you can predict what comes next: the crest advances, troughs fill, foam drifts, highlights slide, and the wave bends around a rock. Although early on this journey, that's exactly what Odyssey-2 Pro does, all learned from decades of video data. This same concept applies generally to dynamics and behaviors of many types.

Language models showed how far a simple next-step objective can go—predicting the next word unlocked reasoning and creativity. models take that idea further. By learning to predict the next frame of video, they begin to approximate the rules that govern our world. As these models learn to act and react to the highest levels of realism, they will transition from enabling emergent media to a general-purpose world simulator.

‹ **Introducing the Odyssey API**   **API Quick Start** ›

Powered by **mintlify**