

A New Test Ranks 14 Video Models on Keeping Hidden Objects in Mind

Kabui, Charles

2026-10-28

 Read at ToKnow.ai

A New Test Ranks 14 Video Models on Keeping Hidden Objects in Mind

A 150-scene test, and a 16B model trained on 1.5 million clips

1.5M

clips from 150 Blender scenes

220+

Elo above the next continuation

320x192

video size against 720p rivals

October 28, 2026

ToKnow.ai

A 31-author team built WROP, a test of whether video models keep track of an object while something hides it, as babies do by 3.5 months. It has 150 hand-built Blender scenes in six groups. Three hide objects behind screens, inside tunnels or under lids; three test solidity, like a car stopped by a barrier. Each scene randomises lighting, camera angle and speed, which yields 1.5 million clips and a 300-question exam. Every clip is cut at the physical event: the

model sees 2.5 seconds and generates the next 2.5. Twenty people judged 14 systems in 361 blind pairs.

Wan 3.0 Prime and MiniMax H3 took the top two places, tied at 1723.6 on an Elo rating (a head-to-head voting score). Both rebuild the scene from a reference clip rather than extend it, so they can drift from what the clip showed. The best model that truly continues a clip was the team's own 16B fine-tune, PWM-WROP, built on [Cosmos 3 Nano](#) and trained for one pass over those clips. It scored 1679.5, more than 220 Elo above Grok Imagine's video extend. No untrained Cosmos 3 Nano entered, so training alone cannot take the credit.

The fine-tune placed first when a moving screen covers still objects and eighth when objects collide, and it outputs 320x192 video where rivals emit 720p. The paper says how a model takes its input explains the ranking better than model size, and calls the result preliminary. Clips, exam and weights are for non-commercial use.

Read More: [Trinity of Consistency](#), a survey that measures where world models break across modal, spatial and temporal consistency.

Sources:

- [Training Object Permanence in World Models \(arXiv abstract\)](#)
- [Full paper with result tables \(arXiv HTML\)](#)
- [Human-preference leaderboard for the 300-question exam](#)
- [WROP training corpus on Hugging Face](#)
- [PWM-WROP model weights on Hugging Face](#)

***Disclaimer:** For information only. Accuracy or completeness not guaranteed. Illegal use prohibited. Not professional advice or solicitation. **Read more:** [/terms-of-service](#)*